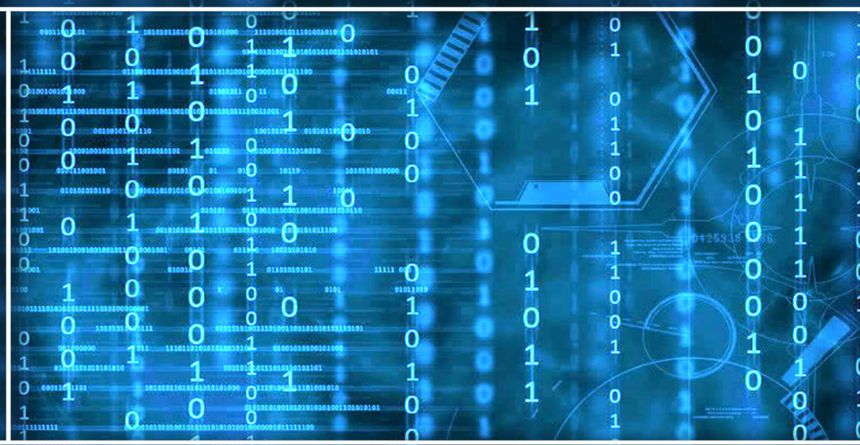
IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 13 Issue 12

December 2022



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)

SAI

www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

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

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

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

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

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

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

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 13 Issue 12 December 2022
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

CONTENTS

A Two-Step Approach to Weighted Bipartite Link Recommendations

Nathan Ma
Adlai E. Stevenson High School Lincolnshire, Illinois, 60069, USA

***Abstract*—Many real-world person-person or person-product relationships can be modeled graphically. Specifically, bipartite graphs are especially useful when modeling scenarios involving two disjoint groups. As a result, existing papers have utilized bipartite graphs to address the classical link recommendation problem. Applying the principle of bipartite graphs, this research presents a modified approach to this problem which employs a two-step algorithm for making recommendations that accounts for the frequency and similarity between common edges. Implemented in Python, the new approach was tested using bipartite data from Epinions and Movielens data sources. The findings showed that it improved the baseline results, performing within an estimated error of 14 per cent. This two-step algorithm produced promising findings, and can be refined to generate recommendations with even greater accuracy.**

***Keywords*—Bipartite graph; weighted graph; link prediction; two-step algorithm; information retrieval**

I. Introduction

Bipartite graphs [1] are graphs that can be split into two sets of vertices such that, within each set, there are no edges. Edges only exist between the vertices of opposite sets. See Fig. 1, for an example of a bipartite graph. Within a weighted bipartite graph, each edge is also given a weight. Weighted bipartite graphs [2] can be used to model a multitude of real-world relationships such as the product-customer relationship. In this scenario, the weighted edges represent how a customer has rated a product.

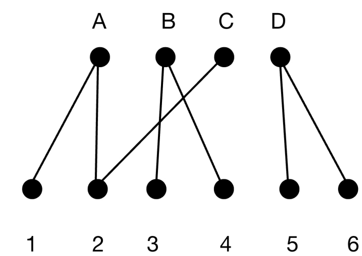


Fig. 1. Example of an unweighted bipartite graph

This research focused on examining the link prediction problem in the context of weighted bipartite graphs. Though not solely limited to weighted bipartite graphs, the link prediction problem [3] takes in an arbitrary graph network and predicts the possibility of an edge between unpaired vertices. When applied to weighted bipartite graphs, the link prediction problem predicts the weight between unpaired vertices based on the input graph network.

Weighted link prediction has broad applications. For example, it is sometimes analogous to forecasting how customers will rate products based on a network of customer-product ratings. Many platforms, such as Amazon or YouTube, likely employ some form of weighted link prediction algorithm when recommending items to users. The applicability of link prediction can be extended beyond user-item recommendations to image recommendation systems as well [4].

In this paper, we propose an algorithm that performs weighted link recommendation in two steps, which differs from existing multi-step algorithms that utilize techniques derived from the predominant Graph Convolution Networks [5] and Graph Neural Networks [6]. In our two-step approach, the confidence in a potential link occurring between two vertices based on common neighbors was first evaluated. Then, weights were determined based on the information derived from these common neighbors. This algorithm was finally tested on existing datasets that could be modeled using weighted bipartite graphs.

II. Related Work

Given numerous real-world applications, prior papers have examined different approaches to link prediction using either weighted [7] or unweighted [8] bipartite graphs.

One approach to unweighted link prediction is to specifically predict internal links [9]. Internal links are defined as an unpaired top and bottom node in which a hypothetical edge between the nodes does not change the bipartite projected graph. Only internal links may be considered for potential edges because being "internally linked" means that two vertices already share some degree of common neighbors. In turn, this indicates that the two vertices are more likely to have an edge develop between them. Whether an internal link would become an actual edge may be determined from several proposed weighting functions that account for factors such as common neighbors and the overall number of neighbors. (See Fig. 2, for an example of an internal link.)

One idea used in weighted link prediction is the Pearson coefficient [10]. The Pearson coefficient does not predict edges, but determines a similarity between two nodes within the same disjoint bipartite set. Similarity is determined by first finding the set of common neighbors between the two nodes. For each of the two nodes, the weight of the edge with the common neighbor is then subtracted from the average weight of all edges extending from that node. These resulting values are multiplied, and this process is repeated for all common neighbors. A positive final value corresponds to a positive similarity and vice versa.

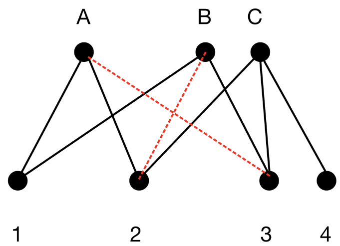


Fig. 2. Example of internal links within a bipartite graph (dotted red edges)

It is also important to note the difference between link prediction and link recommendation. Whereas link prediction involves finding any arbitrary connections that will occur in the future, link recommendation is more specific to the weighted bipartite graph. Instead of dealing with arbitrary connections, link recommendation identifies the most relevant connections between unpaired top and bottom nodes and calculates a weight for a hypothetical edge.

A classical approach specific to link recommendation has been collaborative filtering [11]. The principle underlying collaborative filtering is very similar to that of the Pearson coefficient; if Person A and B rate a product similarly, then Person A should rate another product more similarly to Person B than a randomly chosen person. Thus, collaborative filtering leverages the power of people with mutual interests to generate recommendations.

Collaborative filtering may be applied in many different ways such as: the use of neural networks [12] to establish deeper analysis between the user-user and user-item connections, hybrids between memory and matrix-based ideas [13], and the incorporation of time as a relevant factor in the decision making process [14] [15]. Though collaborative filtering is not the only approach to link recommendation [16], the role of similarity nevertheless plays a vital role.

Traditionally, key factors that affect recommendation algorithms include the size and density of the dataset [17]. This paper builds upon the previous work of both past link prediction and recommendation models to create an algorithm suited for recommendation.

III. Methodology

This recommendation algorithm is comprised of two steps. First, pairs of currently nonadjacent top and bottom nodes were identified for making a prediction. Second, predictions were made after selecting the pairs. Notably, a recommendation could not be made in the case of insufficient data connecting a top node and a bottom node.

Intuitively, the top nodes can be thought of as products and bottom nodes as customers. For ease of understanding, the terms "bottom node" or "customer" and "top node" or "product" are used interchangeably in this paper. Thus, the weights between the top and bottom nodes represent the ratings a customer has assigned to a product.

For each candidate pairing, let t denote the number of neighbors of the original top node, excluding the original bottom node. Next, for each of these neighbors, let n define the number of neighbors that have at least one common top neighbor with the original bottom node. Interpreting $\frac{n}{t}$, a value close to 1 indicates that most of the customers who already purchased the candidate product have purchased items the candidate customer has already purchased. The inverse is also true. Thus, a threshold value for $\frac{n}{t}$ was defined such that all values below were classified as "insufficient" for recommendation, and all values above were classified as "sufficient" for recommendation.

This threshold is expected to vary for every dataset. Generally speaking, the threshold should be lower in datasets with a sparser distribution of edges and greater in more compact datasets. Additionally, datasets with many edges may be "sparse" if they also have many nodes.

Thus, the threshold can be given by the equation $\frac{9}{10} - \frac{4}{x+y}$, where x represents the average amount of edges for a bottom node, and y represents the average amount of edges for a top node.

This threshold was set to have a maximum value of $\frac{9}{10}$. Thus, for larger values of $x+y$, the candidate bottom node must have a common neighbor with at least 90% of candidate-top adjacent bottom nodes. The term $-\frac{4}{x+y}$ serves to compensate for smaller datasets with fewer existing edges. The constant 4 was specifically chosen to accommodate smaller datasets such as the Epinions dataset. After performing some initial testing, the constant values 1, 2, and 3 all returned few or no recommendations because the resultant threshold was too high. The constant value 4 was the first number that consistently produced recommendations.

However, this does not prevent cases where the product and customer have sufficient similarity, but have insufficient data to ascertain that level of sufficiency. For example, if the candidate top node has only one bottom neighbor besides the candidate bottom node, and the candidate bottom node has a common top neighbor with the one candidate top-adjacent bottom node, then $\frac{n}{t} = 1$. This will always yield a value above the maximum threshold set; however, this cannot be definitively determined based on a single data point. Therefore, we also required that the original bottom node and bottom nodes adjacent to the original top node have a number of common neighbors at least equal to the average number of edges of a top node.

Once the unpaired nodes with insufficient data were filtered out, we predicted the expected weight between the candidate top node and the candidate bottom node. This was achieved by returning to the bottom nodes that were neighbors of the candidate top node. For the sake of clarity, these bottom nodes were labeled as candidate top adjacent. We defined the *similarity* of one of these bottom nodes and the candidate bottom node by how closely the common neighbors were rated. In other words, the *similarity* represented how closely the preferences of two customers matched based on the products both had bought. Naturally, a higher similarity means that two customers have similar preferences, which in turn, affects the prediction positively.

Similarity was determined by comparing all products rated by both customers. For each common product, the algorithm computed the difference between the ratings assigned by the

$$\begin{cases} (min(r_1, r_2) - a) \cdot (2 - |r_2 - r_1|) & r_1 \geq a, r_2 \geq a \\ (a - max(r_2, r_1)) \cdot (2 - |r_2 - r_1|) & r_1 \leq a, r_2 \leq a \\ 2 - |r_2 - r_1| & r_1 \geq a, r_2 \leq a \\ 2 - |r_2 - r_1| & r_1 \leq a, r_2 \geq a \end{cases}$$

Fig. 3. Piecewise function for temporary similarity

candidate bottom and the overall average rating assigned by all of its customers. The process was repeated for the other bottom node. The differences were compared, and a temporary similarity for each common top node between the candidate top adjacent bottom node and the candidate bottom node was calculated. This was based on the absolute difference and whether the signs were the same. The piecewise function in Fig. 3 depicts the exact method employed for calculating temporary similarity (r_1 and r_2 denote the two differences; a denotes the average rating of the common top node).

After iterating through each of the common products, the temporary similarity values were averaged to determine the overall similarity between the candidate bottom node and the candidate top-neighboring bottom node.

For each bottom node neighboring the candidate top node, the algorithm computed similarity with the candidate bottom node using the aforementioned algorithm. The predicted rating p was determined by the following formula:

k represents the number of top-adjacent bottom nodes that share at least one common neighbor with the candidate bottom node. s_i represents the similarity score for the ith product-adjacent bottom node. r_i is the analogue for rating.

$$p = \frac{\Sigma_{n=1}^{k} s_i \cdot r_i}{\Sigma_{n=1}^{k} s_i}$$

IV. Results

The algorithm was implemented in Python and tested using two datasets, which were obtained from the Epinions website [18] and the MovieLens website [19], respectively. Both datasets can be modeled using weighted bipartite graphs with weights ranging between 1 and 5. Users in the Epinions dataset rated products, while users in the Movielens dataset rated movies. In both datasets, a 1 represented an utterly unsatisfactory response whereas a 5 represented an extremely satisfactory response.

After reading in the dataset, 80% of the given edges were used to train the algorithm (which provided the averages used in later calculations for similarity) and the remaining 20% of the edges were set aside to test the algorithm.

The algorithm was tested by first running through 20% of edges initially set aside. The first part of the algorithm determined whether the edge had sufficient data to make a prediction. If the edge had sufficient data, the second part of the algorithm was implemented to obtain a predicted weight.

The predicted weight was measured against the actual weight and the percent error was determined. To determine the overall percent error of the algorithm, the average of all individual percent error values for each prediction was used.

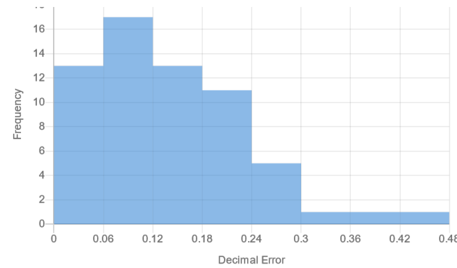


Fig. 4. Histogram of individual error of epinions recommendations

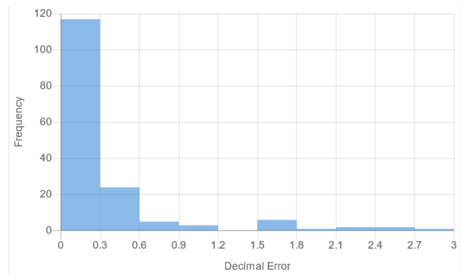


Fig. 5. Histogram of individual error of movielens recommendations

The performance of the algorithm across both datasets is detailed in Fig. 4 and 5 and summarized in Table 1.

Comparing Fig. 4 and 5, we noted that the error in predictions from the Epinions dataset was approximately unimodal, centered between 6 and 18 percent. On the contrary, the error in predictions from the Movielens dataset was much more right-skewed with no clear center value exhibited. The overall percent error was therefore calculated using the mean value in the Epinions dataset and the median in the Movielens dataset.

Considering the two datasets, the overall percent error was approximately 14% which improved upon the 20% error of existing algorithms [20]. This attests to the strength of our two-step approach and encourages further exploration into similar link prediction algorithms. However, though this is a promising finding, it must be noted that each recommendation was "screened"; i.e., a recommendation was only made if sufficient data existed in the first place.

V. Discussion

The two-step algorithm performed fairly well; however, outlier recommendations were notable, particularly with respect to the Movielens dataset. Specifically, the Movielens recommendations generated error values up to 300%. Further analysis revealed that though both datasets employed a similar 1-5 rating system and modelled a type of product-customer relationship, the most pronounced difference was that the

TABLE I. Percent Error of Epinions and Movielens Dataset

Dataset	Percent Error
Epinions	13.8%
Movielens	14.4%

Movielens dataset contained nearly 10 times the number of entries than the Epinions dataset. This size difference may have led to greater diversity in opinions, which in turn may necessitate a different approach to the parameters.

With respect to the size of datasets, this algorithm runs in $O(n^3)$ time. Therefore, testing the algorithm on datasets containing more than approximately $100,000$ entries is infeasible given the unrealistic run time required. A significant number of existing datasets contain millions of entries, which limits the scope of datasets available for testing. However, given the size of the Epinions and Movielens datasets, both datasets may provide solid representations of a real-world product-customer relationship. However, the high time complexity is ultimately a limitation compared to other methods which were able to operate on much larger datasets.

An alternative method for calculating $\frac{n}{t}$ may also be considered. In the current algorithm, n was defined as the number of candidate bottom nodes and top-adjacent bottom nodes that share at least one common top node. However, instead of a single common top node, the two bottom nodes may be required to share at least two common top nodes, three common top nodes, or even more common top nodes.

However, it must be noted that adjusting this parameter would likely necessitate changing the $\frac{n}{t}$ threshold. Requiring a greater number of common top nodes would decrease the value of n and warrant lowering the $\frac{n}{t}$ threshold correspondingly. This makes sense as while there might be fewer pairs of sufficient bottom nodes, there would ultimately be a relatively equivalent amount of data to work with as the number of common top nodes was increased.

Additionally, other conditions can be deliberated on in which the $\frac{n}{t}$ value may be deemed sufficient enough to generate a prediction. Examining the threshold, potential edges were sometimes rejected due to insufficient data. Although the $\frac{n}{t}$ value was high enough, the value t might have been too low to instill confidence in making a prediction. However, instead of rejecting those edges outright, the threshold might have been adjusted so that it increased as the number of top-adjacent nodes decreased.

Furthermore, different functions could be explored that determine similarity between the candidate bottom node and a top-adjacent bottom node. In a manner similar to how the Pearson coefficient was calculated, the piecewise function in Fig. 3 rewards the bottom nodes for rating their common top nodes similarly. It does this by taking into account the differences between how the two bottom nodes rated the common top nodes, and how they typically rated a top node. Matching signs (positive or negative) would contribute more positively towards similarity. One could also explore the effect of slightly modifying the constants of the functions.

Finally, this result continues to highlight the value of intuitive implementation within the context of this problem [21]. This follows a recent trend within the documentation of the link prediction problem of simplifying the problem [22]. Whereas previous papers used more involved methods such as neural networks or time series to generate predictions, this algorithm obtains strong results with two intuitive steps: search and calculation.

VI. Conclusion

In this paper, we presented a two-step approach to the classical link recommendation problem. Extending ideas from previous papers focused on sign prediction or unweighted link prediction, the algorithm utilized the properties of the weighted bipartite graph to determine two parameters: 1) the most relevant edges that could appear based on pre-existing edges, and 2) the weights of the edges based on the idea of *similarity*.

In the first step, the algorithm considered whether the candidate bottom node had common neighbors with bottom nodes adjacent to the candidate top node. The algorithm proceeded to compared this number against the total number of bottom nodes adjacent to the candidate top node. The algorithm then further determined whether any edges considered sufficient were a result of low sample size. The exact tuning parameters were determined by the density and size of the datasets.

In the second step, the idea of *similarity* was used. Similarity between the candidate bottom node and candidate top adjacent bottom nodes was calculated by taking into account how closely common neighbors were rated. The weight of the edge between the candidate bottom node and the candidate top was more heavily influenced by higher similarity pairings and less influenced by lower similarity pairings.

This algorithm was tested using the Movielens and Epinions datasets, and it performed well compared with existing algorithms. Further work will refine and improve the results of the proposed algorithm. In particular, future work can explore means to optimize its performance on larger datasets as well as how to fit the constant terms better for each particular datasets.

Acknowledgment

I would like to thank Dr. Zhiliang Xu, professor of Applied and Computational Mathematics and Statistics at the University of Notre Dame, for introducing me to this topic and providing invaluable guidance and support during the research.

References

[1] R. C. Read and R. J. Wilson. *An Atlas of Graphs (Mathematics)*. Clarendon Press, 2005.

[2] J. A. Bondy and U. S. R. Murty. *Graph Theory With Applications*. Great Britain: Macmillan Press, 1976.

[3] C. C. Aggarwal, ed. *Social Network Data Analytics*. New York, NY: Springer, 2011.

[4] K. Kobysheva, N. Voinova, I. Nikiforova. "Hybrid image recommendation algorithm combining content and collaborative filtering approaches". *Procedia Computer Science*. 2021.

[5] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang. "LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation". *SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020.

[6] H. Zhu et al. "Bilinear Graph Neural Network with Neighbor Interactions" *International Joint Conferences on Artificial Intelligence*. 2020.

[7] P. Bedi, A. Gautam, S. Bansal, D. Bhatia. "Weighted Bipartite Graph Model for Recommender System Using Entropy Based Similarity Measure". *The International Symposium on Intelligent Systems Technologies and Applications*. 2018

[8] P. Kumar, D. Sharma. "A potential energy and mutual information based link prediction approach for bipartite networks". *Scientific Reports*. 2020.

[9] O. Allali, C. Magnien, and M. Latapy. “Link prediction in bipartite graphs using internal links and weighted projection”. *2011 IEEE Conference on Computer Communications Workshops* (INFOCOM WKSHPS). 2011.

[10] N. A. Rahman. *A Course in Theoretical Statistics*. Charles Griffin and Company, 1968.

[11] Wikipedia, ”Collaborative filtering”. https://en.wikipedia.org/wiki/Collaborative_filtering, 2022.

[12] X. He et al. “Neural Collaborative Filtering”. *WWW ’17: Proceedings of the 26th International Conference on World Wide Web*. Apr. 2017.

[13] R. Zhang et al. “Collaborative Filtering for Recommender Systems”. *2014 Second International Conference on Advanced Cloud and Big Data*. 2014, pp. 301–308.

[14] K. Sun, T. Qian, T. Chen, Y. Liang. Where to Go Next: Modeling Long- and Short-Term User Preferences for Point-of-Interest Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020.

[15] Z. Yu, J. Lian, A. Mahmoody, G. Liu, X. Xie. ”Adaptive User Modeling with Long and Short-Term Preferences for Personalized Recommendation”. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. 2020. pp 4213-4219.

[16] H. Ghaleb and M. Abdullah-Al-Wadud. “An Enhanced Similarity Measure for Collaborative Filtering-based Recommender Systems”. *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*. 2019, pp. 1–4.

[17] Z. S. Patrous and S. Najafi. “Evaluating Prediction Accuracy for Collaborative Filtering Algorithms in Recommender Systems”. PhD thesis. Stockholm, Sweden: KTH Royal Institute of Technology, 2016.

[18] P. Massa. http://www.trustlet.org/downloaded_epinions.html. 2003.

[19] F. M. Harper and J. A. Konstan. “The MovieLens Datasets: History and Context”. *ACM Transactions on Interactive Intelligent Systems 5.4*. 2015.

[20] K. Goldberg et al. “Eigentaste: A Constant Time Collaborative Filtering Algorithm”. *Information Retrieval 4*, 2004.

[21] H. Fu, P. Poirson, K. S. Lee, C. Wang. ”Revisiting Neighborhood-based Link Prediction for Collaborative Filtering”. *WWW ’22: Companion Proceedings of the Web Conference 2022*. 2022. pp. 1009-1018.

[22] S. Rendle, W. Krichene, L. Zhang, and J. Anderson. ”Neural Collaborative Filtering vs. Matrix Factorization Revisited”. *Fourteenth ACM Conference on Recommender Systems*. 2020. pp 240–248.

Hyperspectral Image Segmentation using End-to-End CNN Architecture with built-in Feature Compressor for UAV Systems

Muhammad Bilal[1], Khalid Munawar[2], Muhammad Shafique Shaikh[3], Ubaid M. Al-Saggaf [4], Belkacem Kada[5]
Department of Electrical and Computer Engineering[1, 2, 3, 4]
Department of Aerospace Engineering[5]
Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia[1, 2, 3, 4, 5]

***Abstract*—Hyperspectral image segmentation is an important task for geographical surveying. Real-time processing of this operation is especially important for sensors mounted on-board Unmanned Aerial Vehicles in the context of visual servoing, landmarks recognition and data compression for efficient storage and transmission. To this end, this paper proposes a machine learning approach for segmentation using an efficient Convolutional Neural Network (CNN) which incorporates a feature compressor and a subsequent segmentation module based on 3D convolution operations. The experimental results demonstrate that the proposed approach gives segmentation accuracy at par with conventional approaches based on Principal Component Analysis (PCA) to reduce the feature dimensionality. Moreover, the proposed network is at least 35% faster than the conventional CNN-based approaches using 3D convolutions.**

***Keywords*—Hyperspectral images; CNN; dimensionality reduction; segmentation; PCA**

I. Introduction

Remote sensing is an important field which has helped urban and rural planning as well as environmental studies. Originally initiated as a satellite-based technology, it is being fast adopted to be used via sophisticated multispectral sensors that can be installed on Unmanned Aerial Vehicles (UAV) [1]. Other than being equipped with a standard RGB camera [2], several hyperspectral imaging (HSI) sensors [3] are now being deployed on-board these platforms to serve various applications in the field of security surveillance, town management, wild fires and agriculture etc. Another important application is to supplement the on-board navigation algorithms via visual servoing [4]. The spectral information provided by HSI is extremely rich and powerful. However, this raw information has to be processed extensively. Specifically, segmentation (classification) of each spatial data point in HSI has to be performed. Overall, the spatial and spectral data points constitute an information 'cube' [5]. Various researchers have put forward multiple segmentation methods of hyperspectral images to this end [6]. Various open-source datasets have also been made available to facilitate the research and development in this field [7]. Conventionally, HSI segmentation has been done through well-known machine learning methods such as Bayesian [8] distance, nearest-neighbor classification [9] and hand-coded features (e.g. Local Binary Pattern) [10] etc. Although, these methods have shown satisfactory performance on the standard datasets, recently CNN-based methods have shown exceptionally better results with near optimal segmentation [11, 12]. In this regard, 'Indian Pines' is a well-known HIS dataset [7] captured through Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [13]. This scene (Fig. 1) consists of 145×145 pixels (spatial data points) and 224 spectral reflectance bands captured in the wavelength range [0.4 2.5] µ meters. This scene is marked by regions consisting of common agricultural land, a forest as well as wild natural vegetation. Other than this, highways, a railway line and some urban construction also exist. The ground truth (intended segmentation output) consists of sixteen classes (Table I).

Some earlier works have suggested reducing the number of spectral bands to *200* by removing bands corresponding to [104-108], [150-163], the so-called 'water absorption bands' to improve segmentation accuracy [14]. Moreover, it has been observed that there is a considerable spectral redundancy i.e. the observations in different spectral are highly correlated [15]. Thus, it has been found beneficial to consider dimensionality reduction through PCA before application of segmentation algorithms. Furthermore, it has been suggested to employ the spatial variations in spectral data, especially in the case of high-resolution data, to improve the segmentation accuracy. While, the former approach helps in increasing the inference time (lower computational load), the latter adds further computational complexity to the segmentation task. However, in order to improve segmentation performance, it has been proposed to use both spectral and spatial data while segmenting a particular data point [6]. Spatial information improves the segmentation accuracy because further discriminating information could be found in the texture and shape of geological structures. To illustrate this point, we have trained two classifiers on Indian Pines dataset. The results have been given in Table II.

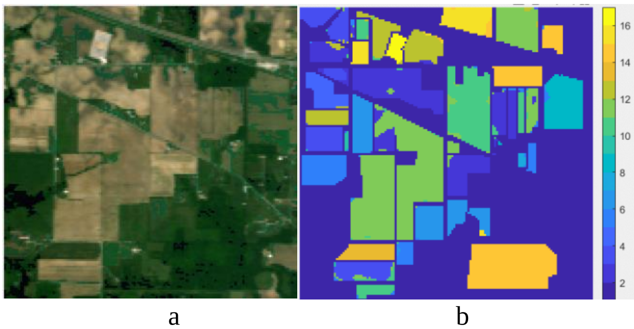


Fig. 1. Segmentation of a hyperspectral image (Indian pines). a) RGB image b) pseudocolor representation of segmented regions for 16 classes.

TABLE I. GROUND TRUTH DATA FOR INDIAN PINES DATASET

#	Class	Number of Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

TABLE II. REFERENCE SEGMENTATION CLASSIFIER ON INDIAN PINES DATASET

Classifier	Dimensionality	Segmentation Accuracy
SVM	220	96.0%
SVM	30	96.1%
CNN	30	99.6%

The first classifier is based on a conventional machine learning classifier, i.e. Support Vector Machine (SVM) [16]. In order to verify the strong redundancy in the spectral bands, two different versions have been trained. The first version uses all 220 hyperspectral bands (including noisy bands) while the second reduces the dimensionality of these spectral bands through PCA. PCA works by projecting the d dimensional data to m dimensional data where $m < d$.

$$\boldsymbol{x}_i \in \mathbb{R}^d, i = 1,2,\cdots N$$

$$\boldsymbol{z}_i \in \mathbb{R}^m, i = 1,2,\cdots N$$

The projection works by exploiting the eigen values and corresponding eigen vectors to build the transform as follows.

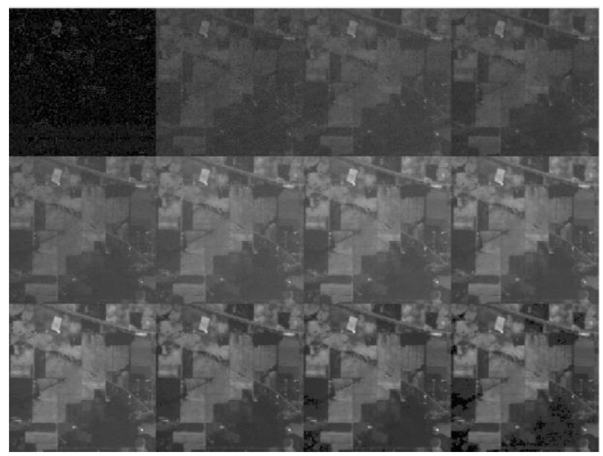
$$\boldsymbol{S}\boldsymbol{u}_{\boldsymbol{d}} = \lambda_d \boldsymbol{u}_{\boldsymbol{d}} \quad (1)$$

Where S is the normalized covariance matrix of the input features, λ_d are the eigen values while $\boldsymbol{u}_{\boldsymbol{d}}$ are the corresponding eigen vectors. The transformation is then given by.

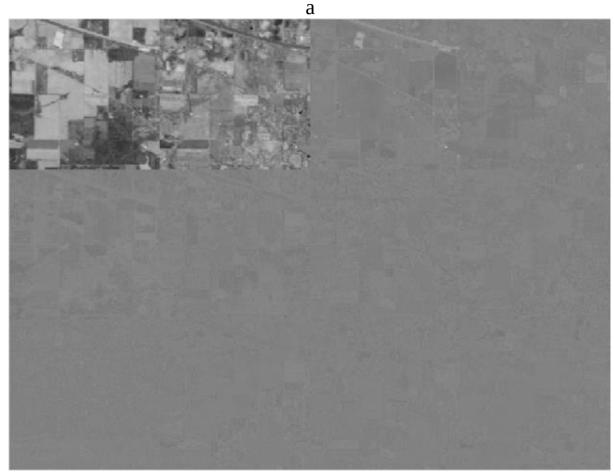
$$z = \boldsymbol{u}_m^T \left(\frac{x_i - \mu}{\sigma}\right) \quad (2)$$

As can be noticed from the data given in Table II, the segmentation accuracy with and without dimensionality reduction is almost identical with the minor difference being without any statistical significance. The accuracy results have been obtained with the provided ground truth as benchmark.

Fig. 2 shows the original first twelve bands in the Indian Pines data as well as the twelve most significant bands after PCA has been applied. It can be clearly seen that the most information is contained within the first three bands only. Thus, the dimensionality reduction through PCA is merited to facilitate the further segmentation task. However, it should be noted that PCA does not ensure that the segmentation accuracy will remain intact in all the cases.



a



b

Fig. 2. Application of PCA on HIS data from Indian Pines a) First twelve original bands b) Twelve transformed bands after PCA.

The second classifier considered in Table II is a CNN [6] classifier which also incorporates the spatial information while performing segmentation. In order to reduce the computational load, CNN has also been fed the same PCA-compressed spectral bands as for SVM classifier. It can be seen that this classifier almost achieves 100% accurate results. For both classifiers, 70% data points have been used for training while the remaining have been used for validation. Moreover, for CNN classifier, image patch size of 25×25 has been used to learn spatial information.

This approach of reducing the spectral information before applying the classifier has been considered by various early researchers [17]. The other approach of inferring spatial information before incorporating spectral data has also been considered in some works [18]. Adaptive Markov Random Fields [19] and graph cut [20] algorithms have also been considered to regularize the segmentation results in the spatial domain. Joint processing of spatial and spectral features has been considered by Zhon et al. [21] and Guijarro et al. [22]. These efforts to combine spatial and hyperspectral data to extract joint features is especially important since latent redundancy could be exploited. As noted earlier, HIS data cube hold valuable information in the local neighborhoods both with respect to spatial and spectral dimensions. However, this also presents a challenge in terms of computation complexity. This is especially relevant for on-board processing of data on battery-powered UAVs.

In contrast to the conventional machine learning models, deep learning or CNN-based approaches have been shown to be exceptionally well in exploring the data dependencies and discriminating features. Thus, several approaches have been proposed by researchers in the recent years related to this technique [23-25]. In this regard, U-Net [26-28] is of special interest because it has been successfully applied to image segmentation tasks other than HSI. However, since like other CNN-based segmentation frameworks, the computational load is a consideration for these methods, PCA is generally applied to reduce the HSI data dimensionality as shown above. However, since PCA and the subsequent CNN deep feature extraction modules are separate, the joint spatial-spectral redundant information [27, 29] may not be captured for optimal segmentation. To this end, Ying et al. [6] proposed using 3D CNN operations for HSI segmentation. 3D convolution kernels can learn the patterns found in both the spatial and the spectral dimensions of the HSI cube in local neighborhoods. This information has been shown to be of crucial value for segmentation. This structure can take the full advantage of all structural similarities within the HIS cube. 3D CNNs have shown promise in a variety of computer vision applications where the input is simple RGB images. Moreover, they have been applied to video frames to make the best use of temporal information. While using 3D CNNs for HSI segmentation, not only the pre-processing (PCA) overhead is reduced, but also the spectral and spatial features are learnt jointly. However, this presents a computational overhead since 3D convolutions are even more complicated than their 2D counterpart. In this paper, we have proposed an approach to circumvent this problem by adding an '*Encoder'* stage before 3D CNNs. The '*Encoder'* stage is inspired by the autoencoders [30] which have been shown to infer even non-linear statistical dependencies. In contrast PCA can only infer linear correlations. The proposed 'Encoder' stage compresses the spectral features, just like PCA, but in tandem with the subsequent segmentation stage based on 3D CNN. In this regard, Khan et al. [31, 32] have proposed using a traditional Autoencoder for the segmentation of HIS but they do not explicitly target the improvement in the inference speed. On the other hand, the experimental results demonstrate that the proposed architecture works at par with the state of the art while being up to 35% faster.

The remaining paper has been structured as follows. Section II describes the proposed architecture in detail with respect to the inclusion of encoder section before a 2D CNN. Section III has been devoted to the provision of results of the proposed architecture on standard datasets followed by a discussion. Section IV concludes the discussion by summarizing the results and contributions of the current work.

II. Proposed Architecture

Autoencoders have been extensively used in computer vision applications to learn the underlying data decencies and have been shown to work better than PCA in the presence of non-linear structures. Autoencoders are made up of two parts i.e. encoder and decoder. The encoder finds a compressed representation of the input while the decoder replicates the input using the compressed (low dimensional) encoding. Inspired by their architecture, we propose adding only the 1D encoding stage before a 3D CNN stage for HIS cube processing. The idea is to find non-linear data decencies in the spectral bands data which is then subsequently processed by the 3D CNN architecture to exploit spatial information as well. However, since the two modules are part of one learnable architecture, both spatial and spectral information are learnt jointly. Moreover, it is faster since the early part of the network (closer to the input) only processes the spectral dimension. In contrast the 3D CNN architecture proposed in [6] processes the whole cube starting from the input which leads to a high computational load.

The proposed architecture depicted in Fig. 3 consists of two stages as mentioned earlier. The 'Encoder' stage reduces the spectral dimension in two stages. The first stage uses 3×3 convolutions with a stride of two to reduce the dimensions from 220 to 110. The second stage uses 3×3 convolutions with a stride of fourj to reduce the dimensions from 110 to 27. The spatial dimensions (25×25) remain the same. Thus, at the end of the 'Encoder' stage, a 25×25×27 cube is presented the 3D CNN stage for further segmentation. This network consists of four 3D convolutional layers i.e. 3×3×7, 3×3×5, 3×3×3 and 3×3×1. There are 8, 16, 32 and 8 filter kernels for these layers respectively. A 256 long vector is generated through a fully connected layer followed by another fully connected layer generating a 128 long vector. The final vector (16 output classes) is generated through another fully connected layer. Thus, the 25×25×220 HSI input cube is gradually processed through convolutions to generate the final segmentations. The hallmark of this architecture, as mentioned earlier, is that the 'Encoder' stage is light-weight since it only processes the spectral data. However, despite only processing spectral data, it is connected with the subsequent spatial processing layers as well through the forward propagation. Thus, the whole network learns spectral and spatial information jointly without being computationally complex. The proposed architecture has been implemented using the Matlab Deep Learning Toolbox and trained in the same environment using NVIDIA GeForce 940 MX GPU on a intel i7 powered CPU (3.48 GHz) with 12 GB RAM. For training and testing of the proposed architecture, Indian Pine, Salinas, SalinasA, Pavia and PaviaU datasets [7] have been used. A sample training session has been shown in Fig. 4. It can be clearly seen that the training process quickly converges for both training and validation sets. Thus, not only is the proposed architecture suitable for the task at hand but also does not overfit showing excellent harmony between spatial and spectral data points.

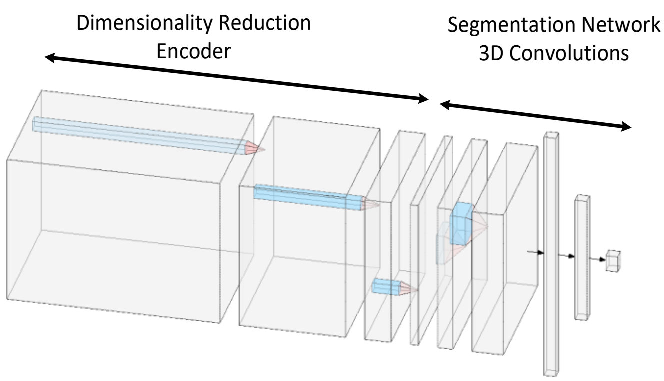


Fig. 3. Proposed HSI cube segmentation through Encoder + 3D Convolution stages.

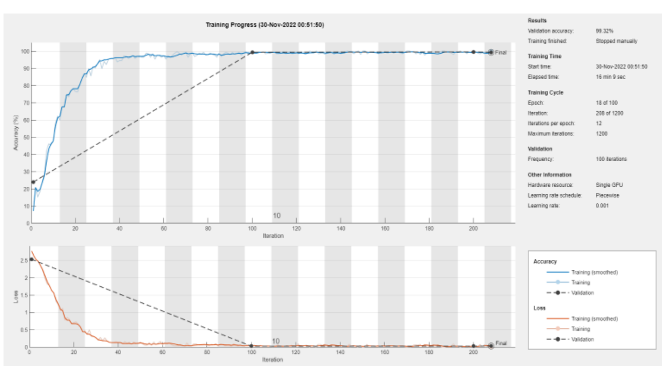


Fig. 4. Training of the proposed network in Matlab environment depicting fast convergence without overfitting.

III. Results and Discussion

The proposed architecture has been trained and tested on Indian Pine, Salinas, SalinasA, Pavia and PaviaU datasets. The former three datasets have been captured by AVIRIS sensor with 220 bands. The latter two datasets are captured by ROSIS sensor and have 102 and 103 bands respectively. To cater for the fewer number of bands on ROSIS datasets, the first stage of the 'Encoder' uses 3×3 convolutions with a stride of two to reduce the dimensions from 102/103 to 50. The second stage uses 3×3 convolutions with a stride of two to reduce the dimensions from 50 to 25. The rest of the architecture remains same.

The test results have been reproduced in Table III. It can be seen that the proposed network either performs better or at par with the 3D CNN network. Moreover, it is worth mentioning that this particular network has been tuned for each dataset separately by the authors [6] while we have used the same universal architecture for all cases. Thus, the proposed architecture is universally applicable to all scenarios.

It is worth mentioning that the proposed architecture also operates faster than the competing 3D CNN architecture due to its simpler 'Encoder' stage. On average, the proposed network processes an HSI cube of size 145×145×220 in *7.14 s* while a 3D CNN without encoder stage processes the same in *9.64 s*. Thus, the proposed approach is at least *35%* faster while being equally accurate.

Fig. 5 shows the reconstruction of spectral data for four example data points in Indian Pine dataset. This data has been generated by a reference Autoencoder with both 'Encoder' and'Decoder' parts for visualization. It can be seen that the spectral information can be fairly accurately recovered even after being compressed to 27 channels from original 220 (12%). However, it must be noted that the purpose in the application at hand is not reconstruction but use in subsequent segmentation where upto 100% accuracy has been achieved on SalinasA dataset even with this much compression. Visual results using psuedocolors for segmentation results on the mentioned datasets have been depicted in Fig. 6, Fig. 7, Fig. 8 and Fig. 9 respectively.

TABLE III. Segmentation Accuracy on Test Datasets

Architecture	Indian Pine	Salinas	SalinasA	Pavia	PaviaU
3D CNN [6]	99.07%	-	-	-	99.4%
Proposed	99.6%	94.3%	100%	98.0%	98.3%

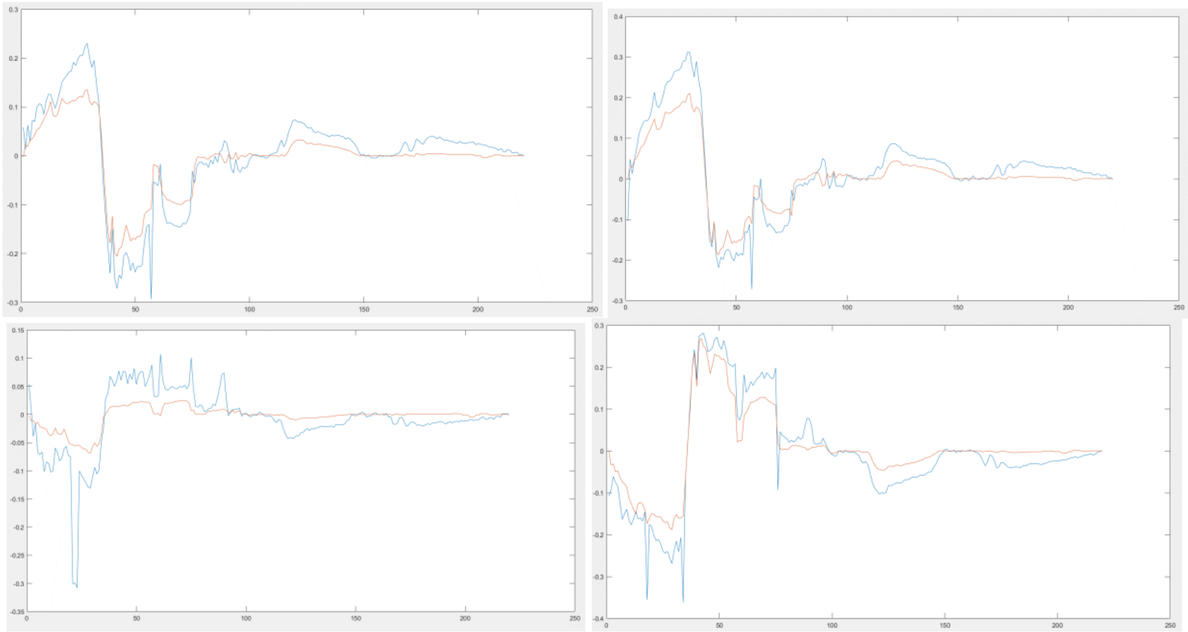
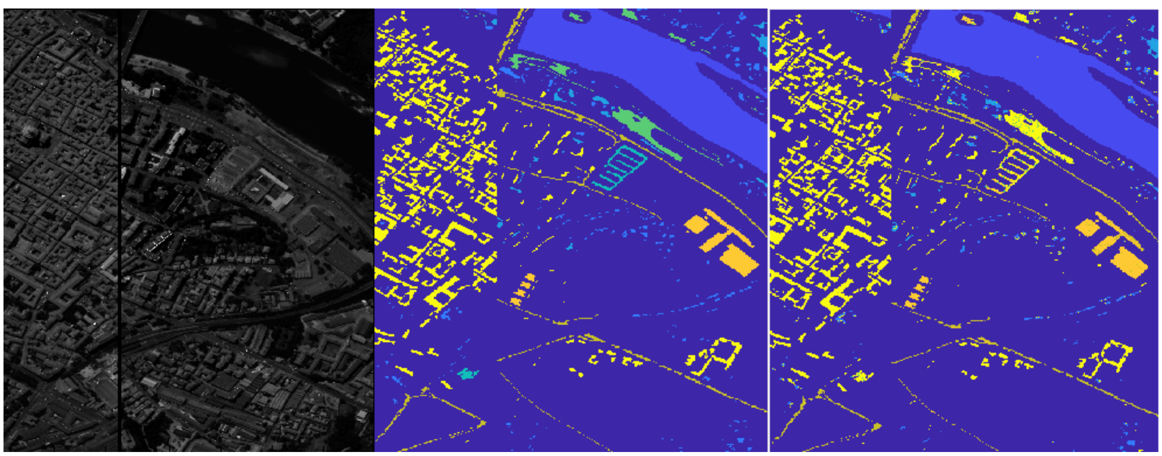
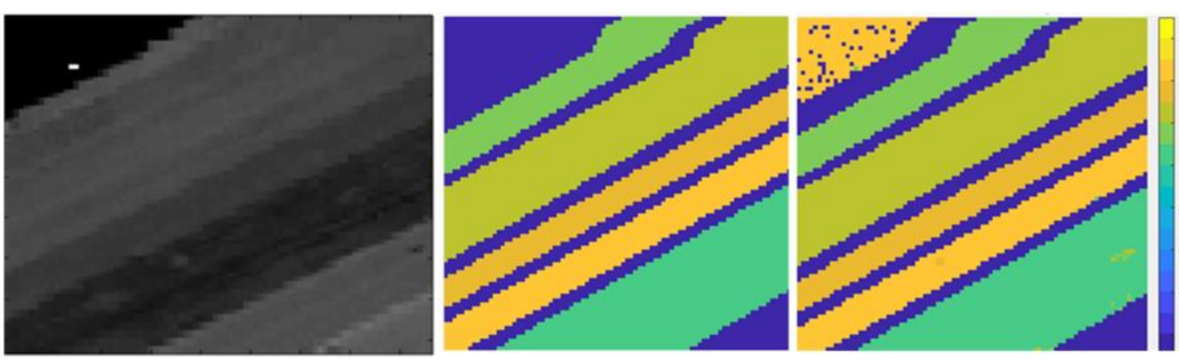


Fig. 5. Reconstruction of the spectral data for 4 sample spatial points. Original-Blue, Reconstructed-Red.



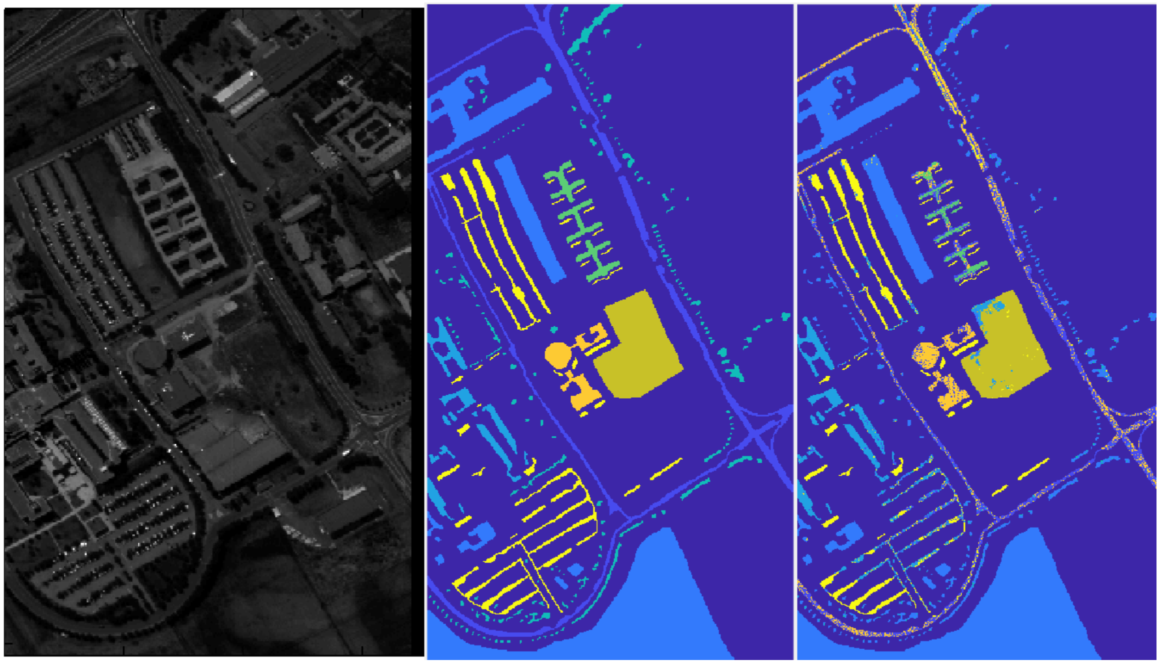
Input Grayscale Ground Truth Proposed Approach

Fig. 6. Pavia centre segmentation.



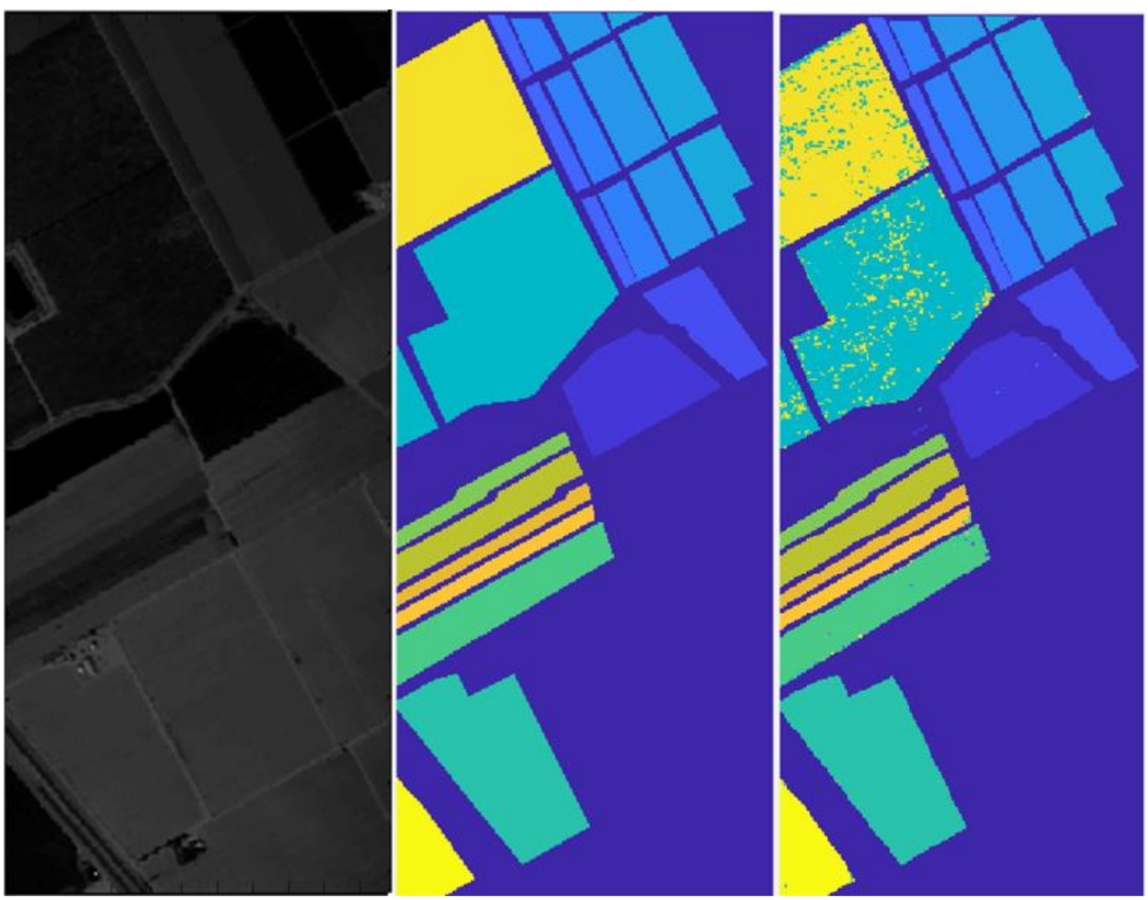
Input Grayscale Ground Truth Proposed Approach

Fig. 7. SalinasA Segmentation.



Input Grayscale | Ground Truth | Proposed Approach

Fig. 8. Pavia University segmentation.



Input Grayscale | Ground Truth | Proposed Approach

Fig. 9. Salinas segmentation.

IV. Conclusion

This paper has proposed a 3D convolution-based segmentation architecture for hyperspectral images. The proposed architecture achieves detection accuracy at-par with the state of-the-art with the added benefit that it achieves up to 35% higher speed due to a feature compressor module at the input. Moreover, this module works almost as good as conventional PCA in compressing the spectral information while being trained in tandem with the spatial data. Thus, the network has the ability to jointly learn spatial and spectral information. In conclusion, the proposed architecture provides a better alternative than both PCA-CNN and 3D-CNN architectures, combining the best of both approaches in terms of computational speed, learning flexibility and segmentation accuracy. For future work, the proposed technique could be applied to HS images from more satellite imagers to test the generalization of the approach.

Acknowledgment

This project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH) – King Abdulaziz City for Science and Technology - the Kingdom of Saudi Arabia – award number (11-SPA2059-03). The authors also, acknowledge with thanks Science and Technology Unit, King Abdulaziz University for technical support.

References

[1] K. A. M. F. Asmala Ahmad, Mohd Mawardy Abdullah, Suliadi Firdaus Sufahani, Mohd Yazid Abu Sari and Abd Rahman Mat Amin, "Noise and Restoration of UAV Remote Sensing Images," International Journal of Advanced Computer Science and Applications, vol. 11, no. 12, 2020.

[2] G. Zhang et al., "UAV Low-Altitude Aerial Image Stitching Based on Semantic Segmentation and ORB Algorithm for Urban Traffic," Remote Sensing, vol. 14, no. 23. doi: 10.3390/rs14236013.

[3] J. M. Haut and M. E. Paoletti, "Cloud Implementation of Multinomial Logistic Regression for UAV Hyperspectral Images," IEEE Journal on Miniaturization for Air and Space Systems, vol. 1, no. 3, pp. 163-171, 2020.

[4] J. Pomares, "Visual Servoing in Robotics," Electronics, vol. 8, no. 11. doi: 10.3390/electronics8111298.

[5] G. Martín, J. M. Bioucas-Dias, and A. Plaza, "HYCA: A New Technique for Hyperspectral Compressive Sensing," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 5, pp. 2819-2831, 2015.

[6] Y. Li, H. Zhang, and Q. Shen, "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network," Remote Sensing, vol. 9, no. 1. doi: 10.3390/rs9010067.

[7] M. Graña, M. Veganzons, and B. Ayerdi. (2022). Hyperspectral Remote Sensing Scenes. Available: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[8] A. Ahmad, H. Sakidin, M. Y. A. Sari, A. R. M. Amin, S. F. Sufahani, and A. W. Rasib, "Naïve Bayes Classification of High-Resolution Aerial Imagery," International Journal of Advanced Computer Science and Applications, vol. 12, no. 11, 2021.

[9] B. Tu, J. Wang, X. Kang, G. Zhang, X. Ou, and L. Guo, "KNN-Based Representation of Superpixels for Hyperspectral Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 11, pp. 4032-4047, 2018.

[10] W. Li, C. Chen, H. Su, and Q. Du, "Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 7, pp. 3681-3693, 2015.

[11] Y. Kong, X. Wang, and Y. Cheng, "Spectral–Spatial Feature Extraction for HSI Classification Based on Supervised Hypergraph and Sample Expanded CNN," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 11, pp. 4128-4140, 2018.

[12] X. Xue, H. Zhang, B. Fang, Z. Bai, and Y. Li, "Grafting Transformer on Automatically Designed Convolutional Neural Network for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022.

[13] C. J. Bruegge et al., "Vicarious Calibration of eMAS, AirMSPI, and AVIRIS Sensors During FIREX-AQ," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 12, pp. 10286-10297, 2021.

[14] J. Wang, J. Zhou, and W. Huang, "Attend in Bands: Hyperspectral Band Weighting and Selection for Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 12, pp. 4712-4727, 2019.

[15] X. Shang et al., "Target-Constrained Interference-Minimized Band Selection for Hyperspectral Target Detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 7, pp. 6044-6064, 2021.

[16] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images," IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 4, pp. 736-740, 2010.

[17] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of Hyperspectral Images by Using Extended Morphological Attribute Profiles and Independent Component Analysis," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 3, pp. 542-546, 2011.

[18] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 3, pp. 480-491, 2005.

[19] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov Random Field Approach for Classification of Hyperspectral Imagery," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 5, pp. 973-977, 2011.

[20] Y. Tarabalka and A. Rana, "Graph-cut-based model for spectral-spatial classification of hyperspectral images," in 2014 IEEE Geoscience and Remote Sensing Symposium, 2014, pp. 3418-3421.

[21] Y. Zhong, A. Ma, and L. Zhang, "An Adaptive Memetic Fuzzy Clustering Algorithm With Spatial Information for Remote Sensing Imagery," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 4, pp. 1235-1248, 2014.

[22] M. Guijarro, G. Pajares, and P. J. Herrera, "Image-Based Airborne Sensors: A Combined Approach for Spectral Signatures Classification through Deterministic Simulated Annealing," Sensors, vol. 9, no. 9, pp. 7132-7149. doi: 10.3390/s90907132.

[23] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 152, pp. 166-177, 2019/06/01/ 2019.

[24] B. Xue et al., "A Subpixel Target Detection Approach to Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 9, pp. 5093-5114, 2017.

[25] C. Yu, R. Han, M. Song, C. Liu, and C. I. Chang, "Feedback Attention-Based Dense CNN for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022.

[26] F. Guo, Z. Li, Z. Xin, X. Zhu, L. Wang, and J. Zhang, "Dual Graph U-Nets for Hyperspectral Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 8160-8170, 2021.

[27] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "Multilevel Superpixel Structured Graph U-Nets for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-15, 2022.

[28] S. Hao, W. Wang, and M. Salzmann, "Geometry-Aware Deep Recurrent Neural Networks for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 2448-2460, 2021.

[29] W. Zhao and S. Du, "Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep

Learning Approach," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4544-4554, 2016.

[30] X. Ma, H. Wang, and J. Geng, "Spectral–Spatial Classification of Hyperspectral Image Based on Deep Auto-Encoder," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 9, pp. 4073-4085, 2016.

[31] S. D. Khan, L. Alarabi, and S. Basalamah, "An Encoder–Decoder Deep Learning Framework for Building Footprints Extraction from Aerial Imagery," Arabian Journal for Science and Engineering, 2022/03/29 2022.

[32] S. D. Khan, L. Alarabi, and S. Basalamah, "Deep Hybrid Network for Land Cover Semantic Segmentation in High-Spatial Resolution Satellite Images," Information, vol. 12, no. 6. doi: 10.3390/info12060230.

Design of Personalized VR Short Video Content Distribution Algorithm based on Artificial Neural Network

Han Zhong[1], Donghyuk Choi[2], Wonho Choi[3], Yu Zheng[4]*
School of Culture and Media, Anyang Institute of Technology, Anyang, China[1]
Dept of Animation & Visual Effects, Dongseo University, Busan, South Korea[2]
Dept of Visual Contents, Graduate School, Dongseo University, Busan, South Korea[3, 4]
School of Media, Anyang Normal University, Anyang, China[4]

***Abstract*—In order to improve the video quality, reduce the number of video jams, and improve the video transcoding rate, a new personalized distribution algorithm of VR short video content based on artificial neural network is proposed. BP neural network is used to compress the original video and determine the execution mode and cache location of VR short video cache file; Transcoding cached high bit rate VR short video stream to generate low bit rate video stream to meet the needs of different network bandwidth; Build a video distribution model, and build a multicast distribution tree based on this model, that is, add some relay servers to minimize the video distribution cost; Finally, through the algorithm of minimizing the distribution cost of VR short video, the bandwidth loss and response delay are effectively reduced to achieve the goal of minimizing the distribution cost. The experimental results show that the video blocking times of this method are always less than six times, which effectively reduces the video blocking times. The PSNR value is high, the increase is large, up to 0.5, and the video transcoding rate is improved, up to 92%.**

Keywords—Artificial neural network; VR short video; BP neural network; video cache; cost optimization

I. Introduction

With the rapid development of network information technology, people rely more and more on the Internet and smart phones for entertainment. The emergence of various short video applications has met the needs of users for fragmented reading, social interaction and personalized expression, but problems have also arisen [1, 2]. At present, the short video industry is still in the rapid development stage. In view of the current situation of the industry environment, it is necessary to analyze the environment of the short video distribution platform to provide a reference solution to the problem. Traditional wireless network access and transmission modes are single, and usually rely on network infrastructure such as base stations and core networks to provide services. Once network congestion or force majeure occurs, users will not be able to obtain corresponding services [3, 4]. In this context, the growing user demand for content services and the limited network carrying capacity have become the main contradiction restricting the development of video. In order to meet the requirements of high quality and low latency for the development of short video, and reduce the network load at the same time, it is important to study the personalized distribution method of short video content.

II. Literature Review

Literature [5] proposed an online short video content distribution method based on federated learning. First, based on the federated learning method, the user group interest vector prediction algorithm is proposed by training the interest prediction model using the user group's local album data, and the user group interest vector representation is obtained; Then, with the user group's interest vector as the input, the corresponding short video content distribution strategy is designed in real time based on the combined confidence upper bound algorithm. The experimental results show that the short video content distribution strategy can effectively improve the accuracy of short video distribution, and help to improve the profits of video content providers. Literature [6] proposed a video distribution method based on QoE awareness. According to the characteristics of optical wireless fusion network structure, considering link state information and scalable video coding structure, a QoE evaluation model was established. Particle swarm optimization algorithm is used to select the optimal video transmission rate, and then analyze the node transmission capacity and node matching degree, so as to select the transmission path for the service and ensure the reliable transmission of the service. The experimental results show that this method can effectively improve the utilization of network resources and reduce network congestion while enhancing the quality of user experience. Literature [7] proposed a video distribution method based on random linear network coding. The source node uses RLNC to segment the video files to be distributed, and the encoded data pieces are distributed to mobile devices on the road. After receiving the data slice, each mobile device will re encode and distribute it to other devices. When the device receives a certain amount of linearly independent data, it will decode the data slice to obtain the original data. Three strategies are designed: mode switching, wireless access point selection, and active re coding to reduce the delay of video distribution. Experimental results show that this method has higher distribution efficiency.

From the above analysis, we can see that the existing methods improve the video distribution effect to a certain extent. In order to further improve the video quality, reduce the

*Corresponding Author.

number of video blocks, and improve the video transcoding rate, this paper proposes a personalized VR short video content distribution algorithm based on artificial neural network. The main research contents of this method are as follows:

1) Solve the problem of VR short video cache reuse by reasonably deploying and placing VR short video files.

2) The drift error adaptive fast video transcoder transcodes the buffered high bit rate VR short video stream to generate a low bit rate video stream to meet the needs of different network bandwidth.

3) Improve the video distribution model, generate a multicast distribution tree, and conduct personalized distribution of VR short video content to effectively reduce bandwidth loss and response delay, so as to minimize the distribution cost.

III. VR Short Video Preprocessing

A. Video Compression

In recent years, with the development of artificial neural network theory and technology, its application in video compression has gradually attracted people's attention. Compared with some traditional compression methods, the artificial neural network technology [8, 9] has good fault tolerance, self-organization and adaptability. Therefore, it is not necessary to use some predetermined data coding algorithm in the video compression process, but can independently complete the video compression completely according to the information characteristics of the video itself. BP neural network [10] is a typical and commonly used network structure in artificial neural network. This paper mainly introduces a video compression method based on BP neural network.

BP neural network refers to a multilayer feedforward network using BP algorithm. Fig. 1 is a simple feedforward network, which consists of an input layer, an output layer and a hidden layer (middle layer). Where, w_{zh} and w_{vh} are the weighted values between the input layer and the intermediate layer (hidden layer), hidden layer and output layer respectively. The neurons in the hidden layer and the output layer first sum the input signals of the previous layer by weighting, and then perform the pre specified transformation to generate the output signals. The output of the hidden layer neurons is also connected to the output layer neurons by weighting.

Let the input vector of the network be $Z=\{z_1, z_2, ..., z_n\}$, where n represents the number of neurons in the input layer. The transformation function adopted by neurons is:

$$D(n)=1+\frac{1}{\exp(w_{zh}-w_{vh})} \tag{1}$$

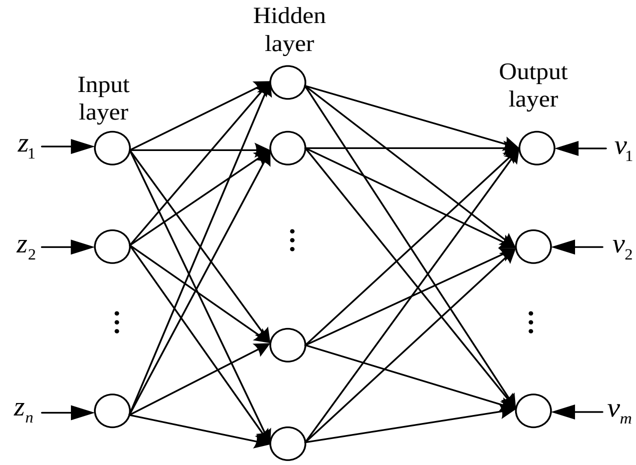


Fig. 1. Multilayer forward network.

The output vector corresponding to Z is $V=\{v_1, v_2, ..., v_m\}$, where m represents the number of neurons in the output layer. If the expected output of the network is required to be $F=\{f_1, f_2, ..., f_m\}$, the error function can be defined as:

$$E(n)=\sum_{i,j=1}^{m}\left(f_{ij}-v_{ij}\right)^2 \tag{2}$$

In the formula, f_{ij} represents the network training speed; v_{ij} represents the network convergence rate.

The BP algorithm uses the gradient descent method to adjust the network weight to reduce the above error function, namely:

$$\eta(n+1)=\eta(n)-\alpha\frac{E(n)}{\lambda(\eta)} \tag{3}$$

In the formula, $\eta(n)$ represents the approximation error of the neural network; $E(n)$ represents the difference function between the actual output and the expected output; $\lambda(\eta)$ represents generalization capability of neural network; α represents the weight adjustment rate, usually $0.01 \le \alpha \le 1$.

The general weight adjustment formula is:

$$\Delta\alpha=\sum_{i=1}^{m}E_{ij}\left|G_{ij}-H_v\right| \tag{4}$$

In the formula, G_{ij} represents the weight correction from the i-th node of a layer to the j-th node of the next layer; H_v represents the output of the node; E_{ij} represents the endpoint equivalent error of node i and node j, which is equivalent back transmitted from the error of the output layer:

$$\eta_v = \Delta\alpha\left(v_i - f_i\right) \tag{5}$$

In the formula, v_i represents that node i is located in the output layer; f_i represents that node i is in hidden layer.

For a VR short video with N frame, when BP neural network is used for compression, the number of neurons in the input and output layers of the network is required to be n^2, while the number of neurons in the hidden layer is much less than that in the input layer. When VR short video is transmitted in the network, the state of each neuron in the hidden layer, as some transformation result of the input information, must contain some information of the original VR short video. Therefore, when the output unit can reproduce the original VR short video, it can be considered that the output of hidden layer neurons is the compression result of the original input VR short video. Fig. 2 describes the VR short video compression process [11].

B. Video Cache

This section will focus on solving the problem of VR short video, and solve the problem of VR short video cache reuse by reasonably deploying and placing VR short video files. After the VR short video compression is completed in Section 2.1, determine the execution mode of the VR short video cache file and whether the cache location is appropriate [12].

Suppose user videos are stored in K clusters, $K = \left\{k_1^1, k_2^1, ..., k_L^1, ..., k_1^r, k_2^r, ..., k_L^r\right\}$, where k_l^i represents the VR short video file l stored in user cache k^i. Update VR short video files and store them in each blank cache. Among them, cluster member k_l^i only stores some files in the VR short video library. Due to storage restrictions, these files cannot contain all VR short video files. Therefore, it is expected to deploy more file contents in k_l^i and cache and reuse them.

For each cluster, selecting the video file to update is also an important issue. In each cluster, cluster file library L_p represents all video file records in the cluster, and $L_p = z_1 \cup z_2, ..., \cup z_l$ represents an independent cluster popularity distribution. On this basis, collect the VR short video file statistics cached by the whole cluster members in the cluster, and then update the missing videos. It is not difficult to find out which VR short video files are missing in the cluster according to the VR short video file statistics [13].

At the same time, considering the problem of video storage capacity, the storage formula of VR short video files is expressed as:

$$V_f = Z \subseteq C_m, m = 1, 2, ..., M \tag{6}$$

Among them, the maximum possibility of updating the number of VR short video files is as follows:

$$\mu_{\max} = M - C_m \tag{7}$$

Therefore, to update the VR short video file library L_p, assume that each file requested to be updated is equal to:

$$\phi_{L_p} = \left\{p \notin L_p, \mu_{\max} \le M\right\} \tag{8}$$

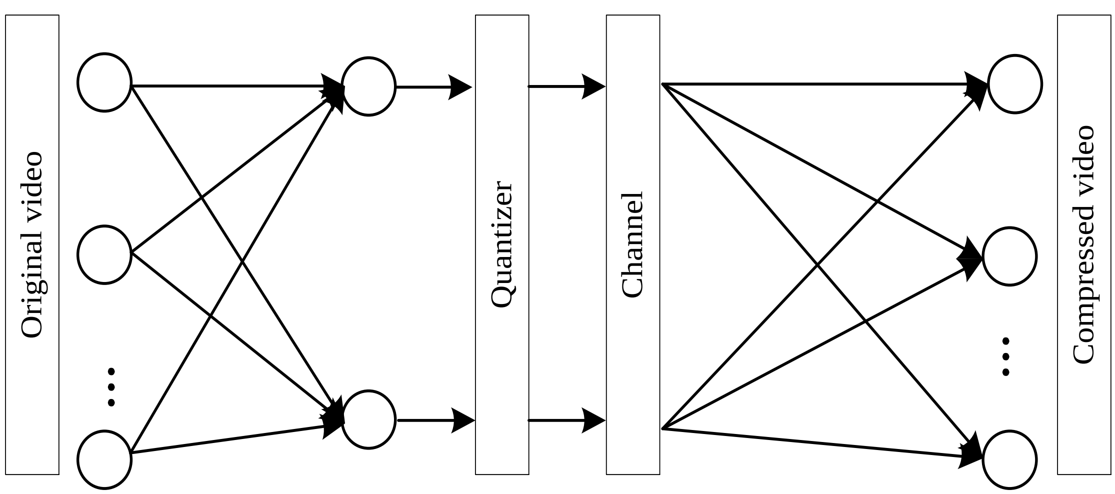


Fig. 2. VR short video compression process.

In the video cache, the most popular VR short video file ϕ_{L_p}, which is not in the original cluster, may be selected to offset the request. At the same time, a popular VR short video file is arranged and stored in the cluster head and the cluster center to facilitate user access and make the user's viewing experience better and smoother. And through constant iteration and change, users can obtain the latest VR short video information without requiring additional micro cache or auxiliary devices to simplify VR short video cache.

C. Video Transcoding

It is one of the most important applications of VR short video transcoding to generate low bit rate video streams to meet the needs of different network bandwidth [14]. Generally, the simplest and most direct way to realize VR short video transcoding is to use the so-called full decoding and full encoding scheme, that is, first decode the encoded video stream to generate a video sequence, and then re encode them to generate a new video stream. Because the recoding process requires a very complex motion estimation and mode selection process, this scheme is difficult to achieve real-time video transcoding, which greatly limits the practical application of this scheme.

To solve these problems, researchers have proposed various video transcoding structures, among which the most influential are the so-called open loop structure and closed loop structure. The open loop structure and the closed loop structure completely reuse the macroblock mode and motion vector information in the original bitstream, so there is no need for high complexity motion estimation and mode selection. Among them, the open-loop structure can directly transcode in the compressed domain (DCT domain) [15]. The transcoding speed is very fast, but due to the existence of drift error, the transcoded video frames are seriously distorted. The closed-loop structure introduces a closed loop in the encoding part to eliminate drift error, which basically eliminates the impact of drift error on the video quality after transcoding. However, the introduction of a closed loop also increases the complexity of the transcoder, making the speed of the closed-loop structure lower than that of the open loop structure. To solve these problems, this paper designs a fast video transcoder with drift error adaptive.

The closed loop for eliminating drift error in the closed-loop video transcoder can be divided into two parts, in which the part before the error buffer input completes the accumulation of drift error; the part after the error buffer output completes the drift error compensation. Suppose Y is the DCT coefficient obtained by decoding in the original code stream, and W is the DCT coefficient obtained by error compensation operation. The new quantized DCT coefficient T obtained by quantization operation is obtained as follows:

$$T = \frac{\left(Y + W\right)^2}{Q^2} \tag{9}$$

In the formula, Q^2 represents quantization operation. If $0 \le \left|Y + W\right| < Q^2$, then the non-zero error W does not affect the value of T. Therefore, if a threshold θ is given in advance, the error compensation operation can be determined according to the size of error W and threshold θ. Based on this idea, this paper proposes the working principle of the drift error adaptive fast video transcoder as shown in Fig. 3 [16]. For VR short video stream, there are three different types of coded video frames (A frame, B frame and C frame). The transcoder of this structure operates as follows:

1) The current transcoding video is A frame: no error compensation operation is required, but error accumulation operation is required.

2) The current transcoding video is a B frame: error compensation and error accumulation operations are required, but error compensation for B frame is performed in blocks.

3) The current transcoding video is C frame: in order to speed up the transcoder, C frame requires neither error compensation nor error accumulation.

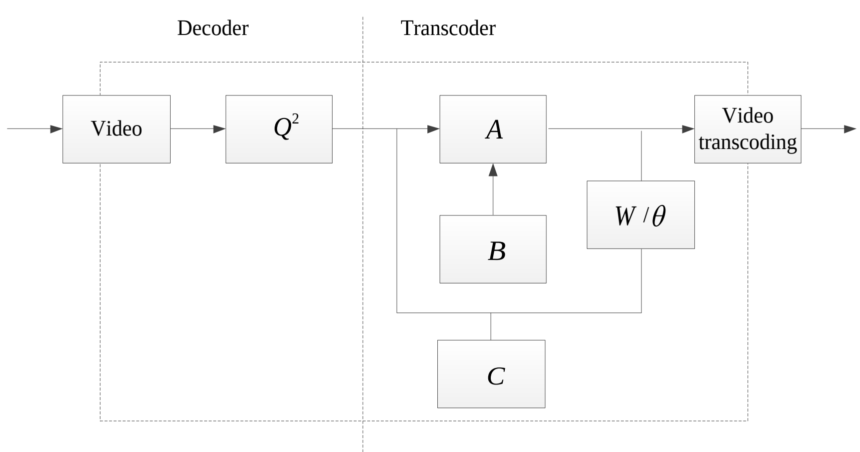


Fig. 3. Working principle diagram of video transcoder.

IV. VR Short Video Distribution Algorithm Considering Cost Optimization

The development of video applications has put forward higher requirements for video distribution in terms of real-time, mobility and interactivity. How to efficiently distribute video, alleviate the contradiction between bandwidth pressure and response delay, and reduce the cost of video distribution has become a difficult and hot spot in video distribution application research. Based on the above video compression, video caching and video transcoding results, this section proposes a cost optimized VR short video distribution method. This method can effectively reduce the bandwidth loss and response delay, thus minimizing the allocation cost. In the process of video distribution, a multicast distribution tree is built to realize the communication and video distribution between the source server and the proxy server.

A. Video Distribution Model

In the VR short video distribution framework, it mainly includes source server, proxy server and relay server. This article assumes that the source server stores all video streaming media, and each proxy server manages at least one area. Users can request and download videos from proxies that are closer to them. In order to meet the needs of users, a VR short video distribution method combining active distribution and passive distribution is adopted. The VR short video distribution model is shown in Fig. 4.

In Fig. 4: 0 represents the source server, 1 represents the relay server, 2 represents the request server, and 3 and 4 represent the proxy server. The directed bold line segment in the figure represents the VR short video distribution process; the directed virtual line segment represents the process that the proxy server sends a request to the source server when there is no video requested by the user; the black directed solid line represents the distribution link.

As shown in Fig. 4, when a user sends a video streaming media request to the local proxy server, the source server quantifies the video based on the user's interest in the video. H_ε represents the maximum number of requests for video streaming media in the streaming media distribution system, H_ω represents the number of requests for video streaming media from the proxy server, and the source server quantifies the user's interest in video streaming media in the proxy server through the H_ε / H_ω ratio.

In the active video distribution strategy, the source server distributes the video streaming media through the multicast distribution tree, and caches it to some proxy servers according to the quantitative proportion. The user directly obtains the video content from the proxy server. The passive video distribution strategy aims at sending video access requests to the proxy server by users, but when they fail to hit, the proxy server sends video requests to the source server. According to the received video request, the source server distributes the video to the request proxy server through the multicast distribution tree, and then forwards it to the requesting user. In the process of multicast distribution, some relay proxy servers can be added to minimize the cost of video distribution. VR short video distribution can effectively reduce user access delay and bandwidth loss.

B. Multicast Distribution Tree

With the rapid increase of types and quantities of VR short videos, users have put forward higher requirements for network bandwidth and access delay. In the process of VR short video distribution, this paper comprehensively considers the impact of personalized demand, link bandwidth and delay on the video distribution cost. In the VR short video distribution model, the source server distributes the video to the proxy server by building a multicast distribution tree. During the construction of multicast distribution tree, based on the 4VR short video distribution model shown in Fig. 4, some relay servers can be added to minimize the video distribution cost. The multicast distribution tree model is shown in Fig. 5, where node 0 represents the source server, node 1, 3, 6, 7, 8 represents the proxy server, and node 2, 4, 5 represents the relay proxy server. The directed line segment between two points represents the communication link between servers. The number on the directed line segment represents the link cost when bandwidth and delay are considered comprehensively. In this paper, we want to find an optimized video distribution path from the source server to the proxy servers to minimize the cost of video distribution.

C. VR Short Video Distribution Cost Minimization Algorithm

In the process of VR short video distribution, the problem of cost minimization is proposed for satisfying users' best experience of video requests and minimizing distribution costs. The performance of the video distribution network discussed in this paper does not calculate the delay and bandwidth cost between the proxy server and the end user. Since users request video from proxy servers in their regions, access time and bandwidth costs will not be affected by video copy allocation.

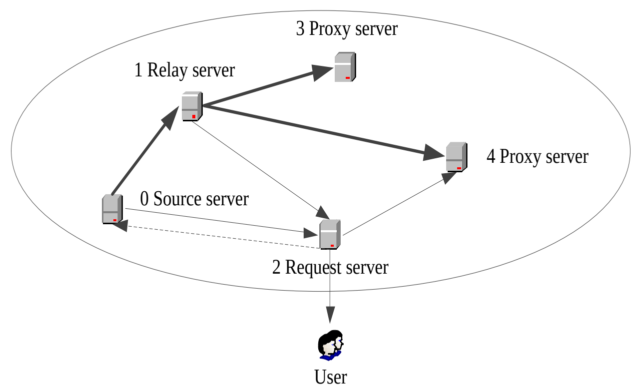


Fig. 4. VR short video distribution model.

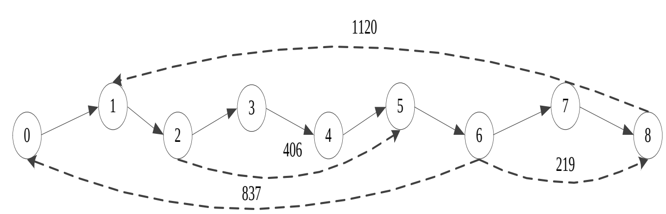


Fig. 5. Schematic diagram of multicast distribution tree model.

The video in the source server is divided into β different video streaming objects. Each video streaming media object is represented by the symbol U, so the video streaming media object set is $U=\{u_1,u_2,...,u_\beta\}$. The storage limit of each proxy server is I_a, and ϑ_a represents the number of video streaming media objects requested by proxy server a. During the active distribution of VR short videos, δ_a represents the degree of interest of users in proxy server a in video streaming media objects, and κ_a represents whether proxy server a caches video streaming media objects:

$$\kappa_a=\begin{cases}1, & a\in U\\ 0, & a\notin U\end{cases} \tag{10}$$

In the formula, 1 shows that the video object is actively distributed to the proxy server; 0 that the video object is passively distributed to the proxy server.

The cost function of VR short video distribution delay is as follows:

$$C_a=\sum_{a=1}^{n}\sum_{i=1}^{m}k^i\left(1-\kappa_a\right)-d_t \tag{11}$$

In the formula, d_t represents the unit retrieval delay of the proxy server for video streaming media objects.

The bandwidth cost function of VR short video distribution is as follows:

$$C_a^2=\sum_{a=1}^{n}\sum_{i=1}^{m}\left(\zeta_{ai}+\chi_{aj}\right)^2 \tag{12}$$

In the formula, ζ_{ai} represents the size of the video streaming media object; χ_{aj} represents the unit bandwidth of the video streaming media object requested by the proxy server.

When bandwidth and delay are considered at the same time, coefficients φ_1 and φ_2 are used to balance the delay and bandwidth costs to meet the needs of multiple applications and systems. Therefore, the cost minimization function of VR short video distribution is as follows:

$$C_{\min}^{'}=\varphi_1 C_a+\varphi_2 C_a^2 \tag{13}$$

The constraints of formula (13) are as follows:

$$\sum_{a,i=1}^{n}\kappa_a k^i\le\tau_i \tag{14}$$

$$\chi_{aj}\le\chi_{\max} \tag{15}$$

$$0.1\le\varphi_1<1;0.1\le\varphi_2<1 \tag{16}$$

$$\varphi_1+\varphi_2\ge 1 \tag{17}$$

Formula (14) is the storage capacity limit of the proxy server. Formula (15) indicates that the link bandwidth between the proxy server and the video streaming media object cannot exceed the maximum bandwidth between two points. Formulas (16) and (17) are the constraints of balance coefficients φ_1 and φ_2.

This paper summarizes the problem of minimizing the distribution cost of VR short video, that is, giving the size of different video streaming media objects and the number of video streams requested, whether there is an allocation strategy to minimize the total cost in formula (13), and whether the video capacity stored in the proxy server is full of formula (14), the size of the link bandwidth meets formula (15), and the relationship between the balance coefficients meets the restrictions of formulas (16) and (17).

V. EXPERIMENTAL VERIFICATION

In order to verify the application effect of personalized VR short video content distribution algorithm based on artificial neural network, experimental analysis is carried out. The methods in literature [5] and literature [6] are used as comparison methods and compared with the proposed methods.

A. Experimental Samples and Indicators

In the simulation experiment, the experimental hardware platform is Pentium IV 2.4GHz CPU with 8GB of memory. The experimental data comes from the COIN (Comprehensive Interactive Video Analysis), an open-source tutorial video dataset, which contains 11827 tutorial videos involving 180 tasks in 12 fields of daily life. Video test sequences are selected from this dataset, all of which are standard test sequences. When analyzing the performance of simulation results, this paper uses Matlab software to convert the experimental data comparison into specific images, which provides strong evidence for the conclusions of this paper.

When comparing the proposed method with traditional methods, the following three performance indicators are mainly considered:

1) Stuck times: the number of times a video stream is stuck during transmission and playback.

2) PSNR: A function of the square difference of pixels between the original video file and the received video file, which is a standard objective indicator to measure the video picture quality.

3) Transcoding rate: it can directly reflect the complexity of video transcoding and indirectly reflect the rate of video distribution.

B. Analysis of Experimental Results

1) Stuck times: The comparison results of the stuck times of the methods in literature [5], literature [6] and the proposed method in the video distribution process are shown in Fig. 6.

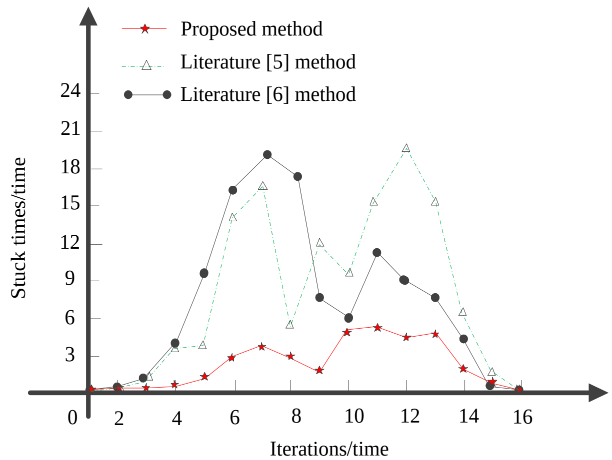


Fig. 6. Comparison results of stuck times.

2) PSNR: The PSNR comparison results of literature [5] method, literature [6] method and the proposed method are shown in Fig. 7.

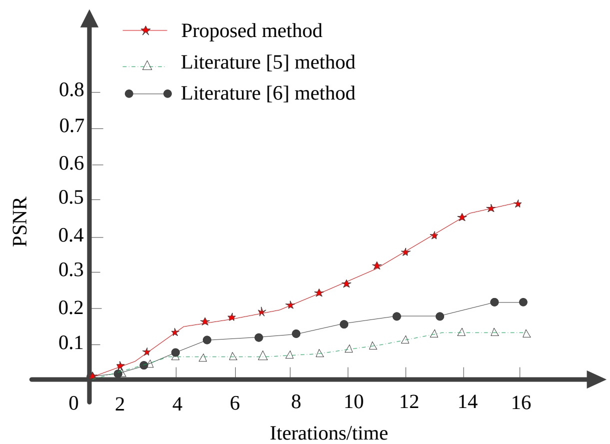


Fig. 7. Comparison results of PSNR.

3) Transcoding rate: Fig. 8 shows the video transcoding rate comparison results of three different methods.

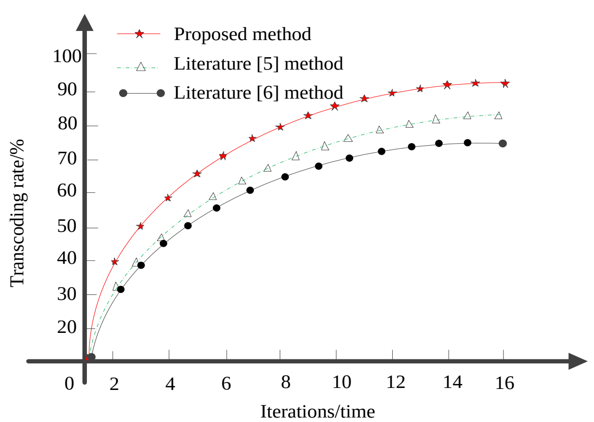


Fig. 8. Transcoding rate comparison results.

VI. DISCUSSION

Through the above experiments, conclusions can be drawn as follows:

1) As shown in Fig. 6, the number of video clicks of the proposed method is always less than six times, while the maximum number of clicks of the method in literature [5] reaches 20 times, and the maximum number of clicks of the method in literature [6] reaches 19 times. Compared with traditional methods, the proposed method can provide users with a higher quality user experience during video distribution, that is, improve the video quality level of users when the video is played smoothly. This is because the design method can select the most popular VR short video files not in the original cluster in the video cache to offset the request. At the same time, the copies of popular VR short video files are stored in the cluster header to facilitate user access, thus making the user's viewing experience smoother.

2) It can be seen from Fig. 7 that although the PSNR values of the three methods have an overall improvement trend, the variation range of the methods in literature [5] and literature [6] is small. The PSNR value of this method is high, with a large increase, up to 0.5, indicating that the average video quality level obtained by users is high, indicating that the video distribution quality is high. The reason for this result is that the design method uses BP neural network to compress the original video and determine the implementation mode and cache location of VR short video cache files; Convert and cache high bit rate VR short video streams to generate low bit rate video streams, which not only meets the needs of different network bandwidth, but also improves the PSNR value.

3) It can be seen from Fig. 8 that the video transcoding speed of this method is significantly higher than that of [5] and [6], and the maximum can reach 92%. There are two main reasons for this result: on the one hand, the new coding coefficients and drift errors in the conversion encoder structure are calculated mainly through look-up tables rather than complex multiplication and division operations, which greatly reduces the overhead of the converter and improves the speed of the converter; On the other hand, the improvement of the speed of fast transcoding structure also comes from the selective error compensation operation of video frames. If the error compensation operation is performed, the table lookup structure cannot be used to calculate new coding coefficients and drift errors. Instead, the table lookup structure can be used to perform corresponding operations. It can be seen that the adaptive drift error control method greatly reduces the complexity of video frame transcoding, and can also improve the speed of video transcoding.

VII. CONCLUSION

In order to meet the requirements of high quality and low delay in short video development and reduce network load, this paper proposes a personalized distribution algorithm of VR short video content based on artificial neural network. Firstly, the original video is compressed by BP neural network, and the implementation method and cache location of VR short video cache file are determined; Secondly, the cached high bit rate VR short video streams are converted to generate low bit rate video streams to meet the needs of different network bandwidth; Third, establish a video distribution model, and build a multicast distribution tree based on this model, that is, add some relay servers; Finally, the minimum allocation cost of VR short video is realized through the algorithm of minimizing the allocation cost of VR short video. The experimental results show that the video blocking times of this method are always less than six times, which effectively reduces the video blocking times. The PSNR value is high, the increase is large, up to 0.5, and the video transcoding rate is improved, up to 92%. The application effect of this method is fully verified, which shows that it has certain application value. However, in the process of experimental analysis, the design method has not been compared with other more methods, and there are still some deficiencies. In the next research process, supplementary research will be carried out to further improve the application effect of this method.

REFERENCES

[1] Cicco L D, Mascolo S, Palmisano V. QoE-driven Resource Allocation for Massive Video Distribution. Ad Hoc Networks, 2019, 89(6):170-176.

[2] Liu W, Zhang T. Joint incentive mechanism based on video content cooperative delivery in mobile edge computing. Application Research of Computers, 2021, 38(9):2803-2810.

[3] Li Y Y, Zhou J H. Integrated ICN Slice and Video Caching as a Transmission Service. Telecommunication Engineering, 2022, 62(4):510-516.

[4] Zhang S H, Luo Y W, Zhang M, Sun Q F, Yang X L. Research progress of mobile video prefetching for mobile content distribution networks. Telecommunications Science, 2021, 37(1):8-21.

[5] Dong W T, Li Z, Chen X. Online short video content distribution strategy based on federated learning. Journal of Computer Applications, 2021, 41(06):1551-1556.

[6] Wang R Y, Yang Y, Wu D P. QoE-aware video distribution mechanism in FiWi. Journal on Communications, 2018, 39(01):1-13.

[7] Chen X, Wang L, Sun Z W. Research on Internet of Vehicles D2D Video Distribution Technology Based on Network Coding. Computer Engineering, 2019, 45(02):58-63.

[8] Pereira S, Abreu E, Iakunin M, Cavaco A, Salgado R, Canhoto P. Method for solar resource assessment using numerical weather prediction and artificial neural network models based on typical meteorological data: Application to the south of Portugal. Solar Energy, 2022, 236(4):225-238.

[9] Altintas E, Tolon M, Karabuga A, Utlu Z. Application of artificial neural network model for forecast energy efficiency of the cryogenic liquefaction system in the meaning of sustainability. International Journal of Global Warming, 2021, 24(3/4):420-444.

[10] Chen W, Li X, Chen X, Xiong Y. Research on influence mechanism of running clothing fatigue based on BP neural network. Journal of Intelligent and Fuzzy Systems, 2020, 40(8):1-11.

[11] Cao Q, Zhou C. Adaptive Compression Method of Embedded Online Video Inter Frame Signal. Computer Simulation, 2021, 38(3):87-91.

[12] Kafiloglu S S, Gur G, Alagoz F. Cooperative Caching and Video Characteristics in D2D Edge Networks. IEEE Communications Letters, 2020, 24(11):2647-2651.

[13] Wu D, Liu Q, Wang H, Yang Q, Wang R. Cache Less for More: Exploiting Cooperative Video Caching and Delivery in D2D Communications. IEEE Transactions on Multimedia, 2019, 21(7):1788-1798.

[14] Yoon J, Banerjee S. Hardware-assisted, Low-cost Video Transcoding Solution in Wireless Networks. IEEE Transactions on Mobile Computing, 2020, 19(3):581-597.

[15] Ahn W, Nam S, Son M, Lee H, Choi S. End-to-end double JPEG detection with a 3D convolutional network in the DCT domain. Electronics Letters, 2020, 56(2):82-85.

[16] Lv Y, Zhu B, Wu D P. Video multipath transmission mechanism with load balancing in the FiWi network. Journal of Xidian University, 2021, 48(3):63-70.

Modeling and Simulation of a Blockchain Consensus for IoT Node Data Validation

Bismark Tei Asare[1], Laurent Nana[2], Kester Quist-Aphetsi[3]
Lab-STICC, CNRS UMR 6285, F-29200, Computer Science Dept, GCTU, Ghana
Cyber Security Division, CRITAC, Directorate of Info. Assurance & Intelligence Research, CRITAC, Ghana
Université de Bretagne Occidentale, UBO, Brest, France[1]
Lab-STICC, CNRS UMR 6285, F-29200, Université de Bretagne Occidentale, UBO, Brest, France[2]
Computer Science Dept, GCTU, Cyber Security Division, CRITAC, Ghana
Directorate of Info. Assurance & Intelligence Research, CRITAC, Ghana[3]

***Abstract*—The classical blockchain developed for the Bitcoin cryptocurrency has evolved since its introduction more than a decade ago. Blockchain exists in different forms for different purposes and operational contexts. There has been a significant growth in the business use cases of blockchain which is based on the unique attributes of the distributed ledger technology. Blockchain provides peer-to-peer distribution of data in a traceable and decentralized architecture that attains data authentication using consensus protocols. Blockchain as a distributed ledger is the fusion of cryptography, peer-to-peer networking technology, distributed system technology, and consensus mechanism to assure information security and digital asset management. Consensus mechanisms are applied to the distributed ledger that operates in a peer-to-peer network where message transmission between peers is validated and stored across all active peers. Reaching an agreement to validate message transmission and maintaining the correctness of the state of data in a network for critical wireless sensor networks have become a necessary requirement for networks that span several subsystems covering a large operational area. Due to the resource constrained nature of the active actors of wireless sensor networks, any cryptographic solution to be adopted must be lightweight and efficient as well. This paper proposes a blockchain-based decentralized mechanism for authentication of node data for storage onto a distributed ledger. The coloured Petri net was used to model and simulate by detailing the critical attributes of the workings of the system that is based on cyber-physical IoT architecture.**

Keywords—Blockchain consensus; ripple consensus algorithm; coloured petri net; cyber-physical system; IoT architecture; node data security

I. Introduction

Reaching agreements to validate the authenticity of node data and subsequent transmission and storage of such network resources for cyber-physical systems have been a challenging and interesting domain for academia and information security industry players in recent times. Distributed ledgers use consensus algorithms to reach agreement among all connected active nodes to validate message transmission in a peer-to-peer approach. Recent advancements in connectivity, artificial intelligence, machine learning approaches although have provided an advantage for the expansion of network coverage and prediction and visualization of network resource sharing for enterprises and institutions, these available passive technologies in the hands of bad actors and hackers could render sophisticated cyber-attack exploitations to networks and user accounts resulting in breaching data, corrupting data, and compromising the security of such network systems [1].

Reaching an agreement to validate messages as well as authenticate the state of a distributed ledger have been the requirement for networks whose major components are resource constrained.

Distributed consensus research has become popular since the Nakamoto Satoshi introduced the Bitcoin blockchain cryptocurrency more than a decade ago. Internet of things have provided a platform for expanding the network resources to secure a new value proposition for scaling the scope of an enterprise’s network. Available reports support that the technological and cybersecurity budgets for business and institutions have increased allocations and are making efforts to include internet-of-things integration, expansion, and management [2].

Cybersecurity investments and the annual budgetary allocations across most enterprises have increased largely due to the increasing number of cyber-attack incidences on enterprise systems which have resulted in data corruption, data theft and huge revenue losses in some cases [3].

During the COVID-19 lockdown period, most businesses and institutions adopted telecommuting as a measure to regulate and manage people in observing physical distancing to avoid possible person-to-person infection. There were reported cases of cyber-attacks during this period where personal and enterprise data were breached as a result [4].

Cascading effect of cyber-attacks on heterogeneous systems for wireless sensor networks like the internet of things has a wide and costly impact due to the critical messages that the resource-constrained devices in such networks transmit [5].

The effect of cyber-attacks on heterogeneous wireless sensor networks resulting in the compromise of critical data in enterprise networks have taken an alarming trajectory due to the complexity of the interconnectedness of the components of the subsystems that make up the internet of things architecture [6].

Availability of pervasive applications and their integration in enterprise networks that have most of its component relying on internet of things architecture could make the security management of such wireless sensor systems a complex challenge to undertake.

A consensus algorithm for decentralized authentication and distributed ledgers for an IoT with heterogeneous system architecture requires a blockchain-based agreement mechanism that operates with relatively less energy, fully scalable and most importantly byzantine fault-tolerant [7].

Wireless sensor networks achieve privacy and integrity for message transmissions using either third-party trust enforcement systems that adopt a centralized entity to authenticate devices and validate messages or a decentralized mechanism for authentication of devices and validating messages. The centralized authentication mechanism is prone to several attacks including single point of failure attacks. In a distributed system that rely on decentralized authentication mechanism, agreement is reached by all active nodes on the network using consensus. Since the introduction of Bitcoin cryptocurrency, several consensus algorithms have been developed. The Bitcoin cryptocurrency uses a distributed consensus mechanism that is based on proof of work [8].

The Coloured Petri Net (CPN) is a modeling and simulation tool for modeling and simulating systems, and verifying their properties (real-time, behavioral, security properties …).

The CPN modeling, simulation, and validation of critical security properties of an efficient blockchain-based consensus mechanism that does not compromise the security requirements of a cryptographic solution and offers low latency with improved resistance to the Byzantine fault tolerance is presented by this paper.

The rest of the paper is structured in sections and represented as follows. In Section II, related work describing the state of the art for CPN in modeling and simulating security protocols for networks, node data security and critical security challenges in cyber-physical systems is presented. In Section III, the Ripple consensus algorithm is described. Section IV outlines the implementation of the consensus algorithm in establishing agreements for storing messages on the distributed ledger. Section V concludes the paper.

II. Related Work

A. Consensus Algorithm

In a stand-alone system, validating transmitted message or any transaction is vested in a dedicated centralized node. Consensus algorithms are useful in networks that do not have a dedicated node to singularly authenticate users, processes and transmitted messages or transactions. A consensus algorithm was employed by the system to agree on a single data among multiple processes and agents. To ensure that situations of some multi-agents failing to agree or be unreliable by not being available for consensus to actively reach agreement, a consensus protocol must adopt mechanisms to make them flexible and fault-tolerant [9].

In [10] a decentralized multi-agent system achieved consensus using consensus problem to control these multi-agents. More than half of all the multi-agents and processes agreed by voting on the state and integrity of a process.

B. Blockchain Consensus Algorithm

Blockchain as a cryptocurrency framework for Bitcoin has evolved since its introduction. Blockchains are uniquely categorized based on the type of consensus algorithm in use. In a Proof-of-Work (PoW) consensus, agreement on the validity of a process was achieved using the computing power challenge. The node that had more computing power achieved consensus through a completely decentralized approach. There is Proof-of-Stake (PoS) consensus that is based on financial power competition where the node that controls more than a third of all the resources within the network gets to validate processes within the blockchain. PoS operates by selecting validators to authenticate transactions within the blockchain based on the quantity of the cryptocurrency holdings forming a stake by a node. The more stake a node possesses the higher the chance of being selected to validate transactions. In a PoS, less computational power is involved since it takes shorter time to reach consensus than in PoW. In Delegated Proof-of-Stake (DPoS), consensus is reached based on election and voting process to guard malicious usage and centralization of blockchain. In DPoS, less computing power and time are involved in achieving consensus [11].

A consensus mechanism must provide a trade-off between performance, fairness, and security. There is Proof-of-Activity (PoA) that is a fusion of PoW and PoS. The PoA operated on an economic phenomenon with the assumption of “Tragedy of the Commons” which described a situation where a limited resource for several agents could be ruined in situations where there is uncontrolled use [12].

C. Ripple Consensus Algorithm

The Ripple consensus algorithm is a permissioned blockchain consensus algorithm that requires access permission for nodes in the network because it is not publicly accessible, and operates in rounds using active nodes as servers. It adopts an approach of closing an active ledger updating session once a consensus is reached to store and maintain an identical state of the ledger on all active nodes. For each round within the ripple protocol consensus algorithm (RPCA) [13];

- End users of the server forwards all new transactions to each server. These valid transactions are compiled and made public in the form of a list to constitute the “candidate set”.
- All the candidate sets from several servers are merged on every server’s unique node list (UNL), to authenticate these transactions.
- The transactions that do not pass the authentication minimal percentage of “Yes” votes are either discarded or included in the candidate set at the commencement of the consensus process for the next ledger. Conversely, transactions that obtain the minimal

percentage of "yes" votes are passed onto the next round of consensus.

The minimum percentage of 80% of a server's UNL is a requirement for agreeing on a transaction and that constitutes the final round of consensus. The final round of transaction closes the ledger after appending the authenticated transactions onto the ledger.

1) Composition of the ripple consensus protocol: The Ripple Consensus Protocol consists of several components: Server, Ledger, Last-closed ledger, Open-ledger, Unique node list (UNL), Proposer.

The server is an entity that runs the Ripple server software.

Ledger is an append only record of the amount of currency in each user's account and represents the ground truth of the network. The ledger grows with updating transactions using the consensus protocol.

The last-closed ledger describes the most recent state of the ledger after the consensus protocol has validated transactions and appended the validated transaction onto the ledger.

The Open ledger is used to represent the current operating status of a ledger on a node.

2) Correctness of the consensus: There is the likelihood of a validating node being compromised to form a cartel of corrupt validating nodes to comprise the byzantine-fault-tolerance integrity of the consensus. The ripple consensus protocol maintains correctness for agreements and resistive to Byzantine failures by adopting a mechanism where a transaction is approved only when 80% of the validating nodes agree using the consensus algorithm. Dishonest agreement to validate a message transmission is possible only after the number of faulty validating nodes exceed 80% of the unique node list. The consensus protocol with honest nodes in the UNL will maintain correctness if the unique node list UNL of n nodes in the network meets this condition:

$$f \leq (n-1) / 5$$

Where f is the number of Byzantine failures. In situations of $(n-1)/5 + 1$ Byzantine failures, the correctness of the consensus is maintained. The consensus will only confirm a fraudulent transaction when there are $(4n+1)/5$ failures or more. The probability of occurrence p^* hereafter, points to the likelihood of growing the size of the nefarious cartel below the maximal threshold of Byzantine failures.

$$p^* = \sum_{i=0}^{\left\lfloor\left(\frac{n-1}{5}\right)\right\rfloor} \binom{n}{i} p_c^i \, (1-p_c)^{n-1}$$

Where p_c denotes the probability of any node colluding with other nefarious cartel.

In Fig. 1, the process for the Ripple consensus is outlined.

The validation nodes are IoT sink nodes from several local IoT networks.

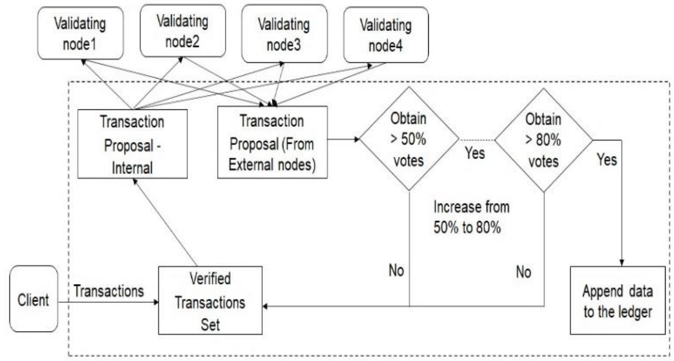


Fig. 1. Ripple consensus data flow diagram

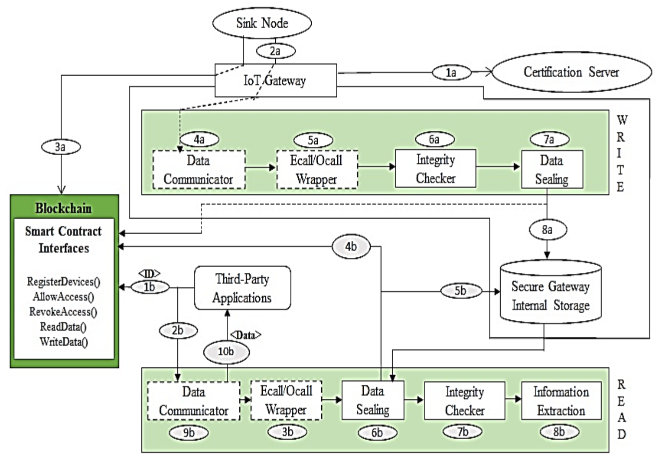


Fig. 2. Smart contract dataflow diagram

In Fig. 2, the smart contract data flow is presented. The smart contract operates between the sink node and the cloud network. The IoT gateway runs the blockchain smart contract. The smart contract ensures that users, devices, and data are verified and validated for data storage operations on the distributed ledger. Data from the sink node get stored onto the distributed ledger through the IoT gateway. The sink node registers itself on the blockchain in step 3a. The IoT gateway always verifies the state of the blockchain using the internal storage distributed ledger as input to validate the integrity of the blockchain using the cloud or remote. Connected sink nodes constitute the consensus nodes for performing user, device and data integrity checking before either writing onto the blockchain or access data from the blockchain. Data writing operations on the blockchain are done by the sink nodes, to append data onto the blockchain. They are referred to as data write operations in step 2a. The hash and encrypted data from the sink node are used in the next phase. The *writedata* function in the smart contract is used to append the hash of the sink node data onto the blockchain. The encrypted data is then written to the gateway internal memory in steps 4a-8a. The Ecall/Ocall wrapper communicates with the gateway internal memory as illustrated in the step 5a. The hash of the data from the sink node is verified by recalculating the hash-based message authentication code (HMAC) based on the encrypted and comparing the given hash with the derived hash. The Integrity Checker verifies and validates IoT data by ensuring that the given hash and the derived hash are the same, the encrypted data is sealed and written to disk in step 7a. If the report from the Integrity Checker shows a difference in the string structure of the derived hash from the given hash, that will result in discarding the data including the hash from the sink node. Step 7a and Step 8a are used in validating the hash and proceeding to either write the encrypted data to disk or disproving the hash and discarding the data from the sink node.

Data accessing activities from the blockchain is done using the data read module. A user module first registers third-party users using the *allowAccess* method with the smart contract. The user calls the *revokeAccess* function to revoke access for a user. Step 1b outlines the interaction of the third-party user with the smart contract in obtaining the hash of the data generated by the sink node after providing the device ID of the sink node. The smart contract checks if the third-party user device ID and the address have the validation necessary to access the data after doing integrity checking for the third-party user ID and address. The hash of the sink node data is only returned from the cloud storage after the integrity checker grants the access permission to the third-party user to enable it to access the data from the IoT gateway persistent storage (IoT gateway internal memory) that represents local storage of the data. The smart contract uses the *READDATA* API as illustrated in step 4b, to confirm if the third-party user has the access permission to read the data hash identifier supplied by the third-party request. In step 5b, it illustrates how data is retrieved from the secured internal gateway storage once data access permission is granted. The data is unsealed in step 6b, and the integrity of the data is checked in step 7b, after recalculating and verifying the digital signature by comparing the given and the derived digital signatures. The sensor data stored in the gateway internal memory is read and returned by the user only after the digital signature verification is completed. Steps 9 and 10 illustrate the data flow for this operation [14].

In Table I, the pseudo-code for the smart contract is presented.

Where:

OwnerAddress: Sink node identity (SNId)

Device: Sensor

DeviceID: SensorID (SsId)

In [15] five main blockchain consensus protocols were examined using the unique properties of type (probabilistic or absolute finality), level of fault-tolerance, power consumption, scalability, and application. The five consensus protocols are: Proof-of-Work (PoW), Proof-of-Stake (PoS), Delegated Proof-of-Stake (DPoS), Practical Byzantine Fault Tolerance (PBFT) and Ripple.

The ripple consensus proved to have a good scalability, involved negligible power consumption, low fault tolerance and operated using permissioned application. In ripple consensus, the entire network will continue to function to support correct consensus even if 20% of the nodes are attacked by Byzantine generals problem [16].

TABLE I. SMART CONTRACT PSEUDOCODE

Algorithm: Smart Contract Pseudo-code

```
HashMap deviceRegistry(key:ownerAddress, value:List[deviceIDs])
HashMap deviceData(key:(ownerAddress, deviceID),
value:List[DataHash])
HashMap DataAccessRegistry(key:(ownerAddress, thirdpartyAddress,
deviceID), value: bool isAllowed)
function REGISTERDEVICE(ownerAddress, deviceID)
  InsertToHashMap(key:ownerAddress, value:List[deviceIDs])
end function
function WRITEDATA(ownerAddress, deviceID, Data)
  if owner == ownerAddress
  deviceData([owner, deviceID], List.InsertData(hash(Data)))
end function
function READDATA(ownerAddress, thirdPartyAddress, deviceID)
if DataAccessRegistry(thirdPartyAddress) == true
  return deviceData[hash(ownerAddress, deviceID])
end function
function GRANTACCESS(ownerAddress, thirdPartyAddress,
deviceID)
  if owner == ownerAddress
  DataAccessRegistry[hash(ownerAddress, thirdPartyAddress,
deviceID)] = true
end function
  function REVOKEACCESS(ownerAddress, thirdPartyAddress,
deviceID)
  if owner == owner Address
  DataAccessRegistry[hash(ownerAddress, thirdPartyAddress,
deviceID)] = false
end function
```

D. Modeling Languages for Verification Systems

Modeling systems exist to provide opportunity for designing, developing, and implementing critical systems. Although there are several kinds of tools and platforms for modeling, simulation, and verification of systems, the coloured petri nets (CPN) is distinguishable in the following aspects: CPN offers several functions and provides a flexible manipulation of the functions in developing a model. The CPN tool has been improved and tested to support the modeling of complex systems [17].

Study [18] surveyed several modeling tools for checking, validating, and some cases improving the design requirements of systems. Notable modeling systems mentioned included the Practical Robust Implementation and Sustainability Model (PRISM), Numeric Symbolic Model Verifier (NuSMV), UPPAAL, Symbolic Analysis Laboratory (SAL), SPIN, Beryl, D-Finder.

The related works showed available research on IoT solutions that is based on blockchain. It however confirmed the absence of an implementation for a blockchain-based solution to authenticate and protect IoT data transmitted between actively connected network elements of the IoT gateway and the cloud.

Additionally, there is an implementation gap for a formal model for IoT systems that directly involved the sensor, sink node, IoT gateway, and the cloud elements in a blockchain-based IoT architecture.

In the next section, the methodology for the paper is presented.

III. Methodology

The target and design principles that formed the basis for the chosen methodology to support a blockchain-based consensus mechanism for authentication of node data for IoT systems are hinged on a security solution appropriate for an environment where the devices are resource-constrained.

The design principles for the methodology are the usage of decentralized authentication, smart contract for consensus among sink nodes, lightweight cryptographic solution, digital signature, smart contract with lightweight cryptographic function, a formal modeling tool that allows for dynamic behavior modeling, and the provision of a visual simulation tool.

The design goals on the other hand involved the elimination of a single point of failure, stronger security, extending data protection with a lightweight cryptographic solution, enforcing authentication with smart contract, the proposal of a formal model for a generic blockchain-based IoT solution, and the validation of a blockchain-based IoT solution through simulation.

A. Heterogeneous IoT Architecture

An architecture consisting of several subsystems was adopted for the implementation of the blockchain-based authentication mechanism.

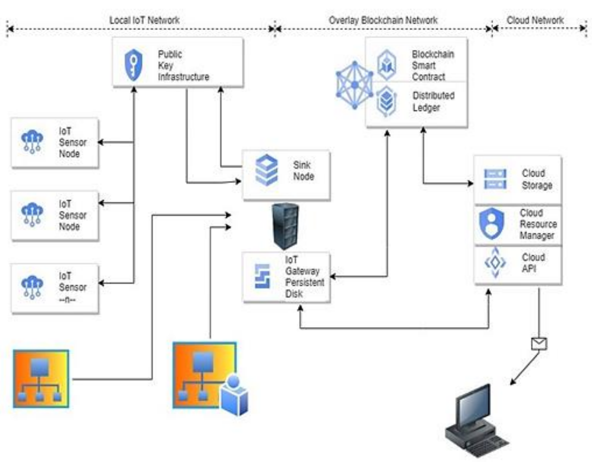


Fig. 3. An IoT architecture

In Fig. 3, an IoT architecture with components for a heterogeneous cyber-physical network is displayed. The architecture has three components involving a local IoT network which consisted of sensor end devices and sink nodes, an overlay network that employs blockchain-based distributed ledger, and a cloud network (remote storage) to receive and store the hash values of the sensor data. The local IoT networks amalgamate validated sink nodes with their validated data and transmit them through an IoT gateway to be stored on the distributed ledger [19].

The local IoT sub model is composed of elements that consist of security management, devices and sensors, internet connectivity things, Application Programming Interface (API) libraries, System Development Kits - SDK. Distributed systems that operate based on a decentralized authentication mechanism is prone to attacks such as the double spending instances where validated messages that represent independent transactions have the possibility of getting used in simultaneous transfers without considering the output of each transfer in the simultaneous transactions [20].

B. Coloured Petri Net

The coloured Petri Net (CPN) is a graphical mathematical modeling language. It is used to describe and check system properties, security requirements and synchronization characteristics for real-time distributed systems, and more generally event-driven systems. CPN comprises essential tools for analyzing boundedness, reachability, resource conflicts, deadlock as well as the structural properties of a real-time system [21].

The formalization of CPN is composed of nine tuples.

$$CPN = (P, T, A, \sum, V, C, G, E, I)$$

Where:

P = {P1, P2, …, Pm} represents a finite set of places.

T = {T1, T2, …, Tn} denotes a finite set of transitions.

A: Directed arc set

$\sum$: A finite set of colour set types

V: Denotes a finite set of variables whose type $\in \sum$.

C: It represents the colour set function from P to $\sum$.

G: Denotes the set of guard functions of transitions.

E: It represents a function that associates an arc expression to each transition.

I: denotes the function that gives the initial marking of each place.

The graphical representation of Petri net comprises of rings representing Places, rectangles denoting Transitions, arrows symbolizing Arcs.

A coloured Petri net is composed of variables, values, and expressions. CPN objects are described using colour domain that comprises variables, data values, operators, a syntax for expressions, and typing rules. An abstract colour domain consists of : Data values $\mathbb{D}$, Variables $\mathbb{V}$ and Expressions ($\mathbb{E}$) [22].

- $\mathbb{D}$ *is the set of data values*; These data values include integer values, Boolean values (True and False), and special undefined value $\perp$;
- $\mathbb{V}$ *is a set of variables*, that are represented using single letters $x, y, \ldots$, or as subscribed letters $x_1, y_k, \ldots$

- $\mathbb{E}$ *is the set of expressions*, that are composed of values, variables, and suitable operators.

Variables or values may form a basic expression. Thus, $D \cup V \subset E$. For example, let $e \in E$, the expression $vars(e)$ denotes the set of variables from V involved in e.

A $binding$ is a restricted function $\beta : \mathbb{V} \to \mathbb{D}$. Let $e \in \mathbb{E}$ and β be a binding. $\beta(e)$ represents the evaluation of e under β; if the domain of β does not include $vars(e)$ then $\beta(e) \stackrel{\text{def}}{=} \perp$.

Both sets and multisets of expressions are subjected to binding evaluations.

For example, if $\beta \stackrel{\text{def}}{=} \{x \mapsto 1, y \longmapsto 2\}$, we have $\beta\,(x + y) = 3$. With $\beta \stackrel{\text{def}}{=} \{x \mapsto 1, y \mapsto 2\}$, according to the colour domain, we may have $\beta(x + y) = \perp$ (no coercion), or $\beta(x + y) =$ "12" (coercion of integer 1 to string "1"), or $\beta(x + y) = 3$ (coercion of string "2" to integer 2), or even other values as defined by the concrete colour domain.

Two expressions $e_1, e_2 \in \mathbb{E}$ are said to be equivalent which is represented as $e_1 \equiv e_2$, if for all possible binding β, the binding for both expressions are the same $\beta(e_1) = \beta(e_2)$. For example, $x + 1, 1 + x \; and \; 2 + x - 1$ are pairwaise equivalent expressions for the usual integer arithmetic.

Definition 1 (Petri nets). A Petri net is a tuple with several elements such as $(S, T, l\,)$ where:

- $S\ is\ the\ finite\ set\ of\ places$; S is also represented as P
- $T, disjoint\ from\ S, is\ the\ finite\ set\ of\ transitions$;
- $l\ is\ a\ labelling\ function\ such\ that$:

$for\ all\ s \in S, l(s) \subseteq \mathbb{D}\ is\ the\ type\ of\ s, i.e., the\ set\ of\ values\ that\ s$ is allowed to carry,

$for\ all\ t\ \in T,\ l(t) \in \mathbb{E}\ is\ the\ guard\ of\ t, i.e.,$ a condition for its execution,

$for\ all\ (x, y) \in (S\ \times T) \cup (T \times S), l(x, y)\ is\ a\ multiset\ over\ \mathbb{E}$ and defines the arc from $x\ toward\ y$.

Definition 2 (Markings and Sequential Semantics) Let $N \stackrel{\text{def}}{=} (S, T, l) be\ a\ Petri\ net.$

A marking M and N is a $function\ on\ S\ that\ maps\ each\ place\ s$ to a $finite\ multiset\ over\ l(s)\ representing\ the\ tokens\ held\ by\ s.$

$A\ transition\ t \in T$ is enabled at $a\ marking\ M\ and\ a\ binding\ \beta, which\ is\ denoted\ by\ M[t, \beta\rangle, iff\ the\ following\ conditions\ hold$:

- $M\ has\ enough\ tokens, i.e., for\ all\ s \in S, \beta\big(l(s,t)\big) \leq M$;
- $the\ guard\ is\ satisfied, i.e., \beta\big(l(t)\big) =$ True
- $place\ types\ are\ respected, i.e., for\ all\ s \in S, \beta\big(l(t,s)\big)\ is\ a\ multiset\ over\ l(s).$

$If\ t \in T\ is\ enabled\ at\ marking\ M\ and\ binding\ \beta, then\ t\ may\ fire\ and\ yield$ a marking $M'\ defined\ for\ all\ s \in S\ as\ M'(s) \stackrel{\text{def}}{=} M(s) - \beta\big(l(s,t)\big) +\ \beta\big(l(t,s)\big).$
$This\ is\ denoted\ by\ M[t, \beta\rangle\ M'.$

The marking graph $G\ of\ a\ Petri\ net\ marked\ with\ M\ is\ the\ smallest\ labelled\ graph$ such that:

- $M\ is\ a\ node\ of\ G$
- $if\ M'is\ a\ node\ of\ G\ and\ M'[t, \beta\rangle M''is\ also\ a\ node\ of\ G\ and\ there\ is\ an\ arc\ in\ G\ from\ M'\ to\ M''\ labelled\ by\ (t, \beta).$

The definition of marking graphs allows the addition of infinitely many arcs between two markings. If $M[t, \beta\rangle$, there might exist infinitely many other enabling bindings that differ from β only on variables not involved in t. Finally only firings $M[t, \beta\rangle$ such that the domain of $\beta\ is\ vars(t) \stackrel{\text{def}}{=} vars(l(t)) \cup \bigcup_{s \in S}(vars\big(l(s,t)\big) \cup vars\big(l(t,s)\big))$ is considered.

In the next section, the results and discussion for the proposed methodology is presented.

IV. Results and Discussion

The use of the mathematical and visualization features of CPN provided the set of mathematical foundations necessary for representing and describing the elements and security properties of the proposed solution that is based on the IoT architecture.

In this section, the simulation results for the blockchain consensus using CPN tools are presented and discussed.

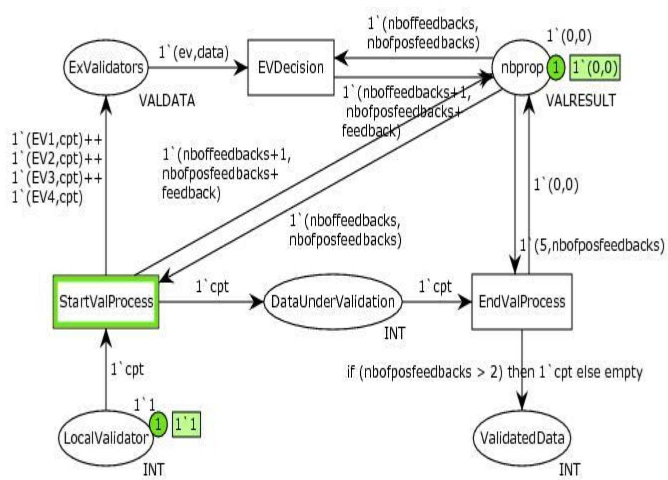


Fig. 4. CPN modeling of the blockchain consensus protocol

In Fig. 4, the CPN modeling of the blockchain consensus mechanism with validators is presented. The validators are an amalgamation of the current sink node which presents the data to be validated as well as other external sink nodes within the hierarchical IoT network. The external sink nodes form the external validators. The sink node whose data is to be validated through the consensus assumes the local validator status. The validators (local and external) reach agreement to validate data if and only if the number of the positive feedbacks are more than half of all the total decisions from the validation voting by all the validators. Once a message does not get at least more than half of the total decisions to be positive feedbacks, that message is discarded. A session for the consensus by the validators is considered closed once the decision on a message has been made in accordance with the consensus correctness criterion of the adopted blockchain consensus. The correctness criterion of the consensus is critical to make the algorithm byzantine fault tolerant.

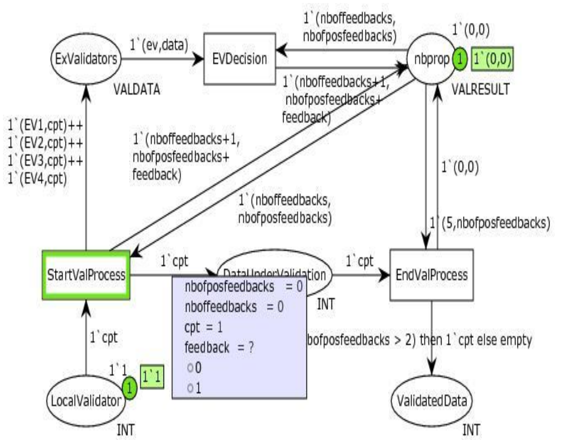


Fig. 5. Local validator feedback decision

In Fig. 5, the consensus commencement for validators is illustrated. The blockchain consensus starting with the local validator to decide by voting by on the feedback of the data under validation is presented. There are two feedback options (0, 1) to be selected by a validator. Option 1 symbolizes positive feedback whereas option 0 denotes non-positive feedback. The "start validation transition" has not been fired yet. Selecting a choice for the feedback will fire the transition. The token (1`1) on the local validator symbolizes a single node data and the specific data to be validated is 1. The update on the number of proposals "nbprop" of 1`(0, 0) shows that voting on the decision feedback on the data under validation has not started (0, 0).

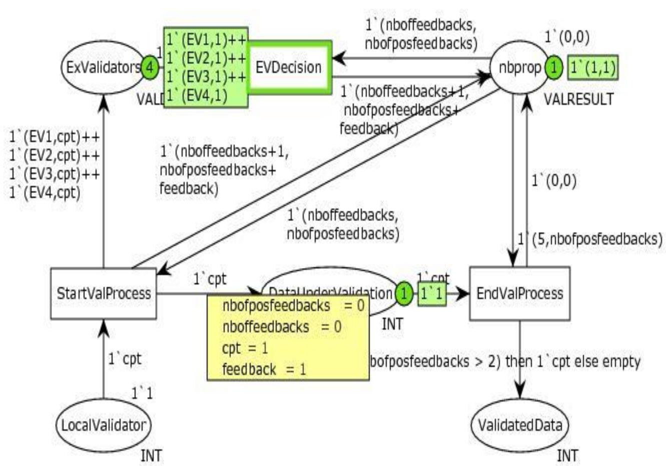


Fig. 6. Local validator feedback decision proposal update

In Fig. 6, a feedback decision of '1' on the data under validation is presented. The feedback from the local validator confirms that the "start validation transition" has been fired. The update on the number of proposals "nbprop" of 1`(1, 1) shows that voting has started on the decision feedback on the data under validation. That only 1 validator has voted on the decision feedback. That decision is a positive decision (1 – 'Number of decisions',1-'number of positive decisions').

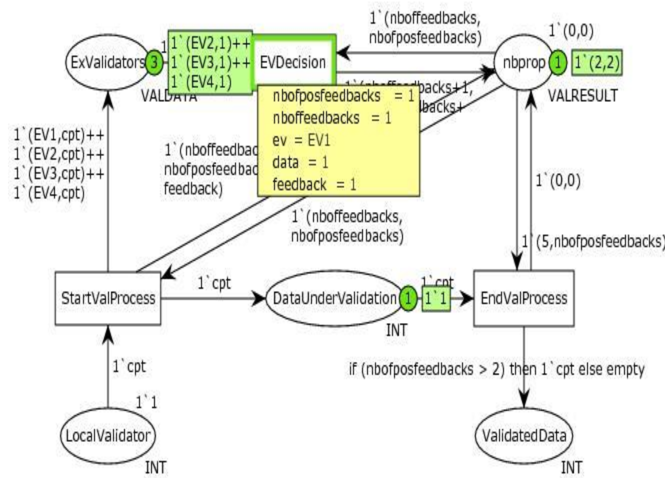


Fig. 7. External validator1 feedback decision

In Fig. 7, the feedback decision from external validator1 on the data under validation is illustrated.

The update on the nbprop place 1`(2,2) shows that there have been two voting decisions and all the decision are positive decisions.

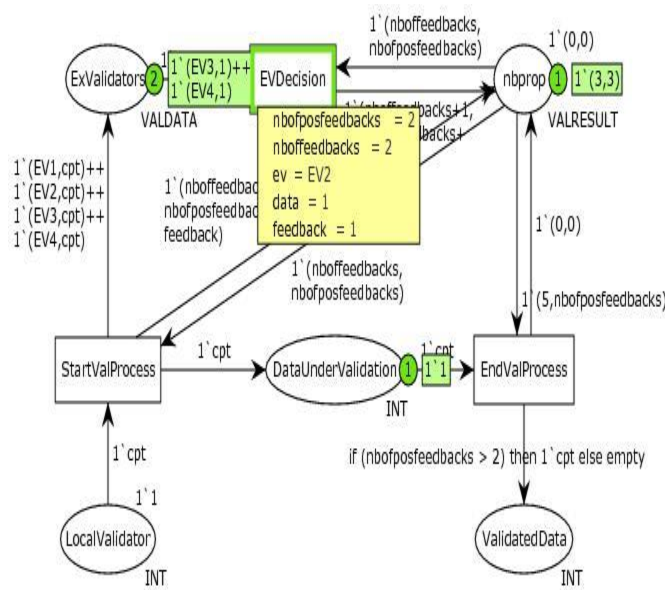


Fig. 8. External validator2 feedback decision proposal update

In Fig. 8, the feedback decision from external validator2 on the data under validation is shown.

The decision feedback voting update on the number of proposals "nbprop" – 1`(3,3) shows that there have been three feedback decisions with all three being positive feedback decisions.

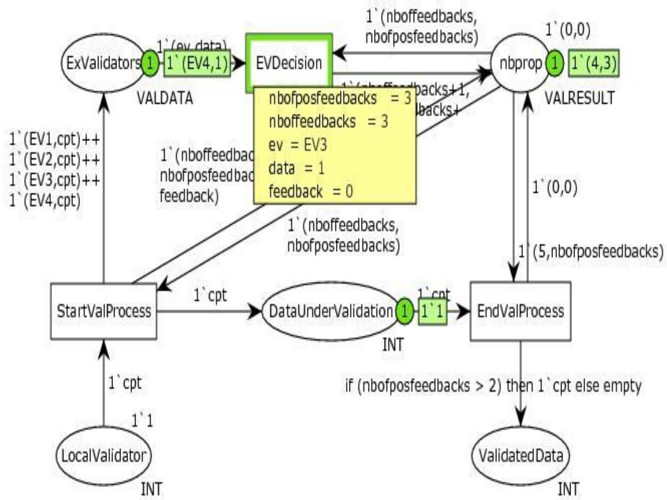


Fig. 9. External validator3 feedback decision proposal update

In Fig. 9, the feedback decision from external validator3 on the data under validation is illustrated. The local validator, external validators 1, 2, 3 have all voted on the decision and have the feedback updated and stored on the nbprop place. The token value has been updated to 1`(4,3) to show that there have been four votes (local validator, external validators 1,2,3). And that three out of the four votes are positive feedback decisions.

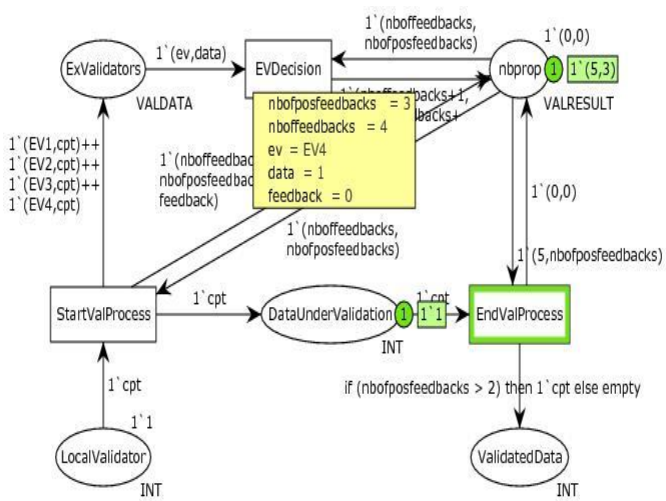


Fig. 10. External validator4 feedback decision proposal update

In Fig. 10, the feedback decision from external validator4 on the data under validation is depicted. Additionally, it provides the update as illustrated in the place for the number of proposals "nbprop" for a total of five decisions, with three positive feedback decisions. The EndValProcess transition is highlighted to show that it is the next action or step to be taken for the simulation.

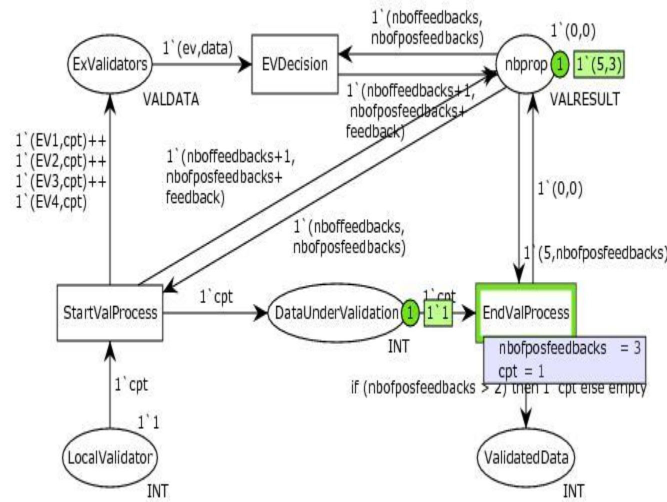


Fig. 11. Decision result after the proposals

In Fig. 11, the data flow CPN simulation on the decision feedback results at the end of the decision voting process is represented. The token value on the nbprop 1`(5,3) and the summary information on the transition confirm that there was one data identity that represented 1 data element to be validated and that there were 3 positive decision feedbacks.

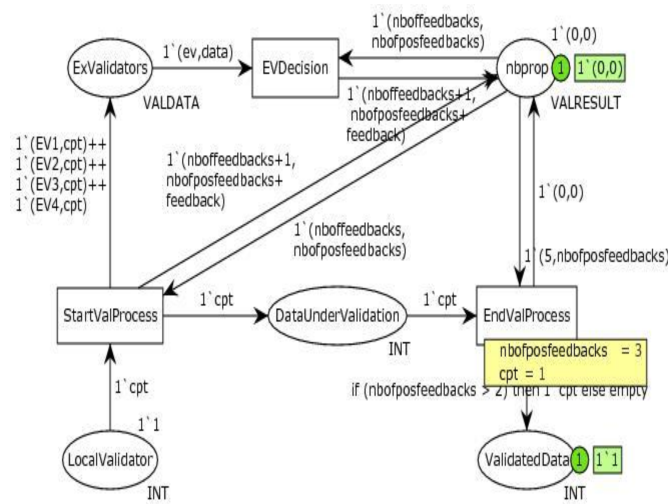


Fig. 12. Consensus decision

In Fig. 12, the decision on the data at the end of the consensus process is shown. The initial token element 1`1 on the local validator place has been moved to the place for the ValidatedData. Additionally, the consensus session is closed and the nbprop token element is reset to 1`(0,0).

The firing of the EndValProcess transition ends and session for the consensus activity. The data is then validated and the colour token 1`1 on the ValidatedData place finalizes the consensus.

TABLE II. DESCRIPTION OF CPN MODELING FEATURES FOR THE BLOCKCHAIN CONSENSUS MECHANISM

Abbreviation	Description	CPN Component
LocalValidator	It represents an internal sink node. The container for keeping a sink node data prior to a validation operation	Places
ExValidators	It denotes external sink nodes that form the external validators. These are all the other sink nodes within the hierarchical IoT network. They join the internal validator to reach an agreement on a message through a blockchain consensus.	
nbprop	A container for keeping all the decisions resulting from validators using the consensus rule to vote on a data under validation.	
ValidatedData	It represents the results after the voting decisions undertaken by all the validators have ended. When the number of positive feedbacks where at least 60% of the total decisions by the validators, the data will be moved to a new state of ValidatedData.	
DataUnderValidation	It denotes a place that specifies the current data being validated is kept. It is represented by the identity of the data which is captured as (`cpt) on the arc inscription.	
StartValProcess	It is an event that signifies the start of the consensus session. The local validator is an input to this event. It fires the data from the local validator to the external validators as well as updates the DataUnderValidation and the “nbprop” places.	Transitions
EVDecision	It is a transition label for the CPN event that fires the decision of each external validator as feedback on a data. The input of the transition is the external validator and the data to be validated based on the consensus rules. The output for this transition is the number of decision feedbacks and the number of positive decision feedbacks.	
EndValProcess	It is a transition to signify the close of a consensus session. It has DataUnderValidation, and nbprop as input. The output of this transition is the ValidatedData.	
cpt	A token for describing the identity of data under validation. The data under validation is submitted by the local validator to the validators where the consensus mechanism is applied on the data using other established rules in the consensus to vote on the data in validating it.	Variables
EVi (i = 1 .. 4)	The token identifying the external validator i.	
data	A data element representing the data under validation by the external validator.	
nboffeedbacks	It is a counter that records the decisions of voting activities by providing an update on the total number of feedback decisions	
nbofposfeedbacks	It is a counter that records the total number of positive feedback decisions. Both the number of feedbacks and the number of positive feedback decisions are stored as a token in the “nbprop” place and are updated each time an external validator decision is taken.	

In Table II, the CPN simulation components for the consensus mechanism for the system are presented. The components for the simulation consisted of CPN places, transitions, arc expressions, and initial marking of places using coloured tokens. The various components used in the simulation of the consensus mechanism were described in the table.

The use of the proposed IoT architecture is an improvement on a related work that used blockchain mechanisms for IoT data security. In [19] the blockchain solution did not indicate how the node data from the sensor was protected as well as an approach to maintain the integrity of the data communicated between the sensor and the sink node. The proposed blockchain-based IoT architecture used a centralized approach with a lightweight-cryptographic mechanism to protect the content of data between the sensor and the sink node. Additionally, the use of a non-monetary-based blockchain consensus mechanism where only the IoT gateway and other sink node clusters formed the consensus nodes is used to implement a smart contract with a lightweight cryptographic function for decentralized authentication of node data.

The use of the decentralized consensus ensured the elimination of a single point of failure situation for the IoT network and supported a distributed ledger that guaranteed the availability of validated node data on the IoT internal storage and the cloud for authorized users in the IoT system.

In the next section, the general conclusion of the work is presented.

V. Conclusion

Distributed systems that rely on decentralized processing for authentication and validation of processes like the blockchain system use agreements through consensus mechanisms to assure and maintain the correctness of decisions, and to guarantee stable systems. For a blockchain mechanism to be deployed in an IoT network where the devices are resource-constrained, an architecture was designed that factored in the challenges regarding memory, computational processing, and energy limitations of sensors and sink nodes. The distinctive security features in the proposed consensus mechanism enabled the core elements of the IoT architecture to reach an agreement among the sparsely fragmented network elements in the IoT architecture. The use of the IoT gateway, PKI, and cloud network in the architecture supported a security solution that provided a trade-off between performance, fairness in load balance, and security.

Additionally, the ripple consensus mechanism provided a byzantine fault-tolerant approach with good scalability to support a large network consisting of several subsystems.

The modeling and simulation of the blockchain-based authentication mechanism provided the possibility of validating the security properties of an IoT security solution that is based on a decentralized authentication approach. The CPN features including places, transitions, arcs, expressions, and initial markings of places with tokens were used to represent the entities or physical attributes of the system as well as the design decisions of the system. The design decisions and the dynamic nature expectations of the distributed ledger system were represented using places, transitions, arc expressions, and tokens.

References

[1] M. Keerthika and D. Shanmugapriya, “Wireless Sensor Networks: Active and Passive attacks-Vulnerabilities and Countermeasures,” Global Transitions Proceedings, vol. 2, no. 2, pp. 362–367, 2021.

[2] W. Fu, X. Wei, and S. Tong, “An improved blockchain consensus algorithm based on raft,” Arabian Journal for Science and Engineering, vol. 46, no. 9, pp. 8137–8149, 2021.

[3] I. Lee, “Cybersecurity: Risk management framework and investment cost analysis,” Business Horizons, vol. 64, no. 5, pp. 659–671, 2021.

[4] H. S. Lallie et al., “Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic,” Computers & Security, vol. 105, p. 102248, 2021.

[5] L. Dong, H. Xu, X. Wei, and X. Hu, “Security correction control of stochastic cyber–physical systems subject to false data injection attacks with heterogeneous effects,” ISA transactions, vol. 123, pp. 1–13, 2022.

[6] B. T. Asare, K. Quist-Aphetsi, L. Nana, and G. Simpson, “A nodal Authentication IoT Data Model for Heterogeneous Connected Sensor Nodes Within a Blockchain Network,” in 2021 International Conference on Cyber Security and Internet of Things (ICSIoT), 2021, pp. 65–71.

[7] S. Zhang and J.-H. Lee, “Analysis of the main consensus protocols of blockchain,” ICT express, vol. 6, no. 2, pp. 93–97, 2020.

[8] E. Zaghloul, T. Li, M. W. Mutka, and J. Ren, “Bitcoin and blockchain: Security and privacy,” IEEE Internet of Things Journal, vol. 7, no. 10, pp. 10288–10313, 2020.

[9] X. Wang, H. Su, X. Wang, and G. Chen, “Fully distributed event-triggered semiglobal consensus of multi-agent systems with input saturation,” IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 5055–5064, 2016.

[10] X. Wang, G.-P. Jiang, H. Su, and Z. Zeng, “Consensus-based distributed reduced-order observer design for LTI systems,” IEEE Transactions on Cybernetics, 2020.

[11] J. Zhang and M. Wu, “Blockchain-Based Authentication with Optional Privacy Preservation for Internet of Vehicles,” Mathematical Problems in Engineering, vol. 2021, 2021.

[12] D. Wang, C. Jin, H. Li, and M. Perkowski, “Proof of Activity Consensus Algorithm Based on Credit Reward Mechanism,” in Web Information Systems and Applications: 17th International Conference, WISA 2020, Guangzhou, China, September 23–25, 2020, Proceedings, Berlin, Heidelberg, 2020, pp. 618–628. doi: 10.1007/978-3-030-60029-7_55.

[13] D. Schwartz, N. Youngs, and A. Britto, The ripple protocol consensus algorithm. Ripple Labs Inc.(2014). 2021.

[14] G. Ayoade, V. Karande, L. Khan, and K. Hamlen, “Decentralized IoT data management using blockchain and trusted execution environment,” in 2018 IEEE International Conference on Information Reuse and Integration (IRI), 2018, pp. 15–22.

[15] S. Zhang and J.-H. Lee, “Analysis of the main consensus protocols of blockchain,” ICT Express, vol. 6, no. 2, pp. 93–97, Jun. 2020, doi: 10.1016/j.icte.2019.08.001.

[16] B. Chase and E. MacBrough, “Analysis of the XRP Ledger Consensus Protocol.” arXiv, Feb. 20, 2018. Accessed: Jun. 02, 2022. [Online]. Available: http://arxiv.org/abs/1802.07242

[17] H. Kaid, A. Al-Ahmari, Z. Li, and R. Davidrajuh, “Single controller-based colored Petri nets for deadlock control in automated manufacturing systems,” Processes, vol. 8, no. 1, p. 21, 2019.

[18] M. Luckcuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher, “Formal specification and verification of autonomous robotic systems: A survey,” ACM Computing Surveys (CSUR), vol. 52, no. 5, pp. 1–41, 2019.

[19] G. Wang, Z. Shi, M. Nixon, and S. Han, “ChainSplitter: Towards Blockchain-Based Industrial IoT Architecture for Supporting Hierarchical Storage,” 2019 IEEE International Conference on Blockchain (Blockchain), pp. 166–175, 2020, doi: 10.1109/blockchain.2019.00030.

[20] U. W. Chohan, “The double spending problem and cryptocurrencies,” Available at SSRN 3090174, 2021.

[21] W. Duo, H. Xin, and M. Xiaofeng, “Formal Analysis of Smart Contract Based on Colored Petri Nets,” IEEE Intell. Syst., vol. 35, no. 3, pp. 19–30, May 2020, doi: 10.1109/MIS.2020.2977594.

[22] C. Gaucherel, C. Carpentier, I. R. Geijzendorffer, C. Noûs, and F. Pommereau, “Discrete-event models for conservation assessment of integrated ecosystems,” Ecological Informatics, vol. 61, p. 101205, 2021.

Microcontrollers Programming Framework based on a V-like Programming Language

Fernando Martínez Santa[1], Santiago Orjuela Rivera[2], Fredy H. Martínez Sarmiento[3]
Universidad Distrital, Francisco José de Caldas, Bogotá, Colombia[1,3]
Corporación Nacional Unificada, de Educación Superior CUN, Bogotá, Colombia[2]

***Abstract*—This paper describes the design of a programming framework for microcontrollers specially the ones with low program and data memory, using as a base a programming language with modern features. The proposed programming framework is named *Aixt Project* and took inspiration from other similar projects such as *Arduino*, *Micropython* and *TinyGo* among others. The project's name is inspired on the weasel pet of the *V* programming language and at the same time it is a tribute to *Ticuna* people who live in the Amazon rain-forest, just between Colombia, Perú and Brasil. *Aixt* comes from *Aixtü* or *Aitü rü* which means otter in *Ticuna* language. The proposed programming framework has three main components: the *Aixt* language based on the *V* syntax, a transpiler that turns the defined V-like source code into *C*, and a generic cross-platform Application Programming Interface (API). The target of this project is obtaining a cross-platform programming framework over the same language modern language an the same API, for programming different microcontrollers especially the ones with low memory resources. *Aixt* language is based on the syntax of *V* programming language but it uses mutable variables by default. *V* language was selected to be used as base of this project due to it is a new compiled programming language with interesting modern features. In order to turn the *Aixt* source code into *C*, a transpiler is implemented using *Python* and the some specialized libraries to design each part of its translation process. The transpiled code is compiled by the native *C* compiler of each microcontroller to obtain the final binary file, that is why the API has to be adapted for each native *C* compiler. The complete project is released as a free and open source project. Finally, different application test were done over the XC8 and XC16 compilers for the PIC16, PIC18, PIC24 and dsPIC33 microcontrollers families, demonstrating the correct working of the overall framework. Those tests show that the use modern language framework to program any microcontrollers is perfectly feasible using the proposed programming framework.**

***Keywords*—Microcontroller; transpiler; API; programming language; V; V-lang; Aixt project**

I. Introduction

The different processor architectures used by the commercial microcontrollers, make the programming process dependent on those architectures and thus not universal. Even, when the microcontrollers are programmed on high level languages, tasks such as peripherals, timers, setup registers, and others, keep depending on the programmer's knowledge of the processor's architecture [1], [2]. There are some different projects which pretends to generate cross-platform programming frameworks [3], using different programming languages like *JavaScript* [4], and other implementations using virtual machines [5], [6], [7]. An example of those programming frameworks (and one of the most popular) is *Arduino*[8], [9], [10], which is based on *C* language in addition to an API which makes the programming process easier. That API works on a predefined hardware setup to reduce the setup process by the programmer. Another popular programming framework for microcontrollers is *Micropython* which implements on several devices a subset of Python language. *Micropython* has specific relatively high memory requirements which makes it impossible to run on small microcontrollers, but it has been ported to a large number of different architectures [11] mainly in internet of things IoT implementations. *Arduino* is compiled but its *C* syntax lacks modern features, on the other hand *Micropython* is interpreted and therefore non time optimized as compiled language, but there is an intermediate framework named *Tinygo* which implements *Go* language on Microcontrollers, offering modern features like *Python* and the advantage of being compiled [12] like *Arduino* (*C*). However, most of the microcontroller with limited memory features does not fit to the memory requirements of the projects previously described, so for those ones it is necessary to use their native *C* compiler.

In order to obtain the best execution times and the best code optimization level [13], [3] it is necessary to use the native *C* compiler of each architecture. Then, if there is a programming framework with an upper modern language layer, a transpiler to *C* and the native *C* compiler as a part of the framework, this could have high level language features along with optimization levels similar to the ones reached with only the native compilers. The described programming framework needs to have a transpiler [14], which is a translator from the upper layer language to the native *C* [15]. Transpilers are highly utilized nowadays [16], [17], in several languages both compiled and interpreted [18], [19], [20], and even in languages based on virtual machines [21]. Those transpilers are mainly used in order to reuse source code that comes from another different language [18], or improve the execution times or another performance feature of the program [22], [23] changing the platform or language (for instance turn *Python* (interpreted) into *Rust* (compiled) [19]), even translating source code to gate-based hardware [24] like FPGAs or other processor-less devices.

Several new programming languages have emerged nowadays, mainly to solve some of the issues of the traditional ones such as safety, memory management among others. Among these new languages are *Go*, *Swift*, *Dart*, *F#* and *Rust*, being this last is one of the most preferred ones[25], having even implementations on microcontrollers [26], [27], [28]. There are some other other languages such as *Peregrine* which is based on the *Python's* syntax and the *V* programming language [29] wish is inspired on *Rust* and other languages. *V* is an statically-typed programming language with several modern features that make the development easy, and a better learning curve than other modern languages like *Rust*.

This paper proposed a programming framework for micro-controllers that is composed by a high level language based on *V* as the main language, a transpiler from this *V-like* language named *Aixt* to *C*, and the microcontrollers' native *C* compiler which finally generates output binary file. In order to generalized the programs across the different microcontrollers, a general API is designed, which is implemented on each *C* compiler of the supported devices (in this first stage for XC8 and XC16 compilers). For the transpiler implementation *Python* and the module *SLY* were used, to write the lexer analyzer and the Parser. This project is based on a previous one named *Sokae* [30] developed by the same authors.

The paper is organized as follows: Section **II** presents the methodology for implementing the overall proposed programming framework, including the *Aixt* language definition (Section **II** - *A*), the *Python* implementation of the Aixt-to-C transpiler (Section **II** - *B*), and the API implementation for the XC16 compiler and PIC24 microcontrollers family (Section **II** - *C*). Section **III** shows the *Aixt* language functionality by implementing several examples, as well as it presents the results of implementing the proposed programming framework by several test source codes. Finally, Section **IV** shows the conclusions about this research's main ideas, including possible future jobs.

II. METHODOLOGY

With the name *Aixt Project*, a microcontroller programming framework is implemented. This framework uses an homonym language which is based on the *V* programming language. A transpiler from *Aixt* language to *C* is the most important block of this framework, as well as an Application Programming Interface (API) written in both languages. As part of the proposed structure, the native *C* compiler of the specific micro-controller finally generates the output binary file, as shown in Fig. 1. Using the proposed framework, the users will be able to write the source code in *Aixt* language using a standard API and obtains the binary file for a specific microcontroller or board without having further knowledge of the programming architecture. This framework pretends to be highly modular and relatively easy to include other microcontrollers or boards. The Fig. 1 shows the general structure of the programming framework indicating that for each new microcontroller to be supported it is necessary to adapt the API (Fig. 1 right) to this and invoke its specific native *C* compiler (Fig. 1 left down). The specific test done for this paper were implemented on some different Microchip® microcontroller families such as PIC16, PIC18, PIC24 and dsPIC33 using the XC8 and XC16 compilers, these microcontrollers were selected because their limited amount of implemented memory.

A. Aixt Language

Aixt is the name given to the proposed language and the overall programming framework. This language is based on the *V* programming language [29] and shares most of its syntax. Due to its relatively short learning curve, *V* language was

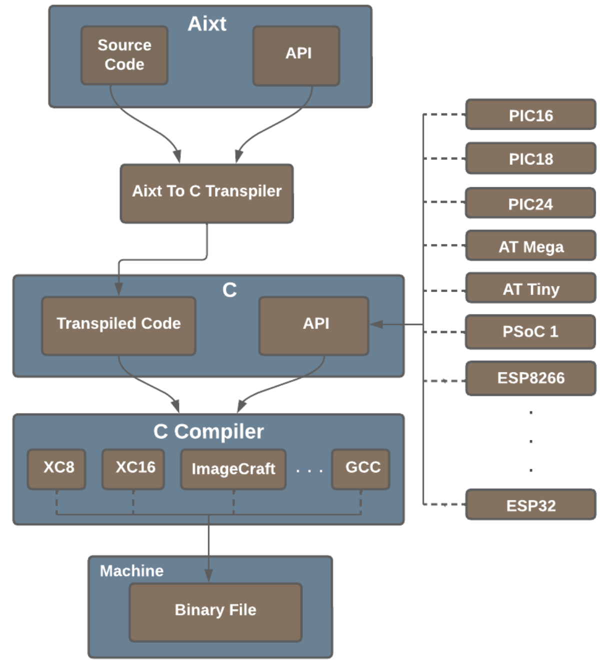


Fig. 1. General programming framework diagram

selected for this implementation instead other new languages like *Rust* [30]. The framework and language name is inspired in the Weasel pet of *V* Language, and at the same time is a tribute to *Ticuna* people who live in the Amazon rain-forest in the borders between Colombia, Brasil and Perú. Weasels are mustelids just like otters, so the name *Aixt* comes from *Aixtü* or *Aitü rü* which is a way to say otter in *Ticuna* language.

Aixt is a compiled and statically typed programming language based on the *V* syntax. This is designed to be used on a wide range of microcontrollers no matter their memory limitations. *Aixt* syntax shares some feature with languages such as *Rust* and *Go*, therefore also it shares syntax features with *C*, which makes *Aixt* easy to understand and transpile.

Listing 1 shows an example code using *Aixt* language and API, which makes blinking a LED for a specific microcontroller's pin. Likewise, the Listing 2 shows the *C* equivalent of the same *Aixt* source code.

Some of the basic features of *Aixt* language are listed as follows:

- the `:=` operator is used for declaring variables.
- Unlike *V*, variables are mutable by default in *Aixt*.
- `ix`, `ux` and `fx` variable types for regular integers, unsigned integers and floating point variables.
- `isize` and `usize` for integers with same size of the processor.
- `rune` type for character variables.
- Default type inference in declaring.
- Underscore character in literals for improving its readability.

- The `main` function is the first entry of a *V* program. In case of having only one source code, the main function definition can be ommited.
- All instructions end with a new line character, whit a semicolon or with a curly bracket close.
- The semicolon is optional. It has to be used when having two simple instructions in the same code line.
- All the code blocks are delimited by curly braces.
- All function declarations start with the reserved word `fn`.
- The names for all the identifiers (variables, constants, functions, etc.) prefer to use snake case as in *V*, for instance the function `pin_low()`. This feature is implemented in order to keep a standard format for all the *V* source code.
- There is only a loop instruction which is used for implementing all the supported loops, changing only its input parameters syntax.
- The reserved word `import` is used for including different complete modules or libraries.
- In order to reduce the *C* obtained code, it is possible to include individual components from a global module using the curly braces following the syntax: `import module { comp1, comp2, ...}`

Listing 1: Blinking LED example in *Aixt*

```
import machine { pin }
import time { sleep_ms }

pin(A6,OUT)
for {
    pin_high(A6)
    sleep_ms(500)
    pin_low(A6)
    sleep_ms(500)
}
```

Listing 2: Resultant *C* code for the Blinking LED example

```
#include "./settings.h"
#include "./machine/pin.h"
#include "./time/sleep_ms.h"

int main(void) {
    pin(A6,OUT);
    while(true) {
        pin_high(A6);
        sleep_ms(500);
        pin_low(A6);
        sleep_ms(500);
    }
    return 0;
}
```

B. Transpiler

A transpiler is a program that translates source code between programming languages with the same abstraction levels, by contrast a compiler translate source code generally to another low level language. The proposed programming framework does not compile directly the *Aixt* source code but transpile it to *C*. The Transpiler from *Aixt* language to *C* is implemented with *Python* and using the PLY module in order to implement the lexer analyzer and parser for the input source code. The complete working diagram of the implemented transpiler is shown in Fig. 2, where and input file with `.v` extension get in to the transpiler and it generates the output `.c` file. The transpiler implementation is based on part of the *V* language grammar, the Listing 3 shows and extract of that grammar in Backus-Naur form (BNF). This part of the grammar shows the definition of the four different ways to do loops in *Aixt* using the reserved word `for`, including infinite loops.

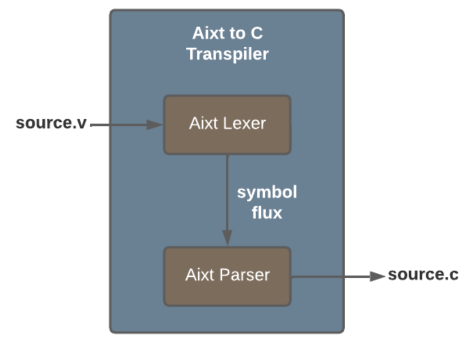


Fig. 2. Transpiler diagram

Listing 3: *Aixt* language BNF definition (extract)

```
...
forStmt ::= for block
          | for expr block
          | for forClause block
          | for inClause block

forClause ::= simpStmt ; expr ; simpStmt

inClause ::= exprList in IDENTIFIER
...
```

For the implementation of the lexer analyzer, all of the tokens of *V* language are supported, such as keywords, operands and other punctuation symbols, as shown in the code extract of Listing 4.

Listing 4: *Aixt* Lexer implementation (extract)

```
...
tokens = {
    I8, I16, I32, I64, ISIZE,
    ...
    F32, F64, BOOL, RUNE,
    IMPORT, IN, MAP, MATCH, RETURN,
```

```
    ...
}

BOOL        = r'bool'    # Types
RUNE        = r'rune'
...
IDENTIFIER  = r'[a-zA-Z_][a-zA-Z0-9_]*'
...
BINARY_LIT  = r'0b[01_]+'
...
literals = { '(', ')', '{', '}', '[',
             ']', ';', ',', '.',
           }
...
```

Once the Lexer analyzer reduced the character flux of the source code to an token flux, the parser analyzes the syntactic rules of language in order to find possible syntactic error and transpile it to *C*. The most of the syntactic rules of *V* are implemented in *Aixt* using the SLY module as shown in the extract source code of Listing 5, which matches with the BNF definition shown in Listing 6.

Listing 5: *Aixt* Parser implementation (extract)

```
...
@_( 'identList DECL_ASGN exprList',
    )
def varDecl(self, p):
...
    return ret_value

@_( 'IDENTIFIER',
    'identList "," IDENTIFIER'
    )
def identList(self, p):
...
    return p[0]
...
```

Listing 6: *Aixt* BNF rules (extract)

```
varDecl  ::= identList DECL_ASGN exprList

identList  := IDENTIFIER
            | identList "," IDENTIFIER
```

SLY library uses *Python's* function decorators to implement the syntactic rules of the language to be compiled or transpiled, applying them to each syntactic production, for example the production varDecl is the implementation of variable declarations in *Aixt* language.

As previously said, the transpiler reads the source code written in *Aixt*, which for compatibility with standard source code editors, the `.v` file extension.

C. Application Programming Interface

One of the main goals of the proposed framework is designing a cross-platform API, which includes the basic features and peripherals of most microcontrollers. In order to make the microcontroller's programming process easier, a general Application Programming Interface is implemented in both the *Aixt* programming language and *C* for the specific native compiler. This API includes the peripherals and features shown in Tables from I to IV.

TABLE I. General Purpose Input/Output

Description	Function name
pin type declaration	`pin()`
setting high and low	`pin_low()` `pin_high()`
setting specific binary value	`pin_value()`
reading an input value	`pin_value()`

TABLE II. Analog to Digital Converter (ADC)

Description	Function name
ADC setting up	`adc()`
ADC reading value	`adc_read()`

TABLE III. Universal Asynchronous Receiver Transmitter (UART)

Description	Function name
UART setting up	`uartx()`
single byte transmitting	`uartx_put()`
single byte receiving	`uartx_get()`

TABLE IV. Timing Features

Description	Function name
delays in microseconds	`sleep_us()`
delays in millisecond	`sleep_ms()`
delays in second	`sleep()`

Table I shows the pin and GPIO functions like setup, input capture and output set. Some devices even could support exchange state functions (`pin_toggle`). The rest of API functions follow the same rules:

- The setup function has the same name of the module.
- The rest of name functions of the same module follow the syntax: `module_function()`. For instance: `adc_read()` function of `machine { adc }` module (Table II).
- Devices with more than one peripheral of the same time follow this name function syntax: `modulex_function()` where the `x` refers to the number that identify each peripheral. For instance: `uart2_get()` as shown in Table III.
- Some API modules refers to a inner features of the device different to hardware peripherals, for instance software delays (Table IV).

The Fig. 3 shows the folder structure designed for the overall API, this structure has to be followed for each of the supported microcontrollers and boards to maintain the compatibility across all the hardware devices. Following strictly this folder structure allow the transpiler to correctly redirect the module including tasks when it is necessary to include to the project isolated components of a module.

As previously mentioned, the module including follows the next syntax in *Aixt*: `import module` for complete modules, which will be transpiled as `#include \./module.h"`. Likewise, the sub-modules or module components including follows the syntax: `import module { sub1, sub2, ...}`, which will be transpiled to `#include \./module/sub1.h"` etc. That is very important to optimize the resultant binary file. On the other hand, when a complete module is included, the `./module.h` header file has to include all of the `.h` files in the correspondent folder on the folder's API structure.

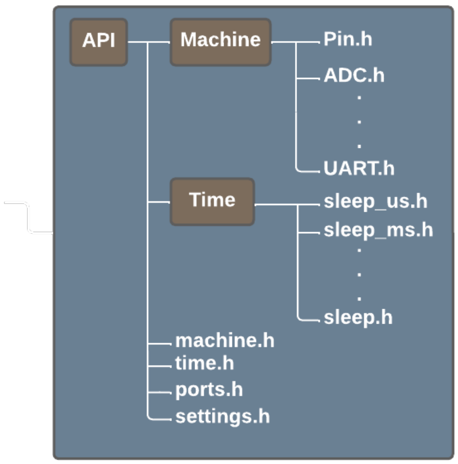


Fig. 3. General API structure

III. Results

The overall project including the *Aixt* language definition, the transpiler from *Aixt* to *C* and the API, is published by the authors as a free software project at the URL https://gitlab.com/fermarsan/aixt-project. The authors hope this project works as a starting point of a great free programming framework for microcontrollers or as seed for other similar projects.

The complete programming framework was successfully tested using some of the 8-bit and 16-bit PIC microcontroller families from Microchip ®. Those devices were selected due to their low amount of implemented data and program memories.

Several different working tests have been performed to check the correctness of most *Aixt* features. Listings 7 and 8 show a comparison of the variable declarations in *Aixt* and the corresponding transpiled *C* code for XC8 and XC16 compilers. In *Aixt* the variable declaration is always along with an assignment. The declaration and assignment process uses the operator `:=` to differentiate with only assignment =. At the same time it is necessary to use the conversion predefined functions such as `i8()`, `u32()` and `f64()` among others, in order to specify the number of bits and the type of integer and floating point variables. One of the benefits of using the conversion functions of *V* for the variable definitions is that each variable is bit-width explicit, independent of hardware device. Listing 7 shows too the use of the underscore symbol "_" for improving the large numbers readability. Also the special notations for hexadecimal, octal, and binary literals, are shown. The only difference with *C* is the octal literals beginning with the sequence "`0o`" (zero + o), instead of only `0` as in *C*.

Listing 7: *Aixt* variable declaring and assignment example.

```
var2 := i8(129)
var3 := i64(-6_835_292)
var4 := u8(0b0011_0101)
var5 := u16(0o073452)
var7 := u64(0xAAFF_7625)
var8 := f32(1_342.56)
var9 := f64(-34.035_440)
```

Listing 8: *C* resultant variable declaring and assignment example.

```
int8_t var2 = 129;
int64_t var3 = -6835292;
uint8_t var4 = 0b00110101;
uint16_t var5 = 0073452;
uint64_t var7 = 0xAAFF7625;
float var8 = 1342.56;
long double var9 = -34.035440;
```

Modern programming languages like *V* has some useful features such as the type inference, which simplifies programming in most cases. Type inference gives programmers peace of mind about variable types when they are not needed, thereby reducing development time. The implementation of this feature in *Aixt* is reached by using the standard types for integer and floating point variables. In the case of XC8 compiler the standard integer type is `int8_t` and for XC16 compiler `int16_t`. For the floating point variables the default type is `float`. Listings 9 and 10 show the transpiling result for some variable declarations by inference, including Boolean, character (named runes), integer and floating point literals, for the XC8 compiler.

Listing 9: *Aixt* variable declaring and assignment by inference example.

```
var0 := true
var1 := false
var2 := 1345
var3 := 71.4
var4 := -457
var5 := -10.445
var6 := `d`
```

Listing 10: *C* resultant variable declaring and assignment by inference example.

```
bool var0 = true;
bool var1 = false;
int8_t var2 = 1345;
float var3 = 71.4;
int8_t var4 = -457;
float var5 = -10.445;
char var6 = 'd';
```

On the other hand, *Aixt* language syntax provides support for some of *V's* looping statements, such as: condition for (`while` in *C*), bare for or infinite loop (`while(true)` in *C*), infinite loops, regular for loop and C-like for loop. Listing 10 shows an example of the loop statements currently supported by the *Aixt* syntax and Listing 11 shows the *C* equivalent of each one. The Aixt-like for loop includes and integer range notation with the syntax: `i..f`, where `i` is the initial value and `f` is the final value.

Listing 11: *Aixt* available loops.

```
// condition for
for a < 10 {
    a += 1
}
// bare for
for {
    a += 1
}
//range for
for i in 0..10 {
    arr[i] = 0
}
//c for
for i:=0; i<=10; i++ {
    arr[i] = 0
}
```

Listing 12: *C* equivalent loops.

```
while(a < 10){
    a += 1;
}
while(true){
    a += 1;
}
for(int i=0, i<10, i++){
    arr[i] = 0;
}
for(int i=0, i<=10, i++){
    arr[i] = 0;
}
```

A. Microcontrollers Setting Up

In order to setup a specific new microcontroller or board added to the *Aixt* programming framework, a configuration file has to be written. The chosen format for this configuration file is *YAML* which means Yet Another Markup Language, and is a very simple format to implement setup file for software projects. In that configuration file the designer can setup features such as: type equivalences between *Aixt* and the native *C* compiler, the microcontroller fuses or configuration bytes, the part or device number, the default header files among others. This configuration file is expected to be modified once by the designer and not to be modified by a regular user. The Listing 12 shows an extract of the configuration file for a PIC24FJ device.

Listing 13: *YAML* microcontroller or board configuration file (extract).

```
i8:       int8_t
...
u16:      uint16_t
...
default_int:    int16_t
...
device:   p24FJ128GA010
...
headers:
  - <xc.h>
  - <stdint.h>
...
configuration:
  - "POSCMOD = XT"
  - "OSCIOFNC = ON"
...
```

On the other hand, a batch file has to be included for each new device. This file works as a Makefile, following the steps and invoking the different component of the framework, in order to obtain the final binary file starting from the *Aixt* source code. The batch file has to be provided in `.ps1` (PowerShell) format for Windows and in `.sh` format for Linux.

IV. Conclusion

Using the proposed programming framework, the microcontrollers programmer can utilize a modern high level language programming environment, using a compiled language with its benefits and at the same time taking advantage of the modern features of the language. *Aixt* Language pretends to be a highly level programming language for microcontrollers with a short learning curve due to its simplicity compared with other modern languages. *Aixt* utilizes modern *V*-based features such as type inference but at the exact same time obtains binary files with similar optimization degrees of standard compiled languages such as *C* and similar execution times. *Aixt* Language and the proposed programming framework could enable programming microcontrollers with ease, as long as they have a native *C* compilers. At the same time *Aixt* does not need a fixed amount of memory to work, the finally binary file depends only the source code. So it has not the problem of the memory needed to run a program written with an interpreted language such as *MicroPython* or *Javascript*.

The transcompilation process between *Aixt* and *C* is successful because both languages are similar, mainly due to all variables in *Aixt* are mutable by default like in *C*, and some other similar features like curly braces and others. Transpile another language such *Python* to *C* for instance, would be a little bit difficult because the differences between both languages.

Aixt Language and programming framework could allow individuals with little electronics know how to program easily embedded systems, just like *Arduino*, *mbed*, and *MicroPython* among other frameworks. Likewise, *Aixt* could enable experienced embedded system programmers only learning one programming language and API, to program a wide variety of microcontrollers no matter their memory sizes.

All of the features in the proposed programming framework was completely tested, however not all the modern features of the *V* programming language were implemented. That means this project can highly improve implementing more features and adapting it to other microcontrollers and boards. It is perfectly able to use *Aixt* language and this framework in the classroom in Basic courses of microcontroller and embedded systems, due to currently this project is highly functional.

In spite of the short learning curve of *V* and therefore *Aixt* languages, it is possible to explore another simple languages to improve the proposed programming scheme, or even giving support to another main languages maintaining the same API. One of the candidates is the *Peregrine* Language who is based on the *Python* syntax.

As future work, the development of other useful features of *V* language are proposed. For example, the array definition, direct array indexing using the array for loop, array interpolation, matching statements among others. Likewise, it is important to keep giving support to other MCUs and board especially those with low program and data memories, which are the motivation for this project. For instance Atmel ® AT mega and AT tiny will be included to the project soon due to they use also the XC8 compiler. Finally, it is possible to combine PC graphical application developed in *V* with embedded application developed in *Aixt*, taking the advantage of learning only one programming basis to develop a complete embedded-based graphical application.

ACKNOWLEDGMENT

This work was supported by Universidad Distrital Francisco José de Caldas and Corporación Unificada Nacional de Educación Superior CUN. The views expressed in this document are not necessarily endorsed by Universidad Distrital or CUN. The authors thank the ARMOS and IDECUN research groups for the simulations and tests.

REFERENCES

[1] A. Radovici and I. Culic, *Embedded Systems Software Development*. Berkeley, CA: Apress, 2022, pp. 27–47.

[2] E. Kusmenko, B. Rumpe, S. Schneiders, and M. von Wenckstern, "Highly-optimizing and multi-target compiler for embedded system models: C++ compiler toolchain for the component and connector language embeddedmontiarc," in *Proceedings of the 21th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, ser. MODELS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 447–457. [Online]. Available: https://doi.org/10.1145/3239372.3239388

[3] A. K. Rachioti, D. E. Bolanakis, and E. Glavas, "Teaching strategies for the development of adaptable (compiler, vendor/processor independent) embedded c code," in *2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2016, pp. 1–7.

[4] K. Grunert, "Overview of javascript engines for resource-constrained microcontrollers," in *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2020, pp. 1–7.

[5] K. Zandberg and E. Baccelli, "Minimal virtual machines on iot microcontrollers: The case of berkeley packet filters with rbpf," in *2020 9th IFIP International Conference on Performance Evaluation and Modeling in Wireless Networks (PEMWN)*. IEEE, 2020, pp. 1–6.

[6] S. Varoumas, B. Pesin, B. Vaugon, and E. Chailloux, "Programming microcontrollers through high-level abstractions," in *Proceedings of the 12th ACM SIGPLAN International Workshop on Virtual Machines and Intermediate Languages*, 2020, pp. 5–14.

[7] R. Gurdeep Singh and C. Scholliers, "Warduino: a dynamic webassembly virtual machine for programming microcontrollers," in *Proceedings of the 16th ACM SIGPLAN International Conference on Managed Programming Languages and Runtimes*, 2019, pp. 27–36.

[8] D. E. Bolanakis, "A survey of research in microcontroller education," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 14, no. 2, pp. 50–57, 2019.

[9] S.-M. Kim, Y. Choi, and J. Suh, "Applications of the open-source hardware arduino platform in the mining industry: A review," *Applied Sciences*, vol. 10, no. 14, p. 5018, 2020.

[10] H. K. Kondaveeti, N. K. Kumaravelu, S. D. Vanambathina, S. E. Mathe, and S. Vappangi, "A systematic literature review on prototyping with arduino: Applications, challenges, advantages, and limitations," *Computer Science Review*, vol. 40, p. 100364, 2021.

[11] V. M. Ionescu and F. M. Enescu, "Investigating the performance of micropython and c on esp32 and stm32 microcontrollers," in *2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, 2020, pp. 234–237.

[12] A. Suarez Ruiz, "Diseño de hardware y firmware para un sistema inalámbrico de adquisición de datos daq de bajo costo," *Departamento de Ingeniería Eléctrica, Electrónica y Computación*, 2019.

[13] H. Wu, C. Chen, and K. Weng, "An energy-efficient strategy for microcontrollers," *Applied Sciences*, vol. 11, no. 6, p. 2581, 2021.

[14] M.-A. Lachaux, B. Roziere, L. Chanussot, and G. Lample, "Unsupervised translation of programming languages," *arXiv preprint arXiv:2006.03511*, 2020.

[15] A. M. Karpiński, "Automatic translation of programs source codes from python to c# programming language," Ph.D. dissertation, Zakład Sztucznej Inteligencji i Metod Obliczeniowych, 2022.

[16] M. Szafraniec, B. Roziere, H. Leather, F. Charton, P. Labatut, and G. Synnaeve, "Code translation with compiler representations," 2022. [Online]. Available: https://arxiv.org/abs/2207.03578

[17] F. A. Bastidas and M. Pérez, "A systematic review on transpiler usage for transaction-oriented applications," in *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, 2018, pp. 1–6.

[18] M. Ling, Y. Yu, H. Wu, Y. Wang, J. R. Cordy, and A. E. Hassan, "In rust we trust – a transpiler from unsafe c to safer rust," in *2022 IEEE/ACM 44th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, 2022, pp. 354–355.

[19] H. Lunnikivi, K. Jylkkä, and T. Hämäläinen, "Transpiling python to rust for optimized performance," in *Embedded Computer Systems: Architectures, Modeling, and Simulation*, A. Orailoglu, M. Jung, and M. Reichenbach, Eds. Cham: Springer International Publishing, 2020, pp. 127–138.

[20] M. Marcelino and A. M. Leitão, "Extending PyJL - Transpiling Python Libraries to Julia," in *11th Symposium on Languages, Applications and Technologies (SLATE 2022)*, ser. Open Access Series in Informatics (OASIcs), J. a. Cordeiro, M. J. a. Pereira, N. F. Rodrigues, and S. a. Pais, Eds., vol. 104. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, pp. 6:1–6:14. [Online]. Available: https://drops.dagstuhl.de/opus/volltexte/2022/16752

[21] B. F. Andrés and M. Pérez, "Transpiler-based architecture for multi-platform web applications," in *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, 2017, pp. 1–6.

[22] T. Würthinger, C. Wimmer, C. Humer, A. Wöß, L. Stadler, C. Seaton, G. Duboscq, D. Simon, and M. Grimmer, "Practical partial evaluation for high-performance dynamic language runtimes," in *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation*, ser. PLDI 2017. New York, NY, USA: Association for Computing Machinery, 2017, p. 662–676. [Online]. Available: https://doi.org/10.1145/3062341.3062381

[23] R. Pereira, M. Couto, F. Ribeiro, R. Rua, J. Cunha, J. P. Fernandes, and J. Saraiva, "Ranking programming languages by energy efficiency," *Science of Computer Programming*, vol. 205, p. 102609, 2021.

[24] K. Takano, T. Oda, and M. Kohata, "Approach of a coding conventions for warning and suggestion in transpiler for rust convert to rtl," in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 2020, pp. 789–790.

[25] W. Bugden and A. Alahmar, "Rust: The programming language for safety and performance," *arXiv preprint arXiv:2206.05503*, 2022.

[26] T. Uzlu and E. Şaykol, "On utilizing rust programming language for internet of things," in *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2017, pp. 93–96.

[27] K. I. Vishnunaryan and G. Banda, "Harsark_multi_rs: A hard real-time kernel for multi-core microcontrollers in rust language," in *Smart Intelligent Computing and Applications, Volume 2*, S. C. Satapathy, V. Bhateja, M. N. Favorskaya, and T. Adilakshmi, Eds. Singapore: Springer Nature Singapore, 2022, pp. 21–32.

[28] J. Aparicio Rivera, "Real time rust on multi-core microcontrollers," Master's thesis, Luleå University of Technology, Computer Science, 2020.

[29] N. P. Kumar Rao, *Getting Started with V Programming*. Packt Publishing, 2021. [Online]. Available: https://www.packtpub.com/product/getting-started-with-v-programming/9781839213434

[30] F. Martínez Santa, S. Orjuela Rivera, and F. H. Martínez Sarmiento, "Rust-like programming language for low-resource microcontrollers," *Advances in Dynamical Systems and Applications*, 2022.

Advantages of Digital Transformation Models and Frameworks for Business: A Systematic Literature Review

Seyedali Aghamiri[1], John Karima[2], Nadire Cavus[3]
Department of Computer Information Systems, Near East University, Mersin 10, TR-99138 Nicosia, Turkey[1, 2, 3]
Computer Information Systems Research and Technology Centre, Near East University, 99138 Nicosia, Cyprus[1, 3]

***Abstract*—Digital Transformation (DT) is a vital change in the way an organization utilizes processes, people, and technology to provide value to its ever changing customer expectations over products and services. Researchers developed models and frameworks to tackle concerns in this area, and existing literature improved our understanding of digital transformation. However, there are not enough comprehensive systematic literature reviews to picture a clear portrait of the advantages of related works and point out the major gaps for future studies. This study aims to evaluate how these models and frameworks affect business while highlighting their advantages and pointing out their gaps for future improvements and studies. A Systematic Literature Review (SLR) applied and collected and reviewed seven models and nine frameworks over five years between 2017 and 2021 from four databases of IEEE, Web of Science, Scopus, and Science Direct. These models and frameworks were reviewed and their advantages for researchers and practitioners were pointed out while picturing a clear vision of what is done in the Digital Transformation development of models and frameworks. The findings in this SLR indicated that the rising trend of DT studies has increased by 275% solely from 2020 to 2021 with 62% of those studies conducted in Europe.**

Keywords*—Digital transformation; digitalization; model; framework; business; SMEs

I. Introduction

Digital Transformation (DT) is a vital change in the way an organization utilizes processes, people, and technology to provide value to its ever changing customer expectations of products and services [1]. Corporations that effectively use DT appreciate enhanced yields of resources and therefore received higher profits [2]. The DT concept was announced in the year 2000 [3], as the DT can be characterized in a broad sense as the adoption or modification of business models or frameworks as an outcome of the fast quotient of technological progress and innovation, which triggers alterations in customer and social behavior. It is necessary to point out that a framework, also called a model in the literature, is a graphical depiction of an occurrence comprising its main factors, variables, and the interactions among them [4]. Today's DT trends are transforming the business landscape, over the past decade, researchers have focused more on designing DT models and frameworks and provided guidelines on how these models and frameworks can function in various businesses [4], mainly the previous studies in this area were to develop or modify DT modes and frameworks to address business needs while explaining the implementation process of DT as a whole or a part of the business model. This study was focused on a novel perspective and hoped to fill the gap between prior studies by gathering the latest DT models and frameworks and identifying their advantages for businesses by tending to the business needs while examining the offered value and applicability of DT models or frameworks in today's business ecosystem. The literature showed that there are numerous DT models and frameworks and it is important to choose the most suitable among all the proposed models, however, this SLR can be a guideline for the leaders and decision-makers in the organization who are responsible for approaching the DT to select the most suitable DT model or framework for their organization based on their DT strategy. As society is leaning toward digitizing countless aspects, businesses feel the rising significance of DT. Small Medium Enterprises (SMEs) may approach DT for numerous causes, DT has the power to enhance customer experience by bringing extra data-based insights for business decision-making. Another reason for taking on DT is the superior collaboration that it brings to the business by offering noble opportunities, agility, and alignment throughout the body of the firm, however, this study approached an unsolved gap in this area, and gathered the most recent DT models and frameworks developed by the researchers [4], [9-23] for businesses and analyzed them to bring their advantages to the surface, this approach helps the practitioners to address related concerns more swiftly, also it will provide the researchers with a more clear vision for their future studies in this area.

The purpose of this systematic literature review (SLR) is to collect and review the proposed models and frameworks for DT and encapsulate their approaches toward businesses, additionally offering a better understanding of the DT for future studies. The significance of improving the DT lies within the modifications of business models or frameworks for responding to technological, digital, and social changes [5], as it can be viewed as a revolution to a fresh organizational arrangement that matches better performance in the digital economy with the rapid growth of digital properties. On the other hand, the scope of this study is to find the answers to the following research questions.

RQ1: What is the distribution of publications over time?

RQ2: What is the geographical distribution of the studies?

RQ3: What are the latest developed models and frameworks for business?

RQ4: What advantages are embodied in each model and framework?

RQ5: How future studies can improve the DT models and frameworks?

II. Literature Review

A considerable amount of literature has been published on Digitalization in SMEs and Industry 4.0 (I 4.0) to determine the salient recommendations for their adoption [27], [30], They made use of the Technology Organization Environment (TOE) framework and determined that SMEs must place a high priority on integration IT systems along the entire value chain and human capital with a focus on proficiency in data analysis and knowledge exploration. A comparative literature review by [28] looked into the DT for transportation with regard to the product-service system lifecycle. Another study [29] only focused on the links between digitalization, company culture, and sustainability in SMEs, researchers were able to identify ten such interdependencies and develop a new analytical framework in the process. The researchers derived six key propositions to aid SMEs in their quest for sustainable digital development via corporate culture development, such as having managers raise awareness and positively shape employee perceptions of the prospect of DT, practicing leadership, culture, and digitalization for sustainability, and prioritizing the integration of DT into the SMEs culture by having the organization's mission and vision statements reflect DT commitments. Another existing study [30] explained the requirements necessary to implement DT in the Smart Manufacturing assessment models. The study [31] undertook a literature review of 204 articles to determine how Big Data and the Internet of Things affect businesses and how they carry out DT. They noted that the growing prevalence of Big Data and IoT increased the amount of disorganized knowledge available to businesses and companies seeking to implement them should learn how to make sense of this noise.

However, these researches do not take into account the advantages of the most recently developed and redefined DT models and frameworks for businesses, also they did not point out the presence of the numerous existing DT models and frameworks waiting to be tested and improved by the practitioners and the enthusiast researchers in this area, this systematic literature review aims to fill this gap.

III. Methodology

The research made use of a systematic literature review to find and choose relevant articles while simultaneously minimizing the chances of conducting the review process [6]. The study consisted of a systematic search undertaken in April 2022. Table I showed a keyword search was employed with the selected keywords being relevant to the subject matter of the study. The keywords used were "digital transformation" OR "digitalization" AND "framework" AND "business" OR "enterprise" in tandem with the synonyms listed in the following table. These keywords were then used to search through four high-quality databases, namely ScienceDirect, Web of Science, Scopus, and IEEE. These were selected as they adequately covered the fields of digitalization, management, and entrepreneurship.

TABLE I. Keywords Search Query

	Keyword	Synonyms
MAIN	"digital transformation" OR	"digitalization"
AND	"model" OR.	"framework"
AND	"business" OR	"enterprise"

Table II portrayed inclusion and exclusion criteria used as part of the selection process to identify articles that were relevant to the research's aims. The inclusion criteria stated that only documents published within the last five years were to be utilized in the study. As it was merely after 2014 that the DT expression was rapidly nurtured in recognition both by experts and researchers [7]. The business world concentration has reformed significantly for facing the businesses DT from 2017 to 2021 [4]. In order to narrow down the search results to the most recent and cutting-edge research in the field of DT, the inclusion criteria are set to be restricted to only journal articles that were published in English and open access type. Records that did not meet these criteria were excluded from the review.

TABLE II. Inclusion and Exclusion Criteria of the Study

Inclusion Criteria
Documents published within the last 5 years from 2017 to 2021. Journal articles published in the English language. Documents that are related to the "Business" and "Management" subject areas. Full-text studies that are accessible and downloadable (open access). Studies that present models or frameworks for digital transformation. Studies that include matching keywords with our research.
Exclusion criteria
Any records published before the year 2017 and after 2021. Only published journal articles and review articles. Journal articles published in any language other than the English language. Relevant studies in which the focus areas are business or management. "Digital transformation" OR "Digitalization" OR "Digitalisation" don't exist in the title, abstract, or keywords of the studies.

The PRISMA statement was used to screen records for the primary study selection process. The acronym PRISMA is an abbreviation of, the "Preferred Reporting Items for Systematic Review and Meta-Analyses" [6]. In [8]author stated that it is capable of exclusion of duplicated studies between databases. It also enabled the researcher to remove studies without clearly defined aims, studies that are not relevant to the research questions, and studies that only focused on one keyword as opposed to all of them, e.g. digitalization, model, and business. In this SLR the updated version of PRISMA, PRISMA 2020 is utilized.

A total of nine thousand, eight hundred and ninety-six (n=9896) articles were retrieved from a preliminary search of the four databases, with the individual databases contributing records as follows: Scopus (n=3921), IEEE (n=5888), ScienceDirect (n=71) and Web of Science (n=16). No duplicate records were found via automation tools, although two duplicate records (n=2) were removed manually by the researchers. Out of these results, records published before

2017 and after 2021 summed four thousand seven hundred and forty-seven (n=4747) were removed. Records other than journal articles were also removed, and these totaled three thousand four hundred and forty-two (n=3442). Records whose full text was not accessible on the Internet totaled eight hundred and forty-three (n=843) and these were excluded as well. Records were excluded as they did not have the main keywords in the abstract or title or keywords section of them were ninety in number (n=90). Records not in the English language totaled nine (n=9), and these were removed in the process. A combined master list of the remaining retrieved articles summed 238 which three (n=3) could not be retrieved over the internet, this left us with 235 studies in number that were consolidated on a Microsoft Excel report after the articles were selected beside the inclusion and exclusion criteria. At this stage, the researchers independently assessed the articles to make sure that they were relevant to the study topic, search terms, and research questions. In order to deal with disputes in the selection process, the researchers conducted video conference sessions for reviewing critical materials and also used a WhatsApp chat discussion to supplement this. As the final results of the screening and eligibility scan following records were excluded, Records that don't have main keywords in abstract/title/keywords (n = 47), Records that are not in the field of business and management (n = 58), Records that do not present a digitalization model or framework (n = 114), ultimately a total of sixteen (n=16) studies remained as principal studies for examination (Fig. 1).

Data extraction was carried out on the remaining articles that met the selection criteria. This was conducted in accordance with a number of parameters as shown in Table III. After the data was collected and extracted from the articles, it was synthesized through the construction of a summary table and subsequently evaluated to better perform a literature review.

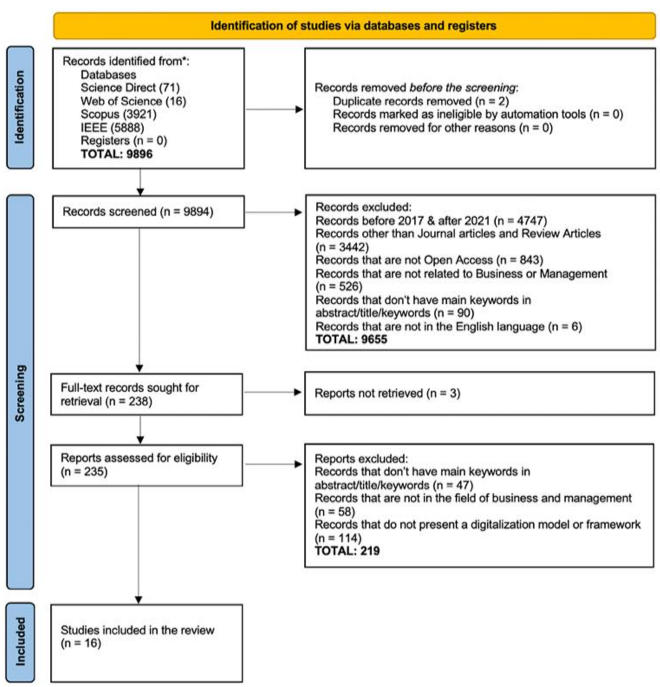


Fig. 1. PRISMA 2020 flow diagram of the study.

TABLE III. DATA EXTRACTION

Data Item	Description
Source	Name of the database the article is from
Author, Year	Give the author names, year of publication
Proposed Model OR Framework	What model or framework was proposed in the study
Aspects and Activities	What aspects and activities were featured in the proposed model or framework
Key findings	Show the main results from the articles
Advantages	Show the advantages of each study for business maturity
Recommendations	Show research gaps and areas of improvement

IV. RESULTS

A. *The Distribution of Publications over Time*

Fig. 2 showed the compiled set of selected studies that dealt directly with DT and proposed a framework or a model. However, in the screening and selection process this SLR didn't find any paper within the research criteria from "IEEE" and "Web of Science", and all the 16 records that were included in the study for further analysis were from "Scopus" and "Science Direct". In the past five years ranging from 2017 to 2021, the trend identified in Fig 2 depicted that within the criteria of this SLR no models or frameworks were proposed in 2017 and 2018. On the other hand, an increase in the number of publications over time from 2019 to 2021 is indicated, with the quantity peaking at 11 total publications in 2021 alone, this shows the focus of researchers on this topic amplified 275 % only over 1 year from 2020 to 2021. Also, displayed that DT and its relationship with businesses and SMEs was a particularly rich area of study for researchers, and deducing this to later years suggests continued interest in the topic. Indeed, [4] highlighted that DT has grown to the point of becoming critical to the survival of companies, with a survey conducted concluding that 84% of such organizations regard it as a necessity for their continued operations in the coming five-year period.

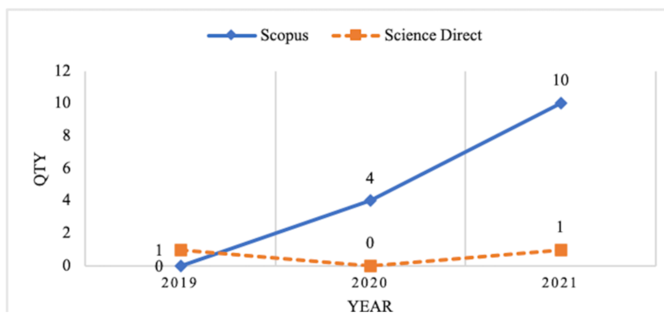


Fig. 2. Publication distribution over time.

B. *The Geographical Distribution of the Studies*

It is depicted in Fig. 3, the studies included in this SLR were a cosmopolitan blend, the 16 proposed frameworks and models presented in the findings were from 5 continents, with studies being authored in eleven (n=11) different countries. Germany was the country with the highest number of articles, with a total of three (n=3) published studies, followed by Italy, Austria, and Brazil each producing two (n=2) studies apiece, also, Finland, Poland, China, Spain, Vietnam, South Africa, and the USA each published one (n=1) study on the subject. This SLR noticed a trending focus on the subject more in European countries with 62% of them conducted in Europe.

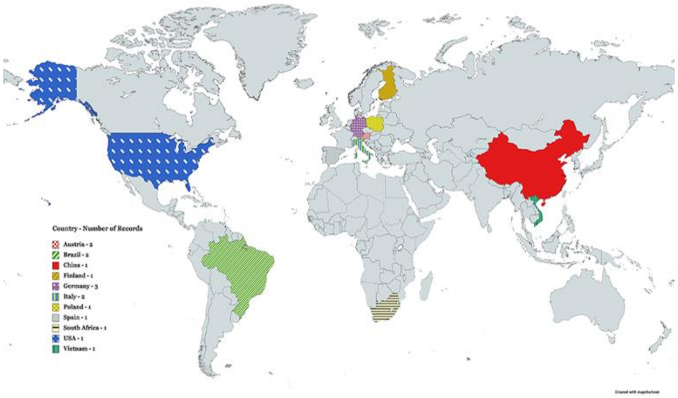


Fig. 3. Geographical distribution of the studies.

C. *The Latest Developed Models and Frameworks*

The results of this SLR showed that various DT models and frameworks were developed, and widely the focus was on tackling the related issues by proposing a novel model and framework, also revising the prior versions to address other concerns in this area. However, there were not enough comprehensive systematic literature reviews to point out their advantages for businesses or organizations or to picture a clear portrait that enables the comparison among the existing models and frameworks. Table 4 depicted the name of the proposed models, and showed the main aspects and activities that researchers presented in their studies on DT models [9–15] Moreover, Table V portrayed the name of the proposed frameworks, and displayed the main aspects and activities that researchers presented in their studies on DT frameworks [4], [16–23]. Aspects and activities in the mentioned tables are stages, or phases that the researchers suggested in their studies, these are the steps to be taken by the practitioners in order to implement the model or framework.

TABLE IV. DIFFERENT ASPECTS AND ACTIVITIES OF DT MODELS

Proposed Model	Aspects and Activities	Ref
A "Unified" Framework of Competing Forces From COVID-19	"Digitalization, Processes, Cost Reduction, Business Opportunities, Risk and Negative Consequences, Surveillance"	[9]
Econometric Model to Assess A Company's Preparedness Levels for Digital Transformation.	"Enterprise Management, Productivity Management, Digital Transformation Platform, Smart Manufacturing"	[14]
The Refined Model of Organizational Competence for Digital Transformation	"Vision and strategy. Digital leadership, Organizational learning process and organizational knowledge, Organizational alignment and organizational structure. Digital maturity, Personal competencies. Importance of personal characteristics, Technology selection, acceptance and use, Information culture. Organizational effectiveness"	[10]
A Digital Competence Maturity Model (DigiCom)	"Digital Content, Human-Machine. Human-Human, Personal"	[11]
Model of The Impact of DT on Project Management	"Disruptions: New technologies, Customer expectations, Competitive landscape, Data and analytics, Change process, Projectification of activities, Effects: process automation implementation of new technologies remote cooperation world sourcing new structures, Virtual project teams Online, constant communication Agile methods, Customer orientation and incremental product delivery, Project manager as a facilitator, Optimization of processes of project delivery, Constant access to data, IT tools"	[12]
Digital Transformation Model	"Positioning, Roadmap, Implementation, Current State Review"	[13]
Development Approach Industry 4.0 Maturity and Realization Model	"Definition, Maturity model's scope, structure and design, Research and population with maturity items, Development assessment tool and structure of maturity report, Transformation maturity items into realization paths, Definition procedure and rules for deriving action-field, Testing and creation of final step-by-step realization model"	[15]

TABLE V. DIFFERENT ASPECTS AND ACTIVITIES OF DT FRAMEWORKS

Proposed Model	Aspects and Activities	Ref
I 4.0 Emerging countries' DT strategy framework (ECDTS)	"Digital platforms, Ecosystems"	[16]
Interpretative Framework	"Business driver, Supply view, Entrepreneurial model"	[17]
Digitainability Conceptual Framework	"Digitalization, Sustainability"	[18]
Multi-Dimensional Framework	"Contextual Conditions, Mechanisms, Outcomes"	[19]
Digital transformation patterns	"Lean Production, Industry 4.0"	[20]
Conceptual Framework of open innovation to support digital transformation.	"Managerial Strategic and Social Drivers, Digital Technology Enablers, Business Partners, Scientific Partners, Facilitating Factors, Limiting or Challenging Factors, Contribution of collaboration for Digital Innovation"	[21]
The interaction-based digital transformation framework	"Digital Technologies, Business, Society"	[4]
Digitalization and public crisis responses theoretical framework	"Digitalization, Dynamic Capabilities, Response Strategies, Response Performance"	[22]
Construct Map Conceptual Framework	"Digital Transformation, Digitalization, Business Model Innovation"	[23]

D. Advantages that Embodied in each Model and Framework

One of the main goals of this SLR was to find the advantage of the proposed models and frameworks, in order to do that this SLR reviewed and summarized the purpose of these studies. Since the Covid-19 pandemic began it affected all aspects of our environment, and DT is one of these areas. One study recognized the disruptive nature of the pandemic and seeks to utilize its pressure in a way that benefits businesses, i.e. by using it as a catalyst for digital transformation, digitalization drivers and barriers are identified and the study gives organizations a clear picture of the benefits and drawbacks of such a shift, the Conceptual Model proposed by this study offers a roadmap that tackles digitalization in the field of uncertainty i.e. the state of the economic crisis during the Covid-19 pandemic [9]. The research [22] examines the correlations between DT, disaster management strategies during the Covid-19 pandemic, and SMEs' disaster response ability. Using online survey data of 518 Chinese SMEs, the findings in the research perfectly illustrate that, in the long term, digitization may assist SMEs in deploying disaster response as well as responding tactically to public crises, contributing to an increase in SME's capability, an advantage the enormous data set of 518 SMEs in China was used for structuring the presented theoretical framework which is considerable and valuable, especially in times of crisis.

Besides the model and framework development, tool development was also an area that received attention from the researchers. One study developed a useful tool for gauging the preparation of an enterprise for Digital Transformation, classifying them as either Newcomer, Learners, or Leaders. This model was used in conjunction with operational efficiency figures and data on the movement of personnel to determine the impact of DT preparation on the sustainability of companies. The analysis showed different results across the three classifications of companies. overall, businesses benefit from tight integration of Digital transformation in the long run as this yields the greatest operational efficiencies and personnel utilization, the benefit is that this model allows for the assessment of a business's preparation and readiness for Digital Transformation on the sustainability of a company [14]. The [10] research led to the creation of an OCDT model that may assist SMEs in identifying and developing DT strengths in order to progress their enterprise model's DT, moreover, the model is more appropriate to the needs of SMEs than other models that enable digital readiness advancement but were intended for large firms, the advantage of this model will enable SMEs to assess and create digital skills that are now lacking, moreover, it will assist in the digital transformation of business models in order to develop competencies that will enable SMEs to effectively adjust to the changing competitive atmosphere created by digital technologies and marked by innovation and rapid change.

This SLR was able to find solely one research that focused on the impact of DT on project management. The author [12] developed a model that is known as the "Model of the Impact of Digital Transformation on project management" and analyzed thoroughly the manifestation of change that was formed into a model explaining the model, things like the "Change Process" which includes objectification of activities, "Disruptions" which includes the following, "Novel technologies", "Competitive landscape", "Customer expectancies", "Data and analytics", and "Outcomes" which includes procedure automation new technology adoption new architecture in the field of distant collaboration, as well as the most essential pro and con characteristics of the changes and identified the influence of Digital Transformation on Project. Moreover, the researchers found that Digital Transformation acts as a significant aspect that is an important element determining contemporary project management, the model that was used in this article has great significance in digital transformation in a general social context, there are several advantages in this study: a. Project effectiveness, which includes cost savings, time reduction on activities that are not useful, proper management and allocation of resources. b. It makes it possible to be able to use the time already saved as a result of the elimination process that took place on repetitive activities and gives more room for more creativity and adding value activities. c. Decrease the delivery time of the project. d. Automation has enabled the duration of delivery projects to be reduced to a minimum through the usage of new automation tools. e. Capacity to depletion of a huge number of existing data, real-time access, and quick and easy data handling. f. Possibility of work and results to remain flexible.

When it comes to DT a significant aspect to focus on is employees, in this regard [23] study has proven that human resources as magnitudes of National Intellectual Capital are cogent features of Digital Transformation and shall be reinforced to enhance Europe's capability to survive the digital world, also it constructed a map towards more robust Business Model Innovation. On the other hand, [11] Construct Map Conceptual Framework provides a historical outlook on European Union Countries in totality and focuses more on working talents as modules of human resource, with more focus on employee training and qualitative vocational training. Also, in another paper researchers proposed the Digital Competency Maturity Model (DigiCom) which consists of 49 attributes in four dimensions which are Digital Content, Human – Machine, Human - Human, and Personal to measure the maturity of individual employee levels in industrial enterprises with five steps developmental methodology in use. The proposed model approach used, makes digital competencies more visible, realistic and measurable and modify the basis for competence development in employees that will enhance the performance of their enterprise. The DT DigiCom model focuses deeply on the digital competency of employees in industrial enterprises because it's essential in the manufacturing sector to have digitally equipped clients with the knowledge of managing machines.

Some researchers worked their idea around Industry 4.0, for instance, Digital Transformation Patterns Framework, pointing out that lean production is a concept that is not considered as often as it should be when it comes to digital transformation, this study is an important step in this new area and shows that I4.0 and LP are related when it comes to transformation and that firms committed to lean achieve better digitalization outcomes than those that are not, and the framework takes into account Lean Production implementation level as a dimension, a novel metric to inform companies on digital transformation patterns with regards to Lean Production [20]. On the other hand, one study proposed an innovative industry 4.0 realization model i.e. industry digital transformation model which proposes a ten steps method that leads the company in a systemic way from their initial interaction by industry 4.0 (transformation model) till industry 4.0 (transformation model) the planning is well defined by roadmaps, action-field, and realization projects, which benefits to the industrial digitalization through its model by creating well-defined structure and enabled the collection of valuation data from manufacturing businesses to amplified precision when benchmarking the company's maturity among others [15]. The third group of researchers focused on digital transformation in emerging countries of Brazil, Russia, India, China, and South Africa, laying out a novel framework for firms to follow that yielded strategic capabilities to the firm, the I4.0 ECDTS framework was backed by a robust analysis of prior research, though its newness could benefit from the application of the proposed framework in the field, the study also reduces the research legwork with regard to policy features in use for strategic pivots. The I 4.0 ECDTS framework enables firms to evaluate the steps taken toward digital transformation, and the framework increased competitive advantage due to flexibility brought on by digitalization and enables strategic capabilities within the firm [16]. However, [21] examined cooperation between scientists and business partners as a key contributor to fostering Industry 4.0 and enhancing digital transformation, plus drawing attention to the fact that technology collaboration drag more benefit on the business success front than technologies developed and adopted solely internally by practitioners, as a benefit this framework analysis how innovation collaborations with different business and scientific partners integrate with digital transformation and Industry 4.0.

In contrast, some researchers followed by another study that they did in 2017 and proposed the model for tackling the DT and in this study, they focused on the step-by-step implementation of the DT model by providing clear explanations and tools and procedure for this purpose, the DT model is an iterative model and needs to be broken down into small pieces of plan and action to be implemented and is not expected to be a model for tackling all the DT difficulties all at once as they have piloted the DT model in 19 SME's and got brilliant results, the presented DT model is a 4 step which is iterative and can be divided into small steps to be taken by the company, also three tools are provided for better guidance and implementation of the DT model, these tools are The DigiMaturity tool, which answer the question, what is the maturity level of the business, the DigiSWOT tool that provides the questions for analyzing the digitalization strengths, weaknesses, opportunities and threats, and The DigiTriangle tool to let the company categorize their vision priorities [13]. Another study by [17] proposed Interpretative Framework investigates the primary legacy from prior industrial revolutions with a particular focus on business drivers, supply view, and entrepreneurial model. The study hopes to act as a guide for international manufacturing SMEs as they tackle digital transformation by providing a comprehensive framework of prior industrial revolutions from them to draw lessons, the features considered crucial were those of business drivers, supply view, and entrepreneurial model, the Interpretative Framework provides a historical outlook on the process of past industrial revolutions and may be used to extrapolate lessons for carrying out digital transformation in the present.

Nevertheless, [19] formed a Multi-Dimensional Framework that seeks to bring together all the aspects of Digital Transformation ranging from contextual conditions, mechanisms, and outcomes of the process, the study proposed a far-reaching conceptual framework that builds on past knowledge of the organizational change, as a result, it identified trends in digital transformation and came up with perspectives on digital transformation perspectives namely holistic co-evolution, systemic shift, technology impact, and compartmentalized adaptation, the Multi-Dimensional Framework identified trends that allow for greater understanding of Digital Transformation, namely, the shift towards malleable organizational designs and their integration into digital business ecosystems. This SLR found another research [4] that proposed a novel framework that works on the understated areas by combining the numerous models discovered in the literature, plainly incorporating the role of society, emphasizing the evolution over time, and incorporating the drivers of digital transformation classified in 23 digital transformation interfaces throughout 6 groups, eventually, this study can be the initial stage of a unified concept of digital transformation, because the researchers looked at the evolution of 41 DT models and found four different research possibilities, they pay homage to the DT research by organizing their findings into a fresh, all-encompassing framework that integrates important shifts in business, society, and technology, also, they further emphasize the processes' diversity by categorizing the drivers of DT into 23 DTI and 6 categories, this research brings a clear perspective on DT research, this could aid in a deeper understanding of how change happens and where certain changes interact with each other, by bridging the gap between business and society.

From a different perspective, [18] acknowledged the dark side of digitalization such as increased carbon footprints and energy consumption by firms, and seeks to mitigate these negative effects through its conceptual framework that lends greater weight to sustainability, as an advantage the Conceptual Framework gives greater focus to sustainability in digitalization, provides new findings for sustainability research and regulations and overall offers companies a greater chance at achieving sustainable digital transformation.

E. How Future Studies can improve the DT Models and Frameworks

To answer the last proposed research question, this SLR took a deep dive into the literature and reviewed the papers to find the most promising areas for future studies. Expansion of the studies to other countries and developments for a more international organizational environment [14], [16-17]. Monitoring and assessment of the applicability of the model/framework is a crucial area to look into, due to the novel nature of these models/frameworks [9-10]. More corporate studies including extending the research to different phases of DT models/frameworks as well as establishing tools, procedures, and methodologies, also trying for linking different understandings of different perspectives of the DT [13], [20]. [10], [12], [21] found another area to conduct research for future studies is an expansion of scope through more interviews and expanding the research through quantitative surveys or consolidation of prior quantitative studies, and trying to catch up with technology as AI is increasingly advancing.

V. Discussion

The results of the SLR found that there are some challenges that affect the success of digital transformation success in businesses. For example, digital technologies [10], the digital skills of employees, the digital transformation strategy [32], and leadership. The business and its employees should have the necessary innovative devices, and digital technologies for the digital transformation process, and of course, the digital skills of the employees should be sufficient to use new technologies effectively. In addition, the business must prepare an effective digital transformation strategy and leadership for digital transformation so that digital transformation processes can be carried out successfully.

The digital world globally needs to achieve smart sustainable development while creating value and wealth for society [24]. Therefore, this SLR took a deep dive into the literature and presented a clear picture of what has been done in terms of DT model/framework development searched for by the researchers. This shortage was one of the reasons why this SLR was carried out so as to fill this gap in the literature, it helps the researchers understand the status quo of DT and enabled them to visualize a clear vision of previous studies in this area. This study was able to identify an uptick in the trend of research papers published in the field of Digital Transformation over the past five years (from 2017 to 2021). The papers steadily increase in number, with the total number peaking at 11 papers published in 2021, which is more than the prior four years combined, which adds up to 5 papers. This growth may be attributed to an increased awareness of the extent to which DT impacts the organization [4]. Put forward the argument that DT is not just a passing trend, but may prove to be key to the continued operation of companies, with several companies acknowledging the new DT status quo, it stands to reason those organizations and researchers alike would take an increased interest in the area, resulting in a rise in the number of published articles over time.

Papers originated from a spread of 11 countries. As 62% (n=11) of studies came from Europe, which is remarkable, as the term Digital Transformation itself was coined in North America, one would deduce that the bulk of research would stem from their [3]. However, this peculiarity is better framed when looked at through an economic lens. Germany alone counts for 25% of Europe's Gross Domestic Product and constitutes the largest market in Europe [25], also, it is the third-largest exporter globally, after the United States and China. These factors may contribute to its organizations swiftly taking up DT so as to survive and thrive in the global marketplace. Researchers may also have been drawn to publish from the country as it is a hub of small and medium-sized enterprises, forming an ideal study and testing ground for proposed DT frameworks [25-26]. It is worth mentioning that the two terms may be used interchangeably when it comes to Digital Transformation [4]. This interchangeability of these terms caused this SLR target search broader than expected which made it harder to find needed results in the literature, but this SLR put both terms in the research agenda to maximize the efficiency of the study and make the road smoother for further studies around the DT area.

Therefore, one limitation, the researchers faced was to conduct this SLR from 4 databases (Scopus, IEEE, Science Direct, Web of Science) meaning that there are other databases that are not covered yet. Another limitation in this SLR was the time boundary, despite the concentration of the studies in recent years we suggest further research on a wider time boundary to collect more relevant studies from before the year 2017. This study focuses on finding the advantages of DT models and frameworks to businesses while gathering proposed models and frameworks that are developed for DT, this can be beneficial to the businesses that are not sure which DT strategy is more suitable to take and also the researchers will have a more clear vision of the road ahead to conduct future studies about DT. We suggest to other researchers who will study this topic, to focus on the negative effects of DT models /frameworks on SMEs and businesses and to utilize this SLR and take additional actions for the development of the DT models and frameworks.

VI. Conclusion

Previous studies in this area mainly focused on the development of DT models and provided guidelines and procedures for the adoption of their presented model. The aim of this SLR was to review and analyze existing DT models and frameworks to point out their advantages for businesses to help them to approach the most suitable DT model or framework for their organization based on their strategy. Moreover, the results of this SLR provide a clearer vision for future studies in this area. This SLR found 16 studies in a period of five years from 2017 to 2021 in four databases of IEEE, Web of Science, Scopus, and Science Direct. The results showed an incremental trend in the time distribution of discovered studies, interestingly only from 2020 to 2021 number of related studies raised by 275% percent. Also this study indicated that 10 out of 16 discovered studies were conducted in Europe. However, the DT area is still seeing novel models and frameworks that are not deeply tested and refined. We hope that the results of this SLR will help researchers and practitioners in this field in their future work and studies.

References

[1] J. Kääriäinen, O. Kuusisto, P. Pussinen, M. Saarela, L. Saari, and K. Hänninen, "Applying the positioning phase of the digital transformation model in practice for smes: Toward systematic development of digitalization," Int. J. Inf. Syst. Proj. Manag., 2020, doi: 10.12821/ijispm080402.

[2] G. Westerman, M. Tannou, D. Bonnet, P. Ferraris, and A. McAfee, "The Digital Advantage: How Digital Leaders Outperform their Peers in Every Industry," MIT Sloan Manag. Rev., 2012.

[3] K. Patel and M. P. McCarthy, "Digital transformation : the essentials of e-Business leadership," p. 134, 2000.

[4] Z. Van Veldhoven and J. Vanthienen, "Digital transformation as an interaction-driven perspective between business, society, and technology," Electron. Mark., vol. 32, no. 2, pp. 629–644, 2022, doi: 10.1007/s12525-021-00464-5.

[5] M. Kotarba, "Digital transformation of business models," Found. Manag., 2018, doi: 10.2478/fman-2018-0011.

[6] B. Kitchenham, "Procedures for Performing Systematic Reviews, Version 1.0," Empir. Softw. Eng., vol. 33, no. 2004, pp. 1–26, 2004, [Online]. Available: https://www.researchgate.net/profile/Barbara-Kitchenham/publication/228756057_Procedures_for_Performing_Systematic_Reviews/links/618cfae961f09877207f8471/Procedures-for-Performing-Systematic-Reviews.pdf.

[7] J. Reis, M. Amorim, N. Melão, and P. Matos, "Digital transformation: A literature review and guidelines for future research," Adv. Intell. Syst. Comput., vol. 745, pp. 411–421, 2018, doi: 10.1007/978-3-319-77703-0_41/COVER.

[8] M. Page et al., "PRISMA 2020 Checklist Section and Topic Item # Checklist item Location where item is reported TITLE Title 1 Identify the report as a systematic review," BMJ, 2021.

[9] J. Amankwah-Amoah, Z. Khan, G. Wood, and G. Knight, "COVID-19 and digitalization: The great acceleration," J. Bus. Res., vol. 136, no. July, pp. 602–611, 2021, doi: 10.1016/j.jbusres.2021.08.011.

[10] N. T. X. Hoa and N. T. Tuyen, "A model for assessing the digital transformation readiness for vietnamese smes," J. East. Eur. Cent. Asian Res., vol. 8, no. 4, pp. 541–555, 2021, doi: 10.15549/jeecar.v8i4.848.

[11] J. M. González-Varona, A. López-Paredes, D. Poza, and F. Acebes, "Building and development of an organizational competence for digital transformation in SMEs," J. Ind. Eng. Manag., vol. 14, no. 1, pp. 15–24, 2021, doi: 10.3926/jiem.3279.

[12] M. Steinlechner, A. Schumacher, B. Fuchs, L. Reichsthaler, and S. Schlund, "A maturity model to assess digital employee competencies in industrial enterprises," Procedia CIRP, vol. 104, pp. 1185–1190, 2021, doi: 10.1016/j.procir.2021.11.199.

[13] A. Kozarkiewicz, "General and Specific: The Impact of Digital Transformation on Project Processes and Management Methods," Found. Manag., vol. 12, no. 1, pp. 237–248, 2020, doi: 10.2478/fman-2020-0018.

[14] J. Kääriäinen, O. Kuusisto, P. Pussinen, M. Saarela, L. Saari, and K. Hänninen, "Applying the positioning phase of the digital transformation model in practice for smes: Toward systematic development of digitalization," Int. J. Inf. Syst. Proj. Manag., vol. 8, no. 4, pp. 24–43, 2020, doi: 10.12821/ijispm080402.

[15] A. Schumacher, T. Nemeth, and W. Sihn, "Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises," Procedia CIRP, vol. 79, pp. 409–414, 2019, doi: 10.1016/j.procir.2019.02.110.

[16] C. R. D. Andrade and C. R. Gonçalo, "Digital transformation by enabling strategic capabilities in the context of 'BRICS,'" Rev. Gest., vol. 28, no. 4, pp. 297–315, 2021, doi: 10.1108/REGE-12-2020-0154.

[17] E. M. Martinelli, M. C. Farioli, and A. Tunisini, "New companies' DNA: the heritage of the past industrial revolutions in digital transformation," J. Manag. Gov., vol. 25, no. 4, pp. 1079–1106, 2021, doi: 10.1007/s10997-020-09539-5.

[18] U. Lichtenthaler, "Digitainability: The combined effects of the megatrends digitalization and sustainability," J. Innov. Manag., vol. 9, no. 2, pp. 64–80, 2021, doi: 10.24840/2183-0606_009.002_0006.

[19] A. Hanelt, R. Bohnsack, D. Marz, and C. Antunes Marante, "A Systematic Review of the Literature on Digital Transformation: Insights and Implications for Strategy and Organizational Change," J. Manag. Stud., vol. 58, no. 5, pp. 1159–1197, 2021, doi: 10.1111/joms.12639.

[20] M. Rossini, F. D. Cifone, B. Kassem, F. Costa, and A. Portioli-Staudacher, "Being lean: how to shape digital transformation in the manufacturing sector," J. Manuf. Technol. Manag., vol. 32, no. 9, pp. 239–259, 2021, doi: 10.1108/JMTM-12-2020-0467.

[21] C. Rocha, C. Quandt, F. Deschamps, S. Philbin, and G. Cruzara, "Collaborations for Digital Transformation: Case Studies of Industry 4.0 in Brazil," IEEE Trans. Eng. Manag., pp. 1–15, 2021, doi: 10.1109/TEM.2021.3061396.

[22] H. Guo, Z. Yang, R. Huang, and A. Guo, "The digitalization and public crisis responses of small and medium enterprises: Implications from a COVID-19 survey," Front. Bus. Res. China, vol. 14, no. 1, pp. 1–25, 2020, doi: 10.1186/s11782-020-00087-1.

[23] C. van Tonder, C. Schachtebeck, C. Nieuwenhuizen, and B. Bossink, "A framework for digital transformation and business model innovation," Manag., vol. 25, no. 2, pp. 111–132, 2020, doi: 10.30924/mjcmi.25.2.6.

[24] N. Siniak and D. K. Lozanoska, "A Review of the Application of the Concept of Economic and Smart Sustainable Value Added (SSVA) in Industries Performance Evaluations," 2019.

[25] Eurostat, "Key figures on Europe 2021 edition," Flagsh. Publ., 2021.

[26] P. Kilimis, W. Zou, M. Lehmann, and U. Berger, "A survey on digitalization for SMEs in Brandenburg, Germany," 2019. doi: 10.1016/j.ifacol.2019.11.522.

[27] Grooss, O. F., Presser, M., & Tambo, T. (2022). Surround yourself with your betters: Recommendations for adopting Industry 4.0 technologies in SMEs. Digital Business, 2(2), 100046.

[28] Isensee, C., Teuteberg, F., Griese, K., & Topi, C. (2020). The relationship between organizational culture, sustainability, and digitalization in SMEs: A systematic review. Journal of Cleaner Production, 275, 122944.

[29] Li, F., Trappey, A. J., Lee, C., & Li, L. (2022). Immersive technology-enabled digital transformation in transportation fields: A literature overview. Expert Systems with Applications, 202, 117459.

[30] Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2018). A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). Journal of Manufacturing Systems, 49, 194-214.

[31] Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. Technovation, 98, 102173.

[32] Teng, X.; Wu, Z.; Yang, F. Research on the Relationship between Digital Transformation and Performance of SMEs. Sustainability 2022, 14, 6012. https://doi.org/10.3390/ su14106012.

Dynamic Time Warping Features Extraction Design for Quranic Syllable-based Harakaat Assessment

Noraimi Shafie, Azizul Azizan, Mohamad Zulkefli Adam, Hafiza Abas, Yusnaidi Md Yusof, Nor Azurati Ahmad
Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

***Abstract*—The use of technological speech recognition systems with a variety of approaches and techniques has grown exponentially in varieties of human-machine interaction applications. The assessment for Qur'anic recitation errors based on syllables utterance is used to meet the Tajweed rules which generally consist of *Harakaat* (prolonging). The digital transformation of Quranic voice signals with identification of Tajweed-based recitation errors of *Harakaat* is the main research work in this paper. The study focused on speech processing implemented using the representation of Quranic Recitation Speech Signals (QRSS) in the best digital format based on Al-Quran syllables and feature extraction design to reveal similarities or differences in recitation (based on Al-Quran syllables) between experts and student. The method of Dynamic Time Warping (DTW) is used as Short Time Frequency Transform (STFT) of QRSS syllable feature for *Harakaat* measurement. Findings from this paper include an approach based on human-guidance threshold classification that is used specifically to evaluate *Harakaat* based on the syllables of the Qur'an. The threshold classification performance obtained for *Harakaat* is above 80% in the training and testing stages. The results of the analysis at the end of the experiment have concluded that the threshold classification method for Minimum Path Cost (MPC) feature parameters can be used as an important feature to evaluate the rules of Tajwid Harakaat embedded in syllables.**

***Keywords*—Speech processing; short time frequency transform; dynamic time warping; human-guided threshold classification**

I. Introduction

The Tajweed Al-Quran guidelines [1], which outline the laws of Makhraj, Sifaat and Harakaat (MSH) scientifically have shown the comprehensiveness of rules and requirements for the proper pronunciation and articulation [2] of each syllable, forming precise recitation for each verse for Al-Quran recitation. This unique information is often embedded inconsistently in the Quranic recitation digital signal, due to different signal representation circumstances and recitation of the Quran by different readers [3]. The preparation of this recitation signal for further analysis, involving Digital Signal (Speech) Processing (DSP) up to the classification of Tajweed rules digitally, is not an easy task as it involves technical solutions, moreover religious opinions and sensitivity need to be factored in.

Difficulties in managing the complexity of recitation signals[3], especially in selecting a reasonable DSP method for the smooth implementation of every stage involved, are among the aspects of the study emphasized in this paper. The discovery of the error of Tajweed-based recitation [4] was completed by contrasting the comparability properties of the discourse signal between the learner and the expert recitation signal. Stochastic reading discourse is a very interesting issue for a person to show or evaluate reading errors based on technology. The specific model should be able to deal with even a relatively miniature form of differences.

The ability of DSP to reveal important features and classification approaches is very promising and becomes an important pillar in facing and solving the complexity of speech or voice signal analysis. Phoneme duration features and speech models are used to recognize some rules of reciting the Quran [5]. Therefore, an approach to reveal the characteristics and classification of time series to represent speech or voice using DSP techniques will be highlighted in this paper.

Harakaat, as one of the Tajweed rules in the recitation of the Qur'an [6], is the process of prolonging the pronunciation of syllables. In general, the law of Harakat Tajweed requires the prolonging to be performed on the particular syllable pronunciation guided by the character of ي (Ya), ا (Alif) and (وْ) (Wau). In addition, the prolongation of the syllable pronunciation [7] at the end of a sentence also involves 2, 4 or 6 Harakaat. The Harakaat rules, known as *Madd* [8] are generally divided into 12 types.

The structure of this paper begins with Section I which explains about the introduction to the complexity in speech or the nature of the voice in reciting the Qur'an. Section II explains about the type of reading rate of the Al-Quran that specifically involves the law of Tajwid Harakaat. Next, the work related to Harakaat analysis is explained in Part III. This is followed by a discussion about the Dynamic Time Warping (DTW) method for feature vectors in Part IV. Descriptions of the extraction design work and a description of the relevant experiments that have been carried out are included in Section V and VI respectively. Part VII of this paper explains the experimental results and the discussion of issues related to the results of the analysis. Finally, the conclusions from the research results in this paper are summarized in Section VIII.

II. Al-Quran Recitation Speed Rate

Islam requires its followers to read the Qur'an in an orderly and prudent manner. Furthermore, reading it with melodious and serious appreciation while following Tajweed rules [5] is very much demanded. There are four types of recitation speed rate which relate to Harakaat Tajweed [9].

A. *At-Tahqiq (التحقيق)*

This is a common reading type for beginners who are just learning Tajweed. The way it is read is similar to at-Tartil, but at-Tahqiq is identified as slower and calmer. The recitation of at-Tahqiq is also as practiced and considered as an initial stage for a novice reciter before the recitation improves to the next level.

B. *At-Tartil (الترتيل)*

In terms of terminology, at-Tartil carries the meaning of slow reading. This is in line with the tafsir stated in Tafsir Ibn Katsir, where tartil means reading the Qur'an in accordance with the law of Tajweed.

This means each of these readings is given the rights and properties of the letters that should be. Basically, reading with the al-Tartil approach allows the reader to understand and reflect on the verses of the Qur'an.

C. *At-Tadwir (التدوير)*

At-Tadwir is a type of reading in which the reading rate is intermediate between slow reading and fast reading. Usually, at-Tadwir recitation is practiced in congregational (Jama'ah) prayers, where the reference can be observed easily based on the recitation of Harakaat, known as Mad Munfasil which is recited not more than 6 Harakat.

D. *Al-Hadar (الحدر)*

Al-Hadar is defined as a way of reading at a relatively fast pace within the laws of Tajweed. In fact, Al-Hadar is the fastest level of reading and usually practiced by memorizers of the Qur'an who requires repetition in the process of memorization. A simple reference can be observed when the reading involved in the law of Tajweed namely Mad Munfasil, the reading is read with only two Harakat only. However, even though the whole recitation is read quickly, the reader adheres to the relevant Tajweed laws in every verse read.

In this paper, the collection of data is implemented for the at-Tadweer category. This important Harakaat attributes are evaluated in the experimental works taking into account the differences in recitation speed rate in every category, producing different Harakaat duration requirements.

III. Related Works

One of Tajweed rules is the prolongation, which is related to rhythm of recitation. There are various types of prolongation or Harakaat. The approach of Harakaat analysis concerns on sequences of voiced or unvoiced sound because of recitation are related to the prosody. The production of voice or speech involves the movement of air from the lung to vocal tract towards the lips. The combination of voice production mechanism produces a variety of vibration and spectral-temporal composition that produce different speech sound and its prosody. Apparently, the Arabic phonetic sound was produced from the specific articulation places or regions in the vocal tract.

MFCC features are widely used to determine the type or error for Al-Quran word pronunciation or Tajweed rules [10]. In addition, there are further discussion on the uses of MFCC as features in Harakaat recognition or identification [11], [12] and the prosody characteristics of speech signals associated with Harakaat [13]. All this emphasizes that it has a high potential to be used to obtain important characteristics. Among the successes reported, one of them is encouraging results in the assessment of English pronunciation [14]. Likewise with the successful use of prosody features in identifying the correct recitation of the Qur'an [15].

IV. Dynamic Time Warping

Speech is a time varying process [16] in which the duration of a word and its sub-words varies randomly. Time alignment may be applied in recitation which is required to find the best alignment and ratio between utterance sequences of expert and learner [17]. The recitation feature vectors represent the syllable similarly length on time occurrence. Dynamic Time Warping (DTW) is used to warp two feature vector sequences in time. DTW is well known in speech recognition to cope with different speaking speeds [18].DTW is used as feature matching with the deployment of the technique of minimum Euclidean distance or another distance-based approach [19]. DTW measures the similarity between the two temporal sequences which may vary in time [17]. This technique is also been used for finding the optimal alignment between the two-times series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis [20][21]. The recitation is also re-mapped as if all utterances were produced from the same vocal channel. The recitation speech signal produced by the learner is warped with respect to the Quranic speech signal that belonged to the expert by using the DTW algorithm. As a result, the sequence of Quranic speech signal of expert reference, X, with the length of |X|, and the learner, Y, with the length of |Y| can be expressed as in equation 1.

$$X = x_1, x_2, ..., x_i, ..., x_{|X|}$$

and

$$Y = y_1, y_2, ..., y_j, ..., y_{|Y|} \tag{1}$$

The warp path, W, is then can be constructed as,

$$W = w_1, w_2, ..., w_K$$

where

$$\max(|X|, |Y|) \le K < |X| + |Y| \tag{2}$$

where K is the length.

And, kth element of the warp path is $w_k = (i, j)$

where i = index of time series of X, and j = index of time series of Y.

The optimal warp path is presented as 'local match' scores matrix by getting the lowest-cost distance warp path and the first frame optimal warp start at $w_1 = (1,1)$ where the distance of a warp path is given by,

$$Dist(W) = \sum_{k=1}^{k=K} Dist(w_{ki}, w_{kj}) \tag{3}$$

$Dist(W)$ is the cosine distance of warp W, and $Dist(w_{ij}, w_{kj})$ is the distance between the two data frame indexes of X and Y in the kth element of the warp path.

The lowest-cost path for the first frame is $D(1,1) = 0$ and can be calculated by,

$$D(i,j) = Dist(i,j) + \min[D(i-1,j), D(i,j-1), D(i-1,j-1)] \tag{4}$$

The minimum-cost alignment is determined from the optimal warp path that ends at $D(|X|,|Y|)$ by looking at the lowest-cost warp path. The lowest-cost warp path is indicated by the ratio of STFT features vector between expert and reciter recitation signals. The search grid is used as a one-to-one monotonic transformation of the time axis in which all the movements have equal weight.

V. Features Extraction Design for Harakaat

To represent the parameters of measurement for the duration of syllable pronunciation, DTW of minimum path cost parameters coefficient is used. DTW calculates the local stretch on the time axis for two-series objects to optimally map one (query) to the other (reference) by calculating the distance between unequal sequences of length [19]. Minimum path cost alignment calculation uses the cosine distance [17] to determine the optimal warp path from the start point of syllable utterance until the end of an utterance. For each of syllable's utterance the search grid is used with one-to-one monotonic transformation along the time axis. The lowest-cost path is calculated from the optimal warp and known as minimum path cost alignment parameters co-efficient (refer to equation 4). The comparison between Quranic syllable-based Recitation Speech Signal (QRSS) syllables is done by the recitation ratio calculation between the expert and the learner recitation based on Short Time Frequency Transform (STFT) features. The STFT is given by

$$X(k,m) = \sum_{n=0}^{N-1} x(n+m)\omega(n)W_N^{nk}$$

Where k, m = 0. 1. …, N-1

ω(n) = time shifted window function.

m = the amount of shift. (5)

The value of the sequence distance represents as the Minimum Path Cost (MPC) for evaluating the syllable characteristics in each period referring to the Harakaat Tajweed rule feature. The minimum path cost (MPC) is used to obtain the optimal alignment by computing all the possible cost path and determine the lowest overall cost from the path to achieve high similarity between the two temporal recitation speech signals. Fig. 1 shows the minimum path cost (MPC) parameters with optimal path. The steps to calculate the minimum path cost parameters are given as follows [22]:

1) Load the QRSS from the expert as a reference template and followed by learner waveform of the same syllable.

2) Calculate the STFT features for both QRSSs (with 25% window overlaps).

3) Construct the 'local match' score matrix as the cosine distance between the STFT magnitudes.

4) Use dynamic programming to find the lowest-cost path between the opposite corners of the cost matrix.

5) Find the cost of minimum path cost parameters of the two signals of recitation speech.

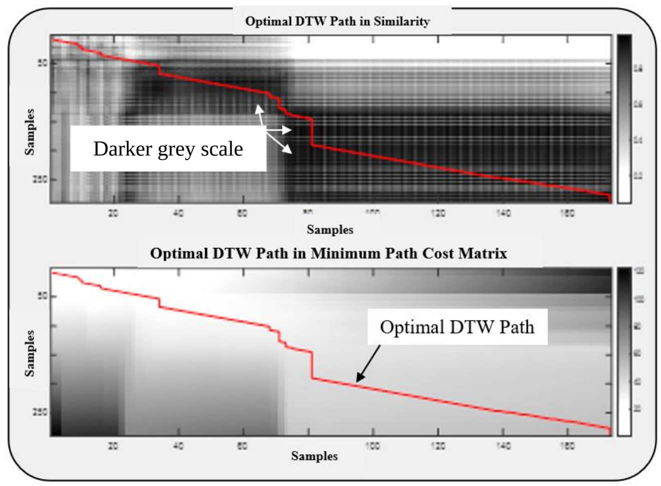


Fig. 1. The minimum path cost (MPC) parameters with optimal path.

The minimum path cost parameters value increases when the duration between two QRSSs is increased. Referring to the optimal DTW path in the similarity matrix, it can be observed that the darker grey scale shows a high similarity between the two QRSS readings. Thus, at the optimal DTW level in the minimum path cost, a lighter grey scale indicates a lower cost path (indicated in red) between the two QRSSs. The minimum path cost parameter is taken at the last value of the lowest cost path.

VI. Experiment

In the context of significant features of QRSS-syllable, the motivation of miniature feature creation remains towards the enrichment of the property characteristics by prosodic features that extracted by DTW. In this section, the *Harakaat* measurement is based on the time length for each syllable measured by using the DTW technique. Firstly, the QRSS-syllable is segmented into a number of frames and uniquely offering the Short Time Frequency Transform (STFT) feature [22] to be extracted using DTW technique. The cosine distance value for each frame is then calculated as the STFT miniature feature property (power energy). Later, the minimum cosine distance is obtained for every frame with the application of dynamic programming.

The minimum cosine distance is finally selected as the minimum path and is represented as the Minimum Path Cost (MPC) value. This means that the MPC parameters are selected to measure the Harakaat for QRSS-syllables. Fig. 2 shows the different values of MPC for different syllables recited by 54 reciters for the Harakaat measurement performed.

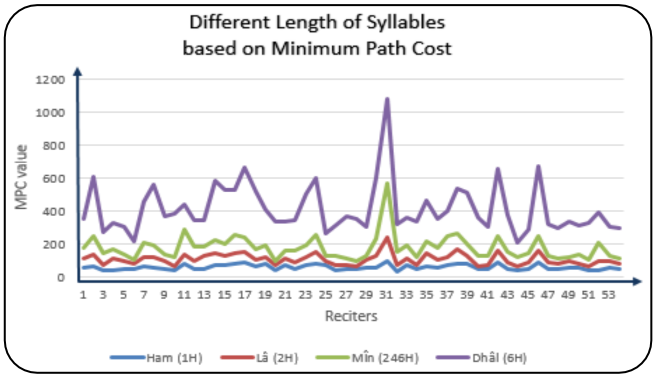


Fig. 2. Different syllables length based on Minimum Path Cost (MPC).

For example, the syllables *Ham*, *Lâ*, *Mîn* and *Dhâl* clearly indicate the lengths of time that differ from each other, which are one Harakaat (1H), two Harakaat (2H), four Harakaat (4H) and six Harakaat (6H) respectively. This is in line with the rule of Harakaat Tajweed for each of these syllables, which have different length of Harakaat respectively.

From the graph shown in Fig. 2, the syllable value of 1H has the lowest MPC value compared to 6H. As for 4H, the MPC values are between 2H and 6H, most of which are in the Tadweer category of speed. The speciality of syllable Min in having the multiple choices of Harakaat (2H, 4H and 6H) is limited to the two, four or six Harakaat (4H) for the measurement simplification.

The computational engine score threshold process is used to evaluate the similarity and the dissimilarity based on human-guided threshold process [23]. This successful threshold process shows how a conventional Talaqi process (experts evaluate the recitation by learners based on how to pronounce syllables in the verses of the Quran) are transformed to a machine evaluation approach. Computational engines must have salient features that can distinguish between correct and incorrect readings. Therefore, in determining the score, analysis of salient features, and matching process is used to obtain reading assessment based on the actual assessment by experts called a human-guided or Talaqi-like assessment [24].

Talaqi-like approach and process have been applied in the training and testing phases in the development of computational engine. This is to ensure that the assessment made by the machine are always guided by the human expertise. In this process, MPC parameters are used as representations to each syllable recited by the learners. Initially, the MPC values from recitations by experts were used as primary references in the threshold range assuming that all MPCs generated by expert readings were within acceptable thresholds by the system. Thereafter, learner’s recitations were assessed by comparison based on this reference threshold range.

In the training phase, the recitation results by learners were also evaluated by experts to confirm whether learners' recitation is acceptable or not, considering that everyone is still able to comply with the Law of Tajweed despite recitation with different levels of voice. Therefore, the initial value of the threshold originally set for the machine will continue to change (i.e. not necessarily fixed) as long as the expert evaluating the recitation complying with the applicable Tajweed Law.

The training process on this machine with human-guided assessment continue to be repeated until all MPC data have been evaluated by a leading expert (human). After the training process is over, the value of the threshold range has been fully obtained and can be used as a benchmark for the recitations made by learners, that is, whether it is correct or otherwise (within or outside the range). The range for this threshold values is then used and tested for accuracy in the testing phase. The testing performance of this Talaqi-like approach is be measured and justified. Fig. 3 shows the Talaqi-Like assessment flow process.

For Harakaat validation, there are four types of classes are tested, namely as 1H, 2H, 4H and 6H according to their prolongation types.

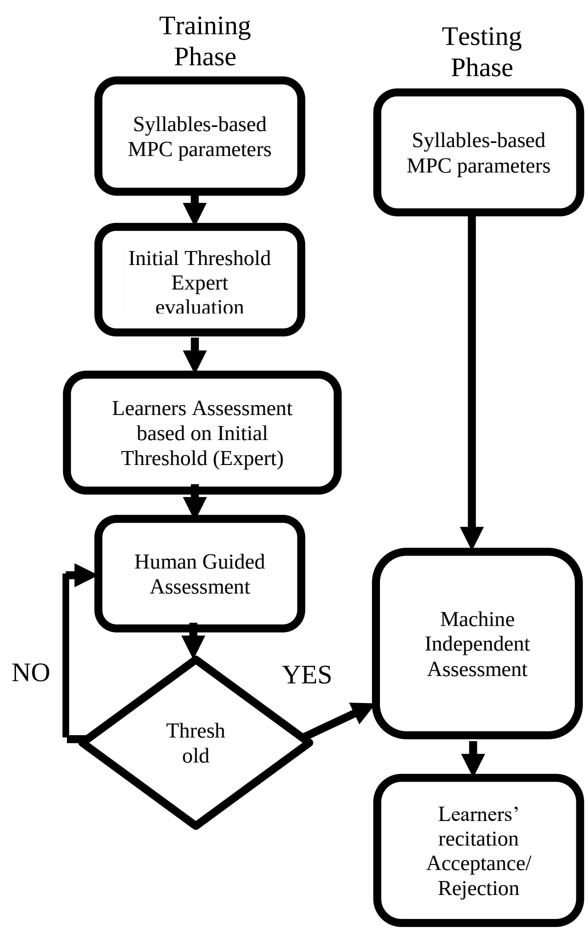


Fig. 3. The Talaqi-like assessment (Human Guided) flow process for Harakaat assessment.

In this classification approach, cross-validation is used to avoid overfitting of the model or class prediction [25]. Cross-validation is a resampling procedure that used to evaluate machine learning models on a limited data sample. It protects the overfitting by partitioning the data set into the fold and estimating the accuracy of each fold. This classification procedure used is k-fold cross-validation which data is divided into k-number of groups. The number of k-group that used in this classification is 5-fold.

The performance of classification is shown with the use of confusion matrix [26] by finding the True Positive Rate (TPR) and False Negative Rate (FNR). Both TPR and FNR are used to indicate the MPC parameters belong to the class or out from the class for each syllable Tajweed rules group or type. TPR shows that the true parameters of MPCs are included in a positive class. While FNR shows that true parameters of MPCs are included in a negative class. In other words, TPR and FNR indicate that parameters of MPCs are classified in the right or wrong classification, respectively.

Confusion matrix is a relatively popular measurement in summarizing the ability or performance of a group of algorithms in classification tasks [26]. In fact, in many cases, the accuracy of a classification result, can produce confusion if it has inconsistencies in the observations that need to be implemented for each classification. The main idea of employing the confusion matrix remains as commonly practice in any machine learning application. The result of the calculation of the confusion matrix is not only able to provide a better perspective on the modelling of a classification, but also at the same time able to identify the form or type of mistakes that have been made. Table I shows the classification performances by Linear Discriminant (LD), Support Vector Machine (SVM) and K-Nearest Network (KNN) for Harakaat measurement using MPC.

TABLE I. CLASSIFICATION PERFORMANCE OF HARAKAAT USING DIFFERENT CLASSIFICATION

Harakaat	LD (%)	SVM ((%)	KNN(%)
1H	98	99.4	96.8
2H	0	1.4	72.6
4H	7.4	13.5	75.5
6H	84.6	94.2	92.3

VII. RESULT AND DISCUSSION

The acquisition of MPC values can be considered as similarity parameters representing MPC the unique significant features of STFT features. Therefore, the training and testing phase classification are developed to assess the *Harakaat* Al-Quran recitation. However, the assessment of this *Harakaat* is limited to recitation from Malay adults because the data used in establishing the threshold range involves only Malay experts and students. This means that the threshold used is limited to the entire voice of Malay adults only.

In this training phase of the classification stage, the initial parameters of MPC are taken from the calculation of 12 Malay expert recitations to create an initial threshold range along with maximum and minimum values. As a result, the Acceptance Threshold (AT) is defined as,

$$\text{MinMPC} << y << \text{MaxMPC} \tag{6}$$

while y = accepted threshold MPC parameters.

Each syllable is represented by MPC parameters. These parameters are used to determine the pronunciation of the syllable is pronounced correctly or not, based on the rules of Harakaat Tajweed. These AT values are used based on the minimum and maximum value ranges of MPC to evaluate the syllables related to Harakaat. The starting point of training phase is when the input given to this designed system begins to create a change of pattern or minimum and maximum value that limits the correctness of a Tajweed in the reading of the surah. This is seemingly caused by the changes of the acceptable lowest and highest values of MPCs that correspondingly due to the variability demonstrated by various reciters but remains accepted (Acceptance Threshold) by the expert.

The process of correcting (or training) the minimum and maximum values (threshold range) is firstly performed on the group of experts' MPCs data. This is the initial threshold range and used as reference values to be compared with the learner recitations. Secondly, the MPCs values obtained from the recited syllables of 40 learners are matched with the expert threshold range. Besides the setting of minimum and maximum values, the indication of True Acceptance (TA), False Rejection (FR) and False Acceptance (FA) of the calculated MPCs are counted and accumulated.

Table II tabulates the MPCs values of syllables verse-2 of Al-Fatihah recited by 40 learners that have been matched with the threshold range of expert's recitations. Referring to the data shown, performance on Harakaat assessment for each syllable are highly accepted for syllable S5. It shows that the assessment based on expert and learners MPCs parameters are acceptable by using this approached.

TABLE II. TABULATES THE MPCS VALUES OF SYLLABLES VERSE-2 OF AL-FATIHAH

Range and Result Indication		Expert Max	Expert Min	TA	FR	FA	Performance (%)
Syllables of Verse-2	S1	68	25	34	6	0	85
	S2	77	39	30	10	0	75
	S3	63	16	36	4	0	90
	S4	53	14	31	8	1	78
	S5	69	25	22	17	1	55
	S6	60	22	33	7	0	83
	S7	36	14	31	9	0	78
	S8	75	17	36	4	0	90
	S9	89	44	31	7	2	78
	S10	61	27	29	11	0	73
	S11	152	30	33	6	1	83

In the testing phase, the main objective is to test the computational engine (from the training phase) that has been designed from the context of reliability of the miniature salient feature, extractor and classifier. The trained range of MPC is used to assess the performance of test data. Each syllable is tested according to the threshold determined based on MPC (maximum and minimum value). A total of 40 different learners from the training phase took their readings and the readings of each syllable in Al-Fatihah were extracted and matched with the reference MPC from the training phase. Each test data is also evaluated manually by an expert and the performance of the reading truth that refers to Tajweed rules is calculated in a technical context, namely true and false positive (TP and FP), false rejection (FR) and false acceptance (FA).

The comparison of errors was made and analyzed between the machine evaluation and human evaluation. From the results, the performance of the machine as an evaluator is then compared with respect to human expert performance.

Table III shows the performance of the parameters of MPCs that can be accepted, which consists of true acceptance (TA) and false acceptance (FA) in the testing phase. From that table, each syllable has been evaluated in terms of good of pronunciation (GOP) [25] for the testing data from Al-Fatihah verses based on Human Guided Assessment threshold from training phase.

In the test phase, the average performance show that each verse in the testing data is lower than training data. However, the average performance for testing data is still above 80% for each MPCs parameter and still shows good performance of Al-Quran recitation assessment. Fig. 4 visualizes the comparison between the performance percentages of training phase and testing phase where the testing gives a lower percentage in performance.

TABLE III. SHOWS THE PERFORMANCE OF THE PARAMETERS OF MPCS TESTING PHASE

Syllables verse 2	Human Machines Performance (%)
S1	87
S2	62
S3	92
S4	85
S5	95
S6	92
S7	85
S8	79
S9	64
S10	85
S11	64
Overall Score (%)	81

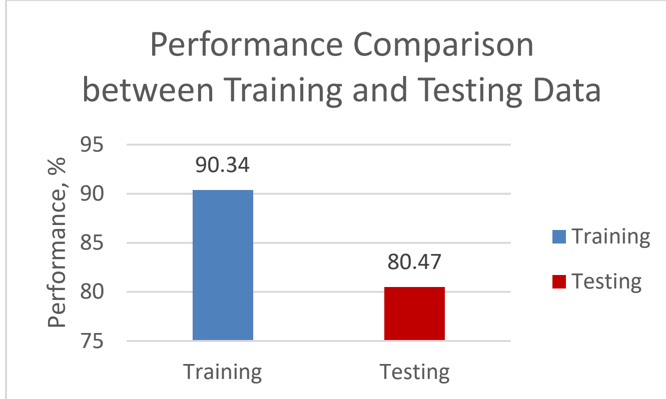


Fig. 4. The performance percentages of training phase and testing phase.

VIII. CONCLUSION

The classification performance using the threshold method for these MPC parameters can be used to evaluate syllable - based Al-Quran recitation where Tajweed Harakaat rules are embedded in the syllable. These features show good performance over 80% in testing phase for representing Tajweed Harakaat rules using DTW distance-features.The threshold method is very reliable in the training and testing phases giving a 90.34% and 80.47% respectively. For future work, the range of thresholds from various countries that are experts in reciting the Qur'an with the best Tajweed-based recitation can be included in experiments to improve computerized evaluations that are guided or learned from human expertise. In addition to that, the threshold for the entire verses of the Qur'an can be expanded.

ACKNOWLEDGMENT

The authors would like to express greatest appreciation to Universiti Teknologi Malaysia (UTM) and Ministry of Higher Education (MOHE), Malaysia for the financial support (grant number: Q.K130000.2456.08G28). We also acknowledged the support from the Cyber-physical Research Group and Advanced Informatics at Razak Faculty of Technology and Informatics (FTIR), UTM in this research work conducting this research.

REFERENCES

[1] T. Ridwan and N. Majid, “Development System for Recognize Tajweed in Qur’an using Automatic Speech Recognition,” in ICSTI 2018, 2018, doi: 10.4108/eai.19-10-2018.2281368.

[2] N. Shafie, M. Z. Adam, H. Abas, and A. Azizan, “Sequential classification for articulation and Co-articulation classes of Al-Quran syllables pronunciations based on GMM-MLLR,” AIP Conf. Proc., vol. 2401, no. October, 2021, doi: 10.1063/5.0072594.

[3] M. M. Al Anazi and O. R. Shahin, “A Machine Learning Model for the Identification of the Holy Quran Reciter Utilizing K-Nearest Neighbor and Artificial Neural Networks,” Inf. Sci. Lett., vol. 11, no. 4, pp. 1093–1102, 2022, doi: 10.18576/isl/110410.

[4] Y. Mohamed, M. Hoque, T. H. S. Bin Ismail, M. H. Ibrahim, N. M. Saad, and N. N. M. Zaidi, “Relationship Between Phonology, Phonetics, and Tajweed: A Literature Review,” in Proceedings of the 4th International Conference on Sustainable Innovation 2020, 2020, vol. 518, no. ICoSIHESS 2020, pp. 407–411, doi: 10.2991/assehr.k.210120.153.

[5] A. M. Alagrami and M. M. Eljazzar, “SMARTAJWEED Automatic Recognition of Arabic Quranic Recitation Rules,” AIRCCSE 2020, pp. 145–152, 2020, doi: 10.5121/csit.2020.101812.

[6] D. Darmawan, A. Hanifan, R. Sri, and R. Waluyajati, “Learning to Recite the Qur ’ an through Mobile Applications,” in 1st International Conference on Islam, Science and Technology, ICONQUHAS, 2020, doi: 10.4108/eai.2-10-2018.2295515.

[7] H. M. A. Tabbaa and B. Soudan, “Computer-Aided Training for Quranic Recitation,” Procedia - Soc. Behav. Sci., vol. 192, pp. 778–787, 2015, doi: 10.1016/j.sbspro.2015.06.092.

[8] M. K. Ilya Husna, Hidayati Azkiya, Suci Ramadhananti Febriani, Rahmawati, “The Analysis of Arabic Language Textbookfor Class II Elementary School,” Lect. Pendidik., vol. 13, no. 2, pp. 234–245, 2022.

[9] A. Mohammed, M. Shahrizal, B. Sunar, and S. H. Salam, “Recognition of Holy Quran Recitation Rules Using Phoneme Duration,” Recent Trends Inf. Commun. Technol., vol. 1, 2018, doi: 10.1007/978-3-319-59427-9.

[10] M. Computing, M. Rashwan, and S. Abdou, “Automatic Speech Recognition (ASR) Systems for Learning Arabic Language and Al-Quran Recitation : A Review,” vol. 10, no. 7, pp. 91–100, 2021, doi: 10.47760/ijcsmc.2021.v10i07.013.

[11] B. Yousfi, A. M. Zeki, and A. Haji, “Holy Qur ’ an Speech Recognition System Distinguishing the Type of prolongation,” Sukkur IBA J. Comput. Math. Sci., vol. 2, no. 1, 2018.

[12] S. S. Alrumiah and A. A. Al-shargabi, “Intelligent Quran Recitation Recognition and Verification : Research Trends and Open Issues,” Arab. J. Sci. Eng., 2022, doi: 10.1007/s13369-022-07273-8.

[13] B. Claire, M. Sawalha, T. Islam, J. Dickins, and E. Atwell, “Automatic Extraction of Quranic Lexis Representing Two Different Notions of Linguistic Salience : Keyness And Prosodic Prominence,” J. Semit. Stud., 2018, doi: 10.1093/jss/fgy005.

[14] F. Chao, T. Lo, T. Wu, Y. Sung, and B. Chen, “Multi-aspect Modeling Approach to English Pronunciation Assessment.”

[15] N. Bettayeb and M. Guerti, “Speech Synthesis System for the Holy Quran Recitation,” Int. Arab J. Inf. Technol., vol. 18, no. 1, pp. 8–15, 2021.

[16] J. Zhao, X. Mao, and L. Chen, “Speech emotion recognition using deep 1D & 2D CNN LSTM networks,” Biomed. Signal Process. Control, vol. 47, pp. 312–323, 2019, doi: 10.1016/j.bspc.2018.08.035.

[17] N. Shafie, M. Z. Adam, and H. Abas, “Al-Quran Recitation Speech Signals Time Series Segmentation for Speaker Adaptation using Dynamic Time Warping,” J. Fundam. Appl. Sci., vol. 10, pp. 126–137, 2018, doi: 10.4314/jfas.v10i2s.11.

[18] M. Mundada and S. N., “Analysis of Speech and its Fluency Disorders,” Int. J. Appl. Inf. Syst., vol. 10, no. 4, pp. 10–12, 2016, doi: 10.5120/ijais2016451484.

[19] T. Li, X. Wu, and J. Zhang, “Time series clustering model based on DTW for classifying car parks,” Algorithms, vol. 13, no. 3, 2020, doi: 10.3390/a13030057.

[20] B. J. Mohan and R. Babu, “Speech Recognition using MFCC and DTW,” in 2014 International Conference on Advances in Electrical Engineering (ICAEE), 2014, vol. 4, no. 3, pp. 1–4, doi: 10.1109/ICAEE.2014.6838564.

[21] N. Shafie, M. Z. Adam, S. Mohd Daud, and H. Abas, “A Model of Correction Mapping for Al-Quran Recitation Performance Evaluation Engine,” Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, pp. 208–213, 2019.

[22] A. H. A. Absa, M. Deriche, M. Elshafei-Ahmed, Y. M. Elhadj, and B. H. Juang, “A hybrid unsupervised segmentation algorithm for arabic speech using feature fusion and a genetic algorithm (July 2018),” IEEE Access, vol. 6, no. July, pp. 43157–43169, 2018, doi: 10.1109/ACCESS.2018.2859631.

[23] M. Maqsood, A. Habib, and T. Nawaz, “An efficient mispronunciation detection system using discriminative acoustic phonetic features for Arabic consonants,” Int. Arab J. Inf. Technol., vol. 16, no. 2, pp. 242–250, Mar. 2019.

[24] M. Z. Adam, N. Shafie, H. Abas, and A. Azizan, “Analysis of Momentous Fragmentary Formants in Talaqi-like Neoteric Assessment of Quran Recitation using MFCC Miniature Features of Quranic Syllables,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 9, pp. 533–540, 2021, doi: 10.14569/IJACSA.2021.0120960.

[25] S. Kanters, C. Cucchiarini, and H. Strik, “The goodness of pronunciation algorithm: a detailed performance study.,” Speech Lang. Technol. Educ. -SLaTE, no. 2, pp. 2–5, 2009, [Online]. Available: http://www.eee.bham.ac.uk/SLaTE2009/papers%5CSLaTE2009-33.pdf.

[26] S. Z. H. Naqvi, S. Aziz, M. H. Tariq, M. U. Khan, H. A. Aslam, and M. A. Imtiaz, “Effect of Al-Quran Recitation on Human Physiology,” 2nd Int. Conf. Electr. Commun. Comput. Eng. ICECCE 2020, no. June, 2020, doi: 10.1109/ICECCE49384.2020.9179455.

A Novel Method for Recognizing Traffic Signs using Color and Texture Properties using the ELM Algorithm

Xiaoda Cao*
Laboratory Management Department
Northeast Agricultural University, Harbin, Heilongjiang, 150030, China

***Abstract*—Road accidents cause a lot of financial and human losses every year. One of the causes of these accidents is human error, and the driver ignores traffic signs. Therefore, accurate detection of these signs will help to increase the safety of drivers and pedestrians and reduce accidents. In recent years, much research has been done to increase the accuracy of panel recognition, most of which are problems that affect the diagnosis, such as adverse weather conditions, light reflection, and complex backgrounds. In the present study, considering the diversity of traffic signs' geometric shapes, the sign detection part has been done using a torsional neural network. Then, in the feature extraction section, we used LBP and HOG techniques, and at the end, the section was identified and classified using the ELM algorithm. The results obtained on 12569 images, 75% of which were used for training and 25% for experimentation, show that the accuracy of this research has improved by 95% compared to the essential work by 93%.**

Keywords—Traffic sign recognition; torsional neural network; HOG Feature; LBP Feature; ELM Algorithm

I. Introduction

One of the most critical issues in car vision and pattern recognition, which has attracted the attention of many researchers today, is detecting traffic signs through images. Ignoring traffic signs is a disservice; many people die in accidents or disabilities yearly. With the advancement of technology and the competition of car factories, much research and work has been done on smart cars. Traffic signs increase traffic safety on roads and streets by giving drivers warnings. Therefore, automatic identification of traffic signs is one of the important components of a driver assistance system. The next generation of vehicles can also be considered an important component of automatic reduction vehicles. This system must have high speed and accuracy and perform real-time detection of symptoms in natural scenes.

Automatic detection of traffic signs is one of the problems of intelligent transportation systems today, which facilitates the recognition and interpretation of signs for users. It helps drivers identify traffic signs and pedestrians by warning them [1, 2].

In recent years, a great deal of research has been done to identify signs by intelligent vehicles. Factors such as bad weather, unfavorable weather conditions, light reflection and misdiagnosis of signs, placement of traffic signs and driving in the shade, placement of signs between leaves of trees, and very complex backgrounds such as the location of buildings, trees, and animals that complicate the background and the background is similar to the panel for detecting problematic symptoms. Because it takes a long time to separate the images from each other and correctly identify the traffic signs, the traffic sign recognition system must be prompt.

The challenge of recognizing traffic signs is their accuracy and short detection time. Much works have been done to detect traffic signs [1, 2], and due to the noise in the image, camera slip, and image quality, all of them have been aimed at accurately detecting traffic signs. In the research reviewed [1, 2], a traffic sign recognition system consists of three main identification steps: image reception, pre-processing, and detection of the area of traffic signs in the image. Early detection of traffic signs and their accuracy can help reduce accidents and human mortality [23], so if traffic signs can be detected intelligently by vehicles, it can be a great help in reducing road accidents.

The motivation of this study is that with the idea of recognizing traffic signs from standard data images. Since traffic sign detection is still a challenging task because of adverse weather conditions, light reflection, and complex backgrounds, it is required to establish a method to provide high accuracy in diagnosis by finding different features in traffic sign images.

The main purpose is to increase the accuracy of existing methods and to detect traffic signs according to the characteristics of color and texture. On the other hand, the use of ELM classification has been used to identify the shape of the symptoms better as well as to extract and pay attention to features such as color, shape, and texture [3].

The major contributions of this study are as, a) proposing a novel feature representation algorithm using color and texture features for finding different features to represent the traffic sign images, b) presenting a method for classifying traffic signs using color and PHOG features the SVM algorithm, c) developing a sign traffic detection method to deal with adverse weather conditions, light reflection, and complex backgrounds.

The rest of this paper is consisted of as, Section II presents the related works. The proposed method is described in Section III. Experimental results and performance analysis are discussed in Section IV. Finally, this paper concludes in Section V.

*Corresponding Author.

II. Related Works

Zhang et al., [29] proposed a cascaded R-CNN to extract the multiscale features in pyramids in order to address undetection and erroneous detection. With the exception of the first layer, each layer of the cascaded network fuses the output bounding box of the preceding layer for joint training. This technique aids in the identification of traffic signs. In order to highlight the features of traffic signs and increase the accuracy of traffic sign detection, a multiscale attention approach is presented to extract the weighted multiscale features using dot-product and softmax. Finally, to reduce the influence from a complicated environment and comparable fake traffic signs, we increase the amount of challenging negative examples for dataset balance and data augmentation in the training.

Avramović et al., [30] presented a traffic sign detection approach based on "You Only Look Once" (YOLO) architectures as baseline detectors for improving the speed and accuracy of traffic sign identification and recognition in high-definition photographs. To meet the real-time performance requirement, a number of preprocessing techniques were suggested. To evaluate the method, tests on a big dataset of traffic signs demonstrate the ability of the method to recognize high-definition images in real time with high recognition accuracy.

Cao et al., [31 proposed an algorithm for traffic sign detection and dealing with the poor real-time performance of deep learning-based traffic sign recognition techniques. In this method, spatial threshold segmentation is firstly performed using the HSV color space, and traffic signs are successfully identified using shape features. Second, utilizing the Gabor kernel as the initial convolutional kernel, adding the batch normalization processing after the pooling layer, and choosing the Adam method. As their results show, the model is much improved over the original LeNet-5 convolutional neural network model. The German Traffic Sign Recognition Benchmark serves as the foundation for the classification and recognition experiments on traffic signs.

III. Proposed Method

Initially, after downloading images from the GTSRB database [5-8], a torsional neural network is used to detect traffic signs. Extraction of shape properties by HOG technique which is a shape descriptor, based on histogram and statistical information of image spectra as well as its inclination and angles, and finally by extracting texture properties along with shape properties by LBP technique which can increase the sensitivity of HOG, extract light reflections and improve cluttered and complex backgrounds. The features are categorized using the ELM algorithm by forming a feature matrix to detect the panel type and normalizing this matrix, which is also the input matrix to the category. Finally, the accuracy is checked [9-14]. Fig. 1 shows the steps of the proposed method as a flowchart.

A. Torsional Neural Network

Twisted neural networks are very similar to artificial neural networks. These networks consist of neurons with learnable (adjustable) weights and biases [15]. Each neuron receives several inputs and then calculates the product of the weights multiplied by the inputs and finally presents a result using a nonlinear conversion (activation) function [16-22]. The entire network also provides a derivative score function, with raw input image pixels on one side and points for each category on the other. These types of networks still have a fully operational cost function (SVM, Softmax) in the last layer, and all the points about conventional neural networks are also valid here. Given the above, the difference between a torsional neural network and an artificial neural network is that torsional neural network architectures explicitly assume that their inputs are images, assuming certain features can be embedded within the architecture. With this action, the forward function can be implemented more optimally, and also, by doing so, the amount of network parameters is significantly reduced [24-26].

Neural networks receive an input (in the form of a vector) and then pass it through several hidden layers. Finally, an output resulting from hidden processing layers appears in the network output layer. Each hidden layer is made up of some neurons that are connected to all the neurons in the previous layer. The neurons in each layer act independently and have no connection. The last layer plays the role of representing the score of each class [27, 28].

B. Characteristics of Texture and LBP Technique

The texture is the measurement of changes in the brightness of surfaces, which determines the quantity level for properties such as smoothness and order. An essential feature of the texture is that it is less sensitive to changes in light intensity than the color feature. Compared to color, texture requires a preprocessing step to produce texture descriptors. There are different texture descriptors. Gray Level Cooccurrence Matrix is a two-dimensional histogram that expresses simultaneous events expressing image intensity values in a given direction and distance.

The local binary pattern (LBP) operator was first proposed by Ojala et al. [26]. Experiments have shown that the LBP operator has a high ability to represent tissue. The LBP algorithm is a local feature extraction algorithm. The LBP algorithm is one of the most potent feature extraction algorithms in machine vision science, and many developments based on this algorithm have been proposed.

As shown in Eq. (1), the local properties of LBP are usually symmetrical circles in a neighborhood, comparing images with its neighbors.

$$\mathrm{LBP(P,R)} = \sum_{i=0}^{P-1} u(g_i - g_c)2^i \tag{1}$$

In Eq. (1), p is the number of neighborhoods, R is the neighborhood radius, gi is the brightness of the *i-th* pixel ($i = 0, \ldots, P-1$), gc is the brightness of the central pixel, and $u(x)$ is a step function that has the relation (Eq. 2) is defined.

$$\mathrm{U(x)} = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

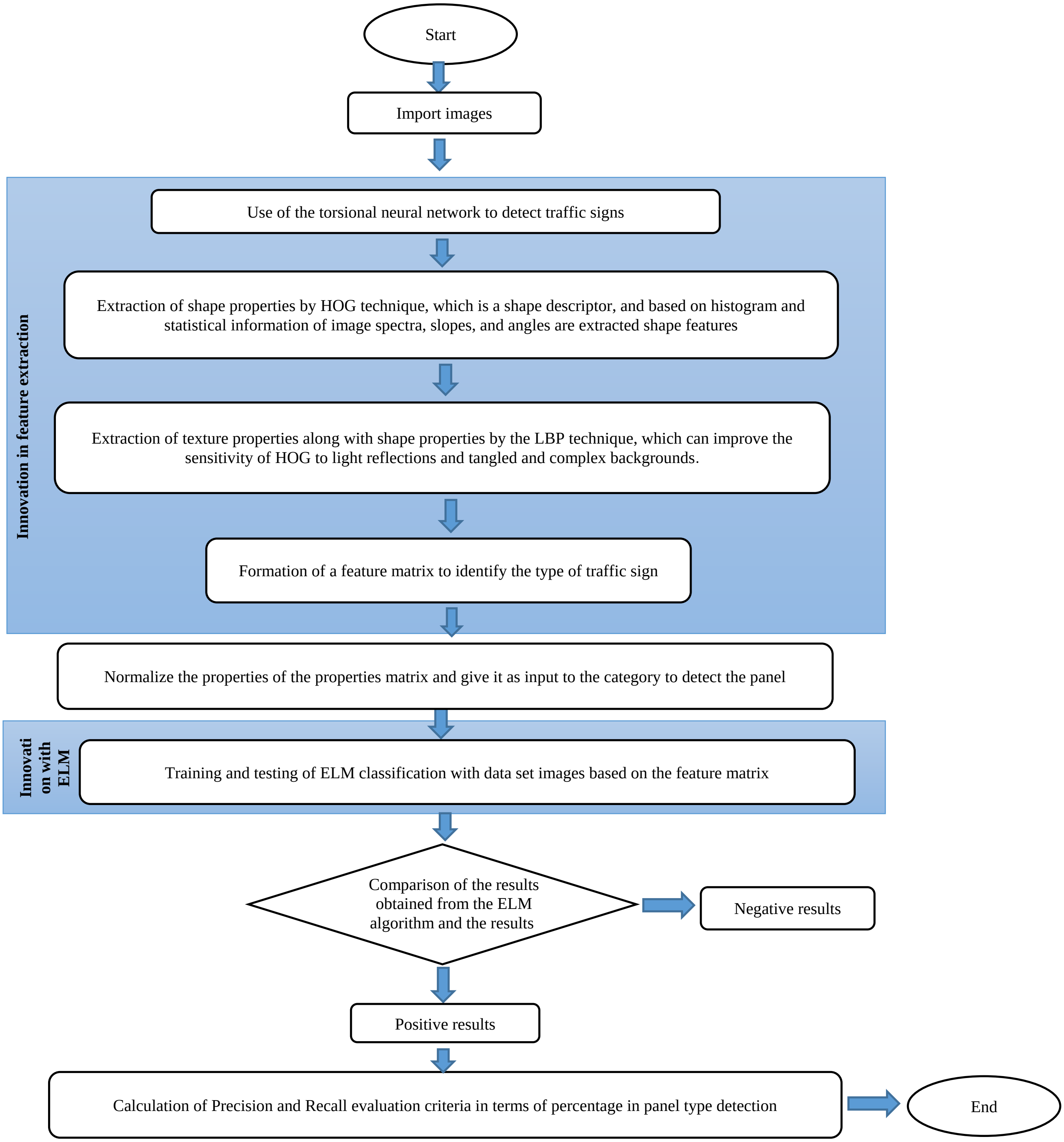


Fig. 1. Flowchart steps of the proposed method.

Suppose eight neighborhoods are used to extract LBP patterns. In this case, 56 uniformly significant patterns and two patterns for the state where the brightness of the neighborhoods is all more or less than the central pixel, and a pattern for the other states is considered. In total, image pixels are tagged with 59 patterns. The normalized histogram of the repetition of these patterns constitutes the LBP features. The minimum amount of binary pattern produced is used to make the patterns more resistant to rotation; therefore, the LBP pattern is resistant to the period in the form of Error! Reference source not found is defined.

$$LBP^{riu2}(P,R) = \begin{cases} \sum_{i=0}^{P-1} u(g_i - g_c) & \text{if } U(LBP(P,R)) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \quad (3)$$

In this case, the number of significant patterns will be seven. As a result, image pixels are tagged with 10 patterns. This paper uses period-resistant LBP patterns. In recognizing traffic signs, searching for a way that is easy to identify and has good accuracy is of great importance. In this paper, the torsional neural network method is used for detection, and then using LBP and HOG; the features were extracted after normalization of the feature matrix as input to the ELM algorithm for classification.

IV. Analysis and Interpretation of Evaluation Results

This research was implemented with MATLAB software version 8.3 and on a system with Intel Core i7 specifications with 2.2GH processing power and 6GB of main memory. The database used in this article is the GTSRB database [5-8]. Considering that the essential work [1] has been implemented on 1000 images from this database while implementing on 1000 images to compare with the essential work on 12569 images in the GTSRB database [5-8], the proposed method has been implemented.

A. Evaluation Data Set

The data used in this study compared with the essential work of 1000 images, of which 800 are noise-free images and 200 are noise-free images to implement the proposed method on all images. The database contains 12569 images downloaded from the GTSRB [5-8], among which 60 are related to speed limit data of 20 km / h, and 720 are related to traffic signs. The speed limit is 30 km / h, 750 pieces of data are related to traffic signs, the speed limit is 50 km / h, 450 pieces of data are related to traffic signs, the speed limit is 60 km / h, and 660 pieces are data. Traffic signs are a speed limit of 70 km / h, 600 pieces of data are related to traffic signs' speed limit of 80 km / h, and 120 data are related to traffic signs with a speed limit of 80 km / h, 480. The number of data related to traffic signs is 100 km / h, 450 of the data related to traffic signs. The speed limit is 120 km / h, 480 pieces of data are related to traffic signs, overtaking is prohibited, 660 images are related to prohibited truck overtaking, 420 pieces of data are related to a two-way intersection, 690 Images of the main street, 720 images of a right of way, 240 pieces of data related to traffic stop signs, 210 pictures of no entry signs, 120 pieces of data related to traffic signs of no-entry trucks, 360 pieces of data related to no-entry traffic signs, 390 pieces of data related to traffic signs, 60 pieces of data related to traffic signs to the left, 120 pieces of data related to traffic signs to the right, 90 images of left turn on the road, 120 pieces of data Bump traffic signs, 150 data slippery road traffic signs, 90 road images approaching from the right, 480 data from workers' traffic signs working, 180 warning images to traffic lights. As we approach the traffic, 60 pieces of data related to traffic signs at the crossing Cloud pedestrian, 180 warning pictures of children playing, 90 warning pictures of cyclists, 150 pieces of data related to frozen road data, 240 pieces of data related to traffic signs for animal crossings, 60 stop signs, 209 number of data related to right-turn traffic signs, 120 data related to left-turn traffic signs, 390 data related to traffic signs for forwarding movement, 120 pieces of data related to two-way intersection traffic signs, 70 news images of entering the street from the right, 690 news images of the one-way road from the right, 90 images of the one-way road from the left, 120 pieces of data related to traffic signs, 60 pieces of data related to traffic signs allowed overtaking, 60 images No truck overtaking is prohibited.

B. Analyze and Evaluate the Results

First, the images are detected using a torsional neural network described in detail. As described in the previous sections, the torsional neural network is an improved version of the artificial neural network. There is a difference in the structure and operation of this type of network. Unlike an artificial neural network, the input matrix is the intensity of image brightness in a torsional neural network. The features are extracted by mapping and reducing the dimension of this matrix, and classification and detection operations are performed. If the feature extraction operation is performed separately in the artificial neural network, the classification operation will be performed after the feature extraction and normalization of the features. In fact, in torsional neural networks, the spatial dependence of the data is essential for network training and feature extraction from the input matrix. Therefore, this type of network can be used in problems where neighborhood information and spatial dependence make sense. Fig. 2 shows how this detection occurs. After the segmentation of images and detection of traffic signs by the torsional neural network, features are extracted from the detected areas to determine the shape of the traffic sign. Properties are extracted using the texture method and LBP technique, and also the HOG method is used to extract the properties and form the properties matrix. In this paper, we compare the primary article method [1], which uses the PHOG method and different color channels for extraction and SVM algorithm for classification [4], and the proposed method, which uses LBP and HOG for feature extraction and ELM for classification. The baseline work used 1000 images for evaluation, using 800 images with noise or negative images and 200 non-noise images or positive images. Table I shows the comparison of the two methods.

In Table I, as can be seen, four different shapes are presented with different proposed methods and different categories, and the results show that the proposed method has very acceptable accuracy.

The following matrix, called the perturbation matrix, has five rows and five columns of the number of data tested, and their results are shown. In this matrix, the horizontal axis of the actual output or the class label is given, and the vertical axis of the output estimated by the category is shown.

TABLE I. Comparison of the Proposed Idea and the Basic Work

Row	Shape type	Color channel method and PHOG with SVM classification [1]	LBP and HOG methods along with ELM algorithm
1	circular	93.52%	100%
2	Triangle	96.45%	95.8%
3	Inverted triangle	94.50%	100%
4	Diamond	95.16%	100%
5	Average	95.02%	98.95%

Fig. 2. Sample GTSRB database images.

An example of the output of a traffic sign detection algorithm is shown in Fig. 3.

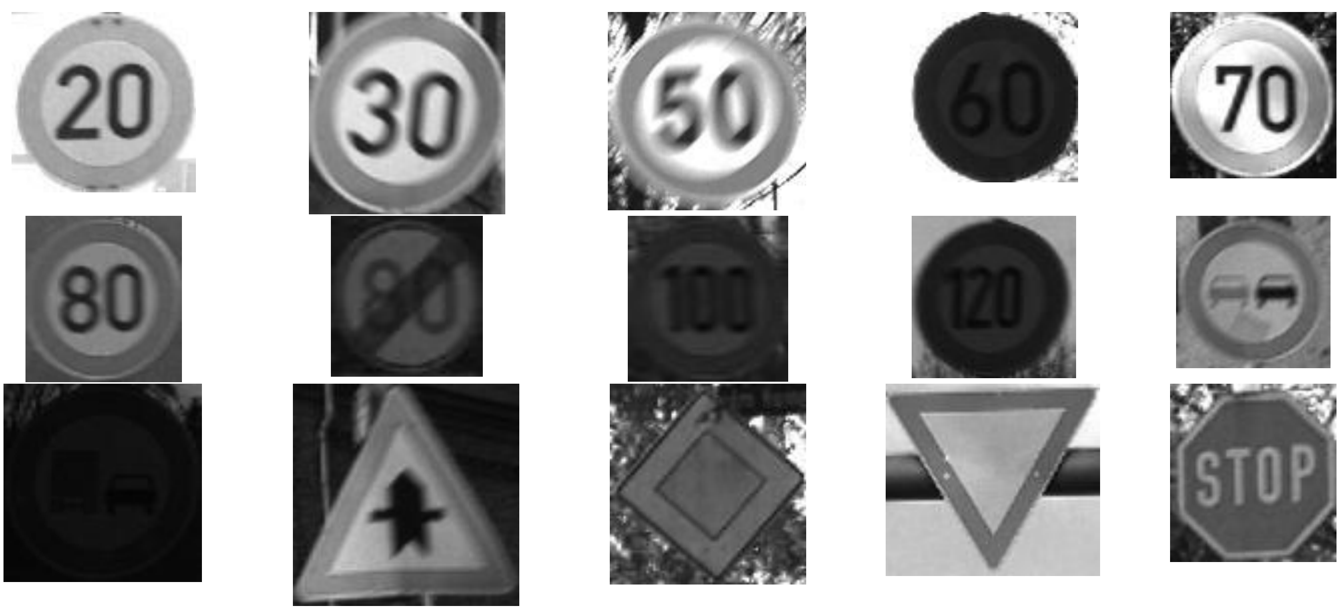


Fig. 3. Traffic signs detected by the proposed method.

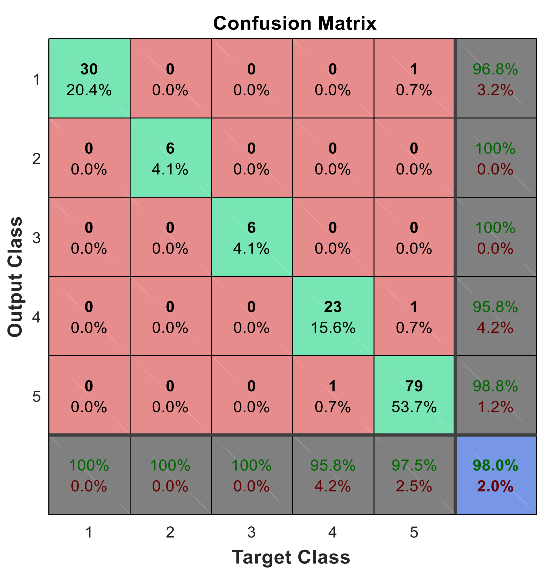


Fig. 4. The proposed method perturbation matrix is based on the basic paper data.

As shown in Fig. 4, in the first 30 categories of data related to circular traffic signs, the category correctly places the data in its group, and the classification accuracy for this group is 100%. In the second group, which is related to rhombus-shaped traffic signs, category six has correctly categorized the data related to this category. The classification accuracy is 100% for this group. In the third group, which is related to triangular traffic signs, and vice versa, six data are categorized correctly, and the classification accuracy for this group is 100%. In the fourth group, related to simple triangular traffic signs, out of 24 data, one case was incorrectly classified in group five and the rest correctly. The classification accuracy for this group is 95.8% out of which 81% data is related to the background model. One case is incorrect in category four. One data is incorrectly categorized in category one, the rest of the data is categorized correctly, and the classification accuracy for this group is 97.5%. The absolute accuracy in this category is 98%. Also, in the following Fig. 5, the ROC diagram of the proposed method is shown on the data set introduced in [5].

Given that in the primary working method [1], the data used is selected. In order to prove the suitability of the proposed methods of extracting the proposed feature and algorithm, extracting all the images related to the GTSRB site, which contains 12569 images, has been used for implementation. A base of 1000 images has been used. After image detection by the torsional neural network, the properties are extracted by LBP and HOG. After normalizing the images and forming the feature matrix, the obtained matrix is given as input to the ELM classifier for classification and then output in 43 classes. In this study, 75% of the images were used for training, and 25% were used for experiments. Table II shows the execution times of the different parts of the proposed algorithm.

As shown in Table II, in addition to the appropriate accuracy and high generalizability of the Maximum Learning Machine category, the speed of immediate implementation of this category has been measured, which has not been considered in the essential work. Another criterion used to evaluate the algorithm's performance is the classification accuracy, which in the proposed method for classifying data into 43 accuracy classes is 95.31%.

For this evaluation, the k-fold approach was used, where the value of k was considered equal to 10. At each stage, 90% of the data is set aside for training and 10% for testing. This process is repeated 10 times to consider all the data for testing and training the system which will be averaged on the accuracy obtained. The ROC diagram for this category is shown in Fig. 6. In this diagram, the larger the area under the curve, the higher the classification accuracy. The pivotal components of this chart are the Misdiagnosis Rate (FPR) and the True Diagnosis Rate (TPR).

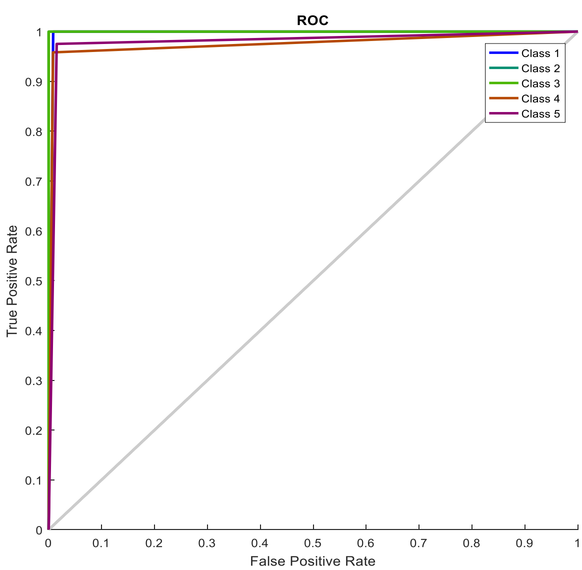


Fig. 5. ROC diagram of the proposed method on reference data [1].

TABLE II. EXECUTION SPEED IN SECONDS

Total data testing time by ELM	Total data training time by ELM	HOG attribute for an image	Local LBP attribute for an image
0.30	13.63	003/0	005/0

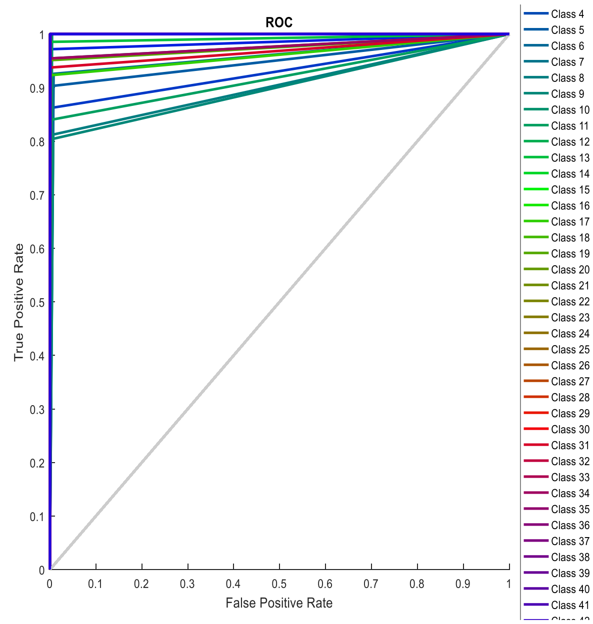


Fig. 6. ROC diagram of the proposed method.

V. CONCLUSION

In this paper, a novel method for classifying traffic signs is presented. The primary work with 1000 images in four classes using color and PHOG features using the SVM algorithm has reached an average accuracy of 95.03%, while the proposed method uses detection. Torsional neural network and feature extraction using tissue feature with LBP technique and HOG feature with ELM classification in 1000 images reached 98.95%, considering that the baseline work did not use all the data related to GTSRB. In this method, extraction of texture properties along with shape properties by the LBP technique can improve the sensitivity of HOG to light reflections and tangled and complex backgrounds. Therefore, feature of the texture is that it is less sensitive to changes in light intensity than the color feature. To prove the superiority of the proposed method, the proposed method is implemented on all data of the GTSRB site, and the results obtained from 12569 images in 43 classes have achieved 95.3% accuracy. Lastly, this study intends to propose a novel feature representation algorithm using color and texture features for finding different features to represent the traffic sign images, presenting a method for classifying traffic signs using color and PHOG features the SVM algorithm, and developing a sign traffic detection method to deal with adverse weather conditions, light reflection, and complex backgrounds. The advantages of this method include the appropriateness of the selected features and the appropriate classification algorithm. However, the weakness is sensitivity in tough light condition and complicated background. For future study, the proposed method can be extended to deal with the tough variation in complex background and light conditions.

VI. ACKNOWLEDGEMENT

This work was supported by Yinhe Metaverse Foundation.

REFERENCES

[1] H. Li, F. Sun, L. Liu, L. Wang, A novel traffic sign detection method via color segmentation and robust shape matching, Neurocomputing, 169 (2015) 77-88.

[2] A. Gudigar, S. Chokkadi, A review on automatic detection and recognition of traffic sign, Multimedia Tools and Applications, 75 (2016) 333-364.

[3] M. Bucurica, R. Dogaru, I. Dogaru, A comparison of extreme learning machine and support vector machine classifiers, 2015 IEEE international conference on intelligent computer communication and processing (ICCP), IEEE, 2015, pp. 471-474.

[4] B. Scholkopf, A.J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press2018.

[5] M. Mathias, R. Timofte, R. Benenson, L. Van Gool, Traffic sign recognition—How far are we from the solution?, The 2013 international joint conference on Neural networks (IJCNN), IEEE, 2013, pp. 1-8.

[6] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, The German traffic sign recognition benchmark: a multi-class classification competition, The 2011 international joint conference on neural networks, IEEE, 2011, pp. 1453-1460.

[7] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Neural networks, 32 (2012) 323-332.

[8] R. Timofte, K. Zimmermann, L. Van Gool, Multi-view traffic sign detection, recognition, and 3D localisation, Machine vision and applications, 25 (2014) 633-647.

[9] E.C. Neto, S.L. Gomes, P.P. Rebouças Filho, V.H.C. de Albuquerque, Brazilian vehicle identification using a new embedded plate recognition system, Measurement, 70 (2015) 36-46.

[10] N. More, B. Tidke, License plate identification using artificial neural network and wavelet transformed feature selection, 2015 International Conference on Pervasive Computing (ICPC), IEEE, 2015, pp. 1-5.

[11] P. Paclık, J. Novovičová, P. Pudil, P. Somol, Road sign classification using Laplace kernel classifier, Pattern Recognition Letters, 21 (2000) 1165-1173.

[12] W.-J. Kuo, C.-C. Lin,Two-stage road sign detection and recognition, 2007 IEEE international conference on multimedia and expo, IEEE, 2007, pp. 1427-1430.

[13] S. Xu, Robust traffic sign shape recognition using geometric matching, IET Intelligent Transport Systems, 3 (2009) 10-18.

[14] S.M. Bascón, J.A. Rodríguez, S.L. Arroyo, A.F. Caballero, F. López-Ferreras, An optimization on pictogram identification for the road-sign recognition task using SVMs, Computer Vision and Image Understanding, 114 (2010) 373-383.

[15] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition—How far are we from the solution?," in Neural Networks (IJCNN), The 2013 International Joint Conference on, 2013, pp. 1-8.

[16] Abukhait, I. Abdel-Qader, J.-s. Oh, O. Abudayyeh, Road sign detection and shape recognition invariant to sign defects, 2012 IEEE International Conference on Electro/Information Technology, IEEE, 2012, pp. 1-6.

[17] A. Gudigar, B. Jagadale, M. Pk, Kernel based automatic traffic sign detection and recognition using svm, International Conference on Eco-friendly Computing and Communication Systems, Springer, 2012, pp. 153-161.

[18] Y. Zhu, X. Wang, C. Yao, X. Bai, Traffic sign classification using two-layer image representation, 2013 IEEE International Conference on Image Processing, IEEE, 2013, pp. 3755-3759.

[19] C. Souani, H. Faiedh, K. Besbes, Efficient algorithm for automatic road sign recognition and its hardware implementation, Journal of real-time image processing, 9 (2014) 79-93.

[20] F. Zaklouta, B. Stanciulescu, Real-time traffic sign recognition in three stages, Robotics and autonomous systems, 62 (2014) 16-24.

[21] M. Mathias, R. Timofte, R. Benenson, L. Van Gool, Traffic sign recognition—How far are we from the solution?, The 2013 international joint conference on Neural networks (IJCNN), IEEE, 2013, pp. 1-8.

[22] R. Timofte, K. Zimmermann, L. Van Gool, Multi-view traffic sign detection, recognition, and 3D localisation, Machine vision and applications, 25 (2014) 633-647.

[23] A. Rezchikov, V. Kushnikov, V. Ivaschenko, A. Bogomolov, L. Filimonyuk, K. Kachur, Control of the air transportation system with flight safety as the criterion, Computer Science On-line Conference, Springer, 2016, pp. 423-432.

[24] H. Balakrishnan, Control and optimization algorithms for air transportation systems, Annual Reviews in Control, 41 (2016) 39-46.

[25] G. Bebis, M. Georgiopoulos, Feed-forward neural networks, IEEE Potentials, 13 (1994) 27-31.

[26] R.R. Trippi, E. Turban, Neural networks in finance and investing: Using artificial intelligence to improve real world performance, McGraw-Hill, Inc.1992.

[27] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence,2002, pp. 971-987.

[28] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893.

[29] Zhang, Jianming, et al. A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection. IEEE access 8 (2020): 29742-29754.

[30] Avramović, A., Sluga, D., Tabernik, D., Skočaj, D., Stojnić, V., & Ilc, N. (2020). Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization. IEEE Access, 8, 189855-189868.

[31] Cao, J., Song, C., Peng, S., Xiao, F., & Song, S. (2019). Improved traffic sign detection and recognition algorithm for intelligent vehicles. Sensors, 19(18), 4021.

Global Pattern Feedforward Neural Network Structure with Bacterial Foraging Optimization towards Medicinal Plant Leaf Identification and Classification

Sapna R[1], S N Sheshappa[2], P Vijayakarthik[3], S Pravinth Raja[4]
Research Scholar, Sir MVIT, Affiliated to VTU, Belgaum, India[1]
Assistant Professor, Department of Computer Science and Engineering, Presidency University, Bengaluru, India[1]
Associate Professor, Sir MVIT, Department of Information Science and Engineering, Bangalore, India[2]
Professor and HOD, Sir MVIT, Department of Information Science and Engineering, Bangalore, India[3]
Associate Professor, Department of Computer Science and Engineering, Presidency University, Bangalore, India[4]

***Abstract*—Medicinal Plant species help to cure various diseases across the world. The automated identification of medicinal plant species to treat disease based on their structure is much required in pharmaceutical laboratories. Plant Species with a complex background in the field will make the detection and classification more difficult. In this paper, optimization of bacterial foraging technique has been employed towards medicinal plant prediction and classification architecture based on feed-forward neural network. It is capable of identifying both complex structures of medicinal plants. Feed-forward Neural Networks are considered to have good recognition accuracy compared to other machine learning approaches. Further bacterial foraging has been implemented to minimize the feature search space to the classifier and provides optimal features for the plant classification. The experimental outcomes of the proposed approach has been analysed by employing the medley dataset and evaluating the performance of the proposed approach with respect to dice similarity coefficient, Specificity and sensitivity towards medicinal plant classification. The findings are very positive, and further research will focus on using a large dataset and increased computing resources to examine how well deep-learning neural networks function in identifying medicinal plants for use in health care.**

***Keywords*—Medicinal plant; feed-forward neural networks; linear discriminant analysis; bacterial forging**

I. Introduction

Most individuals around the world utilise traditional medicines made from medicinal herbs. India is known as the world's botanical garden as well as the country that produces the most medicinal herbs. Due to their natural origins and lack of side effects, herbal medicines have experienced an exponential surge in popularity during the last several years across both developed and developing nations. Plant species have medicinal value and it is considered as the essential resource for curing various diseases around the world [1]. These medications are more widely used since they aim to treat illnesses devoid of causing any negative effects. Indian traditional medicine is renowned for using Naturopathic remedies, Ayurvedic medicine, Unnani, breathing exercises, meditation techniques, yoga, homoeopathy and many more. There are about 8,000 herbal remedies used in these traditional medicines. As per surveyed findings, approximately 75% of migrants utilize herbal plants for medicinal reasons [2, 3].

Manually identifying medicinal plants is still the most used way. Humans gather information about the entire plant or certain parts like leaves, flowers, fruits, or bark using their eye, nose, fingers, or other human organs [4, 5]. Depending on either individual perspectives or referrals, they will decide which species of medicinal plants to use. It has been demonstrated via practise, nevertheless, that this type of identification strategy is time-consuming, inefficient, and heavily dependent on the expertise and subjective experience of the persons using it.

Mainly to determine the variety of medicinal plants, automated models using image processing approaches [6] has been employed to assist botanists in the early identification of plant by employing machine learning architectures [7]. However, machine learning techniques reveal the structural and texture appearance of the plant along with its discriminative visual symptoms. Further detection and classification of the plant species suffer from the noise and illumination effects of the images [8]. In addition, it is mandatory to discriminate the different growth stages of the plant along the virus and disease involvement which was found to be highly complex in identification.

In order to manage these challenges, various image processing steps such as Image segmentation of the seed region [9], feature extraction on the segmented region [10] and optimal feature selection [11] have to be aggregated to achieve high recognition accuracy in prediction and classification of the medicinal plants from large sizes of input images. In this paper, bacterial foraging optimization has been employed to feed forward neural network towards plant classification on basis of pathological properties.

Among numerous machine learning architecture, feed-forward neural networks produce high recognition accuracy in classifying the image dataset. Initially, region-growing segmentation [12] techniques are applied to group the seed points having similar pixel values as the segmented region. The segmented region is analysed using linear discriminant analysis which is considered as a feature extraction technique. Shape descriptors and Texture features are extracted. That extracted feature is reduced using the bacterial foraging technique to generate the optimal features. Finally, feed-forward neural network classifier is employed to optimise features to classify into linear plant classes.

The balance part of the article is sectioned into the following form; Section II provides the review of literature containing similar mechanisms of machine learning for plant identification. Section III defines the proposed architecture containing the Feed Forward Neural Network and bacterial foraging based optimization. Section IV highlights the experimental solutions of the proposed approach along with performance evaluation and at the last, summary of the article with conclusion and future work is provided.

II. Background Study

A. Literature Work

This part of the work describes the problem statement of the research, a review of literature related to the current problem and various deep learning models applied to plant identification using various image datasets has been detailed as follows. The medicinal plant identification model which performs similarly with respect to the proposed architecture has been discussed below.

The Artificial Neural Network Model [ANN] segments the leaf into subfields on different growth stages of the plant. The affinity propagation model computes the variation of the leaf on pixel parameters with reference to plant pathologies and produces the affinity set containing similar leaf properties. Sparse Principle component analysis considered as a feature extraction technique is employed to affinity set to gather the sparse feature and its associations are represented as covariance and correlation matrix [13]. Further Artificial Neural Network is projected to produce the class labels for the sparse features in the feature set. ANN [14, 15] uses a normalization technique to smoothen the edges of the class labels in order to separate plant types.

A Random Forest classifier is implemented to classify the plant on basis of the specified medicinal properties. A classified plant can be employed for disease treatment purposes. Initially, feature extraction methods such as SIFT technique have been employed towards shape extraction, color extraction and texture feature extraction from the plant images. Further genetic algorithm has been applied as a feature selection technique to select effective features towards classification purposes using a Random forest classifier [16].

Convolution Neural Networks (CNN) can be used for the classification of medicinal plant species and they contain built in feature extraction functionality. The shape and texture characteristics are contained in the feature. It classifies the features with respect to the activation function and multiple layers to obtain the effective classes [17]. CNN is also used in various identifications and analysis of images like spatial data, eye tracking data, networks attack and many more [18, 19, 20, 21].

B. Problem Statement

As a critical problem for the preservation, authenticity, and production of herbal medicines, the automatic classification of medicinal plants needs further study [22]. Owing to the difficulties due to illumination effects, shadow effects, nonlinear seed points, overfitting issues and accumulation of reconstruction error in the processing of medicinal plant datasets, the identification and classification of plant leaves will face significant challenges in employing machine learning algorithms. These necessitate a better approach to be developed for identifying and classifying medicinal plants.

III. Proposed Model

This section defines the design procedure towards optimization of Feed-Forward Neural Networks using the bacterial foraging technique to classify the medicinal plant in the Mendeley dataset. In addition, the design procedure of the region growing segmentation approach is to segment the medical plant and linear discriminant Analysis to feature extract is also provided in detail. It is as follows:

A. Image Pre-processing

Image preprocessing is used to increase the image pixel properties on the normalization approach and contrast enhancement approach. A further large number of learning parameters of the modelled have been managed for dataset processing. Image Normalization is employed using histogram stretching. Further to eliminate the image noise and image blur, Contrast enhancement [23] is used in addition to the image thresholding approach [24] to improve the image details which contains an important factor for image classification.

B. Region Growing Segmentation

In this part, seed points illustrating the discriminate image parts of the medicinal plant have been analysed and it is grouped into vectors representing the similar seed points to the entire image. Similar points of the medicinal plant are computed with respect to the homogeneity pixel. Vector represents the segmentation results as a growth process on verification of the homogeneity rule among the pixels. The growth process for the selected regions at every stage S is defined as in (1).

$$G_p i(s) \text{ where } i=1, 2 \tag{1}$$

Homogeneity of the pixel is computed for the seed pixel to verify the presence of unfermented pixels in the neighborhood of the remaining pixel in the particular region. The computation rule is as follows in (2).

$$\text{If } (H\,(G_p i(s)) \text{ for each pixel} = \text{True}) \tag{2}$$

Then compute the Mean M and Standard deviation of each pixel in region R_i is as follows as in (3) and (4).

$$\text{Mean } M = \frac{1}{n}\sum_{i=0}^{n} R(i) \tag{3}$$

$$\text{Standard deviation SD}= \frac{\sqrt{(1\backslash n)R(i)}}{Gpi(s)} \tag{4}$$

Employ strategy to group two regions $G_p(1)$ and $G_p(2)$ on conditions as in (5) and in (6)

If ((Mean of region 1- Mean of Region 2) < Standard Deviation of Region 1) (5)

Verification of the homogeneity of seed points

Else if (pixel intensity of the region 1 is close to mean value of the threshold Ti) (6)

Where Threshold is as in (7)

$$Ti = \{1- \frac{SD_i}{M_i}\} \quad (7)$$

Threshold T_i depends on the variation of the region R_n and the pixel intensity P_i.

Resultant segmentation outcomes are represented as image vectors containing the seed points which are further processed for feature extraction and classification process.

C. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) has been employed to exploit a feature from the image projection using the fisher criterion function [25]. The feature extraction projection in a specific direction extracts properties of homogeneous seeded points in the segmented region vector. Properties of the seed points are represented as a feature. Initially, the homogenous seed point is transformed into scatter matrix to reduce the feature space. Scatter Matrix SM of the vector undergoes the following computation.

Mean of the Vector is as in (8)

$$Mi = =\frac{1}{n} \sum_{x\varepsilon C}^{n} x, n\mathrm{I} \quad (8)$$

Total Mean of the Vector is as in (9)

$$TM =\frac{1}{n} \sum_{x\varepsilon C}^{1} x, N \quad (9)$$

$$\text{Scatter Matrix Sw} =\sum_{i=1}^{c} si\frac{ni}{n} \quad (10)$$

$$\text{Scatter Matrix Sb} = \frac{1}{2}\sum_{i=1}^{c} P\mathrm{iPj(fvi –fvj) (fvi –fvj)T} \quad (11)$$

where (vfi –vfj) is vector feature variation.

The covariance of Mean is as in (12)

$$Si= \frac{1}{n} \sum_{x\varepsilon C}^{1}(x-mi)\ (x\text{-mi})T \quad (12)$$

Aggregate Covariance of Scatter matrixes is to produce an optimal feature set is as in (13).

$$\mathbf{S_T} = \frac{1}{n}\sum ((fv_i –fv_j)\ (fv_i –fv_j)^T \quad (13)$$

Reducing the Feature space is a vital part of the LDA. It can be generated by improving the scatter ratio $|S_b| / |S_w|$. Further, scatter ratio weight will improve on eigenvectors of S_bS_w producing the projection matrix with column matrix W as in (14).

$$W= [W_1, W_2 \ldots\ldots\ldots W_{c-1}] \quad (14)$$

Where W_i are the Eigen Vector of the Scatter matrix S_b and S_w and its value corresponds to reducing Eigen values λ_i

The linear discriminant analysis produces the feature of processing the scatter matrix of the seed points. It is further treated as a reduced feature set on the fusion of two or more vectors. However, this reduced feature set has been employed for feature selection to obtain the optimal feature for classification.

D. Feed-Forward Neural Networks

Feed Forward Neural Network (FFNN) is to determine the plant classes of the image. FFNN is employed to process the spatial and temporal features obtained from the segmented region of the image. An FFNN model is composed of many processing states. FFNN produces the plant classes by representing the feature vector in Direct Acyclic Graph (DAG) structure. In FFNN, seed point features of the segmented region are contained in each layer of the graph network. Table I shows the FFNN parameter components. Extracted spatial and temporal features undergo various computations in a graph structure.

TABLE I. HYPER PARAMETER OF FEED-FORWARD NEURAL NETWORKS

Tuning of Hyper Parameter	Values
Class Batch Size	200
Feature Learning Rate	0.09
Attribute Size	85
Maximum No of pixels to each layer	1000
Length of the pixel in sublayers	250
Loss function	Cross entropy

- Abstraction layer

In this layer, the Abstraction layer is used to collect high-level features in the plant. Vector structure contains the feature. Vector is processed in the activation function. It generates a new state from the old state using epoch functions and a few parameter functions. Matrix Feature weight and Vector feature bias are computed by aggregating the input vectors with output vectors extracted hidden layer. The feature weight of the matrix is computed as in (15).

$$A_t = Sigmoid\ (W(c(t))+h(t-1)) \quad (15)$$

Where A_t represents the feature weight of the Input vector c(t) and hidden vector h(t-1) of the seed points of the plant region.

- Hidden Layer

In this layer, hidden feature information is extracted from the input layer vector seeded point and it is represented as the hidden vector H. The FFNN model uses the hidden layer to compute the feature to be represented in the output layer for a period of data availability. Stored Seeded region will be forgotten in the subsequent process. The FFNN model includes the sigmoid function as an activation function to compute the feature weight to eliminate the less-weighted features. Fig. 1 illustrates the proposed research architecture.

The FFNN hidden layer is concealed with the output information of the output layer illustrated in (16).

$$O(f(t)) = \sigma(G_p\,[A(t)\,C^{n-k} + h(t-1)] + W_f) + \sum_{k=1}^{n} \binom{n}{k} m^k \quad (16)$$

The Hidden layer information of the FFNN is identified as an important process to analyse the structural changes and temporal changes of the plant on various growths stage and disease stages. This information is helpful to classify the medicinal plant. The hidden layer of FFNN identifies which input accommodating the plant features and its concealing is provided by (17).

$$H_f = \sigma\, B_V\, (WmG_i(s);\, Wmd_i(s)] + \sigma\, B_V\, (WmG_i(t);\, Wmd_i(t)] \quad (17)$$

Where H_f represents the model outcome containing features of growth stage and disease stages along the matrix feature weight W_m and bias of feature vector B_V It is further interpreted with pathological information. Ϭ is considered as a sigmoid activation function to extract the disease and growth stages features.

- Activation function

The activation function set the decision point of the learning rate of the proposed model to manage and change the feature weights of the growth and disease feature of the plant of various stages from the hidden layer on employing the sigmoid function [26]. FFNN model activation function is illustrated in form of a tree structure which has a high ability in parsing the feature using the sigmoid function. The activation function is optimized using a hyperparameter to improve the output layer representing the plant classes. The activation function is as follows in (18).

$$A_f = \text{Sigmoid}\ (F_s)^* \sum_{n=1}^{\infty} \left(a_n \cos\frac{n\pi w}{B} + b_n \sin\frac{n\pi w}{B}\right) \quad (18)$$

Activation functions of the hidden layer process the feature set to identify the plant classes on basis of distance measures on feature weight.

- Output Layer

The output layer generates the plant classes. The output layer interprets the feature set of the hidden layer with reference to the pathological information to discriminate the plant classes with high recognition accuracy. Plant classes of the output layer are discriminated using the output function. Function Z(t) uses the feature vector on various stages to classify effectively with the activation function. Output Function is given by (19).

$$\text{Output Function } Z(T) = \text{Sigmoid}\ (\,G(f)^*\,D(f)) \quad (19)$$

The output function of the FFNN contains the sigmoid function to differentiate the growth and disease stage features of the plant. These outcomes are stored as classes with reference to the output function.

- Loss Layer

The loss layer is employed to assure the classification accuracy of plant classes. It is assured on the processing the image data with various folds using the cross entropy function. On identification of irregularity or reconstruction error on the processing layer of the model, hyperparameter tuning is carried out on each layer of the network to reduce the error in the hidden layer, input layer and output layer. The loss function for hidden and output layers are given by (20) and (21) respectively.

$$\text{Loss function Lf for hidden layer} = \text{Softmax}\ (At^*Er^*Ht) \quad (20)$$

$$\text{Loss function Lf for output layer} = \text{Softmax}\ (At^*Er^*Zt) \quad (21)$$

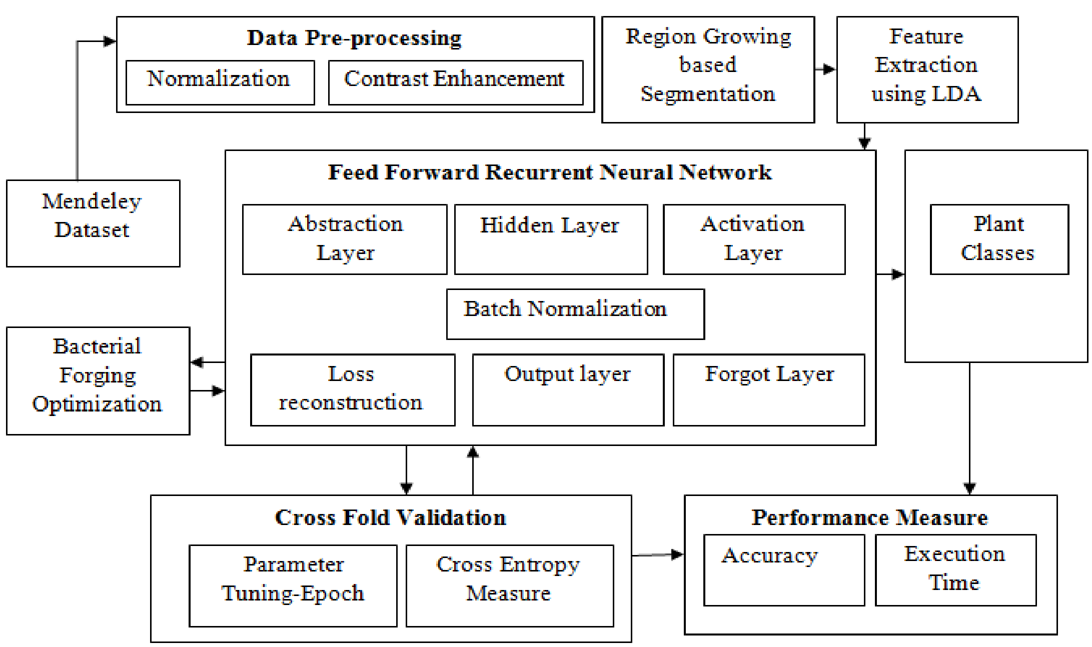


Fig. 1. Proposed architecture.

An accurate neural network model for classification is obtained with the inclusion of the loss function in the layer.

Algorithm 1: Plant Identification

Input: Mendeley Dataset

Output: Plant class

Process

Image Pre-Process ()

Image Thresholding ()

Contrast Stretching ()

Region Segmentation ()

Identify the seed Point

Compute Mean M and Standard deviation SD for Seed point()

Homogeneity Rule ()

$F(s)^n = 1 + \frac{Rx}{2n} + \frac{R(n-1)x^2}{2(n-1)} + \cdots \frac{R(n-1)x^n}{2(n-y)}$

If (Pixel intensity of Region 1 == Pixel Intensity of Neighbour Region of Seeded point 2)

Merge the two pixels into the seed point of the Region as Vector V
Linear Discriminant Analysis ()

Compute Scatter matrix S_m[] of seed region= $\frac{1}{2}\sum_{i=1}^{c} P_i P_j (fv_i - fv_j)(fv_i - fv_j)^T$
FFNN Learning ()
Input layer ( )

Determine the Spatial Feature of Seed Points ()

$F(S) = S_m \sum_{n=1}^{\infty} D(x, y)$

Compute Temporal Features of seed points ()

$F(T) = S_m \sum_{n=1}^{\infty} C(x, y)$

Hidden Layer ()

Identify the hidden feature of the various growth stage

$O(f(t)) = \sigma(G_p [A(t) C^{n-k} + h(t-1)] + W_f) + \sum_{k=1}^{n} \binom{n}{k} m^k$

Activation Layer ()

Use the Sigmoid Function

Output Layer ()

Cross Entropy Layer ()

Loss FunctionL_f Softmax ()

$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$

Identify and display the Plant Class

Algorithm Description:

The classification algorithm initiates with image pre-processing approaches with histogram normalization and contrast stretching for noise removal and image normalization in the Medley dataset. Region-growing segmentation is employed for the preprocessed image. Linear Discriminant Analysis (LDA) has been employed as a Feature extraction approach. Obtained features are processed to classify the feature into classes using Feed Forward Neural Network on the computation of the features in the activation layer of the network using the sigmoid function.

E. Bacterial Foraging Optimization

In this section, bacterial foraging optimization is employed to the hidden layer to obtain the optimal feature space from the hidden layer to enhance plant classification accuracy in processing the features of various plants. The bacterial foraging uses the fitness criteria on the feature space from the hidden layer of the FFNN model [27]. Bacterial Optimization computes the optimal features to classify the plant into categories, especially it is found effective to classify various patterns. The optimal feature set has to be determined using the following fitness function (22).

$$\text{Fitness Function} = \sum_{m=k}^{n-1} \text{Correlation}(\boldsymbol{\delta}_i^j(f_k) - \boldsymbol{\delta}_i^j(f_{m+1})) + O(f_n) \quad (22)$$

Feature set given by fitness function is F(s)

IV. Experimental Results

Experimental analysis has been carried out in the Mendeley plant dataset [28]. The proposed architecture is simulated in Matlab (version 2017). In processing of the model, training data and validating data have been partitioned. In that, 60% of the image has been utilized for model training and 20 % of the image has been fixed for validation and rest 20% of the image has been fixed for model testing. In 10-fold cross-validation of the learning model, the performance of the classification and segmentation of the leaf region has been enhanced.

In this work, the proposed neural network model achieves high recognition accuracy on the identification of plant classes. The model provides excellent outcomes in computing the leaf region with different image sizes and image characteristics evaluated against conventional methods. Model performance has been evaluated with Dice coefficient, sensitivity and specificity against the conventional approaches for medicinal plant detection with volumetric changes of the plant at various stages of the plant growth.

- Dice similarity Coefficient

It is accessed on basis of the variation of classified results against the pathological classes. This difference computation is carried out using a confusion matrix. The confusion matrix yields True Positive (TP), False Positive (FP) and False Negative (FN) values on the class results. It is denoted as in (23).

Dice Similarity Coefficient

$$= \frac{2\text{True Positive}}{2\text{True Positiv}+\mathit{False\ positve}+\mathit{False\ Negative}} \quad (23)$$

- Sensitivity

It is considered as the ability of a test to correctly classify the plant in terms of various features. It is denoted as in (24).

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive}+\mathit{False\ Negative}} \quad (24)$$

- Specificity

It is considered as the ability of a test to correctly classify the plant without various features. It is denoted as in (25).

$$\text{Specificity} = \frac{\text{True Positive}}{\text{True negatives}+\mathit{False\ Negatives}} \quad (25)$$

The Mendeley medicinal plant leaves dataset containing image samples has been analysed to determine the medicinal plant class using machine learning approaches. The learning model has been assessed using the Dice coefficient measure, Sensitivity and Specificity measures. Its performance value is represented in Table II.

The proposed model exhibits good performance in segmenting the plant regions. Optimization of Feed Forward neural Network has been carried out with bacterial foraging illustrating the excellent outcomes on relating with conventional approaches such as Artificial Neural Networks and Convolution Neural Networks.

The Dice Coefficient generates the best outcomes on accessing with a classification accuracy of plant classes, generated using the machine learning model has been illustrated in Fig 2.

TABLE II. PERFORMANCE EVALUATION OF MEDICINAL PLANT CLASSIFICATION TECHNIQUES

Samples	Technique	Dice Coefficient	Sensitivity	Specificity
1St Fold Validation Set	Bacterial foraging optimized Feed-Forward Neural Network– Proposed model	0.9788	0.9485	0.9792
	Artificial Neural Network - Existing Model	0.9578	0.9122	0.9689
2rd Fold Validation set	Bacterial foraging optimized Feed-Forward Neural Network l – Proposed model	0.9777	0.9194	0.9798
	Artificial Neural Network - Existing Model	0.9665	0.9014	0.9271
3rd Fold Validation set	Bacterial foraging optimized Feed-forward Neural Network – Proposed model	0.9771	0.9115	0.9785
	Artificial Neural Network - Existing Model	0.9656	0.9015	0.9365

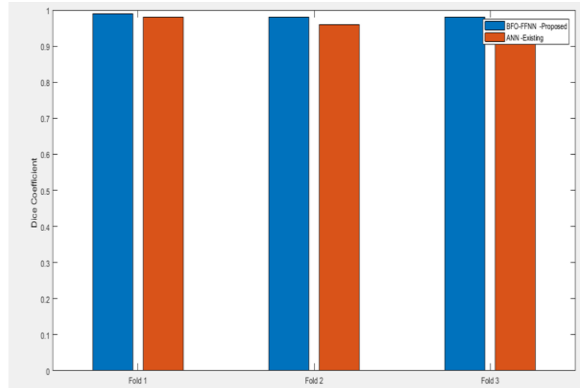


Fig. 2. Performance comparison of the learning model with respect to the dice coefficient.

Regarding the measure of sensitivity, it measures the excellent results in the optimization of the learning model through bacterial foraging to produce optimal features. Fig. 3 shows the performance outcome of the sensitivity measure on the plant classification.

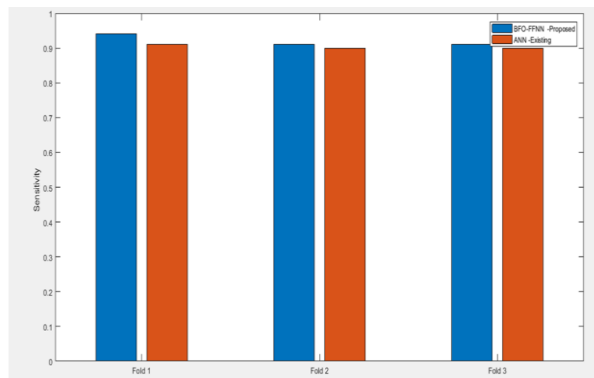


Fig. 3. Performance comparison of the learning model with respect to sensitivity.

The performance of the trained network is adapted to the target domain on producing better outcomes. Fig. 4 illustrates the performance outcome of the specificity. The proposed model is capable of classifying five medicinal plants effectively such as Basella Alba (Basale), Carissa Carandas (Karanda), Citrus Limon (lemon), Ficus Auriculata (Roxburgh Fig) and Rosa-Sinensis (Hibiscus).

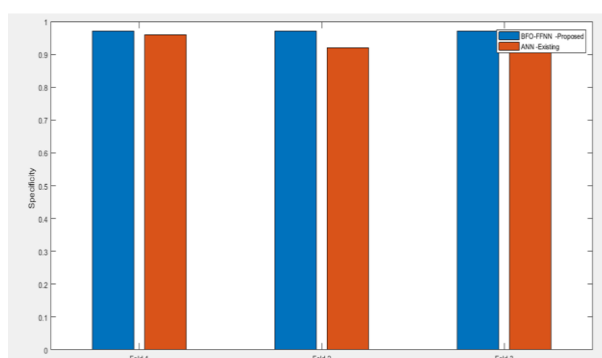


Fig. 4. Performance comparison of the proposed model in terms of specificity.

V. Conclusion

In this work, an optimized framework of Feed Forward Neural Network using bacterial foraging has been designed and implemented to classify the medicinal plant from the medley dataset. The model initiates the segmentation operation to the pre-processed image dataset in order to obtain the segmented region containing seed points in the image. The segment images are further processed using the linear discriminant to extract the spatial and temporal features. The extracted feature set is applied to the feed-forward neural network classifier. The classifier composed of the input layer, hidden layer, output layer and loss layer has fine-tuned with objective and activation function to discriminate the results with high recognition accuracy. The proposed model produces excellent outcomes even presence of image artifacts and volumetric changes in images. Further it is capable of handling the variability in anatomical properties and contrast variations of the dataset in analyzing the modalities and pathologically changed tissues. Further, different posing and structural changes have to be computed to minimize the interpretation issues and enhance the classification accuracy using Gaussian and quantum models on the features of the images to discriminating the medicinal plant classes of leaf images.

The results are very encouraging, and future studies will concentrate on using a huge dataset and more computational power using Quantum Computing to assess how effectively deep learning neural networks work in identifying medicinal plants for use in healthcare [29]. Whilst deep learning is a paradigm for analysis, linked data can be selected as the optimal practice for data representation [30]. The confluence of linked data with the proposed work can further take this study towards more accuracy.

References

[1] Gao, W. and Lin, W., 2012. Frontal Parietal Control Network Regulates the Anti-Correlated Default and Dorsal Attention Networks. Human Brain Mapping, 33(1), 192–202.

[2] Banita Pukhrambam and Arun Sahayadhas, "Advanced Medicinal Plant Classification and Bioactivity Identification Based on Dense Net Architecture" International Journal of Advanced Computer Science and Applications (IJACSA), 13(6), 2022. http://dx.doi.org/10.14569/IJACSA .2022.0130614.

[3] S. Talwar, S. Sood, J. Kumar, "Ayurveda and allopathic therapeutic strategies in coronavirus pandemic treatment", Current Pharmacol Reports, pp. 354–363, 2020.

[4] K. Mohanraj, B. S. Karthikeyan, Vivek-Ananth, "IMPPAT: A curated database of Indian medicinal plants, Phyto chemistry and Therapeutics", 2018.

[5] M. Fitzgerald, M. Heinrich, A. Booker, "Medicinal plant analysis: a historical and regional discussion of emergent complex techniques", Front Pharmacol, Published, 2020.

[6] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2007. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11-16.

[7] Zhang X., Liu Y., Lin H., Liu Y. (2016) Research on SVM Plant Leaf Identification Method Based on CSA. In: Che W. et al. (eds) Social Computing. ICYCSEE 2016. Communications in Computer and Information Science, Vol 624, Springer, Singapore.

[8] Hossain, J. and Amin, M.A., 2010. Leaf Shape Identification Based Plant Biometrics. 13th International Conference on Computer and Information Technology, Dhaka, Bangladesh, 458-463.

[9] Du, J.X., Wang, X.F. and Zhang, G.J., 2007. Leaf shape based plant species recognition. Applied Mathematics and Computation, 185, 883-893.

[10] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. 5th International Conference on Emerging Intelligent Computing Technology and Applications, Ulsan, South Korea, 627-634.

[11] Herdiyeni, Y. and Wahyuni, N.K.S., 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. International Conference on Advanced Computer Science and Information Systems (ICACSIS), West Java, Indonesia, 301-306.

[12] Prasvita, D.S. and Herdiyeni, Y., 2013. MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. International Journal of Advanced Science, Engineering and Information Technology, 3, 5–8.

[13] Le, T.L., Tran, D.T. and Hoang, V.N., 2014. Fully Automatic leaf-based plant identification, application for Vietnamese medicinal plant search. Fifth Symposium on Information and Communication Technology, Hanoi, Vietnam, 146-154.

[14] Arai, K., Abdullah, I.N. and Okumura, H., 2013. Identification of Ornamental Plant Functioned as Medicinal Plant Based on Redundant Discrete Wavelet Transformation. International Journalof Advanced Research in Artificial Intelligence, 2(3), 60-64.D. K. Berger, ``Grey leaf spot disease of maize and food security research,'' South Afr. J. Botany, vol. 100, no. 109, p. 327, Jan. 2017.

[15] Sheela sobana Rani, K., Pravinth Raja, S., Sinthuja, M., Vidhya Banu, B., Sapna, R. and Dekeba, K., 2022. Classification of EEG Signals Using Neural Network for Predicting Consumer Choices. Computational Intelligence and Neuroscience, 2022.

[16] Nayana G. Gavhale, Dr.A.P.Thakare "Identification of Medicinal Plant Using Machine Learning Approach "International Research Journal of Engineering and Technology, Volume: 07 Issue: 07 | July 2020.

[17] Bhanuprakash Dudi,Dr.V.Rajesh" Medicinal Plant Recognition based on CNN and Machine Learning "International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.4, July – August 2019.

[18] Praveena, K.N. and Mahalakshmi, R., 2022. Classification of Autism Spectrum Disorder and Typically Developed Children for Eye Gaze Image Dataset using Convolutional Neural Network. International Journal of Advanced Computer Science and Applications, 13(3).

[19] Pallavi M, Thivakaran T K and Chandankeri Ganapathi, "Evaluation of Land Use/Land Cover Classification based on Different Bands of Sentinel-2 Satellite Imagery using Neural Networks" International Journal of Advanced Computer Science and Applications(IJACSA), 13(10), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0131070.

[20] Sulabh Katiyar and Samir Kumar Borgohain, "Comparative Evaluation of CNN Architectures for Image Caption Generation" International Journal of Advanced Computer Science and Applications (IJACSA), 11(12), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111291.

[21] Singh, K., Mahajan, A. and Mansotra, V., 2021. 1D-CNN based Model for Classification and Analysis of Network Attacks. International Journal of Advanced Computer Science and Applications, 12(11).

[22] Sapna, R. and Sheshappa, S.N., 2022. An Extensive Study on Machine Learning Paradigms Towards Medicinal Plant Classification on Potential of Medicinal Properties. In International Conference on Image Processing and Capsule Networks (pp. 541-555). Springer, Cham.

[23] C. DeChant, T. Wiesner-Hanks, S. Chen, E. L. Stewart, J. Yosinski, M. A. Gore, R. J. Nelson, and H. Lipson, ``Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning,'' Phytopathology, vol. 107, no. 11, pp. 1426-1432, Nov. 2017.

[24] S. Ren, K. He, R. Girshick, and J. Sun, ``Faster R-CNN: Towards realtime object detection with region proposal networks,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[25] P. Xiao, G. K. Venayagamoorthy, and K. A. Corzine, "Combined training of recurrent neural networks with particle swarm optimization and backpropagation algorithms for impedance identification," in Proceedings of the IEEE Swarm Intelligence Symposium (SIS '07), pp. 9–15, April 2007.

[26] K. R. Aravind, P. Raja, K. V. Mukesh, R. Aniirudh, R. Ashiwin, and C. Szczepanski, ``Disease classication in maize crop using bag of features and multiclass support vector machine," in Proc. 2nd Int. Conf. Inventive Syst. Control (ICISC), Jan. 2018, pp. 1191-1196.

[27] Mingjie Lv, Guoxiong Zhou, Mingfang He, Aibin Chen, Wenzhuo Zhang, Yahui Hu "Maize Leaf Disease Identification Based on Feature Enhancement and DMS-Robust Alexnet " IEEE Access, Vol.8,pp:57952-57966, 2020.

[28] Medicinal Leaf Dataset, https://data.mendeley.com/datasets/nnytj2v3n5/1, accessed on Oct 2022.

[29] Renukaradhya, S., Preethi, Bhagawati, R. and Subramanian, T., A Brief Study on Quantum Walks and Quantum Mechanics. In Artificial Intelligence, Machine Learning and Blockchain in Quantum Satellite, Drone and Network (pp. 15-34). CRC Press.

[30] Sapna, R., Monikarani, H.G. and Mishra, S., 2019, February. Linked data through the lens of machine learning: an enterprise view. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-6). IEEE. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

Predicting Employee Turnover in IT Industries using Correlation and Chi-Square Visualization

Bagus Priambodo[1], Yuwan Jumaryadi[2], Sarwati Rahayu[3], Nur Ani[4], Anita Ratnasari[5], Umniy Salamah[6], Zico Pratama Putra[7], Muhamad Otong[8]
Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia[1, 2, 3, 4, 5, 6]
Faculty of Information Technology, Universitas Nusa Mandiri, Jakarta, Indonesia[7]
Department of Electrical Engineering, Universitas Sultan Ageng Tirtayasa, Serang, Indonesia[8]

***Abstract*—Employee turnover in the IT industry is among the highest compared to other industries. Knowing factors that influence the turnover may help reduce this issue in future. One of these factors is job satisfaction, which are composed of two important factors, status and seniority. In this study, the correlation and chi-square visualization are utilized to determine the factors that affect employee turnover. The experiment was carried out to predict turnover using a private IT consultant dataset comparing three classification algorithms (decision tree, Naïve Bayes, and Random Forest). The result shows that job duration and positioning are factors that influence employee turnover in a software company.**

***Keywords*—Employee turnover; turnover factors; chi-square; classification algorithm**

I. Introduction

According to the information and communications technology (ICT) workers survey, 43.6% of ICT professionals work in IT Industries, with an annual growth rate of 14.7% [1]. Given the sharp increase in demand for skilled IT professionals, supply of qualified computer specialists, especially college graduates, is on the rise [2]. This situation has given the IT industry a competitive advantage in hiring an ever-growing number of professional software engineers, as each company tries to lure the best employees by offering them better opportunities in terms of pay, expertise, and working atmosphere. The IT industry therefore has one of the highest turnover rates in comparison to other industries [1].

Poor financial compensation, management style, and career opportunities are the main factors contributing to dissatisfaction with current employer and their decision to leave [3]. Staff turnover is a major disadvantage for the company on direct and indirect costs. The direct costs are related to the time and resources spent on recruiting, hiring and training employees [4]. The indirect costs include the decrease in production and services caused by the company failing to find replacements. Hiring and training new employees is also a large investment because the higher the turnover rate, the higher the cost [5].

In general, turnover is due to employee dissatisfaction with a number of factors, including working conditions, salary, support quality, co-workers, nature of work, employment security, and career prospects. The factors that affect job satisfaction are divided into two factors, namely organizational factors - which include company policies and work environment - and individual factors [6]. The status and seniority are one of these individual factors that affect job satisfaction, as a perception that a lack of employment status causes many workers seeking another job [6]. In addition, attitudes toward human resource management indirectly affect turnover, which is fully mediated by job satisfaction [7].

In this paper, visualization and feature selection are proposed to identify factors that influence employee turnover. The paper is organized as follows. In Section II, this paper first gives an overview of previous work on this topic. This is followed by a presentation of the method used in Section III, followed by a discussion of the experimental results in Section IV. Finally, the paper is summarized in Section V.

II. Related Work

Some turnover studies suggest using machine learning to design retention policies [8]. A study was carried out by [9] to predict employee turnover in one of Indonesia's renowned telecommunication company, comparing three classification algorithm (Naïve Bayes, Decision Tree and Random forest) based on 12 attributes. Another study by [10] implemented twelve features to predict employee turnover, while other use 19 attributes [8]. Similar study by [11] evaluate six classification algorithm (KNN, SVR, Naïve Bayes, Decision Tree, Random forest) to predict employee turnover using a dataset from Kaggle. In contrast to the previous study, this study will evaluate the individual and seniority factors, which are represented by length of service and job level as factor of employee turnover in IT consultant firm.

A. Naïve Bayes

In supervised machine learning [12][13][14][15], Naïve Bayes is a widely used model of classification due to its simplicity and efficiency. Naïve Bayes computes the posteriori probability for a class by observing the churn problem, i.e., the churn and non-churn observation probabilities. The posteriori probabilities are computed by using the rule of Bayes and the assumption of naive Bayesians with unique employee backgrounds for each class. The goal of the Bayesian decision rule is to set a record of a new hire to the class with the highest probabilities posteriori as [16] suitable model for predicting employee turnover.

B. Decision Tree

A decision tree is divided by decision rules into some classes [17]. It represents a fluctuating graph resembling a tree structure that implies a test in an attribute for each internal node (not leaf node). A branch represents a test result, with a class label for each leaf node (or end node). The root node is the top node in a tree [18]. The C4.5 decision tree classifier is employed to build predictive models due to its ability to segment dynamic decision processes [19].

C. Random Forest

Random Forest (RF) proposed by Brieman [20], uses aggregation and bootstrap ideas to introduce random forests based on a decision tree [21]. The trees are constructed separately in RF, using bootstrap samples of various data sets. The concept of RF is to construct multiple decision trees using only a subset of attributes based on sample data [16]. Predictions are taken by taking a simple majority decision by dividing the individual nodes into a subset of predictors using the best distribution. At that node, the predictors are selected randomly and this randomness makes it robust against overfitting [22][23].

III. Methodology

In this section, a proposed framework is presented, which includes preprocessing, relationship visualization, chi-square calculation, correlation, and prediction (Fig. 1).

A. Sourcing and Pre-processing

This is the initial phase of the framework. The purpose of preprocessing the data is to convert the raw data into an easier and more efficient format to use for future processing steps. In the first phase, a min-max method is used to normalize the data. Normalization can shorten the training time as all the data used for training share the same scale. Collecting the dataset from private IT consultants between January 1, 2015 and December 31, 2018, the data consists of gender, entry and exit dates, job title, and department, as shown in Table I.

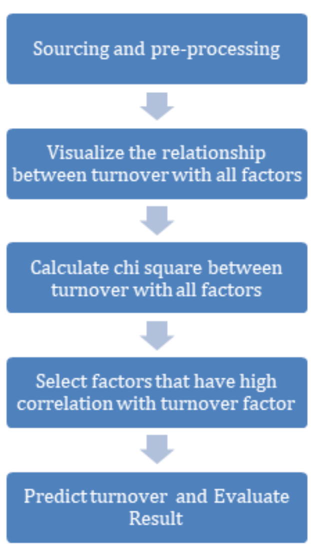


Fig. 1. Framework of predicting employee turnover by identifying the two most influence factors in turnover, i.e., work duration and department.

TABLE I. ATTRIBUTE EMPLOYEE TURNOVER

No	Gender	Date of accession	Date of withdrawal	...	Job title
1	Male	01/12/2016	05/05/2017	...	Head of the Finance Department
2	Female	16/08/2017	29/03/2018	...	Accounting officer
3	Male	01/03/2016	14/01/2017	...	Driver
4	Female	19/09/2016	14/07/2017	...	Recruiter
149	Male	01/07/2015	-	...	Functional consultant
150	Male	01/07/2015	-	...	Software platform manager

B. Visualization of Correlation between Turnover and All Factors

The correlation across pairs of time series indicates the degree to which the variance in one time series may be the same as the variance in the other. This alone does not imply a causal relationship between the two series. However, this is generally grounds for further investigation of a possible relationship. Relationships of correlation are often expressed as values between -1 and 1, figures close to 1 indicating a high correlation, figures close to 0 indicating a low relationship, and figures close to -1 indicating an inverse relationship. An analysis is made of the correlation between turnover and all factors. The visualization is selected to analyze the most correlated factor with the resignation factors.

C. Calculation of the Chi-Square between Turnover and All Factors

For classification problems having both categorical input and output, a chi-square test ("chi") is useful in determining how relevant the input variables are to the output variable. In this study, the correlation between turnover and all factors is analyzed using chi-square, calculating the *p-value* to determine which factor has the lowest *p-value*. The formula for the chi-square is shown on (1).

$$X^2 = \sum \frac{(O-E)^2}{E} \tag{1}$$

Σ means summarize

O = any observed (actual) value

E = any expected value.

D. Predict Turnover

In classification, a mean value of grouping data based on a label or target class is calculated in the study. The three classification algorithms are used to compare the impact of the proposed framework on classification performance. Using Decision Tree, Naive Bayes, and Random Forest as pilot methods, all of them are used to predict turnover based on all factors. Next, a forecast of sales based on selected factors is also made using these three algorithms. Both results are then compared to determine which has the better performance.

IV. Result and Discussion

A. Pre-Processing

Experiments are conducted using several machine learning algorithms with additional tests on the preprocessed data and feature selection to achieve better accuracy. The preprocessed data for job level, department and gender category are converted into numbers as shown in Table II.

B. Visualize the Correlation between Turnover and All Factors

The correlation between gender, department and degree of resignation is visualized in a bar chart (Fig. 2). The correlation between duration and turnover factor (resignation) is visualized by a boxplot. As shown in Fig. 3, gender does not have a significant impact on employee turnover. The trend of gender is similar for terminated and non-terminated employees.

TABLE II. Data Set after Pre-processing

No	Duration (month)	Job level	Department	Gender
1	5	5	13	1
2	7	1	1	0
3	10	1	8	1
4	9	1	17	0
...	...	...	...	...
149	42	1	14	1
150	42	5	41	1

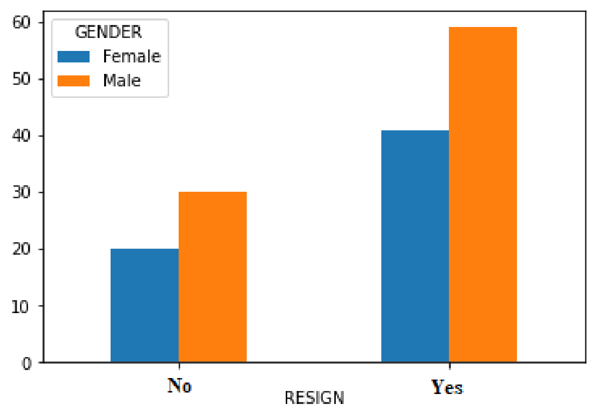


Fig. 2. The relationship between gender and the turnover Factor (resignation).

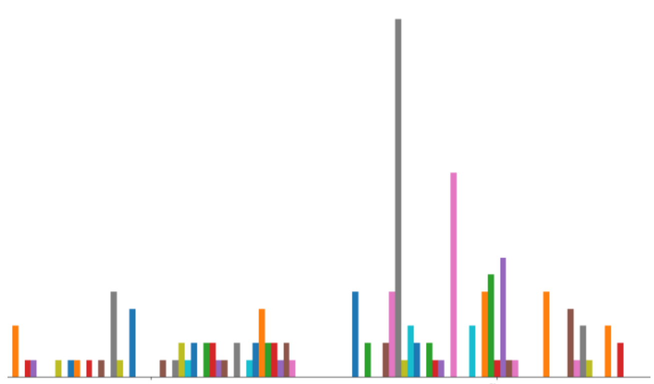


Fig. 3. The relationship between the department and the factor of turnover (resignation).

Fig. 3 shows the relation between the department and the turnover factor and reveals that department factors have a significant impact on employee turnover. The trend of the employee leaves and the employee stays are different from various departments.

Fig. 4 shows the relationship between the workplace level and the turnover factor. It indicate that the workplace level has no significant influence on employee turnover. Further, the trend of the job level is similar for terminated and non-terminated employees.

However, the boxplot relationship between duration and turnover (resignation) factor from Fig. 5 shows that the length of employment (duration) has a significant impact on employee turnover.

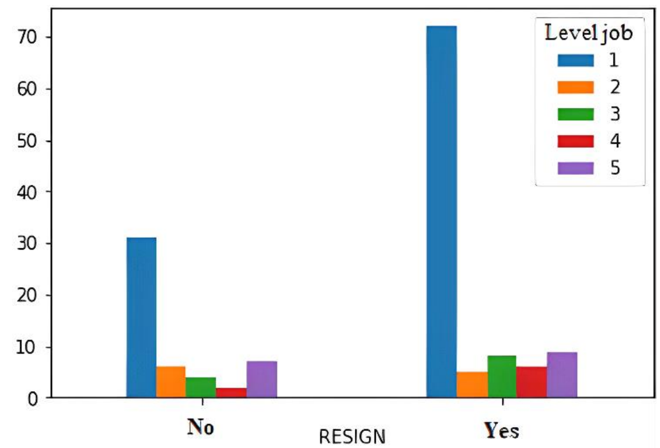


Fig. 4. The relationship between the workplace level and the turnover factor (resignation).

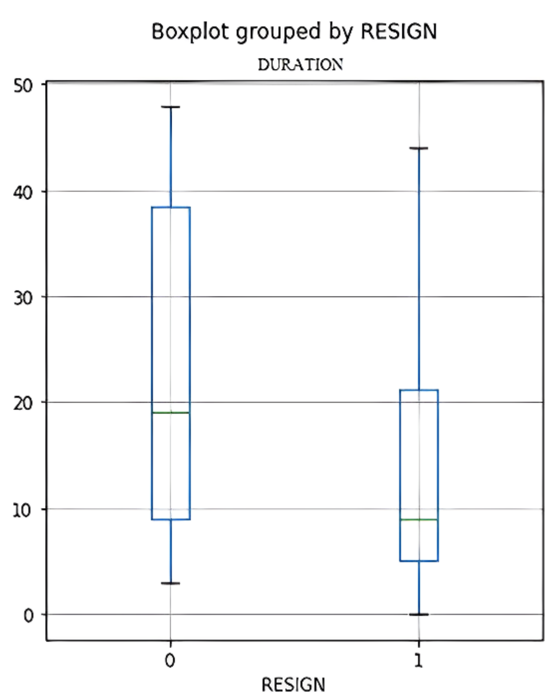


Fig. 5. The relationship between duration and turnover (resignation) factor.

C. Calculate the Chi-Square between Turnover and All Factors

The term statistically significant has become synonymous with a *p-value* <= 0.05. The table of *p-value*s shows that the *p-value* for duration (length of work) and department is less than 0.05, as shown in Fig. 6. Table III and Fig. 6 show that duration of work (length) and department are factors that affect employee turnover.

D. Turnover Forecast and Result Evaluation

The dataset is divided by 80% for training data and 20% for testing data. The result is shown in Table IV. The result of prediction with all factors is shown in the first column (with all factors) and the result of prediction with only selected factors is shown in the second column (only selected factors). Based on the visualization of the correlation between the turnover and all factors and the calculation of the chi-square. Duration and department factor have high correlation with employee turnover. The prediction result shows that the prediction using only selected factors has better performance compared to the prediction using all factors.

TABLE III. P-VALUE BETWEEN TURNOVER AND ALL FACTORS

No	Factors	*p-value*
1	Gender	0.94
2	Duration	**<0.01**
3	Level Job	0.36
4	Department	**<0.01**

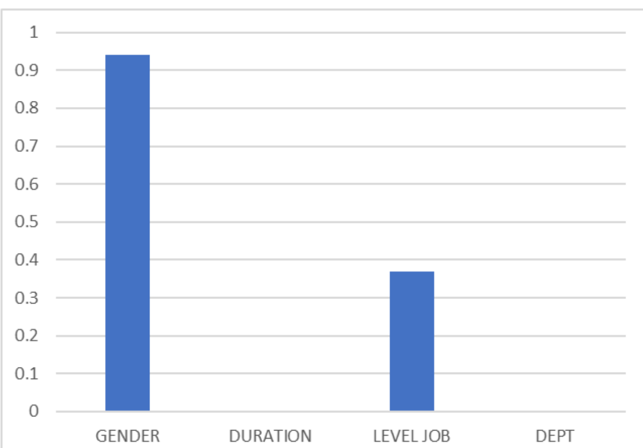


Fig. 6. Bar chart of the p-value between turnover (resignation) and all factors.

TABLE IV. THE PREDICTION RESULT OF THE EMPLOYEE TURNOVER

	With all factors	Selected factors only
Random forest	0.833	0.867
Naïve Bayes	0.77	0.77
Decision tree	0.83	0.90

V. CONCLUSIONS

In this study, correlation and chi-square visualization are applied. Its key idea is based on a two-step process. A correlation between pairs of time series is first used for further investigation of possible relationships. Secondly, Chi-Square is used to determine how relevant the input variables are to the output variables. By forecasting employee turnover - compared to Decision tree, Naïve Bayes, and Random forest - predictions using only selected factors (duration of employment and department) indicated a notable improvement in various performance parameters compared to predictions using all factors. Using a dataset of IT consulting employees in Indonesia, the key factors affecting employee turnover were identified using the chi-square algorithm; specifically, years in the company, job level, department placement, and gender as the most important factors. Therefore, duration of employment and department are unarguably among the most influential factors for retaining employees.

By predicting upfront, the likely turnover of employees, companies are able to prepare plans and initiate measures to reduce that likelihood, hire replacements swiftly, and make other adjustments aimed at retaining individuals in important positions. By using correlation and chi-square visualization approaches, HR managers may anticipate employee turnover better and take action in a timely manner. Similar applications of this algorithm are possible for imbalanced data classification problems such as predicting customer attrition, preventive cancer care, and anomaly detection. These issues share common characteristics, like a focus on minority classes and the condition that "it is preferable to test wrong than to detect wrong." Future work would be to analyze theoretically the characteristics of the data in these areas and to verify and optimize the approach further.

Due to the limitations of the study, the extra process of feature selection and weight calculation implies that the cost of modeling is higher in time compared to other algorithms. Improvements to the operational efficiency and accuracy of the entire prediction can be analyzed in future research. Furthermore, for industries with high turnover rates, the algorithm is based on unbalanced data and therefore may not be suitable. Details on increasing the universality of the algorithm are yet to be studied.

ACKNOWLEDGMENT

This work was supported by Universitas Mercu Buana through KDN Research Grant Scheme [02-5/1355/B-SPK/VI/2022].

REFERENCES

[1] L. Jinadasa and V. Wickramasinghe, "IT Industry Labour Turnover : The Reality," 10th Int. Conf. Sri Lanka Stud., no. December 2005, pp. 0–10, 2005.

[2] Sasmoko, Y. Indrianti, S. A. Widhoyoko, Y. Udjaja, and U. Rosyidi, "Performance Change with or Without ITEI Apps," in The 6th International Conference on Cyber and IT Service Management (CITSM 2018), 2018, pp. 1–4, doi: 10.1109/CITSM.2018.8674055.

[3] R. Korsakienė, A. Stankevičienė, A. Šimelytė, and M. Talačkienė, “Factors driving turnover and retention of information technology professionals,” J. Bus. Econ. Manag., vol. 16, no. 1, pp. 1–17, 2015, doi: 10.3846/16111699.2015.984492.

[4] I. Jeffrey and A. B. Prasetya, “the Employee Performance Assessment and Employee Training, on Employee Intension,” J. Apl. Manaj., vol. 17, no. 1, pp. 56–65, 2019, doi: 10.21776/ub.jam.2019.017.01.07.

[5] R. Ameliya and H. Febriansyah, “The significant factors of employee turnover case study : ABC Hotel,” J. Bus. Manag., vol. 6, no. 2, pp. 239–249, 2017.

[6] Hojops Odoch and S. Nangoli, “The effect of organizational commitment on job satisfaction in Uganda Colleges of Commerce,” Issues Bus. Manag. Econ., vol. 2, no. 10, pp. 165–171, 2014.

[7] C. Maier, S. Laumer, A. Eckhardt, and T. Weitzel, “Analyzing the impact of HRIS implementations on HR personnel’s job satisfaction and turnover intention,” J. Strateg. Inf. Syst., vol. 22, no. 3, pp. 193–207, 2013, doi: 10.1016/j.jsis.2012.09.001.

[8] E. Ribes, K. Touahri, and B. Perthame, “Employee turnover prediction and retention policies design: a case study,” 2017, [Online]. Available: http://arxiv.org/abs/1707.01377.

[9] A. Alamsyah and N. Salma, “A Comparative Study of Employee Churn Prediction Model,” Proc. - 2018 4th Int. Conf. Sci. Technol. ICST 2018, vol. 1, no. 2, pp. 1–4, 2018, doi: 10.1109/ICSTC.2018.8528586.

[10] S. Yadav, A. Jain, and D. Singh, “Early Prediction of Employee Attrition using Data Mining Techniques,” Proc. 8th Int. Adv. Comput. Conf. IACC 2018, pp. 349–354, 2018, doi: 10.1109/IADCC.2018.8692137.

[11] D. S. Sisodia, S. Vishwakarma, and A. Pujahari, “Evaluation of machine learning models for employee churn prediction,” Proc. Int. Conf. Inven. Comput. Informatics, ICICI 2017, no. Icici, pp. 1016–1020, 2018, doi: 10.1109/ICICI.2017.8365293.

[12] I. Nurhaida et al., “Implementation of Deep Learning Predictor (LSTM) Algorithm for Human Mobility Prediction,” Int. J. Interact. Mob. Technol., vol. 14, no. 18, pp. 132–144, 2020, doi: 10.3991/ijim.v14i18.16867.

[13] A. Nugroho, H. L. H. S. Warnars, S. M. Isa, and W. Budiharto, “Comparison of Binary Particle Swarm Optimization And Binary Dragonfly Algorithm for Choosing the Feature Selection,” in 2021 5th International Conference on Informatics and Computational Sciences (ICICoS), 2021, pp. 24–28, doi: 10.1109/ICICoS53627.2021.9651779.

[14] B. Priambodo, A. Ahmad, and R. A. Kadir, “Predicting Traffic Flow Propagation Based on Congestion at Neighbouring Roads Using Hidden Markov Model,” IEEE Access, vol. 9, pp. 85933–85946, 2021, doi: 10.1109/ACCESS.2021.3075911.

[15] B. Jokonowo, M. A. Kulsum, and N. Komala, “Perbandingan Model Proses Algoritma Alpha dan Alpha++ Pada Aplikasi E-commerce,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 6, no. 1, pp. 123–129, 2022, doi: 10.29207/resti.v6i1.3732.

[16] V. V. Saradhi and G. K. Palshikar, “Employee churn prediction,” Expert Syst. Appl., vol. 38, no. 3, pp. 1999–2006, 2011, doi: 10.1016/j.eswa.2010.07.134.

[17] T. M. Mulyono, F. Natalia, and S. Sudirman, “A study of data mining methods for identification undernutrition and overnutrition in obesity,” ACM Int. Conf. Proceeding Ser., pp. 6–10, 2019, doi: 10.1145/3374549.3374565.

[18] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd Editio. San Fransisco: Morgan Kaufmann Publisher, 2014.

[19] H. Jantan, A. R. Hamdan, and Z. A. Othman, “Human Talent Prediction in HRM using C4 . 5 Classification Algorithm,” Int. J. Adv. Trends Comput. Sci. Eng., vol. 02, no. 08, pp. 2526–2534, 2010.

[20] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1201/9780429469275-8.

[21] R. Genuer, J. Poggi, C. Tuleau-malot, and N. Vialaneix, “Random Forests for Big Data,” Big Data Res., vol. 9, pp. 28–46, 2017.

[22] R. Punnoose and P. Ajit, “Prediction of Employee Turnover in Organizations using Machine Learning Algorithms,” Int. J. Adv. Res. Artif. Intell., vol. 5, no. 9, pp. 22–26, 2016, doi: 10.14569/ijarai.2016.050904.

[23] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, “Students’ Orientation Using Machine Learning and Big Data,” Int. J. online Biomed. Eng., vol. 17, no. 1, pp. 111–119, 2021, doi: 10.3991/ijoe.v17i01.18037.

Arabic Location Named Entity Recognition for Tweets using a Deep Learning Approach

Bedour Swayelh Alzaidi, Yoosef Abushark, Asif Irshad Khan
Department of Computer Science, Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

***Abstract*—Social media sites like Twitter have emerged in recent years as a major data source utilized in a variety of disciplines, including economics, politics, and scientific study. To extract pertinent data for decision-making and behavioral analysis, one can use Twitter data. To extract event location names and entities from colloquial Arabic texts using deep learning techniques, this study proposed Named Entity Recognition (NER) and Linking (NEL) models. Google Maps was also used to obtain up-to-date details for each extracted site and link them to the geographical coordination. Our method was able to predict 40% and 48% of the locations of tweets at the regional and city levels, respectively, while the F-measure was able to reliably identify and detect 63% of the locations of tweets at a single Point of Interest.**

Keywords—NER; Named Entity Recognition; NEL; named entity linking; event; location; deep learning; Arabic

I. Introduction

In today's world, information gathering plays an important role in reducing the effects of urgent events, so business managers or emergency agents always focus on gathering up-to-date information to help them make better decisions. So, to assure more precise dealing with situations, different resources of information must be used such as the data posted on social networking sites. The research introduces Twitter as an additional information source, but there are some difficulties and restrictions when using this Information such as data reliability and quality, as tweets may be in different languages and may contain hash tags, internet links, images, or even videos. Also, the geospatial feature of the information is important for disaster response. As for the techniques used to gather information such as Keyword monitoring which is subject and biased and more likely to contain false data. Many models have been developed to extract location-named entities from an informal text [1],[5], however, those models are applied to English text only, And the Arabic language recognition models are not linked to its geographical coordinates and are low accuracy compared to English or other languages that is because there are no functions or converters available to convert the dataset to a format that a model can use it for training. In this paper, we propose a method to extract event location names from unstructured Arabic text documents via NLP methods using the Deep Learning technique [6], first by collecting text data from Twitter and pre-processing it to fit inside the deep learning model and utilizing an Arabic location named-recognition model. In addition, each extracted location will be linked with geographical coordination by using existing knowledge such as Google Maps.

There are several ways to acquire Twitter data which can be divided into two main groups: First way: Purchasing the data from third-party or commercial data vendors [7], which provides a user-friendly GUI and visual environment for better use of the gathered data. Such an option is not recommended for the academic context due to the expenses. Also, users cannot modify the algorithms used for different operations which is a major drawback. Second way: Direct data collection through Twitter API which provides a straightforward way to query and retrieve publicly available tweets.

There are three main types of APIs provided by Twitter which are: REST API [6], Streaming API [9], and Ads API. Both REST API and Streaming API are very similar and provide access to Twitter data, but the author chose to use REST API as it is more useful for this case.

One of the most used REST endpoints is the search API which returns a collection of relevant Tweets matching a specific query. It should be noted that the REST API search service is focused on relevance and not completeness, which means that not all the tweets will be indexed. All interactions with the Twitter API require authentication using the OAuth protocol [10], which provides access to protected API services and server-side resources on behalf of the API owner. For data preparation [11], Twitter API returns tweets in JSON format [12] which is a human-readable data interchange format. Working with JSON tweets requires validation, data cleaning, and data transformation procedures before sorting them in a database. For data cleaning, it is mostly removing unwanted information or redundancy in the raw tweet which can be a group of several tasks that are: (Filtering based on the language "Arabic", Removing usernames, Removing URLs, punctuations, multiple dots, and extra spaces, and Removing words containing non-Arabic characters). Mostly, a tweet in JSON format is composed of several objects which may be more than 70 objects per tweet, some of them are single objects and some are embedded within other objects. Not all the objects will be used but selected objects during different phases for developing and implementing the framework.

Analyzing the Arabic language is challenging due to its complex linguistic structure, also the informal Arabic used in tweets is noisy and poorly structured [13] as in the Arabic region there are more than 25 dialects of Arabic. Text mining is the technique of analyzing mass amounts of text to identify insights and outcomes of new information and it is purely based on Artificial Intelligence (AI). This technique is widely involved with NLP (Natural Language Processing) which is considered an exciting field that enables computers to analyze human language. Previous research proposed a supervised algorithm that produces a weighted average of word embedding [14]. The purpose of the research is to implement text mining and NLP techniques to investigate NER for Arabic tweets using ML (Machine Learning). Recently DL (Deep Learning) has been strongly implemented in NER systems. Named Entity Recognition (NER) is a sub-task of Information Extraction (IE) that aims to look at and classify named entities presented in unstructured tasks and identify useful information such as person name, location, etc. Named Entity Linking (NEL) is the task of assigning a unique identity to the entity. Python supports NER through the library NLTK. NER and NEL can automatically scan entire articles and reveal organizations, locations, and people's names. Google Map API is used for entity analysis as it provides services related to map information.

We propose an approach that is like [15] in that it handles two binary classification tasks. We evaluate our models to classify Arabic tweets about event detectors and identify the heights performance for the training model's accuracy. Although their approach successfully detects the target messages utilizing complex and skill-required machine learning, the results might be further improved by adopting less time-consuming and labor-intensive approaches using both automated and human data labeling methods. The next part provides a summary of the current methodologies and relevant research on location inference of Twitter data after discussing the methods currently used to identify Twitter events.

The objectives of this study are as follows:

- Collect Arabic text data from the Twitter platform using Twitter API.
- Preprocess the collected data in such a way that it is well visualized and can fit inside the deep learning model using data mining, text mining, rescaling, and reshaping techniques.
- Utilize the deep learning algorithms to build an Arabic location named recognition model.
- Design a named entity link method that links the extracted location entities with their geographical coordinates.
- Optimize and retrain the developed models by using several data mining, data preprocessing, scaling data, and many other data cleaning techniques.
- Evaluate the developed models' performance in terms of classification matrix, f1 scores, precision, recall, and accuracy.
- Evaluate performance in real-time to identify issues and the accuracy of the model.

The contents of this paper are organized as follows. In Section II, we give a brief description of the existing related works. In Section III, we describe the proposed framework. In Section IV, the Implementation details of the proposed framework are discussed. In Section V, Result and Evaluation of the proposed model is discussed. Finally, the conclusion is shown in Section VI.

II. Related Work

A. Event Detection on Twitter

An event is often thought of as anything that takes place at a certain time and location [16],[17]. A qualitatively significant change in anything might also be referred to as an event. In the context of this study, the word "event" can refer to any condition in the real world that causes an increase in the frequency and quantity of tweets that are likely to include pertinent information about the circumstance [18]. A critical first step in incorporating Twitter as an extra source of information for any application is the identification of the tweets that were generated in reaction to a real-world occurrence. Event detection methods are often divided into two categories: open events and events that are specific to a certain domain (supervised or unsupervised). As a result, the bulk of research done on Twitter event detection may be grouped according to these criteria. According to the kind of event, the method of detection, and the intended use, (Table I) shows a taxonomy of research on Twitter event detection based on event type and approach.

TABLE I. Research on Twitter Event Detection

Reference	Event Type	Approach	Application
Hagras et al. (2017)[21]	Specific	Supervised	Natural Disaster tweets detection
Ragini et al. (2018) [22]	Specific	Supervised	Crisis Mitigation using Twitter
AL-Smadi et al.(2016)[19]	Open	Unsupervised	General Event Detection
Alabbas et al.(2017) [20]	Specific	Supervised	High-risk floods Detect
Wang et al. (2015) [23]	Open	Unsupervised	General Event Detection
Alomari et al. (2020)[24]	Specific	Unsupervised	Traffic-related event detection
Toujani et al. (2108)[25]	Specific	Unsupervised	Natural Disaster tweets detection
Fu et al. (2020) [26]	Open	Unsupervised	General Event Detection
Rezaei et al. (2022) [27]	Open	semi-Supervised	General Event Detection

Hagras et al. (2017) [21] classified and evaluated tweets related to the Japan Tsunami using the Latent Dircherilet Allocation (LDA) topic analysis approach. Selected 196 tweets for the test set and 6700 tweets for the training set, resulting in 76% accuracy with successful detection. However, this approach may be improved by adding more datasets to speed up processing in real-time and increase accuracy. Emotion analysis and ML algorithms were used in a crisis governance technique that [22] presented. They classified the information depending on its requirements. After using a Support Vector Machine (SVM) to assess the data acquired from Twitter via human attitudes, both positive and negative. Although the suggested approach makes it easier for emergency crews to identify disastrous situations and take appropriate action immediately, it has some limitations when using social network data for crisis mitigation, specifically the ambiguity in obtaining crisis data from various sources and the absence of the proper criterion. However, these issues may be resolved by collecting data from several sources to properly classify the data and improve precision.

AL-Smadi et al. (2016) [19] describe a knowledge-based approach for fostering event extraction out of Arabic tweets. They used an unsupervised rule-based technique for event extraction. Results show that the approach has an accuracy of, 75.9% for event trigger extraction, 87.5% for event time extraction, and 97.7% for event type identification.

Alabbas et al. (2017) [20] suggested a classification model trained on 3700 Arabic tweets to identify high-risk floods. Several machine learning algorithms were employed, which include k-NN, J48, NNET, SVM, and C5.0. With the training matrix containing the chosen terms and their related values, the classification algorithms were trained. TF-IDF (Term Frequency-Inverse Document Frequency) weights. The outcomes demonstrated that SVM performed with better accuracy.

Toujani et al. (2018) [25] suggested a novel method that uses social networks as the primary source to identify event information after a natural disaster. Then, based on the period of risk, they group people into tiers. Reporters can benefit from this clustering process by streamlining the way they obtain information in urgent situations. Additionally, they applied fuzzy theory techniques to these incidents to enhance clustering performance and get rid of opacity in the data that was gathered.

Alomari et al. (2020) [24] created a lexicon that makes it easier to identify traffic occurrences in Saudi Arabia using Twitter and the big data methodology. For Arabic and Saudi dialect terms, they also applied sentiment analysis based on the lexicon method. To improve the indexing and categorization of nonspecific events [23] provide unsupervised algorithms that create structured lexicon-based event information. Additionally, other research has focused on early event prediction using the Twitter data stream.

Fu et al. (2020) [26] perform an open-domain event text generation task with an entity chain as its skeleton. To build this dataset, a wiki-augmented generator framework containing an encoder, a retriever, and a decoder is proposed. The encoder encodes the entity chain into hidden representations while the decoder decodes from these hidden representations and generates related stories. The retriever is responsible for collecting reliable information to enhance the readability of the generated text.

Rezaei et al. (2022) [27] proposed a semi-supervised framework for the detection of data events on Twitter. Then, they used the Hierarchical Attention Network (HAN) method to categorize the data events. A virtual backbone was employed so that the stream data could be divided into one or more classes with various grades.

To extract location-named entities from informal texts like those seen on social media sites, many models have been developed [1],[2],[3],[4],[5]. While other languages have gotten comparatively less attention, those models were created to simply apply to the English text. When evaluated in the informal Arabic language, most location-named entity identification models or algorithms underperformed because they were unable to identify the precise pattern of Arabic keywords in the data. Additionally, the retrieved location's geographic coordinates on those systems are not connected to it. To the best of our knowledge, Arabic text recognition models perform poorly when compared to those in English or other languages. This is because any machine learning algorithm requires mathematical parameters to train. Because there are many programming functions for English tweets, we can readily transform them into mathematical form; but, for Arabic, there are no such converters or functions that can quickly transform our dataset into a format that the model can effectively train on.

The globe has been producing data on social media on the terabyte scale, and this unused data might be the key to many future research projects [8]. Obtaining data on behaviors affected by shifting geographies is one promising study area, allowing researchers to make focused, well-informed judgments [1]. These unstructured Arabic text documents are a significant source of such data, and we specifically want to extract geographic information from them [2]. However, most of this data is not provided in an obvious manner and must be retrieved using different NLP techniques [3]. Machine learning methods have traditionally been used to do this [4]. The technique of employing neural networks to learn a task is known as deep learning. Deep learning significantly outperforms all other conventional learning techniques and aims to replicate how the human brain functions. In this area, deep learning has not yet been applied. However, English language research makes up most of these studies. Arabic is one language that has not received enough attention. The few Arabic works that have been created, have problems with generalization. They were completed by concentrating on a tiny, extremely conservative group, leaving a sizable portion unaffected.

The location extraction and linkage from Arabic-language tweets will benefit from the findings of this study. To the best of our knowledge, there is no clear study that has tackled location extraction and linking from Arabic tweets using deep learning approaches, in addition to the scant amount of research on location extraction from Arabic text.

III. Framework

Framework design. which describes the framework's design proposed by the author to achieve the aim of this research which is only limited to tweets in the Arabic language.

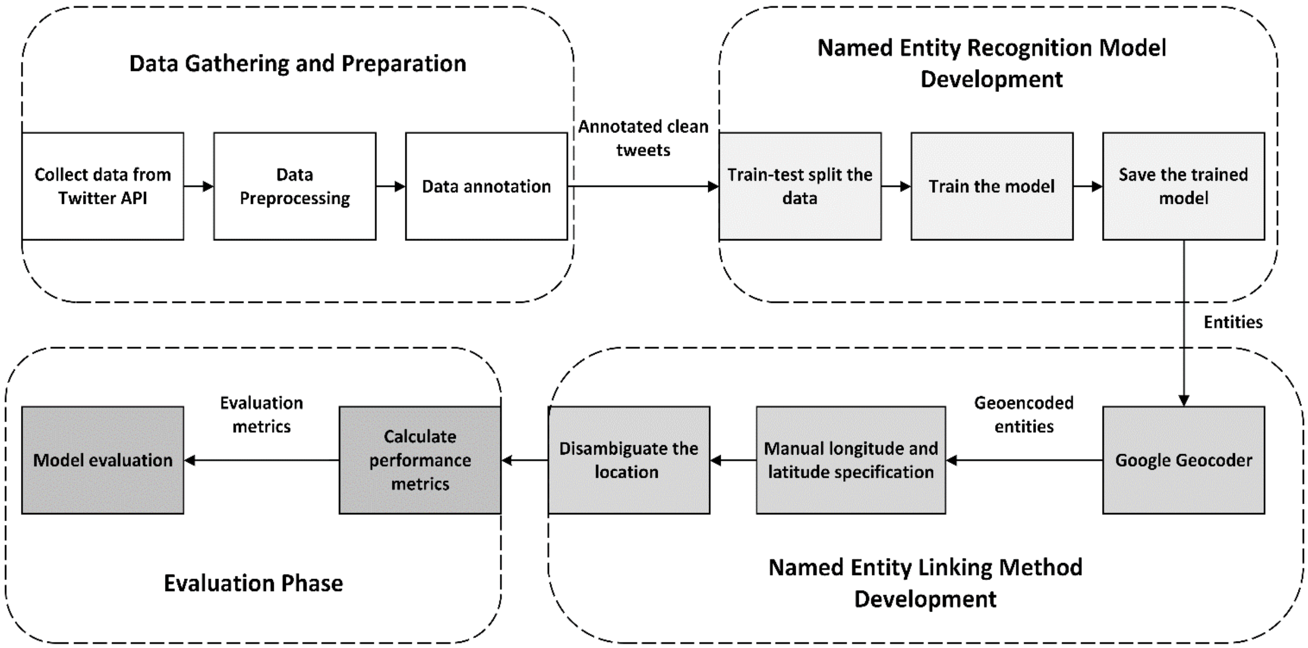


Fig. 1. Framework design.

1) First part: Data collection and Preparation which focus on the collection of Twitter data, the pre-processing, and the labeling of the data, which consists of two main components:

a) The data collection component establishes a persistent connection with the Twitter API and collects public tweets in real time.

b) The Data Preparation component focuses on the cleaning and pre-processing of the collected tweets.

Data preparation includes a few steps:

- Text Processing: a crucial step as the language used by Twitter is informal. Tweets contain noise that must be removed which may be spelling shortcuts, misspellings, new words, URLs, hashtags, tags, emoticons, or HTML characters.
- Data Annotations: an integral part of the ML model training process, because the NN learns from the pattern that exists in the annotated data and uses such patterns in new unseen data. So, we assign various entity tags to the text data.

2) Second part: Location Detection using the NER model which is created by the training set of data obtained in the first part. The NER model analyses the properties of the human Natural Language and the linguistic patterns of the incoming tweet and figures out possible words/phrases that indicate the location. Geocoding service is applied to convert the input location names extracted by the NER to output coordinates and a structured address of the location. The research uses the outputs of the Geo-coding service to hierarchize the location names.

3) Third part: Develop a Named Entity Linking method where the tweet's locations are specified and disambiguated. The location of the non-geotagged tweets can be a challenge in this context, so the Location inference component is designed to predict the location of the tweet in the absence of geo-tagging. The Location inference component uses NEL to perform the task of mapping words of interest from the tweet text to corresponding unique entities in a target knowledge base such as Wikipedia. The location assigned by the location inference component is the inferred location of a tweet.

4) Fourth part: Evaluation Phase in which the results and accuracy level are evaluated using a set of metrics for the assessment of the accuracy and performance of the system. Performance metrics are such as [Recall, precision, and F-measure] to determine the accuracy of the event-relatedness. Also, to evaluate the performance of the location inference component, Mean and Median distance errors are calculated between the inferred and actual locations. The timeliness of the processes done by the prototype is evaluated using several Twitter datasets of different sizes.

IV. Implementation

The system implementation is shown in The implementation architecture, delivers the functionalities of the proposed framework which are:

1) Twitter data collection using Twitter REST API which can cope with a large number of tweets and handle errors automatically.

2) Data preparation which includes tasks of raw data validation, reduction, annotating, and cleaning.

3) Identification of the event-related tweets and inferring their location.

4) Presentation of the results in an appropriate data format.

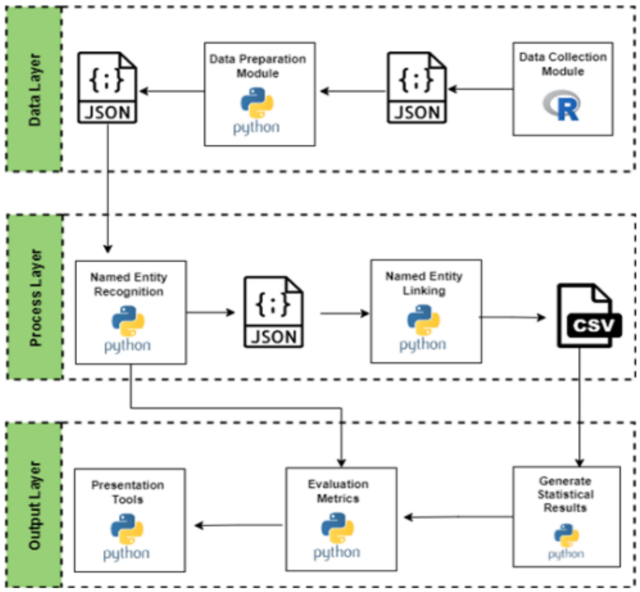


Fig. 2. The implementation architecture.

We followed the Ntair technique for implementing the proposed framework, which is a classic technique for dividing a complex development task into manageable parts (Layered Architecture)[28]. The implementation architecture consists of three layers: Data Layer, the Process Layer, and the Output Layer. The prototype implementation is based on open-source tools using R and Python which support rapid development and fast prototyping [30]. Each section will be explained as shown in Fig. 2, The Implementation Architecture.

A. Data Layer

The data layer performs the tasks of collecting, preparing, and sorting Twitter data. A combination of R and Python packages and libraries were used for the implementation of this layer.

1) Data collection module: Using the Twitter API, Arabic tweets of certain keywords related to different events were collected. The Module performs the following procedure:

- Perform API authentication using Twitter OAuth credentials.s.
- Establish and maintain a persistent connection with the API.
- Infer the approximate location of the event-related tweets from the potential sources of location information embedded within the Twitter data.

Data retrieval on Twitter is done using the web scraping method which aims to obtain information from a website and turn it into a structure that is easy to understand, store and analyze in a database. Data retrieval was done using RStudio with R programming language with the help of the R package “rtweet” which provides different functions to extract data from Twitter REST APIs. The author extracted tweets with Arabic keywords of different events such as (festivals, accidents, tornados, etc.) and stored the tweet data in feather files which are used to exchange data between Python and R. For the authentication process, The OAuth server generates the OAuth credentials (Consumer Key, Consumer Secret, Access Token and Token Secret) which the user must manually paste into the code fields in the OAuth handler of the rtweet library, as shown in the following The OAuth credentials.

The API Listener writes the raw JSON tweets to a temporary dataset which acts as a data buffer between the data collection module and the data collection module. This data buffer is nothing but a simple text-based JSON file. The collected tweets should be structured in the form of an array by the API Listener where each Twitter object is an element of this array within this file.

```
library(rtweet)
```

```
# create token named "twitter_token"
token <- "*********************************************" # From dev.twitter.com
token_secret <- "*********************************************" # From dev.twitter.com

consumerKey <- "*************************************************"
consumerSecret <- "*************************************************"

twitter_token <- create_token(
  consumer_key = consumerKey,
  consumer_secret = consumerSecret,
  access_token = token,
  access_secret = token_secret)
```

Fig. 3. The OAuth credentials.

2) Data preparation module: This module which is implemented in python comprises the following steps of validating and cleaning the collected raw tweets:

- Reads the temporary dataset of the collected tweets.
- Handles the JSON validation for each tweet within the dataset.
- Change the annotation file format to fit the SpaCy requirements.
- Performs data cleaning and pre-processing.
- Writes the cleaned tweets to a structured JSON file.

The temporary dataset is loaded into the python environment, then the module should read JSON tweets and check them for JSON schema conformance. For pre-processing the text as shown in Steps of pre-processing a sample tweet, the first step is iterating over the tweets to remove all numbers, English alphabets, and punctuations such as commas (,), period (.), semi-colons (;), colons (:), question marks (?), Arabic question marks, semicolons and so on, to reduce the size of the feature set. Further, removing the hash (#) and underscore (_) symbols from the hashtags. Also, the author removed Arabic diacritic and vowel marks such as Shaddah, which is a diacritic shaped like a small written "w." After that, the text is divided into words (tokens). The tokens are normalized to replace letters that have different forms into the basic shape. Finally, the Stop Words are filtered using the Arabic stop words list in the Natural Language Toolkit (NLTK). The author modified the list to add the missing word and normalize the words before using them.

Furthermore, checking the result of the pre-processing phase before starting the classification. If the remaining number of tokens is equal to zero, the tweet is excluded from the analysis. The following figure shows the steps of pre-processing applied to a sample tweet:

Fig. 4. Steps of pre-processing a sample tweet.

Once the dataset is cleaned, it is stored in a new JSON file that will be later accessed by the Process Layer for conducting the processes associated with the location extraction using NER.

3) Data annotation: it is an important part of the ML training process as the NN will be trained from the patterns that exist in the annotated data and tries to identify such patterns for unseen data. Data Annotation is carried out using Label Studio which is an open-source data annotation tool whether locally or on a server using a simple user interface. For NER, we would need to assign various entity tags to the text data we have which are used later by the NN to identify a generalized pattern. The author identified three main entities (Point of interest, City, Region) POI, CTY, REG. Then using word frequency analysis, evaluated each tweet against term classes that are associated with practical locations related to certain events to find the tweets are connected to which location. So, the author considered random location-related tweets which are manually annotated then by applying different entity tags to the existing data to perform NER. The model analyses these collections of tweets with their entity tags to find a generalized pattern. Such a setup is simple for NER tagging, with only the names of tags changed on the customization panel. File imported in the Label Studio is of the supported formats (.csv, .json, .tsv), and after the job is done exporting the annotated data in a format of the same previous supported formats. The time-consuming process of data annotations is made simpler with Label Studio.

B. Process Layer

The core implementation layer deals with tasks of location extraction of the collected tweets.

Model Training:

Data is split into two sets: training and testing sets. The annotated tweets are used for the training set to train the model. Then the model is used to identify the location entities in the testing set. The NER Model development process is carried out using the SpaCy library. The training data is converted in a form of tuples containing text data and a dictionary which is a format supported by SpaCy. The following steps are carried out to train the model:

1) Load the model or create an empty model using spaCy blank with the ID of desired language (Arabic in this case).

2) Add the new entity label to the entity recognizer using the add_label method.

3) Loop over the examples and call the Natural Language Processor (NLP) which steps through the words of the input. Predict each word. Then consult the annotations, to see whether it was right. If it was wrong, it adjusts its weights so that the correct action will score higher next time.

4) Save the trained model.

5) Test the model to make sure the new entity is recognized correctly.

6) Explore the results of the evaluation.

Location Named Entity Linking

Using a Geo-coding service, it takes an input text such as an address or the name of a place and returns a latitude/longitude location on the Earth's surface for that place. The locations are decoded using geocoding APIs offered by Google Places and OpenStreetMap. The geocoded service results in a JSON-formatted list of geocoded addresses representing possible matches, including longitude and latitude coordinate and well-formatted address with all the super regions included. If the extracted location is linked to more than a location, The Linker will disambiguate the locations by estimating the distance between the geospatial location and candidate geospatial locations.

C. Output Layer

This layer uses the structured data of the processed tweets to display the results by generating statistical results and evaluation metrics. Results are output from two main operations:

1) Displacement Computations where the displacement between Twitter's entity location points and the coordinates that result from Google Geocoder API, to assess the Geocoding accuracy.

2) Generating statistical results: the frequency distributions of displacement values for SpaCy are computed. Then various groupings of SpaCy entities are evaluated to find the entity grouping giving the highest accuracy in predicting.

3) Evaluation metrics: Three metrics are commonly used for NER and NEL tasks which are Precision, Recall, and F1.

V. Results and Evaluation

A. Data Collection and Preparation

Twitter REST API was used to collect tweets with a specific keyword, where the author conducted the experiment using "accident" as the initial keyword.

Data collection continued up to weeks, during which 11134 unique tweets were collected and stored in different feather files. Then the feather files were concatenated to construct the initial dataset that is processed by the Data preparation module. After the data preparation module, the initial dataset was reduced to 9894 unique tweets after pre-processing and further reduced to 1982 tweets after annotation. A reference dataset of 320 tweets can be defined as a representation of the correctly classified sample tweets for verifying the model is manually annotated by experts.

B. Evaluation of the NER Model

For measuring the performance, each output word was compared to the truth file and assigned a value of True Positive (TP), False Positive (FP), True Negative (TN), or False Negative (FN).TP, FP, TN, and FN counts for all tweets were added and Precision, recall, and F1 scores were calculated per entity (POI, REG, and CTY as the core entity sets for evaluation). The author only relied on F-score to measure the optimal performance of the results as it combines both precision and recall metrics. The F-measure was able to correctly identify and detect 63% of the location of tweets in a certain Point of interest while on the region and city level, the system was only able to predict 40% and 48% of the location of tweets, respectively, as described in Table II.

TABLE II. RESULTS OF THE NER MODEL

Entity	Precision	Recall	F1
POI	77	54	63%
REG	60	30	40%
CTY	57	41	48%

C. Evaluation of the NEL Model

Second phase in the proposed framework is NEL, this phase aims to extract the geolocation for the extracted entities. Table III represent the evaluation of the NEL model, where:

1) The (Label) field represents the corresponding location name label (POI, CTY, REG) assigned to each tweet.

2) (Entity), (lat pred) and (lng pred) fields represent the name and the geo-coordinates (latitude and longitude) of the extracted location.

3) (lat act) and (lng act) fields correspond to the geotagged coordinates of the tweets that extracted manually (Ground truth).

4) The (Distance_km) field represents the distance error (the distance between the actual location and the inferred location of a tweet) which is calculated using the Haversine formula. This field is used as the evaluation metric to measure the accuracy of the proposed solution results. As for the results, in 191 out of 320 tweets with a percentage of (61.3%), the inferred location was at a distance equal to or smaller than 10 km from their actual location. And in 88 tweets (27.5%) the inferred location was located within 10 to 50 km of their actual location. And among the remaining tweets, the locations of 10 tweets (3.1%) were located within 50 to 100 km. While 26 tweets (8.1%) have a distance error greater than 100 km. Fig. 5 below shows the accuracy of the inferred location based on distance error.

TABLE III. RESULTS OF THE NEL MODEL

	Lat act	Lng act	label	entity	Lat pred	lng pred	Distance km
1	30.04	31.23	['POI', 'CTY']	['القاهرة','طريق رئيسي']	30.04	31.23	0.17
2	24.66	46.68	['POI']	['طريق الملك سعود']	21.81	39.08	838.17
3	29.30	48.029	['POI']	['شوارع ميدان حولي']	29.33	48.02	2.84
4	30.01	31.58	['POI']	['الأوسطي']	30.01	31.58	0.004
5	30.01	31.58	['POI']	['الأوسطي']	30.01	31.58	0.004

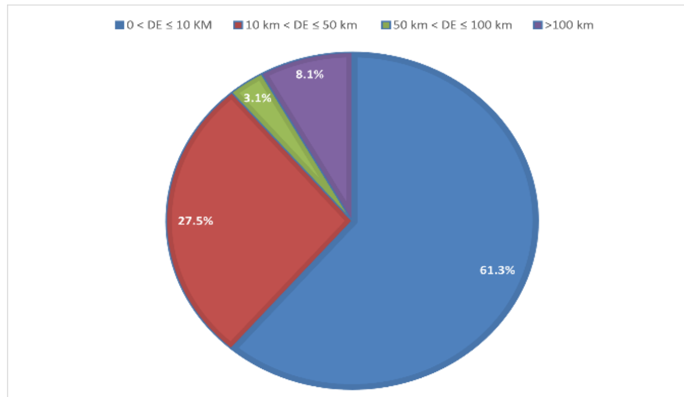


Fig. 5. The accuracy of the inferred location.

VI. CONCLUSION AND FUTURE WORK

The framework proposed can assist emergency/event managers to locate necessary warnings and monitor situations. Also, it can provide information on the type and scope of damage that resulted from a certain event. Twitter data collection and preparation is considered one of the foremost contributions of the research. This study proposes a method to process tweets to determine their degree of relatedness to accident events and infer their approximate location. First of all, getting to know the nature and structure of Twitter data together with the identification of the optimum approaches to its collection, storage, and preparation by the emergency response context, by itself, is an essential knowledge area and can be considered as one of the foremost contributions of this research. However, this assertion, along with the possibility of the adoption of the framework in other types of incidents such as bushfires or terrorist attacks, is subject to further research. Another point is that the framework only focused on Arabic tweets. This could be seen as a major drawback in using the framework on a global scale or in countries that speak a language other than Arabic, especially languages that use non-ASCII characters like in our case (e.g., Turkish and Chinese). Therefore, exploring possible solutions to this issue could form a future research topic. The next section outlines several future research directions.

For future research, the manual annotation process is time-consuming and laborious, so an investigation into automation methods is much recommended. Also, the research only considered the textual part of the tweet excluding any images, videos, or links that may likely provide valuable information, so providing visualization of the tweet's content may be suggested for future work. Also, semantic-based solution for accurate assessment of emoticons, acronyms, or slang. As mentioned earlier, extending the applicability of the proposed framework to other commonly spoken languages can improve the performance of the framework. For location accuracy, integration of other sources of information such as remote sensing data, sensor network data, and crowd-sourced mapping repositories can improve the measurement and minimize the error. For Geo-coding, using other sources for location names and geographical databases other than Google Map API such as GeoNames may enhance the applicability of the proposed framework at larger geographic scales.

REFERENCES

[1] Etaiwi, W., Awajan, A. and Suleiman, D. (2017), Statistical Arabic name entity recognition approaches: A survey, Procedia computer science 113, 57–64.

[2] Feriel, A., & Kholladi, M. K. (2015). Automatic Extraction of Spatio-Temporal Information from Arabic Text Documents. Int. J. Comput. Sci. Inf. Technol, 7, 97-107.

[3] Qiu, Q., Xie, Z., Wu, L. and Tao, L. (2019), Gner: A generative model for geological named entity recognition without labeled data using deep learning, Earth and Space science 6(6), 931–946.

[4] Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering, 34(1), 50-70.

[5] Ali, M. N., Tan, G. and Hussain, A. (2018), Bidirectional recurrent neural network approach for Arabic named entity recognition , Future Internet 10(12), 123.

[6] Qiu, Q., Xie, Z., Wu, L., & Tao, L. (2019). GNER: A generative model for geological named entity recognition without labeled data using deep learning. Earth and Space science, 6(6), 931-946.

[7] Kim, A. E., Hansen, H. M., Murphy, J., Richards, A. K., Duke, J., & Allen, J. A. (2013). Methodological considerations in analyzing Twitter data. Journal of the National Cancer Institute Monographs, 2013(47), 140-146.

[8] About Twitter | Our company purpose, principles, leadership [n.d.]. URL:https://about.twitter.com/en/who-we-are/our-company.

[9] API reference index | Docs | Twitter Developer Platform [n.d.]. URL: https://developer.twitter.com/en/docs/api-reference-index.

[10] OAuth 2.0 — OAuth [n.d.]. URL: https://oauth.net/2/.

[11] Thakare, S. B. and Gawali, S. Z. (2010), 'An effective and complete preprocessing for web usage mining', International Journal on Computer Science and Engineering 2(03), 848–851.

[12] JSON [n.d.]. URL: http://www.json.org/json-en.html.

[13] AlOtaibi, S. and Khan, M. B. (2017), Sentiment analysis challenges of informal Arabic, IJACSA) International Journal of Advanced Computer Science and Applications 8(2).

[14] Gupta, S., Kanchinadam, T., Conathan, D. and Fung, G. (2020), Task optimized word embeddings for text classification representations, Frontiers in Applied Mathematics and Statistics 5, 67.

[15] Ameen, Y. A., Bahnasy, K., & Elmahdy, A. Classification of Arabic Tweets for Damage Event Detection.

[16] Allan, J., Papka, R. and Lavrenko, V. (1998), On-line new event detection and tracking, in 'Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval', pp. 37–45.

[17] Nallapati, R., Feng, A., Peng, F. and Allan, J. (2004), Event threading within news topics, in 'Proceedings of the thirteenth ACM international conferenceon Information and knowledge management', pp. 446–453.

[18] Dou, W., Wang, X., Skau, D., Ribarsky, W. and Zhou, M. X. (2012), Leadline: Interactive visual analysis of text data through event identification and exploration, in 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), IEEE, pp. 93–102.

[19] Mohammad, A. and Qawasmeh, O. (2016), Knowledge-based approach for event extraction from arabic tweets, International Journal of AdvancedComputer Science & Applications 1(7), 483–490.

[20] Alabbas, W., al Khateeb, H. M., Mansour, A., Epiphaniou, G. and Frommholz,I. (2017), Classification of colloquial arabic tweets in real-time to detect high risk floods, in 2017 International Conference On Social Media,Wearable And Web Analytics (Social Media), IEEE, pp. 1–8.

[21] Hagras, M., Hassan, G. and Farag, N. (2017), Towards natural disasters detection from twitter using topic modelling, in '2017 European Conference on Electrical Engineering and Computer Science (EECS)', IEEE, pp. 272–279.

[22] Ragini, J. R., Anand, P. R. and Bhaskar, V. (2018), Big data analytics for disaster response and recovery through sentiment analysis, International Journal of Information Management 42, 13–24.

[23] Wang, Y., Fink, D. and Agichtein, E. (2015), Seeft: Planned social event discovery and attribute extraction by fusing twitter and web content, in Proceedings of the International AAAI Conference on Web and Social Media, (Vol. 9, pp. 483–492).

[24] Alomari, E., Mehmood, R. and Katib, I. (2020), Sentiment analysis of arabic tweets for road traffic congestion and event detection, in Smart infrastructure and applications, Springer, pp. 37–54.

[25] Toujani, R. and Akaichi, J. (2018), Event news detection and citizens community structure for disaster management in social networks', Online Information Review 43(1), 113–132.

[26] Fu, J., Liu, P. and Neubig, G. (2020), Interpretable multi-dataset evaluation for named entity recognition, arXiv preprint arXiv:2011.06854.

[27] Rezaei, Z., Eslami, B., Amini, M. A. and Eslami, M. (2022), Event detection in twitter by deep learning classification and multi label clustering virtual backbone formation, Evolutionary Intelligence pp. 1–15.

[28] Chengjun, W. (2009), The analyses of operating system structure, in 2009 Second International Symposium on Knowledge Acquisition and Modeling, (Vol. 2, IEEE, pp. 354–357).

[29] Ginige, A. (1998), Web engineering: Methodologies for developing large and maintainable web-based information systems, in 'IEEE International Conference on Networking', pp. 89–92.

[30] Wallace, B. C., Dahabreh, I. J., Trikalinos, T. A., Lau, J., Trow, P. and Schmid, C. H. (2012), Closing the gap between methodologists and end-users: R as a computational back-end, Journal of statistical software 49, 1–15.

Low Complexity Classification of Thermophilic Protein using One Hot Encoding as Protein Representation

Meredita Susanty[1], Rukman Hertadi[2], Ayu Purwarianti[3], Tati Latifah Erawati Rajab[4]*
School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia[1, 3, 4]
School of Computer Science, Universitas Pertamina, Jakarta, Indonesia[1]
Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia[2]
Center for Artificial Intelligence (U-CoE AI-VLB), Institut Teknologi Bandung, Bandung, Indonesia[3]

***Abstract*—The laborious, and cost-inefficient biochemical methods for identifying thermophilic proteins necessarily require a rapid and accurate method for identifying thermophilic proteins. Recently, machine learning has become a more effective method for identifying specific classes of extremophiles. There is still a need for a low-cost method for identifying thermophilic proteins, despite the fact that studies employing machine learning yielded superior results to conventional methods. Here, we avoid the problem of manually crafted features, which involves experts defining and extracting a set of features using only protein sequences as input for various computational methods. This study classifies thermophilic proteins and their counterparts using only protein sequences in one-hot encoding representation and the bidirectional long short-term memory (BiLSTM) model. The model achieved an accuracy of 92.34 percent, a specificity of 91 percent, and a sensitivity of 93.77 percent, which is superior to other models reported elsewhere that rely on a number of manually crafted features. In addition, the more trustworthy and objective data set and the independent data set for evaluation make this model competitive with other, more accurate models.**

Keywords—Thermophilic; classification; one-hot encoding; BiLSTM

I. Introduction

Extremophiles are microorganisms that have adapted to inhabit ecological niches deemed "extreme" due to unfavorable environmental conditions, such as excessively high or low temperatures, extreme pH values, high salt concentrations, or high pressure. Proteins isolated from extremophiles are biomolecules that possess unusual properties. One exceptional property enables proteins to function under extreme conditions. Researchers use extremophilic proteins in industrial applications since they can resist harsh conditions, which presents new opportunities for biocatalysts and biotransformation [1]–[4]. Thermophiles, extremophiles that thrive at temperatures between 41 and 122 degrees Celsius, with optimal growth temperatures between 60 and 108 degrees Celsius, are among the most studied.

Identifying thermophilic proteins' biochemical and physicochemical properties is critical because this serves as the foundation for designing and engineering proteins and enzymes. However, the biochemical identification method for thermophilic proteins is time-consuming, labor-intensive, and costly. Therefore, a rapid and accurate method for identifying thermophilic proteins is urgently required.

The increasing availability of data on extremophilic proteins enables computational methods to predict protein classification and identify the key characteristics that define that class [5]. These computational methods provide a more effective means of identifying specific classes of extremophiles than biochemical methods that require wet lab experiments. Many researchers have recently attempted to distinguish thermophilic organisms from their counterparts using machine learning [6]. In general, classifying thermophilic proteins using an approach based on machine learning involves six steps: dataset collection, data pre-processing, feature extraction, feature or dimension reduction, classification, and evaluation. Numerous techniques have been developed for researching and identifying thermophilic proteins. Kumar et al. (2000) identified and categorized thermophilic proteins based on structural differences at room temperature [7], whereas Gromiha et al. (2001) studied the properties of amino acids and the effects of amino acid residues on protein heat resistance [8]. These studies rely on expensive and ineffective biological techniques from the past. In other studies, Zhou et al. (2008) used an amino acid coupling model to identify thermophilic proteins [9]. Zhang et al. (2006) utilized dipeptide composition and amino acid composition to differentiate between mesophilic and thermophilic proteins [10] and achieved an accuracy of 86.6% for five-fold cross-validation. Using neural network-based amino acid composition [11], Gromiha and Suresh (2008) increased the computational complexity of five-fold cross-validation by 89%. Using decision tree methods, Wu et al. classified and identified thermophilic proteins with an accuracy of over 80% [12]. The accuracy of the k-nearest neighbor classifier used by Zuo et al. (2013) to classify thermophilic proteins was 91.02 percent [13].

To classify thermophile proteins and their counterparts, most previous studies have used hand-picked features calculated as input features for various machine learning models. During the feature extraction stage, various features are extracted using various computational methods based on amino acid sequence and transformed into numerical vectors. Protein features have been used in previous studies [6], [11]–

*Corresponding Author.

[19]. Even the most recent study [20], [21] still employs this type of protein representation. However, their scalability is limited because some of these input features are not always available and are computationally expensive to generate. Wu et al. (2009) predicted protein thermostability using protein structure and sequence characteristics, and suggested that although sequence and structural models had slightly higher accuracy, sequence-only models can provide sufficient accuracy for thermostability prediction [12]. Inspired by Wu's claim, we use only protein sequence as input to avoid the problem of obtaining handcrafted features, which requires experts to define and extract a set of features using various computational methods. This study improves accuracy by using sequence-only, removing all complexity to calculate derived features, and bidirectional long short-term memory (BiLSTM).

II. Methods

The core structure of the present research consists of the five processes listed below: (1) dataset collecting, (2) feature extraction, (3) classification, and (4) performance evaluation. The framework's flowchart is presented in Fig. 1. One-hot representation is used to encode protein sequence. This representation is input to three distinct classifiers: Multi-layer Perceptron (MLP), Convolutional Neural Network (CNN), and Bidirectional Long Short-Term Memory (BiLSTM).

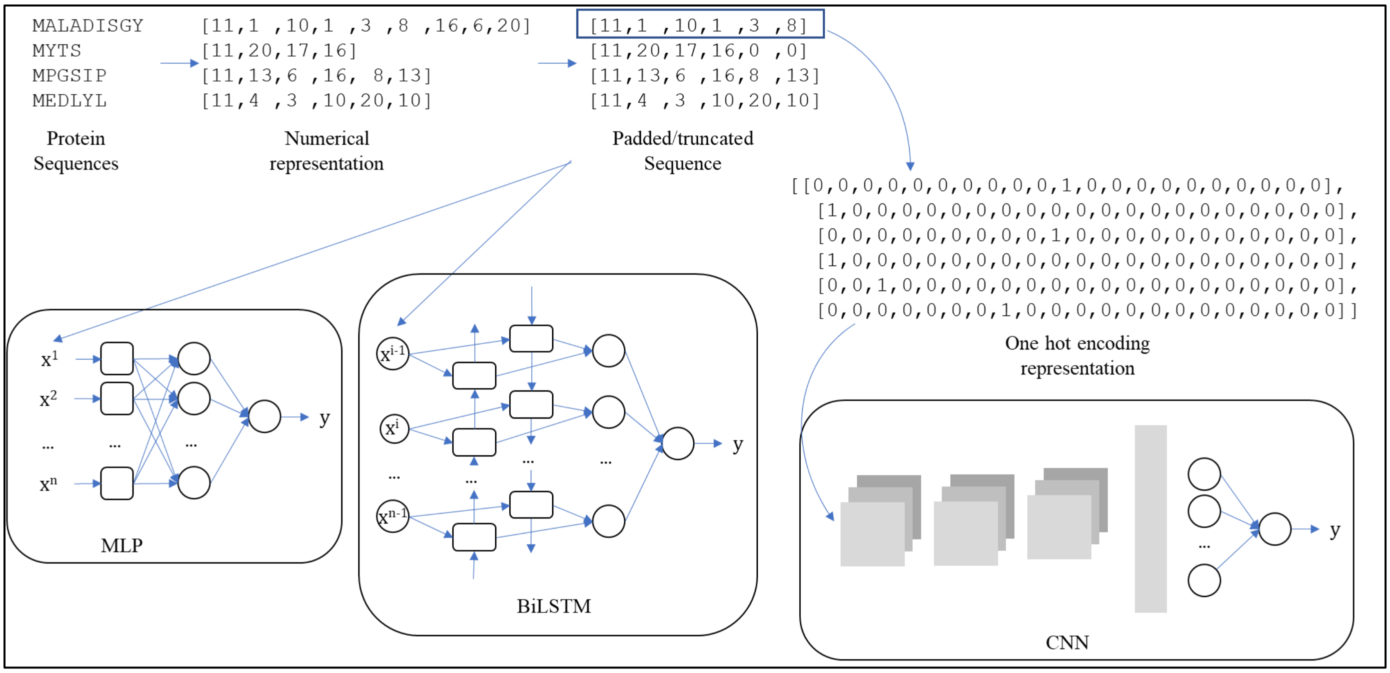


Fig. 1. Flow chart of a framework for predicting thermophilic proteins

A. Datasets

We utilized Ahmed et al. (2022) thermophilic and mesophilic benchmark dataset [21]. The dataset originated from the Universal Protein Resource (http://www.uniprot.org). To provide reliable data, previous studies retained only proteins that had been manually reviewed and excluded proteins with ambiguous residues, sequences that were fragments of other proteins, and proteins inferred from prediction or homology. With the CD-HIT algorithm [22] and a sequence identity threshold of 30%, redundancy and homology bias have been eliminated. There were 1,443 non-thermophilic proteins and 1,367 thermophilic proteins in the final benchmark dataset. The training and testing datasets are identical to those used in the study by Ahmed et al [21].

B. Feature Extraction

Here, instead of describing a protein with physical attributes, we directly encode its amino acid sequence. This approach of vectorizing categorical data is called one-hot encoding. An $L \times n$ matrix encodes a protein sequence of length L, where n is the number of amino acids. Each row of the matrix consists of $(n-1)$ 0s and a single 1, with the location of the 1 representing the amino acid residue in the protein at that position.

Using the constructed code dictionary, a 1 letter code is substituted for an integer value for each unaligned amino acid sequence. If the code is not included in the dictionary, the value is replaced by 0 and only the 20 most common amino acids are considered. This phase will transform the 1 letter code sequence data into numerical data. Next, post padding is performed with a maximum sequence length of 1024, either padding with 0 if the entire sequence length is less than 1024 or truncating the sequence to a maximum length of 1024.

C. Classification

We examined several classifiers, including MLP, CNN, and BiLSTM, to find the best model for classifying thermophilic proteins. **MLP** is a feed-forward neural network having input, hidden, and output layers, which are responsible for receiving, processing, and final prediction, respectively. The network is trained using backpropagation with the supervised learning technique. Each trained neuron's output is defined by equation (1), where x_i represents the firing neuron's input values, w_i their weights, f the activation function, and b the neuron's activation threshold [23], [24]. In the current study, the hidden layer activation function was a rectified linear activation unit (ReLU), while the outer layer activation function was sigmoid. To train the model, input, hidden, and output layers with one neuron, respectively, were used. Table I. shows the hyperparameters in detail.

$$f(\propto) = f(\sum_{i=1} w_i x_i + b) \quad (1)$$

CNN employs layers with convolving filters applied to local features [25]. The use of convolutional neural networks is the same as that of picture data. The sole difference is that 1D convolutions are applied as opposed to 2D convolutions. The convolution network is made of the following: (1) A Conv1D with 128 units, with the relu activation function and a kernel size of two to extract basic properties, (2) another Conv1D with 128 units with relu activation function, kernel size of two, and l2 regularizer, (3) a dropout layer that randomly drops nodes during training, (4) a MaxPooling1D layer that down-samples the input by taking max over the steps that are constrained to a pool_size in each stride to reduce the spatial size of the representation, (5) another dropout layer, (6) a Dense layer with 250 units for the fully connected layer, and (7) an output layer with the sigmoid activation function because this is a binary problem. Table I. shows the hyperparameters in detail.

LSTMs are excellent at maintaining long-term memory dependencies. The LSTM architecture accomplishes this through the use of input gate, the forget gate, and the output gate [26]. The inputs of unidirectional LSTM, as we can see, are from the past. Hence, it only retains past knowledge. On the other hand, a bidirectional LSTM is able to preserve contextual information from both the past and the future at any time because it run the inputs in two directions, one from the past to the future and one from the future to the past [27]. The first layer of the model is the embedding layer which uses the 256-length vector, and the next layer is the bidirectional LSTM layer which has 256 neurons. These will work as the memory unit of the model, which has a vocab size of 21 representing 20 unique amino acids and one padding character. L2 regularization is added to prevent model over-fitting. After bidirectional LSTM, the dense layer is an output layer with sigmoid function. The hyperparameter used in this study is summarized in Table I.

D. Classification

In order to evaluate the overall model performance, the following parameters were used

$$Sn = \frac{TP}{TP+FN} \quad (2)$$

$$Sp = \frac{TN}{TN+FP} \quad (3)$$

$$Acc = \frac{TP+TN}{TP+FN+FP+TN} \quad (4)$$

$$F1 = 2 \times \frac{Sn \times (\frac{TP}{TP+FP})}{Sn + (\frac{TP}{TP+FP})} \quad (5)$$

where Sn, Sp, Acc, F1 denote sensitivity (or recall), specificity, accuracy, and F1-score. In this study we use the F1-score to make sure that when the accuracy is high, it correctly predicts both classes. Thermophilic proteins classified as thermophilic are designated as TP (true positive), non-thermophilic proteins labeled as non-thermophilic are TN (true negative), non-thermophilic proteins classed as thermophilic are FP (false positive), and thermophilic proteins regarded as non-thermophilic are FN (false negative).

III. Results and Discussion

In this study we solely used protein sequences to classify thermophilic proteins. This chain of amino acids was represented using one-hot encoding which is the input for the classifier. Three machine learning techniques, MLP, CNN, and BiLSTM were used to identify thermophilic and non-thermophilic proteins. Hyperparameters tuning was performed to find the optimal model. The best optimized hyperparameters for each model are described in Table I. The result of each model is shown in Table II. The result of each classifier is measured using the following performance metric; accuracy, specificity, recall, and F1-score. The macro average values of specificity, recall, and F1-score are presented in Table II.

TABLE I. Best Hyperparameter

Hyperparameter	MLP	CNN	BiLSTM
Number of Layer	2	4	3
Number of Neuron	100-1	250-1	256-256-1
Number of Filter	-	128	-
Activation Function	Relu	Relu	Relu
Optimizer	Adam	Adam	Adam
Learning Rate	0.0001	0.001	0.001
Dropout	-	0.6 - 0.8	-
Regularizer	0.4	0.9	0.1
Early Stopping	True	True	True
Batch Size	60	256	256
Epoch	250	250	500

TABLE II. Performance Result

Model	Sn (%)	Sp (%)	Acc (%)	F1(%)
MLP	2.19	97.23	51.06	4.18
CNN	90.10	89.96	90.03	89.76
BiLSTM	**93.77**	**91.00**	**92.34**	**92.25**

[a] Note: Sn: sensitivity (or recall), Sp: specificity, Acc: accuracy, and F1: F1-score

A. Performance Comparison on Different Algorithms

We compared the performance of three machine learning methods: MLP, CNN, and BiLSTM. The same features were used to train and test these methods. Based on the result as shown in Table II, the highest accuracy, sensitivity, specificity, and f-measure were achieved using BiLSTM, at 92.34%, 93.77%, 91.00%, and 92.25% respectively. The MLP, on the other hand, was only capable of achieving an accuracy of 51.06%.

TABLE III. PERFORMANCE OF CNN ON VARIOUS K-MERS

k-mers	Sn (%)	Sp (%)	Acc (%)	F1(%)
1	61.53	53.63	57.47	58.43
2	82.78	89.27	86.12	85.28
3	**90.10**	**89.96**	**90.03**	**89.76**
4	79.85	94.11	87.18	85.82
5	86.81	89.27	88.07	87.61

b. Note: Sn: sensitivity (or recall), Sp: specificity, Acc: accuracy, and F1: F1-score

MLP can theoretically approximate any function to any precision. Using the same dataset, MLP architecture in other studies which used protein sequence and several other features as the input could achieve accuracy of 99.26% [21]. However, MLPs were not ideal for processing patterns with sequence. Because we rely on the sequence of amino acids in this study, MLP strives to remember patterns in sequential data to discover reliance on historical data, which is extremely useful for prediction.

CNN learns to recognize spatial patterns. CNN worked remarkably well, however, when applied to specific NLP issues [28]. When applying CNN to sequential data, the result of each convolution will be triggered when a unique pattern is identified. By altering the size of the kernels and concatenating their outputs, it is possible to identify patterns of numerous sizes, such as 2, 3, or 5 adjacent amino acids (k-mers), which are analogous to the term n-grams. Therefore, CNN can distinguish this k-adjacent amino acid regardless of its position in the sequence. The results of our experiments with various k-mers are summarized in Table III, which reveals that prediction using three consecutive residues (amino acid 3-mers) provides the highest performance with an accuracy of 90.03 percent, a sensitivity of 90.10 percent, a specificity of 89.96 percent, and an f-measure of 89.76 percent.

TABLE IV. PERFORMANCE COMPARISON

Studies	Features[c]	Algorithms[d]	Sn (%)	Sp (%)	Acc (%)
Liu[19]	AAC	SVM	98.88	1.0	99.44
Feng [15]	AAC, reduced DC, physicochemical	SVM	98.2	98.2	98.2
Ahmed[21]	AAC, tPseAAC, aPseAAC, CKSAAP, DC, DDE, CTD	MLP	96.34	96.16	96.26
Guo[29]	AAC, DC, DDE, CTDC, CTDT, CTriad, CKSAAP, GTPC, GDPC, TPC	SVM	96.22	95.85	96.02
Wang[17]	pseAAC, AAC, PC, CTD	SVM	96.17	95.69	95.93
Sunny and Saleena[30]	AA frequency, pI, protein binding domain, disorder regions, conserved residues, buried exposed regions	RF	95.53	96.01	95.71
Tang [31]	5-mers AA	SVM	94.8	94.1	94.4
Charoenkwan [20]	AAC, DPC, CTDC, CTDD, CTDT, AAI, aPseAAC, pseAAC, PSSM_COM, RPM_PSSM, S_FPSSM	Meta-predictor optimization	95.1	93.3	94.2
Fan et al.[14]	AAC,pseAAC,pKA, PSSM-400	SVM	89.50	95.64	93.53
Nakariyakul [16]	AAC, DC	SVM	93.0	93.7	93.3
Wang [32]	AAC, PCP, GDC, entropy density, autocorrelation coefficient	SVM	91.68	93.44	92.56
Zhang and Fang [33]	AA Sequence	SVM	92.8	92	92.4
BiLSTM (Our Model)	**AA sequence**	**BiLSTM**	**93.77**	**91.00**	**92.34**
Albayrak [34]	RAAA, N-grams	SVM	92.1	91.4	91.796
Zuo[13]	AAC-based similarity distance	KNN	88.37	92.24	91.02
Lin and Chen [35]	AAC, GDC	SVM	85.40	93.60	90.8
Gromiha and Suresh [11]	AAC	NN	83.30	92.00	89
Zhang and Fang [18]	AAC, DC	LogitBoost	87.34	90.77	88.94
Wu[12]	secondary and tertiary structural features	DT	73.1	88.5	83.6

c. amino acid composition (AAC), traditional pseudo amino acid composition (tPseAAC), amphiphilic pseudo amino acid composition (aPseAAC), pseudo amino acid composition (PseAAC), the composition of k-spaced amino acid pairs (CKSAAP), dipeptide composition (DC), dipeptide deviation from the expected mean (DDE), composition, transition, and distribution (CTD), CTD Composition (CTDC), CTD Transition (CTDT), Cojoint Triad (Ctriad), Grouped Dipeptide Composition (GTPC), Grouped Tripeptide Composition (GDPC), Tripeptide Composition (TPC) , Physicochemical Properties (PCP), g-gap Dipeptide Composition (GDC), Position Specific Scoring Matrices (PSSM), Physic Chemical (PC), Isoelectric Point (pI), Amino Acids (AA), Acid Dissociation Constant (pKa),Reduced Amino Acid Alphabet (RAAA), composition- transition-distribution -based features containing CTD (CTDC), distribution in CTD (CTDD), transition in CTD (CTDT), physi- cochemical property-based features containing amino acid index (AAI), evolutionary information-based features containing position-specific scoring matrix composition (PSSM_COM), PSSM of log-odds score of each amino acid in each posi- tion (RPM_PSSM) and PSSM based on the matrix transformation (S_FPSSM)

d. Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF), Bidirectional Long Short-Term Memory (BiLSTM), K-Nearest Neighbour (KNN), Neural Network (NN), Decision Tree (DT)

Convolutions and pooling processes eliminate information about the local order of amino acids, making it more difficult to implement sequence tagging inside a pure CNN architecture. Additionally, pooling minimizes output dimensionality while maintaining the most essential information. Each filter recognizes a unique characteristic. If this specific feature appears somewhere in the sequence, applying the filter to that region will result in a large value, while applying the filter to other regions will result in a small value. By performing the max operation, we retain information regarding whether or not the feature existed in the sequence, but we lose information regarding where the feature appeared. Although it operates wonderfully, it is incapable of interpreting temporal data. In natural language processing, the resemblance between proteins and sentences explains why CNN performs worse than BiLSTM, which can learn from both past and future data in the sequence.

Among all models, the BiLSTM Model produces the best results. Because it knows what amino acids follow and precede an amino acid in the protein sequence, BiLSTM effectively increases the amount of information available to the network, improving the context available to the algorithm. BiLSTM mines the relationships between contexts in both directions, resulting in a better recognition for classifying thermophilic proteins.

B. Comparison to Other Models

Comparison with previously published methods is required to establish the superiority of the new method. It is widely known that the prediction results would be exaggerated if the suggested technique was trained and evaluated using a benchmark dataset containing a large proportion of homologous sequences. If two protein sequences possess a sequence similarity of greater than 40 percent, they are considered homologous. When a model is trained and evaluated on such a dataset, the vast majority of predictors always attain high levels of accuracy. For instance, previous studies [33] did not eliminate very identical or homologous sequences with 40% sequence identity from their datasets. Other studies only used a 40% cutoff eliminate the homologous sequence [11]. This study's dataset, which uses a 30% cutoff, is therefore judged to be more objective and dependable.

Numerous models have been developed to identify thermophilic proteins [6], [11], [14]–[17], [19], [29], [31]. All of the proposed models were developed using machine learning techniques and assessed using cross-validation. Nevertheless, our model was evaluated using independent data.

While state-of-the-art (SOTA) methods typically use various features derived from amino acid sequences as input to identify thermophilicity, the method(s) presented here use encoded single protein sequences as one-hot encoding, which serve as the only input feature for the prediction. One-hot encodings are sparse, memory inefficient, and high-dimensional by definition. In one-hot encoding, there is no concept of similarity between sequence or structure pieces; they are either identical or dissimilar. In Table IV we summarize the comparison with other models. From the comparison, we can see that using BiLSTM we can still achieve competitive results compared to other models which use various additional features as input, such as physicochemical properties, amino acid composition, and dipeptide compositions. This model even achieved better accuracy compared to another model from Wu et al.[12] which also consider protein structure in classifying the protein.

Although this study only uses the amino acid sequence without any derived features and combines it with the bidirectional sequential model, we can obtain a competitive classification result. Some other studies that use various handcrafted features provide higher performance. So, including the semantics information in the protein representation might increase the performance. In natural language processing, we can extract features using the Language Model (LM), referred to as embeddings. Instead of manually calculating each feature, LMs offer a potential alternative to this increasingly time-consuming database search as they extract features directly from single protein sequences [36]. Using embedding that has semantic information as input for classification can improve classification performance [37]. Use of embedding to represent proteins for classification as a downstream task can be an exciting topic to explore in the future.

IV. Conclusions

This research demonstrates that models trained using simply amino acid sequences perform comparably to and frequently exceed models trained using multiple feature representations. Comparing different machine learning algorithms demonstrates that a sequential model with a bidirectional mechanism is applicable to all protein attributes, and that the position of amino acids in the protein sequence has a significant predictive function. This research facilitates the development of predictive models by bypassing many of the challenges associated with generating the biological, chemical, and physical attributes that describe protein sequences.

Acknowledgment

This work was supported by Indonesian Ministry of Education, Culture, Research and Technology for Fiscal Year 2022.

References

[1] K. Dumorné, D. C. Córdova, M. Astorga-Eló, and P. Renganathan, 'Extremozymes: A Potential Source for Industrial Applications', J. Microbiol. Biotechnol., vol. 27, no. 4, pp. 649–659, Apr. 2017, doi: 10.4014/JMB.1611.11006.

[2] M. De Champdoré, M. Staiano, M. Rossi, and S. D'Auria, 'Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest', J. R. Soc. Interface, vol. 4, no. 13, p. 183, Apr. 2007, doi: 10.1098/RSIF.2006.0174.

[3] S. Tapadar et al., 'Role of Extremophiles and Extremophilic Proteins in Industrial Waste Treatment', Remov. Emerg. Contam. Through Microb. Process., pp. 217–235, 2021, doi: 10.1007/978-981-15-5901-3_11.

[4] D. Zhu, W. A. Adebisi, F. Ahmad, S. Sethupathy, B. Danso, and J. Sun, 'Recent Development of Extremophilic Bacteria and Their Application in Biorefinery', Front. Bioeng. Biotechnol., vol. 8, p. 483, Jun. 2020, doi: 10.3389/FBIOE.2020.00483/BIBTEX.

[5] P. Charoenkwan, N. Schaduangrat, M. M. Hasan, M. A. Moni, P. Lió, and W. Shoombuatong, 'Empirical Comparison and Analysis of Machine Learning-Based Predictors for Predicting and Analyzing of Thermophilic Proteins', EXCLI J., vol. 21, pp. 554–570, 2022, doi: 10.17179/excli2022-4723.

[6] H. Lin and W. Chen, 'Prediction of thermophilic proteins using feature selection technique', J. Microbiol. Methods, vol. 84, no. 1, pp. 67–70, Jan. 2011, doi: 10.1016/J.MIMET.2010.10.013.

[7] S. Kumar, C. J. Tsai, and R. Nussinov, 'Factors enhancing protein thermostability', Protein Eng., vol. 13, no. 3, pp. 179–191, 2000, doi: 10.1093/PROTEIN/13.3.179.

[8] M. M. Gromiha, 'Important inter-residue contacts for enhancing the thermal stability of thermophilic proteins', Biophys. Chem., vol. 91, no. 1, pp. 71–77, Jun. 2001, doi: 10.1016/S0301-4622(01)00154-5.

[9] X. X. Zhou, Y. B. Wang, Y. J. Pan, and W. F. Li, 'Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins', Amino Acids, vol. 34, no. 1, pp. 25–33, Jan. 2008, doi: 10.1007/S00726-007-0589-X.

[10] G. Zhang and B. Fang, 'Application of amino acid distribution along the sequence for discriminating mesophilic and thermophilic proteins', Process Biochem., vol. 41, no. 8, pp. 1792–1798, Aug. 2006, doi: 10.1016/J.PROCBIO.2006.03.026.

[11] M. M. Gromiha and M. X. Suresh, 'Discrimination of mesophilic and thermophilic proteins using machine learning algorithms', Proteins Struct. Funct. Bioinforma., vol. 70, no. 4, pp. 1274–1279, Mar. 2008, doi: 10.1002/PROT.21616.

[12] L. C. Wu, J. X. Lee, H. Da Huang, B. J. Liu, and J. T. Horng, 'An expert system to predict protein thermostability using decision tree', Expert Syst. Appl., vol. 36, no. 5, pp. 9007–9014, Jul. 2009, doi: 10.1016/J.ESWA.2008.12.020.

[13] Y. C. Zuo, W. Chen, G. L. Fan, and Q. Z. Li, 'A similarity distance of diversity measure for discriminating mesophilic and thermophilic proteins', Amino Acids, vol. 44, no. 2, pp. 573–580, 2013, doi: 10.1007/s00726-012-1374-z.

[14] G. L. Fan, Y. L. Liu, and H. Wang, 'Identification of thermophilic proteins by incorporating evolutionary and acid dissociation information into Chou's general pseudo amino acid composition', J. Theor. Biol., vol. 407, pp. 138–142, 2016, doi: 10.1016/j.jtbi.2016.07.010.

[15] C. Feng, Z. Ma, D. Yang, X. Li, J. Zhang, and Y. Li, 'A Method for Prediction of Thermophilic Protein Based on Reduced Amino Acids and Mixed Features', Front. Bioeng. Biotechnol., vol. 8, no. May, pp. 1–10, 2020, doi: 10.3389/fbioe.2020.00285.

[16] S. Nakariyakul, Z. P. Liu, and L. Chen, 'Detecting thermophilic proteins through selecting amino acid and dipeptide composition features', Amino Acids, vol. 42, no. 5, pp. 1947–1953, 2012, doi: 10.1007/s00726-011-0923-1.

[17] D. Wang, L. Yang, Z. Fu, and J. Xia, 'Prediction of Thermophilic Protein with Pseudo Amino Acid Composition: An Approach from Combined Feature Selection and Reduction', Protein Pept. Lett., vol. 18, no. 7, pp. 684–689, 2011, doi: 10.2174/092986611795446085.

[18] G. Zhang and B. Fang, 'LogitBoost classifier for discriminating thermophilic and mesophilic proteins', J. Biotechnol., vol. 127, no. 3, pp. 417–424, Jan. 2007, doi: 10.1016/J.JBIOTEC.2006.07.020.

[19] X.-L. Liu, J.-L. Lu, and X.-H. Hu, 'Predicting thermophilic proteins with pseudo amino acid composition:approached from chaos game representation and principal component analysis', Protein Pept. Lett., vol. 18, no. 12, pp. 1244–1250, Oct. 2011, doi: 10.2174/092986611797642661.

[20] P. Charoenkwan, N. Schaduangrat, M. A. Moni, P. Lio', B. Manavalan, and W. Shoombuatong, 'SAPPHIRE: A stacking-based ensemble learning framework for accurate prediction of thermophilic proteins', Comput. Biol. Med., vol. 146, no. June, p. 105704, 2022, doi: 10.1016/j.compbiomed.2022.105704.

[21] Z. Ahmed et al., 'iThermo: A Sequence-Based Model for Identifying Thermophilic Proteins Using a Multi-Feature Fusion Strategy', Front. Microbiol., vol. 13, p. 82, Feb. 2022, doi: 10.3389/FMICB.2022.790063/BIBTEX.

[22] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, 'CD-HIT Suite: a web server for clustering and comparing biological sequences', Bioinformatics, vol. 26, no. 5, pp. 680–682, Jan. 2010, doi: 10.1093/BIOINFORMATICS/BTQ003.

[23] M. Popescu, V. Balas, L. Perescu-Popescu, and N. Mastorakis, 'Multilayer perceptron and neural networks', WSEAS Trans. Circuits Syst. , vol. 8, no. 7, 2009, doi: 10.5555/1639537.1639542.

[24] M. Popescu, V. Balas, O. Olaru, N. Mastorakis, and O. Olaru, 'The Backpropagation Algorithm Functions for the Multilayer Perceptron', 2009.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, 'Gradient-based learning applied to document recognition', Proc. IEEE, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

[26] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[27] A. Graves and J. Schmidhuber, 'Framewise phoneme classification with bidirectional LSTM networks', Proc. Int. Jt. Conf. Neural Networks, vol. 4, pp. 2047–2052, 2005, doi: 10.1109/IJCNN.2005.1556215.

[28] Y. Kim, 'Convolutional Neural Networks for Sentence Classification', EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf., pp. 1746–1751, Aug. 2014, doi: 10.48550/arxiv.1408.5882.

[29] Z. Guo, P. Wang, Z. Liu, and Y. Zhao, 'Discrimination of Thermophilic Proteins and Non-thermophilic Proteins Using Feature Dimension Reduction', Front. Bioeng. Biotechnol., vol. 8, no. October, pp. 1–10, 2020, doi: 10.3389/fbioe.2020.584807.

[30] J. S. Sunny and L. M. Saleena, 'Amino acid frequency and domain features serve well for random forest based classification of thermophilic and mesophilic protein; a case study on serine proteases', 2021.

[31] H. Tang, R. Z. Cao, W. Wang, T. S. Liu, L. M. Wang, and C. M. He, 'A two-step discriminated method to identify thermophilic proteins', Int. J. Biomath., vol. 10, no. 4, pp. 1–8, 2017, doi: 10.1142/S1793524517500504.

[32] X.-F. Wang, P. Gao, Y.-F. Liu, H.-F. Li, and F. Lu, 'Predicting Thermophilic Proteins by Machine Learning', Curr. Bioinform., vol. 15, no. 5, pp. 493–502, 2020, doi: 10.2174/1574893615666200207094357.

[33] G. Zhang and B. Fang, 'Support vector machine for discrimination of thermophilic and mesophilic proteins based on amino acid composition', Protein Pept. Lett., vol. 13, no. 10, pp. 965–970, Nov. 2006, doi: 10.2174/092986606778777560.

[34] A. Albayrak and U. O. Sezerman, 'Discrimination of Thermophilic and Mesophilic Proteins Using Reduced Amino Acid Alphabets with n-Grams', Curr. Bioinform., vol. 7, no. 2, pp. 152–158, 2012, doi: 10.2174/157489312800604435.

[35] H. Lin and W. Chen, 'Prediction of thermophilic proteins using feature selection technique', J. Microbiol. Methods, vol. 84, no. 1, pp. 67–70, 2011, doi: 10.1016/j.mimet.2010.10.013.

[36] D. Ofer, N. Brandes, and M. Linial, 'The language of proteins: NLP, machine learning & protein sequences', Comput. Struct. Biotechnol. J., vol. 19, pp. 1750–1758, 2021, doi: 10.1016/j.csbj.2021.03.022.

[37] A. Elnaggar et al., 'ProtTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing', bioRxiv, Jul. 2020, Accessed: Mar. 29, 2021. [Online]. Available: http://arxiv.org/abs/2007.06225.

Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis

KaiSiang Chong, Nathar Shah
Faculty of Computing & Informatics
Multimedia University, Cyberjaya, Selangor, Malaysia

***Abstract*—This paper compares the performance of Naive Bayes and SVM classifiers classification based on sentiment analysis of healthcare companies' stock comments in Bursa Malaysia. Differing from other studies which focus on the performance of the classifier models, this paper focuses on identifying the hyperparameters of the classifier models that are significant for sentiment analysis and the optimization potential of the models. Grid Search technique is used for the hyperparameters tuning process. The performance such as precision, recall, f1-score, and accuracy of Naive Bayes and SVM before and after hyperparameter tuning are compared. The results show that the important hyperparameters for Naive Bayes are alpha and fit_prior, while the important hyperparameters for SVM are C, kernel, and gamma. After performing hyperparameters tuning, SVM gave a better performance with an accuracy of 85.65% than Naive Bayes with an accuracy of 68.70%. It also proves that hyperparameter tuning is able to improve the performance of both models, and SVM has a better optimization potential than Naive Bayes.**

***Keywords*—Machine learning; sentiment analysis; opinion mining, Naïve Bayes; SVM Classifier; grid search technique; hyperparameter tuning**

I. Introduction

Sentiment analysis, often known as opinion mining, is a natural language processing (NLP) technique that determines the sentiment behind a body of text. This is a common method for businesses to determine and categorize customer views about a product, service, or idea. Data mining, machine learning (ML), and artificial intelligence (AI) are involved in analyzing the texts and finding out the sentiment.

There are too many ways to perform sentiment analysis by using different machine learning algorithms such as Naive Bayes, Support Vector Machine, K-Nearest Neighbour, and so on. This has made it difficult for the researchers to determine which classifier should be used as the performance of these algorithms is usually dependent on the datasets used. Most studies [7][9][3][6][4], concluded that the Naive Bayes and SVM classifiers outperform all other algorithms in evaluating the sentiment of the text. However, it seems that the performance of Naive Bayes and SVM is very similar. Depending on the datasets used, the performance of the classifiers is affected. In [7] and [9], SVM has a better performance than Naive Bayes. The dataset used in [7] is Amazon product reviews, and in [9] is Twitter reviews. However, in [3], Naive Bayes has a better performance than SVM, and the dataset used in [3] is about e-sport education. Most importantly, these researches only used default hyperparameters for the classification models, and so far the best performance of the classifier obtained is from [7], which is the SVM with 84% of accuracy. The problem of current research is that most of the papers use the default hyperparameter for the sentiment classification. The results might be good, but there should still be some potential for the models to perform better.

Hence, the purpose of this paper is to compare the performance of Naive Bayes and SVM classifiers based on sentiment analysis of healthcare companies' stock comments, justify which model is best suited for this case, and justify the optimization potential of the models by hyperparameter tuning using the Grid Search approach. The data is collected from the I3investor website and preprocessed by using text preprocessing techniques such as removing stopwords, lemmatization, tokenization, and so on. After that, the preprocessed data will be used to train the Naive Bayes and SVM, and the results will be evaluated. Section II will include some background studies of several similar works, Section III will be the methodology which includes the detailed steps of conducting the research, and Section IV will be the evaluation results of both classifications.

II. Background Study

Basically, this section of the paper includes the review of several research papers with similar works to ours. According to these research papers, data preprocessing like stopwords removal, stemming, and tokenization are necessary steps before performing the classification. First, the comparison study of Naive Bayes with SVM is reviewed. It appears that the performance of both models is dependent on the datasets used. Second, comes the review of the Naive Bayes classifier and it seems that Naive Bayes outperforms other classifiers. After that, the study about SVM is reviewed, and it shows that the performance of SVM is affected by the dataset used, and using the Grid Search approach for SVM optimization, the performance of SVM can be improved as well.

A. Comparison Study of Naive Bayes and SVM using Different Datasets

Firstly, this section will review the comparative study of Naive Bayes and SVM.

Sanjay Dey et al. [7] conducted a comparison study of two machine learning algorithms for sentiment analysis of Amazon product reviews. Naive Bayes and SVM were used in this paper. The preprocessing steps such as tokenization, removing stopwords, filling missing values, and feature extraction are applied. The result shows that SVM has a slightly better performance with 84 % accuracy than Naive Bayes with 82.875 % accuracy.

Abdul Mohaimin Rahat et al. [9] worked on a research paper to conduct sentiment analysis on the review from Twitter. The dataset collected is preprocessed by stop word removal, hashtag removal, POS tagging, and so on. Two algorithms, which are Naive Bayes and SVM were applied to classify the positive and negative sentiments. As a result, SVM gets a better accuracy of 82.48 % than the Naive Bayes with an accuracy of 76.56 %.

Rian Ardianto et al. [3] performed sentiment analysis toward e-sport education. The data was collected from Twitter. Naive Bayes and SVM are used in this research as a comparative study. Synthetic Minority Over-Sampling Technique (SMOTE) is used in the evaluation of the two algorithms. As a result, Naive Bayes with SMOTE has a better performance with an accuracy of 70.32 % as compared to SVM with SMOTE with an accuracy of 66.92 %.

B. Study on Naive Bayes

Second, this section will review the research on the Naive Bayes classifier.

The research done by Lopamudra Dey et al. [6] focuses on the comparison of two supervised machine learning approaches which are K-Nearest Neighbour and Naive Bayes based on the sentiment analysis of movie reviews as well as hotel reviews. The accuracy, precision, and recall of these models are evaluated. In short, Naive Bayes has a better performance for movie reviews with an accuracy of 82.43 % than K-NN with an accuracy of 69.81 %, while having a similar performance for hotel reviews with an accuracy of 55.09 % as compared to K-NN with an accuracy of 52.14 %. The researchers concluded that Naive Bayes performs better than K-NN in analyzing the movie reviews.

Achmad Bayhaqy et al. [4] focused on comparing three different classification algorithms, which are Decision Tree, K-NN, and Naive Bayes, by the sentiment analysis about the tweets/reviews of E-commerce in Tokopedia and Bukalapak on Twitter. Text preprocessing techniques are applied to the data collected. The comparison of the three algorithms is done with the assistance of Rapidminer. The results show that the accuracy of the Decision Tree is 80%, K-NN is 78%, and Naive Bayes is 77%. The results for precision for Decision Tree is 79.96%, K-NN is 85.67 %, and Naive Bayes is 88.50 %. Although the accuracy of Naive Bayes is 77 % which is the lowest among others, the researchers concluded the Naive Bayes as the most suitable classifiers for use with their datasets as it has the highest precision of 88.50 % which means it provided more accurate and precise predictions.

C. Study on Support Vector Machine

Lastly, this section will review the research on SVM classification.

Munir Ahmad et al. [1] have chosen to use SVM for the sentiment analysis with WEKA. There are two datasets included which are the tweets about self-driving cars and Apple products, and the data are pre-labeled with the sentiments. In short, the accuracy for the self-driving cars dataset is 59.91 %, and the accuracy for the Apple products dataset is 71.2 %. The outcomes are not very good, demonstrating the dependency of SVM performance on the input dataset. The habits of most Twitter users to use short forms or informal language might be the reason for the difficulty for the SVM to learn successfully.

Besides, using the Grid Search approach for SVM optimization, Munir Ahmad et al. [2] have achieved better results. The precision of SVM is increased from around 70% to 80%. With the Twitter data about the topics of Apple, Google, Microsoft, and Twitter, the potential of SVM optimization is highlighted in this paper.

III. METHODOLOGY

I3investor is a popular stock investment platform for independent stock traders and investors. Every month, the I3investor [5] community creates over 50K comments and posts. In order to be trained by the supervised learning algorithms, the datasets collected from I3investor needed to be preprocessed and labeled. After the preprocessing and labeling, a portion of data is selected from the dataset and is split into train-set and test-set. The words are vectorized by using the TF-IDF vectorizer, and the dataset is used in running both Naive Bayes and SVM classifiers. Moreover, the Grid Search technique is used for hyperparameter tuning to improve the accuracy of both classifiers. The experimental results are then evaluated. Fig. 1 shows the workflow of the methodology. The details of the process will be explained.

A. Data Pre-processing

Every comment may include some words that are neither significant nor beneficial for sentiment analysis. Hence, text preprocessing is a necessary step to obtain a clean dataset, and have better outcomes. The preprocessing steps included:

1) Removing of URL: The URL in the comments which basically links users to other websites is meaningless for sentiment analysis and is removed.

2) Removing of Other Languages: The data collected will include the comments from Malaysians which means there will be several languages such as Chinese, Malay, and English. Hence, the CLD3 package is used in this case to detect and remove the Chinese and Malay comments, only remain the English comments. Since the classifiers are not trained to assess the sentiment of comments in multiple different languages, the removal of other languages will then have a significant influence on the classification process outcomes.

3) Removing of Punctuation: Punctuation has no value for the sentiment analysis and is removed. It is also a step needed for the ease of tokenization.

4) Removing of Stopwords: Stop words are function words that have no sentiment yet are regularly used. If these terms are not eliminated, they will have no effect on the analysis's efficiency. These words are known as "noise." For example, frequently used terms are "a," "of," "the," "I," "it," "you," and "and."

5) Lemmatization: This stage condenses words into their stem or root forms. For example, “evaluate” and “evaluation”, the root of the word “evaluation” is “evaluate”, and having both terms in the data increases the algorithm's effort to interpret their sentiment. As a result, lemmatizing the token to its root type is required to minimize the complexity of the comment and reduce processing time, hence enhancing the model's performance.

6) Lower casing text: All the text in the datasets is changed to lower case to have a consistent format.

7) Tokenization: A method to divide the entire comment into many individual words for convenience of analysis.

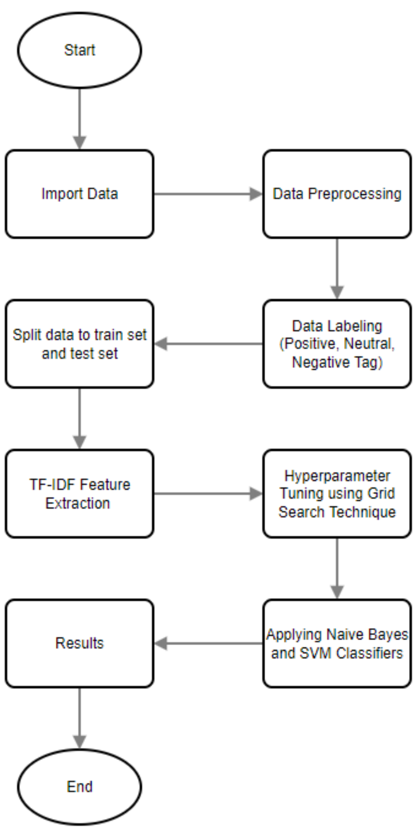


Fig. 1. Methodology workflow.

B. Data Labeling

It is impossible for a human being to manually label the data as the datasets consist of a large number of comments. Hence, the Opinion Lexicon created by Minqing Hu and Bing Liu [8] which contains positive and negative words is being prepared. The preprocessed data is next reviewed to see if the words match those in the positive or negative Opinion Lexicon. The number of positive and negative words for each row of data will next be calculated. The score for each row of data will be computed by subtracting the number of negative words from the number of positive words. As a consequence, data with scores more than 0 will be labeled as positive, data with scores less than 0 will be labeled as negative, and data with scores equal to 0 will be labeled as neutral.

C. Feature Extraction using TF-IDF

The term frequency-inverse document frequency (TF-IDF) was used to extract the feature of the dataset. When retrieving information, the TF-IDF technique weights the frequency of a phrase (TF) and the inverse frequency of documents (IDF). Each word or phrase is given a TF and IDF score. The TF and IDF product results of a word, on the other hand, correspond to the phrase's TF-IDF weight. As a result, the TF-IDF score (weight) rises in tandem with the phrase's rarity and vice versa. As a consequence, the TF of a term denotes its frequency, whereas the IDF denotes its importance across the corpus. If a term's content TFIDF weight is high, the content will always show among the top search results, allowing anybody to avoid stopwords while simultaneously finding words with higher search traffic and lower competition.

D. Hyperparameters Tuning using Grid-Search Technique

Hyperparameters are variables whose values influence the learning process and affect the model parameters that a learning algorithm learns. Grid Search is a technique for optimizing hyperparameters. It prepares the machine learning algorithm for every potential combination of hyperparameters. Cross-validation is used to guide the training process, ensuring that the trained model can extract the majority of the patterns from the dataset. The best set of hyperparameter values from Grid Search is then used in the real model. In summary, the optimal hyperparameters are assured, and the model's accuracy can be enhanced.

IV. Experimental Result

The research is conducted using the Google Colab environment. A portion of the preprocessed data which consists of 20000 comments is used. There are 6219 positive comments, 6196 negative comments, and 7585 neutral comments.

Grid Search approach is used to find out the best hyperparameters of the models. Both original and tuned versions of Naive Bayes and SVM are trained and tested. The precision, recall, f1-score, and accuracy of each model are evaluated.

Table I shows the hyperparameters setting for Naive Bayes. There are three parameters which are alpha, fit_prior, and class_prior. Parameter alpha refers to the additive smoothing parameter, parameter fit_prior control whether to learn class prior probabilities or not, and parameter class_prior refers to the prior probabilities of the classes. The default hyperparameter for alpha is 1.0, and for fit_prior is ‘True.’ After performing the Grid Search, it appears that the best hyperparameter for alpha is 1.4, for fit_prior is ‘False’, and the parameter for class_prior is ‘remain unchanged.’

TABLE I. HYPERPARAMETERS SETTING FOR NAIVE BAYES

Parameters	alpha	fit_prior	class_prior
Default	1.0	True	None
Best	1.4	False	None

TABLE II. HYPERPARAMETERS SETTING FOR SVM

Parameters	C	kernel	gamma
Default	1.0	rbf	scale
Best	7.0	linear	auto

Table II shows the hyperparameters setting for SVM. Basically, SVM has a total of 15 hyperparameters. After performing Grid Search, there are only 3 hyperparameters that have changed which are C from a value of 1.0 to a value of 7.0, the kernel of 'rbf' to kernel of 'linear', and gamma of 'scale' to gamma of 'auto' while the other 12 hyperparameters showing default is the best option to choose. Parameter C refers to the regularization parameter, parameter kernel specified the kernel type to be used, and parameter gamma refers to the coefficient of the kernel. With the other hyperparameters for SVM remaining unchanged, this proves that SVM is already a good model for performing sentiment analysis without tuning the hyperparameter and can usually obtain good performance as in papers [7], [9], [3], and [1].

Precision is a metric used to quantify how many correct positive predictions have been made. It is derived by dividing the number of accurately predicted positive cases by the total number of positive examples predicted. The precision shows the model's accuracy in classifying samples as positive.

Table III shows the comparison of precision for each model before and after hyperparameters tuning. The precision of Naive Bayes has increased from 71.63% to 83.22%, and the precision of SVM has increased from 81.64% to 87.60%. In short, SVM has higher precision than Naive Bayes before and after tuning. Fig. 2 shows the bar chart of precision comparison for Naive Bayes and SVM.

The proportion of valid positive predictions made out of all feasible positive predictions is calculated as recall. The recall metric evaluates the model's ability to detect positive samples. The higher the recall, the more positive samples are discovered.

Table IV shows the comparison of recall for each model before and after hyperparameters tuning. The recall of Naive Bayes has decreased from 79.85% to 75.36%, and the recall of SVM has increased from 86.56% to 88.47%. SVM has a higher recall than Naive Bayes after tuning. Fig. 3 shows the bar chart of recall comparison for Naive Bayes and SVM.

TABLE III. PRECISION COMPARISON FOR NAIVE BAYES AND SVM

Model	Without Tuning	Tuned
Naive Bayes	71.63	83.22
SVM	81.64	87.60

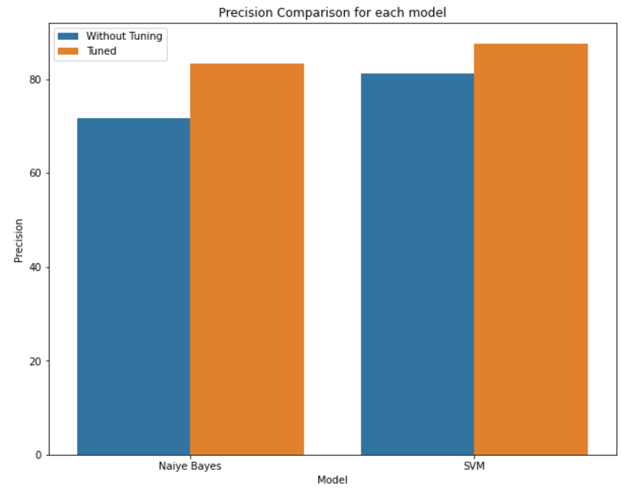


Fig. 2. Bar chart of precision comparison for Naive Bayes and SVM.

TABLE IV. RECALL COMPARISON FOR NAIVE BAYES AND SVM

Model	Without Tuning	Tuned
Naive Bayes	79.85	75.36
SVM	86.56	88.47

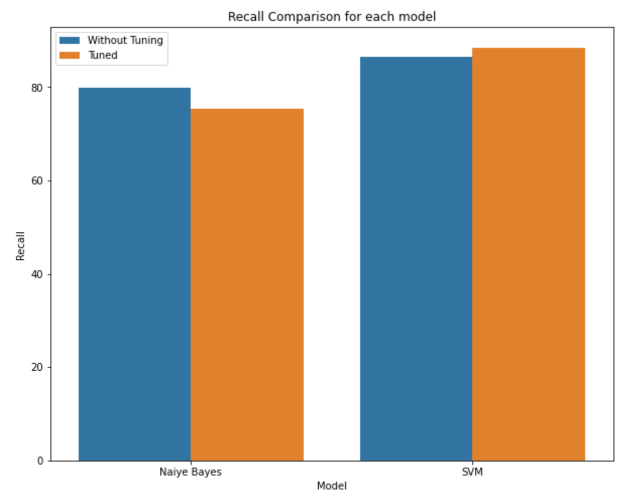


Fig. 3. Bar chart of recall comparison for Naive Bayes and SVM.

The f1-score is a method for combining precision and recall into a single metric that combines both characteristics. We might have good precision with poor recall or vice versa. The f1-score allows you to convey both concerns with a single score.

Table V shows the comparison of f1-score for each model before and after hyperparameters tuning. The f1-score of Naive Bayes has increased from 75.74% to 79.29%, and the f1-score of SVM has increased from 83.90% to 88.04%. SVM has a higher f1-score than Naive Bayes after tuning. Fig. 4 shows the bar chart of the f1-score comparison for Naive Bayes and SVM.

TABLE V. F1-SCORE COMPARISON FOR NAIVE BAYES AND SVM

Model	Without Tuning	Tuned
Naive Bayes	75.74	79.29
SVM	83.90	88.04

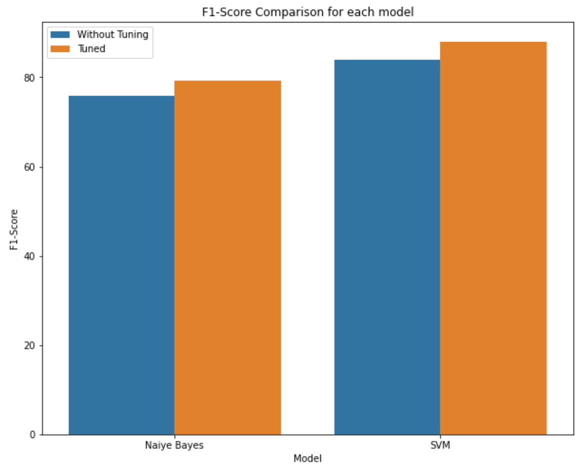


Fig. 4. Bar chart of F1-score comparison for Naive Bayes and SVM.

A model's accuracy is a metric that assesses how well it performs in all classes. This is advantageous when all of the classes are equally important. The ratio between the number of right predictions and the total number of predictions is used to evaluate it.

TABLE VI. ACCURACY COMPARISON FOR NAIVE BAYES AND SVM

Model	Without Tuning	Tuned
Naive Bayes	67.65	68.70
SVM	81.73	85.65

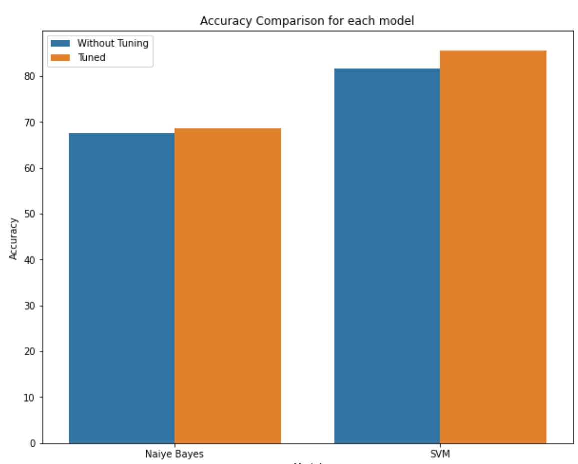


Fig. 5. Bar chart of accuracy comparison for Naive Bayes and SVM.

Table VI shows the comparison of accuracy for each model before and after hyperparameters tuning. The accuracy of Naive Bayes has increased from 67.65% to 68.70%, and the accuracy of SVM has increased from 81.73% to 85.65%. SVM has higher accuracy than Naive Bayes after tuning. Fig. 5 shows the bar chart of accuracy comparison for Naive Bayes and SVM.

In short, the hyperparameters of Naive Bayes that have a significant effect on sentiment analysis are “alpha” and “fit_prior”, while the hyperparameters of SVM that have a significant effect on sentiment analysis are “C”, “kernel”, and “gamma”. SVM has a better performance than Naive Bayes before and after hyperparameters tuning. It appears that SVM has a better potential for optimization with an increase in accuracy of about 4% than the Naive Bayes with an increase in accuracy of about 1%.

V. CONCLUSION

In conclusion, the research has done a comparative study for Naive Bayes and SVM and found out that SVM has a better performance than Naive Bayes based on sentiment analysis of healthcare companies' stock comments. Grid Search approach is used for hyperparameter tuning and is able to identify the hyperparameters of both models that are significant for sentiment analysis. The research is able to prove that hyperparameters tuning can increase the model’s accuracy, and SVM has a better potential for optimization as compared to Naive Bayes. There are still many things to be improved in the future such as adding more datasets, using different classifiers, and using different hyperparameter tuning techniques.

REFERENCES

[1] Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment analysis of tweets using svm. Int. J. Comput. Appl, 177(5), 25-29.

[2] Ahmad, M., Aftab, S., Bashir, M. S., Hameed, N., Ali, I., & Nawaz, Z. (2018). SVM optimization for sentiment analysis. Int. J. Adv. Comput. Sci. Appl, 9(4), 393-398.

[3] Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., & Gata, W. (2020). Sentiment analysis on E-sports for education curriculum using naive Bayes and support vector machine. Jurnal Ilmu Komputer dan Informasi, 13(2), 109-122.

[4] Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018, October). Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes. In 2018 international conference on orange technologies (ICOT) (pp. 1-6). IEEE.

[5] Bursa Malaysia (KLSE) market summary. Bursa Malaysia (KLSE) Market Summary. (n.d.). Retrieved April 2, 2022, from https://klse.i3investor.com/web/index.

[6] Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982.

[7] Dey, S., Wasif, S., Tonmoy, D. S., Sultana, S., Sarkar, J., & Dey, M. (2020, February). A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews. In 2020 International Conference on Contemporary Computing and Applications (IC3A) (pp. 217-220). IEEE.

[8] Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA.

[9] Rahat, A. M., Kahir, A., & Masum, A. K. M. (2019, November). Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (pp. 266-270). IEEE.

Intrusion Detection System using Long Short Term Memory Classification, Artificial Raindrop Algorithm and Harmony Search Algorithm

Meghana G Raj[1], Dr. Santosh Kumar Pani[2]
Assistant Professor[1], Professor[2]
Department School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to be University), Bhubaneswar, India

***Abstract*—Nowadays, various technological advancements in Intrusion Detection Systems (IDS) detects the malicious attacks and reinstate network security in the cloud platform. Cloud based IDS designed with hybrid elements combining Machine Learning and Computational Intelligence algorithms have been shown to perform better on parameters, such as Detection Rate, Accuracy, and the False Positive Rate. Machine Learning algorithms provide effective techniques for classification and prediction of network attacks, by analyzing existing IDS datasets. The main challenge is selection of appropriate data dimensions to be used for detection of attacks, out of the high number of data dimensions available. For the selected data dimensions, Computational Intelligence Algorithms provide effective techniques for hyper-parameter tuning, by optimizing on reiterative basis. The main challenge is selection of appropriate algorithm which offers optimal performance results. In this research, Hybrid Meta-heuristic approach, which combines a Long Short Term Memory (LSTM) classification model in dimension selection, with the application of Artificial Raindrop Algorithm- Harmony Search Algorithm (ARA-HSA) for hyper-parameter tuning, in order to achieve a high performance IDS in cloud environment. The performance validation of the hybrid LSTM-ARA-HSA algorithm has been carried out using a benchmark IDS data set and the comparative results for this algorithm along with other recent hybrid approaches has been presented.**

Keywords—Artificial raindrop algorithm; cloud computing; harmony search algorithm; hybrid meta-heuristic algorithms; hyper-parameter tuning; intrusion detection system; long short term memory classification model

I. INTRODUCTION

The Cloud Computing (CC) platform, which is on-demand network access for several pattern of computing asset, such as servers, applications, networks, and services. Security is the major difficulty for the organizations to accept the cloud enabled solutions. Due to cloud infrastructure (open and fully distributed), which makes it vulnerable to attacks and threats. Thus, it creates incentives for intruders for initiating attack focused devices to permit the data stored in cloud. The threats are confidentiality, availability, and integrity of cloud resources and services. To solve the security challenges, the IDS should be integrated within the cloud environment [1].

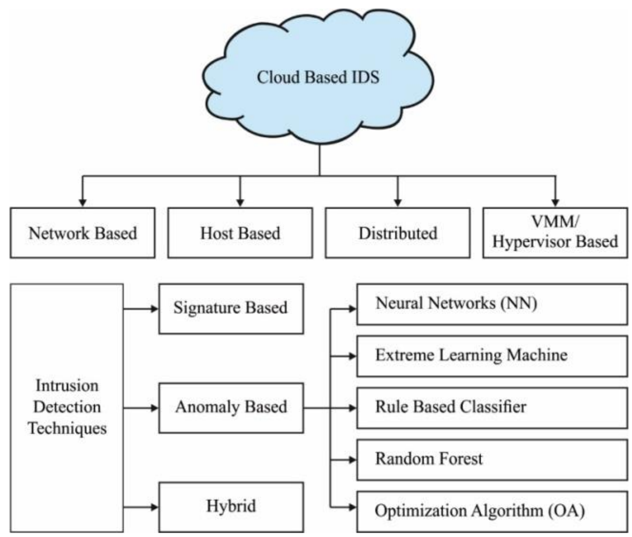


Fig. 1. Types of cloud based IDS

There are several kinds of attacks that could probably arise in a network, such as User to Root (U2R), Remote to Local (R2L), and the Denial of Service (DoS) attacks. In U2R attacks, a local user who is the attacker gets unauthorized access to the routing privilege and access control list [2]. In R2L attacks, the attacker is not a local user but remotely forwarded a set of packets to server or computer over the network, to attempt unauthorized access [3]. In DoS attacks, the attacker attempts to stops the standard service in the network by over-whelming the network with high number of request data packets [4].

Machine learning (ML) algorithms offer effective techniques to improve a efficiency and effectiveness of IDS, due to the capability to identify potential attacks through sophisticated classification of network states described by data dimensions, based on comparison with known network states during previous network attacks [5]. Several ML techniques [6], such as Extreme Learning Machine, Random Forest Classifiers, Multi-layer Perceptron Network etc. have been designed and their performance in detecting attacks evaluated [7]. Fig. 1 provides an overview of types of IDS, based on the different ML algorithms adopted.

The main challenges in design of IDS are:

- Selection of appropriate number data dimensions to be used for detection of attacks, out of the high number of data dimensions available
- Hyer-parameter tuning for the selected data dimensions
- Improving performance in classification

The challenge of feature selection is sought to be overcome by adoption of appropriate ML algorithms. The challenge of hyper-parameter tuning is sought to be overcome by adoption of appropriate Computational Intelligence algorithms. The challenge of improved performance is sought to be overcome by adoption of appropriate ML algorithms. The proposed solution therefore, consists of a hybrid approach, in which both ML and Computational Intelligence algorithms incorporated in order to deliver improved performance compared to existing approaches in the literature.

This paper presents a Hybrid Meta-heuristic approach which uses ML techniques- ExtraTrees Feature Selection Method and LSTM classification model- along with Computational Intelligence Algorithms- Artificial Raindrop Algorithm and Harmony Search Algorithm. For feature selection, ExtraTrees (ET) method is used to choose a proper set of data dimensions to be employed for classification [8][9] and prediction [10] of network attacks, out of maximal number of data dimensions available in the standard IDS data set of known attacks from the past. Artificial Raindrop Algorithm (ARA) and Harmony Search Algorithm (HSA) are used for hyper-parameter tuning (i.e. optimizing the parameter values to be used) for the data dimensions selected. Finally, classification is achieved using the LSTM classification model.

To estimate the resulting performance of IDS which uses hybrid combination of LSTM-ARA-HSA approach, the paper presents the results of simulations using the benchmark data set.

In summary, the paper's contributions can be summarized as follows.

- Use an ML technique- ExtraTrees Feature Selection method for selecting optimal set of data dimensions for classification and prediction of network attacks, out of the high number of data dimensions available for analysis.
    - Perform min-max based data normalization to achieve standardization of the different scales for the available data dimensions.
    - Use of ET Feature Selection method for the optimum selection of data dimensions, to enhance the efficiency of IDS.
- Adopt combination of ARA-HSA algorithms for hyper-parameter optimization of the data dimensions selected.
- Perform final classification using the LSTM classification model.
- Design a Hybrid Meta-heuristic approach for IDS using a combination of LSTM-ARA-HSA, as described above.
- Test the performance of Hybrid Meta-heuristic LSTM-ARA-HSA, using simulations on the benchmark IDS data set.

The paper is organized as follows. Section 2 illustrates a brief literature review of some of recent Hybrid Meta-heuristic approaches used by IDS. Section 3 provides an overview of LSTM-ARA-HSA methodologies. Section 4 discusses an overview of the results of testing a performance valuation of proposed hybrid approach. Section 5 lists the key findings and conclusions of this paper.

II. Literature Review

Review of recently developed IDS approaches for cloud environments is described below.

Sethi *et al.* [11] presented a DRL enabled adoptive cloud IDS model that implements fine-grained classification and precise detection of complex and new attacks.

Ji *et al.* [12] introduced a network IDS by integrating asymmetric convolution AE and RF. This method could integrate the benefits of Shallow and Deep Learning.

Singh and Ranga [13] deliberated a robust network driven IDS employing ensemble-based ML method with four classifications, namely voting scheme, RUSBooted, bagged tree, boosted tree, and subspace discriminant. The voting method was integrated with architecture for obtaining the final prediction.

Alkadi *et al.* [14] devised DBF method for offering privacy with blockchain and security-based distributed IDS by the smart contracts of IoT networks. IDS were utilized using BiLSTM-DL technique for managing the consecutive network data, which is computed by the datasets of BoT-IoT and UNSW-NB15.

Zhong *et al.* [15] presented a novel method using features of model. The method would gather features in network layer through tcpdump packet and a application layer through system routine. GRU and Text-CNN approaches are selected since they could process consecutive data as a language system.

Samriya and Kumar [16] proposed a hybridization method for IDS for improving entire security level of cloud computing platforms. Additionally, this technique assists in handling different kinds of security problems i.e., detection of fake identity and data leakage Phishing attacks to retain the security over cloud environment.

Abusitta *et al.* [17] presented an ML based IDS that effectively uses the previous feedback information to improve a capacity for proactive decision making. Especially, presented method was depended by Denoising Autoencoder (DA) that is employed as an element to create DNN method.

III. PROPOSED MODEL

In this research, use of a hybrid meta-heuristic approach-combining Machine Learning techniques, named ExtraTrees Feature Selection method and LSTM classification model with Computational Intelligence Algorithms named Artificial Raindrop Algorithm and Harmony Search Algorithm- is proposed for optimal classification and prediction of network attacks in the cloud environment.

The overall model encompasses the following processes:

1) Pre-processing of all available data dimensions in the data-set

2) Selection of appropriate data dimensions using ExtraTrees feature selection method

3) Hyper-parameter optimization for the selected data dimensions using Artificial Raindrop Algorithm and Harmony Search Algorithm (ARA-HSA)

4) Final LSTM based classification by IDS into Attack/ Not an Attack, based on above processes

Fig. 2 illustrates the overall working process of LSTM-ARA-HSA technique for IDS. The detailed working of each of the above processes is elaborated in the succeeding sections.

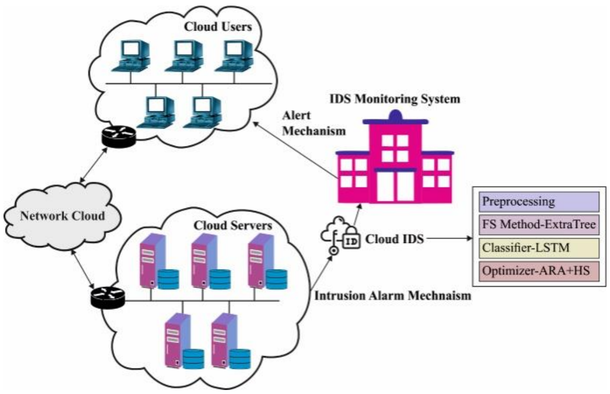


Fig. 2. Overall process of LSTM-ARA-HSA technique

A. *Pre-Processing of All Available Data Dimensions*

During this process, Min_max normalization was utilized to standardize the different scales for all the available data dimensions in the benchmark IDS data set. The normalization process transforms the data points with reference to the data range to which they belong, by implementing linear transformation. The value of each data point in a dimension was normalized to a standard range between zero and one, utilizing Min_max normalization as per the following formula [18]:

$$t = \frac{v - \min_d}{\max_d - \min_d}(tran\_\max_d - tran\_\min_d) + tran\_\min_d \quad (1)$$

where t refers to the transformed value of data value v from dimension d, $\min_d$ refers to the existing minimum value for the data range for dimension d, $\max_d$ refers to the existing maximum value for the data range for dimension d, $tran\_\min_d$ refers to the standardized minimum value (i.e. zero) for the transformed data range for dimension d, and $tran\_\max_d$ refers to the standardized maximum value (i.e. one) for the transformed data range for dimension ***d.***

B. *Selection of Appropriate Data Dimensions using ExtraTrees Feature Selection Method*

ExtraTrees (ET) is ML type technique that aggregates several de-correlated decision trees together in “forest” to output it’s classified results, presented by Geurts et al. [19]. ET is used during this process for the purpose of feature selection, i.e. choosing a sub-set of proper dimensional data from the full set of data dimensions.

Each Decision Tree (DT) in ExtraTrees Forest is created from raw training sample, and every node, every tree is given by random sample of the number of features to select from full set- say “k” features from full features-set where every tree must choose optimal feature to divide the data using defined mathematical criteria. Thus, this random feature led to the generation of several de-correlated decision trees. To fulfil the feature selection using process through this forest structure, the defined total minimization in the defined mathematical factors (such as the Gini coefficient) applied to the decision of feature of split is calculated for every feature throughout the forest construction [20].

The benefits of ET model are reduced variance of the DT and computation effectiveness. For each of the available data dimensions, the standardized reduction of the Gini coefficient utilized for splitting the feature decision is estimated. Afterward, the Gini coefficient is ranked in descending order, and the first k features can be chosen. In this way, the number of the sub-set of data dimensions- specified as k in number- can be selected out of the full set of available data dimensions in the benchmark IDS data set.

C. *Hyper-Parameter Optimization for the Selected Data Dimensions using Artificial Raindrop Algorithm and Harmony Search Algorithm (ARA-HSA)*

For the hyper-parameter optimization or tuning process, a combination of the ARA-HSA algorithms is proposed to be utilized.

The natural occurrence of rainfall served as the inspiration for the ARA, which has been proven to be an effective tool for resolving single-objective optimization issues [21]. The fundamental premise is to model the altitude of the raindrop from where it is initiated to a final state when has minimal altitude and hence has its minimal energy state. The state of minimal altitude is considered to be an optimal state or solution. ARA begins with initiation of the primary population of N raindrops as arbitrarily placed N vapors in the search space, with all vapors having a vapor place determined as follows:

$$Vapor_i = (x_i^{(1)}, \cdots, x_i^{(d)}, \cdots, x_i^{(D)}), i = 1,2,\cdots,N. \quad (2)$$

Where N refers to the population size, D refers to the number of dimensions selected for hyper-parameter optimization, and $x_i^{(d)}$ refers to the place of i^{th} vapor from the d^{th} dimension.

Since raindrops are created by constantly absorbing ambient water vapor naturally, the place of each raindrop is determined as follows:

$$Raindrop = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^{(1)}, \cdots, \frac{1}{N}\sum_{i=1}^{N} x_i^{(d)}, \cdots, \frac{1}{N}\sum_{i=1}^{N} x_i^{(D)}\right) \quad (3)$$

If the effect of external factors is not taken into account, each raindrop falls from its initial altitude in the cloud, to the ground with free-fall due to gravity. This can be modeled as place of each Raindrop being modified to a new place, represented as $New_$Raindrop. Assuming that Raindro$p^{(d_i)}$ is the place of a Raindrop from $d_i th$ dimension, where $d_i (i = 1,2,3,4)$belongs to the set $\{1, 2, \cdots, D\}$,New$_Raindrop^{(d_1)}$ is attained by linear combination of Raindro$p^{(d_2)}$, $Raindrop^{(d_3)}$ and Raindro$p^{(d_4)}$, and determined as follows:

$$\begin{cases} New\_Raindrop^{(d)} = Raindrop^{(d_2)} + \varphi \cdot (Raindrop^{(d_3)} - Raindrop^{(d_4)}), if\ d = d_1; \\ New\_Raindrop^{(d)} = Raindrop^{(d)}, otherwise. \end{cases} \quad (4)$$

where φ refers the arbitrary number from the range -1 and 1 , $d = 1,2, \cdots, D$. Once the $New\_$ Raindrops contact the ground, it is split to small raindrops due to speed as well as quality. Afterward, these smaller raindrops ($Small - Raindrop_i, i = 1,2, \cdots, N$) may move randomly to any direction. For this reason, the place of $Small\_Raindrop_i$ is modeled as follows:

$$Small_{Raindrop_i} = New_{Raindrop} + sign(\alpha - 0.5) \cdot \log(\beta) \cdot (New\_Raindrop - Vapor_k) (5)$$

where$k$refers to the arbitrarily selected index in the set $\{1, 2, \cdots, N\}$, α and β refer to two uniformly distributed arbitrary numbers from the range zero to one, and the sign $(\cdot)$ refers to the sign function [22].

During the act of gravity, the movement of $Small\_Raindrop_i (i = 1,2, \cdots, N)$ is from higher altitude to lower altitude, and they finally stop at places with minimal altitude (i.e., an optimum solution). The Raindrops Pool (RP) is modeled to track raindrops to the places of minimal altitude as follows:

1) RP is initiated randomly to initial places in the search space;

2) A better solution for the RP is rationalized at the end of every iteration;

3) Once the RP size increases the threshold specified initially, any less optimal solution is deleted from the RP, to maintain the RP's size stable and lower the computation required.

The movement of raindrop d_i for $Small\_Raindrop_i (i = 1,2, \cdots, N)$ is modeled as dependent upon linear group of 2 vectors $d1_i$and $d2_i$, where d_i, $d1_i$ and $d2_i$ are explained as:

$$d1_i = sign(F(RP_{k_1}) - F(Small\_Raindrop_i)) \cdot (RP_{k_1} - Small\_Raindrop) \quad (6)$$

$$d2_i = sign(F(RP_{k_2}) - F(Small\_Raindrop_i)) \cdot (RP_{k_2} - Small\_Raindrop) \quad (7)$$

$$d_i = \tau_1 \cdot rand1_i \cdot d1_i + \tau_2 \cdot rand2_i \cdot d2_i \quad (8)$$

where RP_{k_1} and RP_{k_2} refer to two candidate solutions from RP ($k_1, k_2 \in \{1,2, \cdots, |RP|\}$), τ_1 and τ_2 refer to two step parameters of $Small\_Raindrop_i$ flowing, $rand1_i$ and $rand2_i$ refer to two uniformly distributed arbitrary numbers from the range zero to one, F indicates the Fitness Function. Thus the outcome, $New\_Small\_Raindrop_i (i = 1,2, \cdots, N)$ is determined as:

$$New\_Small\_Raindrop_i = Small\_Raindrop_i + d_i \quad (9)$$

For improving the computational efficiency and convergence rate of ARA, the N optimum solutions in $New\_Small\_Raindrop \cup Vapor$ were chosen utilizing sort technique, as the next vapor population. The flowchart of ARA is described in Fig. 3 [22].

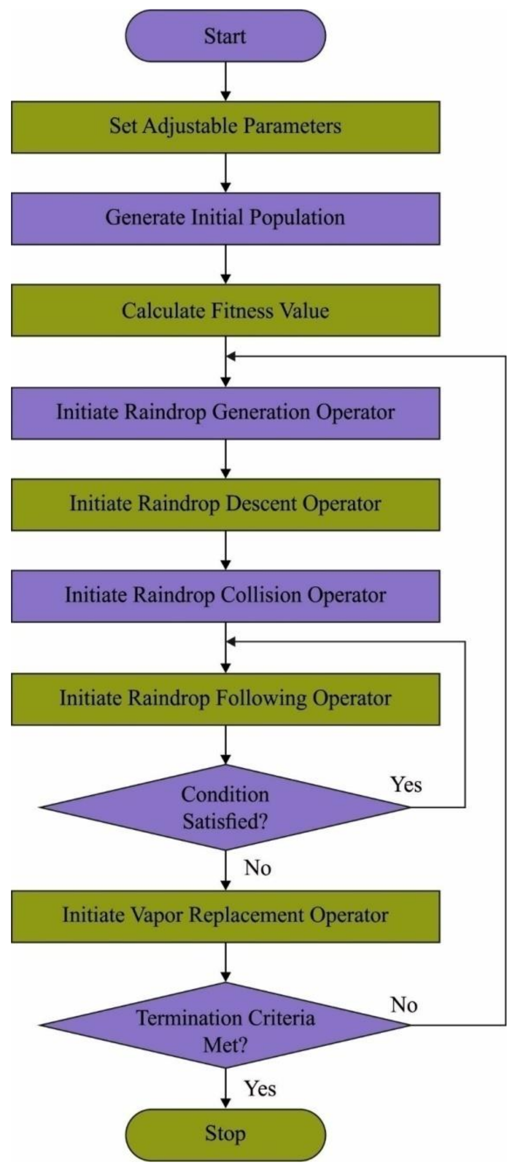


Fig. 3. Flowchart of ARA

The performance of the ARA can be improved by updating every raindrop by the use of the individual model and Harmony Search Algorithm (HAS) based Harmony Search Memory (HM) operator.

HSA attempts several feasible optimal solutions based on use of the HM operator, similar to how multiple musicians attempt to achieve harmony in music based on their collective memory [23]. HSA has been simulated by employing rules of harmony improvisation [24]. It involves the steps as described below.

Step 1. Initializing of HS Memory (HM).

A primary HM has a particular number of arbitrarily created solutions for the optimized solution. For a n dimensional problem, the HM with size of N is modeled as:

$$HM = \begin{bmatrix} x_1^1, x_2^1, \ldots, x_n^1 \\ x_1^2, x_2^2, \ldots, x_n^2 \\ \vdots \\ x_1^{HMS}, x_2^{HMS}, \ldots, x_n^{HMS} \end{bmatrix} \quad (10)$$

where $[x_1^i, x_2^i, \ldots, x_n^i]$ $(i = 1, 2, \ldots, HMS)$ is a solution candidate. Here, the HM was generally fixed between [50, 100].

Step 2. Improvising a solution $[x_1^{'}, x_2^{'}, \ldots, x_n^{'}]$ in HM.

All the elements in the solution $x_j^{'}$ are attained dependent upon HM Considering Rate (HMCR). Here, the HMCR is determined as the possibility of choosing a module in HM member, and the 1-HMCR is, so, the possibility of creating it arbitrarily. Once $x_j^{'}$ appears into the HM, it can be selected in j^{th} dimension of arbitrary HM members and is mutated based on Pitching Adjust Rate (PAR) that defines the probabilities of the candidate in HM is mutated. Moreover, the improvisation of $[x_1^{'}, x_2^{'}, \ldots, x_n^{'}]$ is related to the generation of issues from Genetic Algorithms (GA), and mutation as well as crossover functions. But, while GA generates novel chromosomes utilizing 1 (mutation) or 2 (simple crossover) pre-defined ones, the generation of novel solutions from HSA utilizes every HM member completely.

Step 3. Upgrading the HM. In this step, the Step 2 solution is estimated. Once it creates an optimum fitness which is worse than member from the HM, the existing member will be replaced. Else, the new member is discarded.

Step 4. Repeating Step 2 to Step 3 until a current end circumstance, i.e., a high number of iterations are met. The HSA is an arbitrary search approach and only employs a single search memory for evolving.

In order to optimize the hyper-parameter values for the selected data dimensions, the combination of ARA-HSA is used. The ARA-HSA combination computes a fitness function to accomplish optimization of the hyper-parameters. The value of the fitness function is a positive integer value, which represents the effective outcome of the candidate solution. In case of intrusion detection, the classifier error rate can be treated as the fitness function, as provided in Eq. (11) below. The optimal solutions hold minimum value of error rate and the non-optimal solutions hold maximum value of error rate.

$$fitness(x_i) = Classifier\ Error\ Rate(x_i)$$

$$\frac{number\ of\ misclassified\ instances}{Total\ number\ of\ instances} * 100 \quad (11)$$

D. LSTM based Classification

In this final process, the LSTM model can be utilized for the purpose of detecting and classifying the intrusions in cloud environment. LSTM is a kind of Recurring Neural Network (RNN) i.e., used for processing consecutive data and, it addressing a long-term memory problems of vanilla RNN. The LSTM expands the structure of RNN by a gating method and the standalone memory cell that normalizes the data flow in all over the networks. Here, a gating method includes output, input, and forget gates [25]. This gate controls the flow of data over the network to enables which data needed to continue or the period it would persevere afterward sensing it from the memory. The LSTM network is able to discard the insignificant data and preserves critical data. The memory cell provides a recurrent self-connected unit termed Constant Error Carousel (CEC), which offered a state vector to retain long-term dependency. Fig. 4 depicts the infrastructure of LSTM.

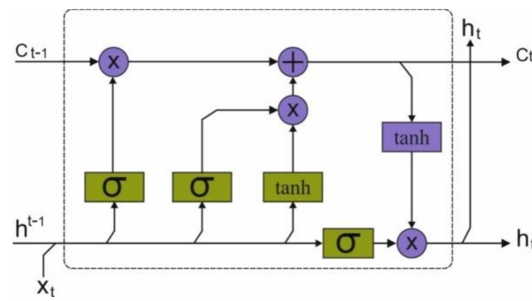


Fig. 4. LSTM structure

Then, to differentiate self-contained cell memory from state h_t in LSTM and, it is represented as c_t. The forget gate f_t obtains input x_t and h_{t-1} to define which data requires to be preserved in c_{t-1}. Activation function for gates i_τ, o_t and f_t are sigmoid layer whereas all the values are predicted among zero and one in which c_{t-1} provides the data preservation to determine the scale. But the above mentioned procedure is determined by Eqs. (12) to (16):

$$i_t = \sigma(W_{xi}x_\tau + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (12)$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f\right) \quad (13)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ c_\tau + b_0) \quad (14)$$

$$c_t = f_\tau \circ c_{t-1} + i_t \circ \emptyset(W_{xc}x_\tau + W_{hc}h_{t-1} + b_c) \quad (15)$$

$$h_t = 0_\tau \circ \emptyset(c_t) \quad (16)$$

Now, $W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{cf} W_{xo}, W_{ho}, W_{co}, W_{xc}$, and W_{hc} indicates weight matrices for the gate and cell memory state, whereas h_{t-1} indicates a preceding hidden state, and c_t denotes a cell state. The bias of the gate is denoted by b_i, b_o, b_f , and b_c whereas 0 indicates an element-wise multiplication process. Likewise, σ displays the logistic sigmoid function and $\emptyset$ signifies tangent function.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (17)$$

$$\emptyset(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad (18)$$

IV. Experimental Validation

The experimental validation of the proposed hybrid meta-heuristic approach LSTM-ARA-HSA technique was done by two benchmark data sets; they are NSL-KDD and KDDCup'99 datasets. The NSL-KDD dataset holds 41 features, in which 2 features are symbolic records and other 39 features are numeric record. The data set contains Basic features: 1–9, Content features: 10–22, Traffic features: 23–31, and Host features: 32–41. The results were analyzed along different aspects. Comparative analysis of the proposed model with other recent techniques is also described below.

A. Results Analysis on NSL-KDD Dataset

This section elaborates the intrusion detection results of the proposed LSTM-ARA-HSA on the test NSL-KDD data set.

Fig. 5 showcases the FS result of the proposed model on the test NSL-KDD dataset. The figure displays the set of features elected by the proposed model.

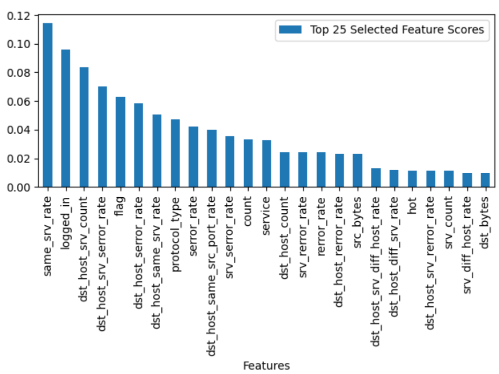


Fig. 5. Selected features and its scores of NSL-KDD dataset

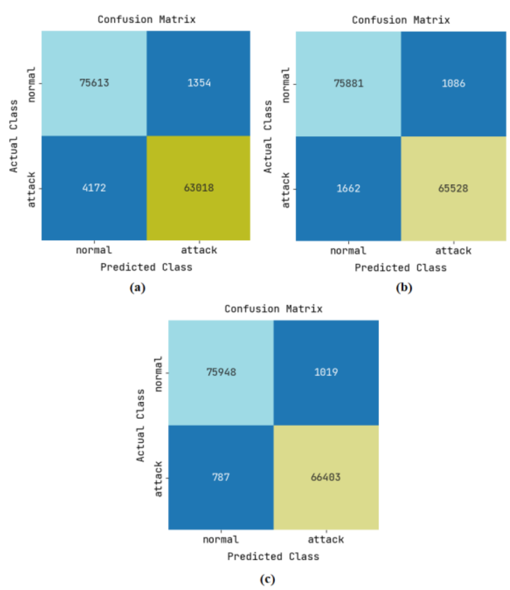


Fig. 6. Confusion matrix of NSL-KDD dataset a) LSTM b) FS+LSTM c) ARA+HS+FS+LSTM

Fig. 6 illustrates the confusion matrix in the proposed model with LSTM and FS+LSTM model. The figure indicates that the LSTM model classified a total of 75613 instances into Normal class and 63018 instances into Attack class. In addition, the FS+LSTM model classified a set of 75881 instances into Normal class and 65528 instances into Attack class. Lastly, the ARA+HS+FS+LSTM model classified a total of 75948 instances into Normal class and 66403 instances into Attack class.

The intrusion detection results of the proposed ARA+HS+FS+LSTM model on the NSL-KDD dataset are offered in Table I. From the table, it is proved that the LSTM model achieved an $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score} of 96.170%, 97.900%, 93.790%, 95.800%, and 96.020% respectively. The FS+LSTM model achieved slightly enhanced outcomes with the $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score} of 98.090%, 98.370%, 97.530%, 97.950%, and 98.060% respectively. Finally, the ARA+HS+FS+LSTM model achieved highest performance with the $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score} of 98.750%, 98.490%, 98.830%, 98.600%, and 98.750% respectively.

The ROC analysis of the ARA+HS+FS+LSTM model with its earlier versions on the test NSL-KDD dataset is shown in Fig. 7. From the figure, it is apparent that LSTM and FS+LSTM models provided reasonable ROC values of 0.9682 and 0.9886. The ARA+HS+FS+LSTM model provided ROC of 0.9875.

TABLE I. Result Analysis of ARA+HS+FS+LSTM Technique on NSL KDD Dataset

Methods	LSTM	FS+LSTM	ARA+HS+FS+LSTM
Accuracy	96.170	98.090	98.750
Precision	97.900	98.370	98.490
Recall	93.790	97.530	98.830
F1-Score	95.800	97.950	98.660
AUC Score	96.020	98.060	98.750

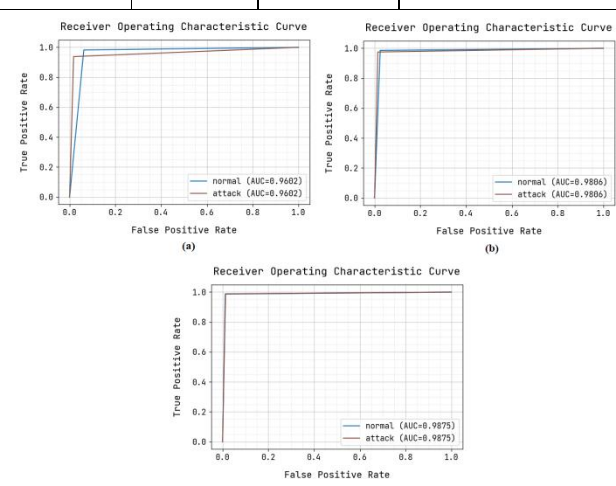


Fig. 7. ROC of NSL-KDD dataset a) LSTM b) FS+LSTM c) ARA+HS+FS+LSTM

A comprehensive comparative result analysis of the ARA+HS+FS+LSTM method with other existing techniques is made in Table II and Fig. 8 [26, 27]. The results indicated that DNN and DT models achieved the least intrusion detection performance. The PCA+DNN and RF models achieved moderately better intrusion detection performance. The XGBoost and Hybrid KPCA-SVM+GA techniques were next in terms of achievement. The Hybrid GA-SVM+PSO technique achieved near optimal outcomes. However, the ARA+HS+FS+LSTM technique achieved the highest performance results compared to other techniques, with the $accu_y, prec_y$, $reca_l$, and $F1_{score}$ of 98.75%, 98.49%, 98.83%< and 98.66%, respectively.

TABLE II. COMPARATIVE ANALYSIS OF ARA+HS+FS+LSTM TECHNIQUE WITH EXISTING APPROACHES ON NSL-KDD DATASET

Methods	Accuracy	Precision	Recall	F1-Score
ARA+HS+FS+LSTM	98.75	98.49	98.83	98.66
PCA+DNN	93.80	93.40	91.80	93.70
DNN Model	91.40	89.10	88.20	90.50
Random Forest	93.60	90.00	82.00	94.60
XGBoost	95.50	92.00	98.00	95.55
Decision Tree	92.89	90.20	85.00	94.50
Hybrid GA-SVM+PSO	96.38	94.90	97.77	96.11
Hybrid KPCA-SVM+GA	95.26	94.03	96.39	95.47

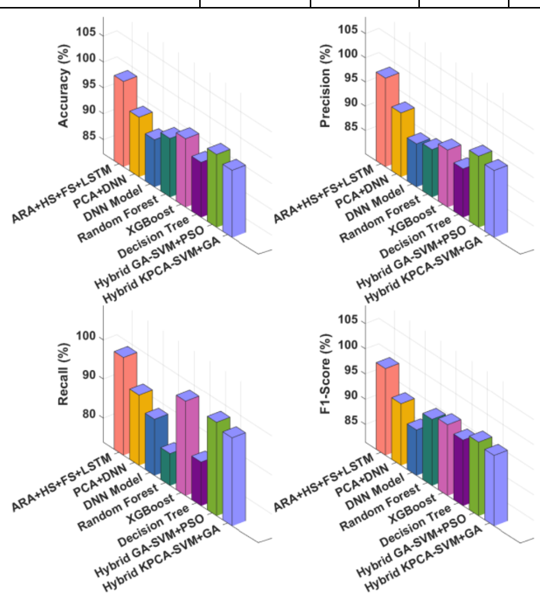


Fig. 8. Comparative analysis of proposed method on NSL-KDD dataset

B. Results Analysis on KDDCup Dataset

This section elaborates the intrusion detection results of the proposed LSTM-ARA-HSA on the test KDDCUP99 data set.

Fig. 9 depicts the FS results of the proposed technique in test KDDCUP99 dataset. The figure indicates the set of features chosen by the presented method.

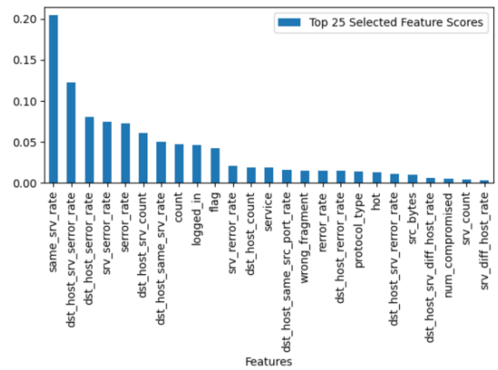


Fig. 9. Selected features and its scores of KDDCup99 dataset

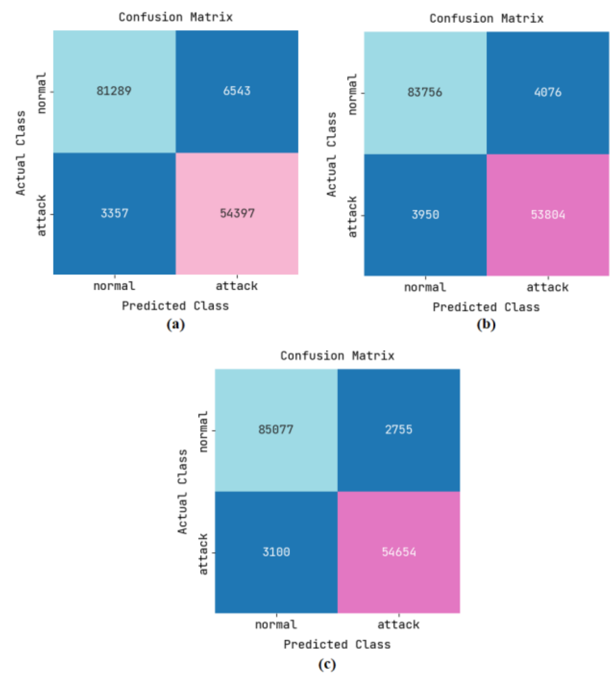


Fig. 10. Confusion Matrix of KDDCup 99 dataset a) LSTM b) FS+LSTM c) ARA+HS+FS+LSTM

Fig. 10 demonstrates the confusion matrix of the proposed algorithm with LSTM and FS+LSTM method. The figure showcased that the LSTM model has identified a total of 81289 instances into Normal class and 54397 instances into Attack class. Also, the FS+LSTM model has categorized a set of 83756 instances into normal class and 53804 instances into Attack class. At last, the ARA+HS+FS+LSTM model has classified a total of 85077 instances into Normal class and 54654 instances into attack class.

The intrusion detection outcomes of the proposed ARA+HS+FS+LSTM approach on the KDDCUP99 dataset is indicated in Table III. From the table, it is apparent that the LSTM model achieved $accu_y$, $prec_n, reca_l$, $F1_{score}$, AUC_{score} of 93.200%, 89.260%, 94.190%, 91.660%, and 93.370% correspondingly. The FS+LSTM method achieved slightly improved outcomes with $accu_y$, $prec_n, reca_l$, $F1_{score}, AUC_{score}$ of 94.490%, 92.960%, 93.160%, 93.060%, and 94.260% correspondingly. Finally, the

ARA+HS+FS+LSTM algorithm achieved the highest performance with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, AUC_{score} of 95.980%, 95.200%, 94.630%, 94.920%, and 95.750%, respectively.

TABLE III. RESULT ANALYSIS OF ARA+HS+FS+LSTM TECHNIQUE ON KDDCUP99 DATASET

Methods	LSTM	FS+LSTM	ARA+HS+FS+LSTM
Accuracy	93.200	94.490	95.980
Precision	89.260	92.960	95.200
Recall	94.190	93.160	94.630
F1-Score	91.660	93.060	94.920
AUC Score	93.370	94.260	95.750

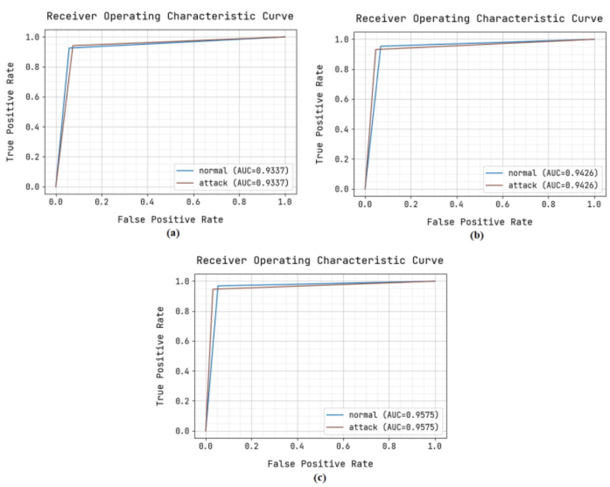


Fig. 11. ROC of KDDCup 99 dataset a) LSTM b) FS+LSTM c) ARA+HS+FS+LSTM

The ROC analysis of the ARA+HS+FS+LSTM technique along with other techniques on the test KDDCUP99 dataset is exhibited in Fig. 11.

From the figure, it is evident that the LSTM and FS+LSTM methods achieved ROC values of 0.9337 and 0.9426. The ARA+HS+FS+LSTM technique achieved ROC of 0.9575.

TABLE IV. COMPARATIVE ANALYSIS OF ARA+HS+FS+LSTM TECHNIQUE WITH OTHER EXISTING METHODS ON KDDCUP99 DATASET

Methods	Accuracy	Precision	Recall	F1-Score
ARA+HS+FS+LSTM	95.98	95.20	94.63	94.92
PCA+DNN	89.80	88.40	89.80	88.20
DNN Model	90.90	89.60	90.90	89.40
Random Forest	93.66	92.46	93.12	94.14
XGBoost	94.06	92.94	93.53	94.50
Decision Tree	93.00	91.28	92.48	92.62
IARADR	94.59	92.90	94.05	94.24
ML-SHNS	93.91	92.68	93.41	94.22

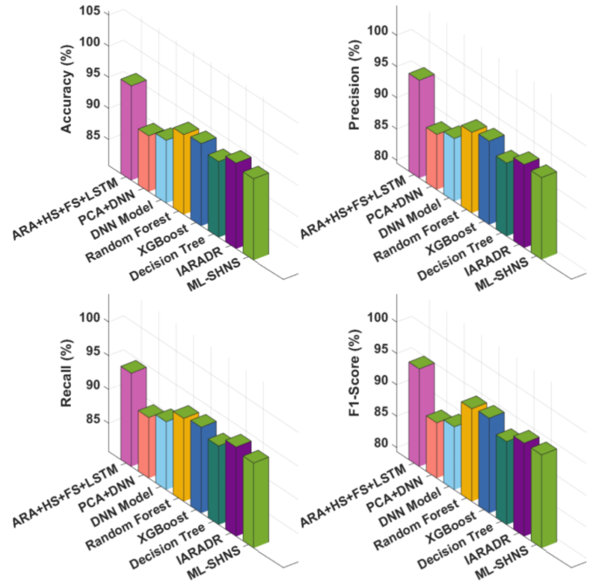


Fig. 12. Comparative analysis of proposed method on KDDCUP99 dataset

A detailed comparative result analysis of the ARA+HS+FS+LSTM method with other existing algorithms is made in Table IV and Fig. 12 [28, 29].

The results indicated that DNN and DT techniques achieved the least intrusion detection performance. The PCA+DNN and RF approaches achieved moderately improved intrusion detection performance. The XGBoost and IARADR techniques were next in terms of achievement. The ML-SHNS approach achieved near optimal outcomes. The ARA+HS+FS+LSTM technique achieved the highest performance results with $accu_y$, $prec_y$, $reca_l$, and $F1_{score}$ of 98.98%, 95.20%, 94.63%, and 94.92% correspondingly.

After evaluating the performance results of the proposed LSTM-ARA-HSA model along with other techniques and approaches, it is apparent that this approach is a feasible hybrid meta-heuristic approach that can be used by cloud-based IDS.

V. CONCLUSION

This paper has proposed a hybrid meta-heuristic approach combining ML techniques of ExtraTrees Feature Selection method and LSTM classification model with the Computational Intelligence Algorithms ARA-HAS, to develop the performance of cloud-based Intrusion Detection Systems and detect network attacks accurately and efficiently. These hybrid methods encompasses different processes such as Pre-processing of data dimensions, ET based feature selection, ARA-HSA based hyperparameter tuning, and LSTM based classification.

The adoption of ARA-HSA enables optimization of hyper-parameters for the dimensions selected using the ET Feature Selection method, in turn enabling more accurate LSTM classification.

The experimental outcome shows that the proposed LSTM-ARA-HSA has better performance than other techniques in terms of accuracy and efficiency. Therefore, this hybrid approach can be utilized by IDS for detecting and classifying the intrusions in the cloud environment. The objective of the current paper is to only establish that the proposed hybrid approach results in more accurate classification of attacks and the scope is limited to experimental validation using the two main datasets. As a part of future scope of research, the detection efficiency of this hybrid approach can be improved through appropriate design of clustering and outlier detection techniques.

REFERENCES

[1] K. Jain, S. Jain, A. Guha, and A. Patra, “An approach to early stage detection of atherosclerosis using arterial blood pressure measurements”, Biomedical Signal Processing and Control, vol.68, pp.102594, 2021.

[2] V.S.H. Rao, M.N. Kumar, “Novel approaches for predicting risk factors of atherosclerosis”, IEEE J. Biomed. Health Inform, vol.17, pp. 183–189, 2012.

[3] A.K. Gárate-Escamila, A.H. El Hassani, and E. Andrès, “Classification models for heart disease prediction using feature selection and PCA”, *Informatics in Medicine Unlocked*, no.*19*, pp.100330, 2020.

[4] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, “A novel medical diagnosis support system for predicting patients with atherosclerosis diseases”, Informatics in Medicine Unlocked, vol.21, pp.100483, 2020.

[5] R. Rajagopal, and V. Ranganathan, “Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification”, Biomed Signal Process Contr, vol.34, pp.1–8, 2017.

[6] U.R. Acharya, V.K. Sudarshan, J.E. Koh, R.J. Martis, J.H. Tan, S.L. Oh, A. Muhammad, Y. Hagiwara, M.R.K. Mookiah, K.P. Chua, et al. “Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals”, Biomed. Signal Process. Control, vol.31,pp.31–43,2017.

[7] M.A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. Yager, “Deep learning approach for active classification of electrocardiogram signals”, InfSci, vol.345, pp. 340–54, 2016.

[8] K. Jain, S. Maka, and A. Patra, “Modeling of cardiovascular circulation for the early detection of coronary arterial blockage”, Math. Biosci. 304 (2018) 79–88.

[9] 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, vol. 19, pp.197, 2019.

[10] K. Jain, A. Patra, and S. Maka, “Modeling of the human cardiovascular system for detection of atherosclerosis”, IFAC-PapersOnLine, vol 51,pp.545–550, 2018.

[11] K. Sethi, R. Kumar, N. Prajapati, and P. Bera, “Deep reinforcement learning based intrusion detection system for cloud infrastructure”, In proceedings of International Conference on COMmunication Systems &NETworkS (COMSNETS), pp.1-6, 2020.

[12] S. Ji, K. Ye, and C.Z. Xu, “A Network Intrusion Detection Approach Based on Asymmetric Convolutional Autoencoder”, In proceedings of nternational Conference on Cloud Computing, Springer, Cham, pp. 126-140, September 2020.

[13] Singh P. and Ranga V. “Attack and intrusion detection in cloud computing using an ensemble learning approach”, International Journal of Information Technology, vol.13, no.2, pp.565-571, 2021.

[14] O. Alkadi, N. Moustafa, B. Turnbull, and K.K.R. Choo, “A deep blockchain framework-enabled collaborative intrusion detection for protecting IoT and cloud networks”, IEEE Internet of Things Journal, vol.8, no.12, pp.9463-9472, 2020.

[15] M. Zhong, Y. Zhou, and G. Chen, “Sequential model-based intrusion detection system for IoT servers using deep learning methods”, Sensors, vol.21, no.4, pp.1113, 2021.

[16] J.K. Samriya, and N. Kumar, “A novel intrusion detection system using hybrid clustering-optimization approach in cloud computing”, Materials Today: Proceedings, 2020.

[17] A. Abusitta, M. Bellaiche, M. Dagenais, and T. Halabi, “A deep learning approach for proactive multi-cloud cooperative intrusion detection system”, Future Generation Computer Systems, vol.98, pp.308-318, 2019.

[18] N. Khare, P. Devan, C.L. Chowdhary, S. Bhattacharya, G. Singh, S. Singh, and B. Yoon, “Smo-dnn: Spider monkey optimization and deep neural network hybrid classifier model for intrusion detection”, Electronics, vol.9, no.4, pp.692, 2020.

[19] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees”, Mach. Learn, vol. 63, pp.3–42, 2006.

[20] E.K. Ampomah, Z. Qin, and G. Nyame, “Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement”, Information, vol.11, no.6, pp.332, 2020.

[21] Q. Jiang, L. Wang, X. Hei, G. Yu, and Y. Lin, “The performance comparison of a new version of artificial raindrop algorithm on global numerical optimization”, Neurocomputing, vol.179, pp.1-25, 2016.

[22] Y. Huang, and Y. Qiao, “Artificial raindrop algorithm for optimal parameter preference in digital IIR filters”, Adv. Model Anal. C, vol.72, pp.114-39, 2017.

[23] Z.W. Geem, J.H. Kim, and G.V. Loganathan, “A new heuristic optimization algorithm: harmony search”, simulation, vol.76, no.2, pp.60-68, 2001

[24] X.Z. Gao, V. Govindasamy, H. Xu, X. Wang, and K. Zenger, “Harmony search method: theory and applications”, Computational intelligence and neuroscience, 2015.

[25] N. Khan, I.U. Haq, F.U.M. Ullah, S.U. Khan, and M.Y. Lee, “CL-Net: ConvLSTM-Based Hybrid Architecture for Batteries”, State of Health and Power Consumption Forecasting, Mathematics, vol. 9,pp.33262021,2015.

[26] Moukhafi M, El Yassini K, Bri,S, “A novel hybrid GA and SVM with PSO feature selection for intrusion detection system”, Int. J.Adv. Sci. Res. Eng, vol.4, pp.129–134, 2018.

[27] F. Kuang, W. Xu, and S. Zhang, “A novel hybrid KPCA and SVM with GA model for intrusion detection. Appl. Soft Comput. J. 2014, 18, 178–184.

[28] M. Pozi, M.N. Sulaiman, N. Mustapha, and T. Perumal, “Improving Anomalous Rare Attack Detection Rate”, Neural Process. Lett., vol.44, pp.279–290, 2015.

[29] A. Khan, S. Abbas, A. Rehman, Y. Saeed, A. Zeb, M.I. Uddin, N. Nasser, and A. Ali, “A machine learning approach for blockchain based smart home networks security”, IEEE Netw. vol.35, pp. 223–2292021.

Social Media Multimodal Information Analysis based on the BiLSTM-Attention-CNN-XGBoost Ensemble Neural Network

Ling Jixian[1], An Gang[2], Su Zhihao[3], Song Xiaoqiang[4]
School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia[1]
School of Languages, Literacies and Translation Universiti Sains Malaysia, Penang, Malaysia[2]
Computer Science College, Guangzhou College of Applied Science and Technology, Guangzhou, China[3]
Ningxia University, Ningxia China[4]

***Abstract*—Social media users internalise information in a multimodal context. Social media functions as a primary information source for disaster situational awareness encompassing texts, photographs, videos, and other multimodal information widely used in emergency management. Applying ensemble learning to social media sentiment analysis has garnered much scholarly attention, albeit with limited research on rescue and its sub-domain, which is characterised as a major complexity. A multimodal information categorisation model based on hierarchical feature extraction was proposed in this study. The information of multiple modes is first mapped to a unified text vector space in modelling the semantic content at the sentence and multimodal information levels in the multimodal information. Multiple deep learning (DL) models were subsequently applied to model the semantic content at the aforementioned levels. This study offers a BiLSTM-Attention-CNN-XGBOOST ensemble neural network model to acquire extensive multimodal information characteristics. Based on the empirical outcomes, this method precisely extracted multimodal information features with an accuracy exceeding 85% and 95% for Chinese- and English-language datasets, respectively.**

Keywords—Natural language processing; social media sentiment analysis; multimodal information processing; ensemble neural network; emergency management

I. Introduction

Social media platforms involving Weibo, WeChat, Facebook, Instagram, and Twitter have evolved into (1) key real-time information acquisition channels in disaster management and (2) information communication tools in physical and virtual contexts following the advent of the mobile Internet [1-3]. A vast amount of rescue and assistance information is promptly shared on digital platforms. Users post many real-time words, photographs, and videos about casualties, facility damage, and emergency assistance on social media [4]. In disaster management, efficiently identifying complex multimodal information on social media and emergency information has garnered much attention during emergencies. Such multimodal information proves crucial in managing emergencies [5]. Natural language processing (NLP) and computer vision technologies based on machine learning (ML) and DL significantly improved their performance and effect in multimodal data-processing tasks owing to breakthroughs in technologies (artificial intelligence and big data), which primarily catalysed multimodal information analysis in social media [6]. Research on social media multimodal information analysis in the context of emergencies is a relatively novel study topic that has garnered much scholarly interest and developed a theoretical and methodological system [7]. Current studies on social media multimodal information in emergencies provide theoretical and methodological support for social media multimodal information analysis for emergency management [8].

Before the study model development, specific limitations in using Twitter to monitor situational awareness in a disaster event were identified in multimodalities and social media network types. For example, Twitter users only represented some residents, as not everyone uses Twitter. Generally, some disadvantaged (low-income, low-education, and elderly) groups without the devices or motivation to access social media applications are less inclined to publish or receive disaster-relevant information [9]. Current works only utilised Facebook and Twitter data regarding the perception and humanitarian challenges identified on Twitter. Current research combining images and text to classify social media emergency management information has focused on English datasets and Twitter platforms [10], with less research on cross-social media platforms and cross-language datasets. In addition, the classification accuracy of current research needs to be improved [11].

This study employed data from WeiBo, WeChat, Facebook, and Twitter. The dataset includes English and Chinese. Recent Studies highlight the need for a multimodal and complex information-processing model that gathers, models, and employs feedback information in the social media interaction with users and continually iterates and optimises the information [12]. Such an approach could passively or actively detect errors, implement online learning and dynamic updating mechanisms based on the faults, and construct a set of self-learning frameworks to alleviate information processing issues in emergency management.

The main contribution of this paper is to use the Bi-LSTSM-Attention-CNN-XGBOOST ensemble neural network model to resolve the intricate processing of multimodal information in social media. The model uses recurrent neural networks to improve the adaptive and migration capabilities of the model. Multimodal information obtained from different platforms is first mapped to a unified text vector space to solve the cross-platform cross-language problem. Various DL models are used to model the semantic content at the sentence and multimodal information levels. This model is divided into six layers: (1) Embedding layer: maps words into low-dimensional vectors; (2) BiLSTM layer: generates deeper semantic vector to represent each word with BiLSTM network; (3) Attention layer: achieves the attention weight for all word; (4) Pooling layer: engage k-max pooling to extract the top 'k' words, which are ranked in step 3; (5) CNN layer: input to the CNN network in performing convolutional operations and extracting features through first k-word vectors; (6) X GBOOST layer: the feature vectors extracted by the final CNN are classified by the X GBOOST integrated classifier.

This paper covers six sections. In Section II related work is presented. Section III covers the method while Sections IV and V explain results and discussions, while Section VI concludes the overall contributions and suggested directions for future work

II. Related Work

Given the advancements in mobile internet, social media platforms (WeiBo, WeChat, and Twitter) have gained prominence in real-time information acquisition channels for disaster management and function as information communication tools in physical and virtual worlds. Only 2% of the extracted works directly encompassed the term 'multimodal' in the title or abstract. All the studies were published in the past five years. Notably, social media data on multiple modalities were analysed. A total of 216 articles potentially incorporating the idea of multimodal fusion accounted for 27% of all the documents. The most common publications involved information analysis of text and image modalities.

A study by [13] identified temporal changes in the social media sites extensively utilised for disaster recovery, their usage patterns by catastrophe type, and the geographic areas other studies have emphasised. Empirical outcomes on social media use in multiple disaster recovery aspects, such as (1) financial support and donations, (2) solidarity and social cohesion, (3) infrastructural services and post-disaster reconstruction, (4) socioeconomic and physical well-being, (5) information support, (6) mental health and emotional support, and (7) business and economic activities were also highlighted. Twitter users from Australia and beyond have aided bushfire victims through positive messaging regarding contributions, relief assistance, news updates, and animal welfare. Research by [14] examined how Twitter was utilised to publicise blaze recovery in Australia. Concerns about climate emergency and the perceived lack of political action entailed the primary sources of undesirable attitudes. In [15], researchers proposed fine-tunned BERT for multimodal analysis. While in [16], trustworthy summaries from crisis-related microblogs are predicted.

Modern deep learning methods were recommended by [17] to learn an integrated representation of social media data involving text and picture modalities. Specifically, a multimodal deep learning architecture with a modality-agnostic common representation was defined with convolutional neural networks. The proposed multimodal architecture outperformed the models developed using a single modality (text or image) based on studies that employed data from actual disasters. The model outperformed the image-only model by approximately 1% in all the informativeness task metrics and 2% in all humanitarian task measurements. Concerns about people's situational awareness during the occurrence of a natural disaster vary based on the stakeholder.

In regarding perception-, humanitarian-, and action-level situational awareness into account, [10] structured a Twitter-based analytic framework for damage estimation: the most prevalent use of Twitter in disaster management. Social media platforms could promptly ascertain damage to save people in danger, determine evacuation routes, and plan countermeasures for potential catastrophes. A study by [18] computed the extent to which volunteering content from one crisis is transferrable to another through language consistency analysis in volunteer- and donation-related social media material across 78 crises. Particular techniques were also provided to offer computational assistance in this emergency support role and establish semi-automated models in classifying social media information on volunteering and donating in the wake of a new crisis. Resultantly, the social media materials associated with volunteering and donations were sufficiently comparable between disasters and disaster types to justify transferring models across the disasters. These models were subsequently assessed with direct resampling procedures.

In comprehending the implications of social media to influence the relief and recovery process, the thematic and emotional nature of Twitter content discussing the 2019–2020 Australian bushfire disaster and its associated wildlife damage was discussed [19]. Thus, the value of social media (and Facebook in particular) has been demonstrated for knowledge sharing, volunteer coordination, fundraising, peer motivation, and accountability. Social media also makes it more challenging to distribute aid and coordinate relief efforts based on disinformation and duplication.

Following [20], past models' categorization of catastrophe signals primarily depends on unimodal techniques. Although specific approaches utilized multimodality to manage data, their accuracy could not be ascertained. A multimodal fusion strategy was developed to identify the relevant catastrophe photos from social media networks and merged the image and text information. The textual characteristics were subsequently elicited from social media with a FastText framework after the visual data were extracted through a deep learning technique. In classifying pertinent catastrophe photographs, a novel data fusion model combining linguistic and visual elements were structured. Well-executed experiments on the Crisis MMD dataset of actual disasters were also conducted.

An efficient DL algorithm was presented in [21] to manipulate multimodal information sources (words and photos) and disseminate helpful information during natural disasters. The programme divided the tweets into seven crucial and actionable groups, including reports of 'hurt or dead individuals' and 'infrastructure damage'. Based on research incorporating a benchmark dataset, integrating text and picture data from multiple sources proved more successful than employing data from a single source: one of the two in extracting pivotal information in crisis circumstances. Using visual and linguistic inputs, the recommended multimodal architecture in [22] classified damage-related postings using ResNet50 and BiLSTM recurrent neural networks using attention mechanisms. The MTLTS, the first end-to-end method for gathering reliable summaries from a substantial number of tweets on disasters, was introduced to supervise methodology and improve the applicability of solutions to unprecedented situations. This innovative approach involving the extractive document summarization technique summarised tweets on specified events. By concurrently learning the structural properties of information cascades and response stances, the credibility verifier is optimized with advanced components to detect rumours.

Information processing is crucial for large-scale data visualisation, which could be performed through data harmonisation [23] and implies the exact representation as multimodal data. Multimodal information-processing methods are currently categorised into classical and DL methods. Conventional methods primarily include support vector machines (SVM), naïve Bayes (NB), KNN, and LDA [24] topic model algorithms. Deep learning methods have recently undergone rapid development. For example, Mikolov et al. [25] suggested the Word2vec model for word quantisation in 2013, substituting the simple one-hot vectorisation method and resolving data sparsity issues. The CBOW and Skip-gram models in Word2vec were identified to train word vectors. Numerous optimisation methods were recommended for training details to enhance the training speed significantly. An attention mechanism method for deep learning in 2014 to extract essential information [26]. The RNN network-attention mechanism integration outperformed most single models within the image domain.

The author in [27] applied the long short-term memory (LSTM) network to the text classification field in the same year, thus mitigating the lack of contextual information in CNN. The training time is longer for long text datasets as LSTM takes the whole text as input without extracting vital information. A study by [28] suggested a recurrent convolutional neural network structure (RCNN) that connected BiLSTM and the maximum pooling layer. The BiLSTM network captured contextual information while the pooling layer extracted vital features. Hence BiLSTM is fully utilised with the integration of both structures. Extracting contextual information that only retains the pooling layer in the CNN part weakens the CNN's ability to extract features. Noda proposes a model for the multimodal robot in [29] and [30] proposes DeepCEP in 2019 with multimodal information streams and complex, spatial, and temporal dependencies.

The DL method of complex multimodal information based on the neural network has been widely acknowledged for its ability to simulate the cognitive function of the human brain. Regardless, this method (not a mathematical model based on the workings of the human brain) only denotes a formal mathematical description of the neuron structure and signal transmission method. It is also deemed challenging to eliminate the reliance on large-scale training samples. In 2020, Bejan described the MemoSys system submitted in Task 8 of SemEval 2020 to classify Internet memes sentiments [31]. In 2021, Zahera and colleagues recommended I-AID, a multimodal approach to automatically categorize tweets into multi-label information types [15].

Reference [32] highlighted various multimodal learning challenges and ways to train deep networks to learn features for task management. Cross-modality feature learning was also demonstrated, which allows for better features for a single modality (video) when other modalities (audio and video) are present when the features are being learned. The means to train a classifier with audio-only data to be tested with video-only data and evaluated on a specific task using shared representations between modalities was also denoted. Research by [33] proposed a model and neural network that demonstrated how image-text modelling could mutually learn word representations and image attributes. This methodology provides sentence descriptions for images without templates, structured prediction, or syntax trees in contrast to many other existing techniques.

Essentially, [34] developed and assessed advanced neural network techniques that incorporate input from several modalities to investigate the impact of complex interactions between textual, visual, and metadata on project success prediction. The method requires data gathered from the pre-posting profile to enable pre-posting prediction. Following [35], a multi-label and multimodal framework using text, audio, and visual input modalities was offered to categorise unbalanced data and automatically create static and temporal features using spatiotemporal deep neural networks. A weighted multi-label classifier functioned to manage data using non-uniform distributions.

A revolutionary deep dual recurrent encoder model in [36] was proposed to comprehend speech data fully and concurrently manipulate text data and audio signals. As spoken and musical content constitutes emotional dialogue, this model employed RNNs to encode audio and text sequences information before integrating it to determine the emotion class. Additionally, [37] proposed a deep multimodal attentive fusion (DMAF), a model for image-text sentiment analysis, to manipulate the discriminative features and internal connection between visual and semantic content using a hybrid fusion framework for sentiment analysis. The multimodal information proves crucial in emergency management. In the 'scenario-response' approach, social media provides additional data sources for emergency management. The information content in multiple modalities is interrelated, albeit with contextual differences.

In a text containing events, the content, image, and video details genuinely depict the scenes of the event. Such aspects complement and confirm one another to reflect the development status of real-world events. It is deemed pivotal to abstract, generalize, and integrate the complex social media multi-modal information from different levels and aspects and summarize and extract more accurate and comprehensive information than a single modality for users' holistic and in-depth understanding of real-time emergencies to reduce emergency management ambiguities. Information extraction from social media is implied as a binary text classification issue with the labels 'relevant' and 'irrelevant' [15]. Nevertheless, effective methods to connect relevant postings to finer-grained classifications remain scarce. Such fine-grained labelling proves useful for crisis responders who need promptly provide catastrophe reactions and update vital information. Mainly, categorizing tweets on disasters with multiple labels facilitates the rapid identification of tweets with helpful information. The model performed optimally when breaking down disaster-related tweets into specific information types.

In research of [12] proposes a multimodal deep-learning framework to identify damage-related information. This framework combines multiple pre-trained unimodal convolutional neural networks that independently extract features from raw text and images, followed by a final classifier that labels posts based on both modalities, ultimately achieving an accuracy of about 92.6% in the English dataset. An approach to classifying thematic social media by integrating visual and textual information through a fused CNN architecture is proposed [38]. Specifically, two CNN architectures are used, targeting visual and textual information of social media posts, respectively. The outputs of the two CNNs, i.e., the features extracted from the social media posts representing visual and textual features, are further connected to form a fused representation. Research in [5] proposed a multimodal approach to identify disaster-related information content from Twitter streams using text and images. The method is based on long and short-term memory and VGG-16 networks, which significantly improve performance. In addition, [39] investigated the extent to which integrating multiple modalities is essential for classifying crisis content. In particular, a multimodal learning pipeline was designed to fuse textual and visual inputs, utilise both, and then classify that content based on a specified task. The average F1 performance in two critical tasks (relevance and humanitarian category classification) was 88.31%. A Python-based data pipeline application, SMDRM, for processing social media data points was proposed in [7].

However, this current work mainly explores the classification of English datasets, and more research needs to be done on Chinese contingency management datasets in social media. Notwithstanding, most multimodal information processing algorithms in social media, which are almost global and static, did not learn from failures and user input in real-time. In research of [40] proposed a multimodal network (MDMN)-based approach for rumour detection, including a text feature extractor, a visual feature extractor, and a fusion classification network, applying a multitask sharing layer, a task-specific converter encoder, and a selection layer to improve the diversity and stability of the text. Domain adaptation involves training adaptive models to extract visual representations. Adaptive models can encode task data better than fine-tuned pre-trained models. Then, experiments with two fusion strategies to fuse multimodal representations of multimodal datasets collected from tweets and microblogs show that the proposed MDMN can outperform the baseline approach. The decision-level fusion strategy achieves more than 92% Recall. In classifying emergency management information in social media, it is nearly impossible for existing algorithms to imitate the intelligent behaviour of human interaction and lifetime learning. It highlights the need for a multimodal and complex information-processing model that gathers, models, and employs feedback information in the social media interaction with users and continually iterates and optimises the information. Such an approach could passively or actively detect errors, implement online learning and dynamic updating mechanisms based on the faults, and construct a set of self-learning frameworks to alleviate information processing issues in emergency management.

All related literature shows that social media sentiment analysis plays a vital role in the advanced technological world. However, it is also found that the existing DL models contribute a lot. However, no such model developed which cross-platform, cross-lingual, multimodal and used multimodal social media data with high accuracy.

III. METHOD

This study proposed the Bi-LSTSM-Attention-CNN-XGBOOST ensemble neural network model (see Fig. 1) to address the complex processing of multimodal information in social media. The model manipulated ensemble learning as presented in Fig. 2. The information of multiple modes was initially mapped to a unified text vector space. Various deep learning models were used to model the semantic content at sentence and multimodal information levels. The MPOD textual data were used for model training and testing. As the name implies multimodality, English- and Chinese-language corpora from different social media platforms were selected for the recommended model. This model is divided into six layers:

1) Embedding layer: the word embedding matrix is regarded as the neural network model parameters and input for the BiLSTM neural network model to encode a sentence.

2) BiLSTM layer: BiLSTM constitutes two independent LSTMs that could summarise information from the forward and backward directions of a sentence. Essentially, the information stemming from both directions could be merged. At each time 't', the forward LSTM computed the hidden vector 'fht' based on the past hidden vector 'fht−1' and the input word embedding 'xt'. The backward LSTM calculated the hidden vector 'bht' based on the opposite (past) hidden vector 'bht−1' and the input word embedding 'xt'. Subsequently, the forward hidden vector 'fht' and backward hidden vector 'bht' were merged into the final hidden vector of the BiLSTM model. In this model, the parameters of two opposing directions proved independent albeit sharing the same word embedding of a sentence to generate a deeper semantic vector and represent each word with the BiLSTM network.

3) Attention layer: the self-attention mechanism functions to extract the more important information by providing them with a higher weight to elevate their significance. This layer primarily aimed to achieve the attention weight for all words and their ranking.

4) Pooling layer: this layer accepts the temporal sequence output by the LSTM layer and performs temporal max-pooling with sole reference to the non-masked portion of the sequence.

The pooling layer converted the entire variable-length hidden vector sequence into a single hidden vector for its output to be fed into the dense layer. The aforementioned layer engages the k-max pooling to extract the top k-words, which are ranked in step 3.

5) CNN layer: CNN layers are added to the front end, followed by the LSTM layers with a dense layer on the output. This layer provides input to the CNN network to perform convolutional operations and extract the key features with first k-word vectors.

6) X GBOOST layer: the XGBoost method enhanced the features extraction mechanism for DL models. This layer generates the feature vectors that are extracted by the final CNN and classified by the XGBOOST integrated classifier.

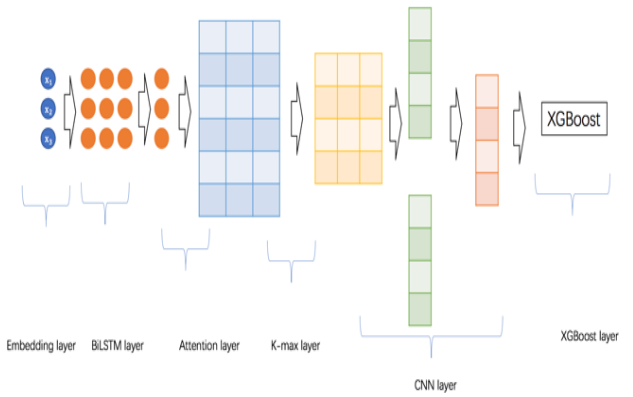


Fig. 1. The BiLSTM-attention-CNN-XGBOOST structure

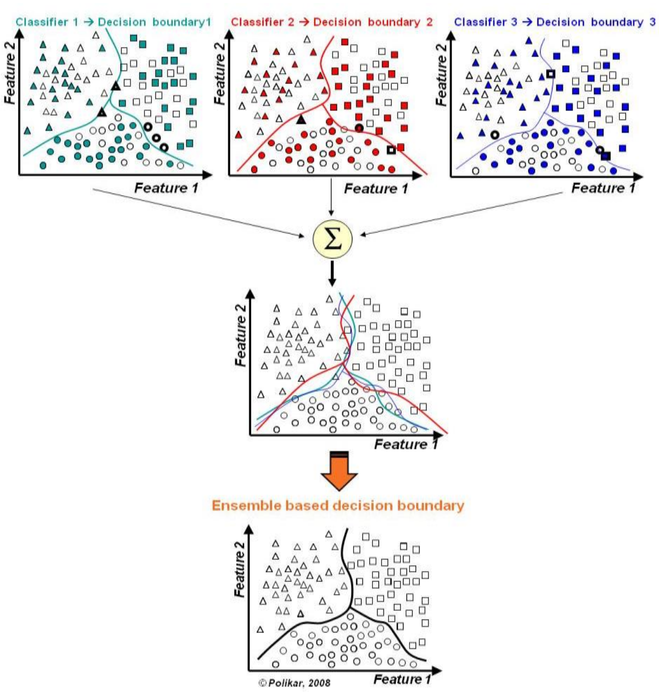


Fig. 2. The ensemble learning process [15]

A. Embedding Layer

Word embedding implies the representation method of word vectors in NLP. Although most conventional word vector representation methods utilise one-hot encoding, this method easily results in a high vector dimension. In the proposed study model, the Skip-gram model in Word2vec was utilised in word embedding to pre-train the words.

B. BiLSTM Layer

The LSTM layer utilises BiLSTM for deeper semantic vectors of words. This structure could combine the current word context to prevent future words in RNN from becoming more influential than their previous counterpart.

C. Attention Layer

Videos can be converted into text in the NLP domain. As the basic unit of text implies words, current models are primarily trained and operated on them. The attention mechanism suggested in this study mainly computed the attention of words in a given text. High levels of attention received by the word denoted its pivotal role in the task. Overall, extracting words with high attention could effectively classify multimodal information.

D. K-max Pooling Layer

The pooling layer in the NLP domain is primarily used to extract features and convert multimodal information of different lengths into fixed-length vectors. A common pooling operation denotes the max pooling operation, which retains one of the features. Meanwhile, the multimodal information ‘t’ typically denotes several words or their integration to express the meaning underlying the whole multimodal information. This model adopted k-max pooling, retained the first k-word vectors with larger attention scores, and obtained the combined feature vector ‘y’ of the first ‘k’ words. A larger weight implies higher levels of attention. Based on the outcomes derived from various experiments, the value of ‘k’ implied 8.

E. CNN Layer

This layer adopts the original Text CNN model for feature extraction. The first (network) layer takes the first k-word vectors retrieved by the k-max pooling layer as input. The second (convolution) layer performs convolution on input word vectors with several filters of varying sizes. The third (pooling) layer performs the highest number of pooling operations to create a novel feature vector.

F. XGBoost Layer

The BiLSTM-Attention-CNN-XGBoost ensemble neural network model proposed in this study implied the XGBoost ensemble classifier. If the length of a sentence is ‘n’, word embedding is performed through the pre-trained k-dimensional word vector with each word represented as a k-dimensional vector. Furthermore, the sentence could be converted into an n*k -dimensional data matrix as input. The convolutional layer employs 3*k, 4*k, and 5*k convolution kernels to extract features, pool the extracted features, and fully connect the features extracted by convolution kernels of varying sizes. Lastly, the multimodal information multi-classification processing is completed through the XGBoost classifier. In assuming that a sentence constitutes six words with each word denoting a four-dimensional word vector, the sentence could be characterised as a 6×4 data matrix as the convolution layer input through 3×4, 4× 4. The 5×4 convolution kernel performs feature extraction to obtain 4×1, 3×1, and 2×1 feature maps. A three-dimensional feature map is derived from maximum pooling while the XGBoost classifier is employed for classification processing.

IV. Experiment and Results

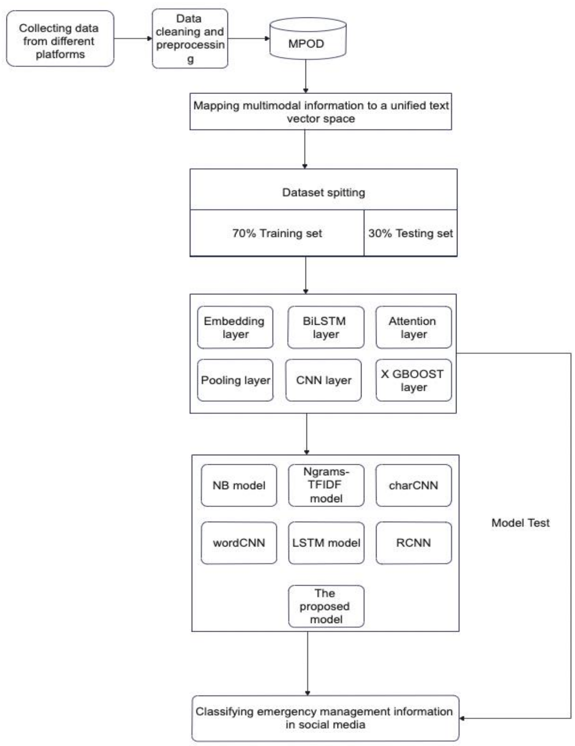


Fig. 3. The experiment framework

The experiment was conducted on ubuntu16.04 with Intel(R) Xeon(R) E5-265 as the CPU, 2.4GHz as the frequency, 32GB as the memory size, Python3.0 as the experimental programming language, and Tensorflow1.12.0 as the deep learning framework. The experimental dataset implied the news public opinion corpora in Weibo, Wechat, Twitter, and Facebook. As shown in Fig. 3, this self-collected dataset containing 6,700 text samples, 342 Image samples and 67 sounds samples was referred to as the multilingual public opinion dataset (MPOD). Notably, 70% of the samples were arbitrarily selected as a training set while the remaining 30% were chosen as a test set. The training set was manipulated to train the proposed model and evaluate it with the test set.

In Table I, the overall comparison with existing models implemented on the selected dataset are presented. The performance of the proposed model is more efficient and accurate then all with having 85 to 97%. The NB method performance varies from 78 to 93. Whereas N-gram with TF-IDF have 78-94, as well as charCNN have slightly incremental up to 95. The wordCNN have 95.66% and LSTM have 95.78% and last but not the least RCNN jumps to 96%.

TABLE I. The Comparison of Different Methods in MPOD

Method	**Chinese language set**		**English language set**	
	WeiBo	**WeChat**	**Twitter**	**Facebook**
NB model	84.12%	78.56%	93.21%	93.02%
Ngrams - TFIDF model	84.78%	78.94%	93.55%	94.67%
charCNN	85.23%	79.04%	93.67%	95.05%
wordCNN	84.97%	79.89%	94.87%	95.66%
LSTM model	85.05%	80.98%	95.67%	95.78%
RCNN	85.24%	81.45%	95.98%	96.02%
The proposed method	**87.64%**	**85.65%**	**98.34%**	**97.62%**

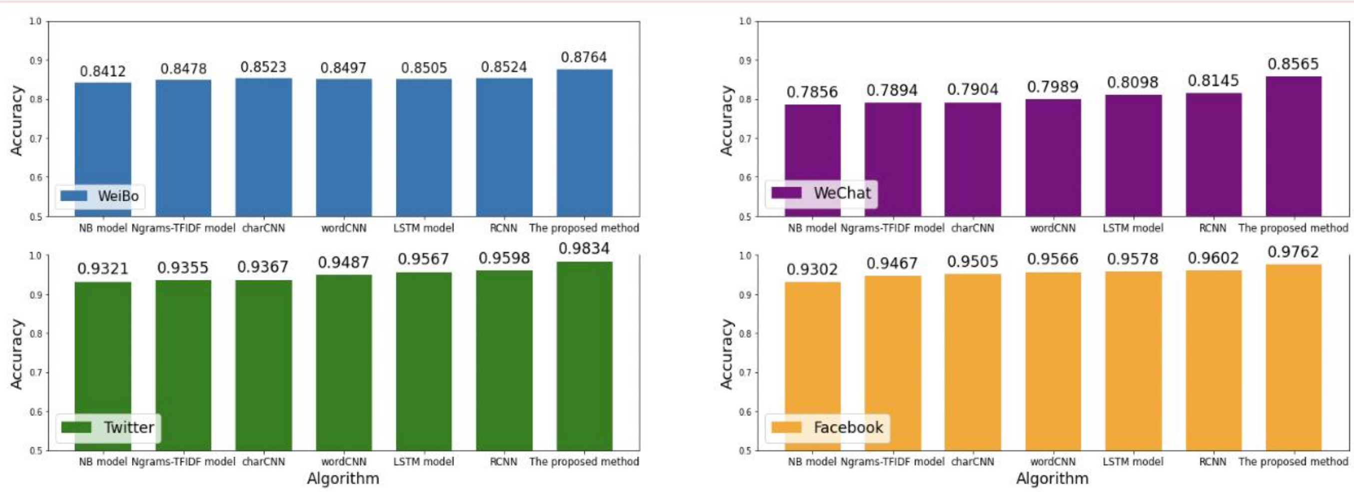


Fig. 4. The comparison of different methods in MPOD

As shown in Fig. 4, the BiLSTM-Attention-CNN-XGBOOST model proposed in this paper shows better accuracy than NB MODEL, Ngrams-TFIDF model, charCNN, wordCNN, LSTM model, RCNN model, regardless of Chinese or English data, whether from WeiBo, Twitter, Wechat or Facebook. The highest accuracy rate of Chinese language dataset is from WeiBo, reaching 87.64%, and the highest accuracy rate of English language dataset is from Twitter, reaching 98.34%.It can be clearly seen that the accuracy rate in the English data set is significantly higher than that in the Chinese data set by about 10%.In the English data set, Twitter performed better than Facebook, and in the Chinese data set, Weibo performed better than Wechat. Facebook and WeChat are biased towards acquaintance connections, while Twitter and Weibo are biased toward media connections.

V. Discussion

Six multimodal information classification methods were selected in this study to verify and discuss the model superiority. The conventional multimodal information classification method employed the NB and Ngrams-TFIDF models, while the deep learning method utilised the charCNN, wordCNN, LSTM, and RCNN models. Research by [32-37] similarly highlighted the multimodal model utilised in various textual, image and video datasets for different domains. Likewise, [15, 17, 20-22] employed multimodality to develop DL-based models in facilitating disaster management and recovery. Table I depicts the comparison of different classification algorithms and their effects on each model.

This study emphasised three multimodal information categories: (1) natural language, which can be both written and spoken; (2) visual signals that are typically represented by images or videos; (3) sounds that encode auditory and paralinguistic information, such as prosody and vocal expression signal. Following the data analysis in Table I, the classification effect of the study model on the same dataset is better for multimodal information. The two conventional methods, NB and Ngrams-TFIDF, depend on statistics, whereas charCNN, wordCNN, LSTM, and RCNN are classic neural network deep learning algorithms. Owing to the multi-layered nature of deep neural networks, a multimodal representation was constructed with several separate neural layers for each modality, followed by a hidden layer that projected the modalities into the joint space to employ neural networks.

The joint multimodal representation was conveyed through multiple hidden layers or directly used for prediction. In this study, the ensemble learning method served to extract the feature vector or latent semantic information of the multimodal counterpart by (1) studying the language cognitive mechanism of the brain and (2) analysing the relationship between cognitive mechanism and multimodal information computing methods. The four algorithms Bi-LSTM, attention mechanism, CNN, and XGBoost were integrated to extensively derive the advantages underlying each algorithm. For example, l-max pooling could reduce the number of model parameters that facilitates the mitigation of model overfitting issues. The attention mechanism could employ the human visual mechanism for intuitive interpretation, neural network interpretability, an in-depth understanding of the inner neural network mechanisms, and connect the neural network model structured based on the workings of the human brain. The weight-sharing network structure of CNN implies its mechanism.

The CNN resembles the biological neural network, thus reducing the complexity of the network model and number of weights. For example, XGBoost provides the model with a larger learning space and utilises feedback information for continuous model optimisation. As a result, the model classification accuracy in this study was 2% to 3% higher than that of the general method. This elevation resulted from using the k-max approach in the pooling layer in this study as the ‘k’ value substantially affected the experiment. Thus, several tests proved necessary to determine the most appropriate parameters and increase the experimental effect at the parameter adjustment stage of test training.

VI. Conclusion and Future Work

This study proposed a multimodal information classification method based on the hybrid BiLSTM-Attention-CNN-XGBoost neural network to classify multimodal information and maximise the benefits derived from each network model. Based on the experimental outcome, the classification accuracy was considerably improved upon adding the attention mechanism despite consuming much time as the experiment was performed on a single machine. The current study model would be tested on a distributed platform in the future to shorten the classification time. Based on the self-collected dataset called MPOD, the aforementioned hybrid neural network outperformed other models, such as CNN and Bi-LSTM, RCNN and Highways, LDA and SVM, attention mechanism, and GRU network. Further research would emphasise the multiple model fusion impacts on the experimental results with multimodal data.

References

[1] Gialampoukidis, Ilias, et al. "Multimodal Data Fusion of Social Media and Satellite Images for Emergency Response and Decision-Making." 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, 2021.

[2] AlAbdulaali, A., Asif, A., Khatoon, S., & Alshamari, M.. Designing Multimodal Interactive Dashboard of Disaster Management Systems. Sensors, 22(11), 4292, 2022.

[3] Lin, J. Social Media Multi-modal Processing Mode for Emergency. In International Conference on Multi-modal Information Analytics (pp. 52-58). Springer, Cham, 2022.

[4] Dominguez-Péry, C., Tassabehji, R., Vuddaraju, L. N. R., & Duffour, V. K.. Improving emergency response operations in maritime accidents using social media with big data analytics: a case study of the MV Wakashio disaster. International Journal of Operations & Production Management, 2021.

[5] Kumar, A., Singh, J. P., Dwivedi, Y. K., & Rana, N. P. A deep multi-modal neural network for informative Twitter content classification during emergencies. Annals of Operations Research, 1-32,2020.

[6] Zhang, C., Yang, Z., He, X., & Deng, L. Multimodal intelligence: Representation learning, information fusion, and applications. IEEE Journal of Selected Topics in Signal Processing, 14(3), 478-493, 2020.

[7] Lorini, V., Panizio, E., & Castillo, C. SMDRM: A Platform to Analyze Social Media for Disaster Risk Management in Near Real Time. In Workshop Proceedings of the 16th International AAAI Conference on Web and Social Media. Retrieved from https://doi. org/10.36190, 2022.

[8] Imran, M., Ofli, F., Caragea, D., & Torralba, A. (2020). Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions. Information Processing & Management, 57(5), 102261,2020.

[9] Ruiz Soler J. Twitter research for social scientists: A brief introduction to the benefits, limitations and tools for analysing Twitter data[J]. 2017.

[10] Zhai, W. A multi-level analytic framework for disaster situational awareness using Twitter data. Computational Urban Science, 2(1), 1-15, 2022.

[11] Zahera, H. M., Elgendy, I. A., Jalota, R., & Sherif, M. A. Fine-tuned BERT Model for Multi-Label Tweets Classification. In TREC pp. 1-7, 2019, November.

[12] Alam, Firoj, et al. "Deep learning benchmarks and datasets for social media image classification for disaster response." 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2020

[13] Ogie, R. I., James, S., Moore, A., Dilworth, T., Amirghasemi, M., & Whittaker, J. Social media use in disaster recovery: A systematic literature review. International Journal of Disaster Risk Reduction, 102783, 2022.

[14] Dahal, L., Idris, M. S., & Bravo, V. "It helped us, and it hurt us" The role of social media in shaping agency and action among youth in post-disaster Nepal. Journal of Contingencies and Crisis Management, 29(2), 217-225, 2021.

[15] Zahera, H. M., Jalota, R., Sherif, M. A., & Ngomo, A. C. N. I-AID: Identifying Actionable Information from Disaster-related Tweets. IEEE Access, 9, 118861-118870, 2021.

[16] Mukherjee, R., Vishnu, U., Peruri, H. C., Bhattacharya, S., Rudra, K., Goyal, P., & Ganguly, N. MTLTS: A Multi-Task Framework To Obtain Trustworthy Summaries From Crisis-Related Microblogs. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining ,pp. 755-763, 2022, February.

[17] Ofli F, Alam F, Imran M. Analysis of social media data using multimodal deep learning for disaster response[J]. arXiv preprint arXiv:2004.11838, 2020.

[18] Mousavi, P., & Buntain, C."Please Donate for the Affected": Supporting Emergency Managers in Finding Volunteers and Donations in Twitter Across Disasters. In Proceedings of the 19th ISCRAM Conference–Social Media for Crisis Management 605-622, 2022.

[19] Willson, G., Wilk, V., Sibson, R., & Morgan, A. Twitter content analysis of the Australian bushfires disaster 2019-2020: futures implications. Journal of Tourism Futures, 2021.

[20] Zou, Z., Gan, H., Huang, Q., Cai, T., & Cao, K. Disaster Image Classification by Fusing Multimodal Social Media Data. ISPRS International Journal of Geo-Information, 10(10), 636, 2021.

[21] Ahmad, Z., Jindal, R., Mukuntha, N. S., Ekbal, A., & Bhattachharyya, P. Multi-modality helps in crisis management: An attention-based deep learning approach of leveraging text for image classification. Expert Systems with Applications, 195, 116626, 2022.

[22] Hossain, E., Hoque, M. M., Hoque, E., & Islam, M. S. A Deep Attentive Multimodal Learning Approach for Disaster Identification From Social Media Posts. IEEE Access, 10, 46538-46551, 2022.

[23] Kumar G, Basri S, Imam A A, et al. Data Harmonization for Heterogeneous Datasets: A Systematic Literature Review[J]. Applied Sciences, 2021, 11(17): 8275.

[24] Wang Jinhua, Yu Hui, Chan Wen, et al Automatic text classification technology based on KNN + hierarchical SVM [J] Computer applications and software, 2016, 33 (2): 38-41

[25] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in neural information processing systems, 2013, 26.

[26] Mnih V, Heess N, Graves A. Recurrent models of visual attention[J]. Advances in neural information processing systems, 2014, 27.

[27] ZhouCT , SunCL , LiuZY , etal . AC-LSTM neural network for text classification[EB]. arXiv:151.08630, 2015.

[28] Chen Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.

[29] Noda K, Arie H, Suga Y, et al. Multimodal integration learning of robot behavior using deep neural networks[J]. Robotics and Autonomous Systems, 2014, 62(6): 721-736.

[30] Tianwei Xing; Marc Roig Vilamala; Luis Garcia; Federico Cerutti; Lance M. Kaplan; Alun D. Preece; Mani B. Srivastava; DeepCEP: Deep Complex Event Processing Using Distributed Multimodal Information, 2019 IEEE INTERNATIONAL CONFERENCE ON SMART COMPUTING (SMARTCOMP), 2019. (IF: 3)

[31] Bejan I. MemoSYS at SemEval-2020 task 8: Multimodal emotion analysis in memes[C]//Proceedings of the Fourteenth Workshop on Semantic Evaluation. 2020: 1172-1178.

[32] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. Multimodal deep learning. ICML.2011.

[33] Kiros R, Salakhutdinov R, Zemel R. Multimodal neural language models[C]. International conference on machine learning. PMLR, 2014: 595-603.

[34] Cheng C, Tan F, Hou X, et al. Success Prediction on Crowdfunding with Multimodal Deep Learning[C]. IJCAI. 2019: 2158-2164.

[35] Pouyanfar S, Wang T, Chen S C. A multi-label multimodal deep learning framework for imbalanced data classification[C]. 2019 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, 2019: 199-204.

[36] Yoon S, Byun S, Jung K. Multimodal speech emotion recognition using audio and text[C]//2018 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2018: 112-118.

[37] Huang F, Zhang X, Zhao Z, et al. Image–text sentiment analysis via deep multimodal attentive fusion[J]. Knowledge-Based Systems, 2019, 167: 26-37.

[38] Huang X, Li Z, Wang C, et al. Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture[J]. International Journal of Digital Earth, 2019.

[39] Afyouni I, Al Aghbari Z, Razack R A. Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey[J]. Information Fusion, 2022, 79: 279-308.

[40] Zhou, H., Ma, T., Rong, H., Qian, Y., Tian, Y., & Al-Nabhan, N. . MDMN: Multi-task and Domain Adaptation based Multi-modal Network for early rumor detection. Expert Systems with Applications, 2022, 195, 116517.

Design of Human-Computer Interaction Product Interface of Intelligent Service System based on User Experience

Xiaoli Xiong[1], Yongguang Hou[2]*
College of Education Science, Weinan Normal University, Weinan 714099, China[1]
School of Computer Science and Technology, Weinan Normal University, Weinan 714099, China[2]

***Abstract*—The current intelligent service platform for human-computer interaction products and services on the user experience is not comprehensive enough, resulting in user satisfaction cannot reach the ideal level. Therefore, a new human-computer interface of intelligent service system based on user experience is designed. Build a user experience based intelligent service system hardware platform, the introduction of the full name of hypertext markup language, so that more personalized design to be met. Design user experience PC terminal and Bluetooth/RS-485 gateway module to achieve two-way signal conversion between Bluetooth and RS-485. Based on ARM processor, the speech recognition of human-computer interaction is completed, the features of collected data are extracted, and the hand gesture recognition is completed. In order to optimize the human-computer interaction effect, Kinect is used to track and identify moving objects, and the 3D interactive image is simulated by fused texture. Experimental results show that the proposed method has a higher probability of receiving data, and the recognition rate of gesture features and recognition accuracy can reach more than 90%.**

Keywords*—User experience; intelligent service system; human-computer interaction; interface design; ARM processor; gateway module

I. Introduction

With the rapid improvement of computer's storage capacity and processing speed, the performance of service robot is getting better and better. More and more service robots are applied to all walks of life [1]. In order to make the service robot enter into people's daily life from the laboratory, some key technologies must be broken through, such as path planning, environment representation, control system, human-computer interaction, mechanism design and so on. As an interactive channel between user and robot, the human-computer interaction of intelligent service system is a crucial technology [2]. The human-computer interaction system of a friendly and natural intelligent service system is the key to the success of service robot. The purpose of the human-computer interaction design of intelligent service systems is to make the communication between humans and robots more convenient, more reliable and more in line with human interaction habits, as well as to reduce users' psychological and physical burdens [3].

The most important thing in the human-computer interaction of intelligent service system is how to make the user complete the task best, rather than how to design the best human-computer interaction of intelligent service system. In applications, the current research direction is no longer the pursuit of function, but "user-centric" design [4-5]. In modern software development, the research and development of human-computer interface of intelligent service system is of great importance. It connects people with computer technology and realizes humanization of computer technology. The human-machine interface of intelligent service system is the most closely related part of application system, so the workload of this part accounts for a large proportion of the whole development work [6-8].

Three core technical modules of intelligent robot: interaction+perception+operation control. Among them, the full name of interaction is human-computer interaction and recognition module. The functions of this module mainly include speech synthesis, speech recognition, image acquisition, image recognition, etc. The technologies related to speech recognition include speech recognition, natural language understanding, natural language generation, speech synthesis and dialogue engine. Image recognition technology includes image processing, analysis, understanding and other different types of technology. But at present, people pay more and more attention to the cross-platform of programs, that is, programs with the same function can run on different system platforms, and keep the same function and interface style, thus enhancing the portability and flexibility of programs. It should be noted that the current intelligent service robot cannot achieve this goal, and the existing technology does not support the robot to be truly customer-centric and truly intelligent and humanized. The innovation of the research is to realize voice recognition and gesture recognition with (Advanced RISC Machines) ARM processor, and realize tracking and recognition of moving images with Kinect technology, thus truly realizing human-computer interaction centered on user experience.

The organizational structure of the paper is as follows. The second part analyzes the current situation of human-computer interaction design of intelligent service robots at home and abroad. The third part designs the human interaction interface of the intelligent service robot with user experience as the center. In the fourth part, the feasibility and reliability of the proposed method are verified by using the comparative design method. The fifth part analyzes the research results and gives a conclusion.

*Corresponding Author

II. Related Work

Robots are high-end emerging technologies that integrate machinery, information, images, materials and other disciplines. Among them, intelligent service robots supported by machine learning, deep learning, artificial intelligence and other technologies have been gradually applied to cleaning, rehabilitation, medical, home services and other fields.

The author in [6] pointed out that the home service robot is a special robot serving human beings, which can replace people to complete home service work, mainly engaged in anti-theft monitoring, security inspection, cleaning, cargo handling, home appliance control, home entertainment, pathological monitoring, children education, alarm clock timing, home statistics, etc. According to the degree of intelligence and the purpose of use, home service robots can be divided into primary small home robots, children's early education robots and human-computer interactive home service robots. The author in [7] pointed out that service robots not only have the characteristics of strong mobility and flexibility, but also can provide a solid hardware foundation for their application in the service industry. They also have the characteristics of convenient mechanical structure, strong human-computer interaction, etc. People can obtain a good user experience in the process of using service robots. The author in [8] proposed a combination of hidden Markov model and Gaussian mixture model to achieve high-precision recognition in the process of dialect interaction. External tools can also be used in the recognition process. This technology has very strong practical value.

The author in [9] tracks interactive targets in real time by improving the efficiency of tracking algorithms. In the case of dialogue analysis, it truly realizes human-computer interaction of intelligent robots by analyzing dialogue video data information. The author in [10] found out how failures occurred and how to solve them in the process of multi-mode voice communication between the interactor and intelligent service robot. Such interdisciplinary work provides opportunities for the communicators and robots to gain new insights into communication problems, so as to provide resources for the mechanisms that can later realize complex human-computer interaction. The author in [11] believed that in the interaction process, the interactor could convey a total of eight positive and negative emotions to the intelligent service robot through touch, and the research results provided a possibility for the follow-up people-centered service. The author in [12] found through experiments that women can communicate emotions for a longer time by using more diversified interaction methods and touching more areas on the robot than male participants. The author in [13] proposed a visual saliency evaluation method of multimedia human-computer interaction interface based on human vision. The adaptive Gaussian filter is used to filter the human-computer interaction interface. The gradient direction of the interface image is converted into the derivative of the horizontal and vertical directions through the Guassian function to determine the size of the interface gradient direction, and the corrected interactive interface image is obtained. According to different co-occurrence concepts of interface image visual saliency, the influence index of human visual saliency is quantified, and the spatial position function of interactive interface is obtained Set the number of direction types and edge points in each image block of the interactive interface, and obtain the texture complexity function of a pixel in each sub image of the interface The weighting coefficient is given, the interface spatial position function and the texture complexity function of pixels are used as evaluation indexes for weighted output, and the evaluation results of visual saliency of interactive interface are given. The author in [14] analyzes the ergonomic criteria and the aesthetic evaluation criteria of interface layout, establishes four basic principles of interface layout: hierarchy, relevance, simplicity and comfort, and constructs a multi-objective optimization mathematical model of interface element layout according to the principles. Based on this model, an improved genetic algorithm is adopted, a multi-objective optimization method of interface element layout based on genetic algorithm is established. The basic principles of product operation interface layout design are given, and a multi-objective optimization method and process of product operation interface element layout based on genetic algorithm are proposed.

However, the methods mentioned above are not comprehensive enough to consider user experience, leading to the failure of user satisfaction to reach the ideal level. Therefore, a new human-computer interaction product interface of intelligent service system based on user experience is designed. Research and creatively put forward the human-computer interaction key points of the intelligent service robot with user experience as the centre, providing technical support and guarantee for the intelligent robot to develop in the direction of intelligence, interaction and personalization.

III. Design of Human-Computer Interaction Product Interface of Intelligent Service System based on User Experience

A. *Hardware Platform Construction of Intelligent Service System based on User Experience*

The introduction of the full name of the Hypertext Markup Language, or HTML, as a way of building and presenting Internet content, will transform the "Web" from a mere tool for presenting content into a full-fledged application platform that will become the standard for a new generation of the Internet. "HTML" 5 adds the ability to interact and present multimedia. Relatively speaking, the user experience has a more specialized level of functionality that allows more personalized design to be satisfied, unlike the H5.

The design of human-machine interface system of intelligent service system based on user experience is shown in Fig. 1.

The framework shown in Fig. 1 is mainly composed of 360PC terminal, gateway module and controller, and the above three parts are briefly introduced respectively. 360 (Program Counter) PC terminal is intended for operation of tablet computers. The gateway module is Bluetooth/RS-485, which is used to integrate Bluetooth and RS-485 chips, so as to achieve the goal of simultaneous Bluetooth wireless communication and RS-485 communication. The controller is programmable and used in PLC industry. Components 1 and 2 are mainly used to realize information exchange under Bluetooth wireless communication. At the same time, the module and the controller achieve bi-directional data communication through Modbus (Remote Terminal Units) RTU, mainly by shielding the communication interface of the module on the controller. Module and RS-485 signal bidirectional conversion, in order to achieve 360PC and (Programmable logic Controller) PLC information exchange, the need for the data link between the two to be set up.

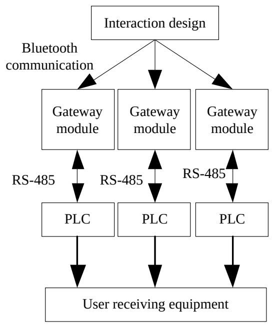


Fig. 1. System platform framework

1) User Experience PC Side:

The UX PC side, also known as epub360, is a very useful tool for H5 interaction design, and is used on Windows platforms, where designers can design professional-level H5 works online without any programming required. But the epub360 idea is the request specialized level function, it may better achieve the designer to the specialized design request.

a) Professional Animation Control. Epub360 is an H5 design tool, is the only SVG path and deformation animation, and control fine sequence frame animation design tools, it can not only relational control, but also a true sense of the control of the interactive animation, is professional.

b) Professional Interactive settings. It combines dozens of Epub360 provided by the trigger control, can be triggered by gestures, can also shake, related control, personalized interaction design, and can be perfect to meet the designer's needs.

c) Professional Social application. Our country is the first to apply WeChat's JsSDK advanced interface, which can complete the functions of personality nickname, avatar picture, friend circle photo, small video, and easily achieve the H5 design of social interaction class such as friend circle comment, click like.

d) Professional Data application. Some advanced data components, such as parameter variables, database, etc., can complete the H5 design of light game, such as small test, puzzle game, etc., and WebApp level professional application can also realize the visual design in the near future.

i) Top toolbar: Add basic and advanced components for the middle section; publish and manipulate the general preview on the left; and operate shortcuts such as save on the right.

ii) List of pages: Overview of the entire page; Add or remove pages; adjust the order of the page; manage the page group.

iii) Canvas: The main design areas are the safety zone inner frame and the bleeding zone outer frame, which are 640 x 960 px and 740 x 1136 px, respectively.

iv) Mobile phone adaptation: The canvas area is specially designed for mobile phone adaptation and consists mainly of an inner and outer frame. Area A is 640 x 960, visible on any phone, and area B is 740 x 1136, and its primary function is to ensure that no white parts are visible on the phone screen.

2) Bluetooth/RS-485 Gateway Module

The role of the module is to carry out Bluetooth and RS-485 two-way signal conversion to achieve the user experience between the PC and PLC can communicate with each other. Gateway module is mainly used for Bluetooth 4.0 and user experience PC in the wireless band, the frequency of 2.4 GHz, and then realizes the operation of Bluetooth 4.0 protocol; Modbus RTU protocol is required through the RS-485 interface to connect with the PLC, the former is wired connection. This module carries on the simultaneous communication between the user experience PC and PLC. The module interprets the data packet sent by the receiving 360 PC terminal, interprets it as the Bluetooth 4.0 protocol, obtains the required and effective data, and then carries out the control instruction or data conversion. If it is necessary for the user to experience the instruction that is to be sent by the PC terminal, it needs to convert it into the Modb protocol, and then transmit it to the PLC, so that the PLC can effectively identify the above instruction. At the same time, the module needs to read, analyze and convert the data sent by PLC and the data collected by PLC according to Modbus RTU protocol, and send the converted data to 360 PC terminal.

B. User Experience

1) Layout Design based on User Sensory Experience

In the process of layout design, we should fully consider the visual flow of the user, follow the left to right and top to bottom order, based on this to highlight the content, so as to attract the user's attention. First, the main menu is set at the top, the left side of the bottom is a secondary menu, the right side is a column link, the middle is the main recommended content, this layout is clearer, but also very convenient to operate, with the advantages and characteristics of symmetrical point of view, so this layout is currently widely used [15-17]. For example, Jing Dong Mall is the way to use screen segmentation for typesetting, to achieve a flexible combination of content, the display of pictures and text to facilitate. At the same time, KnowNet, douban and other websites are also used in this format. Next is the grid multi-column page layout method, this can present more content at the same time in the layout, not only the form is more beautiful, reading is also very convenient, has the very strong maneuverability, quite suits in the online shopping platform foundation demonstration page.

2) Color Design based on the User's Sensory Experience

In the interface design of Internet products, the reasonable combination of colors can give users a very good impression, so as to attract more users to use. In general, the color design of an interface based on user experience needs to consider the following contents: First, the color design needs to have sufficient pertinence, understand the groups that the Internet products face, and understand the habits and emotions of the groups, so as to identify the colors suitable for the products and user experience. For example, if KnowNet is primarily for the sharing of knowledge and experience, then the blue color representing rationality will be used as the primary color of the interface. Secondly, we should have enough comfort in color design, according to the physiology and engineering of human body, according to the physiological characteristics of human vision to choose the color, to reduce fatigue, and improve the comfort of the product.

3) Font Design based on the User's Sensory Experience

In the interface design of Internet products, font plays a key role in information transmission and content description. Whether the design is reasonable or not directly determines the user's reading experience. The font design based on the user experience needs to start from the following aspects: First, for the selection of font, the fonts commonly used in the interface of Internet products include bold type, Song type, round type and calligraphy type. In the selection of font, it is necessary to take into account the user's usual reading habits, but also to combine certain aesthetic habits. In order to make it more convenient for users to find, some content with larger headings will choose bold type, and the remaining text will generally choose bold type or Song type. For example, the major domestic news portal is the application of bold, so that people click and browse. Secondly, the text layout, generally speaking, the interface design of Internet products will involve the font, font size and text color, and so on. It requires designers to make reasonable arrangement of the position of various words, so that they can find the required content more quickly. For example, the use of visual physiology to arrange content in order from high to low purity colors make the user attracted to the color text. Colors can also be used to classify and facilitate user operations [18]. In addition, we can also use word spacing, line spacing and other paragraph settings to highlight the content, so as to greatly simplify the user reading process, shorten the reading time.

C. Speech Recognition based on ARM Processor

Nowadays, there are many kinds of embedded processors, such as MCU, ARM, embedded X86 and so on. The processor is the center of system operation and control, and it tends to adopt RISC instruction set Harvard architecture. ARM (Advanced RISC Machines) is an embedded microprocessor that is also the company's name. In November 1990 ARM was founded in Cambridge, UK. Unlike other semiconductor companies, ARM is an intellectual property (IP) provider, working on intellectual property designs for 16/32-bit RISC microprocessors (reduced instruction set computer, RISC). It does not produce chips by itself, but rather manufactures distinctive chips from partners through the transfer of design proposals. Embedded in full swing today, ARM company has accounted for 75% of the market share. Its microprocessors and technology have been used in all walks of life, almost all kinds of electronic products microprocessors use ARM technology, such as image, consumer entertainment, automobile, security, industrial control, wireless, and mass storage market. In 2001, almost monopolized the global RISC chip market has become the field of RISC chip standards. ARM's success, on the one hand thanks to its unique mode of operation, on the other hand, an important aspect is its own excellent performance. ARM processor has the following characteristics:

1) Low power consumption, small size, low cost, high performance.

a) Support for ARM (32-bit)/Thumb (16-bit) dual instruction sets, with good compatibility with 8-bit/16-bit devices.

2) Using a large number of registers, the instructions are executed faster.

3) A variety of addressing methods, that is flexible and efficient.

4) Fixed instruction length.

Because this paper is based on embedded ARM system intelligent service system human-computer interaction interface design, not a single application system customization, so here is a general pattern design. As a voice interface of service robot, it is usually used by ordinary users, the language is rich and diverse, and the same meaning may have different ways of expression. Moreover, dynamic keyword importing is used to enlarge the recognition range, which is more helpful to set up multiple expressions and make speech recognition more natural.

Fig. 2 is a design flow chart of a common pattern of human-computer interaction in intelligent service system. Both the speech recognition module and the speech synthesis module are encapsulated as an interface (API) that can be invoked by other applications for voice interaction, and the application performs actions based on the return structure of the speech module (i.e., the recognition results).

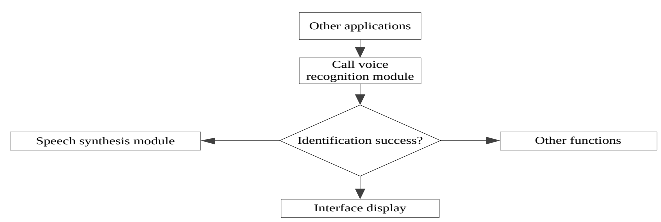


Fig. 2. Flow chart of voice interaction

D. Gesture Recognition based on User Experience

Hand gesture recognition is the basis of human-computer interaction in intelligent service system, which enhances the user experience to extract the features of collected data and complete the recognition according to the features to realize the human-computer interaction. Gesture acceleration features usually include frequency-domain features, time-domain features and other domain features [19-22]. The frequency domain features include the frequency domain entropy, FFT coefficient and other features extracted by using wavelet or Fourier transform; the time domain features include signal amplitude area, statistical maximum value, activity window length, mean value, axis correlation coefficient and standard deviation; the other domain features include the features extracted by using LDA or PCA, etc.

There are $s \times k$ pixels in each gesture motion image, and the values of all pixels are determined by the following formula:

$$\begin{cases} 0 & \text{Point does not pass through the data waveform curve} \\ \sum_{i\in[1,3]} 2^i \times 16 & \text{Point does not pass through the data waveform curve} \end{cases} \tag{1}$$

Where, i represents the number of waveform curves passed by the pixel position. The gesture image is obtained according to the gesture acceleration action as a column in matrix $V_{n\times m}$:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \cdots & \cdots & \cdots & \cdots \\ x_{s,1} & x_{s,2} & \cdots & x_{s,k} \end{bmatrix} \rightarrow \left[x_{1,1} \cdots\ x_{s,1} x_{1,2} \cdots\ x_{s,2} \cdots\ x_{1,k} \cdots\ x_{1,k} \right]^T \tag{2}$$

Transform the data existing in the gesture motion training set to obtain the gesture motion matrix $V_{n\times m}$.

Nonnegative matrix decomposition, which belongs to multivariate data analysis method, uses the product of two low rank nonnegative matrices to replace a nonnegative matrix, and extracts gesture features through nonnegative matrix decomposition method.

The nonnegative matrix algorithm is described by the following formula:

$$V_{n\times m} \approx W_{n\times r} H_{r\times m} \tag{3}$$

In the formula, H represents the weight matrix, W represents the base matrix, V represents the nonnegative matrix, and r represents the rank of the decomposition matrix.

The expression of the optimization objective function is as follows:

$$F = \sum_{i-1}^{n} \sum_{u=1}^{m} [V_{iu} \log(WH)_{iu} - (WH)_{iu}] \tag{4}$$

The iterative algorithm decomposes the base matrix W and the weight matrix H, and obtains the weight matrix $H_{r\times m}$ and the weight matrix $W_{n\times r}$ by decomposing the hand gesture action matrix $V_{n\times m}$.

Input the acquired weight matrix and base matrix into the following classifier to realize gesture recognition and complete human-computer interaction of the intelligent service system:

$$h_j(x) = \begin{cases} 1 & p_j g_j(x) < p_j \theta_j \\ 0 & others \end{cases} \tag{5}$$

In the expression, j represents the feature in the eigenvector; p_j represents the direction of the inequality; $g_j(x)$ represents the eigenvalue of the j rectangle in the sub window to be detected; θ_j represents the threshold of the classifier, the value is 2, $h_j(x)$ represents the classifier.

E. 3D Interactive Image Simulation of Interface based on Fused Texture

When Kinect is used to track and identify moving targets, the problem of target recognition needs to be solved first. However, because Kinect does not have recognition function, it needs to use color histogram for target recognition. At the same time, Kinect cannot directly track human bones. The bone data is acquired by denoising, segmentation and other related technologies on the basis of depth images. Among them, target identification and tracking mainly include virtual target identification and target personnel location tracking, and the specific operation steps are shown in Fig. 3.

In order to obtain better recognition results, the following methods based on image information are mainly used to analyze and research, using Kinect to shoot the color images in the study area, and accurately identify the virtual targets in the study area by color histogram matching results. Among them, color histogram is mainly used to describe the percentage of each color in the whole image. The calculation for color histogram matching is:

$$\begin{cases} d_{corre}\left(H_1, H_2\right) = \dfrac{\sum_i H_1(i) H_2(i)}{\sqrt{\sum_i H_1^2(i) H_1^2(i)}} \\ H_k^{'}(i) = H_k(i) - \dfrac{1}{N}\left(\sum_j H_k(j)\right) \end{cases} \tag{6}$$

In the above formula, $d_{corre}\left(H_1, H_2\right)$ represents the matching result; H stands for component; H_1 and H_2 stand for histogram functions; $H_k(i)$ and $H_k(j)$ represent the matching degree of the i and j histogram respectively. $H_k^{'}(i)$ stands for perfect match value; N represents the number of equal components of H.

The corresponding calculation formula of chi-square matching and intersection matching is as follows:

$$\begin{cases} d_{chis}\left(H_1, H_2\right) = \sum_i \dfrac{\left(H_1(i) - H_2(i)\right)^2}{H_1(i) + H_2(i)} \\ d_{intersection} = \min \sum_i H_1(i) H_2(i) \end{cases} \tag{7}$$

Through the formulas (6) and (7), it can be seen that the larger the value of the calculation result of the correlation and intersection matching algorithm is, the higher the histogram matching degree is, and the perfect matching value is 1; the smaller the value of the calculation result of the chi-square matching algorithm is, the higher the histogram matching degree is, and the value of the perfect matching is 0.

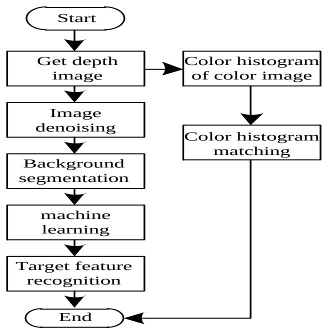


Fig. 3. Flow chart of target identification and tracking

In order to eliminate all the noise in the image, the median filtering method is used to denoise. Among them, the median filter formula is:

$$f(x,y) = \underset{(x,y)\in S_{xy}}{median}\{g(x,y)\} \tag{8}$$

In the above expression, $f(x,y)$ represents the filtered output value, S_{xy} the number of all pixels in the image, and $g(x,y)$ the gray value of the pixel.

To draw a structure image of a virtual object, it is necessary to first separate the moving object from the depth image. The following methods are mainly used for object detection and background segmentation based on user index numbers. The specific operation process is shown in Fig. 4:

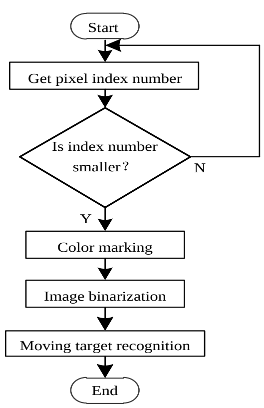


Fig. 4. Operational flow of moving target detection and background segmentation

1) Prioritize the use of Kinect to obtain depth images of each research scenario.

2) Extracting the depth value and user index number of pixels from the depth image.

3) An index number obtained through step (2) marks the moving object.

4) Binary processing shall be used to collect images, and the pixel values in the study area shall be set to 255, and the pixel values of the remaining nodes shall be set to 0, so as to separate the target from the background.

In the early stage of motion recognition, it is necessary to define the starting point and ending point of virtual target motion. Among them, the starting point and the ending point of the moving target are mainly judged by the radius of the palm ball. The radius of the palm ball can be obtained by (Software Development Kit) SDK, and the threshold value is 30mm. When the radius of the center of the hand is higher than the threshold, it indicates that the palm of the target is open, and vice versa, it is closed. If the radius of the palm ball gradually increases from less than 30mm to 30mm, the starting point is indicated, and recognition stops when Leap Motion again captures a threshold of less than 30mm.

The realization of moving target recognition and tracking localization is to recognize each structure accurately in the existing target contour image, and to determine the specific coordinates of the nodes. The above operation process is mainly completed through machine learning technology. Firstly, the sample of decision tree is trained, and then the classification model based on decision tree is obtained. Secondly, the feature value of depth image is classified and evaluated by machine learning. In order to eliminate the influence of external and internal factors on the data noise, the Kalman filtering algorithm is used to eliminate the data's optimal value and measured value iteratively. The Kalman filtering algorithm performs filtering operations on a time-dependent dynamic system, so a discrete control system needs to be added to the algorithm, and the formula is as follows (9):

$$x_{k+1} = F_{k+1}x_k + B_{k+1}u_{k+1} \tag{9}$$

In the above expression, x_{k+1} represents the running state of the system at $k+1$, x_k represents the running state of the system at k, F_{k+1} represents the transformation matrix acting on x_k, B_{k+1} represents the control parameters of the system, the value is 1.5, and u_{k+1} represents the control matrix.

At time $k+1$, the measurements in the system need to meet the following conditions:

$$Z_{k+1} = H_{k+1}x_{k+1} + v_{k+1} \tag{10}$$

In the above formula, H_{k+1} represents the measurement matrix projected into the measurement space by the state at moment x_{k+1}; v_{k+1} stands for measurement noise.

The state at moment $k+1$ can be deduced by formula (10), and the specific calculation formula is as follows:

In the above expression, $\Re(k+1|k)$ represents the optimal value of the system at time $k+1$, and $x(k|k)$ represents the optimal value of the system at time k.

$$\Re\left(k+1|k\right)=\frac{F_{k+1}\cdot\Re\left(k|k\right)}{u\left(k+1\right)} \tag{11}$$

In the phase of Kalman filter prediction, the LEAP SDK is used to obtain the position and pose information of the next period of time. Given the moment, the central coordinate of the moving object is O, and the following formulae can be obtained in combination with the law of motion:

$$\begin{cases} \aleph = v_k t + \dfrac{a_k t^2}{2} \\ p_{k+1} = \Re\left(k|k\right)\cdot\aleph \end{cases} \tag{12}$$

In the above formula, v_k represents the running speed of the moving target; a_k represents acceleration; p_{k+1} represents the coordinate position at time t, and $\aleph$ represents the motion change law of the virtual target.

Based on the above analysis, the moving target structure information is extracted through the depth image, and all the extracted structure information is fused to track and identify the human-computer interaction target of the intelligent service system.

IV. Analysis of Experimental Results

In order to verify the overall effectiveness of the design method of the human-computer interaction interface of the intelligent service system, it is necessary to test the design method of the human-computer interaction interface of the intelligent service system under the enhanced user experience. The experimental platform for this test is Division Mockup, and the probability of data received by the host computer is used as a test indicator to test the design method of the human-computer interaction interface, the design method of the human-computer interaction interface based on the evaluation of visual significance and the multi-objective optimization design method of the layout of product operation interface elements. The higher the probability of data received by the host computer is, the better the human-computer interaction effect is. The test results are shown in Fig. 5.

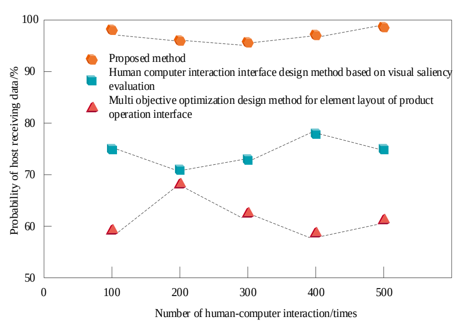


Fig. 5. Test result of probability of receiving data by host

By analyzing the data in Fig. 5, it can be seen that in the process of human-computer interaction, the host receiving data probability of the proposed method is more than 90%, and most of the relevant data can be received to realize human-computer interaction, while the host receiving probability of the human-computer interface design method based on visual significance evaluation and the multi-objective optimization design method of product operation interface element layout fluctuate around 80% and 70% respectively, The data received by the above two methods is not enough to effectively realize human-computer interaction. Comparing the test results of the above methods, it can be seen that the host receiving data probability of the proposed method in the process of human-computer interaction is high, because the method constructs a data transmission model based on game theory, collects relevant data and information in the process of human-computer interaction, and improves the host receiving data probability.

The proposed method, the human-computer interaction interface design method based on visual saliency evaluation and the multi-objective optimization design method of product operation interface element layout are used to recognize human gestures. The feature recognition rate and recognition accuracy are used as test indicators. The test results are shown in Fig. 6 and Fig. 7, respectively.

Through analyzing the data in Fig. 6 and 7, we can see that the recognition rate and accuracy of the proposed method are high in many experiments, which shows that the proposed method can recognize human gestures accurately and comprehensively. When the method of human gesture recognition based on visual saliency assessment is used, the feature recognition rate and feature recognition accuracy are both low, which indicates that the method of human gesture recognition based on visual saliency assessment is not effective and comprehensive.

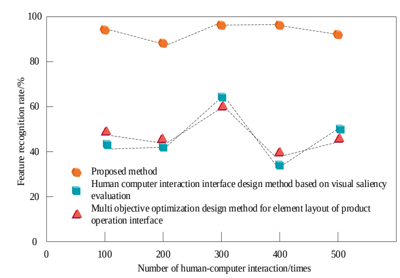


Fig. 6. Test results of feature recognition rate

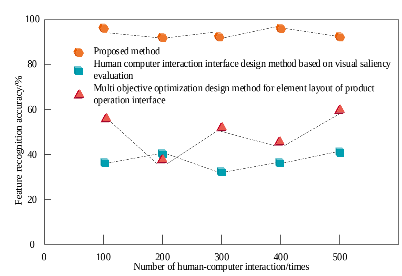


Fig. 7. Test results of feature recognition accuracy

Compared with the test results of the proposed method, the human-computer interaction design method based on visual significance evaluation and the multi-objective optimization design method for product operation interface layout, the proposed method has the best performance in the human-computer interaction, because the proposed method uses the data transmission model to collect the relevant information of the human-computer interaction, adopts the non-negative matrix decomposition method to extract the features of the gesture information, and inputs the features into the classifier to accurately realize the recognition of the gesture, thus improving the feature recognition rate and the feature recognition accuracy.

In the process of human-computer interaction, user satisfaction is the focus of attention. The overall performance of the method is tested by using user satisfaction as a test index. The test results are shown in Fig. 8.

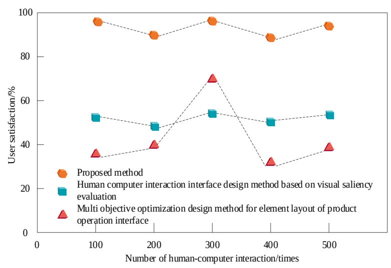


Fig. 8. User satisfaction test results

Through the analysis of the data in Fig. 8, it can be seen that the user's satisfaction with the proposed method is more than 80% in many iterations, and the user's satisfaction with the human-computer interaction design method based on visual significance evaluation and the multi-objective optimization design method for element layout of product operation interface fluctuate around 40% and 60%, respectively. Through the above analysis, the human-computer interaction effect of the proposed method is better, and the user's satisfaction is the highest, because the proposed method constructs the data transmission model through game theory, and can accurately collect the user's static information and dynamic information in the process of human-computer interaction, and feed the fused dynamic information and static information back to the human-computer interaction system. The human-computer interaction system responds to the acquired data, completes the human-computer interaction process of the intelligent service system, meets the user's needs, and thus improves the user's satisfaction.

Fig. 9(a) and Fig. 9(b) refer to the performance of different human-computer interaction methods in indoor and outdoor environments respectively. It can be seen from the figure that the error of the three human-computer interface optimization methods decreases with the increase of the data size. When the data size reaches about 25, it tends to be stable, while the accuracy shows the opposite change rule. In the indoor environment, the convergence errors of the proposed method, the human-computer interaction interface design method based on visual significance evaluation, and the multi-objective optimization design method of product operation interface element layout are 0.156, 0.278, and 0.415, respectively. The accuracy was 97.58%, 95.21% and 92.24%, respectively. In outdoor environment, the convergence errors of the three methods are 0.217, 0.489 and 0.658, respectively. The accuracy was 96.12%, 93.21% and 91.25%, respectively. Therefore, the proposed human-computer interaction method has ideal performance in both indoor and outdoor environments.

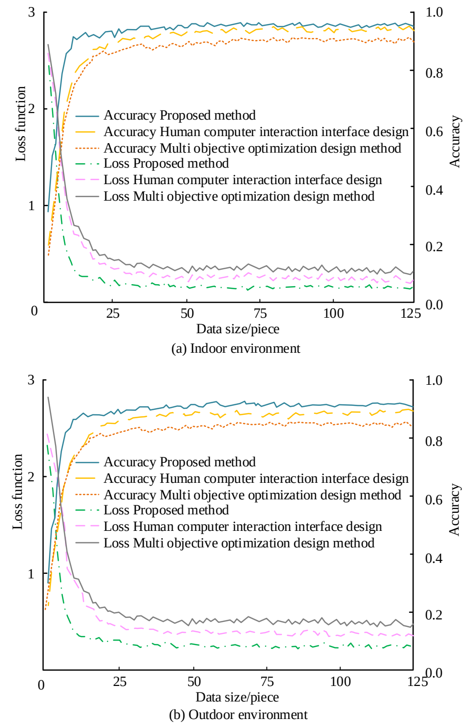


Fig. 9. Performance of different human-computer interaction methods in indoor and outdoor environments

V. Discussion and Conclusion

In order to solve the problems of the current intelligent service platform, such as the low probability of receiving data, the low recognition rate and the low recognition accuracy, a new intelligent service system based on user experience is designed. This introduces the full name of the Hypertext Markup Language and design user experience PC and Bluetooth/RS-485 gateway module. Speech recognition module based on ARM processor is designed to realize human-computer interaction of intelligent service system.

Compared with the human-computer interaction interface design method based on visual significance evaluation and the multi-objective optimization design method of product operation interface element layout, the proposed optimization design method of human interaction interface has higher action recognition accuracy and satisfaction. A large part of tasks completed by computers are completed by human-computer cooperation, which drives the emergence of human-computer interaction systems. However, the traditional human-computer interaction relying on external devices has been difficult to meet the needs of today's society. This has promoted gesture recognition based on machine vision to a certain extent. It is a novel human-computer interaction technology, which has received widespread attention from scholars at home and abroad. Kinect sensor technology is used to realize the tracking and recognition of moving objects. The device includes three functions, namely sensor data flow, bone tracking and advanced audio function. Sensor data stream mainly refers to accessing low-level streams from depth, color camera sensors and microphone arrays; Bone tracking can mainly track the images of the application driven by the postures of 1-2 people in the field of vision; Advanced audio function, which can process some complex audio sound sources, such as echo, noise and beam forming, and can also integrate with the audio recognition library of Windows to collect infrared data and generate unit images. It can promote Kinect bone tracking to be more robust to users' dynamic gestures in different environments. In complex environments, Kinect uses a separation strategy for human segmentation. Its advantage is that it can build a corresponding segmentation mask for each tracked object and retain the human image. This makes it only necessary to transmit the human body image in the subsequent data stream processing, so as to reduce the amount of calculation of body sensing data.

The advantage of Kalman filter is not that its estimation deviation is small, but that it skillfully integrates observation data and estimation data, conducts closed-loop management of errors, and limits the errors to a certain range. Imagine that if there is no information fusion between the two, and only estimation data, the error will accumulate more and more with time, and the uncertainty will increase with time. For the control of landing on the moon so far and for a long time, It will lead to uncontrollable error when arriving near the moon, and the introduction of observation data will correct the estimated data to prevent the error of estimated data from being too large to be spectral. The estimation data fusion of observation data is equivalent to closed-loop feedback management of the former's estimation. It is undeniable that the Kalman filter still has errors, and its advantage is that it can still maintain stable errors when the time is long, because it relies on an observation data in the information source when making decisions. If you listen to one side, you will be dark; if you listen to both sides, you will be clear. As for whether the error is lower than that of the individual estimated data or the individual observed data, further discussion is needed. At the same time, the color histogram used in the research can directly describe the information features of the target through the analysis of local gradient or edge direction density, greatly reducing the impact of local shape features, not only avoiding the overlap of the original feature vectors to a certain extent during the image formation process, but also preventing the impact of individual movement on the detection results, thus ensuring the accuracy of the collected image. It can be seen from the above analysis that the proposed method has higher probability of receiving data, higher recognition rate and accuracy of gesture features, and higher user satisfaction. The above experimental results show that the proposed method has better performance. However, there are still shortcomings in the research. In the process of data acquisition and subsequent action mapping, the time complexity of the algorithm leads to the jumping phenomenon in the action simulation process. Although the optimized Kalman filter eliminates the discontinuity of the action to a certain extent, when the speed becomes faster, the phenomenon of jamming will still occur.

Acknowledgment

Research on the cultivation and improvement path of information teaching ability of normal university students based on professional certification (No.21BY149); Research on the integration of information technology and ideological and political education in universities in the era of fragmentation (No.2022WSYJ100646); Research on strategies for improving the quality of digital online teaching of teachers in the context of new liberal arts (No.202102594042); Research on Digital online Teaching Quality Improvement Strategy(No.2022JYKX12); The 2021 Higher Education Scientific Research Project of Shaanxi Higher Education Association "Research and Practice on the Cultivation System of Normal University Students' Informatization Teaching Ability under the Context of New Liberal Arts" (No.XGH21212).

References

[1] X. Li, J. Tu, H. Zhang, "Game User Experience Measurement Method Based on CV," Packaging Engineering, vol. 40(8), pp. 136-140, 2019.

[2] U. Tandon, R. Kiran, A. N. Sah, "Customer satisfaction as mediator between website service quality and repurchase intention: An emerging economy case," Operations Research, vol. 59(1-2), pp. 155-156, 2019.

[3] Z. Guo, D. Zhou, Q. Zhou, et al., "A hybrid method for evaluation of maintainability towards a design process using virtual reality," Computers & Industrial Engineering, vol. 140(Feb.), pp. 106227.1-106227.14, 2020.

[4] T. Vuletic, "Systematic literature review of hand gestures used in human computer interaction interfaces," International Journal of Human-Computer Studies, 2019.

[5] S. Ahmed, K. D. Kallu, S. Ahmed, et al., "Hand gestures recognition using radar sensors for human-computer-interaction: A Review," Remote Sensing, vol. 13(3), pp. 527, 2021.

[6] D. Belanche, L. V. Casaló, C. Flavián, et al, "Service robot implementation: a theoretical framework and research agenda," The Service Industries Journal, vol. 40(3-4), pp. 203-225, 2020.

[7] J. Hwang, J. Kim, H. Choi, "A review of magnetic actuation systems and magnetically actuated guidewire-and catheter-based microrobots for vascular interventions," Intelligent Service Robotics, vol. 13(1), pp. 1-14, 2020.

[8] E. S. Cross, R. Ramsey, "Mind meets machine: towards a cognitive science of human–machine interactions," Trends in Cognitive Sciences, vol. 25(3), pp. 200-212, 2021.

[9] M. Klumpp, H. Zijm, "Logistics innovation and social sustainability: How to prevent an artificial divide in human–computer interaction," Journal of Business Logistics, 2019, 40(3): 265-278.

[10] B. A. Erol, A. Majumdar, P. Benavidez, et al, "Toward artificial emotional intelligence for cooperative social human–machine interaction,". IEEE Transactions on Computational Social Systems, vol. 7(1), pp. 234-246, 2019.

[11] R. Belk, "Ethical issues in service robotics and artificial intelligence," The Service Industries Journal, vol.41(13-14), pp. 860-876, 2021.

[12] S. Diederich, A. B. Brendel, S. Morana, et al, "On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research," Journal of the Association for Information Systems, vol. 23(1), pp. 96-138, 2022.

[13] S. Wang. "Simulation of visual saliency evaluation method for multimedia human-computer interaction interface," Computer Simulation, vol. 37(3), pp. 161-164, 2020.

[14] H. Kang, S. Yang, S. Deng, et al., "Multi-objective optimization design of product operation interface element layout," Packaging Engineering, vol. 41(8), pp. 149-153,172, 2020.

[15] D Hazer-Rau, S Meudt, A Daucher, et al., "The uulmMAC database—A multimodal affective corpus for affective computing in human-computer interaction," Sensors, vol. 20(8), pp. 2308-2319, 2020.

[16] Y. Li, F. Zou, W. Xie, "Customer satisfaction measurement combining with preference and unbalanced linguistic evaluation," Computer Integrated Manufacturing Systems, vol. 25(1), pp. 165-172, 2019.

[17] T. Guzsvinecz, E. Orbán-Mihálykó,·C. Sik-Lányi, et al., “Investigation of spatial ability test completion times in virtual reality using a desktop display and the Gear VR,” Virtual Reality, vol. 2021(3).

[18] E. Brivio, S. Serino, E. N. Cousa, et al., “Virtual reality and 360° panorama technology: a media comparison to study changes in sense of presence, anxiety, and positive emotions,” Virtual Reality, vol. 2020(3), pp. 1154-1169.

[19] F. G. Pratticò, F. Lamberti, “Towards the adoption of virtual reality training systems for the self-tuition of industrial robot operators: A case study at KUKA,” Computers in Industry, vol. 129(5), pp. 103446, 2021.

[20] E. Petrick, “A historiography of human-computer interaction,” IEEE Annals of the History of Computing, vol. 78(9), pp. 158-69, 2020,

[21] Y. Zhao, P Ruan, X Liu, X Shan, “Study on user satisfaction evaluation based on online comment,” Management Review, vol. 32(3), pp. 179-189, 2020.

[22] M. Klumpp, M. Hesenius, O. Meyer, et al., “Production logistics and human-computer interaction-state-of-the-art, challenges and requirements for the future,” The International Journal of Advanced Manufacturing Technology, vol. 2019(9), pp. 1154-1168.

Age Estimation on Human Face Image Using Support Vector Regression and Texture-Based Features

Jesy S Amelia, Wahyono*
Department of Computer Science and Electronics
Universitas Gadjah Mada, Yogyakarta
Indonesia

***Abstract*—This paper proposed a framework for estimating human age using facial features. These features exploit facial region information, such as wrinkles on the eye and cheek, which are then represented as a texture-based feature. Our proposed framework has several steps: preprocessing, feature extraction, and age estimation. In this research, three feature extraction methods and their combination are performed, such as Local Binary Pattern (LBP), Local Phrase Quantization (LPQ), and Binarized Statistical Image Feature (BSIF). After extracting the feature, Principle Component Analysis (PCA) was performed to reduce the feature size. Finally, the Support Vector Regression (SVR) method was used to predict age. In evaluation, the estimation error will be based on mean average error (MAE). In the experiment, we utilized the well-known public dataset, face-age.zip, and UTK Face datasets, containing 15,202 facial image data. The data were divided into the training of 12,162 images and the testing of 3,040 images. Our experiments found that combining BSIF and LPQ with PCA achieved the lowest MAE of 9.766 and 9.754. The results show that the texture-based feature could be utilized for estimating the age on facial image.**

Keywords—Age estimation; LBP; BSIF; LBQ; MAE; PCA; preprocessing; feature extraction; Support Vector Regression (SVR)

I. Introduction

The face is a part of the human body that has an important role. Characteristics of the face such as eyes, nose, mouth, eyebrows, and wrinkles or aging provide much information. One of them is age estimation. It is possible to limit that under age to accessing content (related to violence) in the web browser through age estimation. Shopping centers can use age estimation to determine advertising strategies based on the age of visitors and provide recommendations. It can choose the right advertising strategy. It is necessary to have a method used to detect the observer's age. We must have an extensive collection of facial images that will be used in age detection and pay attention to the face, such as wrinkles or aging that will be used in the age detection process to produce better estimation.

Generally, the age estimation process has several steps: preprocessing, feature extraction, and age estimation. The first is preprocessing (resizing and grayscale image) to be uniform. Then, the feature extraction step is to find the unique facial features in the age estimation process. The features were the Binarized Statistical Image Feature (BSIF), Local Binary Pattern (LBP), and Local Phrase Quantization (LPQ) methods and their combination: BSIF + LBP, BSIF + LPQ, and LBP+LPQ. In addition, Support Vector Regression (SVR) was used in the age estimation stage. SVR is a machine learning algorithm that provides reliable performance in predicting time series data and can classify data that cannot be separated linearly [1].

Several previous studies have been proposed for age estimation, including Ingole et al. [2] and research by Thrukral et al. [3] and analysis by Khunteta, Ajay, et al. [4]. However, these three studies only predict a group of ages (e.g., with categories 25, 35, 45, 55, 65, and 75 years old), not the age. For this reason, in this study, age estimation was conducted not based on age group but on exactly one age estimation value. Furthermore, the texture of the face is one of the essential features in determining age. This study will detect age on the image containing a single face. The age estimation is carried out only at 1-70 years old due to limited data on ages over 70.

Our research aims to find the best feature extraction method based on texture in detecting age with the lowest Mean Absolute Error (MAE) level from the facial images. Several texture-based feature extraction methods obtain good accuracies, such as LBP, BSIF, and LPQ. The BSIF and LPQ methods provide the best results in the age estimation process with an MAE of 3.58 using the MORPH II dataset and an MAE of 9.07 with the same extraction feature method (BSIF+LPQ) using a Subset of LFW [5]. Meanwhile, LBP is one of the best texture description methods for texture classification, face detection, face recognition, gender classification, and age estimation [6]. In addition, LPQ is proven robust for blur images and is considered to show better performance in texture classification, and the combination method gives better performance results [7].

The paper will be divided into several sections. The second section will explain in more detail how the proposed framework has been applied. The third section will discuss the result of the experiment. Then, the fourth section will conclude our work.

II. Proposed Method

Age estimation is one of the most challenging and crucial issues in utilizing the facial area to produce helpful information. Age estimation can be used in solving scientific problems as well as in subproblems of facial recognition. The prediction of age is formed because the human brain cannot predict age correctly, so it is necessary to have a system that

*Corresponding Author.

can predict age correctly [8]. Many factors can be used in predicting the age of the facial area, including aging, wrinkles on the face, wrinkles in the eye, mouth area, and other features in the facial area. The estimated age itself can be detected using two methods. These methods are classification and regression methods [9]. Age estimation using classification can only estimate based on age range, while age extraction using the regression method can estimate age with exactly one output value of age. It is the difference in age detection using classification and regression methods, as shown in Fig. 1.

Our proposed framework was divided into several significant steps: image preprocessing, feature extraction, and age estimation.

A. *Preprocessing*

Preprocessing is an initial process that aims to improve an image. It is for eliminating noise in the image. In preprocessing stage, we also uniform the image in color and size [10]. The purpose of this process is to make it easier to perform feature extraction at the next stage and to create an age estimation model. This research has two preprocessing steps: resizing and gray scaling.

1) Resizing: Since the dataset contains images with different sizes, the image should be resized to a specific size. The resizing process is used to change the size of the image to either enlarge or reduce size. The resizing function is carried out using the interpolation method. The following is a calculation of the resizing process using the linear interpolation formula:

$$\frac{y-y0}{x-x0} = \frac{y1-y0}{x1-x0} \tag{1}$$

where (x_0, y_0) is two points of initial coordinates, (x_1, y_1) is destination coordinates, and (x, y) is coordinate prediction.

2) Grayscaling: Grayscale is one of the preprocessing steps to change the image from an RGB image to a gray color (grayscale). RGB is an image with three combination colors: Red, Green, and Blue (RGB). The value of RGB is represented in three dimensions XYZ, consisting of lightness, chroma, and hue. The process aims to convert the value of RGB (24-bit) into grayscale value (8-bit) [11]. A grayscale image is an image that has only one value for each pixel. This value is used to indicate the level of intensity. The three colors in the grayscale process are white, black, and gray. The following is a formula for the grayscale process:

$$W = 0.299R + 0.5870G + 0.1140B \tag{2}$$

where W, R, G, and B are grayscale, red, green, and blue values of the image, respectively. Fig. 2 illustrates how the RGB image was converted into a grayscale image.

B. *Feature Extraction*

Feature extraction is a process used to generate unique feature vectors from facial images that will be used to create age estimation models. This feature extraction stage uses three different texture-based methods, including Local Binary Pattern (LBP), Local Phrase Quantization (LPQ), and Binarized Statistical Image Features (BSIF), and their combination, BSIF+LBP, BSIF+LPQ, LPQ+LBP.

1) Local Binary Pattern (LBP): LBP gives the best results in various applications, especially for texture classification, segmentation, face detection, face recognition, gender classification, and age estimation [12]. The LBP method is grayscale invariant and can be combined easily with simple contrast by calculating each gray level of each pixel [13]. LBP considered each pixel has a code. It is called Local Binary Pattern code or LBP code. It works by comparing the value of the center image (pixels) with each neighborhood which amounts to 8 pixels, and thresholding of 3×3 neighborhood [14] divided into eight blocks. The following is a simple example of calculation using LBP at 3×3 pixels:

Given a pixel at (x_c, y_c), the resulting LBP can be expressed in decimal form as follow:

$$LBP = \sum_{P=0}^{7} S(H_P - H_c)2^p \tag{3}$$

where H_c is a central pixel, H_P start from (p = 0, 1…....,7), it is neighborhood from H_c, and S is a function of thresholding, which is defined as:

$$S(H_P - H_c) = \begin{cases} 1 & H_P \geq H_c \\ 0 & H_P < H_c \end{cases} \tag{4}$$

After comparing the center image with its neighborhood, the result will be converted into a binary number. As shown in Fig. 3, a binary number generated from pixel 3×3 is 01010100 = 84. The value will be inserted at the center image. Each value of LBP is formed into 256 histograms which will be used as a texture descriptor.

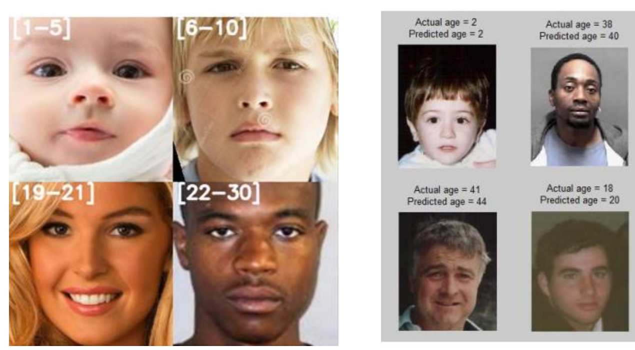


Fig. 1. Age estimation using classification in the left side and regression in the right side.

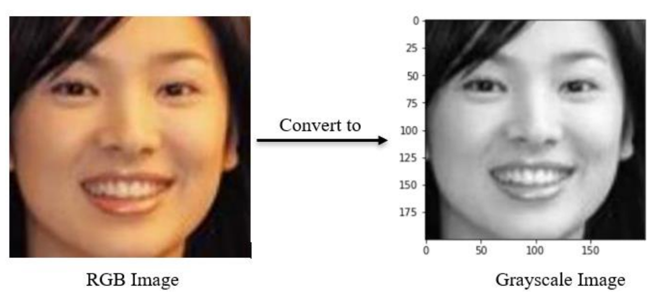


Fig. 2. Image conversion sample from RGB to Grayscale.

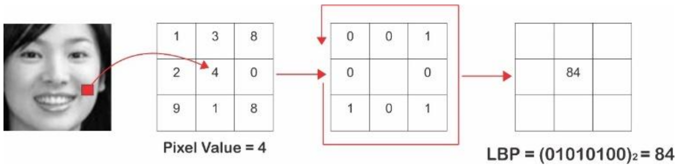
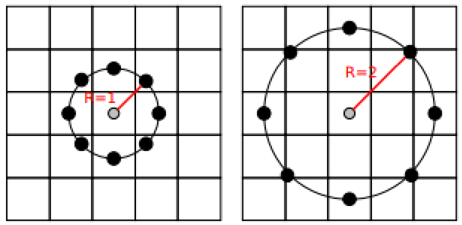


Fig. 3. Example of the basic LBP Operator.



(a) $R = 1, P = 8$ (b) $R = 2, P = 8$

Fig. 4. Circularly symmetric LBP image with different radius.

In addition, LBP has two essential parameters, P and R. P is the total number of neighborhoods, and R is Radius. Radius is measured from the center point distance. There are various kinds of radius in LBP, including 1, 2, 3, etc. Fig. 4 illustrates LBP symmetric sampling with different radius (R) values.

2) Local Phrase Quantization (LPQ): The LPQ descriptor is used in the texture descriptor and classification for blur images [15]. LPQ method is efficiently used to solve the problem of variation expression in the face verification system. LPQ aims to maintain images in local invariant information with various blurring images. Fig. 5 shows the stages of the LPQ. As can be seen in Fig. 5, the q(x) is defined as the following formula:

$$q_j(x) = \begin{cases} 1 \ if \ q_j \geq 0 \\ 0 \ \ if \ q_j < 0 \end{cases} \quad (5)$$

Lastly, the LPQ generates a binary number by combining the pixel values obtained from Re{F×} and Im{F×}. Fig. 5 received a binary number of 89 for the center image.

3) Binarized Statistical Image Features (BSIF): Inspired by LBP and LPQ, Kannala et al. [16] proposed a new local descriptor called BSIF (binarized Statistical Image features). The basis vectors of a subspace into which local image patches are linearly projected are obtained from images by using Independent Component Analysis (ICA). The coordinates of each pixel are threshold, and thus a binary code is computed. A value represents the local descriptor of the image intensity patterns in the neighborhood of the considered pixel [17]. The following is the formulation of BSIF:

$$S_i = \sum_{v,u} W_i(i,v) * X(u,v) = W_i^{\ T} x \quad (6)$$

where X is image with size of M×M, *Filter* W_i learning from *Independent Component Analysis* (ICA) by maximizing the statistical independence of S_i.

$$b_i(x) = \begin{cases} 1 & S_i \geq 0 \\ 0 & S_i < 0 \end{cases} \quad (7)$$

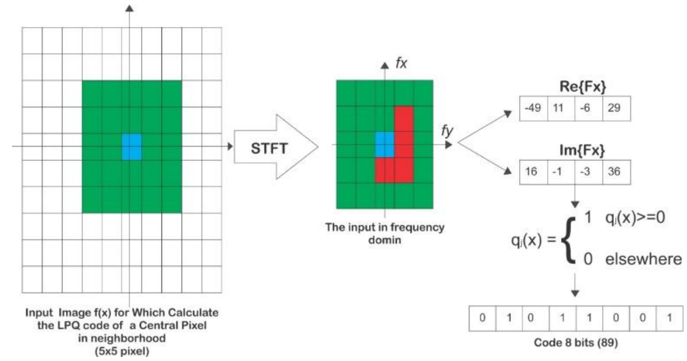


Fig. 5. Example of the LPQ Operator [10].

where b_i is a binary number obtained from S_i. In all our experiments, the BSIF descriptor has been used with filters of size 9×9 and 8-bit.

C. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical method to distinguish patterns and signal processing. PCA reduces dimensional data, feature extraction, and facial recognition [18]. PCA is a powerful method, especially for high-dimensional data. That technique reduces the dimensions of the dataset by extracting the essential components.

D. Age Estimation Using Support Vector Regression (SVR)

Support Vector Regression (SVR) is a supervised learning algorithm to predict discrete values. SVR can solve linear and non-linear problems using kernel functions [19]. The advantage of SVR is that the computational complexity of SVR does not depend on the dimensions of the input space. In addition, SVR has excellent generalization capabilities and high prediction accuracy.

The SVR method is effective for solving problems related to the estimation function. So SVR is the best method for age estimation. SVR is the same as SVM. The idea of SVR is to determine the best fit line. In SVR, the most appropriate fit line is the hyperplane which has the maximum number of points. So, the SVR tries to match the best line within its threshold value. The threshold value is the distance between the hyperplane and the boundary line. So, the regression problem is to determine a function that approximates the mapping from the input domain to real numbers based on the training sample.

E. K-Fold Cross Validation

Cross Validation is a testing method in data mining and machine learning where the dataset will be divided into k parts or folds randomly. Cross Validation is used to evaluate the performance of the algorithm. The data will be separated into two subsets: the learning subset (training) and the validation/evaluation subset [20]. First, the algorithm will be trained using a learning subset, and then the model will be validated using a validation subset. The illustration of the cross-validation scheme is shown in Fig. 6.

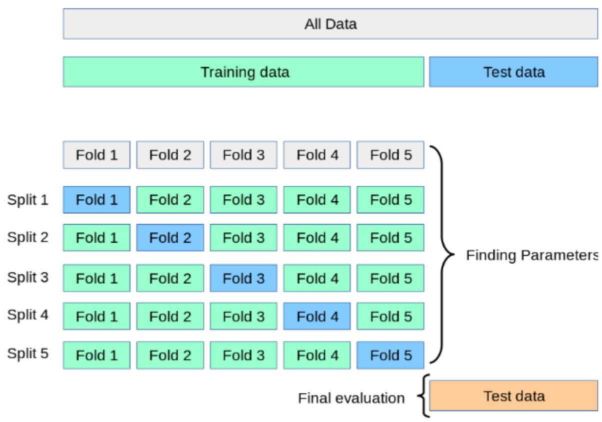


Fig. 6. Cross-validation illustration.

III. Experimental Result

In this section, we first introduce the databases and the protocol evaluation of our experiments. After that, we will describe experimental setups and their results.

A. Dataset

The facial dataset used in this research was downloaded from face-age.zip [12]. The facial image size is 200×200 pixels with a total of 9,778 RGB face images. In this face-age.zip face image, there are facial images with an age range of 1-110 years, but this research uses face images from 1-70 years of age with the female and male gender. First, a selection process is carried out on the facial image, only selecting a face without glasses and a watermark. In addition, we only utilized images with ages between 1-70 years due to data limitation in the face image dataset from face-age.zip for ages above 70 years. From the face-age.zip dataset with the selection process, we choose a total of 3,040 images for our experiment.

Furthermore, the face image dataset was added using the UTK Face dataset. We then selected 12,162 face images from this dataset. The UTK Face dataset is also obtained from downloading through the same source on the face-age.zip dataset, and the total facial image dataset owned is 15,202 images. Finally, the dataset is divided into training and testing data, where the training data is 12,162 face images, and the testing data is 3,040 facial images (20% of the total facial images) which contain facial images 1-70 years old for training and data testing. Table I shows the selected images used in our experiment.

B. Experiment Scenario

The scenario of the system is divided into two main stages: training and testing. The training stage is applied to find the relationship between one face and another and create a model for age estimation. In contrast, the testing stage is performed to detect the estimated age of the face image. Fig. 7 shows the scenario of the experiment for our proposed framework.

In this research, to produce the best age estimation model, experiments were carried out on feature extraction methods based on texture, such as LBP, LPQ, BSIF, and the combination of these methods, BSIF + LBP, BSIF+LPQ, and LPQ + LBP. From the three extraction methods and the combination, it will be seen which method will produce the lowest MAE level in the age estimation process. In addition, there are several experiments: feature reduction extracted using PCA with values of 40, 50, 60, 70, 80, and 90 features. For age estimation using the SVR method.

Since the estimation output will not be in discrete values, for evaluation purposes, we rounded the estimation output using two strategies. The first strategy is rounding the value up from the detected age. For example, the detected ages 32.5, 17.8, and 23.2 will be 33, 18, and 24. The second strategy is rounding the age value down from the detected age. For example, the detected ages 32.5, 17.8, and 23.2 will be 32, 17, and 23.

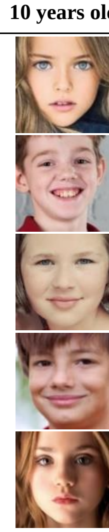
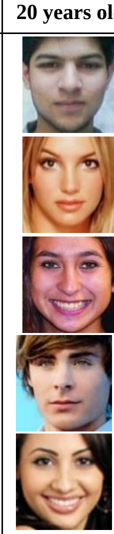
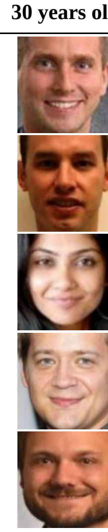
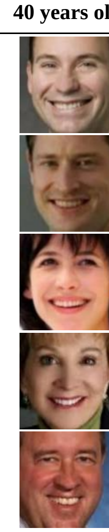
C. Evaluation Matrics

To measure the performance of our approach, we adopted Mean Absolute Error (MAE). MAE is a method used to evaluate the performance of age estimation. MAE is the average absolute error between the ground truth age and the predicted one [5]. The age estimation model can be measured by using MAE. The following is the formulation of MAE:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - G_i| \quad (8)$$

where P_i is the estimated age, G_i is the corresponding ground truth, and N is the total number of samples.

TABLE I. The Sample Images in the Dataset

	10 years old	**20 years old**	**30 years old**	**40 years old**
Age				

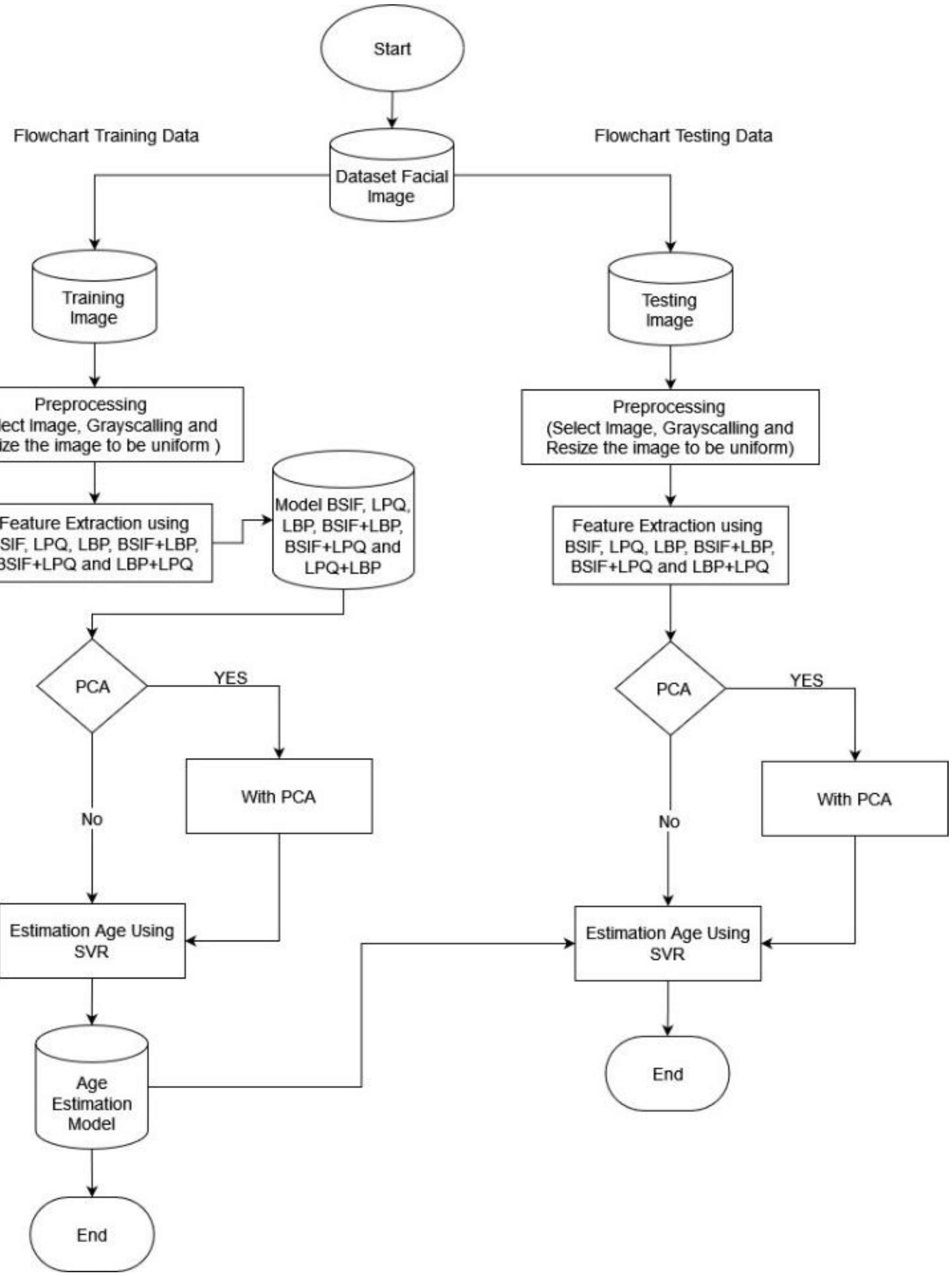


Fig. 7. General process of the proposed framework.

D. Results and Discussion

1) Experiment scenario: In this study, age estimation was carried out with various scenarios of experiments as follows:

a) Resizing the image: 180×180 pixels, 160×160 pixels, 140×140 pixels, and 120×120 pixels.

b) Feature Extraction using different methods and the combination (LBP, LPQ, BSIF, BSIF+LBP, BSIF+LPQ, and LPQ+LBP).

c) LBP Feature extraction using different radius (radius = 1, radius = 2, radius = 3 with same P (total number of neighborhood).

d) Using the first strategy and second strategies for the age estimation, as explained in section III.B.

e) Comparing MAE from extracted features without PCA and with PCA.

f) Tried several variations of PCA, such as 40,50,60,70,80, and 90 features and compare MAE.

g) Find the optimum parameter of BSIF Filter by ICA.

2) Evaluation with different images scale and without PCA: In this paper, experiments were carried out on various image sizes such as 180×180, 160×160, 140×140, and 120×120 with the first strategy and second strategy of age estimation using LBP, LPQ, BSIF, BSIF+LBP, BSIF+LPQ, and LPQ+LBP. Fig. 8 and 9 show the MAE results when applying the proposed framework without using PCA on several image sizes. From these results, it can be concluded that by resizing images to 180×180, 160×160, 140×140 and 120×120 with feature extraction without PCA, the lowest MAE when using the BSIF method with an image size of 180×180 of 10.612 for the first strategy and 10.600 for the second strategy for age estimation. Furthermore, from the results, we could also find that in most scenarios, if the image is resized to a smaller size, the MAE decreases for LBP, LPQ, and BSIF. However, the results are not better compared to the combination of features.

3) Evaluation of LBP using different radius without PCA: In this study, experiments were also carried out on the radius of the LBP. Radius tested are 1, 2, and 3 to obtain the best age detection results with the lowest MAE. As shown in Fig. 10 and 11, it is found that the lowest MAE in age detection using LBP radius 3. Therefore, the LBP with a radius of three will be utilized in all experiments.

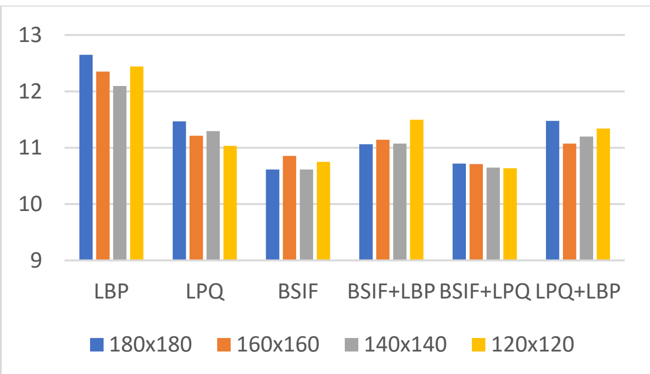


Fig. 8. Comparison results of different image size without using PCA in the 1st strategy.

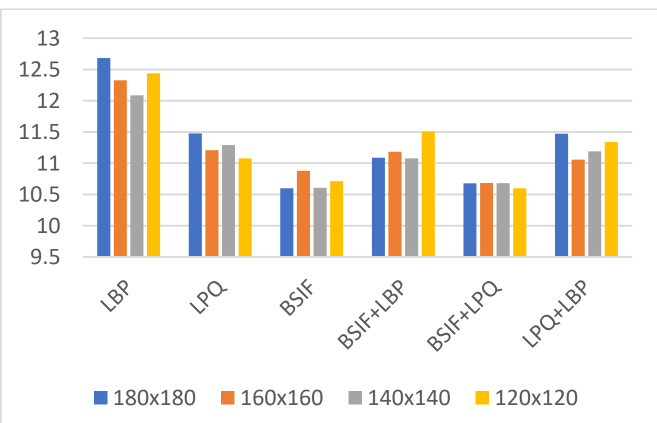


Fig. 9. Comparison results of different image size without using PCA in the 2nd strategy.

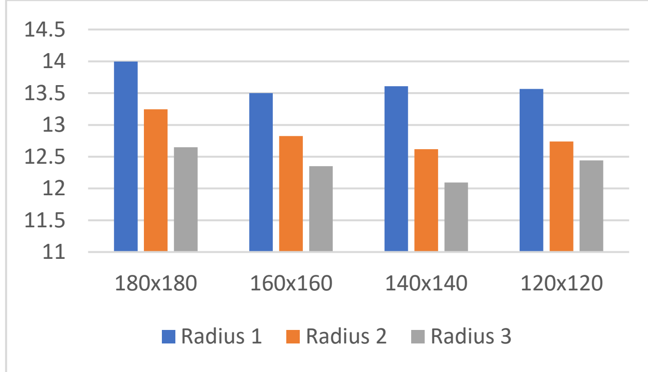


Fig. 10. Comparison results of LBP radius without using PCA in the 1st strategy.

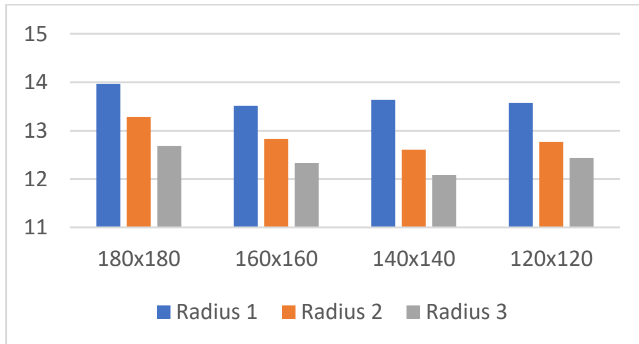


Fig. 11. Comparison results of LBP radius without using PCA in the 2nd strategy.

4) Optimal parameter evaluation for LPQ: Feature extraction on Face recognition uses LPQ with various window sizes: 5, 7, and 9, and the best results were obtained using the LPQ method with window size 7 [10]. For this reason, this study used a window size of 7.

5) Optimal parameter evaluation for BSIF: Parameters that can affect the BSIF method are texture filters. Research for Face Recognition using texture filters BSIF with a size of 7×7, 9×9, and 11×11 in 8 bits and the best result using 7×7 [10]. Meanwhile, in the other research, the BSIF descriptor has been used with filters of the size of 13×13 and 8 bits for age estimation [5]. So, we will test texture filters of size 7×7, 9×9, and 13×13 with 8 bits. From Tables II and III, it can be found that BSIF Texture Filter 9×9 8 bits has the lowest average MAE. This texture filter will be used in the entire experiment.

TABLE II. BSIF FILTER COMPARISON WITHOUT PCA IN 1ST STRATEGY

Image Size	MAE Values		
	BSIF 7×7	**BSIF 9×9**	**BSIF 13×13**
180×180	10.771	10.612	11.015
160×160	10.997	10.855	10.627
140×140	10.442	10.612	11.269
120×120	10.715	10.749	11.063
Average	10.731	10.707	10.994

TABLE III. BSIF FILTER COMPARISON WITHOUT PCA IN 2ND STRATEGY

Image Size	MAE Values		
	BSIF 7×7	BSIF 9×9	BSIF 13×13
180×180	10.762	10.6	11.004
160×160	10.982	10.877	10.634
140×140	10.402	10.607	11.272
120×120	10.719	10.714	11.07
Average	10.716	10.700	10.995

6) Evaluation age estimation based on extraction feature: We use several variations of PCA feature vector sizes, such as 40, 50, 60, 70, 80, and 90, with various image sizes of 180×180, 160×160, 140×140, and 120×120 pixels with optimal parameters from LBP, LPQ, BSIF, and the combination BSIF+LBP, BSIF+LPQ, and LPQ+LBP. However, based on the result of our experiment, we will only display the results at 160×160 pixels because this image size produces the lowest MAE. Tables IV and V show the comparison results of several feature extraction methods for estimating. Based on these tables, it was found that the lowest MAE is generated using BSIF+LPQ with PCA 70. It achieved an MAE value of 9.766 in the first and 9.754 in the second strategies.

7) Age estimation of public figures: After obtaining the best model from the previous experiment, we then utilized the model for estimating the age of two public figures: Ralph Fiennes and Indro Warkop. We utilized the BSIF+LPQ model with an image size of 160×160. Due to limited data on these two public figures, the PCA was not performed. Fig. 12 and 13 show the age estimation of these public figures at different ages. Our model achieved MAE of 10.428 and 11.142 at the first and second strategies for Ralph Fiennes. On the other hand, the model achieved an MAE of 12.142 and 13.142 in the first and second strategies for Indra Warkop.

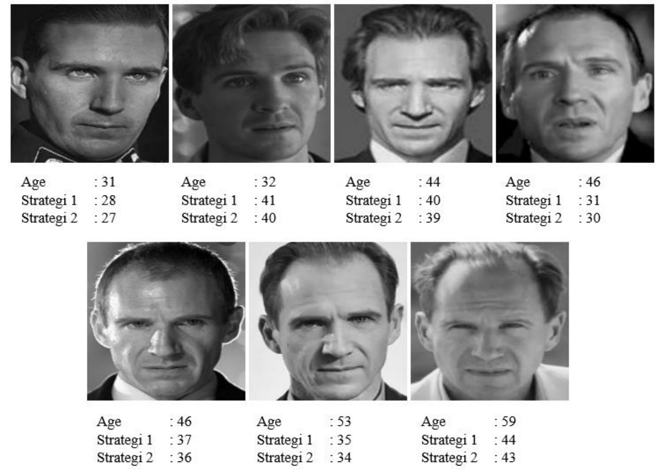


Fig. 12. Estimated age of Ralph Fiennes.

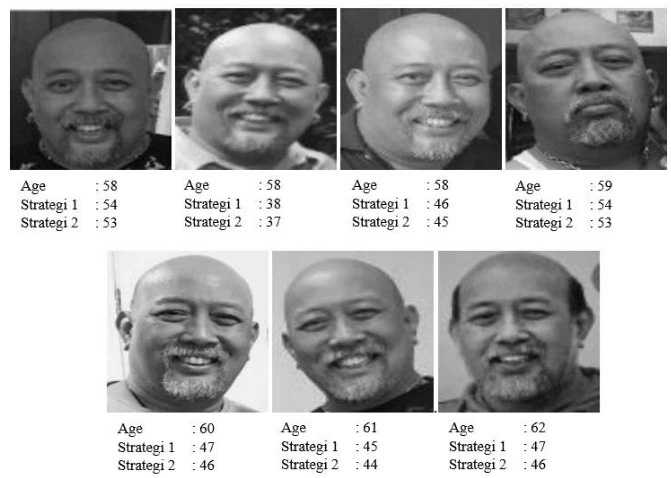


Fig. 13. Estimated age of Indro Warkop.

TABLE IV. MAE COMPARISON WITH IMAGE SIZE OF 160×160 PIXEL IN 1ST STRATEGY

Method	Without PCA	PCA 40	PCA 50	PCA 60	PCA 70	PCA 80	PCA 90
LBP	12.350	11.632	11.674	11.586	11.449	11.762	11.736
LPQ	11.297	10.986	10.635	10.709	10.95	10.696	10.672
BSIF	10.855	10.411	10.448	10.326	10.106	10.578	10.599
BSIF+LBP	11.140	10.678	10.692	10.695	10.698	10.482	10.468
BSIF+LPQ	10.711	10.323	9.953	10.197	9.766	10.344	10.134
LPQ+LBP	11.073	10.507	10.486	10.506	10.662	10.861	10.759

TABLE V. MAE COMPARISON WITH IMAGE SIZE OF 160×160 PIXEL IN 2ND STRATEGY

Method	Without PCA	PCA 40	PCA 50	PCA 60	PCA 70	PCA 80	PCA 90
LBP	12.327	11.628	11.628	11.598	11.448	11.749	11.720
LPQ	11.274	11.018	10.620	10.707	10.921	10.706	10.699
BSIF	10.877	10.387	10.436	10.334	10.087	10.642	10.622
BSIF+LBP	11.184	10.718	10.733	10.720	10.680	10.471	10.490
BSIF+LPQ	10.683	10.333	9.968	10.226	9.754	10.345	10.157
LPQ+LBP	11.057	10.515	10.486	10.495	10.651	10.833	10.769

IV. Conclusion

In this research, the age estimation framework using Support Vector Regression (SVR) with texture-based feature extraction has been successfully implemented. The proposed framework can detect the exact age, not based on the age range. Furthermore, a comprehensive experiment has been carried out to determine the optimal model of age estimation, such as the different sizes of the image, utilization of PCA, and applying several feature extractions and their combination. Based on the experiment, we have found several essential findings in constructing of age estimation model. First, image size 160×160 produced the lowest MAE level compared to image size 120×120, 140×140, and 180×180 pixels. Second, utilizing the PCA after feature extraction gains computation time fast and gives the lowest MAE level in age estimation. Third, the optimal parameters for LBP, BSIF, and LPQ are a radius equal to three, a texture filter size of 9×9 8-bits, and a window size set equal to seven. Finally, the optimal model which produces the lowest MAE is the combination of BSIF and LPQ methods with a PCA dimension of 70. The model obtained MAE values of 9.766 and 9.754 for the first and second strategy is 9.754, respectively. However, the model has limitations for estimating the age of facial images from Asian countries since the dataset used for the experiment is mostly from Western countries. Therefore, future research might consider the additional dataset covering various ethnic groups. In addition, we may use another method like ANN or CNN and compare that with this method for age estimation.

Acknowledgment

The research and publication were funded through the Final Assignment Recognition Program (Program Rekognisi Tugas Akhir) Batch II Year 2022, Universitas Gadjah Mada with Assignment Number 633/UN1.P.III/KPT/HUKOR/2022 and 5722/UN1.P.III/Dit-Lit/PT.01.05/2022.

References

[1] T. Alnuayri, A. L. H. Martinez, S. Khursheed, and D. Rossi, “A Support Vector Regression based Machine Learning method for on-chip Aging Estimation,” in Proceedings - 2021 IEEE 4th International Conference on Computing and Information Sciences, ICCIS 2021, 2021. doi: 10.1109/ICCIS54243.2021.9676376.

[2] Ashwini L. Ingole and Kailash J. Karande, Automatic Age Estimation from Face Images Using Facial Features. 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), 2018.

[3] P. Thukral, K. Mitra, and R. Chellappa, “A hierarchical approach for human age estimation,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2012, pp. 1529–1532. doi: 10.1109/ICASSP.2012.6288182.

[4] R. Gupta, M. Tech, and A. Khunteta, “SVM Age Classify based on the facial images,” 2012.

[5] M. E. nadhir Zighem, A. Ouafi, A. Zitouni, Y. Ruichek, and A. Taleb-Ahmed, “Two-stages based facial demographic attributes combination for age estimation,” J Vis Commun Image Represent, vol. 61, pp. 236–249, May 2019, doi: 10.1016/j.jvcir.2019.03.025.

[6] S. N. Gowda, “Age estimation by LS-SVM regression on facial images,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 10073 LNCS, pp. 370–379. doi: 10.1007/978-3-319-50832-0_36.

[7] B. Yuan, H. Cao, and J. Chu, “Combining local binary pattern and local phase quantization for face recognition,” in Proceedings - 2012 International Symposium on Biometrics and Security Technologies, ISBAST 2012, 2012, pp. 51–53. doi: 10.1109/ISBAST.2012.14.

[8] R. Gupta, M. Tech, and A. Khunteta, “SVM Age Classify based on the facial images.”

[9] S. E. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi, “Facial age estimation using BSIF and LBP.” [Online]. Available: http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html.

[10] M. Belahcene, M. Laid, A. Chouchane, A. Ouamane, and S. Bourennane, “Local descriptors and tensor local preserving projection in face recognition,” in Proceedings of the 2016 6th European Workshop on Visual Information Processing, EUVIP 2016, Dec. 2016. doi: 10.1109/EUVIP.2016.7764608.

[11] C. Saravanan, “Color image to grayscale image conversion,” in 2010 2nd International Conference on Computer Engineering and Applications, ICCEA 2010, 2010, vol. 2, pp. 196–199. doi: 10.1109/ICCEA.2010.192.

[12] S. E. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi, “Facial age estimation using BSIF and LBP.” [Online]. Available: http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html

[13] T. Ojala, M. Pietikainen, and D. Harwood, “A Comparative Study of Texture Measures with Classification Based on Feature Distributions,” 1996.

[14] P. Khoi, L. H. Thien, and V. H. Viet, “Face Retrieval Based On Local Binary Pattern and Its Variants: A Comprehensive Study,” 2016. [Online]. Available: www.ijacsa.thesai.org.

[15] S. R. Zhou, J. P. Yin, and J. M. Zhang, “Local binary pattern (LBP) and local phase quantization (LBQ) based on Gabor filter for face representation,” Neurocomputing, vol. 116, pp. 260–264, Sep. 2013, doi: 10.1016/j.neucom.2012.05.036.

[16] Kannala Juho and Esa Rahtu, BSIF: Binarized statistical image features. Tsukuba, Japan: 21st International Conference on Pattern Recognition (ICPR 2012), 2012.

[17] S. E. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi, “Facial age estimation using BSIF and LBP.” [Online]. Available: http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html.

[18] O. R. Shahin, R. Ayedi, A. Rayan, R. M. Abd El-Aziz, and A. I. Taloba, “Human Face Recognition from Part of a Facial Image based on Image Stitching.” [Online]. Available: www.ijacsa.thesai.org.

[19] S. Kavitha, S. Varuna, and R. Ramya, “A comparative analysis on linear regression and support vector regression,” in Proceedings of 2016 Online International Conference on Green Engineering and Technologies, IC-GET 2016, May 2017. doi: 10.1109/GET.2016.7916627.

[20] X. Yao, “Application of Optimized SVM in Sample Classification,” 2022. [Online]. Available: www.ijacsa.thesai.org.

A Fast and Effective Method for Intrusion Detection using Multi-Layered Deep Learning Networks

A. Srikrishnan[1], Dr. Arun Raaza[2], Dr. Ebenezer Abishek. B[3]
Dr. V. Rajendran[4], Dr. M. Anand[5], S. Gopalakrishnan[6], Dr. Meena. M[7]
Research Scholar, Department of ECE, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India[1]
Deputy Director of CARD, Department of ECE, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India[2]
Associate Professor, Department of ECE Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India[3]
Director and Professor, Department of ECE, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India[4]
Professor, Department of ECE, Dr.M.G.R. Educational and Research Institute, Chennai, India[5]
Assistant Professor, Department of MEE, Sengunthar College of Engineering, Tiruchengode[6]
Associate Professor, Department of ECE, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India[7]

***Abstract*—The practise of recognising unauthorised abnormal actions on computer systems is referred to as intrusion detection. The primary goal of an Intrusion Detection System (IDS) is to identify user behaviours as normal or abnormal based on the data they communicate. Firewalls, data encryption, and authentication techniques were all employed in traditional security systems. Current intrusion scenarios, on the other hand, are very complex and capable of readily breaching the security measures provided by previous protection systems. However, current intrusion scenarios are highly sophisticated and are capable of easily breaking the security mechanisms imposed by the traditional protection systems. Detecting intrusions is a challenging aspect especially in networked environments, as the system designed for such a scenario should be able to handle the huge volume and velocity associated with the domain. This research presents three models, APID (Adaptive Parallelized Intrusion Detection), HBM (Heterogeneous Bagging Model) and MLDN (Multi Layered Deep learning Network) that can be used for fast and efficient detection of intrusions in networked environments. The deep learning model has been constructed using the Keras library. The training data is preprocessed and segregated to fit the processing architecture of neural networks. The network is constructed with multiple layers and the other required parameters for the network are set in accordance with the input data. The trained model is validated using the validation data that has been specifically segregated for this purpose.**

Keywords*—Intrusion detection system; knowledge discovery and data mining; transmission control protocol; adaptive parallelized intrusion detection; constrained-optimization-based extreme learning machine

I. Introduction

IDS models can serve a wide range of purposes and requirements when applied in business settings. One of the most popular applications is the method of intrusion detection in personal systems or distributed settings [1]. The design of modern operating systems includes the implementation of technology that detect and prevent intrusions. However, the handling capabilities of these systems are currently unknown. As a consequence of this, the majority of customers choose to invest in expert intrusion detection solutions for enhanced levels of protection. In addition, there is a considerable need for IDS that may be implemented in clustered systems and used in servers [2, 3]. There are many commercially available intrusion detection systems, some of which include the Bro intrusion detection system, which was developed by VISTAS Labs and the School of Engineering, the Snort intrusion detection system, which is distributed under the GNU licence [4], Network Protocol Analyzer [6], Multi Router Traffic Grapher (MRTG) [7], and a few other options. On the other hand, the computing requirements of the majority of these systems, as well as their accuracy, might be enhanced.

A. Motivation of this Research

This research was motivated by the fact that the majority of currently available intrusion detection systems do not handle the issues listed above as part of their operational process. As a result, there is a need to design an effective intrusion detection system with mechanisms to handle data imbalance and concept drift while simultaneously achieving better accuracy and faster detection of intrusions. This research was carried out to fulfil this need.

B. Objectives

The primary purpose of this work is to create an efficient intrusion detection system that is capable of the identification of intrusion signatures in network data in real time. The secondary goals are as follows:

- To develop a method of intrusion detection that is capable of dealing with the inherent data imbalance that is present within the domain in an efficient manner.
- In order to effectively deal with concept drift, which is an essential component of the domain.

- To incorporate feature reduction in order to lessen the demands placed on the model's computational resources.
- To enable quicker detection of intrusions, parallelization is going to have to be incorporated into the detection process.
- To execute detection of intrusions in real time in order to minimise financial damage as much as possible.

II. Literature Survey

One of the essential components of today's continuously networked world is the presence of an intrusion detection system. As a result, there have been a few examination promises made in this area. In this section, we will discuss what are the most recent commitments in the field of intrusion detection. The investigation of the models will be carried out in three important stages. "The primary section discusses the function of AI-based models in identifying intrusions, the subsequent section investigates the significance of component choice in this space, and the concluding section discusses the function of adaptable models in the field of intrusion detection. H. Yang et al. [8] presented a novel and factual model that makes use of Least Square Help Vector Machines (LS-SVM) in order to recognise intrusions. This model divides the data into subgroups that are not consistent with one another. A delegate test will be selected from within this subgroup in order to prepare the model. SVM-based intrusion detection systems are incorporated into another comparative SVM-based model for network intrusion detection models [9]. The most important AI methods have been implemented in a number of the models, and those models also demonstrate intriguing expectations. These models include hereditary and fluffy calculation based models [10], grouping and k-closest neighbour based models [11], IDS utilising Backing Vector Machines (SVM) for preparation [12, 13], and a lot of other similar models. These models additionally rely on signature-based detection of intrusions in order to function well. In most cases, they are made as twofold classifiers, and they are prepared on both typical and irregular marks. Typical marks are the ones that are used. It was determined that the models were computationally incomprehensible, which resulted in significant time requirements. A model of IDS with several layers was suggested [14]. This is a component choice based model, which considers attacks to be layers and selects highlights for each of the layers in order to construct the detection model. It was proposed [15] to use an ongoing-based irregularity detection model for the purpose of network intrusion detection. Keeping up with adaptive mark databases that are not difficult to renew and reproduce under continuous settings is essential to this concept. In addition to this, the model suggests a multi-objective component determination method for the practical selection of attributes that will result in increased levels of precision. An earlier version of this model that deals with the detection of intrusions on systems that have been implanted was proposed [16]. The study [17] presents a proposal for a staggered intrusion detection paradigm that is based on peculiarity detection. A comparable methodology for the detection of peculiarities based on trees was suggested [18]. In order to identify intrusions, this model relies on a combination of calculations based on the Firefly and Hereditary algorithms. There was a proposal made for a grouping-based intrusion detection approach [19]. The Semi-Directed Multifaceted Grouping Model (SMLC) that was proposed in this work makes use of named data in some capacity for the preparation process, which enables an adjustable detection procedure. Other semi-directed intrusion detection models include a group-based IDS Al- [21], a normal neighbour based model [22], and a semi-regulated model [20].

The models make use of fundamental techniques, which leads to diminished performance when applied to unbalanced data. The author [23] made a suggestion for a model that handled imbalances in the IDS. This is a real-time model that is based on clustering and use the RIPPER algorithm for the detection procedure. The research [24] presented a model with a similar structure that was based on the RIPPER algorithm. The study [25] presented an idea for an intrusion detection model that was based on principal component analysis (PCA). This paradigm is a profiling-based one, and it constructs profiles by making use of the intrusion signatures. Because this model is based on several classifications of classes, it is intended to recognise a wide variety of intrusion signs so that it can provide accurate categorizations. The study [26] presented a proposal that included an in-depth investigation of the classification models that can be utilised for intrusion detection in the most efficient manner.

An efficient methodology for the selection of features has been proposed [27] for use with the KDD CUP 99 dataset. The programme was able to make fairly accurate predictions despite having only six characteristics to work with. In a similar vein, the Flexible Neural Trees model [28] was able to attain an accuracy of 99.19% with just four features. A model for the identification of intrusions that was written in A C# was suggested [29]. Using this method results in the creation of a packet sniffer that has the capacity to effectively gather packets from an interactive TCP session and inspect them. Attackers will frequently engage in packet chaffing whenever they are dealing with models of packet sniffers. Researchers can more easily identify these types of packets with the help of the model, which works by injecting more packets into the network. The study [30] presented a strategy for identifying stepping stone intruders in their research. This model is responsible for carrying out the process of intrusion detection by contrasting the contents of a host's incoming and outgoing traffic packets. When the contents are examined, the model is evaluated to determine whether or not it can serve as a stepping stone. Because of this, it is possible to make a clear distinction between a typical packet and an invasive packet. On the other hand, if the packets are encrypted, this paradigm for detecting intrusions may not be successful. Because of this, inspecting packets cannot be called a model that is 100 percent reliable for spotting invasions. The research [31] presented a model for the identification of intrusions that was very comparable. The stepping stone attack has also been suggested as being detectable by using this concept. In contrast, this model identifies the packet source based on the timestamp, size, and sequence number of the data packets rather than by inspecting the contents of the packets themselves. Even after they have been encrypted, these parameters are still legible; hence, this model is seen to have superior performance when compared to the model that came before it. Examining the information included within the packet header was also the subject of a suggestion made by [32]. Both [33] and [34] presented an additional method that makes use of the information regarding the packet count in order to identify the stepping stone assault. When it comes to fending off stepping stone attacks, connection chain difficulties are seen as being among the most important components. The author [35] presented a model that determines the stepping stone attack by estimating the length of the connecting chain as a starting point for the calculation. The study [36] presented a model with the aim of precisely determining the connecting chain more of the time. Even while these methods assist cut down on the amount of time needed for training and detection, because they are unable to properly deal with concept drift, they are not ideal for real-time intrusion detection.

III. Methodology

Detecting an intrusion into a system typically entails searching through a vast repository for intrusion signatures that are particularly sophisticated. For this purpose, a complicated model that recognises these signatures is required. This research provides a neural network model that is based on deep learning and has the capability of performing efficient intrusion detection on network transmission data. The Multi Layered Deep Learning Network that has been suggested is a deep learning network since it is made up of a number of hidden processing layers at various depths across the network. It was discovered that detection through the use of the deep network exhibited effective performances when it came to detecting the intrusion signatures.

A. Multi-Layered Deep Learning Networks (MLDN)

In this paper, a Multi-Layer Deep Learning Network (MLDN) model is presented for the detection of intrusions in a timely and accurate manner. The neural network is one of the primary models that is utilised in the quest to make accurate forecasts. Deep network intrusion detection is capable of identifying a great number of intrinsic patterns, in addition to addressing issues of data imbalance and concept drift. The deep learning architecture that has been proposed for use in intrusion detection is broken down into four distinct stages. These stages are data pre-processing, data separation, network development, and model fitting. Fig. 1 illustrates the suggested model's design, and the corresponding pseudocode may be found below.

Architecture of the Algorithm for the MLDN:

1) Input transaction data.

2) Perform data pre-processing to eliminate inconsistencies.

3) Separate the data into three categories: training, testing, and validation.

4) Construct the neural network using the input data as a basis.

a) Create Layer and assign activation function.

b) Determine Epoch.

c) Assign Learning rate.

d) Assign Optimizer and Loss function.

e) Data Shuffling.

f) Create Layer and assign activation function b.

5) Get the process of network training off the ground.

6) Validate the trained model by using the data from the validation.

7) Using the results of the tests, determine the final forecast.

B. Data Collection

The model has been validated through the utilisation of industry-standard benchmark datasets such as NSL-KDD, KDD CUP 99, and Koyoto 2022+ datasets.

C. Data Pre-Processing Phase

In most cases, neural networks are unable to effectively manage all of the different types of data that are incorporated into network data. They are only able to deal with data of the double type, and they require all of the data to fall within the same range in order for it to be relevant to both sets of characteristics. These criteria are dealt with during the pre-processing phase. The procedures that were carried out during the pre-processing phase are illustrated in Fig. 2. Data normalisation follows data imputation. One of the necessary pre-processing steps that must be completed before working with real-time data is known as data normalisation. The operational nature of machine learning models makes it necessary to implement a significant amount of standardisation.

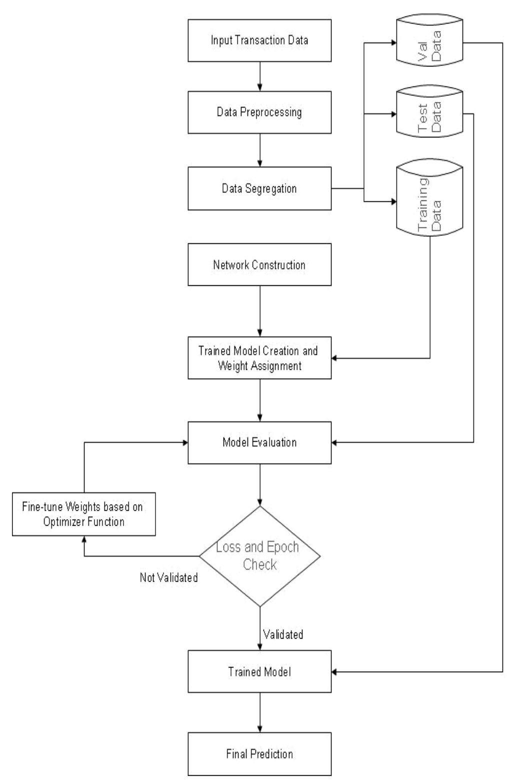


Fig. 1. Proposed MLDN architecture.

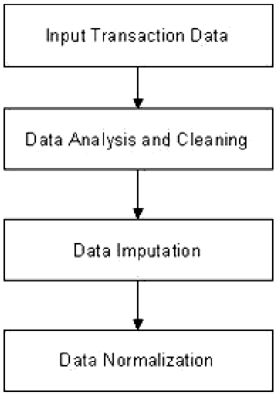


Fig. 2. Data pre-processing.

Every machine learning model has the tendency to fit functions to the data that is provided. In most cases, the weights that are applied to an attribute are what decide the level of relevance that the attribute has. The process of fitting the functions has a tendency to become more difficult when the training data comprises values that fall within a wide range of values. If any of the attributes include significant values, this will ultimately result in the actual value being the one to determine the significance of the variable. Because of this, the weights that were assigned are likely to be useless. As a consequence of this, it becomes vital to transform all of the data into comparable ranges so that consistency can be achieved throughout the process of prediction. The three normalising techniques that are employed the most frequently and extensively are the min-max normalisation, the z-score normalisation, and the decimal scaling. The original data range is transformed in a linear fashion with the application of the Min-Max normalisation. This is demonstrated by:

$$x' = \left(\frac{x - \min(A)}{\max(A) - \min(A)}\right) * (D - C) + C \qquad (1)$$

where "x" represents the normalised value and "x" represents the actual value of the attribute A. The data will be scaled between the predetermined limits [C, D], which are denoted by the letters C and D. Another method that can be utilised in the normalisation process is known as the Z-score normalisation method. The data are normalised between the intervals of 0 and 1 using this model. This follows logically from the formula.

$$x_i' = \frac{x_i - \bar{A}}{std\ (A)} \qquad (2)$$

where xi' and xi are the normalised and actual values of the attribute A, A is the mean value for the attribute A, and std(A) is given by where xi' and xi are the normalised and actual values of the attribute A.

$$std(A) = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{A})^2} \qquad (3)$$

where "n" refers to the total number of rows or instances contained inside the data. The decimal scaling approach is the simplest one, and it yields results that are dependent on both the current value and the highest value that can be found in the property. This is demonstrated by

$$x' = \frac{x}{10^j} \qquad (4)$$

If x' represents the normalised value, x represents the actual value of the attribute A, and j represents the number of digits that make up the highest possible number in the variable A.

For the purpose of normalisation, this study makes use of Min-Max Normalization because it provides the benefit of being able to set both the minimum and the maximum values.

D. Data Segregation Phase

Now that the data have been standardised, the models may be trained using them. On the other hand, it is essential to keep in mind that model validation is a requirement that must be met by any machine learning model. Because of this, the normalised training data is separated into three distinct components: the training data, the testing data, and the validation data. The data is divided in accordance with the proportions 7:2:1. Seventy percent of the total data set is used to sample the training set, twenty percent of the total data set is used to sample the test set, and ten percent of the total data set is used for validation purposes. After the data have been separated, the training data will be utilised in order to construct the trained model.

E. Network Construction Phase

During this phase, a deep neural network model is utilised to facilitate the efficient identification of intrusion signatures derived from the transmission data. In order to construct the neural network, the deep learning library known as Keras was utilised. A neural network, often called an artificial neural network, is a network of neurons that collaborates to perform effective machine learning. Neural networks are also sometimes referred to by its other name, natural neural networks. Neurons, often referred to as perception, are the individual processing pieces that are used to construct neural networks. As can be seen in Fig. 3, a single neuron has numerous inputs coming into it, but it only produces a single output.

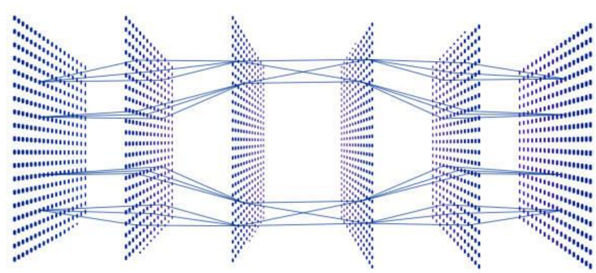


Fig. 3. Neuron: A view.

On the other hand, these inputs cannot be independently acted upon. As a consequence of this, the relative significance of each input is reflected in the weights that are assigned to it. Real numbers (w1, w2,...) are the typical notation used to express weights. The output of a neuron is often a weighted aggregate of the neuron's input value and the weights that accompany that value. This is demonstrated by

$$Output = \varphi \sum_{i=1}^{n} w_i\, x_i \qquad (5)$$

where "the activation function, *w* and *x* are the weights and inputs of the neuron", and the symbol for the activation function. An input layer, one or two processing or hidden layers, and an output layer are the typical layers that make up a neural network. In most cases, the network also contains an output layer. Each layer is made up of multiple neurons, each of which is responsible for processing the information received in that layer and producing the appropriate outputs. Fig. 4 is an illustration that provides a general representation of a neural network.

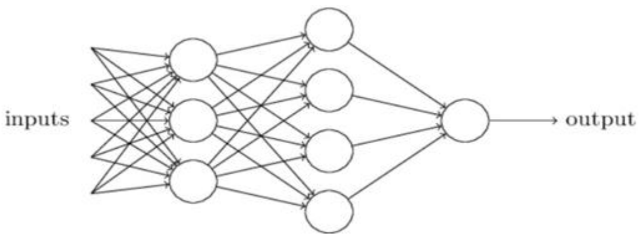


Fig. 4. A simple neural network.

Every layer in the network operates based on the input that was provided by the layer that came before it, carries out operations as specified by Eq. (5), and then passes along its output to the layer that comes after it. One variety of artificial neural network is known as a deep neural network. This network has several hidden layers within its structure. The technique of running a deep neural network is identical to the process of running an artificial neural network; however, the addition of more layers results in the provision of improved prediction capabilities on vast and difficult situations. They are typically utilised in fields that call for the analysis of massive amounts of complex data that contain a number of properties. Because the field of network intrusion detection entails the processing of massive volumes of data with complicated patterns, a deep neural network would be an effective solution to the issue. Keras is a deep learning package that is open source and based on the programming language Python. It makes it possible to create neural networks. The fact that Keras is a high-level library means that it can simply and efficiently integrate with a number of different low-level systems. This is the primary benefit of using Keras. In its current state, Keras is compatible with TensorFlow, Theano, MXNet, and Microsoft Cognitive Toolkit. Keras is typically combined with low-level base libraries like as TensorFlow and Theano. These are the two most popular low-level base libraries. The fact that Keras was designed from the ground up to be extensible, modulatory, and minimalistic is undoubtedly its most significant selling point. When used with TensorFlow, Graphics Processing Units (GPUs) can also be used in conjunction with Tensor Processing Units (TPUs) for improved and more rapid processing. This is yet another advantage of the Keras framework, which was developed with the facilities to include GPUs from the start.

Both the sequential API and the functional API can be used to construct Keras models. The sequential API is the more traditional method. Models can be built layer by layer using the sequential application programming interface. This type is suitable for the vast majority of the applications that are now available. Nevertheless, the approach does not perform very well when used to applications that share layers or that have several inputs or outputs. The functional API makes it possible to create networks that are more adaptable and diverse. It enables connections with layers at any level, which in turn makes it possible to create networks with a greater degree of complexity. Image processing, audio and video processing, and natural language processing are just few of the applications that typically make use of them. To construct the network, this work makes use of the sequential Application Programming Interface.

The neural network model is constructed with the help of the sequential application programming interface. An input layer, several hidden layers, and an output layer make up the network. Each of these levels is sandwiched between two other layers. In this particular investigation, the learning rate is 0.3. The rate at which the model must advance in order to become closer and closer to the correct response is known as the learning rate. A slower rate of convergence is indicated by smaller numbers, while a faster rate of convergence is indicated by larger values. Larger values may cause the model to bypass the optimal solution. As a consequence of this, it is necessary to identify the option that offers the highest value taking into account the circumstances. At this time, level 50 has been selected for the period. Epochs are the intervals of time during which the neural network model is provided with training data so that it can acquire new skills. Higher epochs produce better models. On the other hand, extreme care needs to be taken to prevent the model from being overfit. The parameters that were utilised are detailed in Table I.

TABLE I. NEURAL NETWORK PARAMETERS

Parameter	Value
Network Type	Sequential
Batch Size	64
Epochs	50
Learning Rate	0.1
Shuffle	True
Validation Data	Provided
Optimizer	Adam

The data that is being sent to the neural network has been shuffled to ensure that it is not arranged in any particular way before it is sent there. The network will benefit from this in the form of generic training. The validation data has been added to the network in order to ensure accurate prediction and also to prevent the network model from being overly tailored to the data. In order to perform the process of iteratively adjusting the weights of the neural network model depending on the data, an Optimizer algorithm is required to be utilised. The Adam optimizer is utilised right here. The Stochastic Gradient Descent algorithm has been expanded upon in order to create the Adam optimizer. The learning rate determines the level of update that is available. In order to generate adaptable learning rates, the algorithm modifies the levels of the learning rate at increasingly frequent intervals throughout the training process. This helps in better identifying the best possible solution to the problem. As a result, the Adam optimizer continues to be one of the optimizer algorithms that is utilised the most in neural networks.

The neural network that is suggested will have different layer configurations constructed into it depending on the dataset that is being analysed. It is intended for each of these layers to have a substantial thickness. When all of a layer's nodes are connected to all of the nodes of the layer that follows it, we refer to that layer as dense. This contributes to the construction of a network that broadcasts all of its discoveries to every accessible node on the network.

The input layer is the first layer that is created. The total number of attributes that are included in the training data is typically used to determine how many nodes should be present in the input layer. The model that has been proposed is made up of two discrete levels. Multiple neurons are incorporated into the design of the hidden layers. In this particular experiment, the successive layers each make use of 100 and 50 neurons. One neuron is present in the output layer that comes last. Since the issue that is being worked on is a binary classification issue, it would be sufficient to use a single neuron that was programmed with the output probability.

In addition to these qualities, activation functions are an extremely important factor in determining the effectiveness of the neural network. The activation function of a node in a neural network is what decides what the output of that node will be given the set of inputs for that node. This output is used as an input by the node that is located on the layer below. In most cases, the value that is produced by an activation function falls somewhere in the range of 0 to 1 or -1 to 1.

The activation function that is used is what determines the actual output that is produced. Other activation functions are available, however the sigmoid (Shown in Fig. 5), hyperbolic tangent (tanh), Rectified Linear Units (ReLU), and linear activation functions are the most frequent ones used in neural networks. Other activation functions are also available. In most cases, the sigmoid activation function has the form of the equation below:

$$f(x) = \frac{1}{1+\exp(-x)} \tag{6}$$

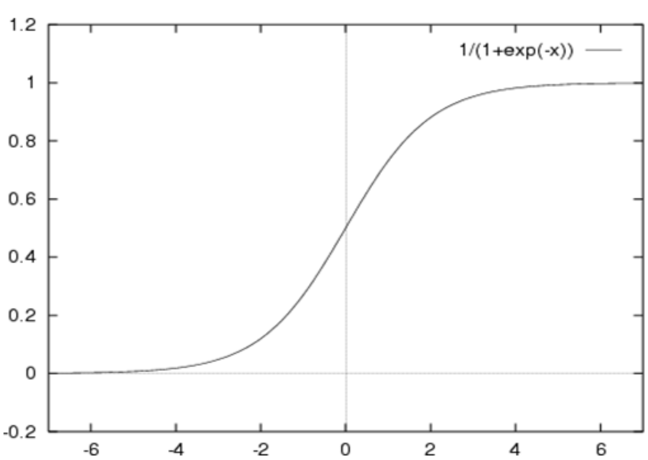


Fig. 5. Sigmoid activation function.

The range of the curve, which is S-shaped, is between 0 and 1, and the range of the curve itself is between 0 and 1. The fact that this function's range is [0, 1], which makes optimization more difficult, is the function's primary drawback. Because of its slow convergence, it is particularly well-suited for issues involving binary categorization. It has a problem with the gradient disappearing into nothingness. The activation function of the hyperbolic tangent, abbreviated as tanh, takes the form

$$f(x) = \frac{1-\exp(-2x)}{1+\exp(-2x)} \tag{7}$$

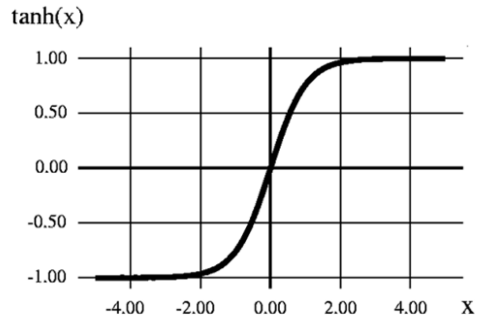


Fig. 6. TanH activation function.

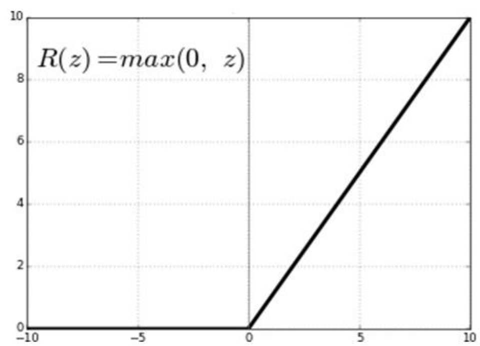


Fig. 7. ReLU activation function.

The ReLU function ranges between 0 and 1 (Fig. 7). It provides six times better convergence compared to TanH. The model is recommended for usage in intermediate layers, as it might lead to dead neurons if the input contains negative values. In the proposed approach the input and hidden layers use ReLU activation function, while the output layer uses linear activation. The S curve can also be seen in this function's plot. On the other hand, the output values are in the range of -1 to 1. This makes it much simpler to perform optimizations. However, this function also has a problem called vanishing gradient, which makes it difficult to evaluate (Fig. 6).

The phenomenon known as the "vanishing gradient problem" typically occurs in models such as neural networks that permit the backpropagation of errors. The fact that errors are typically calculated in the final output layer is the most significant problem. As a result, the layer that comes immediately before the final layer is responsible for handling the faults that have the greatest impact. However, despite the fact that faults are passed on to early layers, each layer is responsible for handling problems and only passes on errors that are still present to subsequent layers. As a consequence of this, the earliest layers do not typically get a significant amount of influence from the faults. Because of this, the early layers are the ones that take the most time to train. On the other hand, the early layers are the ones in charge of recognising fundamental patterns, which serve as the fundamental constituents of the neural network model. The whole neural networks model converges more slowly as a consequence of this issue. This problem was addressed with the development of the ReLU function. This is the form it takes.

F. *Model Fitting Phase*

The building of a functional model or architecture is required before the neural network can be created in the phase that came before this one. Fig. 8 depicts the order in which an artificial neural network model is developed and put into operation. The process of model fitting is the activity that actually carries out the training of the network. Before moving on to the next step, the data is first partitioned into two distinct sections: the data section and the labels section. The data part contains the attributes, with the exception of the class attribute; the class attribute is located in the labels section. The data from the training session are input into the neural network model so that it can perform the necessary analysis. The epoch value determines the total number of training iterations that the backpropagation network undergoes before being considered fully trained. The training continues until an error rate that meets the requirements is achieved. The data that needs to be forecasted is sent over this network, and then the conclusions drawn from those predictions are obtained.

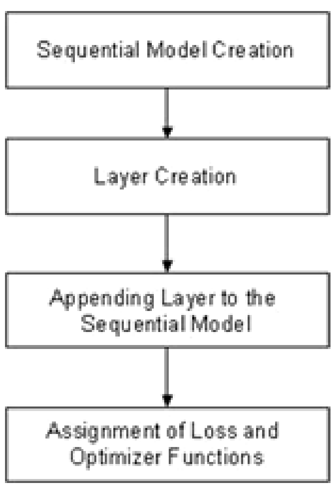


Fig. 8. Neural network operational sequence.

IV. RESULTS AND DISCUSSION

Python and the Keras library suite were used to create the MLDN architecture that has been presented. The model has been validated through the utilisation of industry-standard benchmark datasets such as NSL-KDD, KDD CUP 99, and Koyoto 2022+ datasets. The network is constructed with the help of Sequential API, and then the layers that make up the neural network are added. To properly analyse each dataset, a unique network architecture will need to be developed. Tables II, III, and IV present the structures that were developed for each of the datasets that were utilised.

TABLE II. NEURAL NETWORK CONFIGURATION FOR NSL-KDD

Layer (Type)	Activation Function	Input Dimension	Output Shape	No. of Parameters
Input (Dense)	Linear	41	(None,80)	3360
Processing 1 (Dense)	ReLU	80	(None,100)	8100
Processing 2 (Dense)	ReLU	100	(None,50)	5050
Processing 3 (Dense)	ReLU	50	(None,50)	1530
Output (Dense)	Linear	30	(None,1)	31
Total No. of Parameters				18,071
Trainable Parameters				18,071
Non-trainable Parameters				0

TABLE III. NEURAL NETWORK CONFIGURATION FOR KDD CUP 99

Layer (Type)	Activation Function	Input Dimension	Output Shape	No. of Parameters
Input (Dense)	Linear	38	(None,50)	1950
Processing 1 (Dense)	ReLU	50	(None,250)	12,750
Processing 2 (Dense)	ReLU	250	(None,100)	25,100
Processing 3 (Dense)	ReLU	100	(None,50)	5050
Output (Dense)	Linear	50	(None,1)	51
Total No. of Parameters				44,901
Trainable Parameters				44,901
Non-trainable Parameters				0

TABLE IV. NEURAL NETWORK CONFIGURATION FOR KOYOTO 2022+

Layer (Type)	Activation Function	Input Dimension	Output Shape	No. of Parameters
Input (Dense)	Linear	18	(None,50)	950
Processing 1(Dense)	ReLU	50	(None,200)	10200
Processing 2(Dense)	ReLU	200	(None,100)	20220
Processing 3(Dense)	ReLU	100	(None,20)	2020
Output (Dense)	Linear	20	(None,1)	21
Total No. of Parameters				33,291
Trainable Parameters				33,291
Non-trainable Parameters				0

Fig. 9 presents the results of an evaluation of how well the suggested MLDN model performed on the NSL-KDD, KDD CUP 99, and Koyoto 2022+ datasets in terms of their respective ROC charts. The False Positive Rate (FPR) is represented by the x-axis, and the True Positive Rate is represented by the y-axis in this graph (TPR). It is anticipated that an effective model will demonstrate high levels of TPR while exhibiting low levels of FPR. After the points have been plotted, the graph demonstrates that the suggested model has ROC curves that are located at the (0, 1) or top-right position. This indicates that the proposed model is effective. Fig. 10 depicts the PR curve, which represents the accuracy and recall levels of the proposed model over all three datasets. An efficient model should display high values on the x-axis for recall and also display high values on the y-axis for precision (y-axis). The graph that represents the PR plot demonstrates that all three datasets have high levels of accuracy and recall, which demonstrates that the suggested model has a high level of prediction performance.

On the NSL-KDD, KDD CUP99, and Koyoto 2022+ datasets, the values that were obtained for a variety of performance measures such as FPR, TPR, Recall, and Precision, among others, are provided in Table V. The MLDN model that was proposed has levels of TPR and Precision that are extremely high, which is an indication of its effectiveness in predicting intrusion signs. In a similar vein, a high TNR level is indicative of the fact that the presented model demonstrates excellent prediction efficiency when it comes to anticipating normal transmission signals. In a similar vein, low FPR and FNR levels of less than one percent imply that the model has exceptionally low levels of incorrect predictions. As a result of this, it is abundantly clear that the MLDN model that was proposed is efficient and offers good performance.

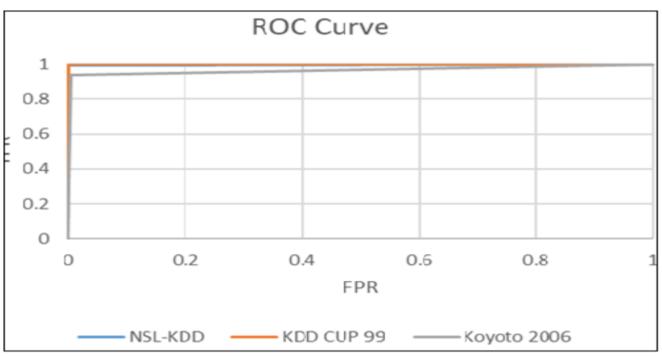


Fig. 9. ROC curve comparison of MLDN.

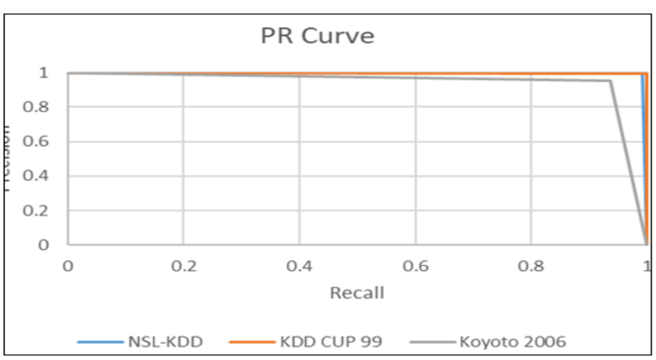


Fig. 10. PR curve comparison of MLDN.

Alterations were made to the settings of the parameters, and a sensitivity analysis was carried out on each of the three datasets in order to determine how the learning rate and the number of epochs affected the results. During the study, multiple parameter pairs were employed, and the accuracy that was acquired for each parameter combination was used during the analysis.

Table VI contains the acquired results for perusal. Within the first five sets, the learning rate is manipulated while the epoch is held constant (P1 to P5). It was possible to see that, as the learning rate was decreased, the performance on all tree datasets tended to decline to some degree, and this tendency increased as the learning rate was decreased further (P1 and P2). The learning rate is decreased, and as a result, the model takes increasingly minute steps in the direction of the best answer. As a result, 50 epochs were insufficient to accomplish the goal of achieving convergence. Taking the example of P11, where the number of epochs is increased, demonstrates that the model was able to reach convergence. When the learning rate is increased, such as in P4 and P5, the results demonstrate a decrease in performance.

TABLE V. PERFORMANCE ANALYSIS OF MLDN

Measures	NSL-KDD	KDD CUP 99	Koyoto 2022+
FPR	0.001934	0.001254	0.005605
TPR	0.991718	0.998519	0.93578
Recall	0.991718	0.998519	0.93578
Precision	0.997917	0.995079	0.953271
TNR	0.998066	0.998746	0.994395
FNR	0.008282	0.001481	0.06422
Accuracy	0.995	0.9987	0.988012
F-Measure	0.994808	0.996796	0.944444
AUC	0.994892	0.998632	0.965087

TABLE VI. SENSITIVITY ANALYSIS RESULTS

Parameter Set	Learning Rate	Epochs	NSL-KDD	KDD CUP 99	Koyoto 2022
P1	0.1	50	0.94	0.937	0.913
P2	0.3	50	0.97	0.959	0.944
P3	0.5	50	0.995	0.999	0.979
P4	0.7	50	0.991	0.999	0.973
P5	1	50	0.89	0.926	0.851
P6	0.5	10	0.72	0.69	0.583
P7	0.5	25	0.79	0.829	0.811
P8	0.5	70	0.995	0.999	0.979
P9	0.5	100	0.995	0.999	0.979
P10	0.5	200	0.995	0.999	0.979
P11	0.3	200	0.995	0.999	0.978

This is because an increased learning rate results in big steps, and as a consequence, the model has a tendency to miss the ideal convergence point. Epochs are changed while the learning rate remains the same in parameter sets P6 to P10. It is clear from the reduced epochs (P6 and P7) that the model is not being given a long enough period of time to converge. As a result, the highest possible degree of precision is not achieved. The highest levels of precision can be attained when the period is advanced to 50 (P3) and beyond (P8 to P10). It has been noticed that using 50 epochs provides the highest level of accuracy. After reaching this point, increasing the number of epochs will have no effect on the performance because convergence will have already been reached by that point.

It is possible to summarise that the rate of learning plays an essential part in the achievement of successful results. It is vital to locate the best convergence level and the sweet spot that corresponds to it. Any number less than this point necessitates additional time for the model to converge, and any value greater than this point will cause the model to miss the point at which it converges. Epochs represent the number of times that the training data should be iterated through by the model in order to reach convergence. If there are fewer epochs than necessary, the model will not have enough time to converge, and if there are more epochs than necessary, there will be an additional time overhead with no improvement in performance. In addition to that, it will also result in overfitting, which is why it ought to be avoided. Comparisons are made between the HBM model proposed in Part 3 and the APID model proposed in Part 4 in terms of TPR, TNR, Precision, F-Measure, and AUC on the NSL-KDD, KDD CUP 99, and Koyoto 2022+ datasets, which are depicted in Fig. 11 to 19.

A comparison of TPR, TNR, Precision, F-Measure, and AUC on NSL-KDD data demonstrates that the MLDN model exhibits better prediction levels when compared to APID and HBM (Fig. 11 to 13). This is shown by the fact that the MLDN model has a higher AUC.

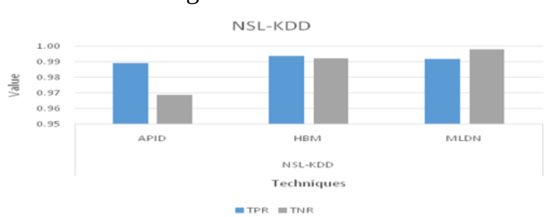


Fig. 11. Comparison of TPR and TNR of MLDN, HBM and APID on NSL-KDD.

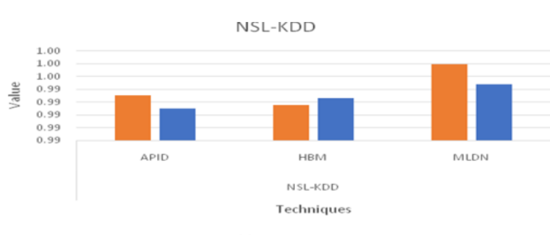


Fig. 12. Comparison of precision and F-Measure of MLDN, HBM and APID on NSL-KDD.

Fig. 13. Comparison of AUC of MLDN of MLDN, HBM and APID on NSL-KDD.

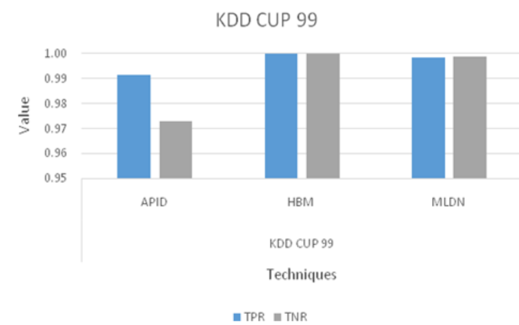


Fig. 14. Comparison of TPR and TNR of MLDN on KDD CUP 99.

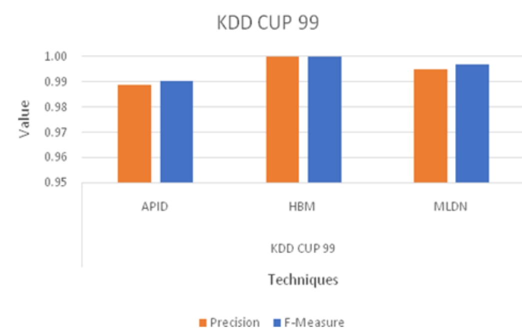


Fig. 15. Comparison of precision and F-Measure of MLDN on KDD CUP 99.

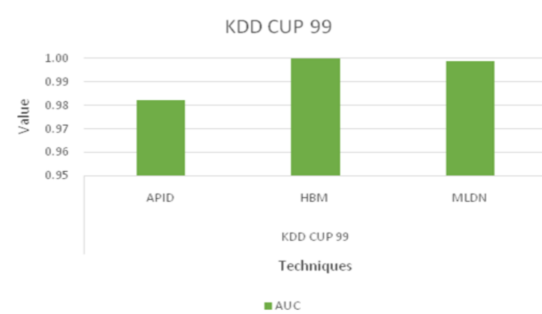


Fig. 16. Comparison of AUC of MLDN on KDD CUP 99.

Analysis of performance on KDD CUP 99 dataset shown in Fig. 14 to 16 demonstrate that in comparison to the APID model, the performance of the MLDN model that has been proposed is superior. In contrast to this, the performance levels demonstrate a marginal drop of 0.1% when measured against the HBM model. The levels of reduction are so negligibly very low that they can be ignored as a result.

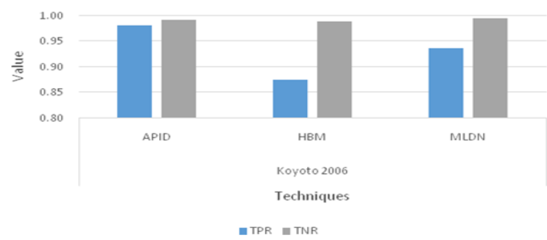


Fig. 17. Comparison of TPR and TNR of MLDN on Koyoto 2022+.

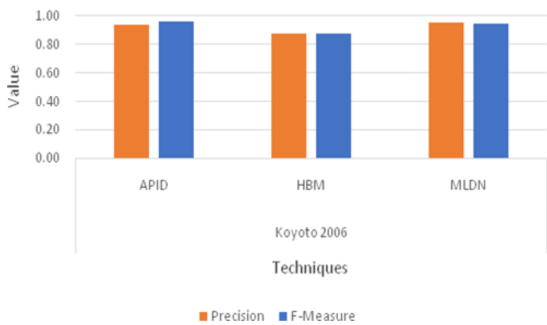


Fig. 18. Comparison of precision and F-Measure of MLDN on Koyoto 2022+.

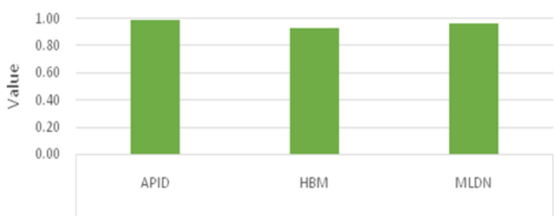


Fig. 19. Comparison of AUC of MLDN on Koyoto 2022+.

When compared to the other models, the performance of the proposed MLDN model on the Koyoto 2022+ datasets shows a slight decrease in the performance with respect to certain metrics such as TPR and F-Measure. Even in this case, the reductions are extremely minimal and, as a result, are insignificant.

Table VII provides a tabular representation of the performance comparisons that were made. The table demonstrates that the overall performance of the proposed models was found to be high and effective, despite the fact that there are slight reductions and elevations in the performance levels of the proposed models.

Table VIII presents a comparison of the amount of time spent training and testing the APID, HBM, and MLDN models. Training is carried out using 70 percent of the records contained in the data, while testing has been carried out using 30 percent of the records across all of the datasets. The HBM model has the most efficiency with regard to its use of time, followed by the APID model and then the MLDN model. Testing requirements for the MLDN model are always less than one second, which is a low requirement that corresponds to a real-time prediction scenario. The training requirements for the MLDN model are quite high. In addition, the little increase in the amount of time needed could be neglected due to the significant boost in terms of performance, which would make the MLDN model the most effective performer when it comes to the detection of intrusions in networks.

The results of a comparative analysis of the proposed MLDN model with the HBM model, the APID model, and models proposed by [5, 6], [7], [8], and [9] are shown in Fig. 20 to 22. This analysis compares the proposed MLDN model with the HBM model and the APID model.

In comparison to other models that are currently available in the literature, the suggested models have a higher level of accuracy in their prediction, as shown in Fig. 20 to 22. The MLDN model produces the best results in terms of performance, followed by the HBM model, which delivers the results with the next best performance. After this comes the APID model, and after that comes the models that already exist in the literature.

TABLE VII. PERFORMANCE COMPARISON OF APID, HBM AND MLDN

Measures	***NSL-KDD***			***KDD CUP 99***			***KOYOTO 2022+***		
	APID	*HBM*	*MLDN*	*APID*	*HBM*	*MLDN*	*APID*	*HBM*	*MLDN*
TPR	*0.99*	*0.99*	*0.99*	*0.99*	*1.00*	*1.00*	*0.98*	*0.88*	*0.94*
Precision	*0.99*	*0.99*	*1.00*	*0.99*	*1.00*	*1.00*	*0.94*	*0.88*	*0.95*
TNR	*0.97*	*0.99*	*1.00*	*0.97*	*1.00*	*1.00*	*0.99*	*0.99*	*0.99*
Accuracy	*0.99*	*0.99*	*1.00*	*0.99*	*1.00*	*1.00*	*0.99*	*0.98*	*0.99*
F-Measure	*0.99*	*0.99*	*0.99*	*0.99*	*1.00*	*1.00*	*0.96*	*0.88*	*0.94*
AUC	*0.98*	*0.99*	*0.99*	*0.98*	*1.00*	*1.00*	*0.99*	*0.93*	*0.97*

TABLE VIII. TIME COMPARISON OF APID, HBM AND MLDN

	NSL-KDD		**KDD CUP 99**		**Koyoto 2022**	
	Training (sec)	**Testing (sec)**	**Training (sec)**	**Testing (sec)**	**Training (sec)**	**Testing(sec)**
MLDN	50	0.259	4	0.316	2	0.279
HBM	2.24	0.21	0.209	0.102	0.107	0.098
APID	2.42	0.302	0.257	0.108	0.0962	0.054

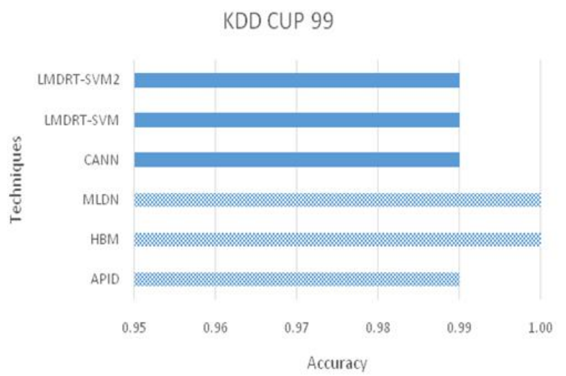


Fig. 20. Comparison of accuracy of MLDN with state-of the-art models on KDD CUP 99.

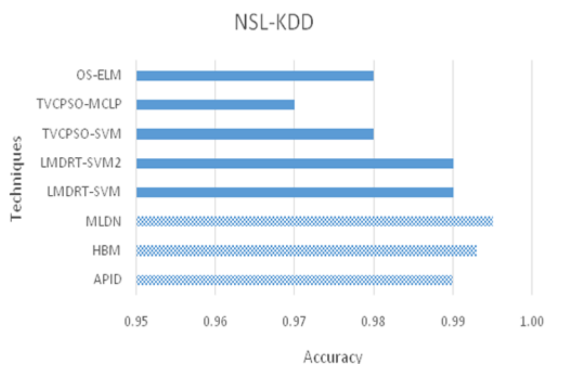


Fig. 21. Comparison of accuracy of MLDN with state-of the-art models on NSL-KDD.

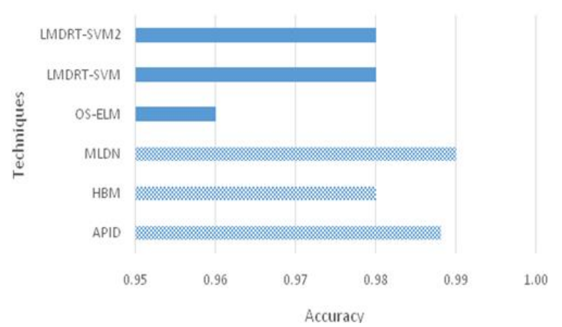


Fig. 22. Comparison of accuracy of MLDN with state-of the-art models on Koyoto 2022+.

A tabular representation of the findings is presented in Table IX. The best results are highlighted in bold below. It was found that the proposed models performed better than all of the existing models that were researched and used as a point of comparison in the literature. In general, methods are constructed with data as their foundation. When utilised with a wider variety of data, such data-specific models are unable to generate results that are beneficial. This particular illustration could be effectively noticed in models from the literature, where models perform better in certain cases while performing worse in others. Because the models that are provided are generic, it is possible to see that the proposed model performs well regardless of the dataset that is being used. This is because the offered models are general.

TABLE IX. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART MODELS

KDD	
Technique	**Accuracy**
APID	0.99
HBM	1.00
MLDN	1.00
CANN	0.99
LMDRT-SVM	0.99
LMDRT-SVM2	0.99
NSL-KDD	
Technique	**Accuracy**
APID	0.99
HBM	0.99
MLDN	1.00
LMDRT-SVM	0.99
LMDRT-SVM2	0.99
TVCPSO-SVM	0.98
TVCPSO-MCLP	0.97
OS-ELM	0.98
Koyoto 2022+	
Technique	**Accuracy**
APID	0.99
HBM	0.98
MLDN	0.99
OS-ELM	0.96
LMDRT-SVM	0.98
LMDRT-SVM2	0.98

V. CONCLUSION

The initial work provides the Adaptive Parallelized Intrusion Detection (APID) model, which is utilised for finding intrusion signs from data that is sent within a network. This model was developed by the researchers. The transmission data that is later sent to the model is first subjected to pre-processing, and then training bags are produced. The training data bags are then given to the learners at the basic level. The training for every base learner is determined by the training bag that is given to it. The learner that demonstrates the best prediction rates in terms of normal data prediction and in terms of the best overall prediction is determined to be the basic learner.

The final predictions are obtained by utilising heuristic-based combiners, after which the test data are forecasted by using all of the basic learners. After that, the ensemble is retrained based on the false predictions to produce an adaptive model that is capable of changing itself to produce better predictions over the course of time. Despite the fact that the model delivers impressive results, it is not appropriate for use for processing large amounts of data because of the significant computational cost connected with it. In the following contribution, a Heterogeneous Bagging based Model, or HBM, is presented with the goal of reducing complexity levels. This model features an improved bagging method that makes it possible to detect intrusions both more quickly and more accurately. The training data is divided up into various bags that also overlap with one another. Both the Decision Tree and Random Forest models are utilised here as the foundation learners for the model. The bagging procedure is altered in such a way that each data bag is made available to both models. Each of the bags that are constructed gets its own unique set of several pairs of base learners.

On the basis of these models, predictions are made, and the results of those predictions are combined with the votes from the voting combiner. While this model does a better job of simplifying complex processes, its performance is marginally inferior. A deep learning network-based intrusion detection model, known as the MLDN, is presented as the last contribution. This is done in order to improve efficiency. The Keras library was utilised during the construction of the deep learning model. In order to accommodate the specific processing architecture of neural networks, the training data is preprocessed and partitioned.

The network is built with various layers, and all of the other necessary parameters for the network are configured based on the data that is entered. Validation of the trained model is accomplished by employing the validation data that has been meticulously isolated for the sole purpose of fulfilling this requirement. Standard benchmark datasets, such as KDD CUP99, NSL-KDD, and Koyoto 2022+ datasets, were utilised in the experiments that were carried out. Comparisons were made with previously published models that were already in existence. The analysis was carried out using the existing standard performance metrics for classifiers, which included TPR, FPR, TNR, FNR, Precision, Recall, F-Measure, and Accuracy. According to the findings, the proposed models appear to have superior performance levels when compared to the standard models that are currently in use.

References

[1] B. Gao, B. Bu, W. Zhang and X. Li, "An Intrusion Detection Method Based on Machine Learning and State Observer for Train-Ground Communication Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 6608-6620, July 2022.

[2] R. Bitton and A. Shabtai, "A Machine Learning-Based Intrusion Detection System for Securing Remote Desktop Connections to Electronic Flight Bag Servers," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 3, pp. 1164-1181, 1 May-June 2021.

[3] Sahu et al., "Multi-Source Multi-Domain Data Fusion for Cyberattack Detection in Power Systems," in IEEE Access, vol. 9, pp. 119118-119138, 2021.

[4] X. Y. Li, R. Tang and W. Song, "Intrusion Detection System Using Improved Convolution Neural Network," 2022 11th International Conference of Information and Communication Technology (ICTech)), Wuhan, China, 2022, pp. 97-100.

[5] C. Chen, X. Xu, G. Wang and L. Yang, "Network intrusion detection model based on neural network feature extraction and PSO-SVM," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 1462-1465.

[6] S. L. Rocha, G. Daniel Amvame Nze and F. L. Lopes de Mendonça, "Intrusion Detection in Container Orchestration Clusters : A framework proposal based on real-time system call analysis with machine learning for anomaly detection," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), Madrid, Spain, 2022, pp. 1-4.

[7] R. Zhang, Y. Song and X. Wang, "Network Intrusion Detection Scheme Based on IPSO-SVM Algorithm," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2022, pp. 1011-1014.

[8] H. Yang, Y. Bai, T. Chen, Y. Shi, R. Yang and H. Ma, "Intrusion Detection Model For Power Information Network Based On Multi-layer Attention Mechanism," 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2022, pp. 825-828.

[9] F. J. Mora-Gimeno, H. Mora-Mora, B. Volckaert and A. Atrey, "Intrusion Detection System Based on Integrated System Calls Graph and Neural Networks," in IEEE Access, vol. 9, pp. 9822-9833, 2021.

[10] X. Gong, X. Chen, Z. Zhong and W. Chen, "Enhanced Few-Shot Learning for Intrusion Detection in Railway Video Surveillance," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11301-11313, Aug. 2022.

[11] P. Freitas De Araujo-Filho, A. J. Pinheiro, G. Kaddoum, D. R. Campelo and F. L. Soares, "An Efficient Intrusion Prevention System for CAN: Hindering Cyber-Attacks With a Low-Cost Platform," in IEEE Access, vol. 9, pp. 166855-166869, 2021.

[12] M. Abdel-Basset, N. Moustafa, H. Hawash, I. Razzak, K. M. Sallam and O. M. Elkomy, "Federated Intrusion Detection in Blockchain-Based Smart Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 2523-2537, March 2022.

[13] Q. Liu, V. Hagenmeyer and H. B. Keller, "A Review of Rule Learning-Based Intrusion Detection Systems and Their Prospects in Smart Grids," in IEEE Access, vol. 9, pp. 57542-57564, 2021.

[14] C. Jichici, B. Groza, R. Ragobete, P. -S. Murvay and T. Andreica, "Effective Intrusion Detection and Prevention for the Commercial Vehicle SAE J1939 CAN Bus," in IEEE Transactions on Intelligent Transportation Systems. doi: 10.1109/TITS.2022.3151712.

[15] J. Shu, L. Zhou, W. Zhang, X. Du and M. Guizani, "Collaborative Intrusion Detection for VANETs: A Deep Learning-Based Distributed SDN Approach," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 4519-4530, July 2021.

[16] M. Nadeem, A. Arshad, S. Riaz, S. S. Band and A. Mosavi, "Intercept the Cloud Network From Brute Force and DDoS Attacks via Intrusion Detection and Prevention System," in IEEE Access, vol. 9, pp. 152300-152309, 2021.

[17] T. Wisanwanichthan and M. Thammawichai, "A Double-Layered Hybrid Approach for Network Intrusion Detection System Using Combined Naive Bayes and SVM," in IEEE Access, vol. 9, pp. 138432-138450, 2021.

[18] G. Apruzzese, L. Pajola and M. Conti, "The Cross-evaluation of Machine Learning-based Network Intrusion Detection Systems," in IEEE Transactions on Network and Service Management.

[19] J. Lansky et al., "Deep Learning-Based Intrusion Detection Systems: A Systematic Review," in IEEE Access, vol. 9, pp. 101574-101599, 2021.

[20] L. Yang, A. Moubayed and A. Shami, "MTH-IDS: A Multitiered Hybrid Intrusion Detection System for Internet of Vehicles," in IEEE Internet of Things Journal, vol. 9, no. 1, pp. 616-632, 1 Jan.1, 2022.

[21] Y. Miao, Y. Tang, B. A. Alzahrani, A. Barnawi, T. Alafif and L. Hu, "Airborne LiDAR Assisted Obstacle Recognition and Intrusion Detection Towards Unmanned Aerial Vehicle: Architecture, Modeling and Evaluation," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 4531-4540, July 2021.

[22] M. A. Siddiqi and W. Pak, "An Agile Approach to Identify Single and Hybrid Normalization for Enhancing Machine Learning-Based Network Intrusion Detection," in IEEE Access, vol. 9, pp. 137494-137513, 2021.

[23] Y. Gao, H. Miao, J. Chen, B. Song, X. Hu and W. Wang, "Explosive Cyber Security Threats During COVID-19 Pandemic and a Novel Tree-Based Broad Learning System to Overcome," in IEEE Transactions on Intelligent Transportation Systems.

[24] Z. Hu, S. Liu, W. Luo and L. Wu, "Intrusion-Detector-Dependent Distributed Economic Model Predictive Control for Load Frequency Regulation With PEVs Under Cyber Attacks," in IEEE Transactions on

Circuits and Systems I: Regular Papers, vol. 68, no. 9, pp. 3857-3868, Sept. 2021.

[25] R. Conde Camillo da Silva, M. P. Oliveira Camargo, M. Sanches Quessada, A. Claiton Lopes, J. Diassala Monteiro Ernesto and K. A. Pontara da Costa, "An Intrusion Detection System for Web-Based Attacks Using IBM Watson," in IEEE Latin America Transactions, vol. 20, no. 2, pp. 191-197, Feb. 2022.

[26] K. Agrawal, T. Alladi, A. Agrawal, V. Chamola and A. Benslimane, "NovelADS: A Novel Anomaly Detection System for Intra-Vehicular Networks," in IEEE Transactions on Intelligent Transportation Systems.doi: 10.1109/TITS.2022.3146024.

[27] C. Kim, M. Jang, S. Seo, K. Park and P. Kang, "Intrusion Detection Based on Sequential Information Preserving Log Embedding Methods and Anomaly Detection Algorithms," in IEEE Access, vol. 9, pp. 58088-58101, 2021.

[28] J. Gao et al., "Omni SCADA Intrusion Detection Using Deep Learning Algorithms," in IEEE Internet of Things Journal, vol. 8, no. 2, pp. 951-961, 15 Jan.15, 2021.

[29] T. Kim and W. Pak, "Hybrid Classification for High-Speed and High-Accuracy Network Intrusion Detection System," in IEEE Access, vol. 9, pp. 83806-83817, 2021.

[30] N. Mishra and S. Pandya, "Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review," in IEEE Access, vol. 9, pp. 59353-59377, 2021.

[31] G. Pu, L. Wang, J. Shen and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," in Tsinghua Science and Technology, vol. 26, no. 2, pp. 146-153, April 2021.

[32] O. Alkadi, N. Moustafa, B. Turnbull and K. -K. R. Choo, "A Deep Blockchain Framework-Enabled Collaborative Intrusion Detection for Protecting IoT and Cloud Networks," in IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9463-9472, 15 June15, 2021.

[33] S. Seth, K. K. Chahal and G. Singh, "A Novel Ensemble Framework for an Intelligent Intrusion Detection System," in IEEE Access, vol. 9, pp. 138451-138467, 2021.

[34] D. Gümüşbaş, T. Yıldırım, A. Genovese and F. Scotti, "A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems," in IEEE Systems Journal, vol. 15, no. 2, pp. 1717-1731, June 2021.

[35] H. Janabi, T. Kanakis and M. Johnson, "Overhead Reduction Technique for Software-Defined Network Based Intrusion Detection Systems," in IEEE Access, vol. 10, pp. 66481-66491, 2022.

[36] L. Vu, Q. U. Nguyen, D. N. Nguyen, D. T. Hoang and E. Dutkiewicz, "Deep Generative Learning Models for Cloud Intrusion Detection Systems," in IEEE Transactions on Cybernetics.doi: 10.1109/TCYB.2022.3163811.

Footwear Sketches Colorization Method based on Generative Adversarial Network

Xin Li[1]*, Yihang Zhang[2]
School of Art and Design, Guangdong University of Technology, Guangzhou, China[1]
College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China[2]

***Abstract*—The coloring of sketches has a constant market demand in the area of research. The difficulty of the coloring sketch outline is its lack of texture and color. Take footwear design as an example, it is difficult for designers to complete a colorful sketch in a limited time, so an artificial intelligence technology for coloring shoes is required. Though we do not build a new GAN, which is based on pix2pix. We try to integrate the existing model in four ways, including generator, discriminator, loss function and comparison. In this paper, given a set of edges-to-shoes that have 50,025 shoe images, our approach produces an image with vivid shoes images. Unlike the recent research, our approach is not based on a unique adversarial training. We show that shoe sketches can be synthesized from simple lines by a GAN into a high-resolution picture. In particular, we offer a new model to synthesize high-resolution photo-realistic images of shoes, and apply a multi-discriminator to train and distinguish the generated images. Our model enables the shoe designer to benefit from the colorization design.**

Keywords—Footwear sketch; generative adversarial network; image to image translation; colorization

I. Introduction

In the footwear industry, color sketch is a time-consuming step in design period for designers. Creating a concept map of colorful shoes requires a professional color, composition and valid use of shade and texture. This process demands experienced drawing expertise and a good sense of design aesthetics. Even the experts would spend much time coloring the sketches.

In the design process, designer may decrease the needless hours if they overcome the colorization issues. As a result, a standalone coloring system can be a suitable solution for the footwear design industry. With the help of this system, the newcomer can be inspired, while professors save more time on color compositions of product sketches. This idea may totally change the fundamental structure of product development, which do optimize the industrial structure.

However, a lot of challenges remain to achieve this process. At first it is hard for the machine to understand the sketches of footwear which have innumerable drawing styles. As well, footwear sketches have a limited expression. What's more, there is no guidance for a machine to make colorizing decisions.

While deep learning in IT vision research is becoming a hot topic, alternative learning-based methods have been developed. For example, Mathias Eitz presented an interactive pattern recognition system that can identify a human sketch object [1]. Christopher Hesse has operated an online system that can generate cat images from the edges [2]. Although these systems can successfully turn a user sketch into a colorful object, such an application still cannot meet the designer's expectation. In addition, it is hard to shape high-resolution images and images with details and texture. Due to the limitation of an experimental topic, there are still has many opportunities to make progress.

In this article, we propose a method that allows users to enrich their design with hand-made shoes, color based on the Generative Adversarial Networks (GAN). GANs can classify the real or fake images, while forming a model that can minimize the loss. We form our network on an open access edge-to-shoes dataset using a new approach that creates high-resolution images. Different from the previous results with little detail and realistic textures, we explore a new, robust adversarial learning method with multi-scale generator and discriminator framework. This framework can produce a result with better visual quality. In this process, we receive the result with adversarial training rather than any loss by hand-made or pre-trained networks. This method shows that it is possible to improve the addition of perception losses from pre-trained networks. What is more, a multi-discriminator allows a better performance in training and alternative quantitative comparison such as PSNR, SSIM and MSE can have a thorough analysis of the coloring footwear results. Our contributions may be summarized as follows:

1) We introduce a new method called local amplifier to synthesize high-resolution photo-realistic images of shoes.

2) We consume the discriminator by extending the GAN into a multiple framework while improving the adversarial loss.

3) A quantitative comparison using the PSNR, SSIM and MSE is included.

II. Related Work

A. Generative Adversarial Network

In recent years, deep learning has unleashed another wave of artificial intelligence. In particular, in the areas of image recognition, the method based on deep learning [3, 4] has much improved over traditional methods. Their accuracy rate is near to or even greater than that of the manual identification.

The most typical task in unsupervised learning is an image generation [5, 6], and the first image generation model based

*Corresponding Author.

on deep learning is Autoencoder [7]. However, Autoencoder do not have a specific link to measure the error between the reconstructed sample and the real sample. The upgrade model VAE, simply makes the generated images more similar to the database images, instead of learning the generative paradigm for obtaining new images.

A new architecture called the generative adversarial network [8] attempts to resolve this problem. The standard GAN model is made up of two parts, one generator and one discriminator. It no longer updates the generator solely by measuring the similarity between the generated image and the actual image. However, it implements adversarial training through a discriminator, so that the generator can learn then latent picture mode. Meanwhile the distribution is nearer to the distribution of reality.

Today, GAN has become the main model for picture generation [9, 10, 11] and even unsupervised learning. It not only occupies the mainstream in academia, but also has made great achievements in fashion, advertising, audio and video industries, etc.

B. Image-to-Image Translation

The conditional generative adversarial network [12] highlights, automatic Image-to-Image translation, which teaches input mapping to output images has been applied to various tasks. For instance, generating photographs from sketches [13] or attributes, semantic layouts [14]. Concretely, the generative model of Image-to-Image translation has two parameters or variants that are not parametric and distributes with special algorithms. In an Image-to-Image translation task, a generative model can utilize the distribution of the target domain by generating perfect "fake" data which named translated images to derive from the target domain's distribution [15]. Popular Image-to-Image translation methods include two-domain and multi-domain [16]. First, two-domain can solve problems like computer vision, image processing by using image style transfer in photo editors to benefit from autonomous driving and image colorization [17]. Secondly, multi-domains focus on creating multiple outputs made up different semantic contents or style textures.

In the future, Image-to-Image translation will increase in resolution and generate variable outputs. And the researchers will generalize the Image-to-Image translation methods within the image field to other aspects such as text, language, speech and also multimodal translation tasks.

C. Colorization

Colorization is a computer assisted method of adding color to an image or film [18]. In 1987, Markle invents a colorization process that paints at least one reference frame. It solved the image problem of visual fatigue, by segmenting images into regions and filling in colors.

In the past, coloring was a time-consuming task that required a professional ability to paint. In order to better reduce processing time of sketching as shown in Fig. 1, several interactive technologies have been offered. Levin uses strokes to both indicate the colors of positive pixels and colorized images with optimization where similar pixels have similar colors [19]. Colorization technologies embrace that image division can be well defined in a different space. Take interactive manga colorization technology as an example, several groups gather in a mensurable cluster before colorization. Luan presents an interactive system through color labelling and color mapping for greyscale images, in the absence of complex texture grouping techniques that further improve the system usability [20]. Cheng has a high-quality comprehensive colorization method, but it needs to train on a huge set of reference image data that contains all of the objects [21]. Gupta introduced a sample-based solution to colorize a grey image, adopting a quick cascade feature mapping scheme to voluntarily match matches between reference and target images. While the process still needs to prepare a color example, which is semantically similar to the grey image [22]. In a word, automatic segmentation often failed to identify the complex drawing color, especially the alternative colorized selection of different experimental subjects. They have multiple sketch lines and gray image feature. For instance, a difference between fashion and sporty shoes. To solve this problem, we are going to invent more and more techniques of colorization.

Nowadays, we are making advances in colorization. The machine can produce a colorful sketch with little mistakes which is impossible in the previous time presented in Fig. 2. Subversive colorization techniques such as CIC, LTBC, Pix2pix and DC used CNNs appear [23]. Generative coloring designs extend the unconditional image generation to a better vision. These optimization-based methods and learning-based methods have put forward to successfully colorize line sketches or grayscale photos. They can create dramatic color images without any hand-made help.

In the coming years, especially in the field of coloring sketches, increasingly efficient method will be taken. And the object will be a variable that can inspire further investigation in this direction in the generative modeling of the sketch.

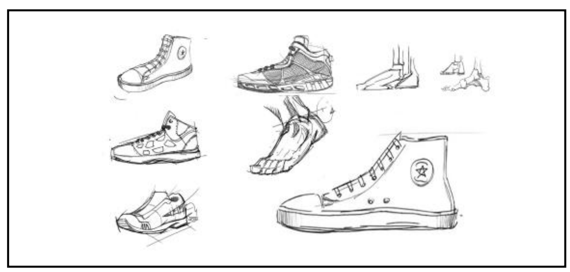


Fig. 1. The footwear handmade sketches.

Fig. 2. Sketches manual colorization.

III. METHOD

To optimize outputs, we construct a conditional adversarial frame to generate a high-resolution shoe image from line sketches. Apart from generation, we use multi-discriminator variants which increase the difficulty of the generator screening sequence. Initially, a basic model of pix2pix is proven. Second, we explain how to accelerate the image reality and resolution of the picture with a GAN frame design. In addition, the adversarial loss is also considered, which can stabilize the training process.

A. The pix2pix Baseline

Pix2pix method is one of the extensions included in the conditional GAN framework for image-to-image translation [24]. The design framework composed of a generator G and a discriminator D. In our research, the aim of our generator G is to colorize the footwear line sketches to complete concept images, while the discriminator D calculated with the purpose of distinguish real images from the forger. The operation consists of a supervised setting. To sum up, the training dataset is provided as a pair of comparative images $\{(s_i, x_i)\}$, where s_i is a purely footwear shape line sketch and x_i is a corresponding real photo. The purpose of conditional GANs is modeling the assumed distribution of real shoe images by enriching the input line sketches via the underlined minimax operation:

$$\min_G \max_D \mathcal{L}_{\text{GAN}}(G, D) \tag{1}$$

And the objection function L_{GAN} (G, D) is given by:

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{(\mathbf{s},\mathbf{x})}[\log D(\mathbf{s},\mathbf{x})] + \mathbb{E}_{\mathbf{s}}[\log(1 - D(\mathbf{s}, G(\mathbf{s}))] \tag{2}$$

The pix2pix method applies U-Net as the generator [25], simultaneously applies fully convolutional patch-base network as discriminator [26]. We put a series of sketches of footwear line and comparative image to the discriminator as our input.

B. Coarse-to-fine Generator

As the resolution of the training images is unstable and the lesser quality, we further improve our pix2pix frame based on the baseline above, the figure presented in Fig. 3. In the beginning, we split the generator into two parts, named G_1 and G_2. Whereas G_1 is a global network of generators, G_2 has a local network of multipliers. Below is an overview of G_1 and G_2. The overall generator network's resolution is 256 x 256, and the local multiplier network is calculated with an output twice the size of the latter. To integrate a better product image resolution, we can join local excess multiplier networks. For example, when using G_1 and G_2 as the generator sections, the resolution of the final output image is G= $\{G_1, G_2\}$ is 512 x 512. However, the resolution grew when G= $\{G_1, G_2, G_3\}$ is 1024 x 1024.

As a vital example of global generator, Johnson et al. [27] constructed a neural network framework for 512 x 512 image style transfer. The framework is composed of three parts including $G_1^{(F)}$, $G_1^{(R)}$ and $G_1^{(B)}$. $G_1^{(F)}$ represented a convolutional front-end, $G_1^{(R)}$ represented as a set of residual blocks and $G_1^{(B)}$ represented a transposed convolutional back-end. The input of the generator is a 256 x 256 image, and output of our process operating three parts is an image of 256 x 256 in resolution.

Apart from global generator, the local enhancer network is composed of three sections as well. And the components, respectively named $G_2^{(F)}$, $G_2^{(R)}$, $G_2^{(B)}$. Firstly, $G_2^{(F)}$ is called a convolutional front-end. Secondly, $G_2^{(R)}$ is called a group of residual blocks, $G_2^{(B)}$ is called a transposed convolutional back-end. We use an input line sketch image to G_2 resolution of which is 512 x 512. Comparing to the global generator, the input of residual block $G_2^{(R)}$ is made up of two feature schedules, the output of $G_2^{(F)}$ and the $G_1^{(B)}$. This manipulation assists to synthesize the entire image information from G_1 to G_2.

In our training course, we operate the global generator before the local enhancer at a certain resolution. After adjusting all the element, the global and local material for shoe line sketches is assembled for image compound. It is obvious that this image generates framework with two-scale is efficient [28]. Other architectures with common usage can be found, such as unconditional GANs or image generator.

C. Multi-scale Discriminators

It's difficult for the GAN discriminator to generate high-resolution images. In order to seperate the high-resolution real and forger images, the discriminator should have a comprehensive production program [29]. Aim to accelerate the framework's ability and decrease the overfitting, our process commands an embedded network and larger convolutional kernels. Our memory space usage, which has a high-resolution image generate function should be large enough for training.

As a maintain problem, we address multi-scale discriminators to cope with it in Fig. 4. We apply multiple discriminators which have a definite network framework but operate multiple image scales. It's obvious that a multi-discriminator which approximate max D V (D, G) can be a good transfer to the generator.

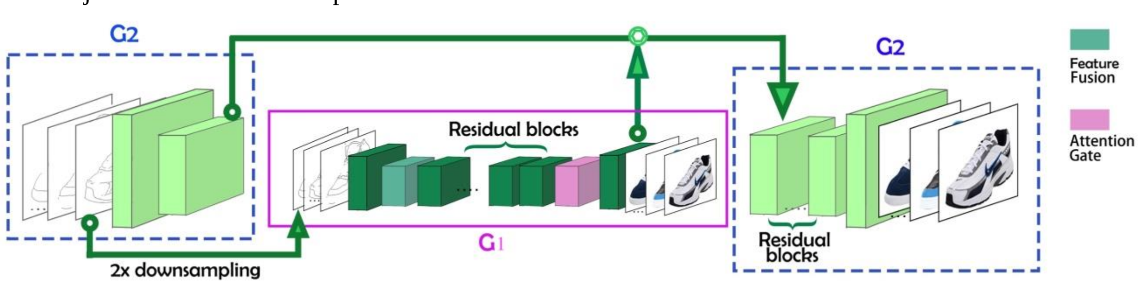


Fig. 3. Two sub-network generators, G_1 for low resolution images, G_2 for high resolution images appending to G_1 input and utilizing the last map from G_1.

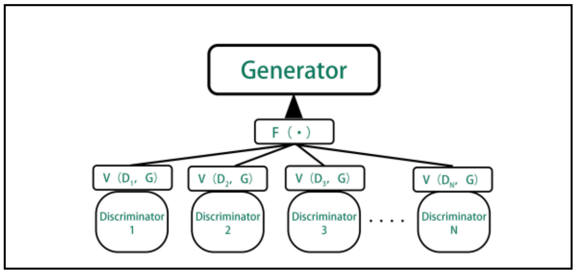


Fig. 4. Multiple discriminators.

To provide a better source for generator, it's necessary to maximize $G(V_i)$. The discriminators are randomly presented, gathering all the optimized V such as the stochastic gradient ascent. $\max_{i\in\{1. N\}} V(D_i, G)$ is the loss of the generator which copes with the problem rendered by the non-convexity of complexes V. In order to minimize the max forces G, we must monitor all of the N discriminators. When we are operating, $\max_{D_i \in D} V(D_i, G)$ doesn't have countless elements, so the above framework is complicated. However, the quantity of the discriminator is still an effective factor.

For example, we construct three discriminators called D_1, D_2, D_3. In particular, we collect the actual images and compound at high-resolution images by several parts to create a pyramid of image of three scales. All of the three discriminators D_1, D_2, D_3 trained to seperate the real and forger at three different scales. To analyze one of the three discriminators, the discriminator that used the roughest scale with the entire field, generating corresponding overall images. Moreover, there is a discriminator with the role of enhancing the finest scale that can produce more specific. This operation allows the calculation to produce higher resolution images that can be an optimized method. By comparing with the previous equation, the multi-scale discriminators equation is:

$$\min_{G} \max_{D_1,D_2,D_3} \sum_{k=1,2,3} \mathcal{L}_{\text{GAN}}(G, D_k) \tag{3}$$

It has been shown that this is a useful method for using multiple GAN discriminators on the same scale. Many researchers have applied the multi-discriminator model to synthesize a better image output [30]. And our purpose is the production of high-resolution images.

D. Improved Adversarial Loss

Our former equation of the objection function LGAN (G, D) has already improved the GAN loss. The steady operation of this loss function helps to produce dramatic results at multiple scales. The process identifies various characteristics of the alternative layers of the discriminators in order to match uncertain representation images. To increase completeness, the following calculation formula is used:

$$\mathcal{L}_{FM}(G, D_k) = \mathbb{E}_{(s,x)} \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^{(i)}(s,x) - D_k^{(i)}(s, G(s)) \right\|_1 \right] \tag{4}$$

where T in the equation means the total quantity of layers and Ni represents the sum of elements in layers. We use the discriminator related to perceptual loss, and we discuss how to make more progress in operational performance. With the combination of GAN loss and characteristic mismatch loss, the equation is:

$$\min_{G} \left(\left(\max_{D_1,D_2,D_3} \sum_{k=1,2,3} \mathcal{L}_{\text{GAN}}(G, D_k) \right) + \lambda \sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k) \right) \tag{5}$$

In this equation, λ is the critical point of two terms, explaining that for loss LFM, Dk is a main extractor.

E. Quantitative Comparison Metrics

To better measure the quality of our experiment, we conduct some evaluation on the synthesized images and true label with these metrics below.

Firstly, SSIM is also a famous quality measurement for testing difference between two image samples [31]. Wang developed the SSIM for calculation related to the image quality of the human visual system [32]. The difference of SSIM is its applications of three factors apart from considering the addition of all image error. The luminance comparison function calculates the similarity of two samples' luminance ($\mu_f\ \mu_g$). The contrast comparison function calculates the closeness of two images' contrast ($\sigma_f\sigma_g$). And the structure comparison function measures the correlation coefficients of image *f* and image *g* noting that σ_{fg} is the covariance. The SSIM index positive values are in [0, 1].

The SSIM can be presented as:

$$SSIM(f,g) = l(f,g)c(f,g)s(f,g) \tag{6}$$

Moreover, the *l (f, g)*, *c (f, g)* and *s (f, g)* It can be described as follow:

$$l(f,g) = \frac{2m_f m_g + c_1}{m_f^2 + m_g^2 + c_1} \tag{7}$$

$$c(f,g) = \frac{2s_f s_g + c_2}{s_f^2 + s_g^2 + c_2} \tag{8}$$

$$s(f,g) = \frac{s_{fg} + c_3}{s_f s_g + c_3} \tag{9}$$

Secondly, MSE (Mean Square Error) is one of the most popular applications which utilized the integrated sample which measured by square intensity differences of various images pixels [33]. In addition, MSE considers the whole reference metric and the final score are better if it values closed to 0. The equation can be defined as:

$$MSE = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} \left[\hat{g}(n,m) - g(n,m) \right]^2 \tag{10}$$

Lastly, PSNR (Peak signal-to-noise ratio) is a common coherence quality measurement for system operation especially in the video or image sample. It is a simple metric that researchers utilize the method to evaluate and develop the visual sample quality. Huynh-Thu et al. show that PSNR could be a useful indicator when the content and codec have analogue. However, the situation changed when the different samples mix together [34]. It is said that the higher PSNR value gains a higher visual sample quality, low quality results from image otherness.

Sometimes noisy samples may affect our evaluation results, and studies have shown that PSNR mixed with MSE have the best performance of evaluating the quality of noisy samples. For example, when applying image degradation PSNR value presented between 30-50 dB for 8-bit data or 16-bit data. The equation can be defined as (11) below, peak value (which we denote by *peakval* in the equation below) means the maximum of the image sample data. For an unsigned 8-bit integer data type, the *peakval* is 255.

$$PSNR = 10\log_{10}(peakval^2) / MSE \tag{11}$$

IV. Experiment

A. Datasets

We perform our method on the Edges2shoes data set, which has been obtained from the official Pix2Pix Datasets directory from Pix2Pix Datasets [35]. The total data set is 50,025, one data pair contains an art line and the color image corresponding to the art line. The task is to map the line art to the color picture, such as the colorization task. To form a model with enhanced generalizability, we randomly sampled 49,825 pairs of images from the original dataset as a training set. After completing the model formation process, we need the test set to evaluate quality, so that the remaining 200 data pairs are used as the test set. Fig. 5 shows edge2shoes sample training information:

B. Implement Details

Our model is improved from pix2pix [15], to check the quality of the model generation, we compared with some models in the colorization task. All of these methods are performed using public codes and the Edges2shoes data set for training and testing in the same experimental environment for comparison purposes. We are experimenting on the computer with an RTX 3090 with 24 GB GPU memory and the PyTorch frame. The model parameters are optimized by the Adam optimizer [36], whose hyper-parameters β_1 and β_2 is set to 0.5 and 0.999, respectively. The training periods are 100 to ensure the convergence of model weights can converge.

Our loss model includes multi-scale GAN loss and feature mapping loss and the latter one plays a significant role in making the training process more stable. We carry out a hyper-parameter search on λ in order to obtain the most effective value for the model. We limit the lambda value between 1 and 10 since λ will bring a blurred image. Fig. 6 shows that all the metrics SSIM, MSE, PSNR achieve best at λ = 5. Consequently, we define λ = 5 in the training process.

Fig. 5. Samples of edge2shoes in the training set.

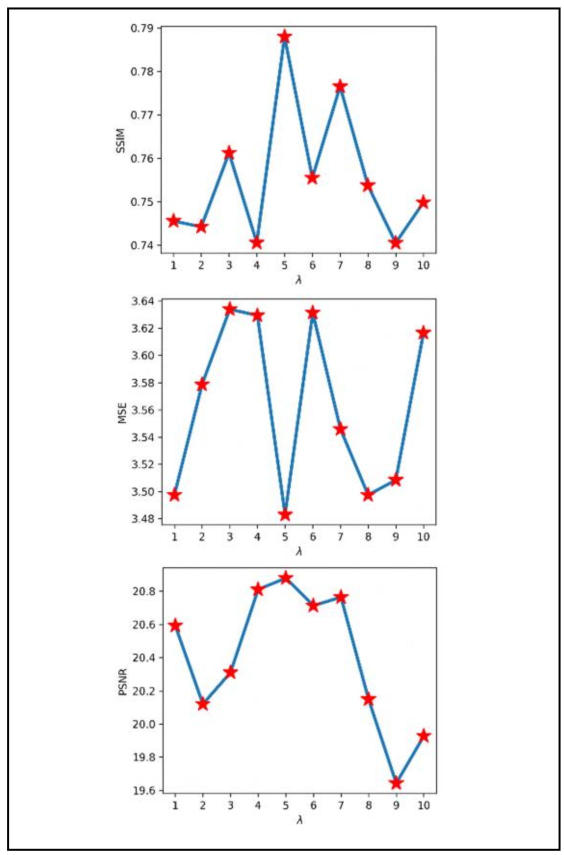


Fig. 6. SSIM, MSE, PSNR with different λ selection.

C. Comparison with state-of-the-arts

From the Table I, we can obtain that our ATTGAN method achieved a significant improvement in SSIM with the SOTA models. Regarding the performance of the PSNR metric for image quality measurement, our method may outperform the other models. In terms of MSE, our ATTGAN method can also achieve the lowest score, which shows that the difference between the image generated and the real image is smaller in pixel. The visualization results can be shown in Fig. 7. It is not difficult to see that our model can get different shaded areas according to precise colors. But they also have local details.

D. Ablation Study

We conduct ablation studies on the Edges2shoes data set to find out how the effectiveness of our upgrades are. From Table II, we can see that all assessment measures are improved, proving that all our improvements are effective. Among them, the metric PSNR has the largest optimization effect.

TABLE I. PERFORMANCE WITH DIFFERENT METHODS OF TRANSLATION ON EDGES2SHOES

Method	Evaluation Metrics		
	SSIM	***MSE***	***PSNR***
CycleGAN	0.5812	4.7832	16.1214
Pix2Pix	0.7192	3.6011	20.1621
NiceGAN	0.6374	4.5172	17.8964
ATTGAN	**0.7881**	**3.4835**	**20.8526**

Fig. 7. Examples of the Edges2shoes colorization by our model.

TABLE II. PERFORMANCE WITH DIFFERENT COMPONENTS OF TRANSLATION ON EDGE2SHOES DATASET

Components			Evaluation Metrics		
Attention	Feature Fusion	Multi-scale D	PSNR	SSIM	MSE
×	×	×	20.1621	0.7192	3.6011
×	×	✓	20.2112	0.7231	3.5813
✓	×	×	20.1922	0.7205	3.5926
×	✓	×	20.5721	0.7649	3.5387
✓	✓	✓	**20.8526**	**0.7881**	**3.4835**

V. DISCUSSION

Here, we try to analyze the influence and importance of several aspects in our experiment.

First of all, the results of the experiment prove that in conditional GANs are able to generate high-resolution images of real footwear photos without any manual control or preparation. The combination of perception loss can improve the image generation results [37]. This approach is effective for studies where the aim is to produce high-resolution image results. At the same time, the researcher has benefited from this method in which pre-training for the exam is prohibited.

What's more, we used multiple discriminators in our GAN network and developed discriminator features exchanging an immersive opponent into effective filters. This GAN network introduces the N-discriminator extension, called Generative Multi-Adversary Networks, which make automatically progress in performance.

Third, the loss matching function fixes the formation where the generator is required to produce image statistics at various scales. This process improves the operational mechanism of the loss function.

Finally, we examine the evaluation measure by PSNR, SSIM, MSE. Treating the pix2pix method as a baseline, we evaluate that ATTGAN has the best performance compared with CycleGAN, Pix2pix and NiceGAN. By three translation methods on Edges2shoes, $\lambda = 5$ in the training procedure gain the best consequence to all. There is no doubt that all measures of evaluation have the highest ratings when it comes to drawing attention, presenting a characteristic and choosing a multi-scale discriminator. While others may have less effort and just below the best.

We believe that the automatic footwear colorization gives designers a chance to minimize the coloring time of the finish drawing line. To implement this automatic colorization footwear design, we are improving the resolution based on an earlier study in another area. At the same time, we make certain multi-discrimination to help the quality of the generation image. We may conclude that the approach for footwear colorization helps to accelerate the effectiveness of the coloring process. However, there are still limitations of the experiment. When the sketch line presented too complex or messy, it would have unavoidable identification errors throughout period. Nevertheless, the color slightly changed which caused by computing constraint.

In the future, designers may have more measures for eliminating colorization time. The development of automatic colorizing solves the problems for selecting color styles and finishing handmade draft. Furthermore, it becomes possible for customers to do DIY based on their individual sketching creation. With the help of artificial intelligence, the distance between manufacturer and consumer brings closeness. It might totally change the current situation of the footwear industry, especially in the post-pandemic era. Consumers may realize their own ideas by interaction media, which even reduce the working-hours of designers. It is no doubt that co-creation of products will be a trend of the times.

VI. CONCLUSION

In this research, we introduce a conditional GAN network for coloring sketches of shoes. The whole process divided into generator and discriminator, which help to perform better for coloring shoes. These methods decrease the difficulties of comparing evaluation measures. The results of our experiments are consistently visual. We believe that our method for coloring shoe sketches can have a good inspiration in the area of research to generate color sketch image samples.

REFERENCES

[1] Eitz M, Hays J, Alexa M. How do humans sketch objects? [J]. ACM Transactions on graphics (TOG), 2012, 31(4): 1-10.

[2] Hesse C. Image-to-image demo: Interactive image translation with pix2pix-tensorflow[J]. Affinelayer. com, February, 2017, 19.

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[5] Schmarje L, Santarossa M, Schröder S M, et al. A survey on semi-, self- and unsupervised learning for image classification[J]. IEEE Access, 2021, 9: 82146-82168.

[6] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of big data, 2019, 6(1): 1-48.

[7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[9] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[10] Wang X, Liu H. Data supplement for a soft sensor using a new generative model based on a variational autoencoder and Wasserstein GAN[J]. Journal of Process Control, 2020, 85: 91-99.

[11] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

[12] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[13] Sangkloy P, Lu J, Fang C, et al. Scribbler: Controlling deep image synthesis with sketch and color[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5400-5409.

[14] Karacan L, Akata Z, Erdem A, et al. Learning to generate images of outdoor scenes from attributes and semantic layouts[J]. arXiv preprint arXiv:1612.00215, 2016.

[15] Pang Y, Lin J, Qin T, et al. Image-to-image translation: Methods and applications[J]. IEEE Transactions on Multimedia, 2021.

[16] Lee H Y, Tseng H Y, Mao Q, et al. Drit++: Diverse image-to-image translation via disentangled representations[J]. International Journal of Computer Vision, 2020, 128(10): 2402-2417.

[17] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks[J]. Advances in neural information processing systems, 2017, 30.

[18] Zhang R, Isola P, Efros A A. Colorful image colorization[C]//European conference on computer vision. Springer, Cham, 2016: 649-666.

[19] Levin A, Lischinski D, Weiss Y. Colorization using optimization[M]//ACM SIGGRAPH 2004 Papers. 2004: 689-694.

[20] Liu X, Luan R, Huang C. A novel encoding method for visual two-dimensional barcode using pattern substitution[C]//2010 International Conference on Machine Vision and Human-machine Interface. IEEE, 2010: 662-665.

[21] Cheng Z, Yang Q, Sheng B. Deep colorization[C]//Proceedings of the IEEE international conference on computer vision. 2015: 415-423.

[22] Gupta R K, Chia A Y S, Rajan D, et al. Image colorization using similar images[C]//Proceedings of the 20th ACM international conference on Multimedia. 2012: 369-378.

[23] Kumar M, Weissenborn D, Kalchbrenner N. Colorization transformer[J]. arXiv preprint arXiv:2102.04432, 2021.

[24] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the

[25] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[26] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[27] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[28] Burt P J, Adelson E H. The Laplacian pyramid as a compact image code[M]//Readings in computer vision. Morgan Kaufmann, 1987: 671-679.

[29] Yildirim G, Jetchev N, Vollgraf R, et al. Generating high-resolution fashion model images wearing custom outfits[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[30] Durugkar I, Gemp I, Mahadevan S. Generative multi-adversarial networks[J]. arXiv preprint arXiv:1611.01673, 2016.

[31] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th international conference on pattern recognition. IEEE, 2010: 2366-2369.

[32] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

[33] Sara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. Journal of Computer and Communications, 2019, 7(3): 8-18.

[34] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment[J]. Electronics letters, 2008, 44(13): 800-801.

[35] Qu Y, Chen Y, Huang J, et al. Enhanced pix2pix dehazing network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8160-8168.

[36] Zhang Z. Improved adam optimizer for deep neural networks[C]//2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Ieee, 2018: 1-2.

[37] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

Multi-Exposure Image Fusion based on Window Segmentation and a Laplacian Pyramid for Chip Package Appearance Quality Detection

Fei Hao[1], Jiatong Song[2], Jiahao Sun[3], Yang Fu[4]
Kangni Industrial Technology Research Institute, Nanjing Institute of Technology, Nanjing, China

***Abstract*—A heterogeneous material image enhancement method based on multi-exposure image fusion is proposed to address the problem of obtaining high-quality images from the single imaging of chips containing two extremely different reflectivity materials. First, a multi-exposure image fusion algorithm based on window segmentation and Laplacian pyramid fusion is proposed. Then, orthogonal experiments are used to optimize the parameters of the imaging system. Next, a method based on information entropy and average gray intensity is utilized to calculate the imaging exposure times of two heterogeneous materials, and two exposure time ranges are obtained that are appropriate for regions with high and low reflectivity. Finally, the subjective and objective experimental evaluations are conducted after the multi-exposure image set has been established. The results show that the fused image has a good visual effect, the information entropy is 6.29, and the average gray intensity is 131.56. In addition, time consumption is reduced by an average of 20.3% compared to the Laplace pyramid strategy. The heterogeneous material enhancement method based on multi-exposure image fusion proposed in this paper is effective and deserving of further research and application.**

Keywords*—Image fusion; multi-exposure; Laplacian pyramid; window segmentation; chip package

I. Introduction

Chip defect detection typically includes size measurement, character detection, stain detection and pin defect detection [1]. It is found that because chips in ordinary cameras contain both high reflectivity metal and low reflectivity black plastic, they cannot capture all information through a single image when collecting images [2]. At present, there are devices for directly acquiring high dynamic images [3], but such hardware devices have high cost and few applications, so software algorithm enhancement methods are widely considered instead.

This paper aims to propose a new multi-exposure image fusion method to improve the image quality of an object composed of a variety of materials with very different reflectivity. The main contributions are as follows:

- An image fusion method based on window segmentation and Laplace pyramid was proposed.
- The optimal imaging parameters of the established machine vision system were determined by the orthogonal experiments.
- A strategy for determining the optimal exposure time based on information entropy and average gray intensity was proposed.
- The experimental results show that the image information entropy and average gray intensity of the fused image by our method were 6.29 and 131.56 respectively and the time was averagely reduced by about 1.26 *s*.

The rest of this paper is organized as follows: the first section is the introduction. The second section reviews the related work in multi-exposure image fusion. The third section describes the image fusion algorithm. In the fourth section, the main factors affecting the imaging quality are determined by orthogonal testing, and the level of the influencing factors is optimized. In the fifth section, through information entropy and average gray intensity, the optimal exposure time for imaging objects with multiple heterogeneous materials is determined. The sixth section discusses the experiment and the analysis of the results. Finally, conclusions are presented.

II. Related Works

Currently, image quality enhancement algorithms are divided into two categories. One is based on single source image enhancement [4], which primarily use the spatial domain and transform domain to enhance a single image to improve the problems of detail loss and uneven lighting. Common spatial methods include grayscale transformation, histogram equalization, contrast enhancement and other methods [5]. Among them, the histogram equalization method is the most commonly used airspace method. The global histogram equalization map is highly efficient, but the enhanced image will easily lose details. Therefore, Celik [6] adopted the local histogram equalization algorithm based on information entropy to solve the problem of image texture loss. The local histogram equalization map can enhance the local details of the image but lacks global integrity. Wavelet transform [7], discrete wavelet transform [8], stationary wavelet transform [9] and other transform domain methods distinguish the basic information and detailed information of the image and perform multi-exposure image fusion. At present, the transform domain algorithm based on the wavelet transform can achieve the effect of enhancing the image [10]. Such methods can enhance the details while taking into account the global whole but also magnify the noise in the image. Therefore, the final result of the single source image enhancement method experiences noise and loss of detail, which is typically accepted in the field of natural images but is not suitable for size measurement and defect detection.

The other category is image fusion based on multi-source images. Image fusion refers to obtaining different image sequences of the same scene with imaging sensors and combining the details and complementary information of the image sequence to obtain a rich and comprehensive image [11]. Multi-Exposure Fusion (MEF) is one of the branches derived from image fusion; that is, it fuses multiple images with different exposures in the same scene into a new image, preserving the brightest and darkest details in the scene [12]. From the perspective of image decomposition, multi-exposure image fusion can be divided into pixel-based and image block-based fusion methods. Goshtasby [13] first proposed image block-based fusion and calculated the information entropy to select the best image block for fusion, but there was some distortion in the results. Qin Lyu used nonuniform triangulation to segment images [14]. T Prabhakara Rao performed feature-level image fusion based on contour blocking, improved the distortion and played a crucial role in developing its subsequent feature extraction and detection [15]. Compared with the method of nonuniform division of images, the fusion algorithm speed block based on uniform division of image blocks is more suitable for the application of industrial detection occasions, but it mainly emphasizes the selection of the optimal block of the image sequence, which frequently results in issues with discontinuous images and obvious stitching traces. In contrast, Mertens [16] proposed a pixel-based multi-exposure fusion method, which can make up for the shortcomings of discontinuous images but has a high computational complexity and low efficiency. An extremely critical method in pixel-based image fusion is pyramid decomposition [17], which constructs an image pyramid from the input image, adds a weight matrix, and finally combines the two to obtain the final fusion result. Ashish et al. adopted multiresolution fusion based on a Laplacian pyramid [18], measured by entropy and contrast, which is characterized by seamless fusion and excellent improved details. Zhong Qu and others improved pyramid decomposition and applied it to image fusion, improving the common artifact problem of multi-exposure fusion [19] and retaining more local details. The pyramid algorithm can typically preserve good image edges and textures, but the processing time is generally very long. The fusion method based on image blocks can greatly improve the computational efficiency and remove random noise. Therefore, it is expected that combining the pyramid algorithm with the image block algorithm will allow the fused image to maintain better details and improve the operational efficiency of the algorithm.

The optimal exposure time should be primarily determined according to the material characteristics of a specific object for developing the image fusion methods. The image gradient was employed in the fusion method proposed by Turgut et al. [20]. Image information entropy was adopted by Kataoka et al. as the only indicator to determine the exposure time for image fusion [21]. Camera response curve was uniquely selected and used for exposure fusion by Zhang et al. [22]. Thus it can be seen that the common indicators of exposure fusion include image gradient, image information entropy and camera response curve, and they were used singly at present. However, these above methods may be invalid and inaccurate for objects with multi-materials. Therefore, this paper will also explore a new multivariate strategy to determine the optimal exposure time for imaging objects with muti-materials.

III. Multi-Exposure Image Fusion Method

A. Image Fusion Algorithm based on Laplacian Pyramid

The source image G_0 undergoes multiple Gaussian fuzzy filtering and downsampling operations and continuously reduces the image size to obtain an image sequence $\left(G_0, G_1, G_2, \ldots, G_N\right)$. The Gaussian pyramid is obtained by arranging G_0 at the bottom and G_N at the top. The essence of the Gaussian pyramid is the multiscale representation of the image signal, as shown in Formula (1). The Laplacian pyramid is obtained by up sampling and interpolation of the Gaussian pyramid [23], and the image sequence can be represented by $\left(LP_0, LP_1, LP_2, \ldots, LP_N\right)$, as shown in Formula (2).

$$G_l(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m,n) G_{l-1}(2i+m, 2j+n) \tag{1}$$

where $w(m,n)$ is a Gaussian convolution kernel with a fixed size.

$$\begin{cases} LP_l = G_l - G_{l+1}^* & 0 \le l < N \\ LP_N = G_N & l = N \end{cases} \tag{2}$$

where G_{l+1}^* is calculated according to Formula (3):

$$G_l^* = 4 \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m,n) G_l\left(\frac{i+m}{2}, \frac{j+n}{2}\right) \tag{3}$$

For multi-exposure image fusion based on the Laplacian pyramid strategy, each layer of images were first processed in multiple Laplacian pyramids according to the strategy in Formula (4) to obtain the fused Laplacian pyramid image sequence, and then obtain the fused source image according to the method in Formula (5).

$$FLP_l = \sum_{k=1}^{H} \lambda_k \cdot \left(LP_l\right)_k \tag{4}$$

where H is the number of multi-exposure images, $\left(LP_l\right)_k$ represents the l-th layer of the Laplacian pyramid decomposition image of the k-th image, λ_k is the weight coefficient corresponding to the k-th image, and FLP_l is the weighted fusion result of the l-th decomposition image of the H multi-exposure image.

$$\begin{cases} FG_N = FLP_N, & 0 \le l < N \\ FG_l = FLP_l + FG_{l+1}^*, & 0 \le l < N \end{cases} \tag{5}$$

In the formula, FG_l is the Laplacian pyramid decomposition image of layer l of the fused image, and FG_{l+1}^* is calculated according to Formula (3).

The image fusion based on the Laplacian pyramid strategy comprehensively uses the information of the source image at different spatial frequency levels. Compared with traditional fusion methods, there is no obvious stitching trace, but it also has the problem of long computing times.

B. Multi Exposure Image Fusion Algorithm based on Window Segmentation

Divide the H multiple exposure light source images evenly to obtain an image block with size $m \times n$. The j-th image block in the i-th source image is represented by $A_{i,j}$. Select an appropriate image quality evaluation index to quantize $\{A_{i,j}\}$, and select the optimal image block B_j. As shown in Formula (6), the fused image is finally obtained by "splicing" all the B_j. The basic process of the multi-exposure image fusion algorithm based on window segmentation is shown in Fig.1.

$$B_j = Optimal\left(\text{Indicator}\left(\left\{A_{i,j}\right\}\right)\right) \tag{6}$$

In Formula (6), $i = 1,2,\ldots,H$, $j = 1,2,\ldots,M$, M is the number of image blocks, $\text{Indicator}(\cdot)$ represents the quantitative evaluation function of image quality, and $Optimal(\cdot)$ represents the optimization function.

Optimizing the exposure quality evaluation index of image blocks and the image block "splicing" algorithm are the core of the multi-exposure image fusion algorithm based on window segmentation. The uniform block-based fusion algorithm has high efficiency, but an inappropriate evaluation index and "splicing" algorithm easily lead to splicing traces. The nonuniform block fusion algorithm has the advantages of obvious edge information and large brightness width, but the efficiency of such methods is usually not high.

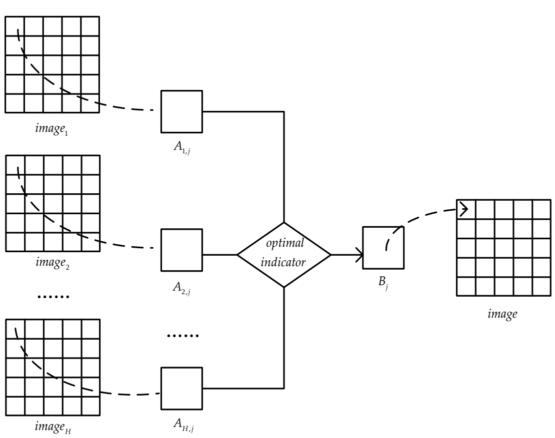


Fig. 1. Window segmentation fusion algorithm

C. Image Fusion Algorithm based on Window Segmentation and Laplacian Pyramid

The image fusion algorithm based on window segmentation and Laplacian pyramid combines the characteristics of high efficiency of window segmentation method and less stitching trace of Laplacian pyramid method. Same as window segmentation, first, the H source image is divided into M image blocks of size $m \times n$, and the j-th image block of the i-th source image is expressed as $A_{i,j}$. Then, the appropriate image quality evaluation index was selected to quantize and sort $A_{i,j}$, whose total quantity is M, and select the best N image blocks $B_{k,j}$. Then, the image fusion algorithm based on Laplace pyramid sideratio is adopted, and B_j is obtained by fusing N image blocks $B_{k,j}$. Finally, the fused image is obtained by "splicing" all the B_j. The basic process of the image fusion algorithm based on window segmentation and Laplacian pyramid is shown in Fig. 2.

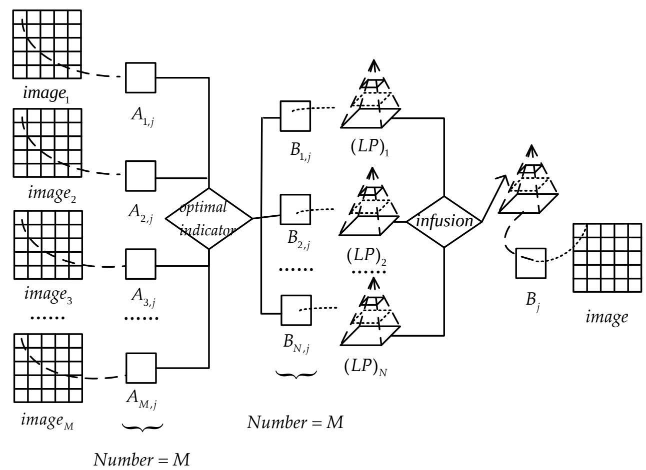


Fig. 2. Image fusion algorithm based on window segmentation and Laplacian pyramid

For vision systems, exposure time is not the only factor affecting the imaging system. Lighting, camera sensors and lenses are key factors that affect imaging. Therefore, before multi-exposure image fusion, this paper innovatively controls the image acquisition conditions and uses orthogonal experiments to optimize the parameters of the multi-exposure imaging system.

IV. Optimal Parameters Selection Method for Imaging System and Its Implementation

A. Machine Vision System

The experiment object of this paper is a semiconductor chip with a size of approximately 4mm × 4mm. A dual camera image acquisition system consisting of two telecentric lenses was built based on the characteristics of small distortion, as shown in Fig. 3.

Fig. 3. Binocular telecentric vision system

B. Effect Factors of Imaging Quality and Its Levels

There are many factors that affect the image quality during the design of a machine vision system. Among them the ambient light, the artificial light and the camera gain are the three most concerned parameters. It is difficult to control the ambient light to reduce the impaction on image quality and thus the dark rooms are typically used to isolate ambient light for image acquisition. The image quality will sharply vary with the light source intensity, light source direction and other illumination conditions. Therefore how to optimize arrangement of artificial lights to ensure the acquisition of high-quality images was studied. In the darkroom, a larger gain may be employed to have more luminous flux to increase the image brightness. However, a larger camera gain may amplify noise and thus reduce the image quality. Accordingly, how to determine the optimal camera gain is also very important. Therefore, in the orthogonal experimental design, the three factors of artificial light source, ambient light and gain must be analyzed to determine their optimal levels to provide stable experimental conditions for multi-exposure fusion research.

TABLE I. EFFECT FACTORS OF IMAGING QUALITY AND THEIR LEVELS

Levels	Factors		
	A	***B***	***C***
1	low	4.00	without a dark room
2	medium	5.00	with a dark room
3	high	6.00	/

The above three factors of imaging quality and their levels are tabulated in Table I. Factor A represents the artificial light and has three levels: low, medium and high light intensity of backlights. Factor B represents the gain and has three levels: 4.00 dB, 5.00 dB and 6.00 dB. Factor C represents the ambient light and has two levels: with or without a dark room.

C. Image Quality Evaluation Indicators

An image collected by the machine vision system is an 8-bit single channel gray image. The image quality evaluation indicators include information entropy, average gray intensity, variance and average gradient, which are defined as follows:

1) Information entropy H

$$H = \sum_{i=0}^{255} p_i \log p_i \tag{7}$$

where p_i represents the proportion of pixels whose gray value is *i* in the image.

2) Average gray intensity U

$$U = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j)}{m \times n} \tag{8}$$

where $I(i,j)$ represents the gray intensity at (i,j) in the image. m and n are the size of the image along two coordinate directions.

3) Variance S

$$S = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} (I(i,j)-U)^2}{m \times n}} \tag{9}$$

where $I(i,j)$ represents the gray intensity at (i,j) in the image. m and n are the size of the image along two coordinate directions. U is the average gray intensity of the image.

4) Average gradient G

$$G = \frac{1}{(m-1)(n-)} \sum_{i=1}^{m-1}\sum_{j=1}^{n-1} \sqrt{\frac{1}{2}\left\{\begin{matrix} \left[I(i+1,j)-I(i,j)\right]^2 \\ +\left[I(i,j+1)-I(i,j)\right]^2 \end{matrix}\right\}} \tag{10}$$

where $I(i,j)$ represents the gray intensity at (i,j) in the image. m and n are the size of the image along two coordinate directions.

5) Comprehensive evaluation indicator W

$$W = k \times H \times U \times S \times G \tag{11}$$

where k is a coefficient that is employed to normalize W to be between 0 and 1.

D. Orthogonal Experiments and Results

The orthogonal experimental scheme was established as shown in Table II according to the orthogonal experimental design method.

TABLE II. ORTHOGONAL EXPERIMENTAL SCHEME L6 ($3^2 \times 2$)

Levels	Factors		
	A	***B***	***C***
1	1	1	1
2	2	1	2
3	3	2	1
4	1	2	2
5	2	3	1
6	3	3	2

According to the above orthogonal experimental scheme tabulated in Table II, the experiments were carried out with the left camera and the right camera. Six images were collected with the left camera, and the evaluation results of image quality are shown in Table III.

TABLE III. RESULTS OF 6 EXPERIMENTS WITH THE LEFT CAMERA

No.	Factors			Image Quality Evaluation Indicators				
	A	B	C	Information entropy	Average gray intensity	Variance	Average gradient	Comprehensive indicator
1	1	1	1	0.63	242.11	2802.76	0.21	0.908
2	2	1	2	0.61	241.61	3026.27	0.18	0.819
3	3	2	1	0.65	243.29	2431.72	0.23	0.860
4	1	2	2	0.61	241.56	3042.61	0.19	0.840
5	2	3	1	0.65	243.34	2411.14	0.23	0.893
6	3	3	2	0.63	242.62	2703.44	0.20	0.830

The results of orthogonal experiments were calculated by using the comprehensive indicator, whose values are tabulated in Table V. k_i is the sum of comprehensive indicators at the i-th level of factors and $\bar{k_i}$ is the mean value of k_i. r is the range of k_i, and σ is the variance of k_i. The larger the range of a factor is, the more significant the impact of the factor on image quality. The larger the $\bar{k_i}$ of a factor is, the more significant the impact of the level on image quality.

TABLE IV. RESULTS OF 6 EXPERIMENTS WITH THE RIGHT CAMERA

No.	Factors			Image Quality Evaluation Indicators				
	A	B	C	Information entropy	Average gray intensity	Variance	Average gradient	Comprehensive indicator
1	1	1	1	0.7000	241.280	2833.36	0.2400	1.1510
2	2	1	2	0.6600	240.640	3087.94	0.1900	0.9460
3	3	2	1	0.7100	243.070	2244.70	0.2600	1.0170
4	1	2	2	0.6700	240.530	3127.14	0.2000	0.9900
5	2	3	1	0.7100	243.050	2253.85	0.2800	1.0800
6	3	3	2	0.6900	242.260	2522.78	0.2200	0.9170

TABLE V. RESULTS OF ORTHOGONAL EXPERIMENTS BY COMPREHENSIVE EVALUATION INDICATORS

Camera	Factors	Evaluation Indexes							
		k_1	k_2	k_3	$\bar{k}_1$	$\bar{k}_2$	$\bar{k}_3$	r	σ
Left	A	1.748	1.712	1.690	0.8740	0.8560	0.8450	0.0290	0.00043
	B	1.727	1.700	1.723	0.8635	0.8500	0.8615	0.0135	0.00011
	C	2.661	2.489	/	0.8870	0.8300	/	0.0570	0.00162
Right	A	2.141	2.026	1.934	1.0705	1.0130	0.9670	0.1035	0.00538
	B	2.097	2.007	1.997	1.0485	1.0035	0.9985	0.0500	0.00152
	C	3.248	2.853	/	1.0827	0.9510	/	0.1317	0.00867

From Table III, Table IV and Table V, it can be seen that the ambient light has the greatest impact on image quality, camera gain has the smallest impact on image quality and backlight intensity has a moderate impact on image quality, i.e., C > A > B. The image quality is the highest when the three influencing factors are set to the first level, i.e., the optimal parameter combination is $C_1A_1B_1$.

V. DETERMINATION OF EXPOSURE PARAMETERS BASED ON INFORMATION ENTROPY AND AVERAGE GRAY INTENSITY

A. *Information Entropy-Exposure Time Model for Images of Single Material Object*

The exposure time required in imaging is related to the reflectivity of the object surface. Dark areas with low reflectivity require longer exposure times. In contrast, bright areas with higher reflectivity require shorter exposure times. When imaging a single material object or an object with multiple materials but little difference in reflectivity, the relation curve of the information entropy of the acquired image to the exposure time is a downward parabola. The information entropy first increases with increasing exposure time. When the information entropy reaches the maximum value, it will decrease with increasing exposure time, as shown in Fig. 4.

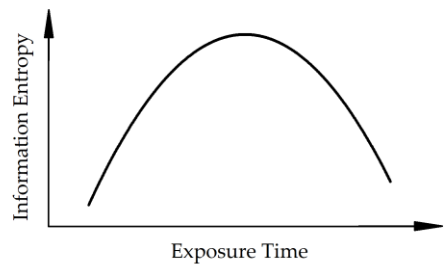


Fig. 4. Theoretical relationship between exposure time and image entropy

According to Fig. 4, when imaging a single material object or an object with multiple materials but little difference in reflectivity, the optimal exposure time can be set to $[t_0-t_1, t_0+t_2]$, where the exposure time is t_0 and the information entropy is maximal. t_1 and t_2 are two positive numbers that are smaller than t_0.

B. *Information Entropy-Exposure Time Model for Images of Multi-Material Objects*

Different from single material object imaging, a multi-material object imaging has multiple peaks in its information entropy-exposure time curve, as shown in Fig. 5. The above method for determining the optimal exposure time cannot be directly used. Therefore, it is necessary to study the method of determining the optimal exposure time according to the characteristics of the information entropy-exposure time curve.

C. *Determination of Optimal Exposure Time based on Information Entropy and Average Gray Intensity*

The chip is composed of pins with high reflectivity and plastic packages with low reflectivity. As shown in Fig. 5, two peaks appear when the exposure time is 34200 μs and 500000 μs. If the selected exposure time is less than 34200 μs or more than 500000 μs, it will not be conducive to high-quality imaging of chips made of both materials. Therefore, the best exposure time must be between 34200 μs and 500000 μs.

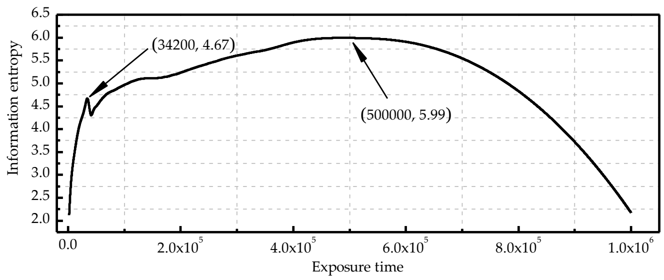


Fig. 5. Information entropy-exposure time curve of objects with two materials

The length of the above exposure time interval determined according to the information entropy is still relatively large. When collecting the pin image and the plastic package image, there is only one lower limit or upper limit of the exposure time. Therefore, the above interval needs to be further refined to obtain two subsections, which are applicable to collecting the pin image and the plastic package image. For this purpose, the relationship between the rate of the average gray intensity of images and the exposure time was analyzed. Fig. 6 and Fig. 7 show the first-order difference and the second-order difference of the model of image average gray intensity and exposure time, respectively.

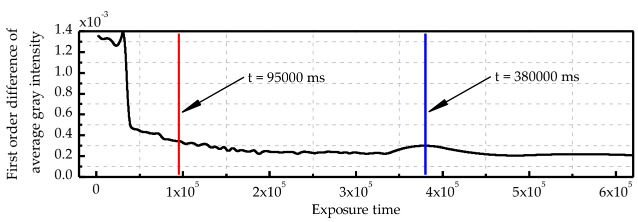


Fig. 6. First-order difference of image average gray intensity vs. exposure time

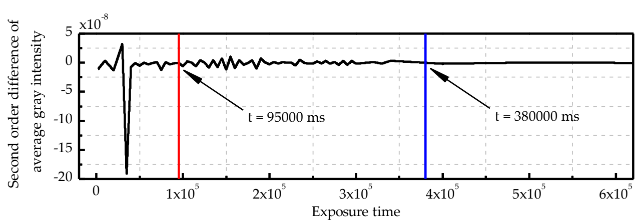


Fig. 7. Second-order difference of image average gray intensity vs. exposure time

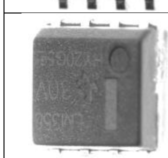
Fig. 6 shows that when the exposure time is greater than 34200 µs but less than 95000 µs, the first-order difference in the average gray intensity of the image drops steadily. However, when the exposure time is greater than 95000 µs, it decreases while there is a certain amount of jitter. In Fig. 7, when the exposure time is between 34200 us and 95000 µs, the second-order difference of image average gray intensity basically shows a horizontal line, although it fluctuates slightly, while when the exposure time is greater than 95000 µs, the second-order difference of image average gray intensity has relatively large fluctuations. With increasing exposure time, low reflectivity targets can be imaged better. When the exposure time is 380000 µs, a peak appears in the first-order difference curve of image average gray intensity vs. exposure time, as shown in Fig. 6. This means that the average gray intensity of the image has a relatively sharp change at approximately 380000 µs. According to the above analysis, the exposure time intervals for two targets with different materials were determined to be [34000, 95000] and [380000, 500000], respectively. It should be noted that the first exposure time is rounded off appropriately to make calculation more convenient.

Then, the above two exposure time intervals were divided into several segments, and multiple images were accordingly collected and fused. The areas with different materials of chips in the fused image are clear. A camera parameter, defined to be the product of exposure time and gain, was used to reasonably determine the segmentation. According to the optimal gain determined by the orthogonal experiments and the exposure times determined by the peak of the information entropy vs. exposure time curve, the camera parameters for two material targets are 136800 and 2000000, respectively. Compared with imaging objects with high reflectivity materials, the camera parameter should be larger when imaging objects with low reflectivity materials. Accordingly, the segmentation of the interval [34000, 95000] should be much smaller than that of the interval [380000, 500000]. Consequently, the former interval was divided into 20 subintervals, and the latter interval was divided into 10 subintervals. The segmentation of the former interval is 3100µs and that of the latter interval is 12000 µs.

VI. Experiments Results and Discussion

Thirty multi-exposure images were collected with the machine vision system shown in Fig. 3. During image acquisition, the previously determined optimal exposure parameters were adopted. The proposed multi-exposure image processing algorithm was implemented by using MATLAB® 2015. An ordinary PC with a 64-bit operating system was used as an image processing system. Some collected images are included in Table VI.

TABLE VI. Source Images with Different Exposure Times

Images	Exposure times (µs)	Descriptions
	45000	The pins can be imaged, but the brightness is slightly dark. The characters are completely invisible.
	85000	The pins are incomplete and the characters are still invisible.
	400000	The pins are missing and the characters are visible.

A. *Subjective Evaluation*

The information entropy method, window segmentation method and the proposed joint method were employed to fuse the images, and the results are shown in Fig. 8.

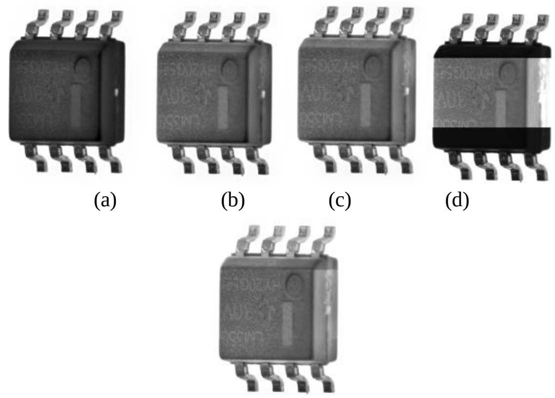


Fig. 8. Results of image fusion. (a) image fusion result based on the Laplacian pyramid strategy with σ = 0.1; (b) image fusion result based on the Laplacian pyramid strategy with σ = 0.2; (c) image fusion result based on the Laplacian pyramid strategy with σ = 0.3; (d) image fusion result based on ordinary window segmentation; (e) image fusion result based on the proposed joint method

Fig. 8(a) to 8(c) show the image fusion results based on the Laplacian pyramid strategy. From the perspective of character clarity, the characters on the plastic package in Fig. 8(a) are very vague. The clarity of characters on the plastic package in Fig. 8(b) is obviously improved compared with that of Fig. 8(a). Compared with Fig. 8(a) and Fig. 8(b), the characters on the plastic package in Fig. 8(c) are the clearest. From the perspective of the pin area contrast, the pin area in Fig. 8(b) has the highest contrast. Fig. 8(d) is the result of image fusion based on window segmentation. Although the contrast of the pin area is high and the characters are relatively clear, there is an obvious "fracture" trace in the fused image, which will cause relatively large interference in subsequent image processing and thus may easily cause measurement errors. Fig. 8(e) is the image fusion result based on the proposed joint method. The characters are clear, the pin area contrast is high, the "fracture" trace is significantly eliminated, the details are well protected, and the visual effect is good.

B. *Objective Evaluation*

The above image fusion results were quantitatively analyzed and compared by using three indices, namely, information entropy, average gray intensity and time consumption. The results are tabulated in Table VII.

According to Table VII, in image fusion based on the Laplacian pyramid strategy, the value of σ will affect the information entropy and average gray intensity of the fused image, while it has little impact on the image fusion efficiency. When σ = 0.1, the information entropy of the fused image is 6.1473, and the average gray intensity of the fused image is 89.9575. Compared to σ = 0.1, when σ = 0.2, the information entropy of the fused image increases by 1.27%, and the average gray intensity of the fused image increases by 33.65%. Compared to σ = 0.1, when σ =0.3, the information entropy of the fused image increases by 2.12%, and the average gray intensity of the fused image increases by 38.09%.

TABLE VII. OBJECTIVE INDICATORS OF DIFFERENT FUSION RESULTS

Fusion Method	**Information Entropy**	**Average Gray Intensity**	**Time (s)**
Laplace pyramid strategy with σ = 0.1	6.1473	89.9575	6.246965
Laplace pyramid strategy with σ = 0.2	6.2252	120.2241	6.245941
Laplace pyramid strategy with σ = 0.3	6.2778	124.2241	6.220413
Ordinary window segmentation	**6.4937**	117.4303	5.112153
Proposed method	6.2922	**131.5629**	**4.967878**

Image fusion based on window segmentation has the largest information entropy, which is 5.63% higher than the minimum value. However, the average gray intensity of the fused image is not high, ranking fourth (penultimate). The efficiency of the fusion algorithm is improved. Compared with the three image fusion methods based on the Laplacian pyramid strategy, the efficiency of the algorithm is increased by 22.20%, 22.18 %and 21.68%.

Image fusion based on the proposed joint method has higher information entropy, which is better than the three image fusion methods based on the Laplacian pyramid strategy and second to the image fusion method based on window segmentation. The average gray intensity of the fused image is the highest. Compared with the other four methods, the average gray intensity is increased by 46.25%, 9.43%, 5.91% and 12.03%. The fusion algorithm takes the least time. Compared with the image fusion method based on the Laplacian pyramid strategy, it takes approximately 1.3 seconds less. Compared with the image fusion method based on window segmentation, the efficiency of the algorithm is also improved, and this algorithm takes 2.90% less time.

VII. CONCLUSIONS

To address the difficulty of acquiring high-quality images with single imaging for chips made of two materials with extremely different reflectivity, an image enhancement method based on multi-exposure image fusion was proposed in this paper. First, a joint image fusion algorithm based on the Laplacian pyramid and window segmentation was proposed, improving image quality and reducing processing time. Then, the factors that affect the imaging quality their levels were analyzed and the imaging parameters were optimized through orthogonal tests. After that, a method of determining the exposure time based on information entropy and first and second order difference of average gray intensity was studied. Finally, multi-exposure image sets were established, and experiments and subjective and objective evaluations were subsequently performed. The results show that the fused image has a good visual effect, its information entropy was 6.29, and its average gray intensity was 131.56. Furthermore, compared with the Laplace pyramid strategy, the time consumed was averagely reduced by 1.26 *s*. The fusion algorithm has the advantages of being less time consuming and high efficiency.

It has been demonstrated that the image enhancement method based on multi-exposure image fusion proposed in this paper is effective for an imaging object composed of two materials with great difference in reflectivity. However, obtaining high-quality images is not the ultimate goal, and it is necessary to further verify the effectiveness of the method in a defect detection or size measurement task. Furthermore, the proposal can be further explored and applied to the objects composed of more than two kinds of materials. On the other hand, although the proposed algorithm takes the shortest time in the above experiments, the image fusion process still takes 4.6 *s*. So the efficiency should be further improved.

ACKNOWLEDGMENT

All authors express their gratitude to the Editor and the anonymous Reviewers for their valuable and constructive comments. And this research was funded by the National Natural Science Foundation of China under grant number 51705238, the Promotion Project for Modern Agricultural Machinery Equipment and Technology Demonstration of Jiangsu Province under grant number NJ2021-58 and Postgraduate Research & Practice Innovation Program of Jiangsu under grant number SJCX21-0913.

REFERENCES

[1] Wang, S.; Wang, H.; Yang, F.; Liu, F.; Zeng, L. Attention-based deep learning for chip-surface-defect detection. *Int.J.Adv.Manuf.Tech.* **2022**, *121*, 1957-1971.

[2] Kapil Joshi; Manoj Diwakar; Joshi, N.K.; Lamba, S. A Concise Review on Latest Methods of Image Fusion. *Recent advances in computer science and communications* **2021**, *14*, 2046-2056.

[3] Lapray, P.-J.; Heyrman, B.; Ginhac, D. HDR-ARtiSt: an adaptive real-time smart camera for high dynamic range imaging. *Journal of Real-Time Image Processing* **2016**, *12*, 747-762.

[4] Zhuang, P.; Ding, X. Correction to: Underwater image enhancement using an edge-preserving filtering Retinex algorithm. *Multimed Tools Appl 79* **2020**, 1729.

[5] Vijayalakshmi, D.; Nath, M.K.; Acharya, O.P. A Comprehensive Survey on Image Contrast Enhancement Techniques in Spatial Domain. *Sensing and Imaging* **2020**, *21*, 40.

[6] Celik, T. Spatial entropy-based global and local image contrast enhancement. *IEEE Trans Image Process* **2014**, *23*, 5298-5308.

[7] Enesi, I.; Çiço, B.; Harizaj, M. Increasing quality in fingerprint images by Implementing wavelet transform based fusion technique. *Association for Computing Machinery* **2021**, 112-117.

[8] Zhang Z; Xu Y; Yang J. A Survey of Sparse Representation: Algorithms and Applications. *IEEE Access*, **2017**, 3:490-530.

[9] Vivone G ; Alparone L ; Chanussot J. A Critical Comparison Among Pansharpening Algorithms. *IEEE Transactions on Geoence & Remote Sensing*, **2015**, 53(5), 2565-2586.

[10] Ma, Z.; Oh, C. A Wavelet-Based Dual-Stream Network for Underwater Image Enhancement. In Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **2022**; pp. 2769-2773.

[11] Fu Z ; Zhu H; Yu S. Multi exposure HDR Image Reconstruction Based on Gray Scale Mapping Function Modeling. *Journal of Data Acquisition and Processing*, **2019**.

[12] Wang, C.; He, C.; Xu, M. Fast exposure fusion of detail enhancement for brightest and darkest regions. *The Visual Computer* **2021**, *37*, 1233-1243.

[13] Ma K; Hui L; Yong H. Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach. *IEEE Transactions on Image Processing*, **2017**, 26(5), 2519-2532.

[14] Lyu, Q.; KinTak, U. A Multi-focus Image-fusion Scheme Based on non-Uniform Triangular Partition. *International Conference on Information Communication and Networks* **2019**, 191-195.

[15] P, R.T.; D, R. A Hybrid Method for Block-Based Feature Level Image FusionTechnique of PAN and MS Image Using Contourlet Transform with Neural Network. *Journal of Emerging Technologies and Innovative Research* **2019**.

[16] Hua K L; Wang H C; Yeh C H. Background Extraction Using Random Walk Image Fusion. *IEEE Trans Cybern*, 2018,1-13.

[17] Singh, S.; Mittal, N.; Singh, H. Review of Various Image Fusion Algorithms and Image Fusion Performance Metric. *Archives of Computational Methods in Engineering* **2021**, *28*, 3645-3659.

[18] Vanmali, A.V.; Gadre, V.M. Visible and NIR image fusion using weight-map-guided Laplacian–Gaussian pyramid for improving scene visibility. *Sādhanā* **2017**, *42*, 1063-1082.

[19] Qu, Z.; Huang, X.; Liu, L. An improved algorithm of multi-exposure image fusion by detail enhancement. *Multimedia Systems* **2021**, *27*, 33-44.

[20] Turgut S S ; Oral M. Multi-Focus Image Fusion based on Gradient Transform. *Computer Vision and Pattern Recognition*, **2022**.

[21] Kataoka K, Kameda Y, Hamamoto T. Adaptive Exposure-time Control Based on Image Entropy for Multiple-exposure-time Image Sensor. ITE Transactions on Media Technology and Applications, 2021, 9(2):128-135.

[22] Zhang W, Sun X, Yu Q. Fast Exposure Fusion Based on Camera Response Function. DEStech Transactions on Computer Science and Engineering, 2020.

[23] X. Deng; Y. Zhang; M. Xu; S. Gu and Y. Duan.Deep Coupled Feedback Network for Joint Exposure Fusion and Image Super-Resolution, *IEEE Transactions on Image Processing* **2021**, 30, 3098-3112.

Image Matting using Neural Networks

Nrupatunga J, Swarnalatha K S
Dept. of Information Science and Engineering
Nitte Meenakshi Institute of Technology, Karnataka, India-560064

Abstract*—Image matting, also refers to picture matting in the article, is the task of finding appealing targets in a picture or sequence of pictures i.e., video, and it has been used extensively in many photo and video editing applications. Image composition is the process of extracting an eye-catching subject from a photograph and blending it with a different background. a) Blue/Green screen (curtain) matting, where the backdrop is clear and readily distinct between the foreground (frontal area) and background (foundation) portions. This approach is now the most used type of image matting. b) Natural picture matting, in which these sorts of photos are taken naturally using cameras or cell phones during everyday activities. These are the present known techniques of picture matting. It is difficult to discern the distinction between the frontal area and the foundation at their boundaries. The current framework requires both the RGB and trimap images as inputs for natural picture matting. It is difficult to compute the trimap since additional framework is required to obtain this trimap. This study will introduce the Picture Matting Neural Net (PMNN) framework, which utilizes a single RGB image as an input and creates the alpha matte without any human involvement in between the framework and the user, to overcome the drawbacks of the prior frameworks. The created alpha matte is tested against the alpha matte from the PPM-100 data set, and the PSNR and SSIM measurement index are utilized to compare the two. The framework works well and can be fed with regular pictures taken with cameras or mobile phones without reducing the clarity of the image.

Keywords*—Picture matting; RGB picture; Blue/Green screen; foreground; background

I. Introduction

With the commonness of cell phones [24], image sensors have expanded strongly, and accordingly expanding the quantity of pictures [19] being captured through electronic devices. Hence, to handle this immense measure of visual information into helpful visual data, image processing has formed into a need in the ongoing moment capturing situation. Image processing has many applications and few of them are up-gradation of image, rebuilding of image, and getting relevant information from an image. Picture matting (or Alpha matting) is a sub-space of image processing that can remove the forefront from a picture, as an alpha matte. It has an assortment of utilization [2][3], like segmentation based on colour in an image [28], removal of reflectional lights, picture colorization, deblurring (bokeh effect), fashion e-commerce [14, 29, 30] and denoising to give some examples.

Numerically, picture matting issue can be demonstrated as:

$$I = \alpha F + (1-\alpha) B$$

where *I*, *F* and *B* signify input picture, frontal area of output and foundation picture separately. α ranges from 0 to 1 which denotes frontal area opacity. 0% opacity of frontal area and 100% opacity of foundation is obtained when $\alpha = 0$ and vice versa when $\alpha = 1$. For every one of the fragmentary estimations of α, these pixels lie in the blended or obscure locales.

In picture matting, most pixels have α esteems either 1 or 0, the essential issue is to gauge the exact α values for pixels in blended locales, obviously to isolate frontal area and foundation locales. Given an information picture *I*, we need to appraise *F*, *B* and *α* at the same time which is a poorly presented issue. To resolve this issue the most generally utilized technique is a pre-characterized trimap [26] as deduced above. Notwithstanding, for humans it takes more time to get the trimap manually like annotating the edges and will be less precision even if captured through a depth camera or else [17, 18] there arises for another framework for getting a trimap. As a result, a few recent studies [16] have attempted to eliminate model reliance upon this trimap by developing trimap-free techniques.

The proposed PMNN, is a lightweight neural net which disintegrates the representation matting process in three corresponding sub-undertakings and streamlines them at the same time through unambiguous constraints, in order to predict a precise alpha matte from just a single RGB picture. There are two bits of knowledge behind PMNN. In the first place, neural networks are better at learning a bunch of straightforward targets as opposed to a mind boggling one. In this manner, addressing the series matting of sub-targets can accomplish better execution. Secondly, by supervising every sub-target allows different components of the system to know decoupled knowledge, allowing the sub-targets being addressed in one framework.

Due to the removal of the trimap input, the tests suggest that PMNN is more robust in practical scenarios. The technique is attempting to determine whether or not a green screen is required for ongoing picture matting. In rundown, will introduce a neural net framework design called PMNN achieving progressively without trimap representation matting.

II. Literature Survey

There are various picture matting approaches [1] (Learning-based matting, propagation-based matting (alpha propagation-based matting), sampling-based matting) have been used to demonstrate the importance of exact alpha matte computation, pixel sample selection, and trimap generation. The interval line based picture matting approach is used to speed up the matte computation. Learning based matting solutions are becoming more popular among established procedures. To improve Matte precision, propagation based techniques and sampling based strategies can be combined, and learning based matting approaches can be used to modify the matte result.

In [4], the author provides a complete semantic matting in the absence of trimap as supplementary information by combining a data set of coarse annotations [20, 25] as information with fine annotation data. To create a prediction of a mask [21, 27] by a network using hybrid information to gauge the mask of coarse semantics, and then introduce a qualitative unification neural net that can bring together the nature of previous output of coarse mask. To estimate the final alpha matte, a matting refinement neural net takes in the combined mask and information picture. By this it expects to predict the opacity of the per-pixel of the frontal area of human locale which is very difficult and for the most part requires trimap and a lot of excellent annotations on the information. Annotation on such information is work concentrated and requires extraordinary abilities beyond the ordinary use, particularly taking into account the exceptionally point by point hair part of people [13].

By incorporating neural net into the process of learning an alpha matte principal propagation, [5] proposes a deep proliferation based picture matting structure. The deep component extraction module, the propagation of matte module, and the learning affinity [15] module are connected to create the deep learning engineering. By using the training process from end to end, these three components can be separated and streamlined in relation to one another. By learning deep picture depictions tuned to propagation matte, the structure creates a semantic-level sequence of comparability of pixels for proliferation. It consolidates the force of deep learning and matte proliferation. The complex of training was approved by the exploratory outcomes from 243K pictures made in light of two benchmark matting data sets. In order to understand [23] deep picture representation with an adaptation to propagation of alpha matte and create more appealing pairwise propagation compatibility, for the design of a DeepMattePropNet.

Past calculations have terrible showing when a picture has comparable frontal area and foundation tones or textures of complication. The main reasons are due to older methodologies. a) using only features of low-level, b) lack of context from high level. In this [6] study, presents a unique deep neural net-based approach that addresses both of these problems. There are two parts to the deep framework. The first section is a convolutional neural encoder-decoder model that predicts an alpha matte of a picture using a picture and the comparative trimap as information sources. The next section is a small convolutional neural net that enhances the prior network's alpha matte forecasts to see more precise alpha quality and finer borders. Moreover, they've also created a massive scope picture matting data - set, which includes training pictures of 49300 and 1000 test pictures. And evaluated the calculation using a picture matting benchmark, a trial set, and a variety of real-life images. The calculation clearly outperforms previous strategies in the trials.

By thoroughly examining numerous differences between the foreground and background images, Jizhizi Li et al.[31] identified the domain gap issue between composite images and real-world photos. They discover that a properly planned composition route RSSN that seeks to lessen the disparities can result in a superior model with impressive generalization ability. Additionally, they offer a benchmark that includes 10,000 portrait photographs with their manually labelled alpha mattes and 2,000 high-resolution real-world animal images, to be used as a test set for determining how well the matting model generalises to real-world images. To fully utilize the trimap information in the transformer block, GyuTae Park et al. [32] suggest a transformer-based image matting model called MatteFormer. The initial step in our procedure is the introduction of a prior-token, which is a global representation of each trimap region (e.g., foreground, background and unknown). As global priors, these prior-tokens take part in each block's self-attention process. PAST (Prior-Attentive Swin Transformer) blocks, which are used at each stage of the encoder and are based on Swin Transformer blocks but differ in a few ways.

III. DATA COLLECTION

It's interesting that the offer the Photographic Portrait Matting benchmark (PPM-100) includes 100 representative photographs with meticulous annotations and various foundations. In order to modify the example varieties in PPM-100 and assure test variety, need to characterise a few ordering rules. For illustration, (a) if the entire human body has been included; (b) is the picture foundation occluded; (c) if the person is carrying any other items. Since this is more in line with the practical uses, need to respect small items that people hold as part of the closer view. The examples in PPM-100 do have flamboyant postures and more regular foundations, as shown in Fig. 1. In this approach, the PPM-100 standard is a more thorough benchmark.

Fig. 1. The PPM-100 data set.

Fig. 1, has more variation in the frontal areas and unique picture foundations. To display experiments having thin hair [10], including more elements [11], and in the absence of bokeh or in the presence of full-body.

IV. METHODOLOGY

PMNN comprises three branches, which gain disparate sub-targets through unambiguous requirements. A low-goal branch for measuring human semantics is present in PMNN and is controlled by the base truth matte's thumbnail. As a result, a high-goal branch that is familiar with the center around the representation limits is controlled by the progress area (α ε (0, 1)) in the base truth matte. In order to predict the final alpha matte, a combination branch (controlled by the full base truth matte) is introduced toward the end of PMNN. So that it can examine the branches employed to address each sub-objective in the subsections that follow.

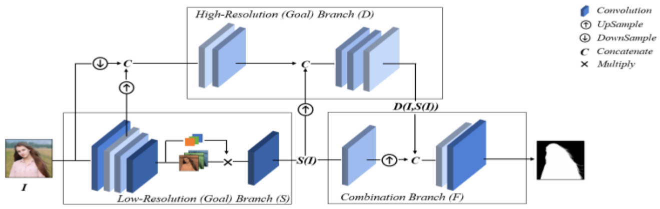


Fig. 2. Block diagram of PMNN.

Fig. 2 uses three associated branches, S, D, and C. PMNN predicts human semantics s_p, edge subtleties d_p, and last alpha matte *p* given an information image *I*. So that it is easy to simplify PMNN from beginning to end because the disintegrating sub-targets are correlated and help reinforce one another.

1) Semantic assessment: The first phase of PMNN is to locate the person information Picture *I*, much like other existing frame-work techniques. It is essential that to only use an encoder, or the low-goal branch *S* of PMNN, to extract the high-level meanings, it has two primary benefits. First, since the decoder is no longer used to perform semantic assessment, it will be more effective. Finally, resulting branches and joint streamlining benefit from the high level depiction *S(I)*. As a backbone for *S*, any CNN [9] [12] can be used. Framework chose the MobileNetV2 [7] technology and [8], a great model created for cell phones, as the *S* in order to operate with continuous collaboration.

S(I) is through into a convolution layers that is initiated by the sigmoid activation function to bring down the channel count to 1 in order to estimate the s_p coarse semantic mask. And to control s_p using a small portion of the fundamental truth matte α_g. So the block employ loss of *L2* here because s_p should be smooth,

like in:

$$L_s = \frac{1}{2}||s_p - G(\alpha_g)||_2$$

Where *G* represents Gaussian blur and also represents the 16x down sampling. It eliminates the fine designs (like hair) that are not fundamental for human semantics.

2) Detail assessment: A high-goal branch *D* that uses *I*, *S(I)*, and the low-level components of *S* as information sources surrounds the frontal area representation in the progress. Reusing the low-level spotlights is done in order to reduce the computational burden on *D*. Additionally, enhancing *D* in the three following perspectives: (a) Compared to *S*, *D* has less convolution layer; (b) The convolution layer in *D* are chosen with lower channels; (c) Through *D*, the block doesn't maintain the initial information resolution. In actuality, *D* has a max channel count of 64 and comprises 12 convolution layer. In the first level, the feature resolution of the map is down-scaled to 1/4 of *I*, and in the following two levels, it is recovered.

The *D* outcome results as *D(I, S(I))*, the intricacy between the sub target human semantic *S(I)* of high level is needed in prior for detail assessment. From *D(I, S(I))*, determine the maximum detail matte d_p and grip it along loss of *L1*, as follows:

$$L_d = m_d||d_p - \alpha_g||_1$$

where md serves as a binary mask for L_d to emphasise the representational boundary. Erosion and dilation on α_g result in the production of m_d. If the pixels are inside the transitional area, its attributes are 1, otherwise they are 0. In actuality, the pixels in the dim area of the trimap are those with $m_d = 1$. Despite the possibility of inaccurate quality in d_p for pixels with $m_d = 0$, it has great accuracy for pixels having $m_d = 1$.

3) Semantic-detail combination: A direct CNN module, the combination branch *C* in PMNN combines semantic and detail. Firstly increase the sample size, to align *S(I)* from with *D(I, S(I))*. At that point, *S(I)* and *D(I, S(I))* is added to estimate alpha matte α_p, which is anticipated by:

$$L_\alpha = ||\alpha_p - \alpha_g||_1 + L_c$$

where L_c is the loss of composition. The background picture ground truth, the frontal area ground truth and the outright difference between the information picture *I* and the composited picture obtained from α_p are all estimated.

Using the appropriate amounts of L_s, L_d, and L_α PMNN is ready from beginning to end.

$$L = \lambda_s L_s + \lambda_d L_d + \lambda_\alpha L_\alpha$$

where, adjusting the three loss: λ_s, λ_d, and λ_α are hyper parameters. The process of training the hyper parameter is robust. Which lay down $\lambda_s = \lambda_\alpha = 1$ and $\lambda_d = 10$.

V. RESULT

The below Fig. 3 represents the outcome from the framework and the Fig. 4 represents the sample comparison between the generated alpha matte and the dataset alpha matte.

The output of a model is tested using an image similarity measures of PSNR - peak signal-to-noise ratio using the formulas as shown,

$$MSE = \frac{\Sigma_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M * N}$$

MSE - Mean Square error. The information picture has *M* row and *N* column. The following formula is used to determine the PSNR:

$$PSNR = 10log_{10}(\frac{R^2}{MSE})$$

Fig. 3. Output of the framework consisting of RGB Image, extracted frontal area and alpha matte (From left to right).

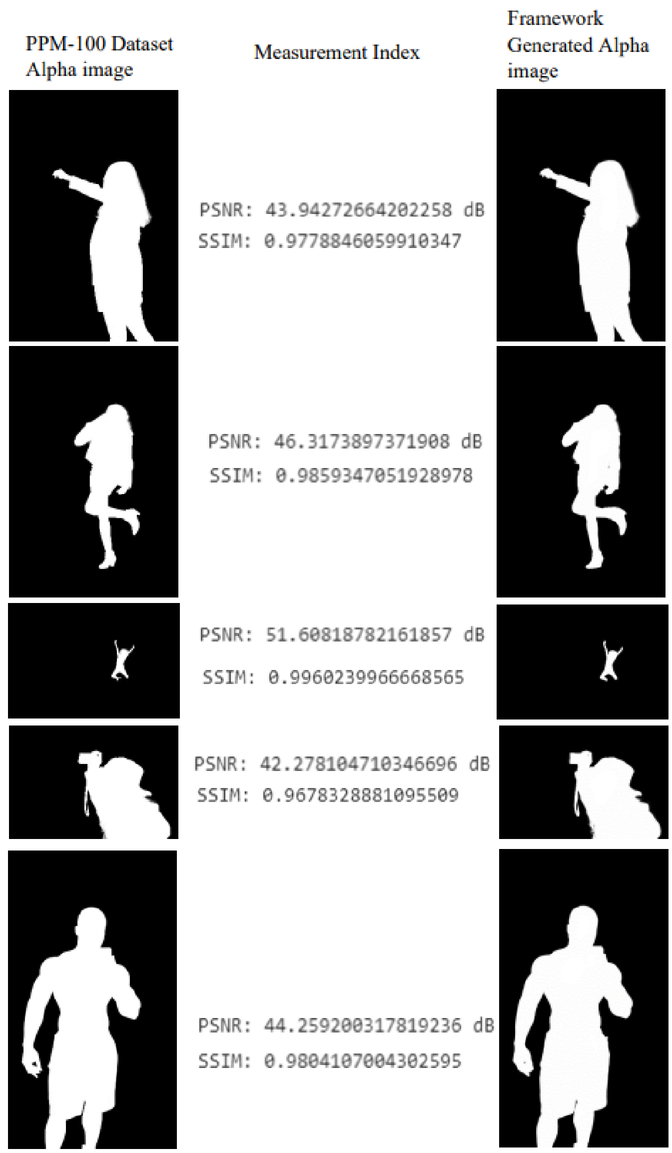


Fig. 4. The Comparison result between the generated and PPM alpha matte.

R is the highest pixel value that the image can contain. Notice that to get decibel value logarithm function is used.

SSIM - Structural Similar Index Measure, is based on structures that may be seen in the photograph. In other words, SSIM assesses the perceived variation between two analogous images. Between -1 and 1, the SSIM score denotes complete structural similarity, where 1 being perfect similarity. The distance *X-Y* across two windows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

In Fig. 4, the first column displays the alpha matte from the PPM - 100 data - set. The second third column picture displays the PMNN generated output of alpha matte. The PSNR and SSIM calculated values are displayed in the second column for the generated picture along with the standard data set of PPM - 100.

The framework not only performs well with the data set but also outperforms well along with the real picture captured with the cell phones or camera irrespective of the dimension or pixels or size of the picture.

VI. Conclusion

A quick, easy, and interesting PMNN is introduced as a result of the framework's aim to avoid the use of a green screen [22]. By relying just on RGB photos as data, the technique permits the forecasting of alpha mattes under changing landscape. The framework operates effectively with the PPM-100 data set, as stated in the result section, and also with captured images, regardless of the file size and dimension of the captured image. The frontal region may now be extracted from the image more easily and without the need for human interaction thanks to the model editing.

PMNN is exhibited to have excellent exhibitions on the meticulously sketched PPM-100 benchmark and various here and now facts. The model's future application must include even more accurate and exact extraction of fine hair and fine details in the image with a move toward zero error. The one significant feature that may be introduced is the user's choice of extraction in the image, similar to how the green/blue screen offers the option for the fine extraction of required frontal area from the device captures. This extraction could be an object in the image, a person with some objects, or a person alone.

References

[1] Qingsong Zhu, Pheng Ann Heng, Ling Shao and Xuelong Li, “What’s the Role of Image Matting in Image Segmentation?”, IEEE International Conference on Robotics and Biomimetics (ROBIO) Shenzhen, 2013.

[2] Jagruti Boda and Dhatri Pandya, “A Survey on Image Matting Techniques”, International Conference on Communication and Signal Processing (ICCSP), 2018.pp vol.

[3] Anil singh Parihar, “A Study on Image Matting Techniques”, 5th IEEE International Conference on Recent Advances and Innovations in Engineering- ICRAIE 2020.

[4] J. Liu et al., "Boosting Semantic Human Matting With Coarse Annotations," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8560-8569, doi: 10.1109/CVPR42600.2020.00859.

[5] Yu Wang, Yi Niu, Peiyong Duan1, Jianwei Lin, Yuanjie Zheng, “Deep Propagation Based Image Matting”, International Joint Conferences on Artificial Intelligence Organization, Twenty-Seventh International Joint Conference on Artificial Intelligence, Pages 999-1006.

[6] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang, “Deep Image Matting”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

[7] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks”, CVPR, 2018.

[8] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.

[9] https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215.

[10] Chang Liu, Henghui Ding and Xudong Jiang, “Towards Enhancing Fine-grained Details for Image Matting”, IEEE Winter Conference on Applications of Computer Vision (WACV), 2021.

[11] Lei Liu and Yingyun Yang, “Image Matting Based On Deep Learning And Image Dissection”, IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) 2019.

[12] Donggeun Yoon, Jinsun Park, and Donghyeon Cho, "Lightweight Deep CNN for Natural Image Matting via Similarity-Preserving Knowledge Distillation", IEEE SIGNAL PROCESSING LETTERS, VOL. 27, Page(s): 2139 - 2143, 2020.

[13] Xuqian Ren, Yifan Liu and Chunlei Song, "A Generative Adversarial Framework For Optimizing Image Matting And Harmonization Simultaneously", IEEE International Conference on Image Processing (ICIP), 2021.

[14] Rishab Sharma, Rahul Deora, and Anirudha Vishvakarma, "AlphaNet: An Attention Guided Deep Network for Automatic Image Matting", International Conference on Omni-layer Intelligent Systems (COINS) 2020.

[15] Guilin Yao, Dongai Jiang, and Jianliang Sun, "An Affinity Based Matting Method Based on Multi-Scale Space Fusion", 33rd Chinese Control and Decision Conference (CCDC) 2021.

[16] Yu Qiao, Yuhao Liu, Xin Yang, Dongsheng Zhou2 and et.al., "Attention-Guided Hierarchical Structure Aggregation for Image Matting", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020.

[17] Bo Liu, Haipeng Jing, Guangzhi Qu and Hans W. Guesgen, "Cascaded Segmented Matting Network for Human Matting", IEEE Access Volume: 9, Page(s): 157182 – 157191, 2021.

[18] YanLong Xu, Hui Fan∗, Jinjiang Li, "DenseNet Matting Algorithm Based on Embedded Improved SKNet", IEEE International Conference on Progress in Informatics and Computing (PIC) 2020.

[19] Haichao Yu, Ning Xu and et.al. "High-Resolution Deep Image Matting", The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), 2021.

[20] Yaoyi Li, Jianfu Zhang, Weijie Zhao, Weihao Jiang and Hongtao Lu, "Inductive Guided Filter: Real-Time Deep Matting with Weakly Annotated Masks on Mobile Devices", IEEE International Conference on Multimedia and Expo (ICME) 2020.

[21] Qihang Yu, Jianming Zhang and et.al., "Mask Guided Matting via Progressive Refinement Network", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[22] Soumyadip Sengupta, Vivek Jayaram, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman, "Background matting: The world is your green screen", CVPR, 2020.

[23] Xiaoyong Shen, Xin Tao, Hongyun Gao, Chao Zhou, and Jiaya Jia, "Deep automatic portrait matting", ECCV, 2016.

[24] R. Dong, B. Wang, Z. Zhou, S. Li and Z. Wang, "Design and Implementation of an Image Matting System on Android Phones," 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2015, pp. 522-526, doi: 10.1109/IHMSC.2015.53.

[25] X. Fang, S. -H. Zhang, T. Chen, X. Wu, A. Shamir and S. -M. Hu, "User-Guided Deep Human Image Matting Using Arbitrary Trimaps," in IEEE Transactions on Image Processing, vol. 31, pp. 2040-2052, 2022, doi: 10.1109/TIP.2022.3150295.

[26] V. Gupta and S. Raman, "Automatic trimap generation for image matting," 2016 International Conference on Signal and Information Processing (IConSIP), 2016, pp. 1-5, doi: 10.1109/ICONSIP.2016.7857477.

[27] C. Orrite, M. A. Varona, E. Estopiñán and J. R. Beltrán, "Portrait Segmentation by Deep Refinement of Image Matting," 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 1495-1499, doi: 10.1109/ICIP.2019.8799367.

[28] B. Qian and X. Gu, "Automatic ID Photos Matting Based on Improved CNN," 2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), 2019, pp. 96-99, doi: 10.1109/DCABES48411.2019.00031.

[29] D. Shin and Y. Chen, "Deep Garment Image Matting for a Virtual Try-on System," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 3141-3144, doi: 10.1109/ICCVW.2019.00384.

[30] B. Yuan, Z. Lu, J. -H. Xue and Q. Liao, "A New Approach to Automatic Clothing Matting from Mannequins," 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019, pp. 880-885, doi: 10.1109/ICME.2019.00156.

[31] Li, J., Zhang, J., Maybank, S.J. et al. Bridging Composite and Real: Towards End-to-End Deep Image Matting. Int J Comput Vis 130, 246–266 (2022). https://doi.org/10.1007/s11263-021-01541-0.

[32] GyuTae Park, SungJoon Son, JaeYoung Yoo, SeHo Kim, Nojun Kwak; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 11696-11706.

New Text Steganography Technique based on Multilayer Encoding with Format-Preserving Encryption and Huffman Coding

Mohammed Abdul Majeed[1*], Rossilawati Sulaiman[2], Zarina Shukur[3]
Center for Cyber Security-Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia (UKM), 43600 UKM, Bangi, Selangor, Malaysia

***Abstract*—Steganography is the process of hiding secret data inside other media or cover media. Balancing the requirements for capacity, security, and imperceptibility is the main challenge for any successful steganography system. In text steganography, the data hiding capacity is limited because of the lack of redundant data compared to other digital media, such as images, video, or audio. Other challenges in text steganography are imperceptibility and security. Poor imperceptibility results from the structure of the text file, which is more visually apparent in terms of syntax and grammar than in other media. Low level of security results from the sequential selection of positions for embedding secret data due to insufficient redundant data in a text file. Therefore, an attacker or a third party would notice slight changes in the text file. This paper proposes a new text steganography method that combines cryptography and compression techniques to deal with these issues. This technique is used to conceal secret data to achieve high data hiding capacity in the cover text while maintaining security and imperceptibility. Multilayer encoding and Format-Preserving Encryption (FPE) with Huffman Coding, are applied to secret data before embedding. Invisible Unicode characters are employed to embed secret data into English text files to generate stego files. Results show that the proposed method satisfies capacity and imperceptibility in the cover file by comparing it with previously developed methods.**

***Keywords*—Text steganography; format-preserving encryption; Huffman coding; unicode characters**

I. Introduction

Modern advancements in digital communication are essential to our everyday lives. The utilization of data transfer has substantially expanded because of developments in web-based technologies and the digitization of information. The data transfer includes audio, video, text, and images among individuals and groups, which has become very convenient [1]. However, the allocation of such massive amounts of data over the Internet makes them vulnerable to attacks. Thus, protecting sensitive data has become an important issue requiring immediate solutions. In general, two techniques are used to preserve the security and privacy of sensitive data: cryptography and steganography. One of the most attractive fields for data security is cryptography. This method uses several data encryption techniques to convert sensitive data into ciphertext, which is an incomprehensible format. Another technique for protecting communications during data communication is steganography. Although they have the same goal, steganography and cryptography use different techniques. Steganography, as compared to cryptography, keeps the original data by concealing it in various medium [2] [3].

Both Greek terms "Stegano" and "Graphy," which make up the name "Steganography," have to do with "Cover Writing". The practice of steganography began centuries ago. For instance, Histiaeus employed steganography to transmit messages by inking (tattooing) on the head of his slave, who would travel after the tattoo had grown enough hair to conceal it. Greeks were famous for transmitting secret messages [4] [5]. Since the text has fewer redundant bits than in other cover media like images, music, and video, inserting hidden data in the cover text is the main challenge in text steganography. Due to the scarcity of redundant bits, any little adjustments made to the cover text will be noticeable. Any steganography system must have three main requirements: capacity, security, and imperceptibility [6] [7]. Steganography highly values the imperceptibility of hiding sensitive data in other media. The hiding process is performed without being noticed by human eyes. The concept of "security" refers to "undetectability", where the concealed data is incapable of being found by statistical methods [8]. A steganography system typically seeks to communicate a significant amount of confidential information using the least-covered media to reduce the risk of being discovered when communicating over an unsecured connection [9].

Let's say there is a need to embed a significant amount of sensitive data in a cover file (which will later be called a stego file). In that instance, altering the cover file is more challenging due to the difficulty in achieving imperceptibility and the possibility of distortion. Therefore, the trade-off between hiding capacity with imperceptibility and hiding capacity with security must be identified [10]. The hiding capacity has an inverse relationship with imperceptibility, which means that when large secret data are hidden within a specific size of cover text, inevitably, the stego file will be distorted. This distortion attracts the intruder's attention, and thus the hidden data is noticeable. In general, Unicode characters are used to embed secret data, which requires modification of the cover file. However, because text media suffer from insufficient locations for hiding secret messages, more text will be needed for embedding. In addition, the selection process of the embedding positions of the secret data is performed sequentially. Sequential patterns of the positions make the algorithm vulnerable to attacks [2]. Moreover, text media is

*Corresponding Author.

naturally bounded in terms of syntax and grammar, making text media more visually apparent in the embedding process.

Several researchers have worked on the relationship between hiding capacity and imperceptibility. Compression techniques are used to reduce the hidden secret data size, simultaneously minimizing the amount of modification in the cover file and increasing the imperceptibility [11], such as work in [12], which compressed the secret message using Huffman coding. Secret messages are also compressed in research found in [13] that combined algorithm with minimum-maximum weight and Huffman coding. Meanwhile, in [14], the secret data are concealed in the forward email IDs platform after compressing it with Huffman coding.

Despite implementing the compression algorithms, there are limitations, especially regarding the compression keys. In steganography, the compressed data will be hidden with the compression information, including the decompression keys, which will be shared separately between authorized participants. However, any third party getting the decompression keys could also obtain the secret data. In other cases, compression also led to low imperceptibility, as found in [15].

In addition, research on hiding capacity and security is being done to provide data protection while maintaining hiding capacity. Work in [16] used the RSA algorithm to provide security by encrypting data with minimum modifications in the presence or characters layout. Another research by [17] employs the Data Encryption Standard (DES), a symmetric encryption key, and combines steganography and cryptography for secure data transfer. However, despite these implemented approaches that improved security, cryptography algorithms also increased the secret data size since encryption generally increases the overall size of the data.

Therefore, any new text steganography technique must consider these issues and propose a method that minimizes the secret message size compatible with the available capacity in the cover text. In addition, changes in the cover text should be minimum since embedding secret data in bulk is more noticeable than in other media. More importantly, the embedding technique used to add layers of security must be in such a way that it does not increase the secret message size.

These issues give the motivation to develop the proposed method described further in Section III, which considers the relationship between hiding capacity and imperceptibility on one hand and hiding capacity and security on the other through the combination of cryptography and compression algorithms. This combination has two objectives. Firstly, is to adjust all steganography requirements simultaneously. Secondly is to preserve the trade-off between high hiding capacity and maintaining the imperceptibility of the stego file.

This paper suggests a new text steganography technique using multilayer encoding with FPE and Huffman Coding, which is applied to secret data that will be embedded using invisible Unicode characters inside the cover text. In addition, this paper explains the nature of the text media and the relationship among factors that must be considered when proposing new techniques for text steganography.

The rest of the paper is organized as follows: Section II presents related works. Concepts in Multilayer encoding with format-preserving encryption (fpe) and Huffman coding illustrated in Section III. Proposed method is illustrated in Section IV. Then, Section V presents the experimental results, followed by the conclusion of the work in Section VI. Lastly, recommendations for future research are presented in Section VII.

II. Related Work

This section analyses related works that share the same fundamental concept as the proposed method. Randomization is a concept that can be used as opposed to the sequential concept. This concept can be seen used in text steganography, as found in [18], which used randomization in selecting the forward emails as a cover text. A randomized index-based word dictionary is used to encrypt the carbon copy field that contains hidden data. A temporary stego key from the public key cryptography generates a system time-based random bitstream and is transmitted separately. This method is considered secure against attacks because noises are excluded from the actual email body content. Moreover, randomizing each word of the index values adds an extra layer of security by inserting an 18-bit key generated using the system time in the "date" column of the forward email format. However, the use of public key cryptography increased secret message size and affected negatively on data hiding capacity.

The author in [19] suggested a set of two letters words based on the Oxford dictionary as indicators to hide the secret data represented through non-printing Unicode characters (UC). The proposed approach maps secret data so that every two bits of secret text with a specific UC generates a UC mapping table that will be shared between the sender and receiver. However, the inserting of secret message after each two letters words leads to increase the changes in a cover text and that minimizes a level of imperceptibility. This work is improved in [20] with the Lempel-Ziv-Welch compression algorithm to minimize the secret data size. The secret data are then represented through non-printing UCs to generate the UC mapping table that will be shared between both sender and receiver. However, the capacity is still low due to mapping representation that used a few bits with non-printing UCs in UC mapping table.

Color coding and LZW compression technique are used by [21] that employed the forward mail as a cover text for the secret data embedding process. The secret data is then compressed by the LZW compression technique and embedded inside the cover text by coloring it based in a color-coding table. However, the capacity is still low due to limitation in mapping representation in color-coding table. Another work that used the Huffman compression can be found in [22] to minimize the secret data. A specific number of characters in an email ID indicates the bits of hidden messages. However, the capacity is still low due to mapping representation that used one bit only in color-coding table.

III. Concepts in Multilayer Encoding with Format-Preserving Encryption (FPE) and Huffman Coding

The rapid increase in the number of covert activities in communication networks has intensified the need to devise an efficient data-hiding method to protect secret information from malicious attacks. One possible solution to this problem is to combine steganography techniques with encryption and compression techniques. As previously mentioned in Section II, several encryptions and compression techniques have been proposed, each with unique advantages and disadvantages. An encryption and compression technique that can provide a high security and compression ratio while maintaining an acceptable imperceptibility for the output file must be adopted. To this end, the proposed model applies multilayer encoding with (FPE) and Huffman Coding, which seeks to encode by encrypting and compressing the secret message before embedding.

This model consists of encryption and compression techniques (as shown in Fig. 1), which are applied to encrypt and compress the secret message and consequently increase the security ratio and hiding capacity.

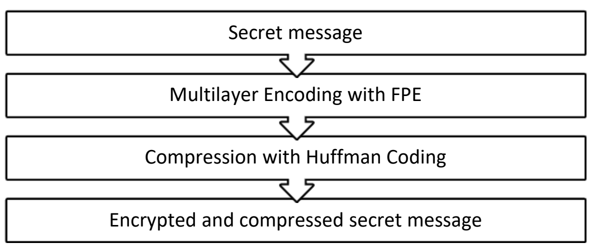


Fig. 1. Multilayer encoding with FPE and Huffman coding process

A. Format-Preserving Encryption (FPE)

A novel symmetric encryption method known as FPE is gaining popularity as a solution to previously mentioned problems. This method differs slightly from traditional encryption protocols like AES and DES [23] [24]. It is a rapidly growing cryptography tool that serves the purpose of secrecy in cryptography by ensuring data security. As the name suggests, format-preserving encryption aims to encrypt data without changing its size or format. It involves encrypting data so that the output matches the input's size and format, which offers several advantages over traditional encryption. Feistel structure-based schemes called FF1 and FF3 are used to implement FPE algorithms. The National Institute of Standards and Technology (NIST) advises only two operating modes: FF1 and FF3. A basic block cipher component called BPS-BC is proposed to be used in Cipher Block Chaining (CBC) mode to encrypt messages of any length. FF1, known initially as format-preserving Feistel-based encryption (FFX), was proposed by [25], and FF3, corresponding to the BPS-BC (Brier Peyrin Stern), was proposed by [26]. Fig. 2's non-binary Feistel structure is the foundation for both operating modes. Both FF1 and FF3 can use electronic code book (ECB) mode to encrypt data blocks.

Despite using AES as the underlying block cipher, operating modes may be considered a type of FPE block ciphers. The encryption of data with changeable forms, such as Primary Account Numbers (PANs) or Social Security Numbers (SSNs) that are not in binary format, is an example of a practical use of FPE [26][27]. FPE can also be used in communication systems when it's crucial to encrypt specific protocols, such as in industry or the military or when encrypting particular image formats [28].

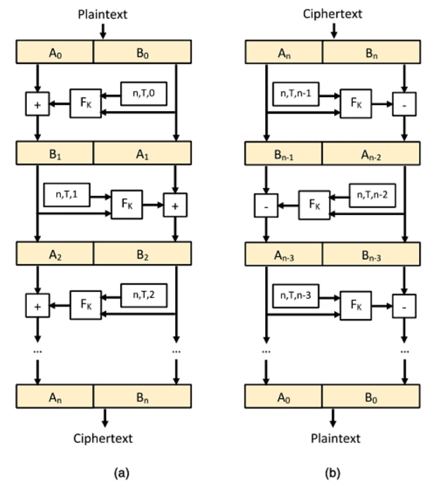


Fig. 2. Generic cipher structure for format preserving encryption: (a) ciphering, (b) deciphering

The present study focuses on the concept of using format-preserving encryption FF1 mode. Data encryption must be considered when developing a conceptual framework because the technique of data encryption leads to an increase in the level of security with maintaining the capacity ratio of stego text files.

B. Compression using Huffman Coding

This section examines the data compression techniques that benefit from current compression techniques related to text files. These techniques can also be combined with steganography to improve the proposed solution. Data compression refers to the process of encoding input data by using a few bits representing the input's original size [29]. In other words, data compression denotes information in a compact format. The data's structure must first be determined to construct such small representations. These data may refer to characters in a text file generated by other processes. Compressed data are only used when both communicating parties are informed of the coding scheme, similar to any other type of communication. Compression is important since it lowers the use of costly resources.

Data compression techniques are highly recommended to increase the hiding capacity in cover files. These techniques, which are popular in computing, require the data transmitted over the Internet to be as compact as possible. Named after the late David Huffman in the 1950s, Huffman coding is a data compression technique that utilizes the greedy algorithm and achieves remarkable savings in capacity (ranging from 20% to 90%).

A variable-length code is employed in Huffman compression coding. Utilizing lower-frequency characters with more extended codes is also preferable in this case. The encoding process involves assigning a numerical string based on the frequency of characters. The Huffman code algorithm takes a string of symbols and transforms them into a varying-length binary string. Then, a binary tree is generated for decoding the binary strings. Each binary character is assigned with 1 for the right child and 0 for the remaining child routes.

IV. Proposed Method

This section describes the proposed method. There are four main components: the encoding phase, embedding phase, extracting phase, and decoding phase, as shown in Fig. 3.

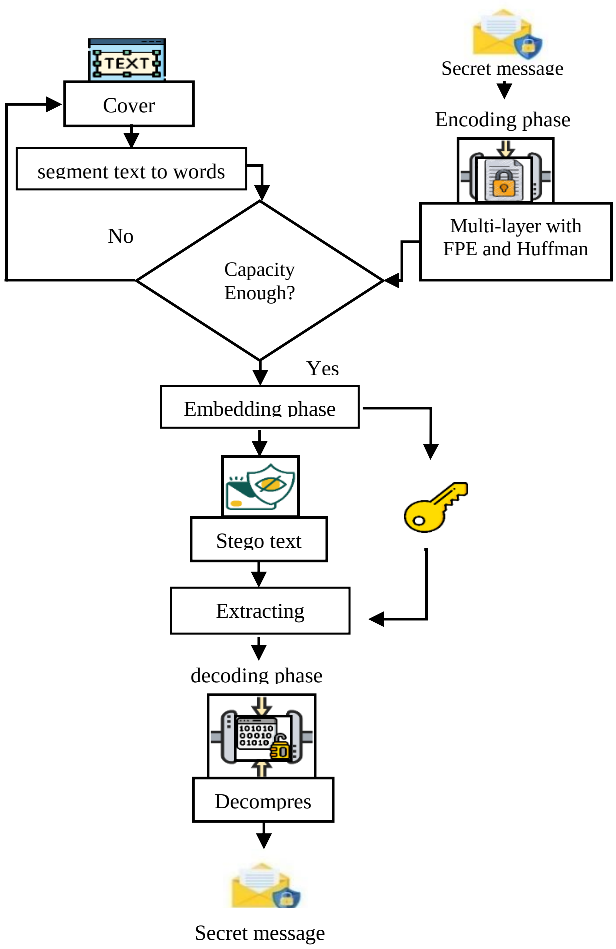


Fig. 3. Block diagram of the proposed scheme

A. Encoding Phase

In this phase, a combination of compression and security is achieved by implementing the multilayer encoding with FPE and Huffman Coding for compression, where secret messages are encoded using FPE. This stage is called multilayer encoding because it involves a few layers of encoding, such as from text to binary, then binary to decimal, and finally applying mod 26 to obtain the final encoding. FPE and the multiple encodings provide multilayer security to the secret message before its actual embedding. Multilayer encoding also reduces the size of hidden messages while providing more hiding spaces, as shown in Algorithm 1 of the multilayer encoding with FPE (Fig. 4). Next, Fig. 5 depicts the algorithm for data compression using Huffman Coding. This algorithm starts by computing the frequency of occurrence of each data (output code from FPE), which is calculated in the input stream. These codes are then arranged from the highest frequency to the lowest.

Algorithm 1: multilayer encoding with FPE
Input: Secret message ($\mathbf{M_S}$), FPE Key (Kp)
Output: $\mathbf{EB_i}$, $\mathbf{k_S}$

Steps:
1. Select Kp
2. For each Char in secret message (M_S)
   Convert M_S into Binary.
   End For
3. Convert the binary value of M_S into decimal (M_d)
4. Divide M_d with x , $Bi = \frac{M_d}{x}$ // x is random length of the divisor.
5. Encrypt B_i with Kp using FPE.
   $EB_i = Eph(B_i,(Kp) \bmod n$
6. Take mod 26 of the Encrypted Block
   (B_i) $EB_i = B_i \bmod 26$.
7. Store the reminder bits of mode 26 as the second key, k_S for each block.

Fig. 4. Algorithm 1: multi-layer encoding with FPE

Algorithm 2: Huffman message compression
Input: Output codes from the multilayer encoding with FPE
Output: Compressed secret message (CM)

Steps:
1. Read the output codes
2. Compute the frequency of occurrence of each output code in a list of (output code, frequency).
3. Arrange the list from the highest to the lowest frequency.
4. Calculate summation of the last two frequency numbers.
5. Rearrange the values in descending order based on their frequency numbers.
6. If there is more than one element in the list
   Repeat Step 3.
   End-If
7. Construct a Huffman tree by assigning the value (0,1) to each pair of branches in the tree.
8. Construct the final table (Huffman coding) that contains the leaf nodes (output codes) and their respective codes according to the Huffman tree.
9. Generate the compressed secret message (CM) by rewriting the output codes using the table in Step 7.

Fig. 5. Algorithm 2: Compression with Huffman coding

In Algorithm 1 (Fig. 4), in line 1, the algorithm selects the value of the FPE key (Kp). The block encoding and encryption processes are shown in lines 2–8. Firstly, each character of the secret message (Ms) is converted into binary. Next, in line 4, a binary string is converted into a decimal string Md, and Md is divided into equal blocks with a random length of the divisor (x), shown in line 5. After that, the encryption phase (labelled as *Eph*) is applied on each block (Bi) with Key (Kp), $EB_i = Eph(B_i, k_p) \bmod 26$ (line 6). In line 7, mod 26 on EB_i is calculated, and finally, the k_S values are determined by collecting the reminder bits of mode 26 for each block (line 8).

The k_S values are used to regenerate encrypted blocks in the extraction process.

In Algorithm 2 (Fig. 5), Huffman coding begins with the input, representing the output codes of the multilayer encoding with the FPE algorithm that need to be compressed. The algorithm then computes the frequency of occurrence of each output code in the input stream. These codes are subsequently arranged from the highest to the lowest frequency. The two codes with the lowest frequencies are treated as children of the node, and the parent node comprises the total frequency of these two child nodes. This node is then inserted back into the list, and the list is rearranged.

The process of applying Multilayer Encoding with FPE and Huffman coding is shown in the following example:

Suppose the Secret message is: "**Universiti Kebangsaan Malaysia**".

1) Multilayer encoding with FPE:

Step 1: Select the encryption key "Kp" value for FPE. Assume Kp=1200.

Step 2: Convert the secret message into random binary stream blocks. The obtained binary blocks are given below:

01010101 01101110 01101001 01110110 01100101 01110010 01110011 01101001 01110100 01101001 00100000 01001011 01100101 01100010 01100001 01101110 01100111 01110011 01100001 01100001 01101110 00100000 01001101 01100001 01101100 01100001 01111001 01110011 01101001 01100001

Step 3: Convert each binary block into a decimal stream to obtain the following decimal stream:

85110105118101114115105116105327510198971101031 1597971103277971089712111510597

Step 4: Convert the decimal stream into random blocks:

8511010511 8101114115 1051161053 2751019897 1101031159 7971103277 9710897121 11510597

Step 5: Encrypt each decimal block using FPE with Kp=1200; the resultant encrypted blocks are given like the following:

1909100954 6927975788 1398002705 6714036020 1293204947 2415200591 2308156747 39114904

Step 6: Apply mod 26 on each encrypted block, and the remaining bits of each block after applying the mod 26 equation are stored as the second key, Ks.

The results of applying mod 26 on each encrypted block are as follows:

[**20 6 21 16 21 11 13 10**]

2) Applying huffman coding:

Step1: Read the output codes of multilayer encoding with FPE, which are:

[**20, 6, 21, 16, 21, 11, 13, 10**].

Step 2: Construct the final table (Huffman coding table) that contains the leaf nodes (output codes), as shown in Table I.

TABLE I. HUFFMAN CODING TABLE

Node	20	6	21	16	21	11	13	10
Codeword	110	1000	11	101	00	010	011	1001

Step 3: Generate the compressed secret message by rewriting the output codes using the table of Huffman coding.

Original secret message: (20, 6, 21, 16, 21, 11, 13, 10)

Compressed secret message: (110, 1000, 11, 101, 00, 010, 011, 1001).

The length of the compressed secret message, as computed using the Huffman algorithm, is 24 bits.

B. Embedding Phase

By using the invisible Unicode characters (UC), the embedding process aims to resolve the cover text limitation in data hiding capacity. In the proposed method, the embedding process is performed using eight Unicode characters to embed the secret data. These characters are inserted into spaces between words of the cover text. This phase proposes a text steganography technique using the property of data redundancy in the English text to improve the imperceptibility of hidden information. This method uses eight invisible Unicode characters UC by mapping three bits in each UC, as shown in Table II.

TABLE II. UC MAP FOR HIDING 3 BITS IN EACH CHARACTER

Unicode characters	**Abbreviation**	**Representation**
Zero width character	ZWC	000
Zero width joiner	ZWJ	001
Zero width no- joiner	ZWNJ	010
Invisible plus	IP	011
Invisible separator	IS	100
Inhabit Symmetric Swapping	ISS	101
Empty string	‘’’’	110
Left-To-Right Embedding	LRE	111

The algorithm uses UC to hide three secret message bits after each word in the cover text. The first step of Algorithm 3 is to read the cover text and start segmenting text into words. Then, secret messages are converted to binary, which is divided into blocks of three bits. Next, every three bits are checked according to UC mapping, as defined in Table II. Then the alternative Unicode character is inserted after each word of cover text. Fig. 6 shows the Embedding algorithm.

Algorithm 3: Embedding process
Input: cover text, secret message, UC
Output: Stego-text file

Steps:

1. Read the secret message.
2. Read the cover text file.
3. Segment cover text to words
4. Divide the hidden data into blocks of 3 bits each.
5. For each block in a secret message
   // there are 8 block options (i.e 000,001,010,100,101,110 and 111) are available to insert UC after each word.
   Check the state of the first 3 bits of block
   5.1 IF 3 bits of block = "000" Insert ZWC after it
   Else Read a new word, Repeat Step 5.1.
   5.2 IF 3 bits of block = "001" Insert ZWJ after it
   Else Read a new word, Repeat Step 5.2.
   End IF.
   5.3 IF 3 bits of block = "010" Insert ZWNJ after it
   Else Read a new word, Repeat Step 5.3.
   End IF.
   5.4 IF 3 bits of block = "011" Insert IP after it
   Else Read a new word, Repeat Step 5.4.
   End IF.
   5.5 IF 3 bits of block = "100" Insert IS after it
   Else Read a new word, Repeat Step 5.5.
   End IF.
   5.6 IF 3 bits of block = "101" Insert ISS after it
   Else Read a new word, Repeat Step 5.6.
   End IF.
   5.7 IF 3 bits of block = "110" Insert '‘’’ after it
   Else Read a new word, Repeat Step 5.7.
   End IF.
   5.8 IF 3 bits of block = "111" Insert LRE after it
   Else Read a new word, Repeat Step 5.8.
   End IF.
   End for
6. Return Stego-text file

Fig. 6. Algorithm 3: Data embedding

C. Extraction Phase

The extraction phase is the third phase implemented in the proposed method. The stego text file is used as input. The extraction phase aims to extract the hidden data from the stego text file by retrieving each invisible UC. The extraction starts by reading the stego text. Next, the alternative Unicode character after each word is extracted. Then, each UC is mapped as defined in Table II to show the three bits of hidden data. Finally, return the recovered encoded secret message bits for the Decoding phase to reconstruct the secret message. Fig. 7 summarizes the extraction process.

Algorithm 4: Extraction process
Input: Stego-text file
Output: Recovered encoded Secret message

Steps:

1. Read the stego text file.
2. Set secret data is null
3. Segment stego text to words.
4. Read each word and extract the alternative UC.
5. For each UC
   5.1 IF UC = ZWC add to hidden data = "000"
   Else Read a new UC, Repeat Step 5.1
   End IF.
   5.2 IF UC = ZWJ add to hidden data = "001"
   Else Read a new UC, Repeat Step 5.2.
   End IF.
   5.3 IF UC = ZWNJ add to hidden data ="010"
   Else Read a new UC, Repeat Step 5.3.
   End IF.
   5.4 IF UC= IP add to hidden data = "011"
   Else Read a new UC, Repeat Step 5.4.
   End IF..
   5.5 IF UC= IS add to hidden data ="100"
   Else Read a new UC, Repeat Step 5.5.
   End IF.
   5.6 IF UC = ISS add to hidden data ="101"
   Else Read a new UC, Repeat Step 5.6.
   End IF.
   5.7 IF UC = '‘’’ add to hidden data ="110"
   Else Read a new UC, Repeat Step 5.7.
   End IF.
   5.8 IF USC= LRE add to hidden data ="111"
   Else Read a new UC, Repeat Step 5.8.
   End IF.
   End for
6. Return secret message.

Fig. 7. Algorithm 4: Data extraction

D. Decoding Phase

The decoding phase includes two processes, namely: the decompression and deciphering of the secret text. The encoded secret message retrieved in the extraction phase is used as the input, while the retrieved secret message text is produced as the output.

1) Decoding of Multilayer Encoding with Format-Preserving Encryption and Huffman code: An encoded secret message retrieved from the extracting phase is decoded using the Huffman coding table to return the indexes. The Huffman coding table is transmitted as a key file. Decoding the retrieved encoded secret message and returning the indexes require rebuilding the Huffman tree based on the Huffman coding table. This process iterates through the binary encoded data.

The process starts traversing from the root until a leaf is found to find the characters that correspond to the current bits. The node on the left of the tree is then approached if the current bit is 0. If the bit is 1, the approach is made to the right node of the tree. When traversing, a leaf node is reached, and its character is displayed. After that, the encoded data is iterated till the end. The significant advantage of Huffman coding is that, although each character is coded with various bits, the receiver will automatically locate the character in order.

In FPE, the encoded data stream is decoded following the same principles applied in the encoding process. The process starts by obtaining the secret message bits from the decompression process with the Huffman code and then applying the decryption process with Ks, which is the key established after applying mode 26 in the multilayer process in the encoding stage. The next step is decrypting the secret message bits from FPE using the Kp key, converting secret bits from decimal to binary to rebuild the characters using ASCII code at the end of this algorithm (Fig. 8).

Algorithm 5: Decoding Multilayer Encoding with Format Preserving Encryption
Input: Encoding data stream , k_S , k_p
Output: Recovered Secret message

Steps:

```
1. Read Encoding data stream.
2. For each number in stream i.e. (20,6,21,16, 21,11,13,10).
        get ND using k_S.
        get NS using , k_p.
   End for
3. For each number in ND stream
        add number in Temp T
        IF T ≥ 64 then
        Add to DI list (T)
        T=0
        Else
        IF T =32 then
        Add to DI (T)
        T=0
        Endif
   End for
4. For each number in DI list
        convert to binary
        retrieve secret message SM
   End for
5. Return secret message.
```

Fig. 8. Algorithm 5: Data decoding

For example:

The retrieved message: (**20,6,21,16, 21,11,13,10**)

Step 1: decrypt (**20,6,21,16, 21,11,13,10**) using $\boldsymbol{k_S}$ which is mod 26 *reminder bits* of each number in encode stage.

The results of decryption will be:

1909100954	6927975788	1398002705
6714036020	1293204947	2415200591
2308156747	39114904	

Step 2: decrypt each block using FPE with Kp; the resultant decrypted blocks will be:

8511010511	8101114115	1051161053
2751019897	1101031159	7971103277
9710897121	11510597	

Step 3: combine blocks of numbers to accomplish the stream of decimal numbers. The stream will be separated according to original decimal numbers that represented each binary of secret bits:

851101051181011141151051161053275101989711010311597971103277971089712111510597.

Step 4: steps to accomplish the separation according to the original of the decimal numbers:

- take two numbers.
- **If** the summation of the two numbers equals 32, then separate
- **else-if** the sum of the two numbers equals or is more than 64, then separate
- **Else** take three numbers and separate

where 32 represents the word space in decimal, which are separate words. This condition identifies the word spaces in the algorithm. At the same time, since each letter in the English language starts with 64 when represented in decimal, and the numbers reflect the secret message letters in decimal, the retrieval numbers must be 64 or more. This condition could be satisfied with two or three numbers from the string while separating the secret message. The output from this stage will be (85 110 105 118 101 114 115 105 116 105 32 75 101 98 97 110 103 115 97 97 110 32 77 97 108 97 121 115 105 97).

Step 5: Convert each decimal number into binary and then retrieve the secret message from the binary. The binary representation will be as follows:

01010101 01101110 01101001 01110110 01100101
01110010 01110011 01101001 01110100 01101001
00100000 01001011 01100101 01100010 01100001
01101110 01100111 01110011 01100001 01100001
01101110 00100000 01001101 01100001 01101100
01100001 01111001 01110011 01101001 01100001

The secret message will be "**Universiti Kebangsaan Malaysia**".

V. Experimental and Analysis

In this section, we analyze the experimental results of our proposed method. The performance of the proposed method is measured in terms of capacity and imperceptibility. The hiding capacity is a major crucial parameter for analysis of the text steganography algorithm performance. We used the Jaro-Winkler distance, as used in [20], which is a similarity metric for the cover text and stego cover (1) and (2). The similarity metrics were used to calculate how similar two strings were to one another, with (0) indicating a difference and (1) indicating equal matching or imperceptibility of strings.

$$Jaro\_Winkler(S,C) = Jaro\_Score + \left(L * P * (1 - Jaro\_Score)\right) \quad (1)$$

$$Jaro\_Score = \frac{1}{3}\left(\frac{m}{Length(s_1)} + \frac{m}{Length(s_2)} + \frac{(m-t)}{m}\right) \qquad (2)$$

L is the length of the common prefix at the start of the string up to a maximum of 4, P is the constant scaling factor $(0.1 \geq P \leq 0.25)$, m is the number of matched characters, s_1 is the first string, s_2 is the second string, and t is the number of transpositions. As in [18, 21], hiding capacity is defined as the hidden data size relative to the size of the stego cover. Formulation (3) can be used for calculating the hiding capacity.

$$Capacity\ Ratio\ = \frac{Size\ of\ the\ embedded\ data\ (byte)}{Size\ of\ the\ coverfile\ \ (byte)} \times 100 \qquad (3)$$

The proposed method is applied to the cover file as shown in Fig. 9 and secret messages shown in Fig 10, which are divided into twelve samples to be embedded into the cover file. Table III shows the results of the experiments, which include the secret messages, size secret message (n) and Jaro-Winkler (JW), for the proposed method and the comparison with previous related studies.

"The loss of tree cover as a result of forests being cleared for other land uses such as farming and logging is called deforestation. Deforestation activities affect carbon fluxes in the soil, vegetation, and atmosphere. However, logging can also lead to carbon emissions if the surrounding trees and vegetation are damaged. Deforestation is defined as the destruction of forested land. It is a major problem all over the world. The causes of deforestation vary from place to place. The most common causes, however, are logging, agricultural expansion, wars, and mining. Deforestation has been the cause of many problems facing the world today such as erosions, loss of bio diversity through extinction of plant and animal species, and increased atmospheric carbon dioxide. In this paper, I will present that we can reduce deforestation by moving from physical letter mail to electronic mail. From the ancient era physical letter mail has come, till now it is going on, but, on the other side due to this everyday trees are being cut i.e., deforestation is taking place by the paper industry, hence increasing CO2 emission and global warming. In place of physical letter mail, we can use electronic mail which will definitely do some reduction in deforestation. There are critical effects of deforestation."

Fig. 9. Cover text (1)

"The importance and size of text data have increased at accelerating pace because the reliance on text based web01234."

Fig. 10. Secret message (1)

As shown in Table III, the experimental results indicate that the proposed method can be applied to embed the secret message in the cover file. The Jaro-Winkler similarity score is 0.984 which is higher than in the two previous studies.

TABLE III. JW OF THE PROPOSED METHOD COMPARED WITH RELATED STUDIES

Secret Message	Message size (Bit)	JW [15]	JW [16]	JW Proposed work
the import	80	0.99	1	1
the importance and s	160	0.98	1	1
the importance and size of tex	240	0.98	0.99	0.99
the importance and size of text data hav	320	0.97	0.99	0.99
the importance and size of text data have increase	400	0.96	0.99	0.99
the importance and size of text data have increased at an ac	480	0.95	0.98	0.99
the importance and size of text data have increased at an accelerating	560	0.95	0.98	0.98
the importance and size of text data have increased at an accelerating pace beca	640	0.94	0.98	0.98
the importance and size of text data have increased at an accelerating pace because the re	720	0.94	0.98	0.98
the importance and size of text data have increased at an accelerating pace because the reliance on	800	0.93	0.97	0.97
the importance and size of text data have increased at an accelerating pace because the reliance on text based	880	0.93	0.97	0.97
the importance and size of text data have increased at an accelerating pace because the reliance on text based web 01234.	936	0.92	0.97	0.97
Average		0.953	0.983	0.984

Next, the performance in capacity is compared with related studies that use the same secret message, as shown in Fig. 11 and the cover file in Fig. 12. Results show that the proposed method also scores higher in hiding capacity results. Table IV compares the hiding capacity of the proposed method with existing techniques. The proposed method achieved 18.4% capacity, which performed better than the other techniques for the same cover text and the secret message. Fig. 13 presents the bar chart comparison of the capacity that is listed in Table IV.

"Behind using a cover text is to hide the presence of secret messages the presence of embedded messages in the resulting stego text cannot be easily discovered by anyone except the intended recipient"

Fig. 11. Secret message (2)

"In the research area of text steganography, algorithms based on font format have advantages of great capacity, good imperceptibility and wide application range. However, little work on steganalysis for such algorithms has been reported in the literature. based on the fact that the statistic features of font format will be changed after using font-format based steganographic algorithms, we present a novel support vector machine-based steganalysis algorithm to detect whether hidden information exists or not. This algorithm can not only effectively detect the existence of hidden information, but also estimate the hidden information length according to variations of font attribute value. As shown by experimental results, the detection accuracy of our algorithm reaches as high as 99.3% when the hidden information length is at least 16" bits."

Fig. 12. Cover text (2)

TABLE IV. CAPACITY OF PROPOSED METHOD COMPARED WITH RELATED STUDIES

Method	Hiding Capacity Ratio %
[22]	7.21
[20]	12.02
[18]	12.17
[21]	13.43
Proposed Method	18.4

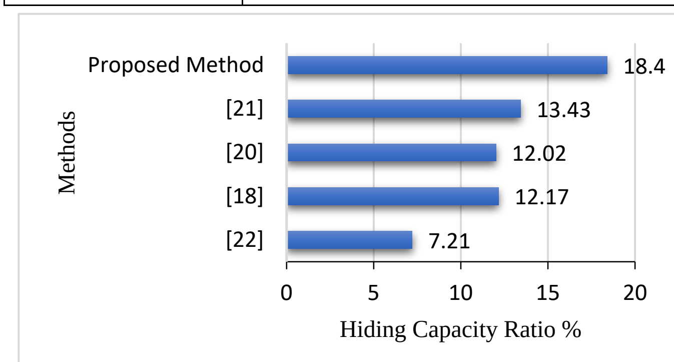


Fig. 13. Capacity of the proposed method compared with related studies

VI. CONCLUSION

This paper proposes a new technique of text steganography using Multilayer encoding with FPE and Huffman coding. The use of UCs shows a significant invisibility due to high imperceptibility after embedding the secret data into the cover text. Each Unicode character represents three bits of data. Before hiding secret data, the proposed scheme minimized the secret data size by applying multilayer encoding and Huffman compression. FPE is applied to secret data to achieve the encryption objective without increasing the size of the secret data. Results show that the proposed method has demonstrated significant improvement when compared with previous studies.

VII. FUTURE WORK

In future work, enhancement can be made by proposing a new method to solve the sequential pattern in the embedding process by introducing a randomization concept. Therefore, the security level will be increased without using any encryption techniques that require sharing of keys between participants. In addition, increasing the number of bits in UC mapping representation will result in increasing the amount of capacity in cover text.

ACKNOWLEDGMENT

The authors would like to acknowledge Universiti Kebangsaan Malaysia for the Research University Grant with code: TAP-K011439 to support this project.

REFERENCES

[1] A. Ditta, M. Azeem, S. Naseem, K. G. Rana, M. A. Khan, and Z. Iqbal, "A secure and size efficient algorithm to enhance data hiding capacity and security of cover text by using unicode," J. King Saud Univ. Inf. Sci., 2020.

[2] M. A. Majeed, R. Sulaiman, Z. Shukur, and M. K. Hasan, "A review on text steganography techniques," Mathematics, vol. 9, no. 21, p. 2829, 2021.

[3] Ibrahim, A. H., & Alturki, A. S. (2020). Computational Analysis of Arabic Cursive Steganography using Complex Edge Detection Techniques. International Journal of Advanced Computer Science and Applications, 11(9).

[4] Shehab, D. A., & Alhaddad, M. J. (2022). Comprehensive Survey of Multimedia Steganalysis: Techniques, Evaluations, and Trends in Future Research. Symmetry, 14(1), 117.

[5] Almayyahi, A. A., Sulaiman, R., Qamar, F., & Hamzah, A. E. (2020). High-security image steganography technique using XNOR operation and fibonacci algorithm. International Journal of Advanced Computer Science and Applications, 11(10).

[6] Yang, Z., Xiang, L., Zhang, S., Sun, X., & Huang, Y. (2021). Linguistic generative steganography with enhanced cognitive-imperceptibility. IEEE Signal Processing Letters, 28, 409-413.

[7] Majeed, M. A., & Sulaiman, R. (2015). AN IMPROVED LSB IMAGE STEGANOGRAPHY TECHNIQUE USING BIT-INVERSE IN 24 BIT COLOUR IMAGE. Journal of Theoretical & Applied Information Technology, 80(2).

[8] Salah, A. H., Hadwan, M., Aqlan, A., Albazel, M., Alqasemi, F., & Al-Sanabani, M. (2021, August). A Survey on Different Arabic Text Steganography Techniques. In 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA) (pp. 1-8). IEEE.

[9] Kumar, R., & Yadav, A. K. (2021). Development of Novel Algorithm for Data Hiding on Mobile Application. International Journal of Advanced Computer Science and Applications, 12(8).

[10] Baziyad, M., Rabie, T., & Kamel, I. (2018, November). Extending steganography payload capacity using the L ab color space. In 2018 International conference on innovations in information technology (IIT) (pp. 1-6). IEEE.

[11] Wahab, O. F. A., Khalaf, A. A., Hussein, A. I., & Hamed, H. F. (2021). Hiding data using efficient combination of RSA cryptography, and compression steganography techniques. IEEE Access, 9, 31805-31815.

[12] B. Khosravi, B. Khosravi, B. Khosravi, and K. Nazarkardeh, "A new method for pdf steganography in justified texts," J. Inf. Secur. Appl., vol. 45, pp. 61–70, 2019.

[13] L. Xiang, W. Wu, X. Li, and C. Yang, "A linguistic steganography based on word indexing compression and candidate selection," Multimed. Tools Appl., vol. 77, no. 21, pp. 28969–28989, 2018.

[14] R. Kumar, A. Malik, S. Singh, and S. Chand, “A high capacity email based text steganography scheme using Huffman compression,” in 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), 2016, pp. 53–56.

[15] Fateh, M., & Rezvani, M. (2021). An email-based high capacity text steganography using repeating characters. International Journal of Computers and Applications, 43(3), 226-232.

[16] N. Alanazi, E. Khan, and A. Gutub, “Inclusion of Unicode standard seamless characters to expand Arabic text steganography for secure individual uses,” J. King Saud Univ. Inf. Sci., 2020.

[17] D. Bhat, V. Krithi, K. N. Manjunath, S. Prabhu, and A. Renuka, “Information hiding through dynamic text steganography and cryptography: computing and informatics,” in 2017 international conference on advances in computing, communications and informatics (ICACCI), 2017, pp. 1826–1831.

[18] G. Maji and S. Mandal, “A forward email based high capacity text steganography technique using a randomized and indexed word dictionary,” Multimed. Tools Appl., vol. 79, no. 35, pp. 26549–26569, 2020.

[19] S. S. Baawi, M. R. Mokhtar, and R. Sulaiman, “New text steganography technique based on a set of two-letter words,” J. Theor. Appl. Inf. Technol, vol. 95, no. 22, pp. 6247–6255, 2017.

[20] S. S. Baawi, M. R. Mokhtar, and R. Sulaiman, “Enhancement of text steganography technique using Lempel-Ziv-Welch algorithm and two-letter word technique,” in Advances in Intelligent Systems and Computing, vol. 843, 2019.

[21] A. Malik, G. Sikka, and H. K. Verma, “A high capacity text steganography scheme based on LZW compression and color coding,” Eng. Sci. Technol. an Int. J., vol. 20, no. 1, pp. 72–79, 2017.

[22] Kumar, R., Malik, A., Singh, S., & Chand, S. (2016, February). A high capacity email based text steganography scheme using Huffman compression. In 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 53-56). IEEE.

[23] Gupta, S., Jain, S., & Agarwal, M. (2018, January). Ensuring data security in databases using format preserving encryption. In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 1-5). IEEE.

[24] Pérez-Resa, A., Garcia-Bosque, M., Sánchez-Azqueta, C., & Celma, S. (2020). A new method for format preserving encryption in high-data rate communications. IEEE Access, 8, 21003-21016.

[25] M. Bellare, P. Rogaway, and T. Spies, “The FFX mode of operation for format-preserving encryption,” NIST Submiss., vol. 20, no. 19, p. 24, 2010.

[26] E. Brier, T. Peyrin, and J. Stern, “BPS: a format-preserving encryption proposal,” Submiss. to NIST, 2010.

[27] Cui, B., Zhang, B., & Wang, K. (2017, July). A data masking scheme for sensitive big data based on format-preserving encryption. In 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) (Vol. 1, pp. 518-524). IEEE.

[28] Oh, I., Kim, T., Yim, K., & Lee, S. Y. (2019). A novel message-preserving scheme with format-preserving encryption for connected cars in multi-access edge computing. Sensors, 19(18), 3869.

[29] B. Carpentieri, A. Castiglione, A. De Santis, F. Palmieri, and R. Pizzolante, “Compression-based steganography,” Concurr. Comput. Pract. Exp., vol. 32, no. 8, p. e5322, 2020.

Secure Palmprint Recognition based on Multispectral Sequential Capture

Amine AMRAOUI[1]*, Mounir AIT KERROUM[2], Youssef FAKHRI[3]
LaRI Laboratory-Faculty of Science, Ibn Tofail University, Kenitra, Morocco[1, 3]
LaRI Laboratory-ENCG, Ibn Tofail University, Kenitra, Morocco[2]

***Abstract*—The security of personal identities is a serious challenge in today's digital world, with so many daily transactions requiring secure solutions. The use of biometric characteristics of the person is presented as the reliable solution to solve this problem. Indeed, this solution is effective, but it hides a weak point which lies in the ability to reproduce certain biometric characteristics for fraud. To overcome this weak point, we propose a secure approach for palmprints that relies on the concept of merging multiple features. Indeed, these features will be extracted from multi-spectral images with different spectra, which allow the extraction of information under the skin of the palm for two different spectrums sequentially in two different times (T1, T2) but instantly. The instant fusion of these characteristics will be impossible to replicate. The images used are grayscale. To satisfy a construction of a reliable and secure system, for this kind of patterns (palmprints), we will use the Compound Local Binary Pattern method, since this method adds an additional bit for each P bits coded by LBP corresponding to a neighbor of the local neighborhood, in order to build a robust system. This feature descriptor, it uses both the sign and tilt information of the differences between the central and neighboring gray values. The reliability of the proposed approach has been demonstrated on the Casia Multi-Spectral database. The final experimental results show reliable recognition rates and these recognition rates vary between 99% and 100% for the left and right palms.**

***Keywords*—Biometrics; multispectral palmprint, local features; fusion; compound local binary pattern**

I. Introduction

The uniqueness of personal identity has always been the major concern of individuals in society. This concern is justifiable given the seriousness of the risks. The identity is the entity requested to ensure the majority of critical transactions. Various applications have been proposed to guarantee this uniqueness, however, in this set of applications, biometrics remains the most appropriate solution [1, 2]. A multitude of scientific research has been undertaken in recent decades in this direction. Researchers have proposed a large set of recognition systems with very promising results. A large number of modalities have been investigated and they have shown a distinctive character that ensures the uniqueness of the identity. Among these modalities, we can cite the use of the fingerprint [3], the iris [4], the finger knuckle [5], the voice [6], the hand geometry [7], etc. However, it should be noted that these different biometric mechanisms offer advantages on the one hand and disadvantages on the other. The suitability of a mechanism for all current applications cannot be judged with certainty. Because of this, scientific researchers have varied their research to try to meet the requirements of different fields. Like all these researchers, the orientation of our work has been guided by constraints. One of the most important constraints that we have taken into consideration relying on the Biometric Zephyr analysis, is intrusion [8]. Indeed, the intrusive nature of the sensors makes the participants uncomfortable during the capture process such as the retina and iris based sensors [9]. This constraint, directed us towards the use of the modalities which reduce the impact of this factor like the descriptors based on the hand. The ease of use and convenience for this kind of descriptors has pushed research in this direction. In the literature, there is a large number of studies that have opted for the use of hand-based modalities and in various fields, such as fingerprints [10, 11], hand geometry [12, 13], finger knuckle print [14, 15] and palmprints [16-18]. Nevertheless, fingerprint-based systems have many challenges that are summed up in the ability of their spoofing easily by third parties [19, 20]. Furthermore, compared to other biometric descriptors, palm prints show low distortion, good stability over time and a highly distinctive character [21, 22]. Unfortunately, palmprint patterns are sensitive to pose orientations, lighting variations, and sensor noise. These factors can lead to misclassification and therefore affect the ability of such systems to recognize users. To overcome these problems, various researches have been done on the extraction mechanisms to improve the relevance of features, which will have a positive impact on the matching of these features. The majority of these researches employed coding-based, statistics-based, subspace-based, and line-based methods.

Coding-based methods are a category widely used in much research on palmprint recognition systems. Various palmprint codes have been generated, among which we cite the fusion code [23], which allows the coding of the phases and the responses of the six Gabor filters. The Robust Line Orientation Code (RLOC) [24], which allows the extraction of palmprint orientation features with a modified finite Radon transform. The feature vector obtained in this technique is used as a competitive code. But, the extracted features have a large size, which leads to a classification over-fitting. The combination of Block Dominant Orientation Code (BDOC) for rough characteristics extraction and Block-based Histogram of Oriented Gradient (BHOG) for fine characteristics extraction [25], Half-Orientation Code (HOC) for palmprint feature extraction, this code employs half Gabor filters [26]. The Double Orientation Code (DOC) of Gabor

*Corresponding Author.

filters for feature extraction with a nonlinear classifier [27]. This category can be expensive in computation time and implementation.

Another category that is also used, this category relies on statistics-based approaches; they have shown good performance [28–29]. A multitude of statistical methods have been employed in this category such as variance, standard deviation, energy and histograms of local binary models [30], Zernike and Hu moments[31]. Some transformations have also been used such as the wavelet transform to convert the palmprint image into a small number of wavelet coefficients, and then calculate the variance and the mean of these coefficients to generate the image characteristics [32]. Other methods have been introduced to calculate the statistical characteristics after the extraction phase with certain filters and transforms such as Gaussian derivative filters, Gabor wavelet, and Fourier transform [33]. The use of principal component analysis (PCA) and wavelet transform to extract features from the quaternion matrix, combined with the Euclidean distance classifier for matching process [34].

Approaches based on the concept of subspace also form an important category that intervenes in the construction of recognition systems. This kind of approach has been widely used in scientific research related to palmprint recognition. Among this research, we can cite the use of approaches in the works such as principal component analysis (PCA) [35], independent component analysis (ICA) [36] and linear discriminant analysis (LDA) [37]. In [38], a method to convert palmprint images into a set of feature space, named eigenpalms was proposed to build a palmprint recognition system. For matching, the Euclidean distance classifier was been employed. Researchers in [39] used Two-dimensional vertical and horizontal LDA (2DLDA) to extract Gabor features and then a distance-based adaptive approach to merge the vertical and horizontal features. Other approaches based on multi-spectral images have been deployed; in [40] kernel discriminant analysis (KDA) was used to reduce feature dimensionality and classification was provided by a KNN classifier and in [41], researchers proposed a multi-spectral method based on a digital Shearlet transform. This category can be affected by environmental variables (light and pose rotation).

A last category which is based on the use of lines to recover outlines and lines of palm prints with edge detection methods. Among the studies based on this concept we can cite [42], in which a technique that employs the Sobel edge detector with morphological operations to ensure the feature extraction phase of lines and [43] who used the Sobel mask to calculate the amplitudes of the lines and projected those amplitudes along the *x* and *y* axes to produce the discriminant histograms for each class. This last category offers medium performance.

In a brief summary of what preceded, we can identify the following issues: the intrusiveness of the modality, the complexity of many methods employed, the sensitivity of some methods to environmental lighting or rotation of employed modality (in the case of the hand-based modalities) and finally the possibility of usurping the personal identity by the reproduction of false descriptors.

To overcome these problems, we propose a reliable approach using local features to build a non-intrusive, secure, robust and efficient recognition mechanism, in which, we take into account the optimization of the computational complexity. We will use palmprints to reduce intrusiveness. To reduce the influence of environment lighting, capture conditions and extract relevant features, we will use method of extracting local features Compound Local Binary Pattern (CLBP) [44, 45], which we will evaluate on different block sizes with different classifiers. The classification phase will be based on the use of classifiers that reduce computation time such as distance-based classifiers. We will test three variants of measurements: Jeffrey Divergence, City-Block and Euclidian Distance. We will also pre-process the Casia Multispectral Palmprint database [46], in order to reduce computation time and improve performance.

This paper is organized as follows: Section II describes the proposed approach and the methods used. Section III reports and discuss the experimental results conducted on the Casia Multispectral Palmprint database. Finally, Section IV draws conclusions.

II. Proposed Approach

The design of the global system went through several stages before arriving at the final system scheme. In the first part, we chose a local extraction method, given the nature of palmprints pattern, which are rich with local information. In addition, this method must be as robust to variations in the lighting of the environment and to rotations. It turns out that the compound local binary pattern method [47] can satisfy these constraints. The objectives of this part can be summarized in two points. In the first point, the goal is trying to demonstrate the reliability of the chosen extraction method.

The second point will concern the classification, which represents a crucial phase to determine the reliability of the systems and their cost in computing time. In this sense and in order to satisfy its needs, experiments will be made with distance-based classifiers, known for their reduced computation time. Several classifiers will be discussed at the beginning, then the choice will be based on the most appropriate with respect to the local extraction method used (ratio optimization: recognition rate and computation time). The concept used for this evaluation is illustrated in Fig. 1.

In the second part, a new gait will be experimented, which aims at the adequacy of the size of the cropping chosen to extract the regions of interest, in order to further improve the reliability and the reduction of the calculation times. During this last point, we will extract the regions of interest with the Principal Based ICP method [48], this cropping will be done with a size of 192x192 pixels; which will generate the first database of ROI images, thereafter, we will proceed to a resizing of the ROI images with the size 128x128 pixels to obtain the second database (Fig. 2). With this approach, it will be discovered that the size 128x128 pixels can offer better results than 192x192 pixels, this can be justified by the fact that if the size of the division block is large and the pattern is poor in information, then the rate inter-class similarity increases, which will reduce the recognition rate. The consequence of this increase will be a reduction in system performance and an increase in computation times, which we aim to avoid with this approach.

Subsequently, in the experimental part, the relationship between cropping resolution and scoring performance will be demonstrated.

Finally, in the last part, the experimental results and the conclusions obtained in the two previous steps will be used to test the reliability and robustness of the proposed approach. In the proposed approach, the choice will be in favor for a sequential capture of two images at two different times T1 and T2, which follow each other with two different spectrums, in our case we will use the 640 nm spectrum for image 1 and 940 nm for the image (we will explain this choice in the experimental part). These two images will form two different descriptors, in addition, with the 940nm spectrum some veins under the skin will be visible and other characteristics will change. The proposed palmprint multispectral recognition system is shown in Fig. 3.

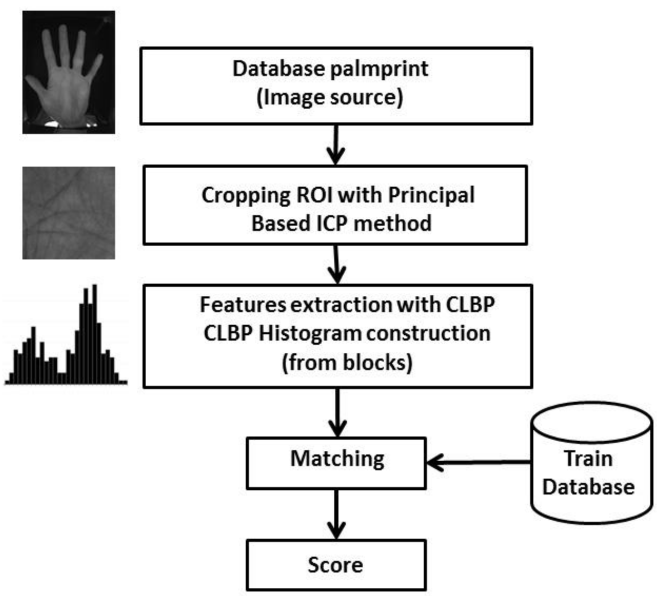


Fig. 1. FKP Recognition System adopted for evaluation.

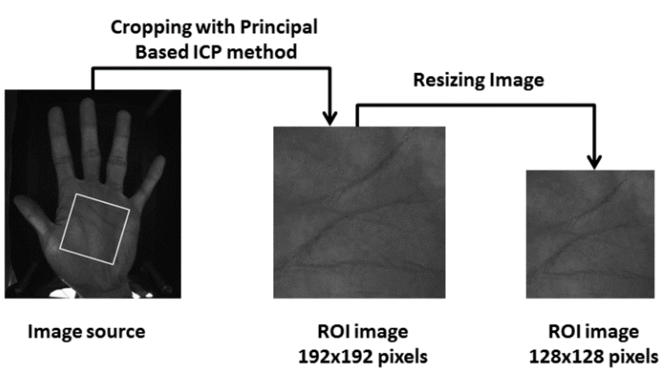


Fig. 2. Cropping and resizing process for ROI image.

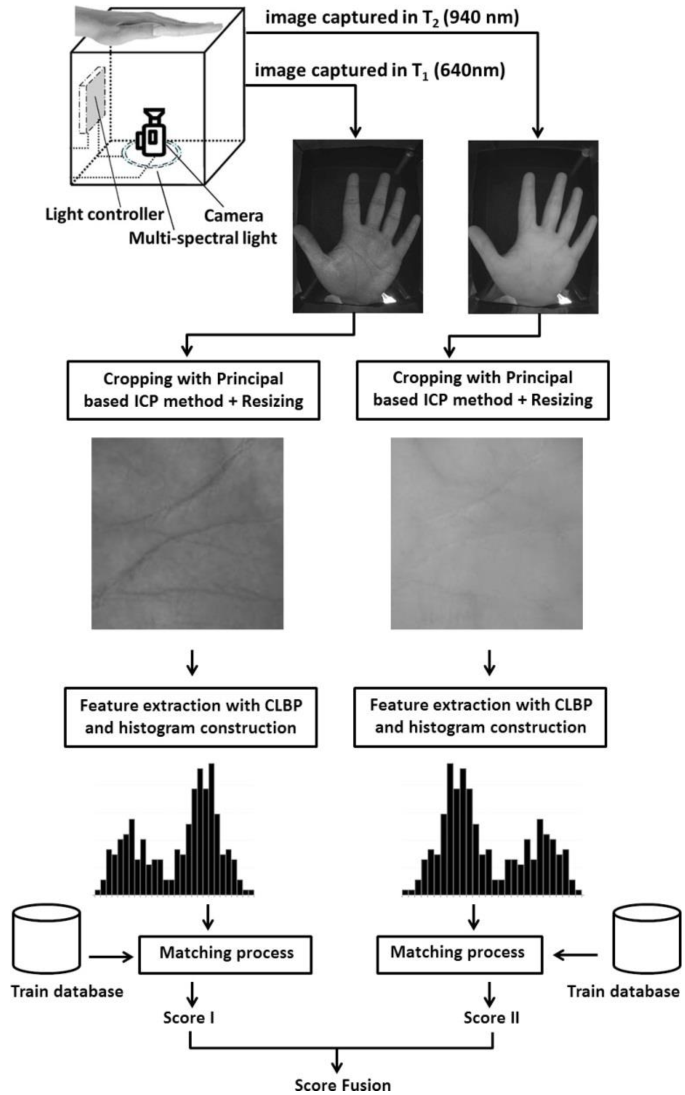


Fig. 3. The proposed multispectral recognition system.

Indeed, this sequential capture in very close times (in ms) will aim; the security of the capture phase against any attempts at fraud or use of false artifacts. Since these fake methods cannot deploy quickly to present two descriptors in a short time (in ms) and therefore, even if descriptor 1 (image1) will be spoofed, descriptor 2 (image2) will not. This will secure the proposed system. This approach will be based on fusion at the score level. The results obtained in this step will be analyzed and compared to the previous step.

A. *Features Extraction Process*

The construction of recognition mechanisms is generally based on two key factors; the first factor depends on the computation time allocated to satisfy the operation, the second factor depends on the efficiency and the recognition rates. It is also necessary to take into consideration the conditions of the environment which can constitute handicaps for the performance. To avoid this kind of handicaps, local methods robust to rotations and light variations will be introduced (which is the case of this database of palm prints).The feature extraction step will be ensured by a robust variant of Local

Binary Patterns (LBP), this method is called Compound Local Binary Patterns (CLBP) [47]. The LBP method was first introduced by Ojala and Al [49], this method is effective for feature extraction from images in real environments. The local binary pattern is recognized by a gray scale texture operator characterizing the local spatial structure of the image texture [50]. Given a central pixel in the image, a pattern code is calculated by comparing it to its neighbors. This method is illustrated in Fig. 4. The LBP operator calculates the signs of the differences of the gray levels of P neighbors equidistant with respect to the central pixel, which will be represented with a binary number of P bits (Fig. 4). If a neighbor is not exactly on a pixel position, then the value of that neighbor will be estimated using bilinear interpolation. The histogram of the coded image block obtained with the LBP operator will then be used as a texture descriptor for this block.

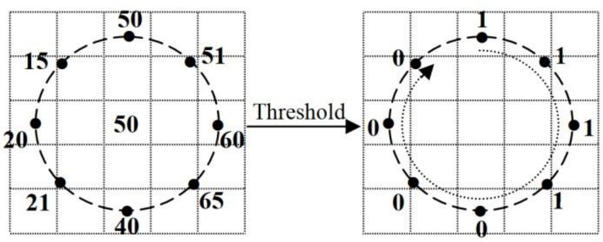


Fig. 4. Local binary pattern operator.

The LBP operator takes the form:

$$\mathrm{LBP}(x_c, y_c) = \sum_{n=0}^{7} 2^n S(i_n - i_c) \quad (1)$$

where in this case n runs over the eight neighbors of the central pixel c, i_c and in are the gray-level values at c and n. Function S(x) is s defined below,

$$S(x) = \begin{cases} 1 & if\ \ x \geq 0 \\ 0 & if\ \ x < 0 \end{cases} \quad (2)$$

The LBP operator uses a method that employs only the use of the sign of the difference between two gray values, which sometimes leads to a failure to generate binary codes consistent with the texture properties for the local region (Fig. 5). To avoid this handicap, a variant of this operator will be used. This variant assigns a 2P binary code to the central pixel according to the gray values of the local neighborhood comprising P neighbors; this method is Compound Local Binary Pattern (CLBP) [47]. The CLBP operator uses two bits for each neighbor to encode the sign as well as the magnitude information of the difference between the center and the neighboring gray values, unlike the LBP which uses only one bit for each neighbor by representing the sign of the difference between the center and the corresponding neighboring gray values. In this method, the first bit represents the sign of the difference between the center and the corresponding neighboring gray values as the basic LBP encoding. The second bit is used to encode the magnitude of the difference with respect to a threshold value, which is the average magnitude M_{avg} of the difference between the central and neighboring gray values in the local neighborhood of interest. The CLBP operator chooses the value 1 for the second bit if the magnitude of the difference between the center and the corresponding neighbor is greater than the threshold M_{avg}. Otherwise, it takes the value 0. Thus, the s(x) indicator of equation 2 is replaced by the function below:

$$s(i_p, i_c) = \begin{cases} 00\ ip - ic < 0, & |i_p - i_c| \leq M_{avg} \\ 01\ ip - ic < 0 & |i_p - i_c| > M_{avg} \\ 10\ ip - ic \geq 0, & |i_p - i_c| \leq M_{avg} \\ 11 & otherwise \end{cases} \quad (3)$$

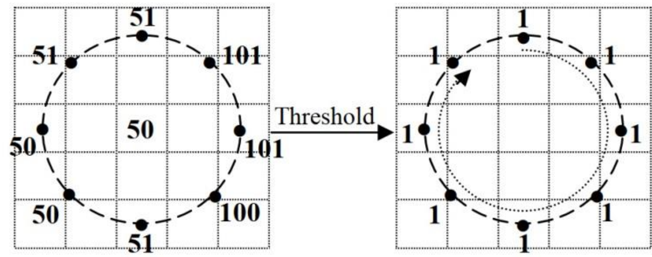


Fig. 5. Generation of inconsistent binary pattern in LPB.

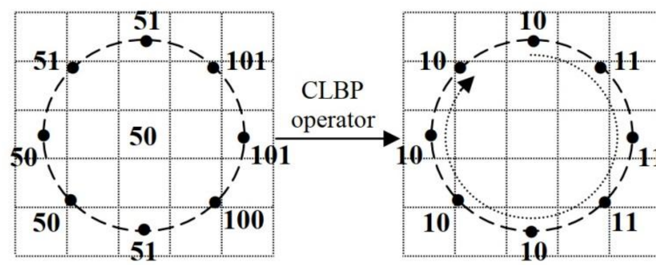


Fig. 6. Compound local binary pattern operator.

where i_c and i_p are the gray values of the central pixel and the neighbors, and the average magnitude of the difference between i_p and i_c in the local neighborhood is M_{avg}. The mechanism of the CLBP operator is shown in the Fig. 6.

B. Matching Process Phase

The choice of classifiers is crucial for a better performance concerning the proposed approach, to ensure this important phase in all recognition systems, we relied on two important factors. The first factor depends on the resolution of the images used, which is low, classifiers like SVM will be disregard, because are more for high resolutions [51, 52]. The second factor is also important, it depends on the computation time costs, which represents a decisive point for the success of a recognition mechanism; these points will direct our choice towards distance-based classifiers. The consideration of the directives already mentioned, directed the choice towards a set including three classifiers based on the distance, which we will use with the method of extraction employed and see their performances for the generation of the recognition rates. These classifiers are based on Euclidean distance, City-block and Jeffrey Divergence.

The Euclidean distance is the best-known distance metric and used in datasets that represent low-dimensional images, examines the root of the squared differences between the coordinates of a pair of objects. This process is generally known as the Pythagorean Theorem. For the tests, we used this classifier, to calculate the minimum distance between the test image and the train image. The Euclidean distance d is as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

The Manhattan distance classifier, city-block distance classifier, also called, rectilinear distance, L1 distance, L1 norm, Manhattan length. It represents the distance between points in a city road grid. It examines the absolute differences between the coordinates of a pair of objects as follows:

$$d(x, y) = \sum_{i=1}^{n}| x_i - y_i| \quad (5)$$

The Jeffrey divergence is a modification of the Kullback-Leibler (KL) divergence, if P = (p_1, . . . p_N) and Q = (q_1, . . . q_N) are two discrete distributions, the Jeffrey divergence between P and Q is defined as:

$$D(P, Q) = \sum_i (p_i \ log \frac{p_i}{m_i} + q_i \ log \frac{q_i}{m_i} \quad (6)$$

Where $m_i = \frac{(p_i + q_i)}{2}$

III. Experimental Result

In this part, an experimental study spread over several stages will be made. The objective of these steps will be articulated on the progressive demonstration of the process reliability for the proposed system. The evaluation of this system will be conducted on the Casia multispectral palmprint database [46]. It should be noted that this database is one of the standards of scientific research in this field.

A. *Casia Multispectral Palmprint Database*

The construction of the CASIA Multi-Spectral Palmprint database is based on the acquisition of 7200 palmprint images from 100 volunteers. This set of images is shared equally between the palms of the right hand and the left hand (3600 images on each hand). It should be noted that the set of palmar images for each hand is captured in two sessions. The time between the two capture sessions has an interval greater than one month. During a session, each volunteer gives three samples for each hand (a total of six images per user in a single session, so 600 images for 100 volunteers). Each image of a sample is captured with six different electromagnetic spectrums (Fig. 7). The wavelengths used for these captures correspond to the following six spectrums: white light (WL), 460 nm, 630 nm, 700 nm, 850 nm and 940 nm. The captures have been made in such a way that the user has a certain freedom for the pose concerning the angle and the rotation of his hand, the goal is to produce variations of hand postures during the capture session. This procedure will simulate a use that is done in the real world and increase the diversity of samples in the same class, which will present interesting challenges for measuring the performance of a biometric system. It is reminded that the palm images captured are 8-bit grayscale JPEG files.

Generally, the biometric database preparation follows a process that relies on two phases; the first concerns the acquisition of images over different time intervals and the second revolves around preprocessing mechanisms, that can vary depending on the modality nature and the objectives targeted by the research conducted.

A careful examination of the database images reveals two important factors that will direct our preprocessing towards an extraction of the regions of interest; the first is seen on the capture of hands that will undergo changes (Hand with Corn, Callosity, Induration) due to manual work or intense physical activities in strength sports such as bodybuilding or powerlifting (Fig. 8).

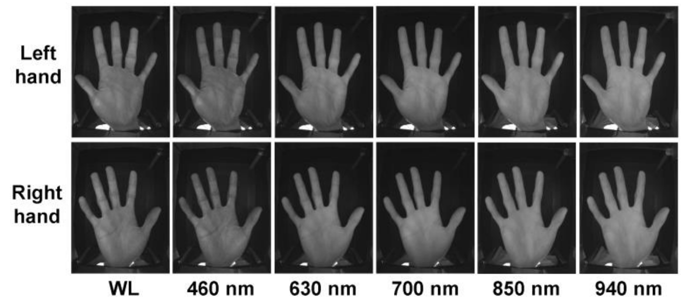


Fig. 7. Casia multipectral palmprint database.

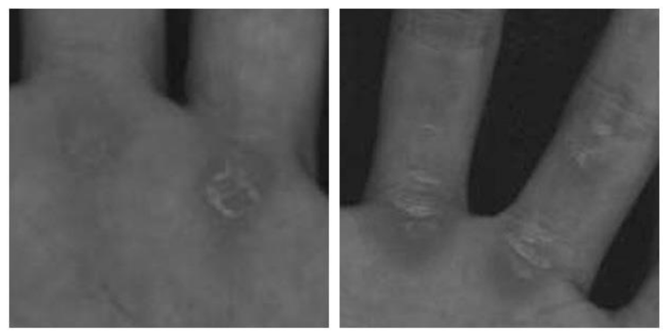


Fig. 8. Images with corn, callosity and induration in casia multipectral palmprint database.

The second factor is noticeable in the images where the pose of the hand for capture is free and not standardized. This hand posing manner (flexion hand) may in some cases present angles with respect to the capture support, which cause the relevant features at the level of the hand palm contour to be inhibited, the same in the case of the poses of the hand which have a concavity (Fig. 9).

Fig. 9. Poses with flexion and concavity in casia multipectral palmprint database.

These two factors may skew experimental results, on the one hand, the change that certain parts of hand palm may undergo, which increases intra-class variations, on the other hand, the absence of relevant features on areas of the hand palm, which may increase the interclass similarities of the histograms with the local methods. To avoid these problems, it is essential to carry out a pre-processing which aims at extracting the regions of interest ROI. This process of regions of interest extraction will be used to solve the problems of the Casia Multi-spectral Palmprint database images already mentioned. Indeed, the principle aims to adjust the rotation of different angles and the normalization of the scales, then to carry out a cropping on the normalized image central region to extract the relevant characteristics. It is obvious that an ROI extraction method that is based on a good algorithm will be decisive in generating palm prints images with zones, which allow the maximum of distinctive characteristics. This effect will have a significant impact on the performance of the recognition system to be built. The majority of existing methods use algorithms that perform an ROI extraction based on the limit of the palm, or key landmarks between the fingers, or some external factors [53, 54].

Many existing palmprint recognition mechanisms assume that palmprint images are aligned before performing feature extraction and matching. Because of this, they are often affected by residual variations in translation and rotation after the alignment process. To avoid this problem, the Principal Based ICP [48] will be employed, this method uses the linear features to refine the alignment of the image before performing the cropping operation to extract Region of Interest (ROI) images (Fig. 10).

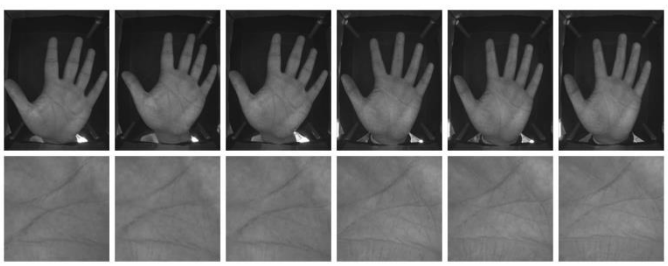


Fig. 10. Casia multipectral palmprint ROI database.

B. Evaluation and Analysis of Experimental Results

To approve the recognition mechanism, we opted for an evaluation process divided into three different parts. The first part focuses on the extraction of ROI images with the Principal Based ICP method for a crop of 192x192 pixels and verifying the ability of the chosen classifiers to offer efficient results with the extraction method adopted. In this phase, several detailed experiments for each hand (left and right) will be carried out and which will have as a goal; the determination of the classifiers most adapted to this kind multi-spectral images, this adaptation will obviously be evaluated on two points: the first concerns the recognition rates obtained and the second point concerns the duration of the calculations necessary for the matching process concerning all the classes. In the second part, the resolution adopted for the ROI images will be changed. The resolution will become 128x128 pixels instead of 192x192 pixels. After obtaining the results of this new phase, we will compare them with the previous phase and we will proceed to the analysis of the data to draw conclusions concerning the impact of the resolution of the global image. In the case where the impact of this operation represents positive points in terms of performance and computation time, it will be adopted for the rest of the experiments. Finally, in the last part, the conclusions obtained in the global approach will be used, the final experimental results of the proposed approach will be examined and the conclusions on the obtained reliability will be drawn.

1) Multispectral palmprint evaluation: During this first step, the system will be evaluated. It should be noted that the Casia database presents images in the same class, which represent concavities during the pose and varying angles depending on the capture pose, these factors increase the intra-class variation. For this, an experimental protocol which will respect the real conditions will be used (Fig. 11). The training database will be composed of five images per user and the test database will be composed of a single image per user (as in the real case to ensure the operation in the devices implemented). Therefore, for each volunteer, there are five training samples and one test sample for each spectrum. The proposed approach is based on local techniques. Therefore, the performance of the mechanism will be examined with different sizes of sub-images.

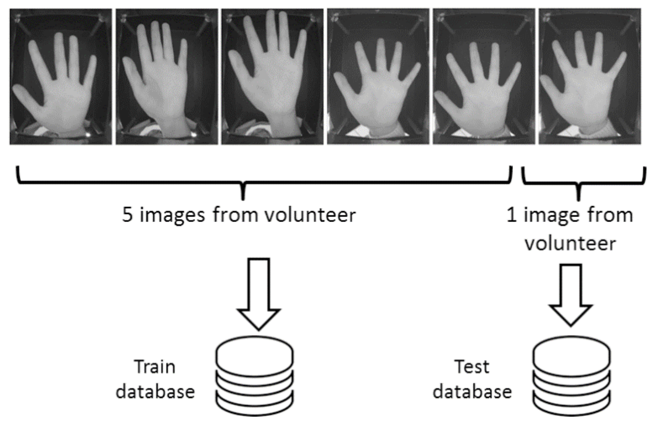


Fig. 11. Protocol used for experiments.

Block subdivision sizes have been classified into three categories: large subdivision, medium subdivision, and small subdivision. The large subdivision is defined by two block sizes: 64x64 and 48x48 pixels, for the medium size: 32x32 and 24x24 pixels, and for the small size: 16x16 and 8x8 pixels. To ensure a better performance of the classification process of the proposed approach, several classifiers based on Euclidean distance, Jeffrey divergence and City-Block was experimented. The recognition rates for each palmprint with the image sub-divisions chosen for the different spectra will be presented later in tables. This comparative evaluation is made in order to show the most adaptive classifiers in our case and the cost in computation time.

a) Result of experiment with cropping 192x192 resolution: In this first phase of the experiments, we will use the diagram Fig. 1 with the resolution 192x192 pixels for the images obtained from the cropping made by Principal Based ICP method. This phase aims to evaluate the extraction method, the classifiers used and their impact on Matching Process Time (MPT) with all the spectrums: white light, 460nm, 630nm, 700nm, 850nm and 940nm.

TABLE I. RECOGNITION RATE FOR LEFT PALM - WHITE LIGHT

Table Recognition rate RR for left palm with white light				
Block size	***RR/MPT***	***Euclidian distance***	***Jeffrey divergence***	***City-Block***
64x64	RR	74%	**82%**	81%
	MPT	0,079923 s	4,801806 s	0,139913 s
48x48	RR	79%	**85%**	83%
	MPT	0,085569 s	8,424969 s	0,215667 s
32x32	RR	83%	**87%**	**87%**
	MPT	0,143411 s	18,388208 s	0,431222 s
24x24	RR	85%	**89%**	**89%**
	MPT	0,224072 s	31,534391 s	0,734840 s
16x16	RR	87%	**92%**	**92%**
	MPT	0,414146 s	64,496038 s	1,608370 s
8x8	RR	78%	**92%**	89%
	MPT	1,774764 s	166,262356 s	6,320234 s

This Table I, clearly shows that the recognition rates (RR) obtained by the classifiers based on the city-block distance and the Jeffery divergence are clearly better than those obtained by the classifier based on the Euclidean distance. The classifiers based on the city-block distance and the Jeffery divergence give the best value of recognition rate 92% for the divisions (sub-images) of 8x8 and 16x16 pixels. We also note that the Matching Process Time MPT increases proportionally to the decrease in the size of the division block. This remark is logical, since the smaller the block size, the more the number of sub-images increases, and consequently the histogram of the image too. It should be noted that despite the similar performance shown by the classifier based on Jeffrey divergence (92%) with the 16x16 pixel block compared to that obtained with city-block (92%), nevertheless, the overall computation time for the set of classes is high (64.496038 s) compared to the city-block distance based classifier (1.608370 s).

We can conclude that even if the rates obtained with Jeffrey diverge and city-block are almost similar, the ratio (recognition rate \ calculation time) remains in favor of city-block. In the following, we are going to join on the same table the recognition rates of the left and right palm prints (resolution 192x192) and no longer record the calculation time, since we already know the behavior of the calculation time with the Table I.

For the right palm (R Palm), Table II shows the same remarks as the left palm (L Palm), the superiority of the classifiers based on the Jeffrey divergence and the City-block, and that the computation time is privileged for the city-block. Note that the value of the high recognition rate is **97%** with Jeffrey divergence for block size 8x8 pixels.

In Table III, the results obtained affirm that the resulting recognition rates with the classifiers based on the Jeffrey divergence and the City-block are the best. Note that the value of the high recognition rate obtained is 91% with city block for a block of size 8x8 pixels with Left palm and 98% with right palm.

In Table IV, we have the same previous remarks, the superiority of the classifiers based on Jeffrey divergence and the City-block is maintained. Note that the value of the high recognition rate obtained is 88% with Jeffrey divergence for a block of size 24x24 pixels with left palm and 100% with Jeffrey divergence for a block of size 8x8 pixels with right palm. City-block also shows similar results: 99% for a block of size 8x8, 16x16 and 24x24 pixels with right palm, and 87% for a block of size 16x16 and 24x24 pixels with left palm.

TABLE II. RECOGNITION RATE - WHITE LIGHT

Table Recognition rate RR for left and right palm with white light						
Block size	***Euclidian distance***		***Jeffrey divergence***		***City-Block***	
	L Palm	***R Palm***	***L Palm***	***R Palm***	***L Palm***	***R Palm***
64x64	74%	63%	82%	80%	81%	81%
48x48	79%	76%	85%	89%	83%	87%
32x32	83%	80%	87%	92%	87%	91%
24x24	85%	87%	89%	95%	89%	94%
16x16	87%	90%	**92%**	96%	**92%**	95%
8x8	78%	87%	**92%**	**97%**	89%	96%

TABLE III. RECOGNITION RATE – SPECTRUM 460

Table Recognition rate RR for left and right palm - spectrum 460nm						
Block size	***Euclidian distance***		***Jeffrey divergence***		***City-Block***	
	L Palm	***R Palm***	***L Palm***	***R Palm***	***L Palm***	***R Palm***
64x64	77%	76%	87%	87%	85%	84%
48x48	81%	81%	88%	92%	87%	90%
32x32	84%	85%	90%	96%	89%	94%
24x24	85%	95%	88%	96%	88%	96%
16x16	87%	96%	89%	97%	89%	97%
8x8	86%	90%	90%	96%	**91%**	**98%**

TABLE IV. RECOGNITION RATE – SPECTRUM 630

Table Recognition rate RR for left and right palm - spectrum 630nm						
Block size	***Euclidian distance***		***Jeffrey divergence***		***City-Block***	
	L Palm	***R Palm***	***L Palm***	***R Palm***	***L Palm***	***R Palm***
64x64	72%	76%	84%	89%	80%	87%
48x48	74%	83%	85%	95%	83%	92%
32x32	84%	90%	86%	97%	87%	97%
24x24	85%	93%	**88%**	98%	87%	99%
16x16	79%	95%	87%	99%	87%	99%
8x8	69%	87%	86%	**100%**	83%	99%

TABLE V. RECOGNITION RATE – SPECTRUM 700

Table Recognition rate RR for left and right palm - spectrum 700nm						
Block size	*Euclidian distance*		*Jeffrey divergence*		*City-Block*	
	L Palm	*R Palm*	*L Palm*	*R Palm*	*L Palm*	*R Palm*
64x64	72%	77%	86%	85%	83%	82%
48x48	80%	81%	86%	91%	85%	86%
32x32	80%	88%	87%	96%	85%	95%
24x24	85%	88%	90%	95%	89%	95%
16x16	80%	88%	**91%**	95%	90%	94%
8x8	73%	81%	88%	**97%**	87%	93%

In Table V, we note again the superiority of the classifiers based on the Jeffrey divergence and the City-block. Note that the value of the high recognition rate obtained is 91% with Jeffrey divergence for a block of size 16×16 pixels with left palm, and 97% for a block of size 8x8 pixels with right palm.

TABLE VI. RECOGNITION RATE – SPECTRUM 850

Table Recognition rate RR for left and right palm - spectrum 850nm						
Block size	*Euclidian distance*		*Jeffrey divergence*		*City-Block*	
	L Palm	*R Palm*	*L Palm*	*R Palm*	*L Palm*	*R Palm*
64x64	73%	75%	92%	94%	87%	92%
48x48	84%	88%	94%	96%	92%	96%
32x32	86%	92%	93%	96%	93%	97%
24x24	89%	94%	94%	**98%**	**95%**	**98%**
16x16	86%	98%	94%	**98%**	94%	**98%**
8x8	68%	81%	**95%**	**98%**	91%	**98%**

In Table VI, the superiority of classifiers based on Jeffrey divergence and City-block is maintained. Note that the value of the high recognition rate is 95% with Jeffrey divergence for a block of size 8x8 pixels and City-block for a block of size 24x24 pixels, the same for the right palm; we notice that the highest recognition rate is obtained with Jeffrey divergence and City-block for the blocks: 8x8, 16x16 and 24x24 pixels.

TABLE VII. RECOGNITION RATE – SPECTRUM 940

Table Recognition rate RR for left and right palm - spectrum 940nm						
Block size	*Euclidian distance*		*Jeffrey divergence*		*City-Block*	
	L Palm	*R Palm*	*L Palm*	*R Palm*	*L Palm*	*R Palm*
64x64	77%	74%	89%	92%	85%	88%
48x48	85%	83%	93%	95%	89%	93%
32x32	86%	92%	**94%**	97%	92%	97%
24x24	91%	94%	93%	**98%**	**94%**	97%
16x16	90%	97%	**94%**	**98%**	**94%**	**98%**
8x8	80%	89%	93%	**98%**	**94%**	**98%**

Finally, for the 940nm spectrum, Table VII again shows the superiority of the classifiers based on the Jeffrey divergence and the City-block. Note that the value of the higher recognition rate is 94% with Jeffrey and City-block divergence for different block sizes: 8x8, 16x16, 24x24 and 32x32 pixels with left palm. We notice the same thing for the right palm, note that the value of the higher recognition rate is 98% with Jeffrey and City-block divergence for different block sizes: 8x8, 16x16pixels.

In summary, we can draw up the following two tables which summarize the recognition rates obtained for all spectrums.

TABLE VIII. BETTER RECOGNITION RATES FOR LEFT PALM – 192X192 RESOLUTION

Table Recognition rate RR for left palm						
Block size	**8x8**	**16x16**	**24x24**	**32x32**	**48x48**	**64x64**
White light	**92%**	**92%**	89%	87%	85%	82%
460 nm	91%	89%	88%	90%	88%	87%
630 nm	86%	87%	88%	86%	85%	84%
700 nm	88%	91%	90%	87%	86%	86%
850 nm	**95%**	94%	**95%**	93%	94%	92%
940 nm	**94%**	**94%**	94%	94%	93%	89%

TABLE IX. BETTER RECOGNITION RATES FOR RIGHT PALM – 192X192 RESOLUTION

Table Recognition rate RR for left palm						
Block size	**8x8**	**16x16**	**24x24**	**32x32**	**48x48**	**64x64**
White light	**97%**	**96%**	**95%**	92%	89%	81%
460 nm	**98%**	**97%**	**96%**	**96%**	92%	87%
630 nm	**100%**	**99%**	**99%**	**97%**	95%	89%
700 nm	**97%**	**95%**	**95%**	**96%**	91%	85%
850 nm	**98%**	**98%**	**98%**	**97%**	**96%**	94%
940 nm	**98%**	**98%**	**98%**	**97%**	**95%**	92%

Both Tables VIII and IX show promising recognition rates for the proposed mechanism which vary between 81% and 100%. We also see the superiority of the results obtained with palm prints of the right hand. We can also say that on average the spectrums that give the best results for both hands are the 850nm and 940nm spectra with the exception of the 630nm spectrum for the right hand. It should also be noted that the (recognition rate / global computation time) ratio remains in favor of City-block.

b) Result of experiment with resizing 128x182 resolution: Based on the previous conclusions concerning the computation time, and the pattern characteristics of the regions of interest of the palmprints which show clearly visible and large-scale features, we thought that there is a relationship between cropping resolution and the performance of the proposed approach which is based on local methods. This assumption comes from the fact that if we have a high resolution with little relevant information, then we will have division blocks that will be similar between the different classes and therefore a reduction in recognition rates. This is what we will try to dismantle with the following experiments.

In the following we will resize images from 192x192 pixels to 128x128 pixels. The results of previous experiments will guide future experiments towards small and medium-sized blocks, which give reliable recognition rates. The second orientation concerns the choice of 940nm, this choice is due to the fact that the 940nm spectrum shows the veins of the palm which makes the system robust to the temptations of fraud and it has shown good results. The last orientation concerns the 460 nm spectrum, this spectrum best show the characteristics of the pattern and it shows a good performance against white light. The results of these choices will help in the construction of the final mechanism.

At the beginning, we will experiment with white light for left palm, to study the improvement brought with this resize.

TABLE X. RECOGNITION RATE FOR LEFT PALM – WHITE LIGHT

Table Recognition rate RR for left palm – white light				
Block size	***RR/MPT***	***Euclidian distance***	***Jeffrey divergence***	***City-Block***
32x32	RR	82%	**88%**	86%
	MPT	0,067901 s	5,478638 s	0,142141 s
24x24	RR	85%	**92%**	91%
	MPT	0,095663 s	10,647785 s	0,257458 s
16x16	RR	88%	**93%**	91%
	MPT	0,160716 s	19,651213 s	0,472513 s
8x8	RR	86%	91%	**93%**
	MPT	0,524071 s	53,258396 s	1,867898 s

The Table X shows a clear reduction in the calculation time compared to the 192x192 resolution, we cite as an example 166.262356 s (8x8 Jeffrey divergence in 192x192) and 6.320234s (8x8 City-block in 192x192) which will be reduced to 53,258396 s (8x8 Jeffrey divergence in 128x128) and 1,867898s (8x8 City-block in 128x128). Similarly the recognition rates have been improved. Note that the value of the high recognition rate is 93% with Jeffrey divergence for block size 16x16 pixels and City-block for block size 8x8 pixels against 92% obtained with Jeffrey divergence for block size 8x8 with 192x192 (cropping resolution) and MPT= 166.262356 s. In the rest of the experiments, we will no longer put the calculation times in the tables; we know very well that it will be significantly reduced.

TABLE XI. RECOGNITION RATES FOR 128X128– WHITE LIGHT

Table Recognition rate RR for left and right palm – white light						
Block size	***Euclidian distance***		***Jeffrey divergence***		***City-Block***	
	L Palm	***R Palm***	***L Palm***	***R Palm***	***L Palm***	***R Palm***
32x32	82%	80%	88%	95%	86%	94%
24x24	85%	87%	92%	97%	91%	95%
16x16	88%	98%	**93%**	98%	91%	98%
8x8	86%	95%	91%	98%	**93%**	**99%**

We note in Table XI, the superiority of the Jeffrey divergence and the City-block classifiers, and the value of the highest recognition rate is 99% City-block for blocks of size 8x8 with right palm and 93% with left palm.

TABLE XII. RECOGNITION RATES FOR 128X128– SPECTRUM 460

Table Recognition rate RR for left and right palm - spectrum 460nm						
Block size	***Euclidian distance***		***Jeffrey divergence***		***City-Block***	
	L Palm	***R Palm***	***L Palm***	***R Palm***	***L Palm***	***R Palm***
32x32	85%	93%	92%	99%	92%	99%
24x24	87%	96%	**95%**	99%	**95%**	99%
16x16	92%	**100%**	93%	99%	93%	99%
8x8	89%	89%	93%	**100%**	**95%**	**100%**

The Table XII shows good value rates, the value of the high recognition rate is 95% with Jeffrey divergence and City-block for blocks of size 8x8 pixels and 24x24, similarly, for the right palm we have perfect value rates, the value of the high recognition rate is 100% with Jeffrey divergence and City-block for blocks of size 8x8 pixels.

TABLE XIII. RECOGNITION RATES FOR 128X128– SPECTRUM 940

Table Recognition rate RR for left and right palm - spectrum 940nm						
Block size	***Euclidian distance***		***Jeffrey divergence***		***City-Block***	
	L Palm	***R Palm***	***L Palm***	***R Palm***	***L Palm***	***R Palm***
32x32	76%	85%	95%	97%	92%	94%
24x24	86%	91%	94%	**98%**	95%	96%
16x16	89%	93%	94%	**98%**	**96%**	**98%**
8x8	81%	92%	93%	**98%**	94%	**98%**

Table XIII also shows good value rates, the high recognition rate value is 96% City-block for blocks of size 16x16 with the left palm. With regard to the right palm, we have effective value rates; the value of the high recognition rate is 98% with Jeffrey and City-block divergence for blocks of size 8x8 pixels, 16x16 and 24x24 pixels.

The results of the experiments confirmed that resizing the resolution of the crop to 128x128 instead of 192x192 pixels improved the performance of the recognition rates and reduced the calculation time. This is normal, since image databases will have small sizes, this resizing will have a significant impact in the real world for large populations. The blocks that show the best performance are 8x8, 16x16 and 24x24 pixels, which is normal since we are using a local method for feature extraction. Large size subdivisions are not efficient for this kind of methods, henceforth we will only use the 8x8, 16x16 and 24x24 pixel blocks for the remaining experiments.

C. *Global Evaluation of Proposed Approach*

In this section, we will conduct our experiments to evaluate the proposed approach (Fig. 3). These experiments consist in exploiting the sizes of the blocks, which have shown their performance previously (24x24, 16x16 and 8x8). In this evaluation, we will use the approach with a score-level fusion, with city-block distance and Jeffrey divergence. We will use the Cumulative Matching Characteristics (CMC) curves for each fusion case to measure the identification accuracy. CMC curves demonstrate the ability of a recognition system to identify a given user in a set of data.

1) Results of fusion left palmprint 460nm and left palmprint 940

a) Results for block size 8x8 pixels: In Fig. 12, we notice that the curves resulting from the fusion of the 460nm + 940nm multi-spectral images with the classifiers with Jeffrey divergence and city-block are clearly higher than the other curves. The recognition rate obtained with the approach equal to 100%, this rate is higher than the best rate obtained for the systems studied with a single spectrum.

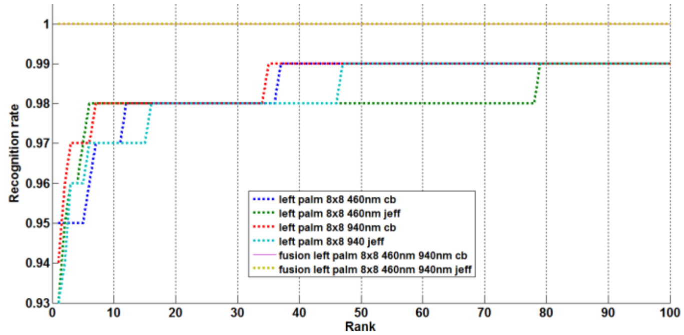


Fig. 12. CMC curve for fusion 460nm+940nm with 8x8 blocks.

b) Results for block size 16x16 pixels: In Fig. 13, we will report the same remark in the case of the block equal to 8x8. The recognition rate obtained with the approach is equal to 99%, this rate is higher than the best rate obtained for the systems studied previously.

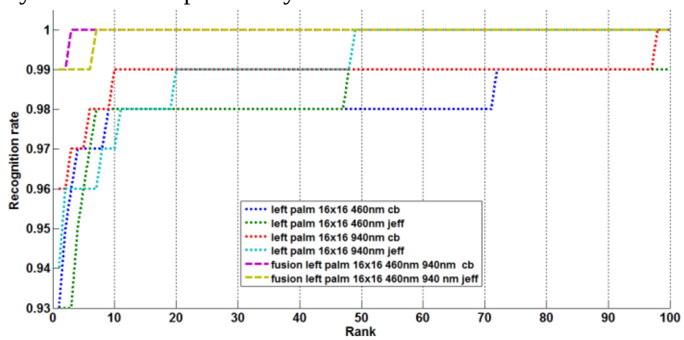


Fig. 13. CMC curve for fusion 460nm+940nm with 16x16 blocks.

c) Results for block size 24x24 pixels: In Fig. 14, we notice that the curve resulting from the fusion of the 460nm + 940nm multi-spectral images with the classifier based on city-block is superior to the other curves, but this is not the case of the fusion curve which uses the classifier based on Jeffrey's divergence. The recognition rate obtained with the approach equal to 100% using the classifier based on the city-block distance, this rate is higher than the best rate obtained for the systems studied with a single spectrum.

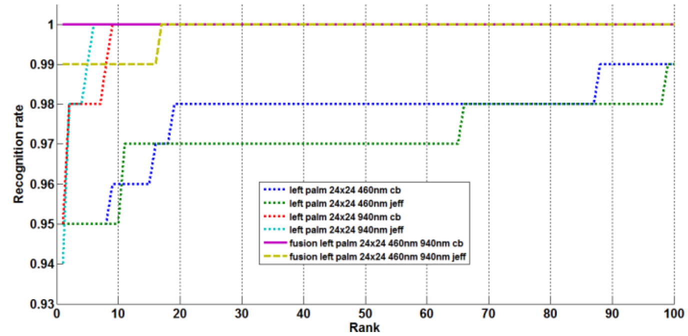


Fig. 14. CMC curve for fusion 460nm+940nm with 24x24 blocks.

TABLE XIV. RESULTS OF GLOBAL APPROACH FOR LEFT PALM

Recognition rate for left palm – fusion with spectrums 460 and 940		
Block size	***Jeffrey divergence***	***City-Block***
24x24	99%	**<u>100%</u>**
16x16	**<u>99%</u>**	**<u>99%</u>**
8x8	100%	**<u>100%</u>**

The Table XIV summarizes the performance obtained by the global approach used.

2) Results of fusion right palmprint 460nm and right palmprint 940

a) Results for block size 8x8 pixels: In the Fig. 15, we have perfect recognition rates for fusion equal to 100%.

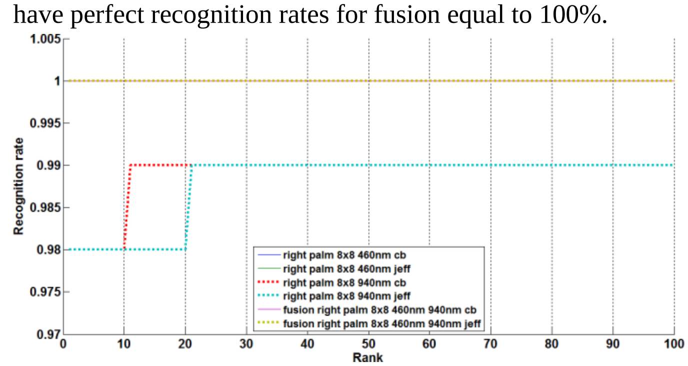


Fig. 15. CMC curve for fusion 460nm+940nm with 8x8 blocks.

b) Results for block size 16x16 pixels: In Fig. 16, we still have perfect recognition rates for fusion equal to 100%, which demonstrates the robustness and efficiency of the proposed approach.

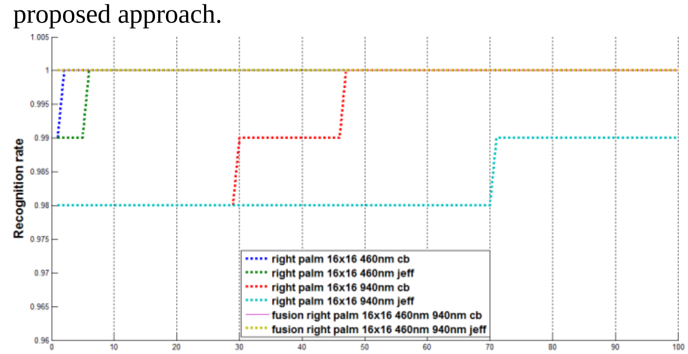


Fig. 16. CMC curve for fusion 460nm+940nm with 16x16 blocks.

c) Results for block size 24x24 pixels: In Fig. 17, we still notice that the curves resulting from the fusion of the 460nm + 940nm multi-spectral images with the classifiers with Jeffrey divergence and city-block are clearly higher than the other curves. The recognition rate obtained with the approach equal to 99%, this rate is higher than the best rate obtained for the systems studied with a single spectrum.

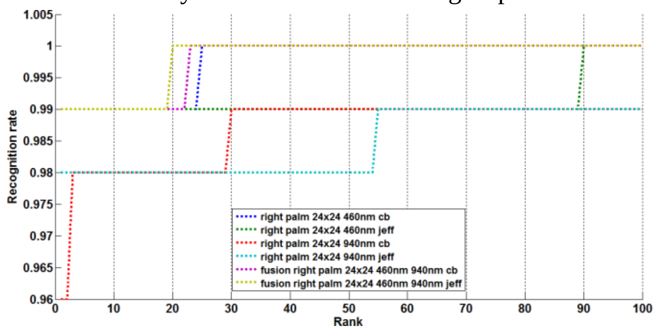


Fig. 17. CMC curve for fusion 460nm+940nm with 24x24 blocks.

TABLE XV. RESULTS OF GLOBAL APPROACH FOR RIGHT PALM

Recognition rate for right palm – fusion with spectrums 460 and 940		
Block size	***Jeffrey divergence***	***City-Block***
24x24	99%	**99%**
16x16	**100%**	**100%**
8x8	**100%**	**100%**

The Table XV summarizes the performance obtained by the global approach used for right palm.

The results demonstrated by the experiments that we see in the CMC curves, on the Tables XIV and XV and the comparison with others methods in Table XVI, reflect the effectiveness of the approach used to build a reliable and robust recognition system.

TABLE XVI. A COMPARISON OF RECOGNITION RATE OF THE PROPOSED APPROACH AND PREVIOUS METHODS

Method	Recognition rate
EigenPalm (EP) [55]	91,25%
FisherPalm (FP) 55]	92,32%
Gabor-based RCM (GRCM) [55]	96%
Enhanced GRCM (EGRCM) [55]	98%
Image level fusion by PCA [56]	95,17%
Quaternion PCA [56]	98,13%
Quaternion PCA+Quaternion DWT [56]	98,83%
PCA on HOG [57]	98,73%
KPCA on HOG [57]	98,737%
Proposed approach (using 24x24 bloc)	**99%**
LDA on HOG [57]	99,17%
KLDA on HOG [57]	99,17%
PCA on HOL [57]	99,73%
KPCA on HOL [57]	99,73%
Proposed approach (using 16x16 and 8x8 blocs)	**100%**

IV. CONCLUSION

In this work, multi-spectral images are used with different spectrums to build a secure system against tampering maneuvers. In this sense, two instances are employed with two spectrums that can be captured at the same time (but sequentially), to be used in a score-based fusion approach. The experimental results were conducted on the casia multispectral database after pre-processing with the Principal Based ICP method, subsequently the cropping used was improved for an optimal construction of the histograms, which made it possible to increase the performance and to reduce the calculation time, Finally, the conclusions drawn were deployed to guide the proposed approach, which clearly showed its effectiveness with rates varying between 99% and 100%. Thus, it is possible to conclude that this approach provides proven reliability and can be used for secure fingerprint recognition systems against fingerprint forgery fraud. This work aims at the possibilities of securing during the "capture" phase, nevertheless this securing will have to affect other aspects such as the base of the images. This perspective will be the subject of future work on securing biometric images with watermarking methods.

REFERENCES

[1] Buciu, I.; Gacsadi, A. Biometrics systems and technologies: A survey. Int. J. Comput. Commun. Control 2016, 11, 315–330.

[2] E. J. Kindt, “An introduction into the use of biometric technology,” in Privacy and Data Protection Issues of Biometric Applications: A Comparative Legal Analysis. Springer Netherlands, 2013, pp. 15-85.

[3] Thejaswini P, Srikantaswamy R S and Manjunatha A S, “Novel Adaptive Auto-Correction Technique for Enhanced Fingerprint Recognition” International Journal of Advanced Computer Science and Applications (IJACSA), 10(8), 2019.

[4] Mostofa, M.; Mohamadi, S.; Dawson, J.; Nasrabadi, N.M. Deep GAN-based cross-spectral cross-resolution iris recognition. IEEE Trans. Biomet. Behav. Identity Sci. 2021, 3, 443–463.

[5] A. Amraoui, M. Ait Kerroum and Y. Fakhri, “Multi-instance Finger Knuckle Print Recognition based on Fusion of Local Features” International Journal of Advanced Computer Science and Applications(IJACSA), 13(9), 2022.

[6] D. Prabakaran and R. Shyamala, “A Review On Performance Of Voice Feature Extraction Techniques,” 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), 2019, pp. 221-231.

[7] Mahalakshmi B S and Sheela S V, “A Novel Feature Extraction for Complementing Authentication in Hand-based Biometric” International Journal of Advanced Computer Science and Applications (IJACSA), 12(9), 2021.

[8] R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha, A.W. Senior, Guide to Biometrics: Selection and System Design, Springer, Boston, MA, 2003.

[9] Ammour, B.; Boubchir, L.; Bouden, T.; Ramdani, M. Face–iris multimodal biometric identification system. Electronics 2020, 9, 85.

[10] J. Priesnitz, R. Huesmann, C. Rathgeb, N. Buchmann, C. Busch “Mobile contactless fingerprint recognition: implementation, performance and usability aspects” Sensors, 22 (3) (2022), pp. 1-21.

[11] Y. Xu, G. Lu, Y. Lu, and D. Zhang, "High resolution fingerprint recognition using pore and edge descriptors," Pattern Recognition Letters, vol. 125, pp. 773-779, Jul. 2019.

[12] S. A. Shawkat, K. S. L. Al-badri, and A. I. Turki, "The new hand geometry system and automatic identification," Periodicals of Engineering and Natural Sciences, vol. 7, no. 3, pp. 996-1008, 2019.

[13] Angadi, S. and S. Hatture, Hand geometry based user identification using minimal edge connected hand image graph. IET Computer Vision, 2018. 12(5): p. 744-752.

[14] Muthukumar, A., Kavipriya, A, "A biometric system based on Gabor feature extraction with SVM classifier for Finger-Knuckle-print", Pattern Recognition Letters, 2019, 125, pp. 150– 156.

[15] Kim, J.; Oh, K.; Oh, B.S.; Lin, Z.; Toh, K.A. A line feature extraction method for finger-knuckle-print verification. Cognitive Computation 2019, 11, 50–70.

[16] Matkowski W.M., Chai T., Kong A.W.K., Palmprint recognition in uncontrolled and uncooperative environment, IEEE Transactions on Information Forensics and Security. 15 (2019) 1601–1615.

[17] S. Zhao and B. Zhang, "Deep discriminative representation for generic palmprint recognition," Pattern Recognition, vol. 98, p. 107071, 2020.

[18] A. Amraoui, Y. Fakhri and M. A. Kerroum, "Unimodal palmprint recognition system based on local features," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2017, pp. 1-5.

[19] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint spoof buster: Use of minutiae-centered patches," IEEE Trans. Information Forensics and Security, 2018.

[20] Chen, H.; Chen, Y.; Tian, X.; Jiang, R. "A Cascade Face Spoofing Detector Based on Face Anti-Spoofing R-CNN and Improved Retinex LBP". IEEE Access 2019, 7, 170116–170133.

[21] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," IEEE Transactions on Instrumentation and Measurement, vol. 59, no. 2, pp. 480–490, 2010.

[22] W. Jia, B. Zhang, J. Lu, Y. Zhu, Y. Zhao, W. Zuo, H. Ling, "Palmprint Recognition Based on Complete Direction Representation," IEEE Transaction on Image Processing, vol. 26, no. 9, pp. 4483–4498, 2017.

[23] A. Kong, D. Zhang, and M. Kamel, "Palmprint identification using feature-level fusion," Pattern Recognition, vol. 39, no. 3, pp. 478–487, 2006.

[24] W. Jia, D.-S. Huang, and D. Zhang, "Palmprint verification based on robust line orientation code," Pattern Recognition, vol. 41, no. 5, pp. 1521–1530, 2008.

[25] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, "A novel hierarchical approach for multispectral palmprint recognition," Neurocomputing, vol. 151, no. 1, pp. 511–521, 2015.

[26] L. Fei, Y. Xu, and D. Zhang, "Half-orientation extraction of palmprint features," Pattern Recognition Letters, vol. 69, pp. 35– 41, 2016.

[27] L. Fei, Y. Xu, W. Tang, and D. Zhang, "Double-orientation codeand nonlinear matching scheme for palmprint recognition," Pattern Recognition, vol. 49, pp. 89–101, 2016.

[28] R. Raghavendra, B. Dorizzi, A. Rao, and G. Hemantha Kumar, "Designing efficient fusion schemes for multimodal biometric systems using face and palmprint," Pattern Recognition, vol. 44, no. 5, pp. 1076–1088, 2011.

[29] S. C. Chen, H. G. Fu, and Y. Wang, "Application of improved graph theory image segmentation algorithm in tongue image segmentation," Computer Engineering and Applications, vol. 48, no. 5, pp. 201–203, 2012.

[30] El-Tarhouni, W., Boubchir, L., Elbendak, M., Bouridane, A. "Multispectral palmprint recognition using Pascal coefficients-based LBP and PHOG descriptors with random sampling," Neural Computing and Applications. 31, 593–603 (2019).

[31] Castro-Ortega, R., et al., "Zernike moment invariants for hand vein pattern description from raw biometric data," Journal of Electronic Imaging, 28(5), 053019 (2019).

[32] J. Gan and D. Zhou, "A Novel Method for Palmprint Recognition Based on Wavelet Transform," in Proceedings of the 2006 8th international Conference on Signal Processing, Guilin, China, November 2006.

[33] A. Giełczyk, M. Choraś, and R. Kozik, "Hybrid Feature Extraction for Palmprint-Based User Authentication", International Conference on High Performance Computing & Simulation (HPCS), pp. 629–633, 2018.

[34] X. Xu, Z. Guo, C. Song, and Y. Li, "Multispectral palmprint recognition using a quaternion matrix," Sensors, vol. 12, no. 4, pp. 4633–4647, 2012.

[35] Genovese A., Piuri V., Plataniotis K.N., Scotti F., "PalmNet: Gabor-PCA convolutional networks for touchless palmprint recognition," IEEE Transactions on Information Forensics and Security. 14 (2019) 3160–3174.

[36] J. P. Patil and C. S. Pawar, "Palmprint based Pattern Recognition Using Fast ICA," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 566-569.

[37] J. Cui and Y. Xu, "Three dimensional palmprint recognition using linear discriminant analysis method," Innovations in Bio-inspired Computing and Applications, International Conference, pp. 107 111, 2011.

[38] G. Lu, D. Zhang, and K. Wang, "Palmprint recognition using eigenpalms features," Pattern Recognition Letters, vol. 24, no. 9- 10, pp. 1463–1467, 2003.

[39] F. Du, P. Yu, H. Li, and L. Zhu, "Palmprint recognition using gabor feature-based bidirectional 2dlda," Communications in Computer and Information Science, vol. 159, no. 2, pp. 230–235, 2011.

[40] W. El-Tarhouni, L. Boubchir, N. Al-Maadeed, M. Elbendak, and A. Bouridane, "Multispectral palmprint recognition based on local binary pattern histogram fourier features and gabor filter," in Proceedings of the 6th European Workshop on Visual Information Processing, EUVIP 2016, fra, October 2016.

[41] X. Xu, L. Lu, X. Zhang, H. Lu, and W. Deng, "Multispectral palmprint recognition using multiclass projection extreme learning machine and digital shearlet transform," Neural Computing and Applications, vol. 27, no. 1, pp. 143–153, 2016.

[42] C. Han, H. Cheng, C. Lin, and K. Fan, "Personal authentication using palm-print features," Pattern Recognition, vol. 36, no. 2, pp. 371–381, 2003.

[43] X. Wu, K. Wang, and D. Zhang, "HMMs Based Palmprint Identification," in Biometric Authentication, vol. 3072 of Lecture Notes in Computer Science, pp. 775–781, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[44] Mukherjee M., Meenpal T., Kinship verification using compound local binary pattern and local feature discriminant analysis, in: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2019, pp. 1–7.

[45] Mohanty, F., Rup, S., Dash, B. (2018). An Improved CAD Framework for Digital Mammogram Classification Using Compound Local Binary Pattern and Chaotic Whale Optimization-Based Kernel Extreme Learning Machine. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science(), vol 11140. Springer, Cham.

[46] http://biometrics.idealtest.org/dbDetailForUser.do?id=6.

[47] Faisal Ahmed, Emam Hossain, A.S.M. Hossain Bari and ASM Shihavuddin, "Compound Local Binary Pattern (CLBP) for Robust Facial Expression Recognition," IEEE International Symposium on Computational Intelligence and Informatics, pp.391-395, Budapest, Hungary, 2011.

[48] Wei Li, L. Zhang, D. Zhang and Jingqi Yan, "Principal line based ICP alignment for palmprint verification," *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 1961-1964, doi: 10.1109/ICIP.2009.5413459.

[49] T. Ojala, M. Pietikainen and D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition, vol. 29, 1996.

[50] Stan Z. Li, ChunShui Zhao, XiangXin Zhu, Zhen Lei, 2D+3D Face Recognition by Fusion at Both Feature and Decision Levels, In Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Ges-tures. Beijing. Oct 16, 2005.

[51] S. Koda, A. Zeggada, F. Melgani and R. Nishii, "Spatial and Structured SVM for Multilabel Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5948-5960, Oct. 2018.

[52] Giannis Lantzanakis, Zina Mitraka, Nektarios Chrysoulakis, X-SVM: An Extension of C-SVM Algorithm for Classification of High-Resolution Satellite Imagery. IEEE Trans. Geosci. Remote. Sens. 59(5): 3805 3815 (2021).

[53] Q. Xiao, J. Lu, W. Jia and X. Liu, "Extracting Palmprint ROI From Whole Hand Image Using Straight Line Clusters," in *IEEE Access*, vol. 7, pp. 74327-74339, 2019.

[54] N. A. A. Khalid, M. I. Ahmad, T. H. Mandeel, M. N. M. Isa, R. A. R. Ahmad and M. Z. Nayef Al-Dabagh, "Palmprint ROI Cropping Based on Enhanced Correlation Coefficient Maximisation Algorithm," *2021 IEEE 11th International Conference on System Engineering and Technology (ICSET)*, 2021.

[55] Lu, Y Zhao, J Hu, "Enhanced Gabor-based region covariance matrices for palmprint recognition," IET Electronics letters, no. 17 : 880-881, 2009.

[56] Xu, Z Guo, C Song, Y Li, "Multispectral palmprint recognition using a quaternion matrix," Sensors 12, no. 4: 4633-4647, 2012.

[57] W Jia, RX Hu, YK Lei, Y Zhao and J. Gui. "Histogram of Oriented Lines for Palmprint Recognition," IEEE Transactions on Systems, Man, and Cybernetics: Systems, no. 3: 385-395, 2014.

Combining the Characteristics of the Buddha Statue from Photogrammetry and 3D Creation to Simulate the Combination of the Art of Creating Buddha Statue

Jirawat Sookkaew[1], Nakarin Chaikaew[2]
Department of Computer graphic and multimedia-School of Information and Communication Technology,
University of Phayao, Phayao, Thailand[1]
Department of Geographic Information System-School of Information and Communication Technology,
University of Phayao, Phayao, Thailand[2]

***Abstract*—The creation of this research was born out of interest in creating the art of carved sandstone into Buddha statue found in the area of Phayao Province, which was part of the Lanna Kingdom that prospered during the 19th-23rd Buddhist century. In the area of Phayao province, the sandstone Buddha was created which is an art and valuable artistic feature that has been shown until now. There are five categories of sand stone Phayao Buddha style that are studied and classified which have distinctive characteristics of each Buddha statue. Nowadays, traditional techniques for making Buddha statue are becoming less and less popular as Buddha statue made of sandstone are not as popular as before. Creating a Buddha statue from stone was a difficult and laborious process. Including craftsmen in carving began to decrease in number. In this research, a tool for collecting data on Buddha statue was used photogrammetry to store and process into 3D objects and use processes and techniques for creating 3D work that has been created, and simulating the Buddha statue by using the outstanding features of the Buddha statue collected as the main part in selecting the proportion of the Buddha statue to combine to form a new Buddha statue in 3D format by simulating the Buddha statue by such methods as a prototype to reproduce the appearance of Buddha statue for use in the creation of works of art that are an important part of history. It is also used to study the characteristics of Buddha statue in combination to create the characteristics, and a new way to preserve art by using technology to transfer and preserve these valuable works of art in another way.**

***Keywords*—3D art; 3D artifacts; creation; blending art; reconstruction artifacts**

I. Introduction

Creating Buddha statue data in a Phayao type Buddha statue entails using the exact qualities and structure of each portion of the Buddha statue collected to build a 3D digital file copy of the Buddha statue that references the Buddha statue's true structure. The right structure and proportions of the sandstone Buddha statue are made by the distinguishing qualities developed for imagination and visualization, which are obtained by analyzing the data from the genuine structure and proportions. The depth of Phayao craftsmen's art has condensed the incredible dimensions found in Phayao sandstone Buddha statue from the beginning through the evolution of countless forms into the unique qualities of each Phayao artisan's work. Beginning with the sandstone Buddha statues discovered in the Phayao an area, the Buddha statues are mixed with characteristics of Buddhist art from other regions of the kingdom and art by other Buddhist artists.

The Buddha statue is rather round and oval, as shown in the art of the Chiang Saen Buddha statue, which was influenced by the U-Thong Buddha statue. The two mouths of the Buddha statue that is frequently seen in the province of Phayao have a tiny and prominent nose; an oval face; a prominent nose; and a mouth with a boundary line. The researcher examined the creation of Buddha statue s by artists who make Buddha statues for temples and princes, as well as the development of Buddha statue s by local craftsmen in the region, in order to build a Buddha statue. The Buddha statues in this study are a mash-up of many types of art.

The art of sandstone carving in Phayao Buddha statue is not popular with people in today. Because the construction of Buddha statue with sandstone is a difficult process. Due to the factors of the material used to build, the weight of the stone and the procurement of stone. The factors of skilled craftsmen in stone carving are in accordance with the correct characteristics and expertise in production. In addition, the characteristics and guidelines for creating Buddha statue should be inherited and have unique prototypes. The problems and factors mentioned above are the problems that should be using digital tools and technology to assist in the creation of step-by-step and scientific and innovative tools will be used to benefit the art for the preservation and development of these values in next generation.

Instruments of technology and multimedia, such as data-collection science and computers, have been expanded and improved in a variety of ways. All suppliers are also working on new developments to make the systems more cost-effective and accurate, while also reducing current limitations. This is, to the best of our knowledge, the first study to survey various smart shelf innovations and application scenarios. [1] Even human life and activities such as living, consumption, and viewing must be integrally tied to and congruent with the technological environment. As a result, the components of the digital and technical world are astounding. Today, the world and human life are extremely important. These digitization initiatives in the human sciences prompted the crossdisciplinary discipline of Digital Humanities to investigate computer based semantic analysis and media processing [2] Computer technology has become a new tool for modern artists in terms of art. As a result, animators can devote more energy to creative work rather than the heavy and tedious work of traditional animation [3]. These technological advances have also helped to preserve and promote art and historical understanding. The method of use is the approved recording technique that makes it easier to digitize archeological data. Photogrammetry, which is image-based, creates three-dimensional images by triangulating points (in these case pixels) and using correlation techniques to reconstruct a subject's volume. This method enables us to establish trustworthy digital archives, providing digital reproductions of the shelters and their etched panels [4].

The aim of this research was to simulate the Buddha statue 3D by combining Buddhist art with technology of computer graphic, and 3D creation. In this research will use photogrammetry and the creative process of changing the 3D model to make a 3D model of the Buddha in the style of the ancient Phayao artists. The use of these digital tools and technologies for developing and generating concepts for creating such Buddha statue is another way to develop creative forms of Buddha statues and works. The art of sculpting sandstone Buddha statues today has many problems that can be solved with digital tools. It is also another method that helps to preserve the value of art and culture for further development and dissemination. With this information, art of Buddha statue can be hacked and changed, or they can be developed digitally for research in the future.

II. Literature Review

The main of research use photogrammetry and creative process of changing the 3D model to make a 3D model of the Buddha in the style of the ancient Phayao artists. Each of them has research and creative works to guide and suggest ideas for this research. Using 3D data for reconstruct 3D shapes from point clouds, despite the fact that many point cloud-based 3D reconstruction frameworks have been proposed in recent years, we discovered that the training loss is usually Chamfer Distance (CD), which assigns equal weights to all points within the point clouds. However, for human visual perception, edges and corners are more important than flat points. According to research, the proposed framework can focus on edges/corners and produce more appealing results. The visual-enhanced approach proposed here can also be used to generate high-quality point clouds in a 3D generative model [5].

Complex deep neural networks are required to grasp and disentangle spatial transformation and image data characteristics inherent in training data in order to effectively recreate 3D images which limit the use of data-driven techniques in many practical applications. That using existing geometric information from imaging technology improves the process of unfolding incoming sensor data into 3D space in the context of computed tomography image reconstruction [6].

In the computer vision and graphic design communities, learning to create three-dimensional (3D) point clouds is becoming increasingly popular. Numerous solutions have been proposed to address the issue of producing 3D point clouds. Because a 3D object can be presented in a variety of formats, such as point cloud, voxel, mesh, and implicit field, these approaches can be classified according to their output format. As a result, manually rebuilding point clouds is the most common method of creating 3D objects. Most techniques start with an encoder that extracts a latent feature from the source images or 3D point cloud, followed by a decoder that maps the latent feature to the ground truth of the 3D point cloud [7].

Visual improvements of 3D point cloud rebuilding from a single perspective is aimed at the construction of 3D Point cloud objects, a type of 3D object creation based on object data collection that is quite common nowadays with numerous types of 3D structure processing and integration procedures involving point cloud., voxel, mesh, and implicit fields. These methods assess processes for extracting 3D structure information from entire 3D objects using point cloud-based processes [7].

The aim of the research was to develop the most effective methodology for virtual 3D reconstruction of damaged archaeological sites, which included construction or building structures, for the sake of cultural heritage conservation. Based on the revised point cloud and excluding affected areas, a surface depicting the terrain shape of the archaeological site prior to mechanical damage was interpolated. Using textured solids and surfaces representing walls, connection channels, and vaults, a rigorous virtual reconstruction of the architectural complex's original condition was carried out using both measured and interpolated surfaces. [8] The goal of this study is to recreate 3D models of ancient buildings that have been severely damaged by using high-quality map data to produce 3D skeletal structures for the preservation of history and culture [8].

In certain circumstances, digital photogrammetry and 3D reconstruction can aid archaeological excavation by reducing time without sacrificing information and even generating unique data, as well as making the site accessible to the public. Using this method, in the research aims to support studies of relative chronology based on the observation of structure textures and the statistical processing of block measurements and presents the results in the case of the St. Maria Veterana complex (Triggiano, southern Italy), a very articulate archaeological site of unknown dating, whose stratification and constructive chronological phases have been very confusing thus far. The statistical analysis of the dimensional dataset revealed indicative correlations between the various rooms of the archaeological site, and thus suggested valuable information on building techniques through comparison with results of lithological and textural observations of areas [9].

The use of photogrammetry technology creates a scalable surveying option that can be deployed faster and at a lower cost than an airplane survey. These rapidly advancing technologies open up exciting new avenues for archaeological inquiry and methodology, with the potential to reach a broader audience by engaging with topics and disciplines outside of archaeology's traditional domains [10].

The technique employed was centered on the use of point cloud 3D modeling to obtain virtual objects of a Roman cornice from the Castulo Archaeological Site (in Spain), followed by the use of those models for material restoration of the losses via 3D printing of the piece for reintegration. The process used focused on using photogrammetric 3D modeling to obtain virtual models of a Roman cornice from the Castulo Archaeological Site (in Spain), followed by the use of those models for material restoration of the losses via 3D printing of the piece for reintegration [11].

Photogrammetry procedures and processes for the digitalization and reconstruction of three-dimensional works in archaeological research and excavations can reduce study time and make knowledge accessible to the public through simulation approaches. The 3D model was created from an archaeological site in Spain using the 3D object modeling method, 3D manipulation of the specimen to generate the updated 3D structure, and 3D printing to validate the modified 3D object data [11].

With a good setup and good quality cameras with a standard lens, photogrammetry-based digitization can be a powerful tool for documenting, conserving, analyzing, and making large swaths of archaeological collections and landscapes available to the public and fellow researchers. Archaeologists and 3D specialists believe that before even taking the object out of the box, one should question their motivations for digitization and ensure the best reproduction quality possible [12].

The use of photogrammetry has aided in the virtual reconstruction of the work for the restitution of the statue, but it has also provided a useful database in 2D and 3D for the documentation of the locations. The combination of these data with historical sources improves the approach for the proposal of a scientific approach useful in restitution work, which is still hypothetical and a topic for debate in the scientific community [13]. In creating a 3D model a method of simulating a similar structure can be used to complete the construction of the object.

In fact, the use of photogrammetry allows for the creation of a detailed virtual model, which has two advantages: on the one hand, it provides valid and, above all, reliable support to historians, archaeologists, and restorers, and on the other, it allows for the dissemination of Cultural Heritage artifacts to a wider audience [14].

III. METHOD

This application, which runs on the Windows 10 operating system, allows the researcher to create and improve 3D objects. The selection of a 3D acquisition method is an important step when designing a digitization plan, and it is highly correlated with requirements such as the purpose of 3D digitization, the final use of 3D models, as well as other aspects of a digitization project such as budget, duration, and available personnel experience [15].

a) Agisoft metashape - Photogrammetry Image processing applications and tools. (see Fig. 2)

b) Autodesk maya 2020 - customization retopology process editing, and improvement tools for 3D models.

This experiment employs both surveys and creative experiments, which are data collection and processing in both the science of producing ancient Buddhist art and the creation and evaluation of multimedia technological tools.

Analysis of the different parts of Phayao Buddha art by looking at and rating the different types of sandstone Buddha statues and picking out the features that make each Buddha statue unique. Fig. 1 explains all the steps involved in the research process.

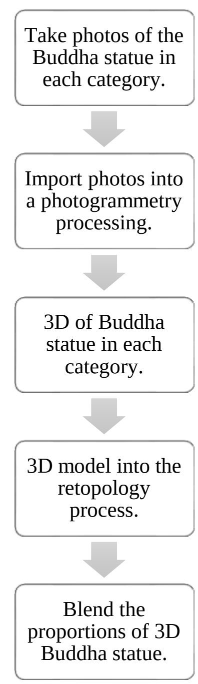


Fig. 1. Research process

1) Take a picture of the Buddha statue in each body by taking a picture around the Buddha statue . This photo recording uses the recording with the use of recording to obtain a digital image for processing. Photogrammetry is the next step. Data collection of Buddha statues digitally utilizing image processing techniques such as photogrammetry, which is a digital data recording procedure that converts images into 3D images. This will make it easier to get the desired surface proportions and surface attributes for the material.

2) Importing 3D image processing data in the form of photogrammetry. Retopology is the technique of decreasing the features of a 3D model to the lowest resolution possible while maintaining texturing and rendering equivalent to a high resolution actual image. This will make image processing in a computer program easier and faster, as well as enable the workpiece to be used in a wider range of applications, which is a crucial component in making 3D things that users of 3D tools grasp. It is important for 3D tools and applications.

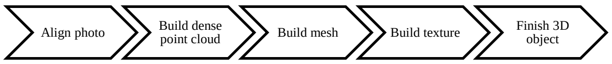


Fig. 2. Photogrammetry processing in agisoft metashape

3) But the workpiece obtained from the above process will have the same result as the real workpiece but with high polygon (high detail), making it difficult to use the 3D workpiece for editing or adjusting. Retopology is the use of the High polygon model to rearrange polygons with fewer polygons to make it easier to In order to create a new Low polygon piece from the original High polygon with the most proportions and appearance equivalent to the original workpiece, the surface treatment technique or Texture UV of the workpiece was also used. The 3D rendering is very close and looks the most similar to High polygon 3D.

4) The Buddha statue s that have been carved into exceptional proportions from the 3D models in each piece can be brought to life by choosing the proportions of the Buddha statue s that have been sketched in a rough outline before joining the Buddha statue s in the next stage.

*A*nalysis of the different parts of Phayao Buddha art by looking at and rating the different types of sandstone Buddha statue s and picking out the features that make each Buddha statue unique. Exploration of alternative means of meaning making in archaeology based on non-linear narratives, three-dimensional perspective, and virtual reconstruction is now possible thanks to 3D visualization and digital archaeological methods. [16]

5) Creating a connection between each statue by using editing software and adjusting 3D pieces by emphasizing the connection to have the least amount of adjustment because it may affect other parts of the Buddha statue and collecting details, such as adjusting and adding texture to the surface for beauty.

By evaluating the structure and proportions of the plinth elements, parts with characteristics that can be connected to each other are used. The parts of the statue will be adjusted to a minimum in order to maintain the original art that has been collected from data as much as possible.

Since then, computer technology has advanced significantly. 3D images created with the right software are truly realistic and look like photographs. They are used for popularization, education, and research. Scientists have recognized the importance of digital 3D imaging as a tool for testing hypotheses and communicating research findings to the public. [17]

IV. Analysis of Proportions and Characteristics in Assembly and Use

Characteristics of the Buddha, the Buddha statue that have been selected and used to create this Buddha statue refer to the concept of creating a new 3D Buddha statue while retaining the distinctive characteristics of Buddha statue in the Phayao art style with a combination of characteristics and widespread beliefs. Therefore, the Buddha statue proportions in the obtained Buddha statue may have different characteristics and feature combinations that are different from those commonly seen in the art of sandstone Buddha statue. But the characteristics of the proportions necessary to take into account the proportions and composition of the sandstone Buddha statue of Phayao are as follows.

1) The appearance of the head is oval, round, and flat.

2) Curved eyebrows collide into a winged shape.

3) Curved and prominent nose shape.

4) The curved mouth resembles a bird's wing.

5) Hair granules are round or square pyramidal or without hair granules.

6) The hair pacifier looks like a lotus flower which comes in both small and large sizes.

7) Body parts are plump, with large breasts protruding strongly and prominently. [18]

The above characteristics are a clear feature that will be an indicator of the creation of sandstone Buddha statue in this research process to be creative as the correct characteristics of traditional art.

V. Result

Using digital tools and methods to convey or convey the meaning of antiques and to interpret or create new ones. This is a simple and convenient method that can be extended more widely.

A touch-object, in the form of a 3D printed facial reconstruction, extends an otherwise visual experience to the visually-impaired and encourages 'embodied knowledge-making' among all visitors. [19]

The proportions and compositions of the Buddha statue s used to assemble to create a new Buddha statue in 3D are taken from the highlights of sandstone Buddha statue of Phayao that have been recorded, and study with the components shown as follows.

1) A clear base is a type of base that is extensively used in the construction of sandstone Buddha statue in Phayao. Because sandstone Buddha statues are manufactured from a single piece of sandstone, building a sandstone Buddha statue with a clean foundation is the method. Considering the techniques used by sandstone sculptors at the time, flat base styles are classified into three types: flat base, short base, tall base, and tidy hexagon base.

2) The appearance of the Buddha statue 's sitting position is a strong influence from the Chiang Saen Buddha statue , which was influenced by the sandstone Buddha statue while making the first Buddha statue in the Phayao craftsman. The diamond meditation pattern seen in the building of the Buddha statue of Chiang Saen Sing 1st Characteristics of the Buddha statue originated in Chiang Saen Kingdom which is a vast ancient kingdom in the north) and also found in the Sukhothai Kingdom art Buddha statue was used in the fabrication. It has a huge, square foot form and is the same size as the toes.

3) On a wavy crown, the head has a Buddha-shaped mouth at the margin of the mouth. The corners of both sides of the lips are swept up; the eye is bent upward to meet the mouth; and the crown is shaped like a square, comparable to a pyramid.

4) His physique is massive. His arms, influenced by the Chiang Saen Singha, 1st Buddha statue, are huge, plump, aligned, and the same size. Above the tan is a short robe (breast).

A 3D visual used in archaeology to reconstruct cultural heritage objects is a software image that reflects the knowledge gained from an analysis of archaeological excavation sources. Spatial models are built using existing documentation and literature. 3D reconstructions are created by analysing and interpreting photographs taken during excavations. [17] In Fig. 3 creating a Buddha structure by using the structure of the Buddha statue 3D from process of photogrammetry, then selecting the proportions of each part of the statue to be adjusted and corrected using the tools from the 3D program, which uses scaling to be appropriate to be able to connect seamlessly.

Fig. 3. Characteristics of the components of the Buddha statue in each part to bring together

The combination of the proportions of each piece from the 3D creation of the Buddha 3D statue from Fig. 4 in each part that creates a smooth connection and is in accordance with the typical characteristics of creating a Buddha statue carved in the Phayao style.

Fig. 4. 3D Buddha statue created by blending proportions in each art style

VI. Discussion And Conclusion

The result of the development and experimenting is a digital work in the form of a 3D piece of Buddha statue that reveals the qualities of a sandstone Buddha statue carved uniquely because the evidence and art of sandstone sculptures have degraded and cracked significantly due to both eroding material and weather. The application of such strategies or technologies to foster honesty and connect thoughts and tales from the technological process is thus extremely significant and required in the manufacturing and fabrication of components to assist in the completion and operation of this work. In this sense, the selection of qualities is critical. The Buddha statue and its dimensions should be as similar as feasible so that they can be related to each other. Too much modification in the proportions of the Buddha statue may impact the features of the Buddha statue in other proportions the fullness of the Buddha statue reproduction while applying the creative concept of the Phayao craftsman in each proportion, which is uncommon nowadays. The prospect of extending and developing the concept of Buddhist art or features from simulations, or it can be used to compare simulated Buddha statue in order to estimate the proportions of shattered Buddha statue. The use of proportions that are the overall characteristics of the unique characteristics of the sandstone Buddha statue of the ancient Phayao Buddha art statue used to assemble the Buddha statue is also the researcher's concept that brings out the outstanding features of the Buddha statue of Phayao craftsmen in each era. Assembled into a Buddha statue without finding such a Buddha statue in the creation because the development of the art of creating a Buddha statue in Phayao sandstone stopped developing the guideline, when the Lanna Kingdom and Phayao's prosperity stopped during the decline of the Lanna Empire and the city of Phayao, was abandoned. The science of developing these arts has also stopped developing. This creation is another concept that brings the remaining works of art to be applied and assembled into a new piece without destroying the original prototype. In various movements to assess the probability of previous features, research and simulation, and perfect Buddha statue or antiquities that desire to be examined and postulated in connected and associated sciences diverse works of art in other multimedia domains via the use of simulation technology. In creating a 3D model of Buddha statue, this piece helps guide the construction and production of Buddha statue or is used to create and support the creation of art to occur more conveniently from technological and multi-tools media. It also reduces problems and steps that hinder the creation and production of work pieces. It may be used to compare and evaluate Buddha statue in order to estimate and evaluate Buddha because it will disrupt the connection of the proportions of the simulated Buddha statue. According to the qualities of Buddhist art, the proportions of the presentation may not be comprehensive and attractive. However, in this creative effort, the image utilized has ties to each component of the Buddha statue. The look that was born in this exhibition demonstrates.

References

[1] M. Kellermayr-Scheucher, L. Hörandner, and P. Brandtner, “Digitalization at the Point-of-sale in grocery retail - state of the art of smart shelf technology and application scenarios,” Procedia Comput. Sci., vol. 196, pp. 77–84, 2022.

[2] M. Deligiorgi et al., “A 3D digitisation workflow for architecture-specific annotation of built heritage,” J. Archaeol. Sci. Rep., vol. 37, no. 102787, p. 102787, 2021.

[3] Z. Hu and L. Liu, “Research on the application of virtual reality technology in 3D animation creation,” Optik (Stuttg.), vol. 272, no. 170274, p. 170274, 2023.

[4] E. Lesvignes et al., “Using digital techniques to document prehistoric rock art: first approaches on the engraved panels of the Paris Basin

shelters," Digit. appl. archaeol. cult. herit., vol. 15, no. e00122, p. e00122, 2019.

[5] G. Ping, M. Abolfazli Esfahani, J. Chen, and H. Wang, "Visual enhancement of single-view 3D point cloud reconstruction," Comput. Graph., vol. 102, pp. 112–119, 2022.

[6] L. Shen, W. Zhao, D. Capaldi, J. Pauly, and L. Xing, "A geometry-informed deep learning framework for ultra-sparse 3D tomographic image reconstruction," Comput. Biol. Med., vol. 148, no. 105710, p. 105710, 2022.

[7] Q. Yu, C. Yang, and H. Wei, "Part-Wise AtlasNet for 3D point cloud reconstruction from a single image," Knowl. Based Syst., vol. 242, no. 108395, p. 108395, 2022.

[8] F. Carvajal-Ramírez, A. D. Navarro-Ortega, F. Agüera-Vega, P. Martínez-Carricondo, and F. Mancini, "Virtual reconstruction of damaged archaeological sites based on Unmanned Aerial Vehicle Photogrammetry and 3D modelling. Study case of a southeastern Iberia production area in the Bronze Age," Measurement (Lond.), vol. 136, pp. 225–236, 2019.

[9] G. Fioretti, C. Campobasso, and S. Capotorto, "Digital photogrammetry as tool for mensiochronological analysis: The case of st. Maria Veterana archaeological site (Triggiano, Italy)," Digit. appl. archaeol. cult. herit., vol. 19, no. e00158, p. e00158, 2020.

[10] J. M. Vilbig, V. Sagan, and C. Bodine, "Archaeological surveying with airborne LiDAR and UAV photogrammetry: A comparative analysis at Cahokia Mounds," J. Archaeol. Sci. Rep., vol. 33, no. 102509, p. 102509, 2020.

[11] M. Higueras, A. I. Calero, and F. J. Collado-Montero, "Digital 3D modeling using photogrammetry and 3D printing applied to the restoration of a Hispano-Roman architectural ornament," Digit. appl. archaeol. cult. herit., vol. 20, no. e00179, p. e00179, 2021.

[12] A. Bisson-Larrivée and J.-B. LeMoine, "Photogrammetry and the impact of camera placement and angular intervals between images on model reconstruction," Digit. appl. archaeol. cult. herit., vol. 26, no. e00224, p. e00224, 2022.

[13] B. Bennoui-Ladraa, Y. Chennaoui, and H. Ainouche, "The virtual archaeology and interpretative process: Case study of the virtual reconstitution of a Hercules marble statue from the nameless temple of Tipasa," Digit. appl. archaeol. cult. herit., vol. 19, no. e00163, p. e00163, 2020.

[14] L. T. De Paolis, S. Chiarello, C. Gatto, S. Liaci, and V. De Luca, "Virtual reality for the enhancement of cultural tangible and intangible heritage: The case study of the Castle of Corsano," Digit. appl. archaeol. cult. herit., vol. 27, no. e00238, p. e00238, 2022.

[15] A. Koutsoudis, B. Vidmar, and F. Arnaoutoglou, "Performance evaluation of a multi-image 3D reconstruction software on a low-feature artefact," J. Archaeol. Sci., vol. 40, no. 12, pp. 4450–4456, 2013.

[16] N. Lercari, "3D visualization and reflexive archaeology: A virtual reconstruction of Çatalhöyük history houses," Digit. appl. archaeol. cult. herit., vol. 6, pp. 10–17, 2017.

[17] M. Markiewicz, "Photography vs. visualisation. Technical images in archaeological research," Digit. appl. archaeol. cult. herit., vol. 24, no. e00213, p. e00213, 2022.

[18] S. Saisingha, The Payao sandstone Buddha statue . Phayao,Thailand, 2020.

[19] K. Smith, M. Roughley, S. Harris, C. Wilkinson, and E. Palmer, "From Ta-Kesh to Ta-Kush: The affordances of digital, haptic visualisation for heritage accessibility," Digit. appl. archaeol. cult. herit., vol. 19, no. e00159, p. e00159, 2020.

Hardware Trojan Detection based on Testability Measures in Gate Level Netlists using Machine Learning

Thejaswini P, Anu H, Aravind H S, D Mahesh Kumar, Syed Asif, Thirumalesh B, Pooja C A, Pavan G R
Department of Electronics and Communication Engineering,
JSS Academy of Technical Education, Bangalore-560060, Karnataka, India

***Abstract*—Modern integrated circuit design manufacturing involves outsourcing intellectual property to third-party vendors to cut down on overall cost. Since there is a partial surrender of control, these third-party vendors may introduce malicious circuit commonly known as Hardware Trojan into the system in such a way that it goes undetected by the end-users' default security measures. Therefore, to mitigate the threat of functionality change caused by the Trojan, a technique is proposed based on the testability measures in gate level netlists using Machine Learning. The proposed technique detects the presence of Trojan from the gate-level description of nodes using controllability and observability values. Various Machine Learning models are implemented to classify the nodes as Trojan infected and non-infected. The efficiency of linear discriminant analysis obtains an accuracy of 92.85 %, precision of 99.9 %, recall of 80%, and F1 score of 88.8% with a latency of around 0.9 ms.**

***Keywords*—Hardware trojan; machine learning; controllability; observability; detection and mitigation**

Abbreviations	
A	Accuracy
P	Precision
R	Recall
F	F1–Score
TP	True Positive
TN	True Negative

I. Introduction

Hardware Trojans are modern-day system attacks that will cause prominent damage to the IC or system in numerous ways. Though the software is considered to be vulnerable, the underlying hardware is generally considered to be safe. However, research has shown that, due to the complex nature of the design, fabrication process, rapid prototype development, and distribution of the final product, new sources of attack are prominent [1]. Speeding up the development cycle and lowering R&D costs is the main goal of most manufacturing companies because the estimated R&D cost is up to $5 billion. Most companies cannot afford to invest such a huge amount from start to finish. So, companies frequently outsource fabrication to a third-party foundry, buy IP cores from third-party suppliers, and employ EDA tools from third-party vendors. Third-party suppliers can readily enter such a model, and the supply chain is currently deemed vulnerable to assaults like Hardware Trojan insertion, reverse engineering, IP theft, IC tampering, IC cloning, and IC overproduction, among others. Hardware Trojans are arguably the most concerning of them, and they have attracted a lot of attention. It will eventually change the functionality of the system and the user will be unable to take any action against it [2].

Hardware Trojans can be defined as malicious components introduced during the design, manufacturing, fabrication, testing, or development phase [3]. Once introduced they can be activated anytime, anywhere, and according to the attacker's interest. The activation mechanism divides Hardware Trojan into two groups: always on and triggered. Always-on Trojans are active as soon as the systems or designs are turned on, whereas triggered Trojans require the activation of some form of condition. A Hardware Trojan circuit is generally designed in two parts; a condition-based circuit (trigger) and an operation circuit (payload) which is interconnected via trigger net [4]. The Trojan will be triggered and activated when the predefined criteria is satisfied. The most dangerous part is that they can be inserted anywhere in the circuit, be it processor, IC power grid, IO, and there is no way to immediately know the source of the threat. By the time it is discovered and neutralized it may be too late as it can change the functionality of the circuit. They can also downgrade its performance, leak sensitive information and finally cause a Denial-of-Service attack [3]. Therefore, to find a solution for many such attacks, various researches are being conducted. Amongst them, the logic function test (LFT) is a traditional method [5]. Most of the existing methods of LFT do not effectively activate any potential hidden Trojans. Researchers use side-channel analysis to detect Hardware Trojans by modelling and analysing electromagnetic information generated during chip operation [6]. Traditional side-channel analyses' effect isn't sufficient, according to researchers [7], due to low sensitivity detection rates for big process fluctuations and a small Trojan footprint. Combining principal component analysis and linear discriminant analysis to analyse chip power is effective in evaluating Trojan detection accuracy [8]. Machine learning techniques for malware detection are the most successful state-of-the-art research topic because of their ability to keep up with malware evolution. They concentrate mainly on two areas, one is feature extraction, and the other is dimensionality reduction. Support Vector Machine algorithm seems to be the go-to algorithm for detection in multiple cases [9]. However other advanced algorithms are also being explored to trigger and detect a Trojan. Triggering a Trojan has an impact on the system's power usage as well as the circuits' delay. This behavior is extensively used to study the power and delay characteristics of hardware Trojans in order to detect them[10].

Even though there is a possibility to detect the Trojan using the different techniques as stated so far, there is still scope for optimization. The contributions made by this paper are as follows:

1) We propose the use of gate-level descriptions of nodes using controllability and observability values to generate a dataset to detect the presence of Hardware Trojan.

2) Machine Learning techniques are used to detect the presence of Trojan.

3) Benchmark circuits-C17, C3540 and C432 are considered to validate our proposed technique.

4) We experimentally prove that our proposed model has an area and power reduction up to 75% and 80% respectively in Trojan free circuits. This performs better in comparison to other state-of-art techniques.

The rest of this paper is organized as follows. Section II and III analyze the related works and motivation. Section IV shows the proposed scheme. The experimental results and discussion are presented in Section V. Section VI shows the conclusion.

II. Related Works

Various techniques ranging from score-based classification [6] for identifying Hardware-Trojan-free or Hardware-Trojan-infected circuits without using golden model-based approach to deep learning techniques are researched. The side-channel analysis and detection method [11], uses dimensional reduction to detect HTs. This causes the loss of important feature information of Hardware Trojans after the principal component analysis or filtering process. To solve this problem, a Hardware Trojan detection technology is proposed based on Extreme Learning Machine (ELM), which can completely retain important information without any inaccuracies caused by modelling. Results show that detecting the Hardware Trojans only used about 0.15% of resources. The accuracy rate was about 90%.

In terms of router looping, traffic diversion, or core spoofing, a trojan attack corrupts the router packet [9] by changing the destination address. As a solution, SVM is used to increase detection accuracy. According to the estimates, the suggested security solution architecture achieves a 93 percent accuracy for seizure detection applications in 4.8μS.

Detection of Trojan using gate-level netlist based on observability and controllability analysis [4] produce sufficient results. When this technique was used on numerous trojans, the findings reveal that even in the worst situation, all Trojans are discovered effectively with zero false positive and negative rates in less than 14 seconds.

Using a specific gate-level netlist that specifies the Trojan nets in full, the review paper [12] covers extracting 51 gate-level Trojan features. The usage of an ensemble- based random forest classifier results in a true positive and true negative rate of 100 percent.

Implementation of SVM using five-dimensional vectors [13] classifies all the nets in an unknown netlist into Trojan affected and normal ones using the Trust-HUB benchmark. Not only SVM but various other machine learning algorithms like Decision Tree, K Nearest Neighbour [14] is applied to identify the Trojan. Further to increase classification accuracies, Hardware Trojans are discretized based on their dominant attributes. The results show that both Machine Learning algorithms when trained on a given dataset perform well. DT and KNN models can accurately predict about 83% of the test data.

In addition, existing deep neural networks security studies are not extensively conducted at these software algorithm levels [15] and the more realistic attacks by third-party vendors are not explored. So, it is successful in demonstrating how an attack is possible. Experiments reveal that the approach can quickly generate and activate a variety of Trojan attacks that can readily overcome existing defenses. This is very important in formulating a solution to the attack.

III. Motivation

The presence of Hardware Trojans (HTs) in circuits causes malfunction on various scales depending on the type of Trojan attack. Amongst the numerous existing state-of- art techniques, the FANCI [16] technique uses a coverage-like approach where it does not require access to any verification stimuli. But the attackers like third-party vendors being aware of this can make the Trojan look benign. The third-party inputs from on-chip IPs can be scrambled to suppress the Trojan triggers. But this would not work for analog triggers. So, the Side-channel techniques can be used to unmask Trojans injected by third parties [17], but this method is unusable until IC is manufactured and inside the supply chain. As a solution, the use of formal proofs enables to development of trusted IPs. However, this assumes that proof is sufficient to rule out injected Trojans and that IP specifications are known. Therefore, there is a need for better-advanced techniques to counter novel malware attacks. One such technique is Machine Learning (ML) [18].

The rise of Machine Learning has profound implications for many industries, including cyber security. ML-based anti-malware tools are generally believed to provide better detection of modern malware attacks and improve scanning methods. Machine Learning algorithms perform better against unforeseen threats as they can be trained to handle unknown potential threats. This is a major advantage over other techniques. They yield more accurate, efficient solutions. As a result, employing ML techniques proves to be beneficial.

IV. Methodology

The proposed Trojan detection technique is a combinational circuit [19] for avoiding unintended malicious activity is as shown in Fig. 1. Controllability and Observability are the two parameters considered for the detection and classification of Trojan in the given circuit [20].

To implement the proposed methodology, let us consider ISCAS benchmark circuits. These are implemented using Verilog. We have designed a Trojan threat model and applied it to ISCAS benchmark circuits. Then we generate a Gate Level netlist. Using Python and gate-level netlist, we obtain values of two parameters, Controllability and Observability to detect the presence of Trojan in the circuit. Hence, we create a dataset using these values as input. The model is trained and tested by applying different Machine Learning techniques. The desired output is obtained and verified using functional verification and overhead analysis.

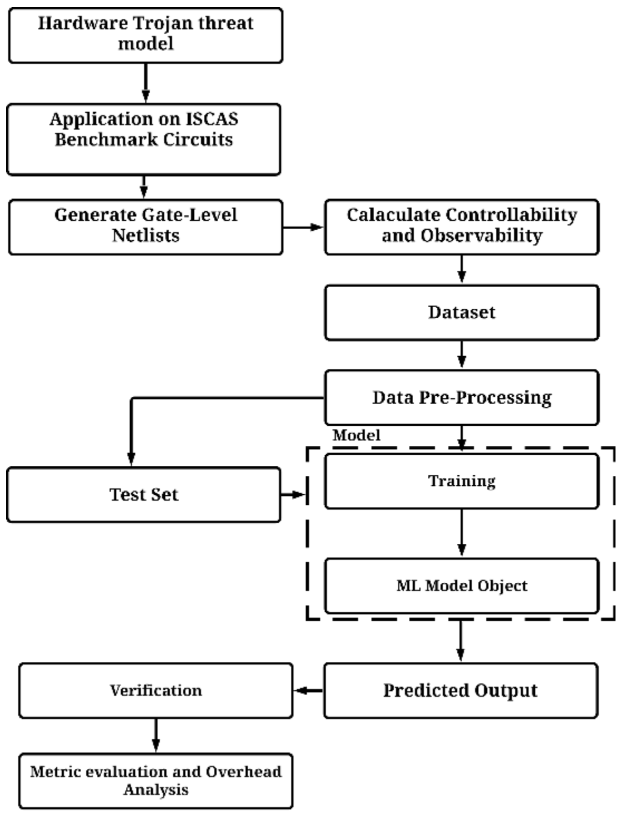


Fig. 1. Block diagram of proposed methodology

A. Design of Trojan Threat Model

A trojan is a unique circuit that performs specific malicious activity [21]. Trojan considered in our proposed work will change the functionality of the circuit. Fig. 2 shows the Trojan threat model designed for our work. Here, 4-bit Linear Feedback Shift Register (LFSR) is used to trigger the Trojan and NOT gate act as payload. One input to the comparator is from LFSR and the other input is a random number generated by the attacker. The output of the comparator is connected to the select line of MUX. If the output of the comparator is one, Trojan is triggered and vice versa.

B. Benchmark Circuit Selection

The proposed method is implemented on a standard ISCAS benchmark circuit [22]. The ISCAS '85 benchmark circuits are combinational circuits that are used by researchers as the basis for performing analysis and comparing results. C17-NAND only circuit, C432-a 27 channel interrupt, and C3540–an 8-bit ALU are the three circuits considered in our work as depicted in Fig. 3, Fig. 4 and Fig. 5.

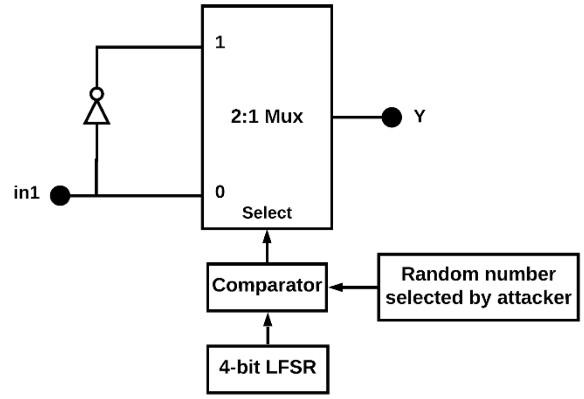


Fig. 2. Design of trojan threat model

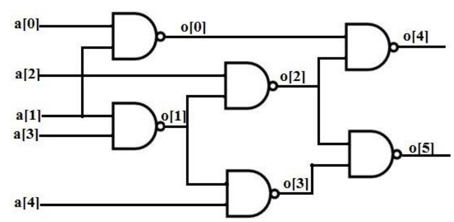


Fig. 3. C17 (NAND circuit)

C. Gate Level Netlist

Gate Level Netlist contains information regarding the logical connectivity of all standard cells and macros [23]. Gate level netlist of C17, C432, and C3540 is generated by synthesizing these circuits using the Cadence tool at 90nm technology. These gate-level netlists are used for calculating observability and controllability values.

D. Controllability and Observability Calculation

The presence of Trojans in the circuits is analyzed by calculating the controllability and observability values.

1) Controllability analysis: Nets with poor testability are identified using combinational controllability. The levels of controllability range from 1 to infinity. Because the possibility of detecting such a node is very low, and controlling that particular node is quite difficult, nodes with high controllability (CC) ratings are more susceptible to having a trojan inserted. All signals from primary inputs to primary outputs have their controllability values determined first. The circuit is initially levelled by giving each gate a level value [24]. Each gate's output controllability is then calculated. Controllability (CC) in general is expressed as:

$$CC(i) = \sqrt{CC0(i)^2 + CC1(i)^2} \quad (1)$$

where, CC0(i) is Combinational Controllability 0 and CC1(i) is Combinational Controllability 1

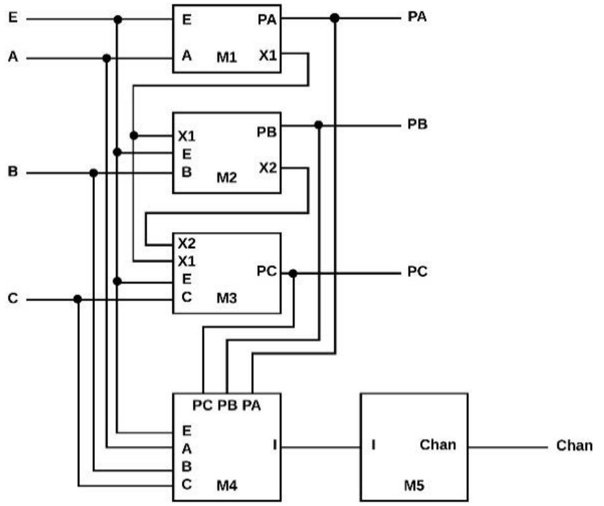


Fig. 4. C432 (27 channel interrupt controller circuit)

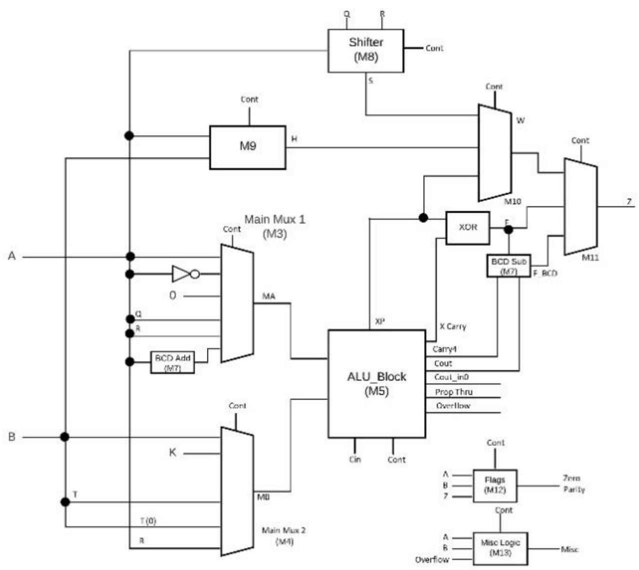


Fig. 5. C3540 (8-Bit ALU circuit)

2) Observability analysis: Observability is a measure of the ease (or difficulty) with which one can calculate the signal value at any logic node in the circuit by controlling its primary input and observing the primary output [4]. The observability values for all signals from primary outputs towards primary inputs are then calculated. The observability of one input of a AND gate with multiple inputs is given by

$$CO(s) = P(x,y)+1 \qquad (2)$$

where, x=output observability and y= CC1 of other inputs. If 'U' is a primary output node of a digital circuit, then the combinational observabilities of node I are defined as, CO(U) = 0. Table II, Table III, and Table IV reports the controllability and observability values obtained.

E. Machine Learning Algorithms and Data SetCreation

The use of machine learning algorithms aids in the better analysis of various Trojan threats because they are capable of processing massive amounts of data with greater precision. Using Python, dataset is created by evaluating the controllability and observability values received from the chosen benchmark circuits. The more advanced algorithms can be trained to detect any kind of Trojan across various platforms. Hence, we employ various Machine Learning Techniques (MLT) [25] along with our proposed technique to detect the presence of Trojan in the circuits by classifying the nodes as Trojan free and Trojan infected.

V. Experimental Results and Discussion

A. Functionality Verification

To understand the impact of Trojans on the circuits let us consider C3540, an 8bit ALU. C3540 consists of Mux, shift register, ALU_Core, xor gate, and BCD subtractor. Trojan can be introduced to any block in the circuit. To understand the effect of Trojan, we are introducing Trojan to the BCD adder block in C3540. Let us consider two and five as inputs to this BCD block. When this block is configured as adder, the output is six in the absence of Trojan. When Trojan is activated, the output changes. This can be observed in Fig. 11, where the Trojan activation at 72ns changes the output value from 6D to 83 and remains in the same state as long as Trojan is activated. Once the Trojan is deactivated, output changes to the original value. As the blocks in C3540 are cascaded as seen in Fig. 11, the final output of ALU will also be changed. This clearly demonstrates that, due to the presence of Trojans, functionality of the circuit will change. Similarly, functionality verification of C17 and C432 benchmark circuits are carried out for both the cases, with Trojan and without Trojan. The obtained simulation results are shown in Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, and Fig. 11. Thus, this shows how the entire functionality of a circuit or a system change upon Trojan activation.

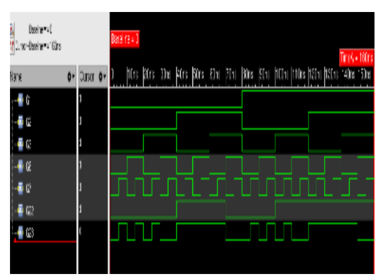


Fig. 6. C17 simulation result without Trojan

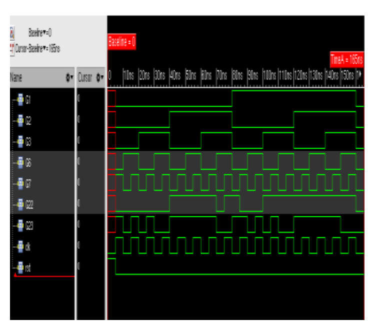


Fig. 7. C17 simulation result with Trojan

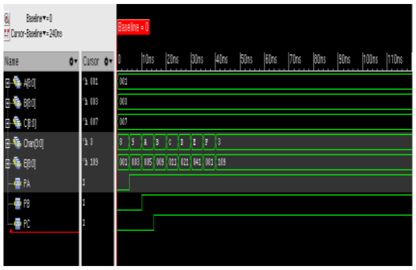


Fig. 8. C432 simulation result without Trojan

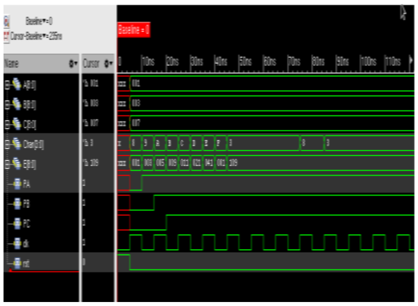


Fig. 9. C432 simulation result with Trojan

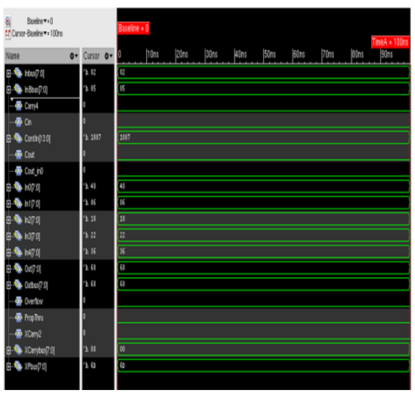


Fig. 10. C3540 simulation result without Trojan

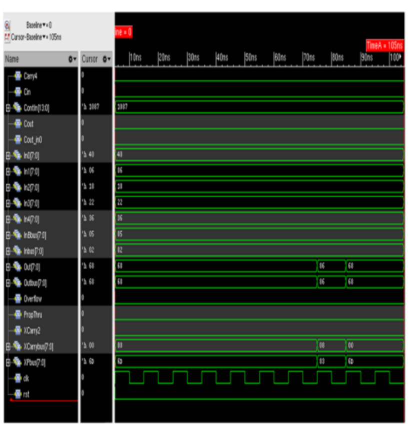


Fig. 11. C3540 simulation result with Trojan

B. Performance Analysis

Performance analysis of the benchmark circuits-C17, 432 and C3540 using area and power metrics are shown in Table I. The circuits are implemented in Verilog HDL using Cadence Incisive and synthesized in 90nm technology using Cadence Genus. It can be observed that, area of the circuits without Trojan is less compared to with trojan. It can be noticed that, power consumption of the circuits in the absence of Trojan is less. From Table I, it is evident that Trojan presence can also be determined by performing area and power analysis. This is due to the fact that, any new addition of unwanted components such as Trojan, increases the area and power consumption in a drastic way.

TABLE I. SYNTHESIS REPORT

Bench mark Circuits	Area(μm^2)		Power(μW)	
	Without Trojan	**With Trojan**	**Without Trojan**	**With Trojan**
C17	**16.62**	**227.07**	**0.425**	**6.76**
C432	**439.002**	**3279.648**	**15.205**	**107.09**
C3540	**4931.203**	**5196.118**	**185.38**	**194.27**

C. Controllability and Observability Calculation

Controllability and Observability are the parameters used for detection of Trojan in the benchmark circuits. Netlist obtained from synthesis of C17, C432 & C3540 is converted into benchmark formats. The benchmark formats are then fed as input to the python code that determines controllability and observability values. The output in the form of a text file is used to create datasets. Combinational controllability is used to identify nets which show difficulty in testability. The controllability values range from 1 to infinity. If 'I' is a primary input node of a digital circuit, then the combinational controllabilities of node 'I' are defined 1 i.e., CC0(I) and CC1(I)=1. Similarly, calculations of combinational controllabilities for various gates are shown in Fig. 12.

An Observability is simply a function of controllability, meaning that it is impossible to observe a given internal node if the circuit is not driven to a given state. The Observability values range from 0 to infinity. If 'U' is a primary output node of a digital circuit, then the combinational observabilities of node 'U' are defined as, CO(U) = 0

The formulation of combinational observabilities for various gates are shown in Fig. 13 where,

CC = Combinational Controllability

CO = Combinational Observability

Trojan value 0 = No Trojan detected

Trojan value 1 = Trojan detected

Table II, Table III, and Table IV represent samples of datasets generated using Controllability and Observability values obtained using the calculations for various gates mentioned in Fig. 12 and Fig. 13. The gate-level netlists are converted to benchmark codes that are fed into the python code that performs the calculations to output files with the testability measures.

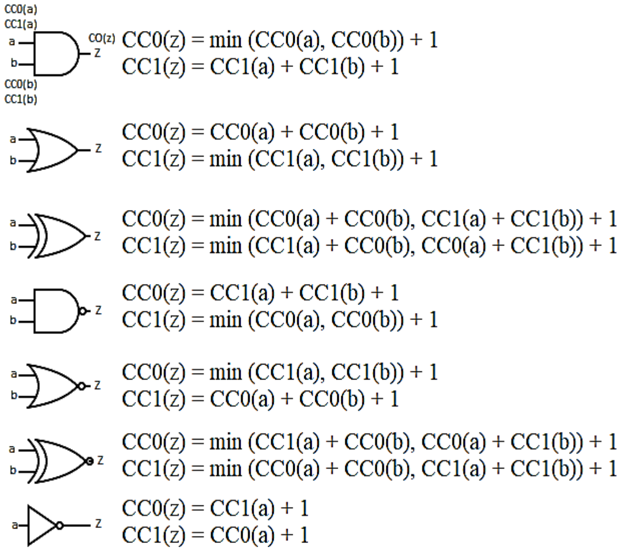


Fig. 12. Combinational controllability calculation for various gates

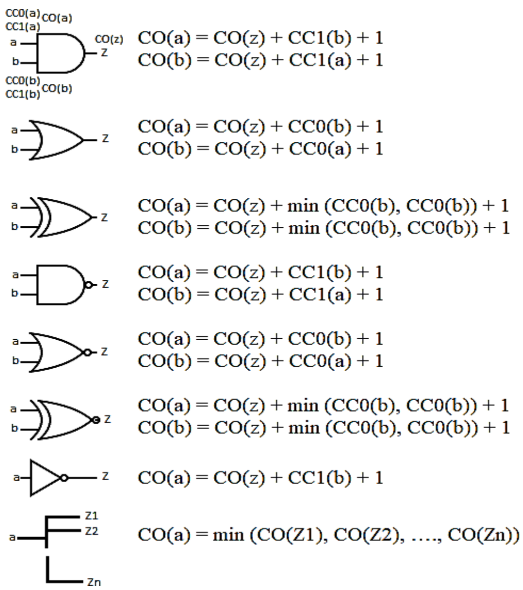


Fig. 13. Combinational observability calculation for various gates

TABLE II. C17 DATASET

Line Name	CC	CO	Trojan
G3gat	4.2	17	0
G1gat	1.4	5	0
G22gat	6.4	0	0
G10gat	3.6	3	0
G19gat	4.5	3	0
G3gat	4.2	54	1
G10gat	38.1	3	1
G19gat	4.5	3	1
G1gat	1.4	42	1
G22gat	7.1	0	1

TABLE III. C432 DATASET

Line Name	CC	CO	Trojan
G159	13.5	74	0
G165	13.5	74	0
G295gat	14.3	152	0
G1gat_1	1.4	38	0
. . .			
G236gat_0	14.2	157	0
G159	13.5	76	1
G165	13.5	76	1
G360gat	15.44	244	1
. . .			
G1gat_1	1.4	56	1
G236gat_0	14.2	163	1

TABLE IV. C3540 DATASET

Line Name	CC	CO	Trojan
G905	22.8	141	
G906	7.6	89	0
G116	5.6	409	0
. . .			
G353	11.7	0	0
G68	8.4	364	0
G905	126.3	141	1
G906	42.1	89	1
G116	5.6	384	1
. . .			
G1018	41.2	82	1
G625	42.1	4	1

D. Machine Learning Techniques (MLT)

MLT are used for detection of Trojans in the benchmark circuits. Controllability and Observability values calculated for C17, C432 and C3540 are used as input to create dataset. Upon pre-processing the data, dataset is split into training set and test set. This data set is used on various MLT for classifying the nodes as Trojan free and Trojan infected. Scatter plot shown in Fig. 14, Fig. 15 and Fig. 16 better visualizes the results. C17 circuit is easier to classify as with Trojan and without Trojan but C432 circuit has considerable overlapping and is harder to classify. Thus, non-linear ML models have to be used in order to separate this nonlinear data. Hence, we propose an MLT model which is a combination of LDA and Naive Bayes. Comparative study of all these MLT along with the proposed model in terms of accuracy, precision, F1 score and Recall is carried out for C17, C432 and C3540 circuits (see Fig. 17 to 22). It is evident from Table V, Table VI and Table VII that our proposed LDA+ Naive Bayes model is best among other MLT with the latency of around 0.9 ms in comparison with the other state-of-art techniques.

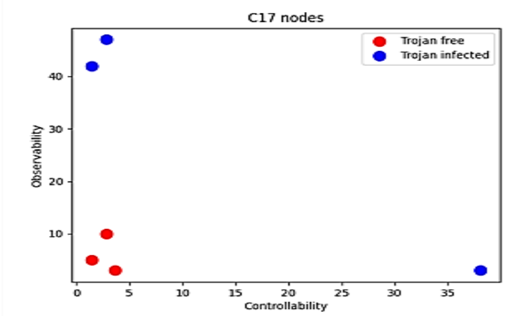


Fig. 14. Observability (Y) vs controllability (X) scatter plot results of C17 nodes

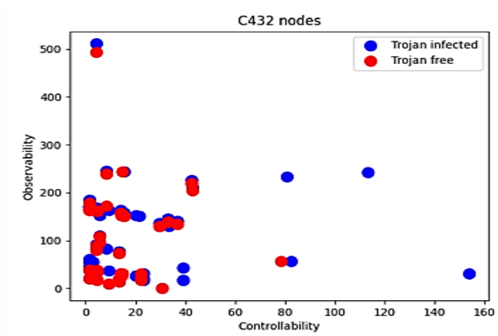


Fig. 15. Observability (Y) vs controllability (X) scatter plot results of C432 nodes

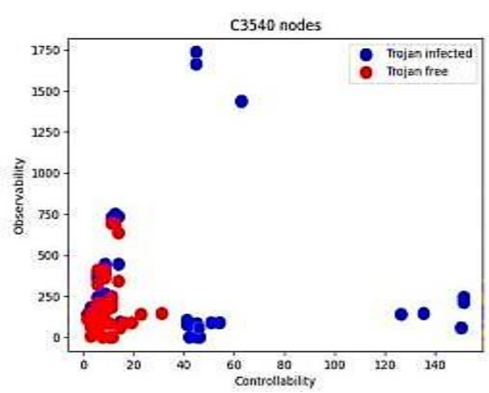


Fig. 16. Observability (Y) vs controllability (X) scatter plot results of C3540 nodes

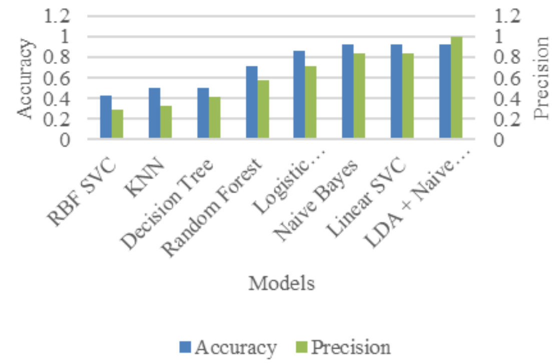


Fig. 17. C17 accuracy and precision metrics

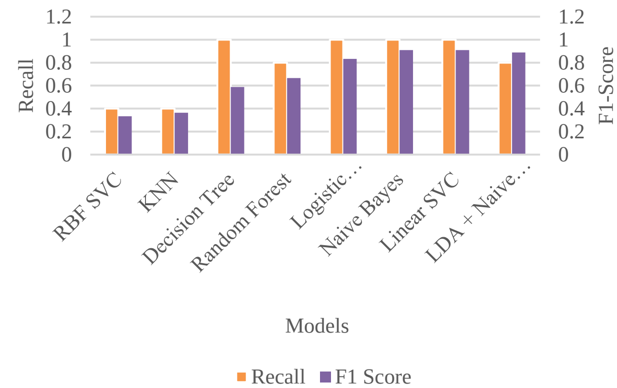


Fig. 18. C17 recall and F1-score metrics

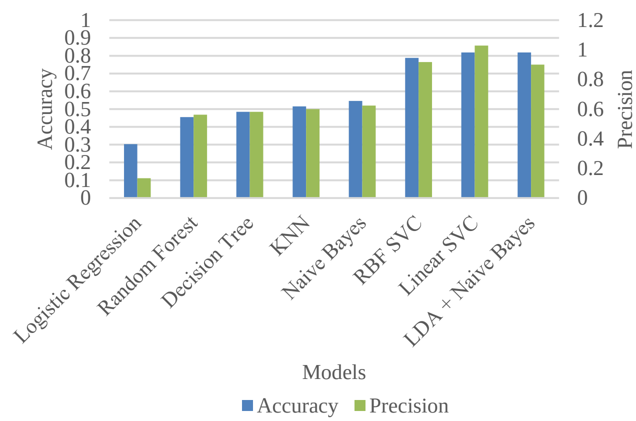


Fig. 19. C432 accuracy and precision metrics

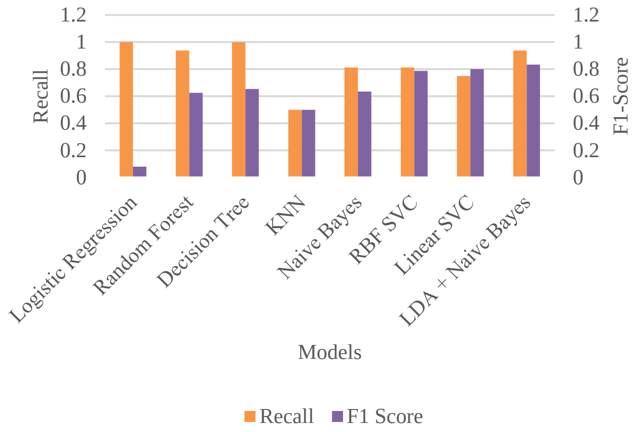


Fig. 20. C432 recall and F1-score metrics

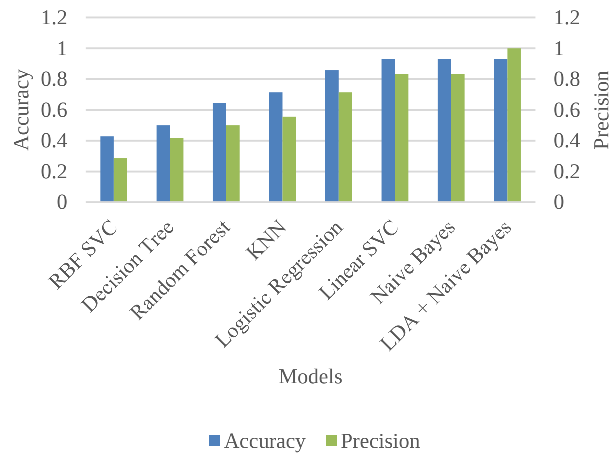


Fig. 21. C3540 accuracy and precision metrics

Fig. 22. C3540 recall and F1-score metrics

The Performance Metrics in Table V, Table VI, Table VII use the following parameters:

TABLE V. C17 PERFORMANCE METRICS

Model	A (%)	P (%)	R (%)	F (%)	TP	TN	Time (ms)
RBF SVC	43	29	40	33	4	2	0.9
KNN	50	33	40	36	5	2	3.9
Decision Tree	50	42	100	59	2	5	0.9
RandomForest	71	57	80	67	6	4	17.9
Logistic Regression	86	71	100	83	7	5	0.9
Naive Bayes	93	83	100	91	8	5	0.9
Linear SVC	93	83	100	91	8	5	0
LDA + Naive Bayes	93	100	80	89	9	4	0.9

TABLE VI. C432 PERFORMANCE METRICS

Model	A (%)	P (%)	R (%)	F (%)	TP	TN	Time (ms)
Logistic Regression	30	11	100	8	9	1	0.0
Random Forest	45	47	94	63	0	15	0.9
Decision Tree	48	48	100	65	0	16	0.0
KNN	52	50	50	50	9	8	1.9
Naïve Bayes	55	52	81	63	5	13	0.0
RBF SVC	79	76	81	79	13	13	11.9
Linear SVC	82	86	75	80	15	12	0.5
LDA + Naïve Bayes	82	75	94	83	12	15	1.9

TABLE VII. C3540 PERFORMANCE METRICS

Model	A (%)	P (%	R (%)	F (%)	TP	TN	Time (ms)
RBF SVC	43	29	40	33	4	2	0
Decision Tree	50	42	100	59	2	5	0
Random Forest	64	50	60	55	6	3	9.9
KNN	71	56	100	71	5	5	2.9
Logistic Regression	86	71	100	83	7	5	0.9
Linear SVC	93	83	100	91	8	5	0
Naïve Bayes	93	83	100	91	8	5	0.9
LDA + Naïve Bayes	93	100	80	89	9	4	0.9

VII. Conclusion

Considering the various threats to the manufacturing process of an IC/system like functionality change, IC tampering, third party vendor attacks etc, a new Hardware Trojan detection technique using Machine Learning is proposed. Controllability and Observability analysis was performed using Gate Level netlists. Based on the values obtained, the various Machine learning models were able to distinguish the nodes in three of the benchmark circuits used as Trojan free or Trojan infected. Amongst them, our proposed model i.e., LDA + Naïve Bayes performed the best when compared to other state-of-art techniques with an accuracy of 92.85%, precision of 99.9

%, recall of 80% and F1 score 88.8%. The latency of the proposed technique was around 0.9ms. Along with this, the Simulation and Synthesis reports obtained using 90nm technology of Cadence tool also proved that presence of Trojan increasingly affects the system in terms of area and power.

References

[1] K. Madden, J. Harkin, L. McDaid, and C. Nugent, "Adding Security to Networks-on-Chip using Neural Networks," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1299-1306.

[2] Noor, N. Q. et al. "Hardware Trojan Identification Using Machine Learning-based Classification." Journal of Telecommunication, Electronic and Computer Engineering 9 (2017): 23- 27.

[3] Z. Huang, Q. Wang, Y. Chen and X. Jiang, "A Survey on Machine Learning Against Hardware Trojan Attacks: Recent Advances and Challenges," in IEEEAccess, vol. 8, pp. 10796-10826, 2020.

[4] H. Salmani, "COTD: Reference-Free Hardware Trojan Detection and Recovery Based on Controllability and Observability in Gate-Level Netlist," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 2, pp. 338-350, Feb. 2017.

[5] J. Cruz, F. Farahmandi, A. Ahmed, and P. Mishra, "Hardware Trojan detection using ATPG and model checking," in 2018 31st International Conference on VLSI Design and 2018 17th International Conference on Embedded Systems (VLSID), 2018: IEEE, pp. 91-96.

[6] Y. Xiang, L. Li and W. Zhou, "Random Forest Classifier For Hardware Trojan Detection," 2019 12th International Symposium on Computational Intelligence and Design (ISCID), 2019, pp. 134-137, doi: 10.1109/ISCID.2019.00037.

[7] Y. Huang, S. Bhunia, and P. Mishra, "MERS: statistical test generation for side-channel analysis based Trojan detection," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016: ACM, pp.

[8] J. He, Y. Zhao, X. Guo, and Y. Jin, "Hardware trojan detection through chip-free electromagnetic side- channel statistical analysis," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 25, no. 10, pp. 2939-2948, 2017.

[9] A. Kulkarni, Y. Pino and T. Mohsenin, "SVM-based real-time hardware Trojan detection for many-core platform," 2016 17th International Symposium on Quality Electronic Design (ISQED), Santa Clara, CA, 2016, pp. 362-367.

[10] A. Amelian and S. E. Borujeni, "A side-channel analysis for hardware Trojan detection based on path delay measurement," Journal of Circuits, Systems and Computers, vol. 27, no. 09, p. 1850138, 2018.

[11] F. K. Lodhi, I. Abbasi, F. Khalid, O. Hasan, F. Awwad and S. R. Hasan, "A self-learning framework to detect the intruded integrated circuits," 2016 IEEE International Symposium on Circuits and Systems (ISCAS), 2016, pp. 1702-1705, doi:10.1109/ISCAS.2016.7538895.

[12] K. Hasegawa, M. Yanagisawa and N. Togawa, "Trojan-feature extraction at gate-level netlists and its application to hardware-Trojan detection using random forest classifier," 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, 2017, pp. 1-4.

[13] K. Hasegawa, M. Oya, M. Yanagisawa and N. Togawa, "Hardware Trojans classification for gate- level netlists based on machine learning," 2016 IEEE 22nd International Symposium on On-Line Testing and Robust System Design (IOLTS), Sant Feliu de Guixols, 2016, pp. 203-206.

[14] Noor, N. Q. et al. "Hardware Trojan Identification Using Machine Learning- based Classification." Journal of Telecommunication, Electronic and Computer Engineering 9 (2017): 23- 27.

[15] T. Liu, W. Wen and Y. Jin, "SIN2: Stealth infection on neural network — A low-cost agile neural Trojan attack methodology," 2018 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), Washington, DC, 2018, pp. 227-230.

[16] A. Waksman, M. Suozzo, and S. Sethumadhavan, "FANCI: identification of stealthy malicious logic using boolean functional analysis," in Proc. ACM SIGSAC Conference on Computer and Communications Security (ACM-CCS), 2013, pp. 697–708.

[17] A. Moradi, D. Oswald, C. Paar, and P. Swierczynski, "Side-channel attacks on the bit stream encryption mechanism of Altera Stratix II," Proc. ACM/SIGDA Int. Symp. F. Program. Gate arrays - FPGA '13, p. 91, 2013.

[18] R. Elnaggar and K. Chakrabarty, "Machine Learning for Hardware Security: Opportunities and Risks," Journal of Electronic Testing, vol. 34, no. 2, pp. 183- 201, Apr. 2018, doi: http://doi.org/10.1007/s10836-018-5726-9.

[19] M. Banga, M.S. Hsiao, "A Region based Approach for the Identification of Hardware Trojan," in Proc. IEEE Intl. Symp. On Hardware Oriented Security and Trust (HOST'08), pp. 40-47, Jun. 2008, doi: 10.1109/HST.2008.4559047.

[20] Goldstein, Lawrence H., and Evelyn L. Thigpen. "SCOAP: Sandia controllability/observability analysis program." Proceedings of the 17th Design Automation Conference. 1980.

[21] J. Wang, S. Guo, Z. Chen and T. Zhang, "A Benchmark Suite of Hardware Trojans for On-Chip Networks," in IEEE Access, vol. 7, pp. 102002- 102009, 2019.

[22] Trust-Hub.org

[23] Martin, Antonio J. Lopez. "Cadence design environment." New Mexico State University, Tutorial paper (2002): 35.

[24] Reshma K., Priyatharishini M., Nirmala Devi M. (2019) "Hardware Trojan Detection Using Deep Learning Technique". In: Wang J., Reddy G., Prasad V., Reddy V. (eds) Soft Computing and Signal Processing. Advances in Intelligent Systems and Computing, vol 898. Springer, Singapore.

[25] K. G. Liakos, G. K. Georgakilas, S. Moustakidis, P. Karlsson and F. C. Plessas, "Machine Learning for Hardware Trojan Detection: A Review," 2019 Panhellenic Conference on Electronics & Telecommunications (PACET), 2019, pp. 1-6, doi: 10.1109/PACET48583.2019.8956251.

Multi-Objective Optimal Path Planning for Autonomous Robots with Moving Obstacles Avoidance in Dynamic Environments

Kadari Neeraja[1], Dr. G Narsimha[2]
Research Scholar[1], Associate Professor[1], Professor[2]
Department of CSE, JNTUH, Hyderabad, India

***Abstract*—Path planning is vital for robust autonomous robot navigation. Driving in dynamic environments is particularly difficult. The majority of the work is based on the premise that a robot possesses a comprehensive and precise representation of its surroundings prior to its starting. The problem of partially knowing and dynamic environments has received little attention. This circumstance occurs when an exploratory robot or a robot without a floor plan or terrain map must move to its destination. Existing approaches for dynamic-path-planning design a preliminary path based-on known knowledge of the environment, then adjust locally by replanning the total path as obstacles are discovered by the robot's sensors, thereby sacrificing either optimality or computational efficacy. This paper presents a novel algorithm. A Near-Optimal Multi-Objective Path Planner (NO-MOPP), capable of planning time-efficient, near-optimal, and drivable paths in partially known and dynamic environments. It is an expansion of our earlier research contributions called "A Multi-Objective Hybrid Collision-free Optimal Path Finder (MOHC-OPF) for Autonomous Robots in known static environments" and "A Multi-Objective Hybrid Collision-free Near-Optimal Path Planner (MOHC-NODPP) for Autonomous Robots in Dynamic environments". In the environment, a mix of static and moving dynamic obstacles are present, both of which are expressed by a hybrid, discrete configuration space in an occupancy-grid map. The proposed approach is executed at two distinct levels. Using our earlier method, A Multi-Objective Collision-free Optimal Path Finder (MOHC-OPF), the initial optimal path is found in environment that includes only known stationery obstacles at the Global-path-planning level. On the second level, known as Local Re-planning, this optimal path is continuously refined by online re-planning to account for the movement of obstacles in the environment. The proposed method, A Near-Optimal Multi-Objective Path Planner (NO-MOPP), is used to keep the robot's sub-paths optimum while also avoiding dynamic obstacles. This is done while still obeying the robot's non-holonomic restrictions. The proposed technique is tested in simulation using a collection of standard maps. The simulation findings demonstrate the proposed method's ability to avoid static as well as dynamic obstacles, as well as its capacity to find a near-optimal-path to a goal location in environments that are constantly changing without collision. The optimal-path is determined by taking into account several performance measures, including path length, collision-free path, execution time, and smooth paths. 90% of studies utilizing the proposed method demonstrate that it is more effective than other methods for determining the shortest length and time-efficient smooth drivable paths. The proposed technique reduced average 15% path length and execution time compared to the existing methods.**

Keywords—Autonomous mobile robots; dynamic environment; planning; collision-free; time-efficient paths

I. Introduction

Path planning for mobile robots, especially when the environment is known, is a well-researched problem [1-8]. However, one issue that arises when putting theory into practice is the fact that incomplete information about the environment is often available. In most cases, it seems unrealistic to expect to have a detailed map with all the obstacles clearly marked. Recent years have seen tremendous progress realized in the realm of path planning in dynamic environments across a wide range of domains. Particularly, mobile robots have found practical use across a wide range of domains. Applications incorporate emergency rescue management in natural disasters [12], planetary exploration [10, 11], inventory control [12], the manufacturing industry [13], etc.

In the 1960s, research began in the arena of path planning for different kinds of robots [1, 2, 12]. The Path-Planning is the procedure of establishing a path-way in an environment that is no-obstacles and that connects a predetermined starting point and an intended ending point [14-15]. The environments in which robots operate can either be static or dynamic. When working in a static known environment, the locations of obstacles remain the same, but when working in a dynamic environment, their positions shift over the course of time. The goal of employing path-planning algorithms is to translate the high-level-specifications which humans execute into low-level-steps [15]. This is accomplished by locating the optimal-path and presenting to the robot in the form of a series of waypoints that it should follow as moving directions.

The dynamic environment comprises moving obstacles, the path-planning algorithm is a necessity not-only to determine the optimal-path but also to observe it. In order to persist responsive to its environments, the approach must know the current position of an obstacle, forecast upcoming paths, and bring up-to-date its path in real-time with sufficient frequency.

In most cases, an autonomous mobile robot is free to follow any one of a number of predetermined routes. The length of the path, the amount of time it takes, and the amount of energy it takes are all factors that go into determining what constitutes the optimal path. Numerous algorithms were developed to handle the problem of path planning; these algorithms can be categorized as either classical or intelligent. Artificial Potential Field [16], Rapid Exploration of Random Trees (RRT) [15, 17] and its variants RRT*, etc. [15,18], Partitioned Learning Traditional methods, such as D* [20], are utilized to solve dynamic path planning problems. As the search space grows in size, however, these methods become inefficient and get stuck at local maxima. Thus, intelligent optimization techniques like the Genetic Algorithm [21-22], Particle Swarm Optimization [23], Bees algorithm [24] [29] and etc. have been employed to solve path planning difficulties.

Real-time dynamic path planning needs more investigation [30], as stated earlier. This study proposes an A*[3] based Near-Optimal Multi-Objective Path Planner (NO-MOPP) to swiftly identify a near-optimal drivable smooth path in a dynamic environment while taking the kinematic restrictions of the robot into account. The following are some of the contributions made in this work:

1) "A Multi-Objective Hybrid Near-Optimal Dynamic Path Planner (NO-MOPP)" is a new dynamic path planning technique that finds no-collision near-optimal-drivable paths in a dynamic environment with a hybrid environment representation. Since A* ensures both optimality and completeness, the A* algorithm will serve as the foundation for this approach. This algorithm performs comparably to A*.

2) The proposed technique performs on two distinct levels. Initially, using global path planning, the optimal-path is determined in an area with known static-obstacles. The second level, known as Local-replanning, adjusts the optimal path online with the assistance of sensors in order to prevent collisions with dynamically generated immovable and moving obstacles. After that, path tracking is performed, and the path is optimized during path tracking without sacrificing the algorithm's real-time performance.

3) Kinematic constraints, like a robot's orientation, are employed in order to find the most efficient driving smooth paths in ever-changing real-time environments.

4) It finds application in a wide range of different dynamic environments. The percentage of successful attempts is 90%.

5) When compared to RRT and RRT*, our suggested method achieves superior outcomes in dynamic environments in relation to the amount of time required for execution, execution time, and the total length travelled path-length.

The paper is organized as follows. Section II explains related research that is pertinent to the techniques for planning paths. The technique and underlying algorithm for path-planning in the presence of static as well as dynamic obstacles are presented in Section III. Section IV looks at how well the suggested strategy works and gives the results of the experiments. Section V brings the article to a conclusion, which also offers guidelines for future work.

II. Related Work

Past decades have seen many path-planning algorithms. Graph-based techniques include Dijkstra's algorithm [4,24], A* [3], D* [20], and etc. After discretizing the path planning state space into a graph structure, they employ graph search to find a feasible path. A* and Dijkstra's algorithms are suitable for lower-dimension static environments. D* is used for dynamic environments. Optimal Path Planning using Memory Efficient A*. Improved A* Path Planning Method Based on the Grid Map. Sensors [25], Time-Efficient A* Algorithm for Robot Path Planning [26], Safe Path Planning of Mobile Robot Based on Improved A* Algorithm in Complex Terrains [27], Optimal Path Planning using Memory Efficient A*[31] and Fast path planning using modified A* method [32].

The graph-based approach is full and resolution optimal, meaning it finds an optimal-path if a viable path-exists and fails otherwise. The graph-based partition of the state-space yields a massive search space, which makes these graph-based approaches unsuitable for large-scale issues. The recent updates on A*, in research papers like Dynamic-Algorithm for Path-Planning using A* with Distance-Constraint [13] and Improved-Analytic-Expansions in Hybrid A* Path-Planning for Non-Holonomic Robots [9]. They are suffering from high computation time.

Another significant kind of path-planning algorithm is the sampling-based path-planning approach. Instead of discretizing the state space, it generates a graph or tree by randomly selecting points. Sampling-based path planning algorithms beat graph-based ones in large-scale situations. The sampling-based path planning strategy is probabilistically complete, thus when the trials number reaches infinite, the likelihood of discovering a suitable path-way approaches one. Sampling-based planners employ RRT [15,17] and PRM [19] algorithms. The RRT, a single-query path planning method that traverses state space by generating a tree rooted at the start state, is faster than the PRM. Despite finding an initial path in high-dimensional space quickly, RRT has many downsides. RRT's path may not be ideal because it is randomly generated. RRT* [18] advanced RRT. The RRT* takes time and memory to identify the best path. RRT* likewise experiences significant search time variability. Though, these techniques perform poorly and trap in local optima when the search space is big.

Hence, intelligent optimization procedures have been employed in the process of solving path-planning problems. Some examples of these algorithms include the Genetic Algorithm [21-22] [28], Particle Swarm Optimization [23], Simulated Annealing [12], Ant Colony Optimization [8], Bees [29] and etc. Even if these algorithms conquered the difficulties of path planning, they still wouldn't be usable without the partitioning and pre-processing of environment maps. This is because such maps need to be prepared in advance. The accuracy is reduced as a result of discretization and pre-processing, which also results in non-optimal pathways.

III. Proposed Path Planning System

The purpose of path-planning is to discover a continuous path that will lead a system from its current state to a desired one. Finding a path across a dynamic environment that a non-holonomic robot or vehicle can follow without colliding with any of the environment's obstacles is the goal of dynamic path planning.

1) Problem formulation: The path-planning issue's main objective is to identify an optimal-path for an autonomous-robot to proceed from a given beginning-point to a certain goal location in a dynamically changing environment containing static and dynamic stationery as well as moving obstacles by satisfying optimization-criteria. The path-planner's goal is to discover the optimal- or near-optimal-path for a mobile-robot that avoids obstacles in the surroundings.

The environment is denoted as a Grid. The initial step in mobile-robot-path-planning is establishing an environment-model for the mobile-robot's 2-dimensional. As identical square cells, grids are used to represent the mobile-robot's workplace. Each grid-cell is either free (logic 0) or forbidden (logic 1) by an obstacle. There are both static as well as dynamic stationary and moving obstacles in this area.

2) Optimization criteria: The proposed path planning system NO-MOPP determines a no-collision smooth path that obeys multi-objective optimization criteria. The criteria is: first one is the Cost objective-function , minimum-cost path for a mobile-robot to move from its start-point to the goal-point, provided that it is a smooth and safe path, i.e., the mobile-robot travels with no collision with obstacles. This measure is specified by Eq. (1) :

$$\text{Cost } f = g + h + SO + DSO + DMO \quad (1)$$

where cost-f is the sum-of-the-costs from the start-node to current-node *(g)*, the estimated-cost *(h)* to the goal from current-node, the additional-cost for changing the orientation angle, SO is the cost for switching orientation, DSO corresponds to the cost of avoiding dynamic stationary obstacles, and DMO indicates the cost of avoiding dynamic stationary obstacles during local replanning.

The second Criterion is the *Execution time objective* needed for finding a minimum length and safe path. This is given by equation (2):

$$T_{total} = T_{Global\text{-}path\text{-}planning} + T_{local\text{-}replanning} \quad (2)$$

where T_{total} is the time essential for the completion of execution of the path planner, $T_{Global\text{-}path\text{-}planning}$ is the time taken by the offline line path planner and $T_{Local\text{-}raplanning}$ is the time needed to update the initial optimal path to skip dynamic stationery obstacles.

The optimal path-planning problem can be stated as:

"Find the lowest cost and least time taking near-optimal smooth path between the start-point and the goal-location, such that the above optimization-criteria Cost function f and Execution time T objective functions given in above equations (1) & (2) are lessened by taking non-holonomic constraints of the robot into account".

3) The proposed dynamic path planning system's architecture: The component structure of the dynamic-path-planning system is depicted in Fig. 1. The system will function efficiently on the two levels. At the first level, global path planning, the optimal route is determined using the information that is currently known about the environment, including the known static obstacles, for instance. After that, the robot will continue along this optimal path. This optimal path is updated online at the second level, which is known as "Local Replanning," in order to skip collisions with-obstacles which are dynamically presented and may be either stationary or dynamic.

Primary components of the dynamic-path-planning system are depicted in Fig. 1. The environment is dynamic and comprises both static and moving obstacles. To represent this ever-changing environment, a binary occupancy grid map is employed. Constraints: Non-holonomic car-like robots or vehicles have kinematic constraints. Optimization criteria: Path smoothness, path length, and the time required to locate a path comprise optimization criteria.

4) Path-planning: The path-planning algorithm is the crucial component of a dynamic-path-planning system that addresses a path-planning issue. In this proposed system, there are two levels. A Multi-Objective Optimal Path Finder (MOHC-OPF) is used to obtain a quick initial optimal-path in an environment that includes known static obstacles only in the First level of Global Path Planning. Local-Replanning is the second level of our proposed dynamic path planning approach, a Near-optimal Multi-Objective Path Planner (NO-MOPP), which is employed to avoid dynamic obstacles in dynamically changing environments.

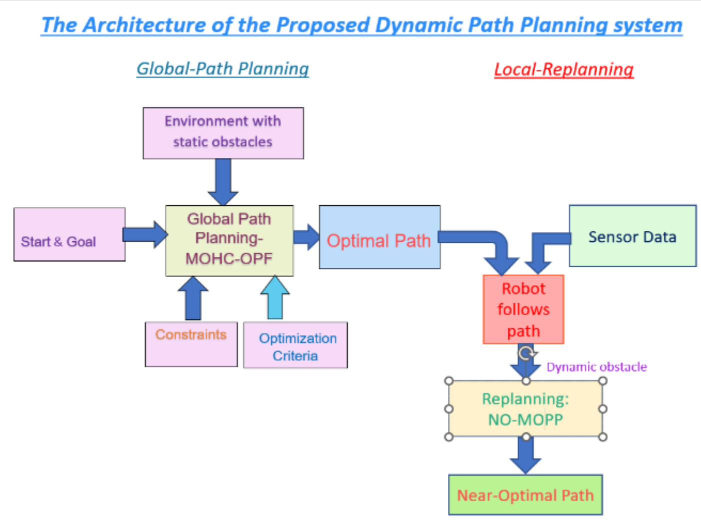


Fig. 1. Block diagram of the dynamic path planning system

A. The Working Principle of NO-MOPP

A novel method called NO-MOPP has been proposed as a way to avoid the limitations of the traditional A* methodology. The kinematics of the car-like robot or vehicle is added to predict the movement of the robot which is dependent on the steering angle in a continuous search space. The proposed system has a collection of continuous states represented by the coordinates (x_p, y_p, θ), here (x_p, y_p) represents the location of the robot or vehicle and θ represents its orientation. Non-holonomic robots and vehicles can benefit from this feature since it helps the path planner choose the best successor state for them to follow. One of the five steering actions, maximum-left, left, maximum-right, right, and no-steering, expands the states and leads to an arc of a circle with a minimum turning radius, in accordance with the kinematic restrictions of the simple car-like robot or vehicle. On the basis of these operations, the proposed method, the NO-MOPP algorithm selects the states depicted in Fig. 2.

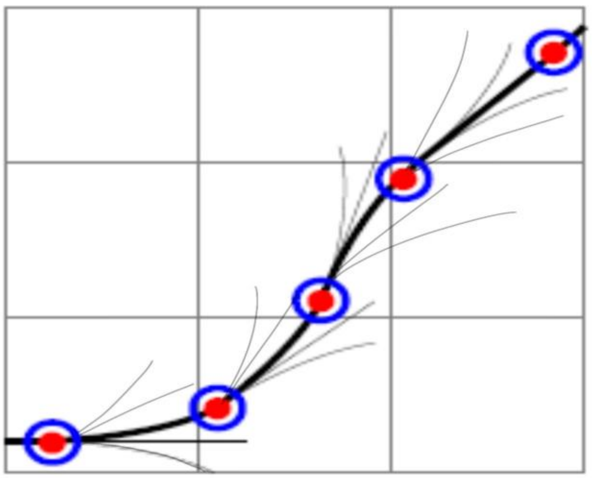


Fig. 2. NO-MOPP incorporates kinematic constraints with 5 steering angles

B. Multi-Objective Functions for Optimization

As part of this effort, different objectives are analysed and taken into consideration so that the updated path can be optimized.

Cost Function: It calculates the cost of driving from the present-point to a neighboring node. This cost f is the total-cost from the start node to the current node (g), the anticipated-cost from the present node to the goal-(h), and the cost for adjusting the orientation angle (SO). Eq. (1) is utilized to calculate this cost.

1) Path length: The final path is made up of a series of path segments denoted by the notation P= {P1, P2, …, Pn}. The ultimate length of the path is equal to the totality of the lengths of all path segments that connect the Start state to the Goal state via any intervening states. This is illustrated in Fig. 3. To get the length of this final path, Eq. (2) is utilized.

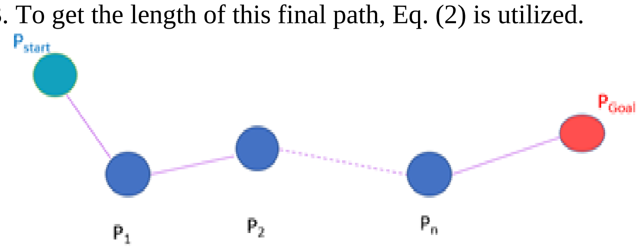


Fig. 3. Path length

$$\text{Path Length} = \sum \text{Pi} \text{ where i= 1 to n.} \qquad (3)$$

At each stage, the next state with the lowest possible cost is chosen, and the arcs in the path are optimized to have the minimum turning radius. As a result, it ensures that the final path will be the shortest and most optimal one.

2) Execution time: The amount of time required to carry out the method that was proposed for discovering a path and Eq. (2) is used for determining this.

The travelled distance must be sufficient in order to exit the current cell, and the equation that describes this requirement is below (4).

$$l > \sqrt{2}\, s \qquad (4)$$

In this equation, l represents the length of an arc, and s indicates the size of a single cell in the grid map. At each node, the continuous state is rounded off to a discrete state in order to prevent the search graph from becoming an increasingly huge structure. This, in turn, results in a reduction in the amount of search time necessary to locate the ultimate path.

3) Smooth path: Kinematic constraints determine the next node in this proposed approach, resulting in minimum turning radius curves. Therefore, the final path produced is smooth.

C. The Heuristic Function

The heuristic function predicts the minimum-cost from any node to the destination on the map. This reduces node exploration. Thus, heuristic function selection directly impacts the performance of path planning approaches. The Euclidean distance is employed as a heuristic function. Equation (5) determines each node's heuristic values.

$$h = \sqrt{(x2 - x1)2 + (y2 - y1)2} \qquad (5)$$

D. The Proposed Path Planning Approach

The proposed approach executes the path-planning procedure in a dynamic-environments with both dynamic as well as static obstacles. The method operates on two levels. Global Path Planning 2. Local Redevelopment Replanning. In the first level, the Global-path-planning method is applied to determine the optimal-path through a static obstacle environment. The attained path is provided for the robot to track during the second level. Simultaneously, the algorithm modifies the path in real-time to prevent a collision by means of any new obstacles, to ensure the sub-paths are optimum. Using proposed path-planning system, the path's optimality is preserved. Fig. 1 depicts this process.

1) The global path planning: By adhering to kinematic constraints, the MOHC-OPF method, which was the objective of our previous research, is used in Global Path Planning to swiftly construct an optimal path for a given environment containing only static obstacles. The MOHC-OPF algorithm employs Open-list and Closed-list. Comparable to the conventional A*, they keep track of the states while searching. The open list includes the neighbours of states that have been expanded during the search process. The closed list comprises all states for which processing has been finalized. Here is a summary of the MOHC-OPF algorithm.

Input: Occupancy grid map of the environment with static obstacles information, starting position (xstart), target position (xgoal), heading θ, and kinematic constraints of simple car-like robot and U(x) actions set with five forward steering angels.

Output: A mobile robot's optimal path from its starting point to its goal, including path length and execution time.

Step 1: Set Open list and Close list

Step 2: Assign Xstart to the start state,

Step 3: For each of the 5 steering angles, find Xstart 's 5 neighbors. Using a simple car-like robot's kinematic model with a global location of (xg, yg, θ) generated using the equation (6) (LaValle, S.M. Planning Algorithms, 2006. [15]).

$$Xg. = ug \cos \theta$$

$$Yg. = ug \sin \theta$$

$$\varphi = us / l \tan \theta$$

$$\rho \min = l / \tan\varphi_{max} \quad (6)$$

Where u is an action set {0,1}, φ is the steering angle, ρ min is the minimum turning radius and l is the front-rear-axles distance of a simple car.

Step 4: If any of the neighbors is a goal state, then quit.

Step 5: Estimate the cost of each neighbor using the cost function equation (1) if they are not likely to collide.

Step 6: Assign Xstart to a neighbor having the minimum cost function f-value, then execute the related action on the map. Keep the former Xstart and the f values in an open list.

Step 7: Repeat from step 3 until Open-list is empty.

2) The local replanning: Global path planning generates the optimal-path in a dynamic-environment with stationary obstacles only. In Local re-planning, the autonomous mobile robot or vehicle takes this optimal path from the starting-point to the target-point in a dynamically changing situation. The robot moves along the course with the aid of surrounding sensors and a scanning procedure. The robot's sensors allow it to survey an area from a 360-degree angle.

The proposed method, NO-MOPP, begins by analysing sensor data to identify any new static or moving dynamic obstacles on the optimal path for tracking. Once an obstacle reaches the robot sensor's exposure range, sensor readings provide all information regarding the robot's movement and location, as well as all obstacles in the surroundings. Using this information, the likelihood of a robot and obstacle collision is evaluated. If there is no collision, the robot will continue along its original path as shown in Fig. 4(a).

However, in the case of a collision, the proposed NO-MOPP technique replans the segment of the path containing a potential collision location. The newly found subpath must be the best and shortest possible. The inventive optimal path was adjusted such that the-robot will track the updated no-collision path. The process of alerting a robot to the presence of new obstacles is known as obstacle detection.

The proposed methodology enables the robot to-move toward the goal though detecting any new obstacles. The following stages are included in this implementation. The measurements of the sensor are recorded. The robot-obstacle distance is then estimated from the robot to the close exterior-surface of the obstacle.

3) Obstacle detection: The obstacle's presence is detected when sensors sense the obstacle. The obstacle-robot distance exceeds a certain threshold. The collision check method is invoked when an obstacle is encountered to find that the robot as well as the obstacle will crash.

4) Collision check: Even if an obstacle is within the sensor range of the robot, not all detected obstacles will cause a collision. If there is no collision as depicted in Fig. 4(a), the robot will follow the optimal reference path. If there is a collision possibility as depicted in Fig. 4(b) and the distance amid the robot & the obstacle is greater-or-equal to the threshold value, then replanning is performed using the proposed local search method called NO-MOPP; otherwise, the robot will halt and pause for the obstacle to go before continuing along the same path.

The likelihood of a collision is computed based on the robot's location and direction angle in relation to the sensed obstacle. Calculating the time and location of the collision is: If the robot's present location is P_{r-p1} (x_{r-p1}, y_{r-p1}) and it is heading toward P_{r-p2} (x_{r-p2}, y_{r-p2}), and if the obstacle's present location is P_{obst1} (x_{obst1}, y_{obst1}) and its goal is P_{obst2}(x_{obst2}, y_{obst2}). The formulas of the robot movement are given by (7).

Calculating robot motion requires the following formula:

$$x_{r-p} = x_{r-p1} + v_{robot}\, t_r \cos \theta$$

$$y_{r-p} = y_{robot1} + v_{robot}\, t_r \sin \theta \quad (7)$$

The equation for the movement of the obstacle could be expressed as (8).

$$x_{obst1} = x_{obst1} + v_{obst}\, t_{obst} \cos \varphi$$

$$y_{obst1} = y_{obst1} + v_{obst}\, t_{obst} \sin \varphi \quad (8)$$

The collision amid the robot and the moving obstacle occurs when the subsequent Eq. (9) is satisfied.

$$x_{robot} = x_{obst1} \quad (9)$$

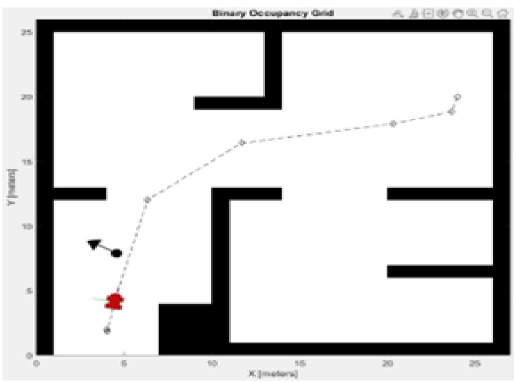
$$y_{robot} = y_{obst1}$$

i, e.

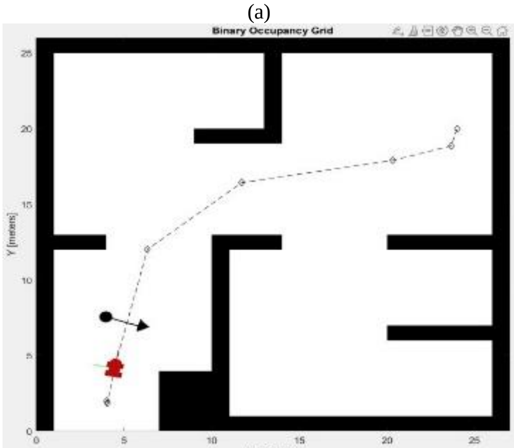
$$x_{r1} + v_r\, t_r \cos \theta = x_{obst1} + v_{obst}\, t_{obst} \cos \varphi \quad (10)$$

$$y_{r1} + v_r\, t_r \sin \theta = y_{obst1} + v_{obst}\, t_{obst} \sin \varphi$$

where θ is orientation and v_{robot} is the robot-velocity, φ and v_{obst} are the orientation and the obstacle-velocity respectively, and tr and t_{obst} are the current time at which the robot and obstacle are there and they are positive. Then the point of the intersection can be determined by replacing Eq. (7) and (8) with Eq. (9) producing equation (10). In the case of a collision, the robot's sub-path consists of 3 locations. The robot's present location is represented by the first point, $X_{present}$ (x_1, y_1), the collision points by X_{colisn} (x_c, y_c), and the next point in its path by X_{new_next} (x_n, y_n). If there is a collision, the obstacle will be on this sub-path. For this reason, a modified local search is proposed, NO-MOPP, for replanning in order to locate an alternate collision-free sub-path. When both of the below circumstances are true Local replanning is done by invoking the proposed method NO-MOPP. The first circumstance is when a new obstacle enters the robot sensor's coverage range. Second, when a collision detection judgment is made favourably. In this local search, NO-MOPP looks for the next neighbour with minimum cost value calculated using Eq. (1) in the Global path planning from its neighbor's list, named Open list. The computational cost is less because the next node is selected from the open list which is readily available. Therefore, the replanning has no impact on the time efficiency of the proposed technique. Once again, the new neighboring point is checked for collisions; if there are none, a minimal turning radius path segment is constructed. This is repeated until the goal is found.



(a)



(b)

Fig. 4. (a) No collision Case (b) Collision case

The shortest path is guaranteed by a Minimum turning radius path segment. The near-optimal sub-path is returned by the proposed method NO-MOPP. The robot is instructed to pursue this new branch of the course. Because the Global-path-planning of the proposed dynamic path-planning system previously computed the costs of five neighbors corresponding to each steering angle for each location on the global optimum path, the cost computing in the proposed method NO-MOPP is no longer required. There is therefore no computational overhead. The time needed for this is likewise quite short because there are just five neighbour points with regard to five steering angles while looking forward and only those points for which the arc length is greater than the cell size Eq. (4) So, search time is also reduced. Therefore, the method has no consequence on the efficiency of time. The following Algorithm 2 summarizes the pseudo-code of the proposed dynamic path planning algorithm NO-MOPP:

Algorithm 2: NO_MOPP
Input: Start, Goal, Closed-list and Open-list
Output: The near-optimal-path for an autonomous mobile-robot fromthe given initial point to the goal-position in Path length and Execution time.

Algorithm NO_MOPP (Open_List, Closed_List, X_{start}, X_{goal})
X_{next}=X_{start}
index=0
optimal_path =Closed_List
Pathlength= length (Optimal-path)
Robot follows optimal_path
for each point in Closed_List ()
{
index=index+1
Robot move forward in the mentioned orientation
X_{next} = Closed_List(index)
if X_{next} ==Xgoal
Print ("Path detected successfully")
Print ("Found near Optimal path", Closed_List);
Sensor data= Sensor reading from its coverage area
If (Sensor data != Obstacle)
Robot moves forward to the X_{next}.
Else if (Collision_detection () == false)
Robot moves forward to the X_{next}
Else if
{
$X_{current}$ = X_{next}
While (neighbours of $X_{current}$ from Open_List ==true)
{
X_{new_next}=$X_{current}$'s neighbours with next minimum cost from Open_List ()
X_{new_next}=$X_{current}$ having next minimum cost from Open_List ()
If Collision_detection (X_{new_next}) == false
Robot moves to X_{new_next} with the given orientation
Update Closed_List
Return
Else
Continue
}
update path length
If (no more neighbors of $X_{current}$ in Open_List)
Print ("There is no path exists")
}
Method: Collision_detection (X_{new_next})
{
d= distance between the robot's current point and collision point
if (d >= threshold value)
return false
else
return true
}

IV. Experimental Results

Using MATLAB 2021a on Windows 10 64bit-with an Intel-core i5 NIVIDIA G5-CPU, the performance of the proposed method, a novel Near-Optimal Multi-Objective Path Planner (NO-MOPP), was evaluated. For testing, sensor views from -45° to +45°, -90° to +90°, and -180° to +180° were captured. In ninety percent of the studies, it was discovered that the dynamic obstacles were successfully avoided. On a variety of dynamic maps, all simulations were executed with varying starting and ending points. Path-Length-Mean and Average-Execution-Time between the provided start and goal locations on the map are the performance metrics considered when evaluating the effectiveness of our proposed system. The execution time and length of the near-optimal-path was recorded.

1) Case-study-1: Complex Map: A more complex dynamic environment is considered; the environment is a complex maze with both static and dynamic obstacles that are stationary and moving. The proposed method was executed one hundred times in order to determine its average execution time. Fig. 5(a), (b) and (c) depicts a nearly optimal path generated by NO-MOPP for this complex environment on two levels known as Global path planning and Local-replanning paths and how robot (arrow) avoiding dynamic stationery(purple color) and moving obstacle(green color). In ninety percent of experiments, it successfully avoids dynamic obstacles.

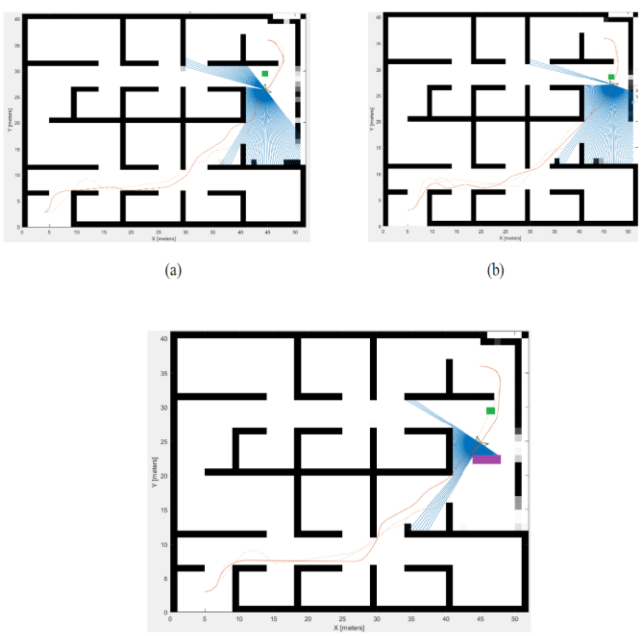


Fig. 5. (a) & (b): No collision and collision avoidance with a moving dynamic obstacle (green color) using NO-MOPP. (c) Collision avoidance with dynamic moving (green color) and stationary obstacles (purple color) using NO-MOPP

The findings for the Complex map's Path Length Mean and Average-Execution-Time are given in below Table I.

TABLE I. Type Results for Complex Map

	Offline path planning	After Local-replanning
Path length_mean in meters	64.5589	67.8374
Execution time in secs	0.457936	1.986376
Direct_Path_Length in meters	51.856	51.856

2) Case-study-2: Package pickup in Warehouse scenario: A package pickup in a warehouse dynamic-environment contains static and dynamic moving and stationary obstacles. Fig. 6(a), (b), and (c) illustrate, NO-MOPP devised a path that was close to optimal for avoiding static (purple color) and moving (green color) obstacles on the way to the package pickup site. In 90% of the experiments, the proposed strategy was effectively avoided.

The Path length mean and execution time outcomes for the Package Pickup scenario are summarized in Table II.

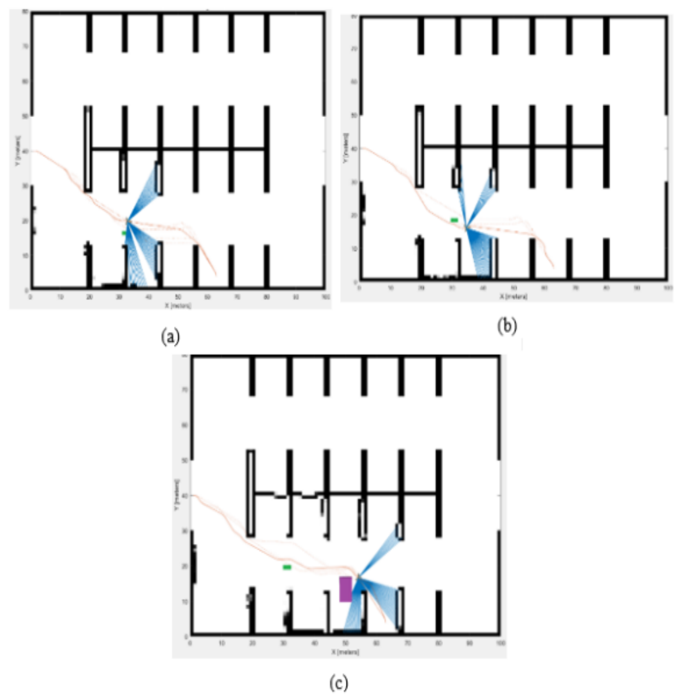


Fig. 6. (a) & (b): No collision, collision avoided with a moving dynamic obstacle (green color) using NO-MOPP. (c) The collision was avoided with the moving dynamic obstacle (green color), and stationery obstacle (purple color) using NO-MOPP

TABLE II. Results for Package Pickup in a Warehouse Scenario

	Global path planning	After Local-replanning
Path length_mean in meters	73.5541	91.5263
Execution time in secs	0.4951	2.953216
Direct_Path_Length in meters	62.1488	62.1488

Performance Evaluation: In the preceding two case studies, we describe the efficacy of the proposed method NO-MOPP and comparability it to the current techniques RRT and RRT*. In this section, we compare NO-MOPP to RRT and RRT*. The proposed method NO-MOPP has been rigorously examined. The comparison of performance is summarised here. In comparison to RRT and RRT*, the path length supplied by the Proposed Method NO-MOPP was superior. The execution time of No-MOPP is considerably shorter.

1) Case Study-1: Complex Dynamic map: Table III provides a visual representation of the results of a performance evaluation that compares the proposed technique NO-MOPP to the existing methods RRT and RRT* when applied to a complex dynamic map. The figures provide abundant evidence that the proposed technique is successful even when the level of map complexity increases.

TABLE III. PERFORMANCE EFFICIENCY COMPARISON IN A COMPLEX MAP

Planner/performance metric	Path length Mean in meters	Avg_Execution Time in secs
RRT	110.873295	4.136285
RRT Star	105.567284	4.513792
Proposed method NO-MOPP	67.8374	1.986376

As is evident from the Fig. 7 below, the proposed method NO-MOPP performed better than the existing methods RRT and RRT* after being run through 100 iterations. The blue line that depicts its performance shows that the proposed method developed the shortest length paths in contrast to other existing approaches in the Complex Dynamic Map. This is proved by the fact that the method created the shortest length paths in each iteration.

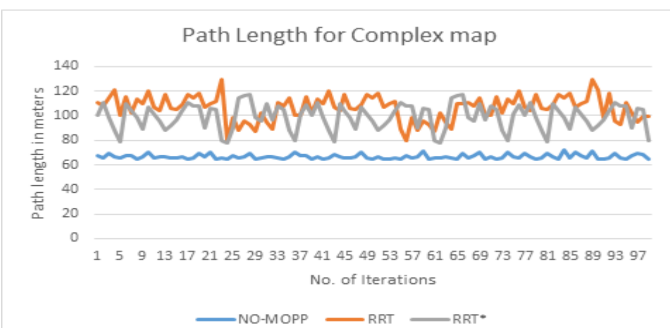


Fig. 7. Path length comparison in complex map

The Fig. 8 displays the efficiency and efficacy of the execution time. The proposed method NO-MOPP is represented by a blue line in virtually all 100 iterations of the Complex Dynamic Map. This method generates paths that require significantly less time consuming compared to the existing techniques RRT and RRT*.

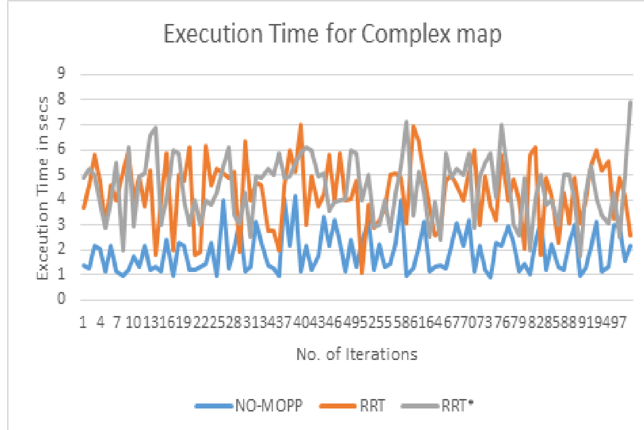


Fig. 8. Execution time comparison in complex map

2) Case Study-2: Dynamic map for Package Pickup: The RRT and RRT* approaches, as well as the proposed approach, were evaluated using the Warehouse scenario as a point of comparison. Table IV presents a comparison of various performance metrics for further consideration. Examining how the proposed solution comes up in path length as well as execution time in comparison to the other available options. The proposed method is carried out fairly well when applied to this difficult warehouse map.

TABLE IV. PERFORMANCE EFFICIENCY COMPARISON IN A COMPLEX MAP IN WAREHOUSE

Planner/performance metric	Path length Mean in meters	Avg_Execution Time in secs
RRT	170.8753	4.596832
RRT Star	150.3547	4.975649
Proposed method NO-MOPP	91.5263	2.953216

The Fig. 9 indicates how the performance of the recommended method NO-MOPP compares to that of the existing methods RRT and RRT* in 100. When compared to the other existing methods in the Warehouse dynamic map, the proposed method NO-MOPP consistently created paths that were the shortest in length. This can be seen from the blue line that depicts the performance of the proposed methodology.

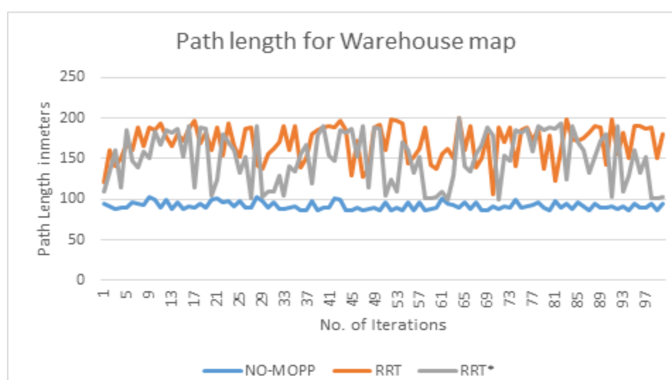


Fig. 9. Performance metric path length comparison in warehouse map

The results of a comparison between the efficiency of our method NO-MOPP and the existing techniques RRT and RRT* can be found in the following Fig. 10, which can be found below. The strategy that was suggested consistently produced the least time-consuming paths when compared to other approaches that were already being used for the dynamic map of the Warehouse.

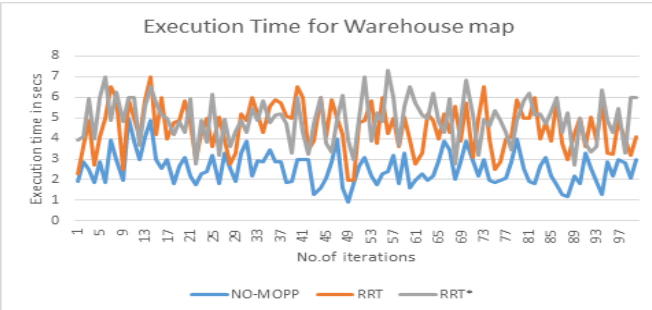


Fig. 10. Performance metric execution time comparison in warehouse map

V. Conclusion and Future Work

In this paper, A Near-Optimal Multi-Objective Path Planner (NO-MOPP) is used to determine the optimal path for mobile autonomous robots operating in dynamic environments. While complying with the robot's kinematic constraints, the robot is able to follow the determined path and avoid new obstacles. Detection of obstacles is dependent on the coverage area of the sensors of a robot and an assessment of the likelihood of a collision. Collision avoidance is achieved through re-planning if the collision check method indicates a collision with a newly introduced dynamic obstacle. Consequently, the development of smooth, drivable paths, which are required for realistic scenarios, is ensured. On average, smoother, collision-free, near-optimal paths may be discovered 90% of the time. NO-MOPP accomplishes Multiple Objective Optimization, which includes Path length, Execution Time, Cost function, and Path Smoothing. Based on the preceding experiments, it is obvious that applying the proposed technique reduced average 15% path length and execution time compared to the existing methods RRT and RRT*. Compared to these existing methods, the proposed method has exhibited superior performance efficiency in complex settings.

Future studies may extend the NO-MOPP method to dynamic situations with higher dimensions for real-time autonomous robots and autonomous vehicles.

References

[1] T. Lozano-Perez and M. A. Wesley, "An algorithm for planning collision-free paths among polyhedral obstacles," Communications of the ACM., vol. 22, no. 10, pp. 560-570, Oct. 1979.

[2] Lozano-Perez, T., "Spatial Planning: A Configuration Space Approach", IEEE Transactions on Computers, Vol. C-32, No. 2, February 1983.

[3] Hart, P.; Nilsson, N.; Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths, IEEE Transactions on Systems Science and Cybernetics 4(2), pp. 100-107, 1968

[4] Dudi, T.; Singhal, R.; Kumar, R. Shortest Path Evaluation with Enhanced Linear Graph and Dijkstra Algorithm. In Proceedings of the **2020** 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), **2020;** pp. 451–456.

[5] Ammar, A.; Bennaceur, H.; Chaari, I.; Koubaa, A.; Alajlan, M. Relaxed Dijkstra and A* with linear complexity for robot path planning problems in large-scale grid environments. Soft Computing - A Fusion of Foundations, Methodologies, and Applications Volume 20Issue 10Oct **2016** pp 4149–4171 https://doi.org/10.1007/s00500-015-1750-1.

[6] Qing, G.; Zheng, Z.; Yue, X. Path-planning of the automated guided vehicle based on improved Dijkstra algorithm. In Proceedings of **the 2017** 29th Chinese control and decision conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 7138–7143.

[7] Syed Abdullah, F.; Iyal, S.; Makhtar, M.; Jamal, A.A. Robotic Indoor Path Planning Using Dijkstra's Algorithm with Multi-Laye Dictionaries. In Proceedings of the 2nd International Conference on Information Science and Security (ICISS), Seoul, Korea, 14–16 December 2015.

[8] Sariff, N.; Buniyamin, N. Ant colony system for robot path planning in a global static environment. In Proceedings of the 9th WSEAS International Conference on System Science and Simulation in Engineering (ICOSSSE'10), Takizawa, Japan, 4–6 October 2010; pp. 192–197.

[9] Dang, C.V.; Ahn, H.; Lee, D.S.; Lee, S.C. Improved Analytic Expansions in Hybrid A-Star Path Planning for Non-Holonomic Robots. Appl. Sci. **2022**, 12, 5999. https://doi.org/10.3390/app12125999.

[10] I. Martin, S. Parkes, and M. Dunstan, "Modeling cratered surfaces with real and synthetic terrain for testing planetary landers," IEEE Trans. Aerospace Electronics Systems, vol. 50, no. 4, pp. 2916–2928, Oct 2014.

[11] X. Ning and L. Liu, "A two-mode INS/CNS navigation method for lunar rovers," IEEE Trans. Instrumentation and measurement, vol. 63, no. 9, pp. 2170–2179, Sep. 2014.

[12] Miao, H. Robot path planning in dynamic environments using a simulated annealing-based approach. Brisbane: Queensland University of Technology. 10th International Conference on Control, Automation, Robotics and Vision 2008.

[13] Damle, Vikas P., and Seba Susan. "Dynamic Algorithm for Path Planning using A-Star with Distance Constraint." In **2022** 2nd International Conference on Intelligent Technologies (CONIT), pp. 1-5. IEEE, 2022. DOI: 10.1109/CONIT55038.2022.9847869

[14] Siegwart, R., & Nourbakhsh, I. (2004). Introduction to autonomous mobile robots (1st ed.). London: MIT Press.

[15] LaValle, S.M. Planning Algorithms; Cambridge University Press: Cambridge, UK, 2006.

[16] Wang Siming; Zhao Tiantian; Li Weijie, Mobile Robot Path Planning Based on Improved Artificial Potential Field Method, IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE), 2018.

[17] LaValle, S.M. Rapidly-Exploring Random Trees: A New Tool for Path Planning. 1998. Available online: http://lavalle.pl/papers/ Lav98c.pdf.

[18] Karaman S, Walter M, Perez A, et al. Anytime motion planning using the RRT∗. IEEE International Conference on Robotics and Automation (ICRA); 2011

[19] L. Kavraki, P. Svestka, J-C. Latombe, M.H. Overmars, Probabilistic roadmaps ˇ for path planning in high-dimensional configuration spaces, IEEE Trans. on Robotics and Automation 12 (1996), pp. 566–580.

[20] A. Stentz. The focussed D* algorithm for real-time replanning. IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2, August 1995 Pages 1652–1659.

[21] Arora, T., Gigras, Y., & Arora, V. (2014). Robotic path planning using genetic algorithm in a dynamic environment. International Journal of Computer Applications, 89(11), 9–12.

[22] Zhang, X., Zhao, Y., Deng, N., & Guo, K. Dynamic path planning algorithm for a mobile robot based on visible space and an improved genetic algorithm. International Journal of Advanced Robotic Systems, **2016**. 1, 1–17.

[23] Khulna. Karami, A., & Hasanzadeh, M. An adaptive genetic algorithm for robot motion planning in 2D complex environments. *Computers and Electrical Engineering*, 2015, *43*, 317–329. https://doi.org/10.1016/j.compeleceng.2014.12.014.

[24] Islam, R., Muftee, H., & Mahfuzul Hossain, S. (2014). Autonomous robot path planning using particle swarm optimization in a dynamic

environment with mobile obstacles & multiple targets. In International conference on mechanical, industrial and energy engineering (pp.1 – 6).

[25] Ou, Y.; Fan, Y.; Zhang, X.; Lin, Y.; Yang,W. Improved A* Path Planning Method Based on the Grid Map. Sensors **2022,** 22, 6198. https:// doi.org/10.3390/s22166198.

[26] Akshay Kumar Guruji, Himansh Agarwal, D. K. Parsediya, Time-Efficient A* Algorithm for Robot Path Planning, 3rd International Conference on Innovations in Automation and Mechatronics Engineering, **2016**. doi: 10.1016/j.protcy.2016.03.010.

[27] Hong-Mei Zhang, Ming-Long Li and Le Yang, Safe Path Planning of Mobile Robot Based on Improved A* Algorithm in Complex Terrains, Algorithms **2018,** MDPI, Sensors doi:10.3390/a11040044

[28] Khulna. Karami, A., & Hasanzadeh, M. An adaptive genetic algorithm for robot motion planning in 2D complex environments. Computers and Electrical Engineering, 2015, 43, 317–329. https://doi.org/10.1016/j.compeleceng.2014.12.014.

[29] Pham, D. T., & Castellani, M. (2015). A comparative study of the Bees Algorithm as a tool for function optimization. Cognet Engineering, 1–28.

[30] Choset, H.M. Principles of Robot Motion: Theory, Algorithms, and Implementation; The MIT Press: Cambridge, MA, USA, 2005.

[31] Noreen, I.; Khan, A.; Habib, Z. Optimal Path Planning using Memory Efficient A*. In Proceedings of the IEEE International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 19–21 December **2016**; pp. 142–146.

[32] Warren, C.W. Fast path planning using modified A* method. In Proceedings of the IEEE International Conference on Robotics and Automation, Atlanta, GA, USA, 2–6 May 1993.

Eye-Vision Net: Cataract Detection and Classification in Retinal and Slit Lamp Images using Deep Network

Binju Saju[1]*, Rajesh R[2]
Research Scholar[1], Associate Professor[2]
Department of Computer Science, Christ (Deemed to be University), Bengaluru, Karnataka 560029, India

Abstract—**In the modern world, cataracts are the predominant cause of blindness. Early treatment and detection can reduce the number of cataract patients and prevent surgery. However, cataract grade classification is necessary to control risk and avoid blindness. Previously, various studies focused on developing a system to detect cataract type and grade. However, the existing works on cataract detection does not provide optimal results because of high detection error, lack of learning ability, computational complexity issues, etc. Therefore, the proposed work aims to develop an effective deep learning techniques for detecting and classifying cataracts from the given input samples. Here, the cataract detection and classification are performed using two phases. In order to provide an accurate cataract detection, the proposed study introduced Deep Optimized Convolutional Recurrent Network_Improved Aquila Optimization (Deep OCRN_IAO) model in phase I. Here, both retinal and slit lamp images are utilized for cataract detection. Then, the performance of these two image datasets are analysed, and the best one is chosen for cataract type and grade classification. By analysing the performance, the slit lamp images attain higher results. Therefore, phase II uses slit lamp images and detects the type and grade of cataracts through the proposed Batch Equivalence ResNet-101 (BE_ResNet101) model. The proposed classification model is highly efficient to classify the type and grades of cataracts. The experimental setup is done using MATLAB software, and the datasets used for simulation purposes are DRIMDB (Diabetic Retinopathy Images Database) and real-time slit lamp images. The proposed type and grade detection model has an accuracy of 98.87%, specificity of 99.66%, the sensitivity of 98.28%, Youden index of 95.04%, Kappa of 97.83%, and F1-score is 95.68%. The obtained results and comparative analysis proves that the proposed model is highly suitable for cataract detection and classification.**

Keywords*—*Cataract detection; grade classification; CRNN; dense CNN; Aquila optimization; BE-ResNet101

I. Introduction

The eyes are the major organs that offer a clear vision to humans and living organisms. A cataract is a severe eye disorder that affects the original vision of the eyes and produces vision distortion. A cataract is the major cause of vision impairment or blindness worldwide [1-2]. Vision or Visual Impairment (VI) in humans refers to the inability to perceive clear vision. A global survey by the WHO (World Health Organization) estimated that among 285 million people, 39 million are blind, and the remaining are reported to other VIs [3]. Some of the retinal disorders are Cataract, DR (Diabetic Retinopathy), Glaucoma, AMD (Aged Macular Degeneration), RP (Retinitis Pigmentosa), [4-5] etc. Cataracts are the leading cause of vision loss in aged and younger individuals [6-7]. The risk factors related to cataracts are people over 40 years, uncontrolled diabetes, steroid usage, family history, trauma in the eye, UV (Ultra-Violet) light exposure, [8] etc.

Early cataract diagnosis can control global vision loss. Based on the location and development, the cataract is divided into different types, namely NC (nuclear cataract), CC (cortical cataract), and PSC (posterior subcapsular cataract) [9]. The severity of the cataract can be identified with the grading procedure. Eye diseases are often diagnosed using several methods such as slit-lamp images [10], VA (visual acuity) examination [11], digital photography [12], retinal images [13] and ultrasonic images [14-15]. Ophthalmologist utilizes retinal and Slit-lamp images to detect the presence and absence of cataracts [16-17]. In ophthalmology, the screening and detection of cataracts have produced robust outcomes with deep learning (DL) models [18-20]. DL offers significant advantages by using several HL (Hidden Layer) sequences to extract useful image features. Moreover, in the medical field, DL learns the image features effectively and automatically. The advent of computer infrastructure has contributed to the faster adoption of DL due to its high level of feature extraction, processing of huge amounts of data and accurate classification.

Motivation: The eyes are the sensory organs that view the elegance of our surrounding environment. In recent years, the research organization are focussed on developing a robust cataract detection and classification methodology to assist affected people. From the recent studies, it is observed that deep learning techniques are attaining great attention for cataract detection. By using different types of datasets, the deep learning approaches gain better results in detecting and classifying cataract diseases. Many existing studies exhibit that using slit-lamp images produces high performance because of reduced cost, easy of maintain, flexibility etc. The neural network based feature extraction techniques can detect the features automatically. Also, such kind of feature extraction methods directly enhances the classification accuracy. Thus, it motivates the author to establish a deep learning based method for cataract detection and grade classification using slit lamp images. The major contribution of proposed work is given as,

- To develop an effective cataract detection and classification in retinal and slit-lamp images by

*Corresponding Author

proposing a new Deep Optimized Convolutional Recurrent Network (Deep OCRN).

- To reduce the detection error rate, the presented loss in the proposed Deep OCRN model is minimized by updating the weight parameters through Improved Aquila Optimization (IAO) algorithm.
- To extract the required features and to effectively identify the cataract type by proposing Dense CNN.
- To develop Batch Equivalence ResNet-101 method for classifying the grades of cataracts.
- To evaluate the performance of the proposed DL model by comparing it with other state-of-art classifiers.

The rest of this paper is structured as follows: Section II discussed recent existing studies on cataract detection, and Section III presents the proposed cataract detection classification model grade classification model. Section IV represents the results and discussion of the proposed study, and Section V deals with the overall conclusion and future scope for further research.

II. Related Works

Some recent research studies which are carried out on cataract detection and classification are listed as,

Pratap and Priyanka [21] presented a computer-aided robust cataract diagnosis using DNN (Deep Neural Network) in retinal fundus images. This diagnosis model involves CFE (Combined Feature Extraction) using dual fine-tuned independent DCNNs (Deep Convolutional Neural Networks). Moreover, multi-class SVM (Support Vector Machine) classifiers were trained on the noise levels from 0–25. The developed model obtained higher accuracy of (93.49%). However, the performance is affected due to noise problems.

Akil et al. [22] introduced the deep learning based CNN model for detecting retinal abnormalities in retinal fundus images. The ensemble learning based DR (Diabetic retinopathy) detection attained 97.7% accuracy and, with the grading achieved 98.5% accuracy. The main drawbacks were caused due to the limited number of training images and the lack of earlier screening.

Hu et al. [23] developed a unified automated NC (Nuclear Cataract) grading in slit-lamp smartphone images. The DL YOLOv3 model was used to localize the ocular lens nuclear region. Moreover, the combination of ShuffleNet and SVM was used in grading cataract severity by measuring the gray conjugate features of the core region. The overall performance obtained was accuracy (93.5%), F1 (92.3%), AUC (0.9198), and Kappa (95.4%). The drawbacks were high computational complexity and convergence loss.

Liu et al. [24] presented the diagnosis framework based on the localization of paediatric cataracts in slit-lamp images using deep feature modelling of CNN. The classification was completed with the combination of SVM and Softmax classifier. The overall classification accuracy was 97.07%. The Performance is not as much effective due to an inappropriate extraction of features.

Zhang et al. [25] developed automatic detection and grading of cataracts using DCNN (Deep Convolutional Neural Network). This DL framework DCNN was used to automatically examine cataracts by visualizing the feature maps with empirical higher-order semantic meaning at the pool5 layer. The overall accuracy performance obtained with cataract detection was (93.52%), and the grading of cataracts was (86.69%). The major limitation was classification and grading were analysed only on the fundus retinal images.

Problem statement: Earlier and accurate diagnosis of cataracts is more essential to limit the risk categories and prevent blindness. In existing, different techniques are developed to afford accurate cataract detection. However, they face several challenges while detecting cataracts from the given input samples. The detection accuracy of existing methods are reduced due to an enhanced computational complexity problem and large number of parameters in the detection models. By concerning this, the proposed study optimally selects the needed parameters using an effective algorithm. In addition, to make effective detection, extraction and selection of optimal features are more important. But, the existing detection techniques cannot extract the relevant features because of reduced efficiency. So, that the existing methods are not as suitable for detecting cataracts from the provided inputs. Also, many of the existing studies prefer retinal images for detecting cataracts and only few of them uses slit lamp images. To fulfil this gap, the proposed study used both retinal and slit lamp images for detecting cataracts. Moreover, robust DL models are utilized in the proposed study to learn appropriate features and make the system to attain higher detection accuracy with reduced complexity.

III. Proposed Methodology

The proposed detection and classification of cataract images involves two phases. In phase 1, the detection of cataracts using both retinal and slit-lamp images are performed. Initially, image acquisition is enabled for collecting retinal and slit-lamp images from the DRIMDB data source and real-time slit lamp images. Next, pre-processing is enabled to refine the quality of images, which involves image resizing, color conversion, removal of uneven illumination, quality selection and image enhancement. Here, score based image quality selection is performed using Hybrid NIQE-PIQE, and image enhancement is processed with Improved Mean Adjustment (IMA). To reduce the complexity and processing time, the proposed study extracts different features like shape, wavelet and texture. The relevant shape features are initially extracted through Pyramidal HOG (Histogram Oriented Gradient). Then, wavelet features are extracted using Haar Wavelet Transform (HWT) and the texture features are extracted using GLCM (Gray Level Co-occurrence Matrix). In order to reduce the high feature dimensionality, selecting an optimal feature set is essential. Thus, Relief Neighbourhood Component Analysis (RNCA) approach is proposed. Finally, the cataract detection is performed using Deep Optimized Convolutional Recurrent Network_Improved Aquila Optimization (Deep OCRN_IAO). Here, the loss in the network is optimized using Improved Aquila Optimization (IAO) and through this learning ability is increased. Moreover, the detection performance can be compared with other DL models to validate the proposed performance. Fig. 1 represents the proposed methodology of the proposed work in both phases.

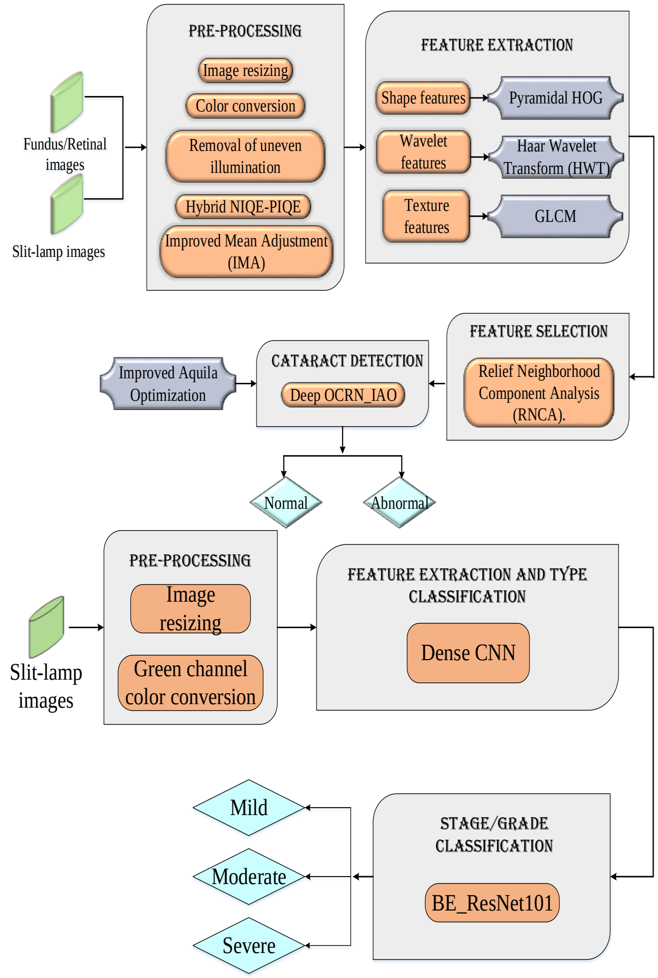


Fig. 1. Block diagram of proposed methodology

In phase 2, the slit-lamp cataract type and stage classification are done. Initially, image resizing and green channel color conversion are performed to enhance the image quality. Next, feature extraction and type classification can be performed using the Dense Convolutional Neural Network (Dense CNN) based deep learning framework. Finally, the classification of stages or grading as mild, moderate and severe can be attained using Batch Equivalence ResNet-101 (BE_ResNet101).

A. Image Acquisition

Image acquisition is the initial step for cataract detection, where various images are collected from different sources. The proposed work collects the retinal and slit-lamp images from popular datasets like DRIMDB and real-time slit lamp images. Fig. 2 shows some of the input images acquired from the dataset.

Fig. 2. Original input images

B. Pre-processing

Pre-processing an image is a significant step because it affords better classification results. The proposed pre-processing stage aims to refine the image quality to obtain effective performance. The images are pre-processed by,

- Image resizing
- Color conversion
- Uneven illumination removal
- Image enhancement
- Quality selection

In the proposed work, the high resolution images of the varied pixel are resized into 224*224 pixels for easy processing. The proposed work used green channel conversion in the pre-processing stage. Due to the presented uneven illumination and the available reflection of the eyes, the accurate classification of the different grades of cataracts is more difficult. Thus, the uneven illumination is removed in the pre-processing stage to enhance the performance of grade classification.

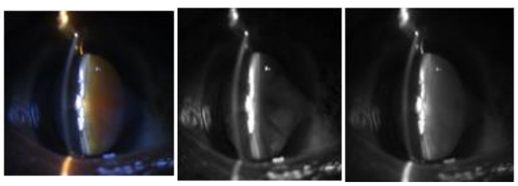
Image enhancement using Improved Mean Adjustment (IMA): Image quality enhancement plays a vital role in each algorithm. The poor quality of images lead to performance degradation of a specific classifier. Therefore, it is necessary to enhance the image quality during the phase of training and testing. This method is performed to elevate the image contrast. Here, the original input image I^O is utilized to generate the enhanced I^E image. The original image is split into 3×3 blocks of non-overlapping, and for each block mean is calculated. The highest mean value is mentioned as $\mu_{\max}$ and the lowest mean is indicated as $\mu_{\min}$. When the variations among $\mu_{\max}$ and $\mu_{\min}$ is small, then it is represented that the image contains a poor contrast. Hence, it can be enhanced by,

$$I^E(k,l) = T\left(I^O(k,l)\right) \tag{1}$$

Where, $$T(f) = \frac{f \times s}{\mu_{\max} - \mu_{\min}} - \frac{\mu_{\min} \times s}{\mu_{\max} - \mu_{\min}} \tag{2}$$

Here, the gray scale value at $I^{O}(k,l)$ is mentioned as f and the maximum gray scale value in the given image is represented as S.

Quality selection using hybrid NIQE-PIQE: The module of image quality selection is performed to filter out the improved quality of images for cataract diagnosis. The proposed work selects the image quality through the NIQE-PIQE model [21]. In NIQE-PIQE, the score value is evaluated for each image in which a low score is considered the best quality of image and the high score of images is assumed to be the worst quality. By [26] and [27], the higher values of NIQE-PIQE for the best image quality are observed to be five and 50 correspondingly. NIQE-PIQE is evaluated on the presented dataset, and the quality selection is performed based on the threshold. Fig. 3 shows the pre-processed images in the proposed study.



(a) (b) (c)

Fig. 3. Pre-processed images (a) Resized image (b) Output image of color conversion (c) Enhanced image

C. Feature Extraction

When the pre-processed images are directly subjected to the classification stage, the system faces several problems in diagnosing cataract disease and also it degrades the performance. Thus, extracting suitable features is essential to attain optimal detection results. The features that extracted by proposed feature extraction stage is described as,

1) Extraction of shape feature through pyramidal HOG: Pyramidal HOG is utilized for the object retrieval process and is highly suitable for extracting shape features from the input samples. In existing several methods are used to extract the shape features but they are not as much effective. The proposed method separates a localized and tracked region into several cells at varying pyramid levels. Depending the direction of the image gradient, the pyramidal HOG is termed an image descriptor. The gradient direction $\theta(a,b)$ and gradient magnitude $m(a,b)$ is evaluated in each cell on every pixel. It is mentioned as,

$$m = \sqrt{g_a(a,b)^2 + g_b(a,b)^2} \quad (3)$$

$$\theta = \arctan \frac{g_a(a,b)}{g_b(a,b)} \quad (4)$$

Where, $g_a(a,b)$ and $g_b(a,b)$ are indicated by the image gradients with the directions a and b correspondingly. The histogram vector is computed based on the direction and size values in all the cells. A large feature vector is generated by concatenating the histogram vectors of cells together, and the last feature vector is the outcome of the pyramidal HOG approach.

2) Extraction of wavelet features through Haar Wavelet Transform (HWT): The Haar transform is one of the simplest methods in wavelet transforms. This HWT method cross multiplies a function over the Haar wavelet with several stretches and shifts. In various existing studies, the wavelet transform is used in the pre-processed retinal image to contrast with the background and blood vessels. Still, they face some troubles while extracting wavelet features. Thus, the proposed study uses a robust HWT to extract the wavelet features in which the retinal images are dissolved into three levels: reduced contrast vessels, the background, and the vessel mentioning the background. Then, it crosses over the sum and continues to present the upcoming scale. The HWT is expressed in the form of a matrix.

$$B = HIH^{T} \quad (5)$$

Where, I states the images of $M \times M$ matrix, H mentions the Haar transformation of $M \times M$ matrix and B denotes the transform of $M \times M$ matrix that involves the Haar basis functions. Thus, the wavelet features are extracted via HWT.

3) Text feature extraction through GLCM (gray Level co-occurrence Matrix): The texture features correspond to the image spatial organization of pixel values. The proposed model prefers the GLCM method to attain text-based features. Because GLCM is one of the powerful method to extract texture based features. The GLCM approach acquires the specific relationship among two pixels spaced by a certain distance in an image. The GLCM techniques contain several features, where the proposed work extracts the statistical features like Angular Second Moment (ASM), correlation, contrast, entropy (ENT_y), energy (ENE_y), homogeneity (HOM_y), inverse difference moment (IDM) and variance (VAR).

The proposed feature extraction methods extract numerous features, which has the chance to enhance the computational complexity. Thus, to minimize the computational complexity, selecting an optimal feature set is essential.

D. Relief Neighbourhood Component Analysis (RNCA) based Feature Selection

The feature selection process reduces the high data dimensionality by choosing the most effective features. In the proposed cataract detection system, RNCA is utilized for feature selection. This method can make optimal feature predictions through feature weights. The aim of this stage is to diminish the computational complexity, and also it helps to boost the classification accuracy. The proposed RNCA is one of the multi-level feature selection approaches which incorporates both ReliefF and NCA algorithms. The mentioned NCA [28] is a feature selection and dimensionality minimization method and is mainly utilized in varied classification studies. The NCA utilizes training data with varied class labels when selecting the projection that will better isolate classes in the prescribed region. The objective function used in the NCA method is expressed as,

$$\Gamma(w) = \frac{1}{m}\sum_{i=1}^{m} Q_i - \psi \sum_{r=1}^{q} w_r^2 \tag{6}$$

Where, ψ represents the regularization parameter, feature weight is indicated as w_r, total amount of images are given as m, the probability score of the image i is mentioned as Q_i, q signifies dimensionality. When the parameter ψ is randomly selected, each of the feature weights can acquire values that are nearer to 0. The ReliefF approach is a method which can generate an accurate prediction of features. With the aid of feature weights, the features are estimated. Initially, the feature weights are set as zero then it randomly chooses the images from the input at every step. After that, the nearest k images from a similar class and the nearest from a varied class are identified. Then, weights are updated for each feature, and the feature that does not reach the particular condition is ignored from the input. The formulation of the ReliefF approach is represented as,

$$\omega(z^b) = \omega(z^b) - \frac{\sum_{j=1}^{k} diff(F, C_i, Y_J)}{nzk} + \sum_{d \neq class(C_i)} \left[\frac{Q(D)}{1 - Q(class(C_i))} z \sum_{j=1}^{k} \frac{diff(F, C_{i,N_j})}{nzk} \right] \tag{7}$$

Where, z^b mentions the b^{th} feature, the feature set is represented as F, C_i and Y_j indicates the feature set instances and the user chosen parameter is signified as n and k. By this, the optimal set of features is selected using the proposed RNCA, and the selected reduced feature sets are subjected to the Deep OCRN_IAO model for cataract detection. By generating low dimensionality features, the process becomes smooth and reduces computational time.

E. Cataract Detection using Deep OCRN_IAO

This stage detects the cataract from the input images using a hybrid mechanism. Recently, recurrent neural network gained more attention for detecting diseases. In existing, several deep learning models are used to detect cataracts. But they cannot achieve higher detection accuracy due to the reduced learning capability. In order to obtain better detection results, the proposed work hybridizes the convolutional recurrent neural network (CRNN) with IAO algorithm. The major module of CRNN is the recurrent convolutional layer (RCL). The parameter setting of CRNN is given as, total number of epochs-10, maximum iteration-500, iteration per epoch-50 and learning rate-0.001. The RCL units develop against discrete time steps. For a unit placed at (i, j) on p^{th} feature map in RCL, its input is represented as,

$$x_{ijp}(t) = (\omega_p^h)^T v^{(i,j)}(t) + (\omega_p^\beta) z^{(t-1)} + b_p \tag{8}$$

Where, $v^{(i,j)}(t)$ indicate the feed forward input and the recurrent input is mentioned as $z^{(t-1)}$, ω_p^h signifies as feed forward weights, ω_p^β mentions the recurrent weights and bias is represented as b_p. The initial term in equation (19) is utilized in conventional CNN, and the second term represents the recurrent connections. The activity of this unit is mentioned as the function of its input and is given as,

$$z_{ijp}(t) = g\left(r\left(x_{ijp}(t)\right)\right) \tag{9}$$

Where, r specifies the rectified linear activation function. The function of local response normalization is represented as g.

$$r\left(x_{ijp}(t)\right) = \max\left(x_{ijp}(t), 0\right) \tag{10}$$

$$g\left(r_{ijp}(t)\right) = \frac{r_{ijp}(t)}{\left(1 + \frac{\alpha}{M} \sum_{p'=\max(0,\, p-M/2)}^{\min(P,\, p+M/2)} (r_{ijp'})^2\right)^{\eta}} \tag{11}$$

Where, P mentions the overall feature maps in the present layer. It is observed that in equation (23), the sum process over M feature maps at a similar region (i, j), the constant that maintains the normalization's amplitude is mentioned as α and η. The exploding state is avoided by using local response normalization. The dynamic characteristics of RCL are described in equations (20) and (21). The CRNN holds a stack of RCLs with max pooling layers. The initial layer is the feed-forward convolutional layer without the connections of recurrent layers. This convolutional layer generates an activation map and transforms all the pixel values into a single value. The second layer is the max pooling layer in which the feature dimensionality is reduced and minimizes the number of parameters to learn. The third is stacked with recurrent layers, and the next layer is fully connected. In the fully connected layer, each input from one layer is linked to each activation function of the upcoming layer. Finally, the cataract is detected in the softmax layer, which is given as,

$$o_p = \frac{\exp(\omega_p^T z)}{\sum_{p'} \exp(\omega_{p'}^T z)} \qquad (p = 1, 2, \ldots\ldots., D) \tag{12}$$

Where, o_p represents the predicted probability from the p^{th} category and feature vector created by max pooling layer is represented as z. However, the available loss function influences the detection performance of CRNN. Thus, the loss function is minimized by updating the weight parameters using the IAO approach. The available loss function is described as,

$$L_F = \frac{1}{T}\sum_{i=1}^{T} \left(a_i - k_i^{'}\right)^2 \tag{13}$$

Where, T represents the total number of iterations, the actual value is signified as a_i and $k_i^{'}$ mentions the predicted value. Cataracts in the retinal images are detected based on the objective function and are expressed as,

$$Objective\ function = Min[L_F] \tag{14}$$

The proposed IAO method aims to minimize the loss function through weight parameter updation. This is one of the population based approaches and is motivated by the hunting behaviour of Aquila. In order to update the weight parameters, Aquila algorithm used the below formulation.

$$L_1(h+1) = L_{best}(h) \times \gamma(P) + L_r(h) + (x-y) \times rand \tag{15}$$

Where, $L_1(h+1)$ represents the upcoming iteration of the solution h, P mentions the position of the feature, levy flight distribution function is denoted as $\gamma(P)$ and $L_r(h)$ represents the random solution. To enhance the efficiency of Aquila algorithm, weight strategy is introduced.

Weight strategy: According to the maximum iteration, the feature's distribution scope is compressed, generating the search space as small. The weight factor is applied to enhance the different characteristics of features and recover the optimization algorithm from the local optima. This is represented as,

$$\varpi_j = \exp(n\ iter_j / n\ iter_{\max} - 1) \tag{16}$$

Where, ϖ_j denotes the weight at j^{th} iteration, the maximum iteration number is specified as $n\,iter_{\max}$ and the current iteration number is signified as $n\ iter_j$ where, $n\,iter_j \le n\,iter_{\max}$. In this, the objective function is evaluated for each iteration, and the attained result from every iteration is compared. By analysing the best objective function, the optimal solution is obtained. Therefore, the proposed Deep OCRN_IAO model identifies the cataract from the given dataset image. However, determining the grade and severity of the cataract is highly important to cure the disease. Thus, the proposed work used the cataract type classification and grade categorization in phase 2 using effective deep learning techniques.

F. *Cataract Type and Grade Classification using Slit Lamp Images through Dense CNN+BE_ResNet101 Model*

The automatic detection of grading and type is challenging for several existing studies. Thus, the proposed work focussed on developing an automatic cataract grade and type classification with the support of deep learning techniques. Here, the Dense CNN is developed to detect the type of cataracts and BE-ResNet101 is performed to classify the cataract grades. The hyperparameters of the proposed type and grade detection model is described as, maximum number of epochs-10, maximum iteration-500, iteration per epoch-50 and learning rate-0.001.

1) Pre-processing of slit lamp images: Initially, the slip lamp images are pre-processed to enhance the classification performance. The pre-processing steps that carried out on slit lamp images are,

- Resizing of slit lamp images
- Green channel color conversion

After pre-processing the input images, the image quality is enhanced and fed to the Dense CNN method for feature extraction and cataract type categorization.

2) Feature extraction and detection of cataract type using dense CNN: The proposed work is done on feature extraction from the pre-processed images to minimize the classification error. For this purpose, Dense CNN is preferred in this study for feature extraction. Because CNN has the ability to extract features automatically, it helps to produce higher classification results. After feature extraction, the Dense CNN model's final layer classifies the cataract type from the provided slit lamp images. The proposed approach contains a set of dense layers that evaluate each input's weight average and is transmitted to the activation function. The input layer is the initial layer of CNN, where the images from the dataset are set as input. Next, the convolutional layer is the second layer designed by various convolutional kernels. The parameters in the convolutional layer are optimized via a back propagation approach. The major intention of this layer is to extract the most needed features from the slit lamp image, which helps to maintain the complexity. Next, a max-pooling layer is presented which generates feature mapping, and it reduces the feature dimensionality. The information from the pooling layer is passed into the set of dense layers. Here, the neurons of the layer are linked to each neuron of its previous layer. The CNN functionality mainly confides in the activation function, a systematic layer. It is noted that the rate of training can prevent over-fitting problems.

$$h(z)\,K \max(z;0) \tag{17}$$

The functionality of the convolutional layer is described as follows,

$$z_p^l = h\left(\sum_{i \in Np} z_i^{l-1} * K_{ip}^{l-1} + b_p^l\right) \tag{18}$$

Where, p mentions feature mapping, Np represents input mapping, K_{ip} is denoted as filter and bias of feature mapping is specified as b_p. The training process of the network is accelerated using the activation layer, and batch normalization is performed among the activation function and convolutional layer. The cross entropy is analysed in the output layer and is defined as,

$$Loss_{class} = -\sum_{i=1}^{m} 1\{f = i\} \log \frac{e^{xi}}{\sum_{l=1}^{f} e^{xi}} \tag{19}$$

Where, q is denoted as the number of classes, x_i is the final layer and f represents the information of each feature from the input image. During the process of CNN training, the feature extraction is performed, and also it diminishes the cross entropy loss and reconstruction error. In the training process, each kernel is trained to acquire cataract types from the input source. The confidence value is enhanced and fine-tuned by minimizing the cross entropy loss function. Thus, the proposed Dense CNN classifies the types of cataracts precisely.

3) Grade or stage classification of cataracts using BE_ResNet101 framework: The proposed study used the BE-ResNet101 approach for cataract grade categorization. The fundamental ResNet is also a type of CNN model, and the structure of ResNet was inspired by the traditional VGG-19 method. The proposed BE-ResNet101 detects the grade by avoiding the issues of attaining information from unbalanced instances. The presented batch equivalence in ResNet101 helps to improve the cataract grade detection results. Also, the proposed model helps to reduce the degradation issue. The basicResNet-101 is a residual network that contains 101 layers and is the advanced version of the ResNet-50 model. The layers available in the ResNet contain a similar amount of filters for a similar size of output feature map. The amount of filter is twice if the feature map's size is minimized to reduce the time complexity. In the ResNet model, down sampling is directly performed by convolving layers with two strides. This ResNet structure ends with a pooling layer and a fully connected layer with a softmax layer. The objective of distributive layers is to avoid the vanishing gradients issues through reutilizing activation from the previous layer till the closest layer of the current layer has learned its weights.

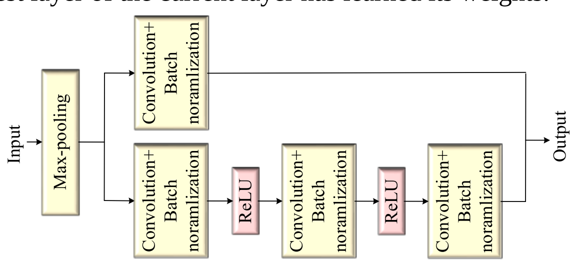


Fig. 4. ResNet-101 structure

The ResNet-101 involves a total of 104 convolutional layers. It contains layers of 33 blocks, and 29 of such blocks utilize the output of the preceding block. It is shown above as remainder connections, and such remainders are used as the initial operand of the addition operator used at the end of all blocks to get the input of the coming blocks. The rested four blocks acquire the output of the preceding block and utilize it in a convolutional layer with 1×1 of filter size and one stride, followed by a batch normalization layer. This layer enables the normalization process, and the attained outcome is transmitted to the addition operator at the output of the block. The structure of ResNet is illustrated in Fig. 4.

Batch equivalence: In the proposed experimental data, the number of images graded from 3.0 to 4.0 is considered the highest. When a mini-batch is designed randomly, the grading model obtains the information from unbalanced occurrences and becomes biased. This issue is solved by applying a strategy of batch balancing. Here, the batch is separated in to five varied groups like [1, 2], [2, 3], [3, 4], [4, 5] and [5, 6]. A similar number of occurrences are randomly chosen during the construction of a mini-batch, i.e., 25 from all the groups and generate the batch fully balanced. Thus, the proposed BE_ResNet101 model classifies and detects the grade of cataracts effectively. Fig. 5 illustrates the cataract grade classified images from the proposed Dense CNN+BE_ResNet101 model.

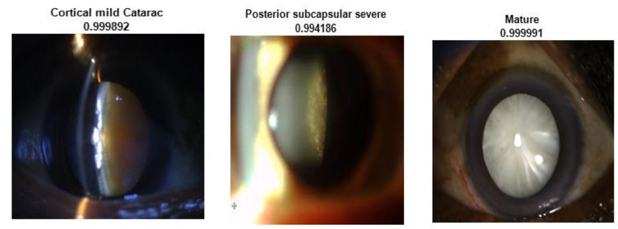


Fig. 5. Classified images

IV. Results and Discussions

This section presents the simulation results of the proposed cataract detection and classification system with clear analysis. The experimental setup was executed through MATLAB software.

A. Dataset Description

The input images utilized in this work are chosen from two different datasets like DRIMDB (Diabetic Retinopathy Images Database)https://academictorrents.com/details/99811ba62918f8e73791d21be29dcc372d660305b and real time slit lamp images. The DRIMDB dataset is utilized in several retinal image analysis studies. In this dataset, a total of 216 consecutive images are taken in the proposed experimental analysis. The real-time slit lamp images are collected from various hospitals containing five types of cataracts: cortical, hyper mature, mature, nuclear and posterior.

B. Performance Metrics

The performance metrics are utilized to evaluate the efficacy of the proposed model. The metrics utilized in the proposed study are expressed below.

$$\text{Accuracy: } A_y = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \tag{20}$$

$$\text{Sensitivity: } S_y = \frac{Tp}{Tp+Fn} \tag{21}$$

$$\text{Specificity: } Sp_y = \frac{Tn}{Tn+Fp} \tag{22}$$

$$\begin{aligned} Youden\ Index &= S_y + Sp_y - 1 \\ &= TPR - FPR \end{aligned} \tag{23}$$

$$F_1 - score = 2\times\frac{\Pr ecision\times Sensitivity}{\Pr ecision + Sensitivity} \tag{24}$$

$$Kappa = \frac{P_{observed} - P_{chance}}{1 - P_{chance}} \tag{25}$$

C. Performance Analysis of Proposed Classification

The performance of the proposed BE_ResNet101 model is analysed by measuring the performance metrics and comparing the attained results with some other approaches. The confusion matrix of the proposed classification is shown in Fig. 6.

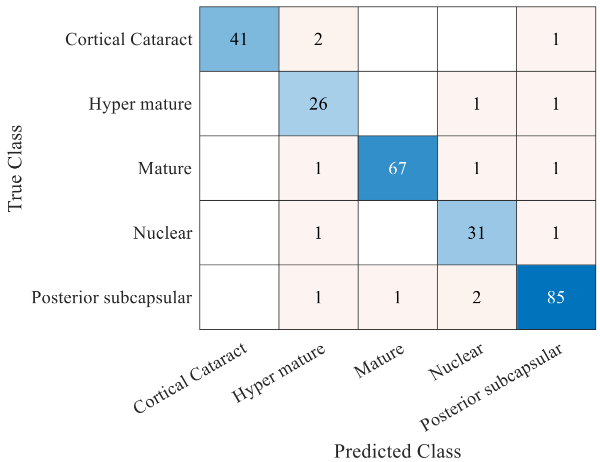


Fig. 6. Confusion matrix of proposed Dense CNN+BE_ResNet101 classification model

The above confusion matrix reveals the efficiency of the proposed model also, it mentions the prediction ability. At each testing phase, different cataract types are detected. During classification, only few classes are misclassified as others. Thus, the confusion matrix exhibits that the proposed model provides accurate classification. Fig. 7 shows the obtained accuracy and loss during training and testing.

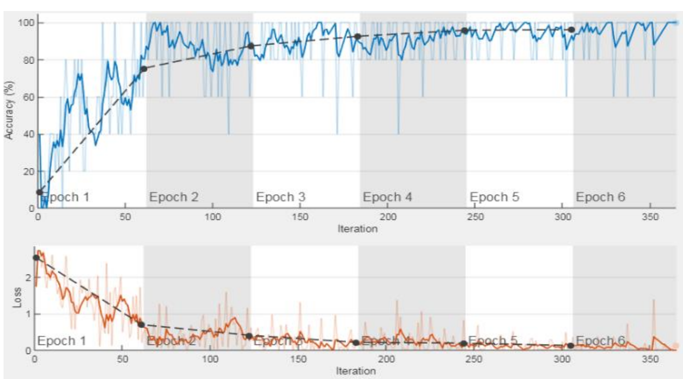


Fig. 7. Accuracy and loss graph of proposed Dense CNN+BB_ResNet101model

The accuracy and loss are measured at each iteration by varying the number of epochs. By analysing the results, it is clear that the accuracy and loss is well maintained in the proposed model. This proves that the proposed classifier is powerful for cataract type and grade classification. The results obtained from phase I by using retinal and slit lamp images are shown in Table I.

TABLE I. PERFORMANCE ANALYSIS IN PHASE I USING RETINAL AND SLIT LAMP IMAGES

Datasets	*Accuracy (%)*	*Sensitivity (%)*	*Specificity (%)*	*Precision (%)*	*F1 score (%)*	*Youden Index (%)*	*Kappa (%)*
(DRIMDB) Retinal images	92.36	94.58	96.57	91.06	91.47	89.65	89.32
Slit lamp images	94.67	96.68	98.29	93.03	93.81	91.23	90.28

In order to determine the strength of datasets, the performance metrics are evaluated, and the obtained results reveals that the slit lamp images obtain a better outcome than retinal images from the DRIMDB dataset. Thus, slit lamp images are preferred in phase II to detect the type and grade of the cataract. The accuracy measure comparison is illustrated in Fig. 8.

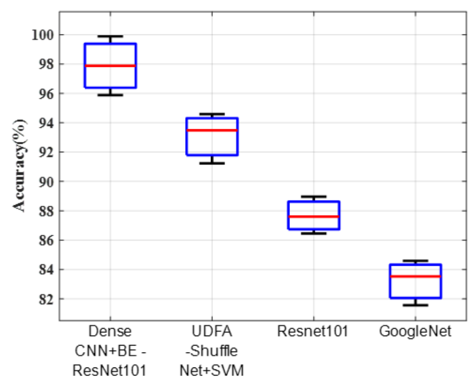


Fig. 8. Accuracy performance comparison

The above analysis clearly shows that, the accuracy of the proposed classifier is enhanced because of the higher efficiency. Besides that, the proposed BE-ResNet101 model diminishes the training error and also it avoids the vanishing gradient issue. But, the existing approaches generates higher classification errors and leads to low accuracy. The accuracy of proposed model is 98.87%, ResNet-101 is 87.6%, GoogleNet is 83.53% and UDFA is 93.48%. The comparative analysis of proposed and existing works in terms of sensitivity is shown in Fig. 9.

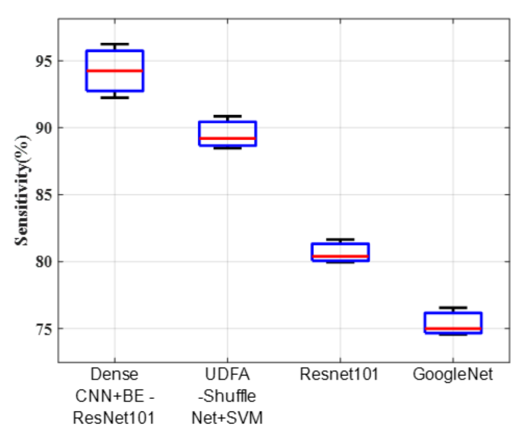


Fig. 9. Sensitivity performance comparison

The above analysis shows that the proposed Dense CNN +BE_ResNet101 model obtains enhanced sensitivity to existing approaches. The average value of sensitivity attained from the proposed model is98.28%, ResNet101 is 80.39%, GoogleNet is 75%, and UDFA is 89.2%. The specificity comparison analysis is illustrated in Fig. 10.

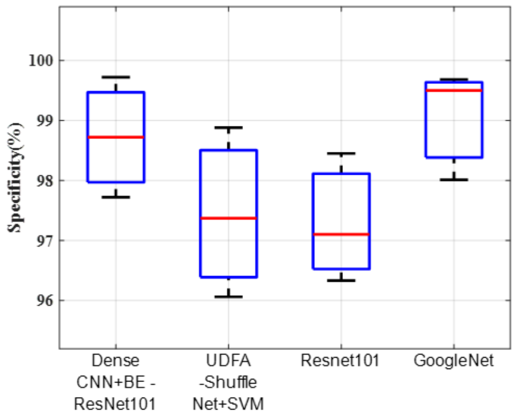


Fig. 10. Specificity performance comparison

The existing approaches face several issues when detecting cataracts from the given image. Thus, the previous studies are unable to provide improved specificity results. The average specificity value attained from the proposed Dense CNN+BE_ResNet101 model is 99.66%, ResNet 101 is 97.1%, UDFA is 97.37% and GoogleNet is 99.5%. The F1-score comparison is depicted in Fig. 11.

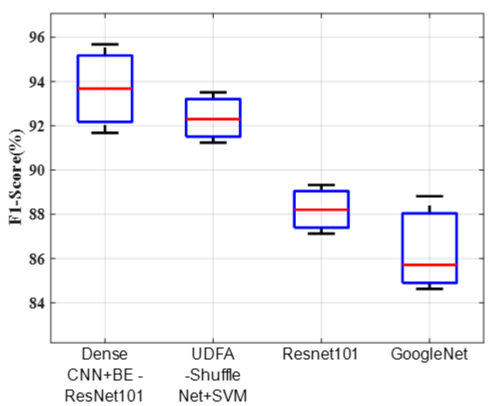


Fig. 11. F1 score performance comparison

The average F1-score of proposed classifier is 95.68%, UDFA is 92.3%, and ResNet101 is 88.2% and GoogleNet 85.71%. Compared with other approaches, the F1score of the proposed classifier is superior. This states that the developed Dense CNN+BE ResNet101 is more suitable for detecting cataract type and grade. The comparative analysis of the kappa coefficient is shown in Fig. 12.

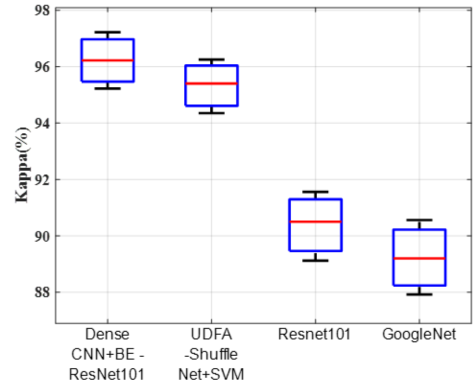


Fig. 12. Kappa performance comparison

The comparison analysis using the kappa measure shows that the proposed techniques are more potent than the others. The average kappa value obtains from the proposed Dense CNN+BE-ResNet101 is 97.83%, UDFA is 95.4%, ResNet101 is 90.5% and GoogleNet is 89.2%. The Youden index comparison of proposed and existing techniques is shown in Fig. 13.

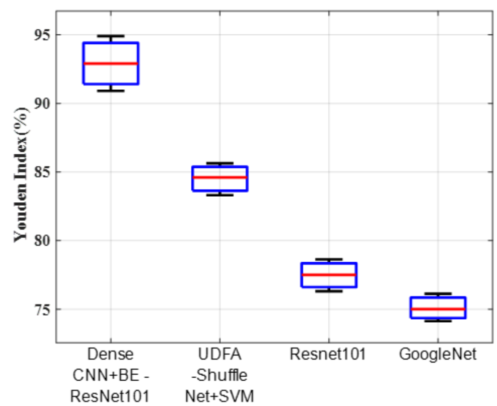


Fig. 13. Performance comparison in terms of Youden index

The average Youden index attained from the proposed model is 95.04, UDFA is 84.6%, ResNet101 is 77.5%, and GoogleNet is 75%. Compared with others, the proposed method is the highly enlarged range in the Youden index. The comparison of the ROC curve and AUC values of different techniques is illustrated in Fig. 14.

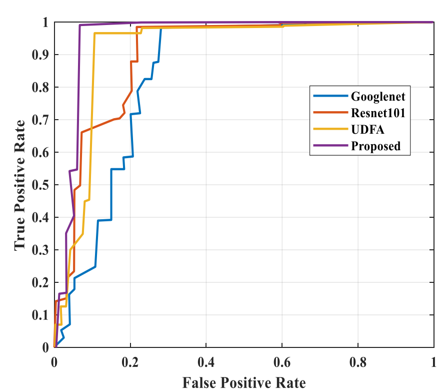


Fig. 14. ROC curve and AUC values of proposed and existing approaches

The ROC curve comparison represents that the proposed model is better than the other techniques. Analysing the ROC curve, the obtained AUC values are superior in the proposed cataract grade detection model. By analysing the proposed work with different approaches, the proposed model obtains improved outcomes in terms of accuracy, sensitivity, specificity, Youden index, Kappa and F1 score.

V. Conclusion

This research presents cataract detection and classification using deep learning approaches through retinal and slit lamp images. Here, the proposed work is carried out in two phases. In the first phase, the cataract is identified, and in the second phase, the type and grade of the cataract is categorized. In phase I, the images are gathered from DRIMDB and slit lamp images. By using these two varied dataset images, the cataract is detected through proposed Deep OCRN_IAO model. The experimental results prove that the slit lamp images are obtained improved performance than the retinal images.

Thus, in phase II, the slit lamp images are preferred for cataract type and grade classification. In the proposed work, the type of cataract is detected using Dense CNN model, where the convolutional layer of CNN reduces the feature dimensionality. Depending on the outcome of convolutional layer, the dense layer categorizes the features. Because of the presence of a dense layer, the CNN network accurately detects the type of cataracts. Also, the proposed BE-ResNet101 model detects the grade of cataracts by reducing the degradation issue. This model classifies the grades as mild, severe and moderate. Therefore, the system obtains enhanced results than the existing ones. The experimental results show that the proposed Dense CNN+BE-ResNet101 model attains good performance results in an accuracy of 98.87%, sensitivity 98.28%, specificity 99.66%, kappa coefficient 97.83%, F1 score 95.68% and Youden index 95.04%. In future studies, the proposed work will be applied to different biomedical images and detect their performance and effectiveness through various datasets.

References

[1] D. Lam, S.K. Rao, V. Ratra, Y. Liu, P. Mitchell, J. King, M.J. Tassignon, J. Jonas, C.P. Pang, D.F. Chang, “Cataract,” Nature reviews Disease primers, Vol. 1, no. 1, pp. 1-15, 2015.

[2] R.J. Olson, R. Braga-Mele, S.H. Chen, K.M. Miller, R. Pineda, J.P. Tweeten, D.C. Musch, “Cataract in the adult eye preferred practice pattern®.” Ophthalmology, Vol. 124, no. 2, pp. P1-P119, 2017.

[3] S.R. Flaxman, R.R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, M.V. Cicinelli, A. Das, J.B. Jonas, J. Keeffe, J.H. Kempen, J. Leasher, “Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis,” The Lancet Global Health, Vol. 5, no. 12, pp. e1221-e1234, 2017.

[4] K. Landau, M. Levin, “Retinal disorders,” Handbook of clinical neurology, pp. 97-116, Vol. 102, 2011.

[5] I.K. Muftuoglu, M. Al-Sheikh, M.A. Rasheed, S.R. Singh, J. Chhablani, Imaging “Imaging in inherited retinal disorders,” European Journal of Ophthalmology, Vol. 31, no. 4, pp. 1656-1676, 2021.

[6] C. Karakosta, A. Tzamalis, M. Aivaliotis, I. Tsinopoulos, “Pathogenesis of Age-Related Cataract: A Systematic Review of Proteomic Studies,” Current Proteomics, Vol. 18, No. 4, pp. 458-466, 2021.

[7] Ç. Öktem, F. Aslan, “Vitamin D Levels in Young Adult Cataract Patients: A Case-Control Study,” Ophthalmic research, Vol. 64, no. 1, pp. 116-120, 2021.

[8] Q. Fan, X. Han, J. Luo, L. Cai, X. Qiu, Y. Lu, J. Yang, “Risk factors of intraocular lens dislocation following routine cataract surgery: a case–control study,” Clinical and Experimental Optometry, Vol. 104, no. 4, pp. 510-517, 2021.

[9] R.E. Ruiz-Lozano, J.C. Hernandez-Camarena, M. Roman-Zamudio, R.J. Alcazar-Felix, O. Davila-Cavazos, J.A. Cardenas-de la Garza, "Three types of cataract associated with atopic dermatitis and chronic topical corticosteroid use: A case report," Dermatologic therapy, Vol. 34, no. 1, pp. e14600, 2021.

[10] S. Hu, H. Wu, X. Luan, Z. Wang, M. Adu, X. Wang, C. Yan, B. Li, K. Li, Y. Zou, X. Yu, “Portable handheld slit-lamp based on a smartphone camera for cataract screening,” Journal of Ophthalmology, Vol. 2020, 2020.

[11] J. Chen, Y. Chen, Y. Zhong, J. Li, “Comparison of visual acuity and complications between primary IOL implantation and aphakia in patients with congenital cataract younger than 2 years: a meta-analysis,” Journal of Cataract & Refractive Surgery, Vol. 46, no. 3, pp. 465-473, 2020.

[12] S. Kumar, S. Pathak, B. Kumar, “Automated detection of eye related diseases using digital image processing,” In Handbook of multimedia information security: techniques and applications, pp. 513-544, 2019. Springer, Cham,

[13] Y. Luo, K. Chen, L. Liu, J. Liu, J. Mao, G. Ke, M. Sun, “Dehaze of cataractous retinal images using an unpaired generative adversarial network,” IEEE Journal of Biomedical and Health Informatics, Vol. 24, no. 12, pp. 3374-3383, 2020.

[14] Y. Wang, C. Tang, J. Wang, Y. Sang, J. Lv, “Cataract detection based on ocular B-ultrasound images by collaborative monitoring deep learning.” Knowledge-Based Systems, Vol. 231, pp. 107442, 2021.

[15] X. Zhang, J. Lv, H. Zheng, Y. Sang, “Attention-based multi-model ensemble for automatic cataract detection in b-scan eye ultrasound images,” In 2020 international joint conference on neural networks (IJCNN), pp. 1-10, 2020. IEEE,

[16] A. Koyama, D. Miyazaki, Y. Nakagawa, Y. Ayatsuka, H. Miyake, F. Ehara, S.I. Sasaki, Y.Shimizu, Y. Inoue, “Determination of probability of causative pathogen in infectious keratitis using deep learning algorithm of slit-lamp images,” Scientific reports, Vol. 11, no. 1, pp. 1-13, 2021.

[17] A. Imran, J. Li, Y. Pei, F. Akhtar, J.J. Yang, Q. Wang, “Cataract detection and grading with retinal images using SOM-RBF neural network,” In 2019 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 2626-2632, 2019. IEEE,

[18] C. P. Niya, T. V. Jayakumar, “Analysis of different automatic cataract detection and classification methods,” In 2015 IEEE International Advance Computing Conference (IACC), pp. 696-700, 2015. IEEE,

[19] L. Guo, J.J. Yang, L. Peng, J. Li, Q. Liang, “A computer-aided healthcare system for cataract classification and grading based on fundus image analysis,” Computers in Industry, Vol. 69, pp. 72-80, 2015.

[20] X. Gao, S. Lin, T.Y. Wong, “Automatic feature learning to grade nuclear cataracts based on deep learning,” IEEE Transactions on Biomedical Engineering, Vol. 62, no. 11, pp. 2693-2701, 2015.

[21] T. Pratap, P. Kokil, “Deep neural network based robust computer-aided cataract diagnosis system using fundus retinal images,” Biomedical Signal Processing and Control, Vol. 70, pp. 102985, 2021.

[22] M. Akil, Y. Elloumi, R. Kachouri, “Detection of retinal abnormalities in fundus image using CNN deep learning networks,” In State of the Art in Neural Networks and their Applications, pp. 19-61, 2021. Academic Press,

[23] S. Hu, X. Wang, H. Wu, X. Luan, P. Qi, Y. Lin, X. He, W.He, “Unified diagnosis framework for automated nuclear cataract grading based on smartphone slit-lamp images,” IEEE Access, Vol. 8, pp. 174169-174178, 2020.

[24] X. Liu, J. Jiang, K. Zhang, E. Long, J. Cui, M. Zhu, Y. An, “Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network,” PloS one, Vol. 12, no. 3, pp. e0168606, 2017.

[25] L. Zhang, J. Li, H. Han, B. Liu, J. Yang, Q. Wang, “Automatic cataract detection and grading using deep convolutional neural network,” In

2017 IEEE 14th international conference on networking, sensing and control (ICNSC), pp. 60-65, 2017. IEEE,

[26] A. Mittal, R. Soundararajan, C. B. Alan, “Making a “completely blind” image quality analyser,” IEEE Signal processing letters, Vol. 20, no. 3, pp. 209-212, 2012.

[27] N. Venkatanath, D. Praneeth, M.C. Bh, S.S. Channappayya, S.S. Medasani, “Blind image quality evaluation using perception based features,” In 2015 Twenty First National Conference on Communications (NCC), pp. 1-6, 2015. IEEE,

[28] E. Başaran, Z. Cömert, Y. Çelik, “Neighbourhood component analysis and deep feature-based diagnosis model for middle ear otoscope images,” Neural Computing and Applications, Vol. 34, no. 8, pp. 6027-6038, 2022.

Fast Comprehensive Secret Sharing using Naive Image Compression

Heri Prasetyo[1], Kukuh Caezaocta Prayuda[2]
Department of Informatics, Universitas Sebelas Maret (UNS), Surakarta, Indonesia

***Abstract*—This paper presents a simple method for performing *(k,n)*-Secret Sharing (SS) with fast computation. It aims to reduce the computational time of the former scheme in the shadow generation process. The former scheme performs SS with the polynomial function computation by involving the color palette. The color palette transforms noisy-like shadow image into more meaningfull appearance. However, this scheme requires a high computational burden on this transformation process. The proposed method exploits naive image compression to decrease the required bit for representing a secret and cover image. It effectively avoids the color palette usage previously used by the former scheme. The proposed method produces a set of shadow images with a cover image-like appearance. In addition, the secret and cover image can be reconstructed by gathering at least *k* shadow images. As documented in the Experimental Results section, the proposed method yields a promising result in the *(k,n)*-SS with reduced computational time compared to that of the former scheme.**

***Keywords*—Comprehensive; image compression; naïve; polynomial; secret sharing**

I. Introduction

Several methods have been proposed for secure secret image communication. The SS is the most popular approach to securely send one or multiple images from the sender to other parties, i.e. called as receiver or participant. The first work in the SS can be traced back to the classical paper [1]. It introduces a SS concept under (k,n)-SS thresholded setting. In this method, a secret method is converted into n shadows images and then transferred to the n participants. The recovery process aims to reconstruct a secret image by collecting k or more shadow images to achieve a correct or lossless result. If the number of collected shadow images is less than k, the recovered secret image is lossy, or nothing is obtained. An improvement of SS method is Visual Cryptography (VC) [2] which performs SS into a grayscale image. This improvement leads the direction for further development of SS methods. On the other hand, the Chinese Remainder Theorem (CRT)-based SS [3] also gain popularity because of its wider application ability. However, the CRT-based SS has a slight limitation in the secret image recovery process. The recovered image is lossy compared to that the original image. While the other methods use a binary set basis [4-5], modular arithmetics [6], general access structure [7], bitwise Boolean operation [8-9], adaptive weight priority [10], etc., to generate a set of shadow images.

In another ways, several techniques have also been developed for the multiple secret sharing [11-14]. Most methods exploit the exclusive-OR operation and CRT computation to generate a set of shadow images. The method in [11] involves a simple image encryption, while scheme in [12] utilizes the generalized chaotic image scrambling. The methods in [13] and [14] use the hyperchaotic image scrambling and improved beta chaotic image encryption, respectively, to yield a set of shadow images. However, all technique produces the noise-like shadow.

The SS and its variants effectively secure secret image communication. But, a set of shadow images generated by these methods are in a noise-like appearance. A malicious attacker can easily recognize these shadow images as a secure image containing some confidential information. This attacker may collect several shadow images to obtain a fake or counterfeit secret image. This situation is unacceptable in secret image communication. Thus, the friendly SS tries to solve this problem by converting each shadow image into a more friendly appearance or cover image-like. An attacker now cannot perceive the noise-like shadow image. The method in [15] is an example of a friendly SS approach. It utilizes the CRT and bitwise Boolean operation to generate a set of shadow images. The methods in [16] and [17]perform thresholded SS and progressive SS, respectively, with the meaningfull shadow images. Meanwhile, the method [18] generates a set of meaningfull shadow images under the multiple secret sharing framework.

Several method have been reported in literature in order to convert the noise-like shadow image into more friendly or meaningfull appearance such as in [15-19]. The method in [19] performs the comprehensive visual SS. A secret image is converted into a set of shadow images with a friendly or cover image-like appearance. This scheme employs the polynomial function computation and color palette in the shadow image generation stage. It can be categorized as (k,n)-SS. This method effectively produces a set of shadow images in the cover-like appearance. The secret and cover images can be recovered from at least k collected shadow images. However, this method requires a very high computational burden in the shadow image generation since it needs to compare the similarity over four bits as mentioned in [19]. It becomes inferior for the practical implementation of SS required fast computation. The method offers a solution to transform the noise-like shadow to be more meaningfull.

Thus, this paper offers a solution to reduce the computational time of [19] using naive image compression. This naive compression or image companding scheme effectively overcomes the former scheme limitation. The proposed work give a significant contribution on reducing the computational time of [19] in the comprehensive secret sharing task. It replaces the color pallete usage with a simple image compression technique. It introduces a new concept for converting the noise-like shadow image into more friendly appearance with noise compression (companding) which can be further utilized for future works, i.e. friendly secret sharing, comprehensive secret sharing, image watermarking [20], etc.

II. Former Scheme of Comprehensive Secret Sharing

This section introduces the former scheme [19] for performing secret sharing. It can be regarded as (k, n)-SS, with $k < n$, since it converts a secret image into n shadow images, while it requires at least k shadow images to obtain a recovered secret image. The method in [19] generates a set of shadow images in which their appearance is maintained as similar as possible to the targeted cover images. It employs a set of cover images in shadow image generation. The secret image can be recovered by using at least k shadow images. In addition, the cover image can also be reconstructed using after obtaining the recovered secret image. This aforementioned method employs the color palette to generate a set of shadow images and to recover secret and cover images.

The detail of the former method [19] can be explained as follow. Let I be a secret image, and $\{C_1, C_2, \dots, C_n\}$ be a set of cover images. This method forces to change I into a set of shadow images $\{S_1, S_2, \dots, S_n\}$. The appearance of shadow image should be as similar as possible to the cover image, i.e. $S_i \approx C_i$ for $i = 1,2, \dots, n$. The method in [19] firstly extracts four bits of each cover image C_i as follow:

$$C_i = \langle c_1, c_2, \dots, c_4 \rangle \quad (1)$$

where $c_1, c_2, \dots, c_4$ denotes four extracted bits of C_i, with c_1 is the most significant bit. These four bits are acquired by using a color palette [19]. Yet, The secret image is regarded as a_0, i.e. $a_0 = I$. Subsequently, the polynomial function computation is applied to perform (k, n)-SS as follows:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} (\mathrm{mod}\, P) \quad (2)$$

where $f(x)$ is the polynomial function order k, for $x = 1,2, \dots, n$. The value of P is a prime number. It is typically set as $P = 257$ in the 8-bits image representation. While the value of a_i is a random number generated in the range $a_i \sim [0, P)$, for $i = 1,2, \dots, k-1$. The temporary shadow image T_i can be obtained by changing the value of x in (2) with the index of shadow image, i.e. $i = 1,2, \dots, n$. The computation of T_i can be conducted as follow:

$$T_i = f(i) \quad (3)$$

for $i = 1,2, \dots, n$. From this process, one obtains a set of temporary shadow images $\{T_1, T_2, \dots, T_n\}$.

Until this process, the appearance of each shadow image T_i is in noise-like form. The appearance of T_i should be exchanged to be more resemble as C_i. An additional step is needed to perform this process. The temporary shadow image T_i should be converted from decimal into 8-bits representation. This binary number extraction process is given as:

$$T_i = \langle t_1, t_2, \dots, t_8 \rangle \quad (4)$$

where t_i is the i-th bit, for $i = 1,2, \dots, 8$, with t_1 is the most significant bit. The proposed method simply compares the four significant bits of T_i with the four significant bits of C_i. If there are all identical, T_i is then regarded as the shadow image S_i. Specifically, if $c_i = t_i$ for $i = 1,2, \dots, 4$, this process is performed:

$$S_i = T_i \quad (5)$$

where S_i denotes the i-th shadow image, for $i = 1,2, \dots, n$. Otherwise, the proposed method needs to recompute (2), i.e. the computation of polynomial function is executed again until the four significant bits are identical to that of the four significant bits of cover image. This process produces a set of shadow images $\{S_1, S_2, \dots, S_n\}$. Now, each shadow image S_i is visually similar to the cover image C_i.

The Lagrange interpolation is utilized to extract a secret image. Herein, the receiver simply collects at least k shadow images in the recovery process to obtain a lossless secret image. One gets a recovered secret image $\tilde{I}$ after applying the Lagrange interpolation. To reconstruct the cover image, the receiver needs to extract four significant bits of each S_i. Then, the inverse process of color palette computation [19] is performed to yield C_i, for $i = 1,2, \dots, n$, by considering four significant bits of S_i. This process produces a set of recovered cover images as $\{C_1, C_2, \dots, C_n\}$. The former scheme performs well in the (k, n) under the comprehensive SS setting.

Even though the former method effectively generates a set of shadow images with a cover-like appearance. However, the computation of similarity matching over four significant bits, i.e. $c_i = t_i$ for $i = 1,2, \dots, 4$, need a high computational burden. The method should recalculate $f(x)$ if the four bits are not identical. It will be inconvenient if a fast computation response is required to generate a set of shadow images from a secret image.

III. Proposed Method

The proposed method offers a simple solution for the limitation of former scheme [19]. It tries to reduce the computational burden of similarity matching for four significant bits. The proposed method avoids this similarity matching to further reduce the computational time. Herein, simple naive image compression is exploited in the shadow image generation and secret image recovery. Sender and receiver do not use the color palette in this SS process. The proposed method is further explained in this section as follows.

A. Shadow Image Generation

As mentioned before, the proposed method converts a secret image I into a set of shadow images. This method involves a set of cover images as $\{C_1, C_2, \dots, C_n\}$. Let r and q be the required bit for compressing the secret and cover images, respectively. The value should satisfy $r + q = 8$, for an 8-bits representation of an image. These two values should be kept for both sender and receiver in the SS process. The proposed method performs naive image compression or image companding process utilizing two specific quantizer values. These two quantizer values can be computed as follows:

$$Q_s = 2^{8-r} \quad (6)$$

$$Q_c = 2^{8-q} \quad (7)$$

where Q_s is the quantizer value for the secret image, and Q_c is the quantizer for cover image. Subsequently, the compression processes for secret image I and cover image C_i are performed as follows:

$$\hat{I} = \left\lfloor \frac{I}{Q_s} \right\rfloor \quad (8)$$

$$\hat{C}_i = \left\lfloor \frac{C_i}{Q_c} \right\rfloor \quad (9)$$

where $\hat{I}$ is a compressed secret image, and $\hat{C}_i$ is the i-th compressed cover image. The symbol $\lfloor \cdot \rfloor$ represents the floor operator. From these computations, the lengths of $\hat{I}$ and $\hat{C}_i$ are now with r and q-bits, respectively. The compressed secret image $\hat{I}$ can be converted into binary form as follow:

$$\hat{I} = \langle \hat{\imath}_1, \hat{\imath}_2, \dots, \hat{\imath}_r \rangle \quad (10)$$

where $\hat{\imath}_i$ is the i-th bit of $\hat{I}$, for $i = 1,2,\dots,r$. While $\hat{\imath}_1$ is the most significant bit. The binary conversion of compressed cover image $\hat{C}_i$ is given as:

$$\hat{C}_i = \langle \hat{c}_{i1}, \hat{c}_{i2}, \dots, \hat{c}_{iq} \rangle \quad (11)$$

where $\hat{c}_{ij}$ is the j-th bit of $\hat{C}_i$, for $j = 1,2,\dots,q$. The most significant bit is represented with $\hat{c}_{i1}$. These binary numbers are used in shadow image generation.

Subsequently, the polynomial function is computed with $P = 2^r$. Herein, the value of P is a non-primary number. In our proposed method, the value of P depends on the required bit of secret image, i.e. r. Similar to the former scheme [19], the proposed method also considers $a_0 = I$. It also generates a set of random numbers in a specific range, i.e. $a_i \sim [0, P)$ for $i = 1,2,\dots,k-1$. The polynomial function for the SS can be calculated as follow:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} (\mathrm{mod}\, P) \quad (12)$$

This computation is for $x = 1,2,\dots,n$. It should be noted that $f(x)$ is in r-bits representations, i.e. the length of $f(x)$ is r under the binary form. The temporary shadow image is then obtained by replacing x with the index of shadow image, i.e. $x = i$. This process is formally defined as follows:

$$T_i = f(i) \quad (13)$$

for $i = 1,2,\dots,n$. Each temporary shadow image is also in r-bits representation, i.e. $T_i = \langle t_{i1}, t_{i2}, \dots, t_{ir} \rangle$. The proposed method performs bit concatenation between all bits in T_i with all bits in $\hat{C}_i$. Then, the final shadow image is obtained as follows:

$$S_i = \left[\hat{c}_{i1}, \hat{c}_{i2}, \dots, \hat{c}_{iq}, t_{i1}, t_{i2}, \dots, t_{ir} \right] \quad (14)$$

where S_i is the i-th shadow image, for $i = 1,2,\dots,n$. The symbol $[\cdot]$ denotes the bit concatenation operator. Herein, the length of S_i is 8-bits. After converting the binary number into the decimal representation of each S_i, one can gain a set of shadow images $\{S_1, S_2, \dots, S_n\}$. The sender sends these shadow images to the receiver via a communication channel.

B. Secret Image Recovery

In the secret image recovery process, the receiver tries to produce a set image and recovered cover image by collecting several shadow images $\{S_1, S_2, \dots, S_K\}$, while $k \le K \le n$ to obtain a perfect reconstruction process. The receiver firstly converts each shadow image into binary representation as follows:

$$S_i = \left[\hat{c}_{i1}, \hat{c}_{i2}, \dots, \hat{c}_{iq}, t_{i1}, t_{i2}, \dots, t_{ir} \right] \quad (15)$$

for $i = 1,2,\dots,K$. Where K denotes the number of collected shadow images. The r least significant bits are extracted from (15) to generate a temporary shadow image. This binary number is then converted into decimal number as follows:

$$T_i = \mathrm{Dec}\langle t_{i1}, t_{i2}, \dots, t_{ir} \rangle \quad (16)$$

for $i = 1,2,\dots,K$. While $\mathrm{Dec}\langle\cdot\rangle$ denotes the operator from binary to decimal number conversion. The Lagrange interpolation as used in [19] is then applied to all T_i to yield I^*, where I^* denotes a temporary secret image. This temporary secret image is still in low dynamic range, i.e. it is still in r-bits representation. The final secret image is then produced using the following process:

$$\tilde{I} = Q_r \times I^* \quad (17)$$

where $\tilde{I}$ represents a recovered secret image. In the proposed comprehensive (k, n)-SS, the cover image can be recovered from the shadow image. The receiver extracts q-bits from (15) to recover the cover image. The process of cover image recovery is given as follows:

$$\tilde{C}_i = Q_c \times \mathrm{Dec}\langle \hat{c}_{i1}, \hat{c}_{i2}, \dots, \hat{c}_{iq} \rangle \quad (18)$$

where $\tilde{C}_i$ denotes the i-th recovered cover image, for $i = 1,2,\dots,n$. Both sender and receiver need Q_s and Q_c in the secret image recovery and cover image reconstruction process. The receiver only keeps the values of r and q for computing Q_s and Q_c. Using this simple approach, the proposed method overcomes the limitation of the former scheme [19] in the high computational burden.

IV. Experimental Results

Several experiments have been conducted to investigate the proposed method performance. All experiments are then reported in this section. This section firstly shows the experimental results under the visual inspection, i.e. the correction of the proposed method is only observed under human investigation. Subsequently, it delivers the performance comparisons under the objective quality assessment. In our experiments, all images are of size 512×512. All images are in color format. The histogram is given at the bottom left of each image.

A. Visual Observation

This subsection reports the proposed method performance for dealing with (3,4)-SS and (2,3)-SS. This experiment only overlooks the generated shadow images and recovered secret image with the visual investigation. Herein, one secret image is involved, and several cover images are used in the experiment. Fig. 1(a) depicts a secret image used in the experiment. This image is in original I, while Fig. 1(b) is the compressed version

$\hat{I}$ with $r = 4$. A set of original cover images is shown in Fig. 2, while the compressed version of all cover images are given in Fig. 3. Herein, the required bit for the cover image is set as $q = 4$.

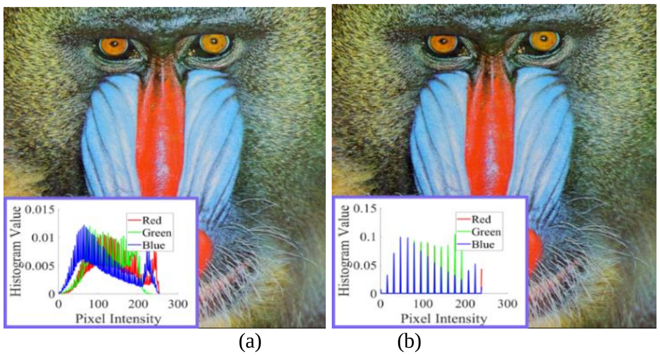


Fig. 1. Secret image used as experiment: (a) original image I, and (b) compressed version $\hat{I}$

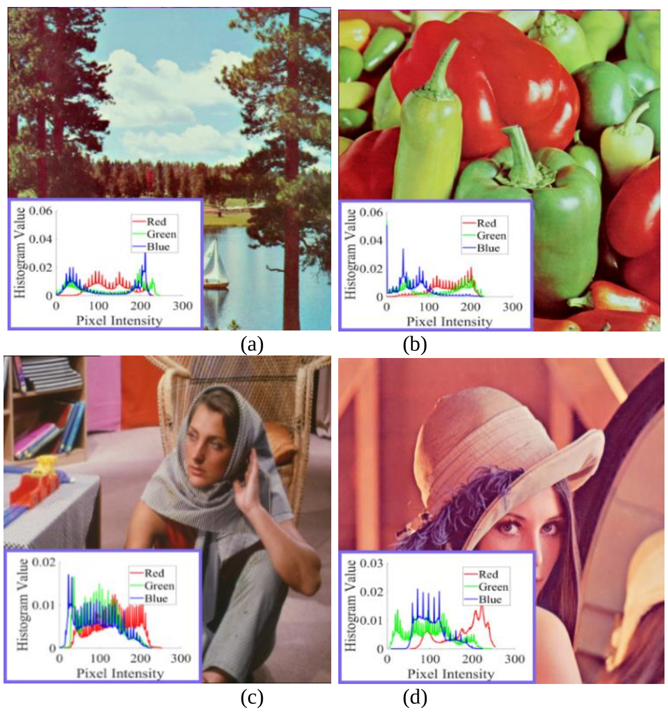


Fig. 2. A set of cover images: (a) C_1, (b) C_2, (c) C_3, and (d) C_4

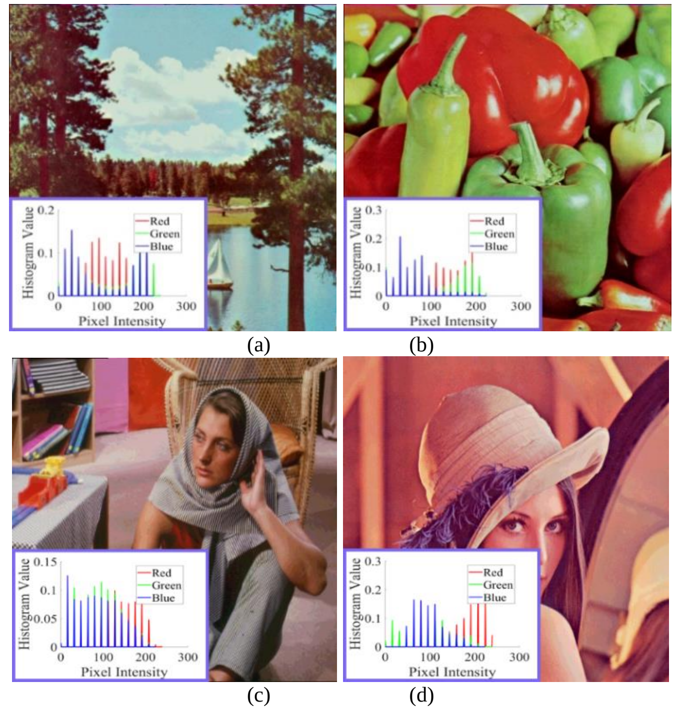


Fig. 3. A set of compressed cover images: (a) $\hat{C}_1$, (b) $\hat{C}_2$, (c) $\hat{C}_3$, and (d) $\hat{C}_4$

This section firstly considers the proposed method performance under (3,4)-SS setting. Fig. 4 is a set of shadow images obtained from our scheme. As shown in this figure, the content of each shadow image is almost similar to that of the original cover image. Thus, the proposed method effectively performs the comprehensive secret sharing with (3,4)-SS approach. Recovered secret images are obtained by performing the secret image recovery process involving several shared images. Fig. 5 delivers the recovered secret image $\tilde{I}$ while two or more shared images are used in the recovery process. The proposed method yields correct results for (3,4)-SS. The secret image can be losslessly recovered while at least three shadow images are used in the recovery process. In addition, a set of cover images can be recovered after extracting the secret image. Fig. 6 reports the result of recovered secret image. Herein, all shadow images are employed for performing the cover image recovery. As demonstrated in this figure, the proposed method is able to recover the cover image with an identical result compared to the compressed version of cover image.

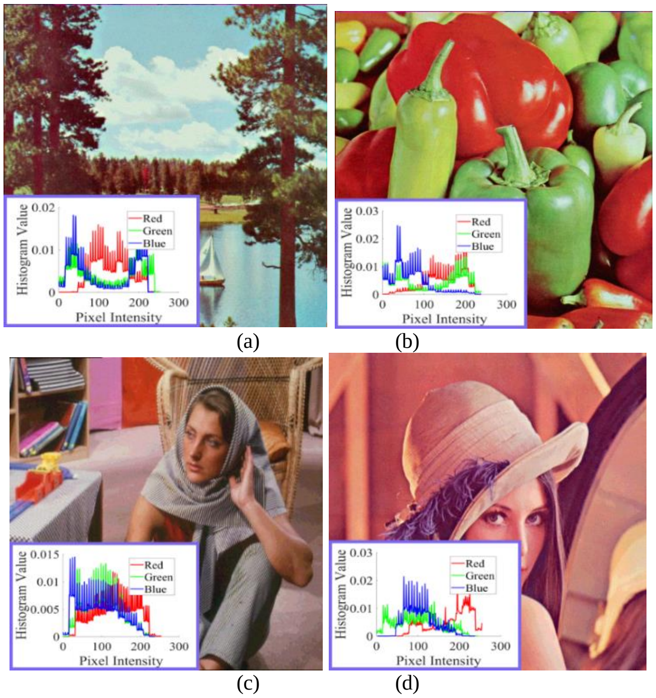


Fig. 4. A set of shadow images from (3,4)-SS: (a) S_1, (b) S_2, (c) S_3, and (d) S_4

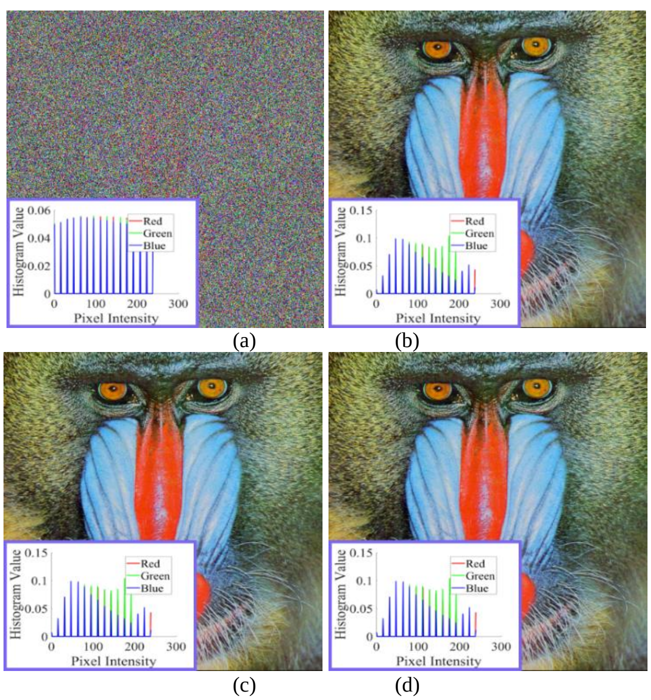


Fig. 5. Recovered secret image $\tilde{I}$ from (3,4)-SS by involving several shared images: (a) $\{S_1, S_2\}$, (b) $\{S_1, S_2, S_3\}$, (c) $\{S_2, S_3, S_4\}$, and (d) $\{S_1, S_2, S_3, S_4\}$.

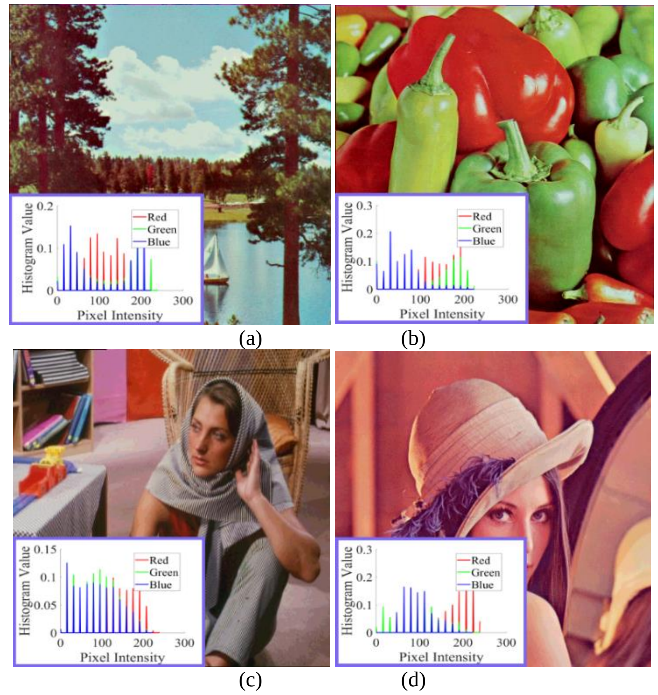


Fig. 6. Recovered cover images from (3,4)-SS: (a) $\tilde{C}_1$, (b) $\tilde{C}_2$, (c) $\tilde{C}_3$, and (d) $\tilde{C}_4$

An additional experiment is also executed further to examine the correctness of proposed method under visual investigation. This experiment inspects the proposed method under (2,3)-SS setting. A secret image and three cover images are from Fig. 1(a) and Fig. 3(a-c), respectively. It also utilizes $r = q = 4$. The proposed method yields a set of shadow images as delivered in Fig. 7. Again, the proposed method performs well in converting the secret image into a set of shadow images whose appearance is similar to that of the cover image. Fig. 8 exhibits the visual result of the recovered secret image $\tilde{I}$ while one or more shared images are involved during the recovery process. The recovered secret image is lossless, while at least two shadow images are used. Thus, the proposed method is correct for performing the (2,3)-SS. A set of recovered cover images can also be obtained using all shadow images during the recovery process. Fig. 9 shows a set of recovered cover images. All of these images are identical to that of the compressed cover image. It can be concluded that the proposed method gives promising results for (k, n)-SS with the comprehensive scenario. In addition, the proposed is a strong candidate while implementing the comprehensive secret sharing compared to the other scheme. It avoids the computation burden as used in [19]. It also requires a simple step for conducting the comprehensive secret sharing.

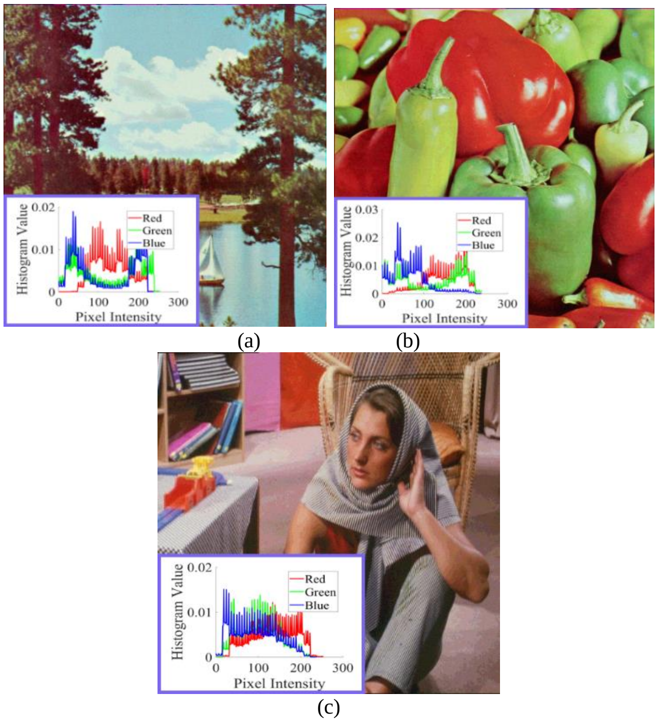


Fig. 7. A set of shadow images generated by (2,3)-SS: (a) S_1, (b) S_2, and (c) S_3

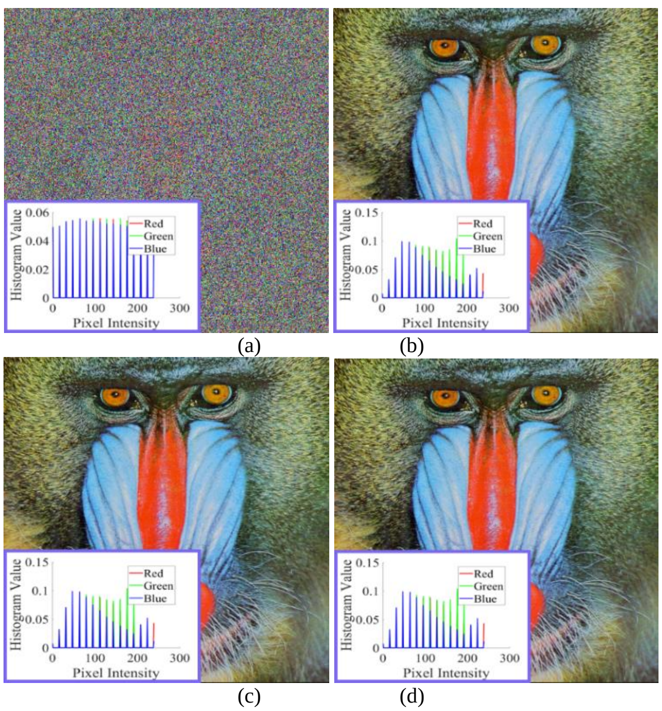


Fig. 8. The visual of $\tilde{I}$ from (2,3)-SS when shared images are used in the recovery process: (a) $\{S_2\}$, (b) $\{S_1, S_2\}$, (c) $\{S_2, S_3\}$, and (d) $\{S_2, S_3, S_1\}$

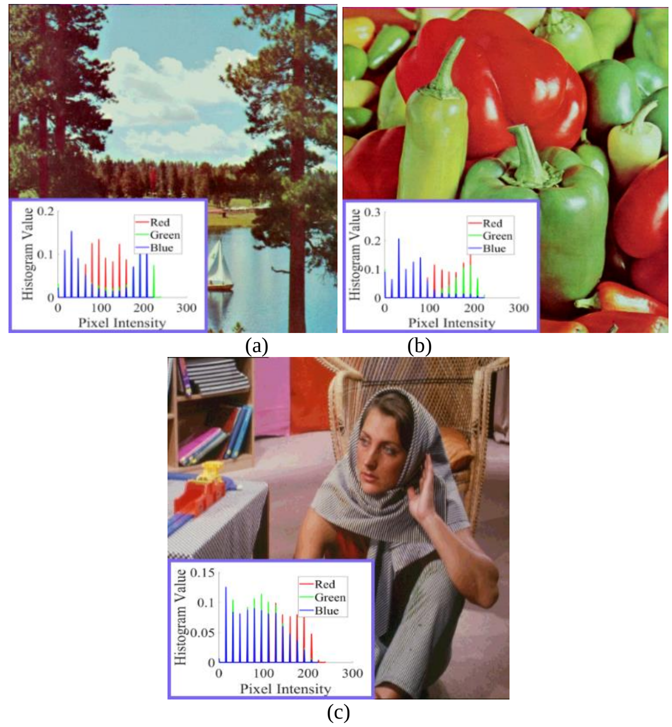


Fig. 9. A set of recovered cover images from (2,3)-SS: (a) $\tilde{C}_1$, (b) $\tilde{C}_2$, and (c) $\tilde{C}_3$

B. Objective Comparisons

This subsection summarizes the proposed method performance under the objective image quality assessment. This experiment only examines the proposed method under (3,4)-SS setting. It observes the performance by investigating the effect of required bit for secret image r. It means that the cover image is compressed with various q with $q = 8 - r$. All shadow images are involved during the secret image recovery process. The similarity between the shadow image and the original cover is measured with Peak-Signal-to-Noise Ratio (PSNR). It computes the average PSNR scores over all four shadow images. The similarity between the recovered cover image and the original version is also observed under the average PSNR score. This calculation is also for the recovered and original secret images. A higher value of average PSNR indicates better performance. Fig. 10 displays the performance comparisons with the average PSNR value over various $r = \{2,3, \ldots ,6\}$. The quality of shadow image and recovered cover image is decreased while applying higher r. But, the quality of a recovered secret image is increased by using a higher value r. The proposed method yields the best performance with $r = q = 4$, as confirmed in Fig. 10.

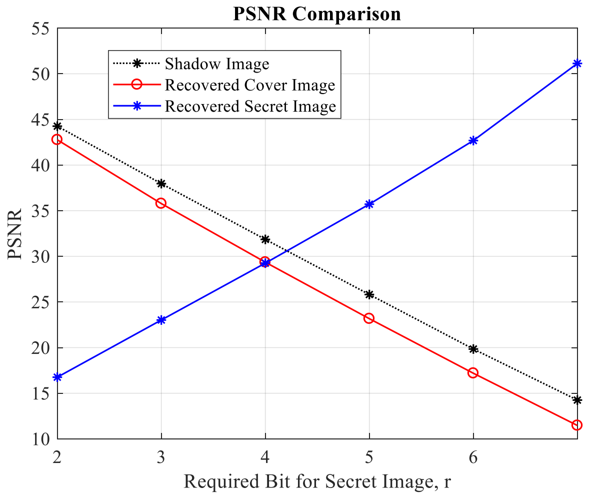


Fig. 10. Performance comparisons in terms of average PSNR value

The evaluations are also observed for the proposed method performance in terms of Structural Similarity Index Metric (SSIM). Herein, this experiment also considers the quality of shadow image, recovered cover image, and recovered secret image. A higher value of SSIM also implies better performance. Fig. 11 demonstrates the proposed method performance under the average SSIM score evaluation over various r, i.e. $r = \{2,3,\dots,6\}$. The quality of shadow image and recovered cover image is reduced while applying a higher of r. However, the quality of recovered secret image is increased with higher r. The proposed method yields the best performance by setting $r = 4$ as demonstrated in Fig. 11. The proposed method effectively (k,n)-SS setting.

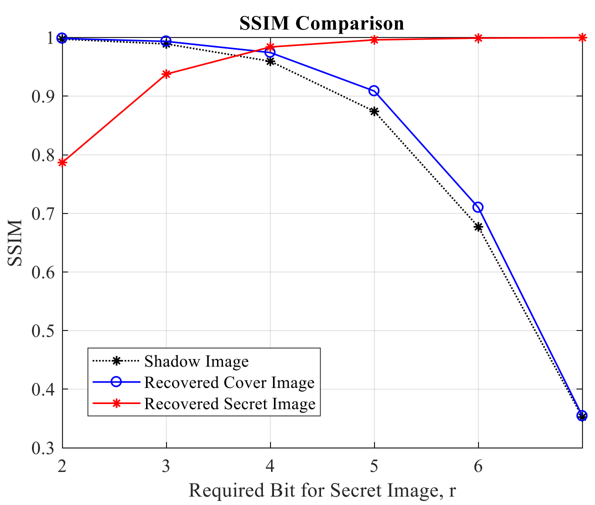


Fig. 11. Performance evaluations in terms of average SSIM score over various required bit for secret image

V. Conclusions

A simple solution for reducing the computational time for the former (k,n)-SS has been presented in this paper. The proposed method utilizes naive image compression to replace the color palette usage. This image compression is a straightforward approach to reduce the required bit of secret and cover image. The proposed method effectively produces a set of shadow images with a friendly appearance. In addition, it can recover the secret and cover images. For future works, the security level of the proposed method can be improved by involving image encryption or hashing functions. It can also be extended for secret video communication. The proposed method can also be applied to multiple secret sharing.

Acknowledgment

This work was fully funded by the Universitas Sebelas Maret (UNS), Indonesia, under the research grant "Hibah Penelitian Unggulan Terapan (PUT-UNS) Tahun Anggaran 2022" with the contract number 254/UN27.22/PT.01.03/2022.

References

[1] A. Shamir, "How to share a secret, " *ACM Communication*, vol. 22, no. 11, pp. 612-613, 1979.

[2] M. Naor M., and A. Shamir, "Visual cryptography," *Workshop on the theory and application of cryptographic techniques*, pp. 1-12, 1994.

[3] W. Yan, W. Ding, and Q. Dongxu, "Image sharing based on chinese remainder theorem," *Journal of North China University of Technology*, vol. 12, no. 1, pp. 6-9, (2000).

[4] H. Prasetyo, and C. H. Hsia, "Lossless progressive secret sharing for grayscale and color images," *Multimedia Tools and Applications*, vol. 78, pp. 24837–24862, 2019.

[5] A. Kanso, and M. Ghebleh, "An efficient lossless secret sharing scheme for medical images," *Journal of Visual Communication and Image Representation*, vol. 56, pp. 245-255, 2018.

[6] C. N. Yang, C. E. Zheng, M. C. Lu, and. X. Wu, "Secret image sharing by using multi-prime modular arithmetic," *Signal Processing*, vol. 205, 2023.

[7] X. Jia, Y. Guo, X. Luo, D. Wang, and C. Zhang, "A perfect secret sharing scheme for general access structures," *Information Sciences*, vol. 595, pp. 54-69, 2022.

[8] Y. Liu, C. N. Yang, S. Wu, and Y. Chou, "Progressive (k,n) secret image shading schemes based on Boolean operations and covering codes," *Signal Processing: Image Communication*, vol. 66, pp. 77-86, 2018.

[9] H. Prasetyo, D. Rosiyadi, and S. J. Horng, "Modified generalized random grids-based progressive secret sharing with lossless ability for binary image," *in Proc. International Conference on Computer, Control, Informatics and its Applications (IC3INA 2018)*, Tangerang, Indonesia, 2018.

[10] H. Prasetyo, C. H. Hsia, and A. W. H. Prayuda, "Progressive secret sharing with adaptive priority and perfect reconstruction," *Journal of Imaging*, vol. 7, no. 4, 2021.

[11] H. Prasetyo, C. H. Hsia, and J. Y. Deng, "Multiple secret sharing with simple image encryption," *Journal of Internet Technology*, vol. 21, no. 2, pp. 323-341, 2020.

[12] H. Prasetyo, C. H. Hsia, "Improved multiple secret sharing using generalized chaotic image scrambling," *Multimedia Tools and Applications*, vol. 78, no. 20, pp. 29089-29120, 2019.

[13] H. P., J. M. Guo, "A note on multiple secret sharing using Chinese remainder theorem and exclusive-OR," *IEEE Access*, vol. 7, pp. 37473-37497, 2019.

[14] J. M. Guo, D. Riyono, H. Prasetyo, "Improved beta chaotic image encryption for multiple secret sharing, " *IEEE Access*, vol. 6, pp. 46297-46321, 2018.

[15] H. Prasetyo, and D. Rosiyadi, "Converting (n, n)-multiple secret sharing into more friendly appearance using chinese remainder theorem and boolean operations," *in Proc. International Symposium on Electronics and Smart Devices (ISESD)*, pp. 1-6, 2021.

[16] C. N. Yang, P. Y. Tsai, and Y. Liu, "A (k,n) secret document sharing with meaningfull shares," *Journal of Information Security and Applications*, vol. 62, 2021.

[17] H. Prasetyo, and J. W. Simatupang, "XOR-ed Based Friendly-Progressive Secret Sharing," *in Proc. International Symposium on*

Intelligent Signal Processing and Communication Systems (ISPACS 2019), Taipei, Taiwan, 2019.

[18] H. Prasetyo, C. H. Hsia, C. Yu, and A. R. Wirawan, “Friendly Appearance of Multiple Secret Sharing,” *in Proc. IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan 2020)*, Taipei, Taiwan, 2020.

[19] J. Cheng, X. Yan, L. Liu, Y. Sun, and F. Xing, “Comprehensive reversible secret image sharing with palette cover images,” *Journal of Information Security and Applications*, vol. 68, 2022.

[20] B. Harjito, and H. Prasetyo, “False-positive-free GSVD-based image watermarking for copyright protection,” *in Proc. International Symposium on Electronics and Smart Devices (ISESD 2016)*, Bandung, Indonesia, 2016.

The Effect of Blockchain using Big data and the Internet of Things in Healthcare

Bassant Nabil Mohamed*, Hatem Abdelkader
Information Systems Department
Faculty of Computers and Information
Menufia University, Egypt

***Abstract*—Modern organizations of all sizes emphasize safeguarding sensitive consumer information. Regardless of the limits given by the degrees they choose to pursue, people are nevertheless required to possess data management skills. In addition, determine if the data should be centralized or decentralized to meet the objective of enhancing accessibility. In addition, you will need to be able to monitor who has access to your data and regulate who has access to your data.**

***Keywords*—Big data; blockchain; internet of things; data security; healthcare; data processing cost; IOMCT; climate change and global warming**

I. Introduction

This article focuses on the Internet of Things (IoT) and massive datasets to investigate the use of blockchain technology, specifically to enhance the overall quality of the healthcare system. The fast growth of the internet of things is inextricably related to the rapid growth of the information technology sector (IT). A Model (Internet Of Medical Thing) IOMCT was created to address these issues since it was difficult to get the necessary insurance approvals and there was a difficulty with the dispersion of information. This concept was created to solve the challenges associated with getting the requisite insurance approvals. In addition, there is a reduced likelihood that sensitive information and data may be stolen by unauthorized individuals as a direct result of being exposed.

The findings of this research indicate that the technique we have developed will contribute to the safeguarding of patients' personal information. In addition, it will improve people's awareness of the connection between paper use and global warming. Therefore, it is fair to assume that individuals would choose the paperless technique, which will contribute to the conservation of natural resources.

It should not come as a surprise that healthcare is a top priority in countries with low per capita income since it has direct effects on the health and longevity of the people. It has the potential to considerably benefit society as a whole by reducing the occurrence of a broad range of illnesses and ailments that are harmful to one's health as shown in Fig. 1. This study should be continued since its potential benefits to society as a whole are substantial. In light of this, it is glaringly evident that recent advances in technological capabilities and ongoing research and development have been crucial to the advancement of the medical sector [1].

In the not-too-distant future [19], it is anticipated that the Internet of Things will become a formidable mechanism for the production of counterfeit wristbands, smart cards, and watches. Using a blockchain ledger in this capacity, data such as a patient's heart rate, blood pressure, and blood sugar levels may be recorded automatically and exactly.

Advanced Technologies including the IoT, Blockchain, and Big Data, can assist doctors and medical sectors in the early diagnosis of various diseases.

As a direct consequence of technological improvements, our lives have grown simpler, more accurate, and more reliable. By the time you conclude this essay, you will understand how to simplify the patient care procedure at any medical institution. The Internet of Things employs a broad range of technologies, including sensors, the cloud, wireless technology, and security, to name a few [10]. Paper organization. Section II explains the essential problems and illustrates the objectives of the paper. Section III explains the literature reviews of the previous researchers. Section IV explains the research methodology. Section V explains the Healthcare proposed model (IOMCT). Section VI explains the proposed Blockchain Algorithms. Section VII illustrates the results of the research. And finally, Section VIII clarifies the conclusions and recommendations of this paper.

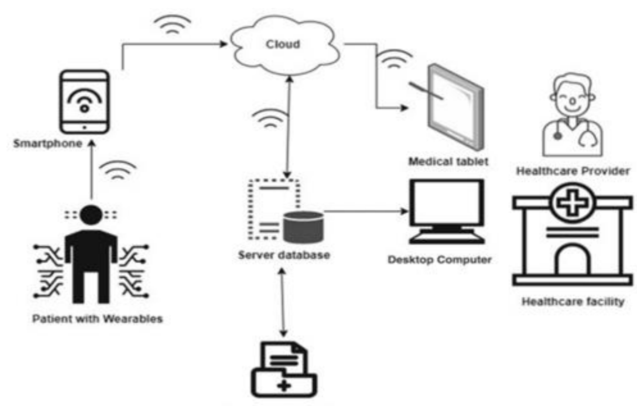


Fig. 1. Conceptual model of IoT in healthcare [1].

II. Problem Definition and Objective

In the HealthCare era, the patient always feels that he wastes a lot of time to form registration or even if wants just a simple consultation. So, he must move through the whole

*Corresponding Author.

hospital procedure, and waste time in waiting. On the other side, for people who face problems with heart pulse rates, sugar rate or blood pressure continuous monitoring and observation are. They don't have a solution except to go to the doctor and wait till they take only consultation, or they can only send an email or message to the doctor and then wait till he answers. In another situation, people don't need their data to be exposed to all hospitals. As he can see the patient history from before. So, according to that patient cannot own and control or secure his data from exposure. In addition, people who work in sensitive positions, and face some health problems, don't have the luxury to save their information. So, in this paper, we will go through all these problems and propose effective solutions.

III. LITERATURE REVIEW

Jalel Ktari and his coworkers came up with the concept of developing a heterogeneous embedded platform to facilitate the construction of electronic health records (EHRs) with multiple inputs. This platform can collect data from a variety of sensors that are interconnected via the internet of things. Utilizing Ethereum's permissioned network, this method encrypts data not once, but twice, thereby enhancing the security of the data [17]. To complete this task, the PoW consensus technique will be utilized. We decided on this course of action to protect the privacy of our patients. This information must be protected from both unauthorized access and prying eyes [3].

Tian-Fu Lee, et al. proposed a mechanism for preserving medical information that considers the entire process of data from data generation through transmission by wearable devices to mobile devices and then to a medical Center server [4]. In a presentation, Gautam Srivastava et al. discussed both the advantages and practical challenges associated with using blockchain-based security solutions for the Internet of Things-based remote patient monitoring. In addition, it assesses the relative benefits and drawbacks of a variety of cryptographic algorithms that might be useful for Internet of Things deployment. In this study, we provide a paradigm for securely transporting data over networks and storing data in the cloud using low-overhead cryptographic techniques, such as the ARX encryption scheme [5].

Partha Pratim Ray and colleagues [6] suggested the use of an IoBHealth data-flow architecture for storing, retrieving, and managing electronic medical information. IoBHealth combines blockchain technology with the Internet of Things. The purpose of this presentation by K. Azbeg and colleagues is to classify the Internet of Things and blockchain technology applications in the medical industry. Six distinct medical service applications were examined, along with a study of the problems that may occur when trying to use Blockchain technology in IoT-based healthcare systems [7]. In a newly published research, Manal Al-rawashdeh and her colleagues categorized the characteristics that influence IoT adoption as follows: personal considerations, technical considerations, security issues, health concerns, and environmental aspects [8].

The authors undertake a literature study on the issue of healthcare IoT application adoption between 2015 and 2021. The purpose of this study was to compile prior studies on the acceptability and deployment of Internet of Things (IoT) technologies in smart healthcare. The research painstakingly explored, obtained, evaluated, and synthesized the relevant material for the study [9].

A. Decentralization

We need a decentralized system in order to avoid many-to-one traffic flows, as well as ensuring that the system is both resilient and scalable. It's possible that using a decentralized approach would reduce the amount of time that passes between blocks of information in Blockchain [13]. An overlay scattered network is used in our approach to the problem.

B. Healthcare Providers and Patients

Patients or insurance companies might hire healthcare professionals to carry out diagnostic procedures or to provide therapy initiatives. When they get a network alert, medical practitioners are able to begin treating patients immediately after receiving the notification [13]. Patients have complete authority over the individuals who are granted access to their medical information and are free to revoke consent at any time [5]. This covers the companies that supply their insurance and healthcare services.

C. Smart Contracts

Any agreement that is made via an Internet of Things device has the potential to be converted into a "smart contract" and put into effect when certain criteria are satisfied [18]. Take into consideration the fact that we determined the highest and lowest values of blood pressure for each of our patients. If the smart contract receives a blood pressure measurement from a wearable device that is outside of the normal range, it will send an alarm to the appropriate individual or healthcare professional and save the abnormal data in the cloud. This will allow the provider to see the data in real time. The contemporary world has given rise to a wider variety of potential uses for RFID technology. This technology has seen widespread use for the automation of industrial processes as a direct result of the complicated manufacture of consumer goods such as motor cars, home appliances, and other products (refrigerators and washing machines) [11]. The Vatican Library, which possesses more than two million books, uses a technology called radio frequency identification (RFID) to speed up inventory and book searches, automate book distribution, and assist in the prevention of theft. Over seven hundred of the major libraries in the world have either already accepted this technology or are in the process of adopting it [2, 3].

RFID chips are being embedded into all newly issued passports by a growing number of countries throughout the globe. These travel papers are referred to as biometric e-passports, and they include a chip that has the same information as the printed pages of the passport. The healthcare industry is "establishing a foothold" for this technology [14]. A woman and her baby might be connected at the hospital via the use of an RFID bracelet. In traditional hospitals, they are often used to keep track of patients who need continuing medical attention by following their movements throughout the facility. The unique usage of the idea of a wireless sensor network [4] that is used to monitor and change physical things is shown by connecting the tracker to the heart rate monitor. The connection between these devices and GPS trackers, mobile phones, social networking platforms, cloud servers, and big data analytics tools is now essential to the functioning of the Internet of Things (IoT) [12].

IV. Proposed Methodology

This research and its methodology, suggest new research methods, newly apply, and interpret the existing methods, also it complements scientific theories, concepts, and models with new data and newly interprets this data, presents new empirical data obtained through the application of the existing and new research methods, and independently collected, processed, and analyzed by the researcher [20].

Methodology of this research contains:

1) Data collection.
2) Data preparation.
3) Data analysis.
4) Data processing Costs.
5) Data Acquisition.

A. Data Collection

Data should be collected from multiple resources, like public and private sectors hospitals and clinics, most of the data could be collected from the ministry of health.

B. Data Preparation

Data should be classified according to each major and specialty in the health industry, like, Cardiology, Dermatology, Chest, ENT, gastroenterology, Nephrology, Orthopedics, and Ophthalmology. In the end, each patient should have like a folder that contains all his history in every specialty.

C. Data Analysis

It was taken from various perspectives through various technologies: The blockchain is a decentralized, auditable, and reliable ledger for storing and transferring data. It is a database that tracks the activity of the various users it serves in its most fundamental form. Utilizing this decentralized database should not put you in danger. Users are supplied with the option to independently check the chain's validity via distributed data storage and verification. [3] Blockchain technology will help in saving all information for one patient in a block taking its hash function as a reference to another block, to build up a chain of blocks that configure a blockchain. Internet of Thing (IoT), which help in collecting data from cloud sources or databases that have the patient information got from sensors applied in bracelets, watches, and electronic cards. Big Data it’s a concept of how to save structured and unstructured data combined from all sources that contain all information needed, through its techniques.

D. Data Processing Costs

Another key challenge that modern healthcare organizations have to overcome is the high cost of gathering and distributing patient data. The vast majority of time is used by sharing the patient's medical history to other facilities. It is both time-consuming and expensive to compile a patient's complete medical history from the several, non-computerized hospital record systems that are available [15]. It is possible that the administrative expenses generated by present healthcare systems for third parties will be reduced if blockchain technology is used. In addition to this, it enables flexible data access to the patient's medical record. The patient's medical record is compiled and kept up to date using information from a range of sources, such as personal wearables, mobile devices, and patient records. It is possible that blockchain technology may reduce the operational costs of healthcare providers since it would consolidate patient data in a single location.

The process of getting useful data from extensive databases is referred to as information extraction [4]. The strategy is dependent on the manner in which the information is gathered [5]. "Having access to the data is fantastic, but it serves no use if we are unable to turn it into something of value." Because of this, it is the most important component of big data, since it is the area in which the majority of businesses invest their money and resources in order to gain knowledge and make money [16].

E. Data Acquisition

The majority of data-collecting systems take analogue wave patterns and convert them into digital values so that they can be processed more easily. A data collection system is comprised of a variety of components all working together to form the whole. Sensors are devices that collect data from the real world and convert it into a format that computers and other electronic devices can interpret. Signal conditioning circuitry that transforms analogue sensor data into a format that can be successfully digitized. Following conditioning, sensor signals are transmitted to an analogue-to-digital converter, which receives the digitized data. Through data acquisition, we can close the circle of integrating the internet of things with blockchain and big data. The majority of the time, software applications used for data collection are managed by programs written in one of several general-purpose programming languages. [2].

V. Healthcare Proposed Model IOMCT

Proposed Model Phase 1

A. Internet of Medical Card Thing (IOMCT) Phase 1

The IOMCT model represents the cycle of patient data collection.

Phase one as shown in Fig. 2 talks about two paths for data collection:

When the IoT that is represented in the patient bracelet detect the patient’s heart pulse, Sugar Rate, and pressure Rate. Then add this information to the patient Block through the cloud storage. So that it will update the patient history, which will be automatically added to the patient card when uploading the updated information through the cloud storage. Data collection is formed in the hospital when the patient presents the card to the reception desk. It only needs just one tap to check the patient’s data.

B. Internet of Medical Card Thing (IOMCT) Phase 2

Phase two as shown in Fig. 3 talks about the rest of entire procedure that represented in validating the patient data in couple of steps.

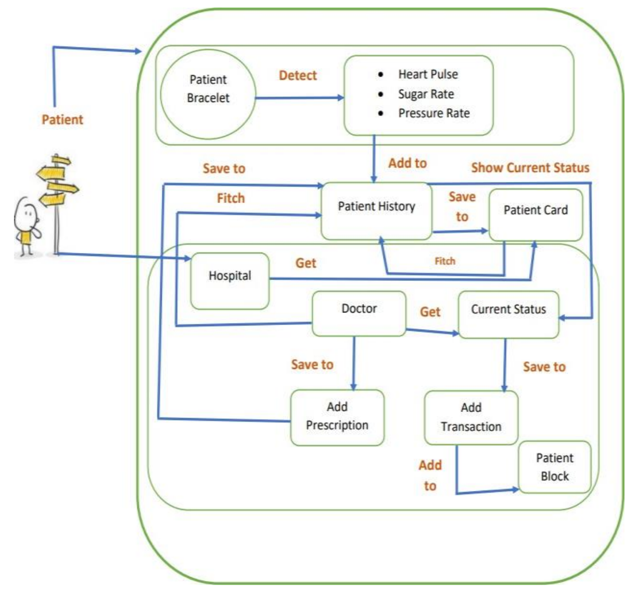


Fig. 2. IOMCT model Phase 1.

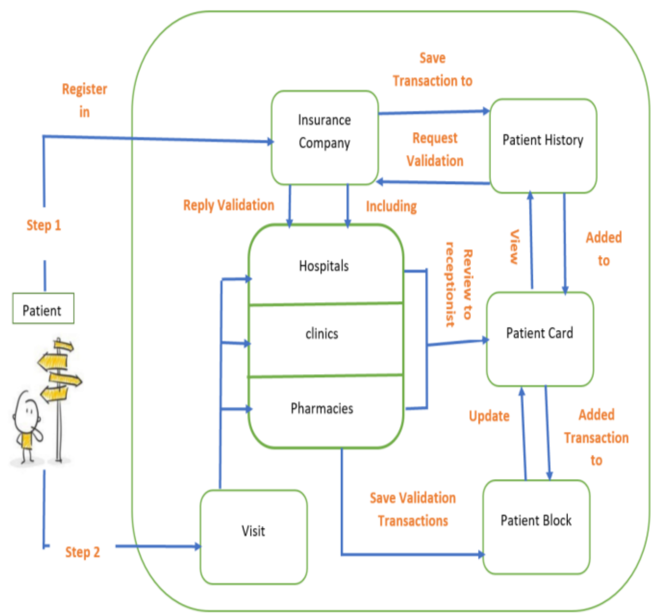


Fig. 3. IOMCT model Phase 2.

First step, with the insurance company whether for hospitals, clinics, or pharmacies and save validations results to the patient history that will be added directly to the patient Block through cloud storage to update the patient Card information.

Second step, when patient visit any of the hospitals, clinics or pharmacies, this visit will be updated in the patient block and added to the patient card so that data will be automatically maintained and updated whenever any new visit happen.

VI. HEALTHCARE PROPOSED BLOCKCHAIN ALGORITHMS

The Tables I and II show the IOMCT model algorithm phase 1 and phase 2 that present the Blockchain creation and the hash creation.

TABLE I. IOMCT ALGORITHM PHASE 1

Algorithm 1: Blockchain Creation

Input: hashlib, json, Time
Output: new blockchain
Procedure:

-Step1: Create essentially chain object. x = essential block. chain
Y = Current transaction z = nodes.
X []
x.y = []
x.z = set []

-Step2: Create new Block for the created chain in step1.
x. new block (p.hash= 1, proof=100) where
p.hash = previous block hash =
SHA1

-Step3: Generate hash for blocks.
Define hash (block):
Block_string = json.dumps (block.
Sort_key=true). encode ()
Where json used for sorting
&exchanging data &
Generate SHA256 value.
Return.hashlib. sha256
(block_string).hexdigest().
Where hashlib is a module for hashing messages.

-Step4: Create new block in the created blockchain
Define new_block (x, proof, p.hash = Name):
Where
Block = {'index' = len(x.chain)+1,
'timestamp'= time(),
'transactions': x.y,
'proof' = proof,
'p.hash': p.hash or x.hash
(x [-1]),
}
x.y = [] x.append
(block)
return block
where x.y [] for resting the current
list of transaction

-Step5: Create last block
Define last_block (x):
Where last_block (x) return the last block in the
chain(x)
return x [-1]

End procedure

TABLE II. IOMCT ALGORITHM PHASE 2

Algorithm 2: Blockchain Creation
Input: hashlib, json, Time, x [essential block.chain], y [current transaction]
Output: add new transactions to blocks & validate their data
Procedure:
-Step1: Define new transactions statement Define N_T (x, sender,recipient, amount):
x.y.append ({“sender”: sender,
“recipient”: recipient, “data”: amount,})
Return int (x.last_block[‘index’]) +1 Where N_T = new_transaction
-Step2: Define proof of work statement.
Define proof_of_work (x, last_proof): Proof = 0
While x. validate_proof (last_proof, proof) is false:
Proof + = 1 Return proof
Where pr is containing 4 zeroes & pr is the previous pr
-Step3: validate proof
Define validate_proof (last_proof, proof): Guess= pr {last_proof}{proof}
.encode()
Guess_hash =hashlib.sha256 (Guess).hexdigest()
Return Guess_hash [:4]==”0000”
Where validate_proof shows if
hash (last_proof, proof) contains 4 zeroes or not?
End procedure

VII. RESULTS

Results in this paper were taken from two stages:

1) Development Stage: that presents code written using the python tool.

2) Interface designed application stage, with the aid of a designed card handles all information about a certain patient that will hold the card: this design is connected with the code forming the results shown below.

- First Stage: Development Stage: The code below will initiate the new blockchain node and create a new globally unique identifier as shown in Fig. 4.

```
app = Flask(__name__)

node_identifier = str(uuid4()).replace('-', '')

blockchain = BlockChain()
```

Fig. 4. Unique identifier.

The blockchain class object definition by initiating:

- New chain array.
- Current transactions array.
- New Node creation.
- New block creation and setting the hash method and proof (success) as shown in Fig. 5.

```
class BlockChain(object):
    """ Main BlockChain class """
    def __init__(self):
        self.chain = []
        self.current_transactions = []
        self.nodes = set()

        self.new_block(previous_hash=1, proof=100)
```

Fig. 5. New block creation.

The code below shown in Fig. 6 describes the new block by setting the index, timestamp, current transaction, proof, and previous hash. According to the code above shown in Fig. 5, the previous hash is passed with the value one. However, if it is passed by none the value will be set according to hashing value done in the previous chain block value.

```
def new_block(self, proof, previous_hash=None):

    block = {
        'index': len(self.chain)+1,
        'timestamp': time(),
        'transactions': self.current_transactions,
        'proof': proof,
        'previous_hash': previous_hash or self.hash(self.chain[-1]),
    }
```

Fig. 6. Hashing value.

```
@staticmethod
def hash(block):

    block_string = json.dumps(block, sort_keys=True).encode()
    return hashlib.sha256(block_string).hexdigest()
```

Fig. 7. Code value.

The code shown above in Fig. 7 was developed to hash the block using the SHA256 protocol .After setting the block, we will empty the current transaction array. Then, adding the new block to the current node as shown in Fig. 8.

```
self.current_transactions = []
self.chain.append(block)
return block
```

Fig. 8. Chain protocol.

The code shown in Fig. 9 and Fig. 10 is for new transaction creation in the blockchain and setting the transaction values. However, before submitting the transaction received, we will check if the required values are passed or not (sender, recipient, and amount).

```
def new_transaction(self, sender, recipient, amount):

    self.current_transactions.append({
        "sender":sender,
        "recient":recipient,
        "data":amount,
    })
    return int(self.last_block['index'])+1
```

Fig. 9. New transaction creation with setting values.

```
@app.route('/transaction/new', methods=['GET'])
def new_transaction():

    values = request.get_json()
    required = ['sender', 'recipient', 'amont']

    if not all(k in values for k in required):
        return 'Missing values.', 400

    index = blockchain.new_transaction(
        sender = values['sender'],
        recipient = values['recipient'],
        amount = values['amount']
    )

    response = {
        'message': f'Transaction will be added to the Block {index}',
    }
    return jsonify(response, 200)
```

Fig. 10. Registration nodes.

Registering new nodes by getting the JSON object and then searching for node values as shown in Fig. 11. The code checks if there are no nodes. So, it will give an error. Otherwise, iterating on nodes and adding them to the current block.

```
@app.route('/nodes/register', methods=['POST'])
def register_nodes():
    values = request.get_json()

    print('values',values)
    nodes = values.get('nodes')
    if nodes is None:
        return "Error: Please supply a valid list of nodes", 400

    for node in nodes: blockchain.register_node(node)

    response = {
        'message': "New nodes have been added",
        'all_nodes': list(blockchain.nodes),
    }

    return jsonify(response), 201
```

Fig. 11. New nodes.

The new node addition will be done by calling the below code shown in Fig. 12.

```
def register_node(self, address):

    parsed_url = urlparse(address)
    self.nodes.add(parsed_url.netloc)
```

Fig. 12. Additional new nodes.

Fig. 13 shows how to find the whole chain in the current blockchain by passing the chain value along with the length of the chain.

```
@app.route('/chain', methods=['GET'])
def full_chain():
    response = {
        'chain': blockchain.chain,
        'length': len(blockchain.chain),
    }
    return jsonify(response), 200
```

Fig. 13. Whole chain value.

Now in Fig. 14 shown below, it finds the latest block in the chain.

```
@property
def last_block(self):

    return self.chain[-1]
```

Fig. 14. Latest block.

The below code as shown in Fig. 15 validates if the proof is like the previous proof containing the same structure with leading four zeros in both.

```
def proof_of_work(self, last_proof):

    proof = 0
    while self.validate_proof(last_proof, proof) is False:
        proof += 1
    return proof
```

Fig. 15. Mining block.

Before mining the block, some inputs need to be validated like the below code shown in Fig. 16 to validate the proof by encoding the last proof values using SHA256 protocol and then validating if the proof does contain four leading zeroes or not.

```
@staticmethod
def validate_proof(last_proof, proof):

    guess = f'{last_proof}{proof}'.encode()
    guess_hash = hashlib.sha256(guess).hexdigest()
    return guess_hash[:4] == "0000"
```

Fig. 16. Input validation.

After submitting the transaction, the next step is mining as shown in Fig. 17. So, the code below will be used to mine the newly submitted block through the following:

1) Running the proof of work.

2) Submitting new transaction.

3) Rewarding the mining server for finding the proof by one coin.

4) Hashing the block.

5) Adding the new block.

- Second Stage: Interface designed application stage

As mentioned in this paper, the data of health records will be distributed and secured through the blockchain, however, we need to show the data using different channels like mobile, tablets or other devices that will be used by medical authorities or doctors. We believe that this data should be managed dynamically and completely by the patient to show only the piece of information required. This will be achieved through a health record card that will be owned by the patient and s/he can control this data through a health record administration app.

The health record will be shown to the doctors, and authorities by tapping the card on an NFC reader that could be mobile, tablet or handheld that has NFC embedded to show the health record as shown in Fig. 18.

The health record will be shown on the device according to a piece of information selected through the medical record administration app as shown in Fig. 19.

```
@app.route('/mine', methods=['GET'])
def mine():

    last_block = blockchain.last_block
    last_proof = last_block['proof']
    proof = blockchain.proof_of_work(last_proof)

    blockchain.new_transaction(
        sender=0,
        recipient=node_identifier,
        amount=1,
    )

    previous_hash = blockchain.hash(last_block)
    block = blockchain.new_block(proof, previous_hash)

    response = {
        'message': "Forged new block.",
        'index': block['index'],
        'transactions': block['transactions'],
        'proof': block['proof'],
        'previous_hash': block['previous_hash'],
    }
    return jsonify(response, 200)
```

Fig. 17. Mining step.

Fig. 18. Patients' health card.

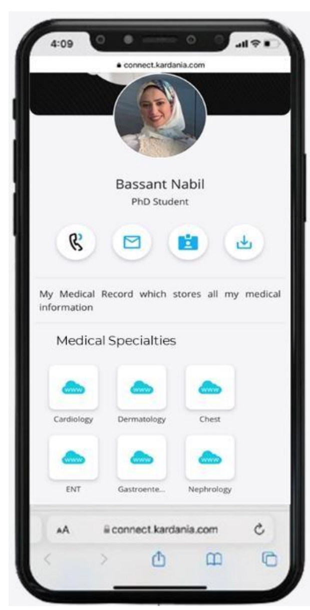


Fig. 19. Patient record.

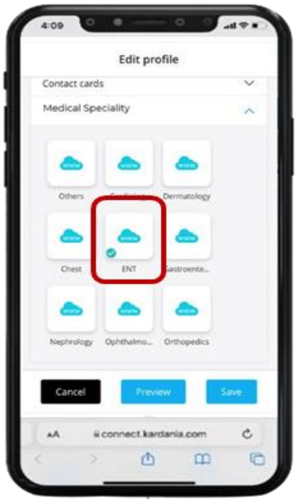


Fig. 20. Segmentation.

The patient can edit his medical profile to select which information is to be shown once the card tapped on the device (ex: ENT) as shown in Fig. 20.

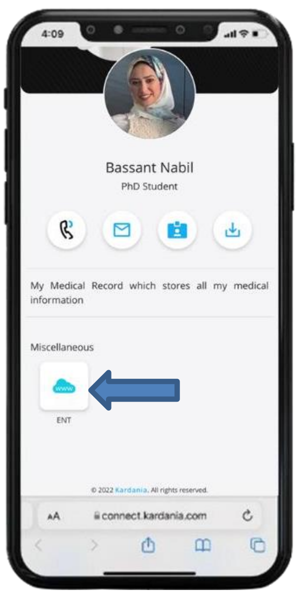


Fig. 21. Profile.

Now, the information that will be shown on the card is what is selected through the administration app. So, the ENT medical record information will be shown, and authorities/doctors can click on it to show the data as shown in Fig. 21.

The document shown in the Fig. 22 mentioned above is the last medication taken by the patient and now the doctor can take a decision accordingly. All this information and documents are synced through the patient block in the ministry of health block chain with all security measures taken benefiting from encryption and distributing data across all servers.

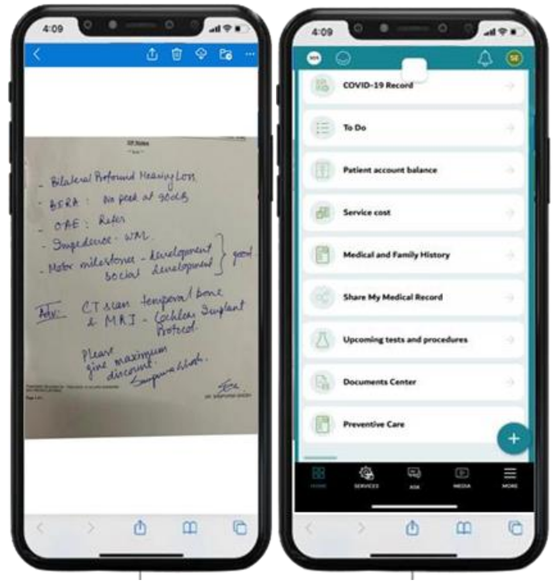


Fig. 22. Documentation.

VIII. Conclusions and Recommendations

Even though blockchain is still thought of as a concept that is in its infancy, a significant amount of study has been carried out on the topic. There are still research holes that need to be filled, even though there has been some development in our understanding of how these things might be used in Hyperledger. The course of this investigation led to the discovery of a potential loophole in the data security that is in place between the participants. The vast majority of the work done in the field thus far has been devoted to investigating a wide variety of subjects, including the transmission of sensitive data, the veracity of distributed systems, various ideas and types of blockchains, blockchain capabilities, and distributed ledger technology.

In addition to above loophole, the policies that should be implemented should be integrated pragmatically to support data compliance with code of ethics in using the online data. A unified health policy based on multidisciplinary partnership is critical to safeguard online data and promote public trust.

During this long journey, there’s a lot of other pillars that will need to be studied like the migration and integration between medical authorities in addition to the cost of ownership and return of investment.

Relying on this paper, a real case is under study now in Abu Dhabi for facilitating to the UAE citizens who are getting some treatments abroad and how they can manage their health record to show the piece of information required from the authorities abroad to avoid sharing the medical history over emails with doctors or through physical documents.

The approach proposed through this paper is serious trail that approaching to connect all stakeholders in one platform taking into consideration the data security and integrity.

Adding to this, there are some benefits will be achieved

- Better healthcare service quality.
- Better patient experience.
- Holistic connected care
- Unified patient medical record.
- Increasing the efficiency of services provided.

ACKNOWLEDGMENT

The research reported in this paper has been created through combined efforts. Thanks, are due to Dr Hatem Abdelkader and Kardania [21] team for their valuable comments and feedback and contribution on the early drafts and ideas for assisting with preliminary queries on the Trove database used to implement the practical part shown.

REFERENCES

[1] Ratta, P., Kaur, A., Sharma, S., Shabaz, M., & Dhiman, G. (2021). Application of blockchain and internet of things in healthcare and medical sector: applications, challenges, and future perspectives. Journal of Food Quality, 2021. https://en.wikipedia.org/wiki/Data_ac question. Last seen on 25/8/2022.

[2] Ktari, J., Frikha, T., Ben Amor, N., Louraidh, L., Elmannai, H., & Hamdi, M. (2022). IoMT- Based Platform for E-Health Monitoring Based on the Blockchain. Electronics, 11(15), 2314.

[3] Lee, T. F., Chang, I. P., & Kung, T. S. (2021). Blockchain-based healthcare information preservation using extended chaotic maps for HIPAA privacy/security regulations. Applied Sciences, 11(22), 10576.

[4] Srivastava, G., Crichigno, J., & Dhar, S. (2019, May). A light and secure healthcare blockchain for IoT medical devices. In 2019 IEEE Canadian conference of electrical and computer engineering (CCECE) (pp. 1-5) IEEE.

[5] Ray, P. P., Dash, D., Salah, K., & Kumar, N. (2020). Blockchain for IoT-based healthcare: background, consensus, platforms, and use cases. IEEE Systems Journal, 15(1), 85-94.

[6] Azbeg, K., Ouchetto, O., Andaloussi, S. J., & Fetjah, L. (2021). A taxonomic review of the use of IoT and blockchain in healthcare applications. Irbm.

[7] Rayan, R. A., Tsagkaris, C., & Iryna, R. B. (2021). The Internet of things for healthcare: applications, selected cases, and challenges. In IoT in Healthcare and Ambient Assisted Living (pp. 1-15). Springer, Singapore.

[8] Al-Rawashdeh, M., Keikhosrokiani, P., Belaton, B., Alawida, M., & Zwiri, A. (2022). IoT Adoption and Application for Smart Healthcare: A Systematic Review. Sensors, 22(14), 5377.

[9] Bazel, M. A., Mohammed, F., & Ahmed, M. (2021, August). Blockchain Technology in Healthcare Big Data Management: Benefits, Applications and Challenges. In 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA) (pp. 1-8). IEEE.

[10] Rayan, R. A., Tsagkaris, C., & Iryna, R. B. (2021). The Internet of things for healthcare: applications, selected cases and challenges. In IoT in Healthcare and Ambient Assisted Living (pp. 1-15). Springer, Singapore.

[11] Srivastava, G., Crichigno, J., & Dhar, S. (2019, May). A light and secure healthcare blockchain for iot medical devices. In 2019 IEEE Canadian conference of electrical and computer engineering (CCECE) (pp. 1-5). IEEE.

[12] Pablo, R. G. J., Roberto, D. P., Victor, S. U., Isabel, G. R., Paul, C., & Elizabeth, O. R. (2022). Big data A synergy with artificial intelligence and blockchain technology. Journal of Integrative Bioinformatics, 19(1).

[13] Adere, E. M. (2022). Blockchain in healthcare and IoT: A systematic literature review. Array, 100139.

[14] Alam, T. (2021). Blockchain cities: the futuristic cities driven by Blockchain, big data and internet of things. GeoJournal, 1-30.

[15] Bazel, M. A., Mohammed, F., Alsabaiy, M., & Abualrejal, H. M. (2021, September). The role of Internet of Things, Blockchain, Artificial Intelligence, and Big Data Technologies in Healthcare to Prevent the Spread of the COVID-19. In 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT) (pp. 455-462). IEEE.

[16] Karafiloski, E., & Mishev, A. (2017, July). Blockchain solutions for big data challenges: A literature review. In IEEE EUROCON 2017-17th International Conference on Smart Technologies (pp. 763-768). IEEE.

[17] Simić, M., Sladić, G., & Milosavljević, B. (2017). A case study IoT and blockchain powered healthcare. Proc. ICET, 1-4.

[18] M. Bublitz, F., Oetomo, A., S. Sahu, K., Kuang, A., X. Fadrique, L., E. Velmovitsky, P., ... & P. Morita, P. (2019). Disruptive technologies for environment and health research: an overview of artificial intelligence, blockchain, and internet of things. International journal of environmental research and public health, 16(20), 3847.

[19] Qadri, Y. A., Nauman, A., Zikria, Y. B., Vasilakos, A. V., & Kim, S. W. (2020). The future of healthcare internet of things: a survey of emerging technologies. IEEE Communications Surveys & Tutorials, 22(2), 1121-1167.

[20] Kelly, J. T., Campbell, K. L., Gong, E., & Scuffham, P. (2020). The Internet of Things: Impact and implications for health care delivery. Journal of medical Internet research, 22(11), e20135.

[21] Damaty, S. B., SD, Agha, T. H., PD, & Abd Al latif, T. A., PhD (1). One Tap, Endless Possibilities. kardania. https://kardania.com/.

Multi Oral Disease Classification from Panoramic Radiograph using Transfer Learning and XGBoost

Priyanka Jaiswal[1], Dr. Vijay Katkar[2], Dr. S.G. Bhirud[3]
Professor[3]
Department of CE & IT, Veermata Jijabai Technical Institute, Mumbai-400019, India[1, 3]
Department of IT, YCCE, Nagpur[1]
Associate Professor, Department of Computer Engineering, Marwadi University, Rajkot, India[2]

***Abstract*—The subject of oral healthcare is a crucial research field with significant technological development. This research examines the field of oral health care known as dentistry, a branch of medicine concerned with the anatomy, development, and disorders of the teeth. Good oral health is essential for speaking, smiling, testing, touching, digesting food, swallowing, and many other aspects, such as expressing a variety of emotions through facial expressions. Comfort in doing all these activities contributes to a person's self-confidence. For diagnosing multiple oral diseases at a time panoramic radiograph is used. Oral healthcare experts are important to appropriately detect and classify disorders. This automated approach was developed to eliminate the overhead of experts and the time required for diagnosis. This research is based on a self-created dataset of 500 images representing six distinct diseases in 46 possible combinations. Tooth wear, periapical, periodontitis, tooth decay, missing tooth, and impacted tooth are all examples of diseases. This system is developed using the concept of transfer learning with the use of a different pre-trained network such as "ResNet50V2", "ResNet101V2", "MobileNetV3Large", "MobileNetV3Small", "MobileNet", "EfficientNetB0", "EfficientNetB1", and "EfficientNetB2" with XGBoost and to get the final prediction The images in the dataset were divided into 80% training and 20% images for testing. To assess the performance of this system, various measuring metrics are used. Experiments revealed that the proposed model detected Tooth wear, periapical, periodontitis, tooth decay, missing tooth, and impacted tooth with an accuracy of 91.8%, 92.2%, 92.4%, 93.2%, 91.6%, and 90.8%, respectively.**

Keywords*—Panoramic radiograph; dentistry; deep learning; ensemble classifier; multi-disease classification and prediction; oral diseases; weighted ensemble module; XGBoost

I. INTRODUCTION

According to the WHO survey, the population of seniors aged 65 and more has surpassed 150 million, and the population of patients with chronic conditions has surpassed 400 million. Due to the increase in periodontal diseases in the 35-44 age group, the percentage of infection is higher in the 65-74 age group. Additionally, proper oral health is essential for maintaining a healthy heart and digestive system. Aging and mouth problems have a long history of interdependence; a lack of attention to oral health might predispose older individuals to other health conditions, such as malnutrition, heart-related issues, Digestion related problems, etc. [2] This has increased the demand for automated medical and healthcare solutions that are secure and of superior quality. Which must be useful for the early detection and classification of the present disorders of an individual. To perform diagnosis of present tooth-related diseases in the mouth, Currently, dental radiography plays a crucial role in diagnosing the condition of oral health. Due to its accurate confirmation of clinical findings, dental panoramic radiography (DPR) images are attracting a rising amount of attention in the diagnostic process. Typically, three types of dental X-ray radiographs are used for diagnosis in dental imaging. Dental radiographs consist of bitewing, periapical, and panoramic types. Bitewing focuses mostly on the coronal half of teeth, periapical encompasses the apical region of teeth, and panoramic X-ray scans consider the entire upper and lower jaw, including teeth and supporting bones. For disease localization diagnosis, periapical radiography is suggested. But when the disease is widespread, panoramic X-rays are used to obtain a thorough perspective for diagnosis and therapy planning. They are easily standardizable and reproducible.

Fig. 1. Different types of dental radiography

Due to the unspecified quality of x-ray images, varying tooth sizes, and other factors, numerous challenges are associated with automatically separating teeth from dental radiograph images. Fig. 1 depicts the three main types of x-ray images of teeth: bitewing, periapical, and panoramic teeth. The dentist manually clarifies radiographs by identifying each tooth and the corresponding issue. But if the x-ray radiography is unclear, it can lead to misinterpretation during analysis [1]. Conventionally, a dentist uses dental radiographs and a clinical evaluation of the patient to make a diagnosis, based on the available infrastructure and expertise. These numerous methods encourage researchers to use and build new machine learning, deep learning, and dental image comprehension techniques more effectively [2][3]. This approach will help to classify multiple oral diseases automatically.

Image processing is currently the most prominent field in digital healthcare applications; it is also utilized well in the medical environment to analyze illness more precisely [4]. As a result of these developments, medical image processing has become prevalent in both medical and health sectors in analyzing diseases with precise sensitivity and measurement [5]. Similarly, medical image processing is applied to dental information to identify and classify disease types and severity [6]. To identify dental problems, all factors must be considered [7]; this is the predominant trend in the detection of dental diseases. The kind and severity of the condition differed between patients [8] due to inconsistencies in the data. Several mathematical models have been created to predict, identify, assess, and monitor the early oral health of teeth [9]. However, the issues are not addressed because of the complexity of the images [10]. Furthermore, dental images differ depending on the patient's genes and body condition. Predicting dental disease is thus a difficult task. Furthermore, a mathematical model has carried out the different cavity analysis processes practically. The cavity severity analysis procedure has varied depending on the patient's body and tooth conditions [11]. Finally, the current tooth diseases and their causes were identified. Nonetheless, those model features are insufficient for real-time data. Because actual time tooth X-ray data contains whiter spaces, and some tooth X-rays have tooth clips [12] disease identification becomes more difficult. Furthermore, those types of images have required more additional features to identify the affected tooth [13]. As a result, those approaches failed to identify the diseased tooth. Furthermore, a standard filtering process was used to remove the noisy features in the trained data, which increased the cost of the resource [29]. Furthermore, when the dataset images were complex, the filtering results were poor. In other domain concepts, deep learning is used to efficiently estimate remote heart rates and to classify breast lesions in mammography using a conventional neural network, which helped us understand the running complexity of different machine intelligence techniques [14]. These issues prompted the current study to develop optimized deep networks with filtering parameters to obtain more precise disease classification results [17].

The current article has descended into a new solution for maximizing disease classification by utilizing the different pre-trained deep learning models. The proposed study has a step-down in the multi-disease classification model, which is defined as multiple parameters in the proposed framework being upgraded to classify multi-disease features. The following section goes over the specifics and importance of detecting tooth wear, tooth decay(cavity), impacted tooth, missing tooth, periapical, and periodontitis at an early stage. The image of a patient suffering from tooth wear is shown in Fig. 2.

A. *The Necessity of Detecting Tooth Wear in Dentistry*

Teeth have specific anatomy consisting of several cusps and fissures to serve the function of chewing food like a mortar and pestle where the approximating cusps work as a pestle and fissures work as mortar crushing the food coming in between. If these anatomical landmarks are worn-out due to wear, this crushing mechanism gets disturbed and food slips from between the teeth causing a decrease in chewing efficiency ultimately leading to nutritional deficiency and decreased quality of life. Hence teeth wear if diagnosed early can be treated conservatively and a patient can be counseled to prevent further damage. Dental wear has several types which include dental attrition, erosion, abrasion, and abreaction. Dental attrition, abrasion, and erosion are commonly found in the Indian population. All these have specific patterns and causative factors following which they can be diagnosed.

Fig. 2. Tooth wear (erosive)

All these wear patterns can cause irreparable harm to a tooth's structure without creating any noticeable pain in the early stages. When wear and tear reached a severe stage, treatment can be quite costly for the patient. It is possible to avoid getting them by getting checked for them early and making some changes to one's way of living. [15]

B. *The Necessity of Detecting Periodontitis*

The following section describes the details of gingivitis, periodontitis, and the process of diagnosing these diseases using radiographs and clinical examination. Gingivitis is an inflammatory disease limited to gums adjoining the teeth in which the primary signs are redness and bleeding from the gums along with their swelling. Whereas periodontitis is when the inflammation of the gums extends to the adjoining tooth-supporting bone it is derived as periodontitis. Primary signs of periodontitis are loss of tooth-supporting bone and tooth mobility. Methods are different to diagnose periodontitis and Gingivitis as gingivitis can be diagnosed with the help of clinical examination, observing for the signs and symptoms of the disease. Like red, swollen, and bleeding gums. Periodontitis can be diagnosed with the help of clinical and radiographic examinations. Clinically Doctors can measure the loss of supporting bone with the help of a periodontal probe measuring the sulcus depth, and radio graphically the amount of horizontal and vertical bone loss can be identified to gauge the severity and prognosis of periodontitis. The reason for not considering Gingivitis in this implementation is as it is an early stage that can be diagnosed only on clinical examination and here in our research, we had considered OPGs as a sturdy base which is the most preferred diagnostic tool to diagnose periodontitis. By considering all mentioned parameters early diagnosis is needed to prevent the further loss of teeth-supporting tissues and plan for the repair. Fig. 3 represents an image with periapical and periodontitis.

Fig. 3. Periodontitis and periapical

C. The Necessity of Detecting Tooth Caries (Dental Cavity)

Dental caries is a disease that can be diagnosed easily on clinical and radiographic examination. Treatment of decayed teeth depends upon the extent of decay which can be viewed only radiographically. Orthopantomograms are used to diagnose caries between the teeth and over the teeth and Intraoral periapical radiographs and bitewing radiographs are the specific radiographs to detect the extent of the decay. A representation of a tooth with a cavity is shown in Fig. 4.

Fig. 4. Tooth caries/dental cavity

D. The Necessity of Detecting Periapical

periapical infections can only be diagnosed radiographically as clinical symptoms often appear in the delayed stage. In OPG the extent and the size of the periapical lesion can be exactly gauged. It is shown in Fig. 2 and Fig. 7.

E. The Necessity of Identification of Missing Teeth

Often one or more missing teeth do make complex changes in patients' quality of life. One missing tooth can lead to decay and periodontal disease to seven teeth as the teeth approximating the edentulous space and opposing arch teeth drift from their natural definite position. These all changes can be diagnosed only through an OPG. Fig. 5 shows missing teeth. it is observed that two teeth are missing and the last is affected by decay.

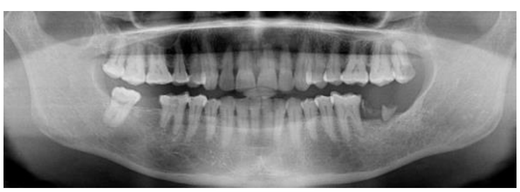


Fig. 5. Missing tooth

F. The Necessity of Identification of Impacted Tooth

Impacted wisdom teeth or any other tooth can’t be seen clinically. OPG gives the exact two-dimensional view of the impacted tooth. A representation of an Impacted tooth is shown in Fig. 6

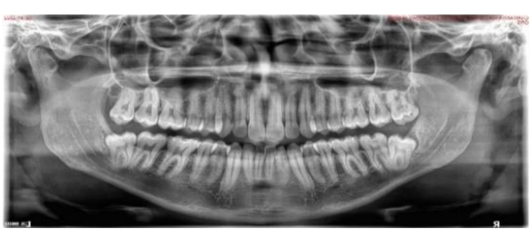


Fig. 6. Impacted tooth

Fig. 7. Wear, decay periapical, and impact

This research work is conceived as a collaborative effort between technical expertise and a dentistry stream in an area of significant interest to the medical community. Literature review reveals that extensive research on tooth extraction or diagnostic purposes used in cybercrimes such as estimation of age, identification of a person based on tooth anatomy, etc, but regrettably, there has been very little work combining the oral health care and engineering fields for continuous engineering and clinical problem-solving. This research began with an emphasis on the various stages of oral illnesses [16]. Such as tooth wear, periapical, periodontitis, tooth decay, tooth loss, and tooth fractures, impacted tooth.

The remaining sections are organized as follows. Section II offers a concise assessment of pertinent literature. It discusses the current state of the art in the use of image processing, machine learning, and deep learning techniques in the domain of dentistry. Understanding different types of radiographs and their importance in evaluating models is discussed using various methods and networks. Section III describes the data collection, radiograph selection, and disease identification processes. The importance of various diseases that can be identified using panoramic radiographs has already been discussed in the introduction section. This section focuses on dataset labeling as well as expert validation of labels. Section IV elaborates on the concept of transfer learning with XGBoost and ensemble learning concerning medical image processing. The proposed methodology is illustrated in Section V, which is followed by the results and discussion and future scope sections.

II. RELATED WORKS

A comparative analysis of the deep learning model for dental segmentation in panoramic radiographs is reported in [18]. The comparison is made between U-Net, DCU-Net, Double U-Net, and NanoNet. On the 1500-image dataset provided by Silva and Olivia. There is also a teeth mask available for this. Their system is designed to identify 32 teeth, fillings, braces, and dental implants. Data enhancement is achieved through random rotation and horizontal inversion. Using a dataset with data augmentation and segmentation without data augmentation based on the segmentation model, the experiment results are shown. The outcome is then compared to the current state of the art.

In this biomedical study [19], the U-Net approach to apical lesion segmentation on panoramic radiographs is discussed. The objective of this study was to extract apical lesions from 470 dental panoramic radiographs. This dataset included 380 images for training, 43 for validation, and 47 for testing. 1140 images were generated after augmentation was applied at the second stage of implementation. For augmentation, horizontal and vertical flips are used. Cropping (pre-processing) is used after augmentation to divide images into four parts: upper right, upper left, lower right, and lower left with size=1000x530. They then multiplied 1140 by 4 for the training group. This dataset is not available to the public.

For apical lesion segmentation from panoramic radiographs, [20] adopts a deep learning technique. There were a total of 1691 images utilized for segmentation. For the training set, all radiographs were manually tagged in red using a polygon labeling tool to construct lesions' contours. The name of the software is "deep stack." The image's original resolution of 1976x976 was reduced to 960x480. Various augmentation techniques, including flip, blur, shift, scale, rotation, sharpness, emboss, contrast, brightness, grid distortion, and elastic transform, are utilized to increase the dataset via online augmentation. Utilizing a pre-trained U-Net CNN for pre-processing and training.

In [21]: Work on a multimodal panoramic x-ray dataset for diagnostic system benchmarking was presented. There are a total of one thousand photos about the labeling of anomalies and teeth. This study was carried out on five levels: anatomical location, peripheral characteristics, radiodensity, surrounding structure effects, and anomaly category. Using eye tracking and a think-aloud methodology, this is the first-time radiologist skill has been captured. This study includes a publicly accessible dataset as well as benchmark performance analyses for a variety of cutting-edge devices. In addition, picture enhancement and image segmentation are carried out. The maxillomandibular region of interest is delineated using x-ray images, gaze maps, and multimodal audio text files. A thousand radiographs and a ground truth mask make up the dataset. 15% is utilized for validation, whereas 85% is used for training. Adam was trained on GPUs with 32 GB of RAM that was NVIDIA Tesla V100 devices. However, in terms of image quality, CLAHE outperforms well with AME Score=24.32 and Log AME=0.3702. Ground truth images were used for segmentation, and the results were compared using UNET, UNET++, PSPNet, DeepLabV3, and DeepLabV3+. VGG19, ResNet18, and ResNet 50 were used to train each of the given networks. The module learning rate is 1x10-3, the weight decay is 1x10-4, the epoch is 150, and the batch size is 16. Pixel accuracy, intersection over union, and dice coefficient are all evaluated. To achieve the desired result, CNN for periodontal bone loss and Squeeze Net for implant fixtures were used. To detect and classify periodontal bone loss, a hybrid framework with Mask RCNN and ResNet101 was used. For intraoral radiographs covering caries, a back propagation neural network is used.

Work on dental caries is being conducted by [c1concepts.]'s (cavity). Utilizing a multi-input deep convolutional neural network ensemble known as MI DCNNE, they were able to obtain accurate results and overcome negative perceptions of panoramic radiography. This system is based on computer-based intelligent vision systems supported by techniques for machine learning and image processing. This new approach accepts both raw periapical images and an improved version of them as inputs. In total, 340 photos were considered for this study. The model has a 99.13% accuracy rate when diagnosing dental caries from periapical pictures.

The primary goal of this work [25] is to extract the area of interest (ROI). Because of background noise, the original image may contain useless information. We occasionally come across images that are of poor quality. Because of this, separation is critical. U-Net was used in three phases to accomplish this: one feature encoder, corresponding decoder, and multipath feature extractor.

The author of [26] proposed an automatic teeth recognition model based on residual network-based faster R-CNN techniques. This model is implemented in two stages. The results obtained in step one, i.e., using R-CNN, are refined further using candidate optimization techniques. These improvements improve the results by about 1%. Ten-fold cross-validation is used in this system to test the model's feasibility and robustness. This model is used to describe the various tooth categories in a panoramic radiograph.

The implementation in [27] is based on the use of u-net architecture to perform panoramic radiograph segmentation. They used CNN to achieve accurate panorama image segmentation. With a 94% accuracy and a 94% dice score. The performance of the u-net is compared with various segmentation techniques in this work. This fuzzy c-means (FCM) is developed using a hybrid of FCM and a neutrophilic approach to segment the jaw and lesion present in the jaw using a panoramic radiogram. This paper investigates the comparison of global thresholding, fuzzy c-means, watershed, canny edge detection algorithms, and the U-NET model. When compared to the other techniques, n-net performed the best. Ivison lab provided the dataset for this study.

In [28], researchers offered an innovative approach to segmenting and classifying data. They have covered 15 various disorders including healthy teeth, missing teeth, dental restoration implants, fixed prostheses, mobile prostheses, etc. To depict all 15 diseases, a total of 2,000 images are made. All these images were taken at separate dental clinics. A convolutional neural network (CNN) was trained to perform semantic segmentation on images from the dataset, and many techniques were employed to enhance the annotated data. Images are segmented and then binarized with multiple thresholds. The detection of teeth is a two-stage algorithmic process. When discrepancies are uncovered during semantic segmentation, refinement techniques are employed to rectify the situation. In this study, the Nobel approach to teeth detection and dental problem classification using deep learning and image processing was introduced.

The paper [22] compares ten different segmentation methods for use with dental imaging. The offered solutions were evaluated and classified using five different metrics: accuracy, specificity, precision, recall, and F1-score. Because of the shattered bones in the buccal cavity, not one of the ten segmentation methods tried was successful in isolating the teeth.

With the help of a mask region-based convolution neural network, the authors of [23] suggested a method for instance segmentation of teeth in panoramic photos. In the wake of Resnet-101-driven feature extraction, a feature pyramid network (FPN) is built with predetermined anchors and extracted regions of interest.

Each region's proposal network consists of the FPN and its anchors (RPN). The regions of interest are then repositioned so that they are all the same size. Additionally, the bounding box coordinates are used to classify each feature as either a tooth or a background and to pinpoint its exact location. Finally, a bounding box is formed around the tooth once it has been segmented.

DeNTNet, described in [24], is a deep neural transfer network trained to identify panoramic dental radiographic evidence of periodontal bone loss (PBL). During detection, several convolutional neural networks are trained. A segmentation network is then trained to extract teeth from the ROI, and a second network is trained to predict areas of periodontal bone loss. Using the encoder portion of the lesion segmentation network as a pre-trained model, a classification network is constructed to predict the presence of PBL in each tooth.

Table I provides a complete overview of the various datasets available, the work done in dental image processing, and the disorders linked with it. It has been observed that work is being done in various disciplines of dentistry employing artificial intelligence techniques. To begin, a private dataset is produced and not made public. Another observation is that most systems are designed to segment teeth. Furthermore, the minimum one and maximum four diseases covered for classification and detection are one. One method is used, in which 14 diseases are investigated, but just the highlighting of diseases is done, rather than classification or detection.

TABLE I. ANALYSIS OF AN AVAILABLE DATASET AND METHODOLOGY USED

Papers	Variable detected	Total Images	Publicly available	Labeled Teeth	Labeled abnormalities	Methodology Used
[36]	Teeth	100	No	Yes	No	CNN
[37]	Teeth	100	No	Yes	No	CNN
[38]	Teeth	1500	Yes	Yes	No	Mask R-CNN
[39]	Teeth	1352	No	Yes	No	CNN
[40]	Endodontic	85	No	No	Yes	CNN
[41]	Endodontic	300	No	Yes	No	CNN
[42]	Teeth	1000[116]	Yes (Kaggle)	No	No	CNN
[32]	Teeth	1200	No	Yes	No	Yolov3
[43]	Teeth	1000	No	Yes	No	Deep Learning and Image Processing Techniques
Own Dataset	Teeth	500	No	No	Yes	Transfer Learning and XGBoost

III. DATASET DESCRIPTION

Researching any specific domain data availability is the primary step, but in the field of dental informatics dataset availability is a major concern. After analyzing all the related work in Dental Image Processing (DIP), we discovered that obtaining a suitable dataset for the early stages of research is challenging. As a result, we collaborated with many dental healthcare sections to acquire data. Rather than waiting for new patients for each ailment, we chose to screen panoramic dental X-rays from electronic medical records to collect data efficiently. There is no ethical issue because these images do not contain any personally identifiable information. The dataset consists of images and a questionnaire, which was supplementary at the initial level of research. Collecting symptoms and images for every patient is a time-consuming task so we focused on a collection of radiographs. Principally there are many types of radiographs available in dental radiography, such as Bitewing X-ray, Periapical X-ray, Occlusion X-ray, Panoramic X-ray, Cephalometric Projection, and Cone Beam X-ray. Among these categories, Panoramic OPG (Orthopantomogram) is considered as is two-dimensional radiograph that covers maximum all diseases with complete coverage. Dental radiographic images in OPG format were collected.

A. Panoramic (OPG) X-Ray and its Benefits

The maxillary and mandibular teeth, as well as the surrounding periodontium and certain anatomical landmarks like the maxillary sinus and the temporomandibular joints, are all visible on an OPG, a two-dimensional radiograph. Oral disorders such as periapical abscess, cysts, osteitis, and various tumors and cysts of mandibular and maxillary origin can be diagnosed using OPG X-ray. OPG is useful for determining how much bone has been lost due to periodontal disease and how severe it is. OPG X-ray is useful for diagnosing a wide variety of dental conditions, including cavities, root fractures, and periodontal inflammation. OPG X-rays can help identify and evaluate dental cusp fracture and enamel deterioration, but their two-dimensional nature has its limitations.

The data was collected from Government Dental College and several private dental clinics and laboratories in Nagpur and cities across India. Here image collection work is focused on the wear of the tooth to find erosion, attrition, and abrasion of the tooth and other oral diseases, including impacted tooth, periapical lesion, fractured tooth, missing teeth, etc.,

B. Process of Collecting, Preparing and Labeling the Data

This section describes the process of collecting, preparing, and labeling the dataset. The main contribution while collecting images and preparing the dataset are expressed with the help of the diagram given below:

a) From 2019 to 2021, we collected approximately 2,000 OPG radiographs from multiple sources.

b) All images were then checked for contrast and illumination quality.

c) Images of poor quality and nonrelevant, such as images with prostheses, multiple caps, braces, etc were removed from the dataset.

d) Diseases that can be seen through OPG are identified and discussed with clinicians.

e) OPG radiographs can be used to diagnose diseases such as caries, periodontitis, periapical infection, impacted teeth, tooth wear, bone fracture, missing teeth, and anatomy of the maxillary sinus and temporomandibular joint.

f) Following disease identification, images are labeled with the assistance of various dental experts (such as Oral and Maxillofacial Radiologists, Prosthodontists, Oral and Maxillofacial Surgeons, Periodontics, and endodontists,).

g) Image labeling is done in two ways: radiograph with multiple diseases and radiograph with a single disease.

h) After the image sorting process, it is discovered that the approximate probability of receiving a relevant radiograph is 50/500. Finally, we can filter 500 useful radiographs from available images.

i) Before developing any module, various pre-processing steps are performed to create the extension of the data set, In this automatic cropping of the image is done to extract the relevant portion by using an automatic cropping algorithm.

j) Our dataset is cross-validated and tested with the help of three different domain experts to make it more consistent, reliable, authentic, and useful.

k) The final label matrix is generated for different six diseases if the disease is present in that image so marked as 1 otherwise 0 which is shown in Table II: Labelled dataset. For example, in image id 92 wear, decay, missing, and impacted tooth diseases are present.

TABLE II. SAMPLE LABELLED DATASET

Image Id	Wear	Decay	Peri-Apical	Periodontitis	Missing	Impacted
1	0	1	0	0	0	0
92	1	1	0	0	1	1
104	1	0	0	0	0	1
238	1	1	0	0	1	0
347	1	1	1	0	0	0
457	1	0	1	1	0	0

To check the validity and authentication of dataset creation complete supplementary material is available. The panoramic dental images were obtained from a variety of sources, including dental laboratories. The image properties due to different acquisition systems vary in terms of contrast and brightness. The pre-processing challenge was to adjust the contrast and illumination levels of all the collected images so that subsequent stages produced the same set of features corresponding to each class or group. We gathered images that were prone to six major issues: dental decay, dental wear, periapical, periodontitis, missing teeth, impacts, and a combination of all six diseases. An experienced medical practitioner in the respective field. While creating the LABEL matrix great care was taken to avoid any false labeling in the LABEL matrix. Images with low contrast and illumination were discarded after being manually sorted one by one. Duplicated images were identified using image comparison and separate codes. Finally, 500 images were taken for this research project, while the remaining unclear images were discarded. Fig. 8 shows the analysis of the total number of images available in the dataset.

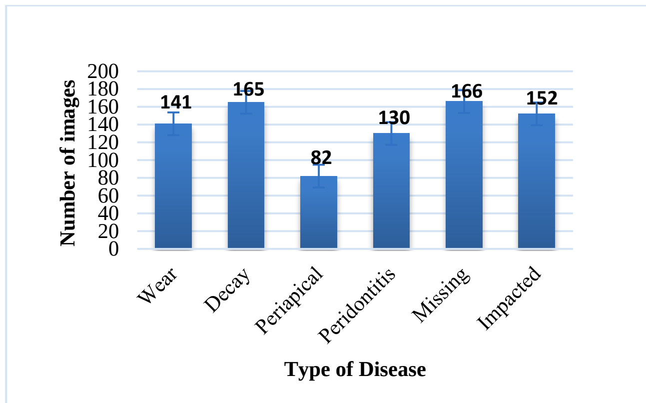


Fig. 8. Analysis of images available in the dataset

IV. TRANSFER LEARNING, XGBOOST, AND ENSEMBLE APPROACH

A. Transfer Learning

Deep feature extraction (DFE) and Transfer learning (TL) are two of the most effective alternatives to employing a small number of images as training samples. When a small sample size of images is present. TL refers to the process of applying previously taught models to unexpected challenges. TL is not a distinct classification of Machine learning algorithms; rather, it is a technique that can be used to create a new Machine learning model. The model will be able to use the knowledge and skills acquired from earlier training in new scenarios. In a manner like the preceding task, data will need to be organized based on the type of data. A further application of TL is the extraction of deep feature data.

Rather than manually modifying the activation layers of the CNN, it is possible to extract feature vectors by utilizing pre-trained CNN models. The deeper layers, which are activated by the activation of the lower-level layers, include the picture classification-critical higher-level features.

Transfer Learning is a technique in which we reuse a previously trained model as the foundation for a new model on a new challenge. In this case, a model trained on one task is repurposed for another. Transfer learning tries to improve target learners' performance on the target domain by transferring knowledge from distinct but related source domains. ResNet50V2, ResNet101V2', 'MobileNetV3Large', 'MobileNetV3Small', 'MobileNet', 'EfficientNetB0', 'EfficientNetB1', and 'EfficientNetB2' are examples of pre-trained networks. [33].

ResNet50V2 [34] is a changed version of ResNet50. It does better on the ImageNet dataset than ResNet50 and ResNet101. ResNet-101 is a 101-layer convolutional neural network.

MobileNetV3 [35] is a convolutional neural network that is tuned to mobile phone CPUs using a combination of hardware-aware network architecture search (NAS) and the NetAdapt algorithm. It was then improved by making new architectural advances.

EfficientNet-B0 is a convolutional neural network that has been trained on more than a million images from the ImageNet database [30]. The network can divide images into 1000 different kinds of objects, like a keyboard, mouse, pencil, and many animals.

B. XGBoost

Extreme Gradient Boosting, which is what XGBoost stands for, is a distributed gradient-boosted decision tree (GBDT) machine learning library that can be used on a large scale. It is the best machine learning library for regression, classification, and ranking problems, and it has parallel tree boosting. It's like Random Forest in that it builds a group of decision trees, but instead of training the models at the same time, XGBoost trains them one at a time. Each new decision tree learns from the mistakes made by the one before it. Boosting is the process of training models one after the other. A gradient in XGBoost stands for a type of boosting that uses weak learners. Weak learners are simple models that only do better than random chance. The algorithm starts with a weak learner at the beginning. Each new model tries to fix the mistakes that the previous decision tree made. This keeps happening until there are no more ways to make the model better. The result is a strong learner model. [31]

C. Ensemble Modeling

Ensemble modeling is a way to predict what will happen based on several different base models. The goal of using a group of models is to reduce the prediction's generalization error. When the ensemble approach is used, the prediction error goes down if the base models are different and can be used on their own. The method tries to figure out what will happen by asking a variety of individuals. Even though the ensemble model is made up of several base models, it works like a single model. [32]

D. Weighted Average or Weighted Sum Ensemble

Weighted average or weighted sum ensemble is a type of machine learning that uses a group of models to make a prediction. The contribution of each model is weighted according to how good it is. The voting ensemble has something to do with the weighted average ensemble. In this method, we didn't tune it; instead, we used the weights that were already there. In other words, a weight tells how much the input affects the output. Biases, which are always the same, are an extra input for the next layer that will always be 1[33][34].

V. PROPOSED METHODOLOGY

We have created a database of 500 images of panoramic radiographs which are explained in section III. Dataset is divided into a training set and a testing set. The training set goes to eight different pre-trained models.

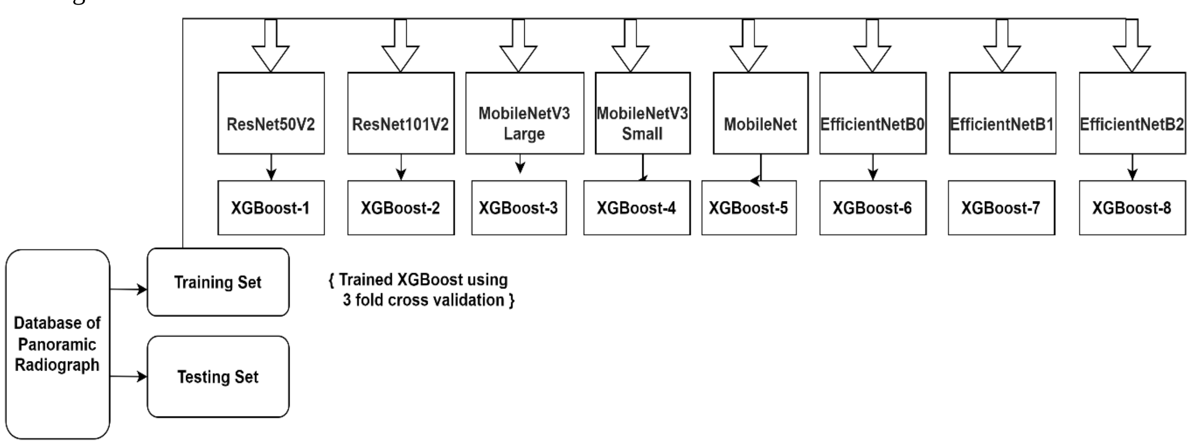


Fig. 9. Proposed methodology (training phase)

Different pre-trained models used for this phase are ‘ResNet50V2', 'ResNet101V2', 'MobileNetV3Large', 'MobileNetV3Small', 'MobileNet', 'EfficientNetB0', 'EfficientNetB1', and 'EfficientNetB2'. Out of 500 samples 80% samples used for training. The detailed process of the training and evaluation phase is given in Fig. 9 and Fig. 10. While training we used 3-fold cross-validation. Cross-validation is a technique for testing models that involves training them on distinct subsets of the available input data and then testing them on the other. Dataset D is divided into three identically sized subsets. The process of fitting and evaluating the model is done three times, and each time a different subset is used as a training sample. At last, in the testing phase given in Fig. 10, we get eight tuned XGBoost. pre-trained networks are used to extract the features and then extracted features are given as input to the XGBoost model for training.

In comparison to other gradient-boosting approaches, XGBoost is nearly ten times faster and has a strong predictive power. Additionally, it contains a range of regularizations that minimize overfitting and enhance overall performance. These eight trained XGBoost will generate eight different classifier prediction outputs. To get the final prediction weighted ensemble module which is an extension of a model averaging ensemble where the contribution of each member to the final prediction is weighted by the performance of the model.

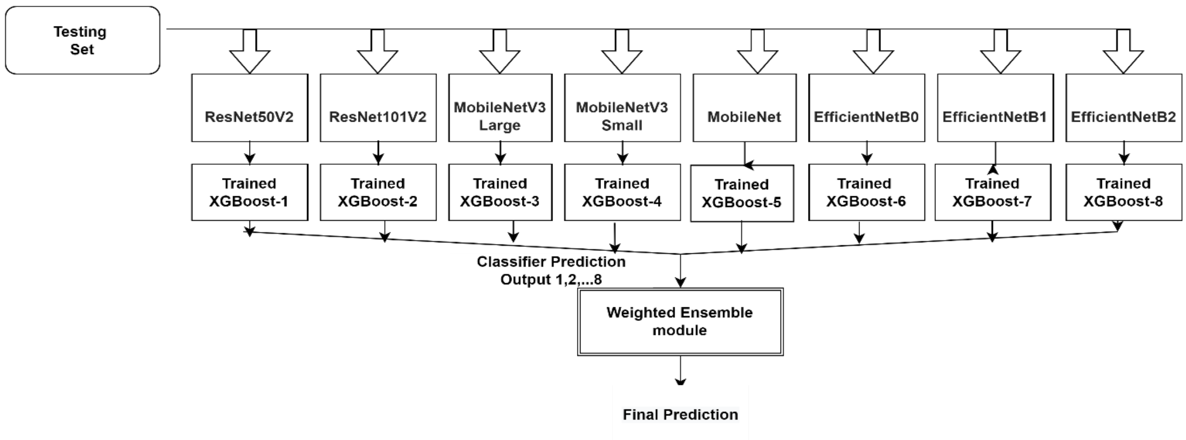


Fig. 10. Proposed methodology (evaluation phase)

VI. Results and Discussion

The developed methodology employing transfer learning is processed using Python; the predicted model's success rate is evaluated utilizing existing mechanisms in terms of accuracy, sensitivity, F-measure, precision, and recall. In this approach, 500 X-ray dental images are utilized by considering 80% for training and 20% for testing. The proposed system can classify multiple diseases which cover six different categories with 46 possible combinations. the dataset covers all types of images like one image with one disease and one image with multiple diseases. For example, this system can identify images where only wear or decay is present, apart from this it is also able to identify and classify images where different diseases are present such as wear, and decay as shown in image id.347. Thus, to prove the sustainability of the proposed work ablation study is performed which is expressed in Tables III, IV, V, and VI by considering different 8 pre-trained networks. Table III shows the performance of the model by considering seven different networks with 91% accuracy. Similarly, Table IV shows the performance of the classification of six different diseases by considering six different networks with 90% accuracy on average.

TABLE III. 7-Pre Trained Models [MobileNetSmall, MobileNet, MobileNetLarge, EfficientNetB0 EfficientNetB1 EfficientNetB2 ResNet50V2]

Performance Measure	Wear	Decay	Periapical	Periodontics	Missing	Impacted
Precision	0.776	0.843	0.717	0.799	0.826	0.773
Recall	0.957	0.912	0.855	0.925	0.91	0.944
F1-Score	0.857	0.876	0.78	0.857	0.866	0.85
Specificity	0.895	0.921	0.933	0.926	0.913	0.888
Accuracy	0.912	0.918	0.92	0.926	0.912	0.904

TABLE IV. 6-Pre Trained Models[MobileNetSmall, MobileNet, MobileNetLarge, EfficientNetB0 EfficientNetB1 EfficientNetB2]

Performance Measure	Wear	Decay	Periapical	Periodontics	Missing	Impacted
Precision	0.767	0.841	0.703	0.793	0.825	0.764
Recall	0.957	0.899	0.855	0.925	0.904	0.944
F1-Score	0.852	0.869	0.772	0.854	0.862	0.845
Specificity	0.89	0.921	0.928	0.924	0.913	0.882
Accuracy	0.908	0.914	0.916	0.924	0.91	0.9

TABLE V. 5-Pre Trained Models [MobileNetSmall, MobileNet, MobileNetLarge, EfficientNetB0 EfficientNetB1]

Performance Measure	Wear	Decay	Periapical	Periodontics	Missing	Impacted
Precision	0.765	0.831	0.689	0.786	0.815	0.756
Recall	0.942	0.899	0.855	0.917	0.904	0.944
F1-Score	0.844	0.864	0.763	0.846	0.857	0.84
Specificity	0.89	0.915	0.923	0.921	0.907	0.876
Accuracy	0.904	0.91	0.912	0.92	0.906	0.896

TABLE VI. 4-Pre Trained Models[MobileNetSmall, MobileNet, MobileNetLarge, EfficientNetB0]

Performance Measure	Wear	Decay	Periapical	Periodontics	Missing	Impacted
Precision	0.765	0.826	0.67	0.786	0.815	0.756
Recall	0.929	0.893	0.852	0.902	0.892	0.932
F1-Score	0.83	0.858	0.75	0.84	0.852	0.834

	9					
Specificity	0.88 9	0.912	0.919	0.921	0.906	0.876
Accuracy	0.9	0.906	0.908	0.916	0.902	0.892

A. Precision

Precision is the proportion of precise teeth affected region computed using Eqn. (1).

$$P = \frac{T_p}{T_p + F_p} \tag{1}$$

Fig. 11 shows the evaluation of precision for different diseases. it is observed that the precision of wear, decay, periapical, periodontitis, missing tooth, and impacted tooth is 79%, 85%, 73%, 82%, 84%, and 78% respectively.

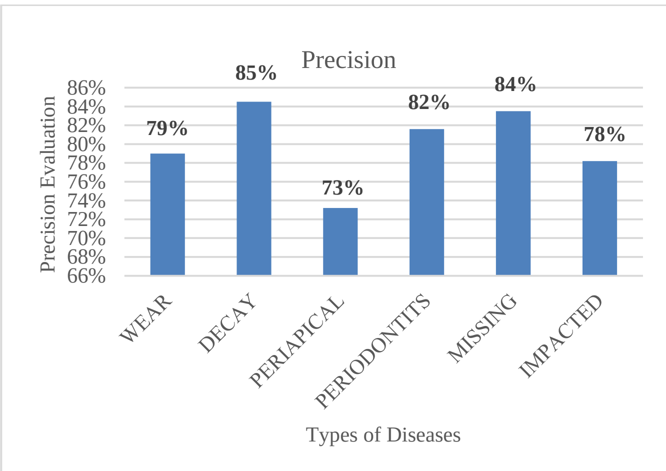


Fig. 11. Comparison of precision for different diseases

B. Recall

It expresses the proportion of predictions that have been correctly diagnosed as expected outcomes and is calculated using Eqn. (2).

$$R = \frac{T_p}{T_p + F_n} \tag{2}$$

It is observed from Fig. 12 that the Recall value for wear, decay, periapical, periodontitis, missing tooth, and impacted tooth is acceptable in the healthcare domain as it is above 95%,93%,86%,93%,91%, and 94% respectively.

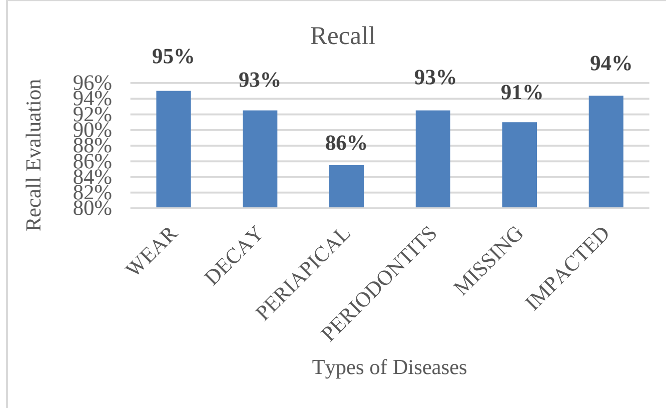


Fig. 12. Comparison of recall for different diseases

C. F-measure

F1-score is measured using Eqn. (3).

$$F1 - score = 2\left(\frac{P * R}{P + R}\right) \tag{3}$$

As per statistics shown in Fig. 13, it is observed that the F1-score for the detection of wear, decay, periapical, periodontitis, missing tooth, and impacted tooth is 86%, 88%, 79%, 87%, 87%, and 86% respectively.

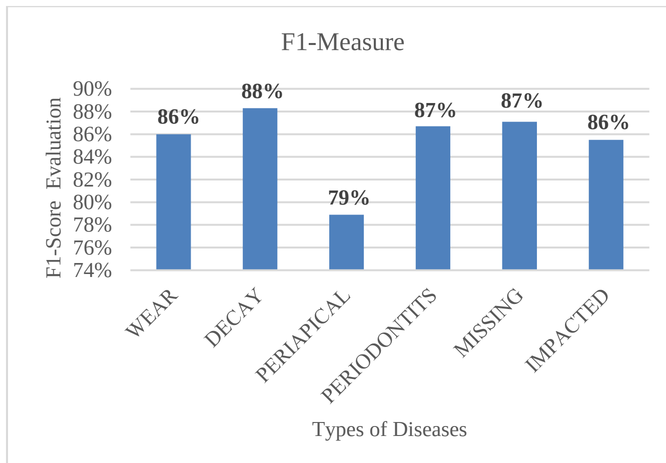


Fig. 13. Comparison of F1-measure for different diseases

D. Specificity

A recall is the probability of a negative diagnosis such as the patient being free from that disease. It is measured using eq. (4).

$$S = \frac{Tn}{(Tn + Fp)} \tag{4}$$

As per the statistical data presented in Fig. 14 specificity for all the diseases covering wear, decay, periapical, periodontitis, missing, and impacted teeth are found to be 90%,92%,94%,93%,92%, and 89% respectively.

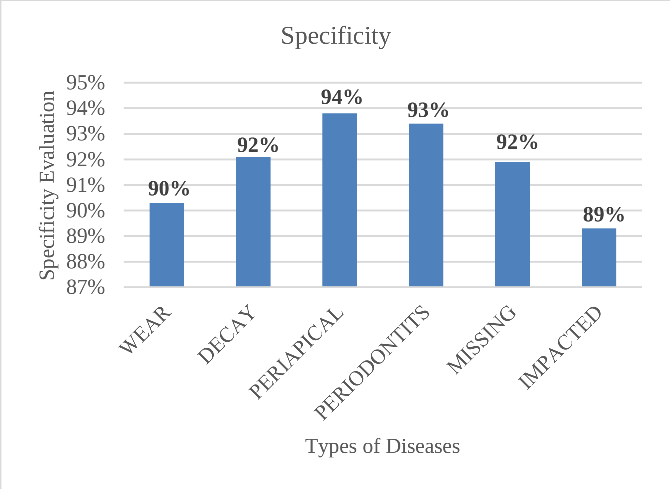


Fig. 14. Comparison of specificity for different diseases

E. Accuracy

Accuracy to the entire observations that are expressed in Eqn. (5),

$$A = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (5)$$

As per statistics shown in Fig. 15, it is examined that accuracy for classification of wear, decay, periapical, periodontitis, missing tooth, and impacted tooth is 92%, 92%, 92%, 93%, 92%, and 91% respectively.

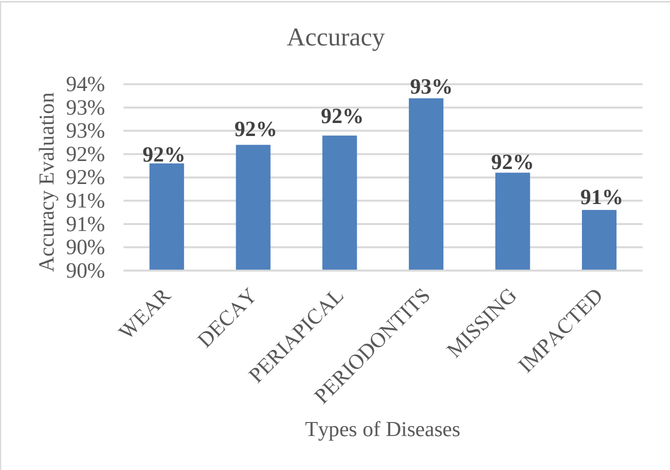


Fig. 15. Comparison of accuracy for different diseases

VII. Conclusion and Future Work

This system is implemented using the idea of transfer learning and different pre-trained networks, such as 'ResNet50V2', 'ResNet101V2', 'MobileNetV3Large', 'MobileNetV3Small', 'MobileNet', 'EfficientNetB0', 'EfficientNetB1', and 'EfficientNetB2', with XGBoost. To get the final prediction, a weighted ensemble module is used. Additionally, the developed technique has achieved better outcomes in terms of accuracy, precision, F-measure, recall, and sensitivity. Thus, it achieved 92-93% of accuracy in classifying multiple dental diseases including tooth wear, periapical, periodontitis, tooth decay, missing tooth, and impacted tooth. This approach is implemented on new diseases and compared the measuring parameters against different diseases. For some measuring parameters, values are less compared which can be improved in future work by using the extension of this dataset. In this present study, the given model is tested by classifying six diseases and their 46 possible combinations. this system can identify one disease present in an image as well as a combination of multiple diseases present in one image. The proposed method performs better in terms of accuracy and different measuring parameters than current state-of-the-art methods and has a variety of uses in computer-assisted multiple tooth disease identification and classification.

Acknowledgment

This work is supported by the following dental clinics for data collection, understanding, annotation, authentication, and validation of images.

a) Government dental college, Nagpur.

b) Dr. Jayaswal's clinic of Dentistry and Geriatric oral health care center, Nagpur, India.

c) Government Dental College, Aurangabad, India

d) Invasion Lab, Nagpur

Data Availability

On request, the corresponding author will provide the data used to generate the study's findings.

Funding Details

This work is being carried out with the assistance of an AICTE financing award (F No B-60/RFID/RPS/Policy-1/2017-18).

References

[1] Gu, Zaiwang, et al. "Ce-net: Context encoder network for 2d medical image segmentation." IEEE transactions on medical imaging 38.10 (2019): 2281-2292.

[2] Shiraisi, Ai, et al. "Improvement in Oral Health Enhances the Recovery of Activities of Daily Living and Dysphagia after Stroke." Journal of Stroke and Cerebrovascular Diseases 30.9 (2021): 105961.

[3] Lakshmi, M. Muthu, and P. Chitra. "Tooth Decay Prediction and Classification from X-Ray Images using Deep CNN." 2020 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2020.

[4] Schwendicke, Falk, et al. "Convolutional neural networks for dental image diagnostics: A scoping review." Journal of dentistry 91 (2019): 103226.

[5] Hashem, Mohamed, and Ahmed E. Youssef. "Teeth infection and fatigue prediction using optimized neural networks and big data analytic tool." Cluster Computing 23.3 (2020): 1669-1682.

[6] Angst, Patrícia Daniela Melchiors, et al. "Oral health status of patients with leukemia: a systematic review with meta-analysis." Archives of Oral Biology (2020): 104948.

[7] Banu, Ancuta, et al. "Dental health between self-perception, clinical evaluation and body image dissatisfaction–a cross-sectional study in mixed dentition pre-pubertal children." BMC Oral Health 18.1 (2018): 1-9.

[8] Liu, Lizheng, et al. "A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal." IEEE journal of biomedical and health informatics 24.3 (2019): 898-906.

[9] Glurich, Ingrid, et al. "Interdisciplinary care model: pneumonia and oral health." Integration of Medical and Dental Care and Patient Data. Springer, Cham, 2019. 123-139.

[10] Abdalla-Aslan, Ragda, et al. "An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography." Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology 130.5 (2020): 593-602.

[11] Grischke, Jasmin, et al. "Dentronics: Towards robotics and artificial intelligence in dentistry." Dental Materials 36.6 (2020): 765-778.

[12] Mahmoud, Nourelhoda M., et al. "Detecting dental problem-related brain disease using intelligent bacterial optimized associative deep neural network." Cluster Computing 23.3 (2020): 1647-1657.

[13] Lee, Jeong-Hoon, et al. "Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks." BMC oral health 20.1 (2020): 1-10.

[14] Sonavane, Apurva, RohitYadav, and AdityaKhamparia. "Dental cavity Classification using Convolutional Neural Network." IOP Conference Series: Materials Science and Engineering. Vol. 1022. No. 1. IOP Publishing, 2021.

[15] Tandon, Divya, JyotikaRajawat, and Monisha Banerjee. "Present and future of artificial intelligence in dentistry." Journal of Oral Biology and Craniofacial Research 10.4 (2020): 391-396.

[16] P. Jaiswal and S. Bhirud, "Study and Analysis of an Approach Towards the Classification of Tooth Wear in Dentistry Using Machine Learning Technique," 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES), 2021, pp. 1-6, DOI: 10.1109/TRIBES52498.2021.9751650.

[17] Nripendra Kumar Singh, Khalid Raza, Progress in deep learning-based dental and maxillofacial image analysis: A systematic review, Expert Systems with Applications, Volume 199,2022,116968, ISSN 0957-4174,

[18] https://doi.org/10.1016/j.eswa.2022.116968. da Silva Rocha, É.; Endo, P.T. A Comparative Study of Deep Learning Models for Dental Segmentation in Panoramic Radiograph. Appl. Sci. 2022, 12, 3103. https://doi.org/10.3390/app12063103

[19] Ibrahim S. Bayrakdar, Kaan Orhan, Özer Çelik, Elif Bilgir, Hande Sağlam, Fatma Akkoca Kaplan, Sinem Atay Görür, Alper Odabaş, Ahmet Faruk Aslan, Ingrid Różyło-Kalinowska, "A U-Net Approach to Apical Lesion Segmentation on Panoramic Radiographs", BioMed Research International, vol. 2022, Article ID 7035367, 7 pages, 2022. https://doi.org/10.1155/2022/7035367

[20] Il-Seok Song, Imaging Science in Dentistry 2022, Deep learning-based apical lesion segmentation from panoramic radiographs.

[21] K. Panetta, R. Rajendran, A. Ramesh, S. P. Rao, and S. Agaian, "Tufts Dental Database: A Multimodal Panoramic X-Ray Dataset for Benchmarking Diagnostic Systems," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 4, pp. 1650-1659, April 2022, DOI: 10.1109/JBHI.2021.3117575.

[22] G. Silva, L. Oliveira, M. Pithon, “Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives”, Expert Systems with Applications, 107, pp. 15-31, 2018.

[23] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, L. Oliveira, “Deep instance segmentation of teeth in panoramic X-ray images” 31 st SIBGRAPI Conference on Graphics, Patterns and Images, October 2018.

[24] J. Kim, H. S. Lee, I. S. Song & K. H. Jung, “DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs”, in Scientific reports, vol.9(1), pp. 1-9, Nov. 2019.

[25] K. Zhang, J. Wu, H. Chen, and P. Lyu, "An effective teeth recognition method using label tree with cascade network structure”, Computerized Medical Imaging and Graphics, vol. 68, pp. 61-70, 2018.

[26] D. Abdelhafiz, C. Yang, R. Ammar, S. Nabavi, “Deep convolutional neural networks for mammography: advances, challenges, and applications”, BMC Bioinformatics, 20(S11), 2019.

[27] [8] M.A. Kadampur, S. Al Riyaee, “Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images”, Informatics in Medicine Unlocked, 18, 100282, 2020

[28] D.V. Tuzoff, L.N. Tuzova, M.M. Bornstein, A.S. Krasnov, M.A. Kharchenko, S.I. Nikolenko, M.M. Sveshnikov, G.B. Bednenko, “Tooth detection and numbering in panoramic radiographs using convolutional neural networks”, Dentomaxillofacial Radiol, 48, 20180051, 2019.

[29] Saleh A. Albelwi, “Deep Architecture based on DenseNet-121 Model for Weather Image Recognition” International Journal of Advanced Computer Science and Applications(IJACSA), 13(10), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0131065

[30] Krois, J., Cantu, A. G., Chaurasia, A., Patil, R., Chaudhari, P. K., Gaudin, R., Gehrung, S., & Schwendicke, F. (2021). Generalizability of deep learning models for dental image analysis. Scientific Reports, 11. https://doi.org/10.1038/s41598-021-85454-5

[31] Xin Yu Liew, Nazia Hameed, Jeremie Clos, An investigation of XGBoost-based algorithm for breast cancer classification, Machine Learning with Applications, Volume 6,2021,100154, ISSN 2666-8270,https://doi.org/10.1016/j.mlwa.2021.100154.

[32] Almalki, Y.E.; Din, A.I.; Ramzan, M.; Irfan, M.; Aamir, K.M.; Almalki, A.; Alotaibi, S.; Alaglan, G.; Alshamrani, H.A.; Rahman, S. Deep Learning Models for Classification of Dental Diseases Using Orthopantomography X-ray OPG Images. Sensors 2022, 22, 7370. https://doi.org/10.3390/s22197370

[33] Guo, Y., Ge, Y., Yang, C., Al-Garadi, M. A., & Sarker, A. (2022). Comparison of Pretraining Models and Strategies for Health-Related Social Media Text Classification. Healthcare, 10(8). https://doi.org/10.3390/healthcare10081478

[34] Raeho Lee and Haewon Byeon, “Application of Stacking Ensemble Machine in Big Data: Analyze the Determinants for Vitalization of the Multicultural Support Center” International Journal of Advanced Computer Science and Applications(IJACSA), 13(10), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0131007

[35] Munindra Lunagaria, Vijay Katkar, and Krunal Vaghela, “Covid-19 and Pneumonia Infection Detection from Chest X-Ray Images using U-Net, EfficientNetB1, XGBoost and Recursive Feature Elimination” International Journal of Advanced Computer Science and Applications(IJACSA), 13(9), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130954

[36] A. B. Oktay, “Tooth detection with convolutional neural networks,” in Proc. IEEE Med. Technol. Nat. Congr., 2017, pp. 1–4.

[37] C. Muramatsu et al., “Tooth detection and classification on panoramic radiographs for automatic dental chart filing: Improved classification by multi-sized input data,” Oral Radiol., vol. 37, no. 1, pp. 13–19, 2021.

[38] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, “Deep instance segmentation of teeth in panoramic X-ray images,” in Proc. IEEE 31st SIBGRAPI Conf. Graph., Patterns Images, 2018, pp. 400–407.

[39] D. V. Tuzoff et al., “Tooth detection and numbering in panoramic radiographs using convolutional neural networks,” Dentomaxillofacial Radiol., vol. 48, no. 4, 2019, Art. no. 20180051

[40] T. Ekert et al., “Deep learning for the radiographic detection of apical lesions,” J. Endodontics, vol. 45, no. 7, pp. 917–922.e5, 2019

[41] M. Fukuda et al., “Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography,” Oral Radiol., pp. 1–7, 2019

[42] K. Panetta, R. Rajendran, A. Ramesh, S. P. Rao, and S. Agaian, "Tufts Dental Database: A Multimodal Panoramic X-Ray Dataset for Benchmarking Diagnostic Systems," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 4, pp. 1650-1659, April 2022, DOI: 10.1109/JBHI.2021.3117575.

[43] M. P. Muresan, A. R. Barbara, and S. Nedevschi, "Teeth Detection and Dental Problem Classification in Panoramic X-Ray Images using Deep Learning and Image Processing Techniques," 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), 2020, pp. 457-463, DOI: 10.1109/ICCP51029.2020.9266244.

Dilated Multi-Activation Autoencoder to Improve the Performance of Sound Separation Mechanisms

Ghada Dahy[1] , Mohammed A.A.Refaey[2], Reda Alkhoribi[3], M.Shoman[4]
Department of Information Technology, Faculty of Computers and Artificial Intelligence, Cairo University,
Cairo, Egypt

***Abstract*—Speech enhancement is the process of improving the quality of audio relative to target speaker while suppressing other sounds. It can be used in many applications as speech recognition, mobile phone, hearing aids and also enhancing audio files resulted from separation models. In this paper, a convolutional neural network (CNN) architecture is proposed to improve the quality of target's speaker resulted from speech separation models without having any prior information about the background sounds. The proposed model consists of three main phases: Pre-Processing phase, Autoencoder phase and Retrieving Audio phase. The pre-processing phase converts audio to short time Fourier transform (STFT) domain. Autoencoder phase consists of two main modules: dilated multi-Activation encoder and dilated multi-Activation decoder. Dilated multi-Activation encoder module has a six blocks with different dilation factors and each block consists of three CNN layers where each layer has different activation function then the encoder's blocks are arranged in reverse order to construct dilated multi-activation decoder. Audio retrieving phase is used to reconstruct audio depending on feature resulted from second phase. Audio files resulted from separation models are used to build our datasets that consist of 31250 files. The proposed dilated multi-activation autoencoder improved separated audios Segmental Signal-to-Noise Ratio (SNRseg) with 33.9%, Short-time objective intelligibility (STOI) with 1.3% and reduced bark spectral distortion (BSD) with 97%.**

Keywords—Speech de-noising; speech enhancement; speech separation; short time Fourier transform (STFT); autoencoder; dilated Convolution neural network; multi-activation functions, convolution neural network (CNN); bidirectional long short memory (BLSTM)

I. Introduction

The field of machine learning proposed different architectures in complex and difficult problems that were unattainable in the field of speech processing. Speech separation and enhancement are considered two of the most important problems in signal processing where machine learning field achieves better performance on dealing with them. Speech enhancement is used as a step of removing or reducing background noise to enhance the main signal of the target speaker. In the current days, researchers do a lot of effort to have a very high quality speech that relative to the target speaker.

In the field of speech analysis, speech enhancement [1] focus on having high quality sound of the target speaker by suppressing other sounds except sound of the target one [2] while speech separation [3] is considered the process of separating mixed sounds.

The main objective of speech enhancement and separation is extracting signals of interest parts from an audio according to critical rules that can be done by post production software that depends currently on deep learning field. Speech enhancement is considered the main stone of many applications as teleconferencing systems, speech recognition [4,5], hearing aids [6-8], speaker recognition [9,10] and also can be used to enhance the resulted signals from speech separation systems.

Speech understanding in noisy environments is still one of the major challenges as we need very high quality audio files with large datasets and different types of noises to train the enhancement models and this will need very high quality resources. To face the previous challenge, we built a dataset with 31250 audio files resulted from separation model that trained on a subset samples from The Oxford-BBC LRS2 Dataset.

Recently, deep learning made a great improvement in the speech enhancement field by using different structures over signal processing methods that based on supervised enhancement techniques.

The main contribution of the proposed model is implementing speech enhancement structures to improve the speech quality of the target's speaker resulted from separation models by analyzing input speech in multiple levels. The proposed model plays an important role in the digital speech signal processing according to the type of degradation and background sounds in the speech signal of the target speaker.

In this paper, we proposed a new approach for speech enhancement which achieved an interesting performance comparing against old methods. It improves the quality of speech resulted from separation model. It consists of three main phases: Pre-Processing phase, AutoEncoder phase and Retrieving Audio phase. It considers low quality audios resulted from separation models as an input and complex ratio mask relative to cleared audio as output. The dilation convolutional neural network is considered as the main stone of our proposed model as it has ability to expand the covered areas of the input audio without needing to apply pooling that maybe cause missing some information. Dilated convolutional network mainly uses dilation factor to control the size of the covered area. It also could analysis input signal to suppress background sounds without increasing kernel parameters.

The rest of this paper is organized in the following order. Section II introduces the related work; Section III introduces the proposed structures and defines the main mechanism of speech enhancement in details. Section IV describes the dataset, network parameters and demonstrates the performance of our network. Finally, we conclude the idea of the proposed model in Section V.

II. Related Work

Neural Networks is considered one of the most common machine learning algorithms. It has been improved over time and also outperforms other algorithms and methods in speed and accuracy. There are different types of neural network as convolutional neural network (CNN) [11.12], RNN (Recurrent Neural Networks) [13], Autoencoders [14], LSTM (Long short Tern Memory) [15,16] and also WaveNet that depends on dilated convolution network [17].

Emad el al. [11] proposed multi-resolution convolution neural network that support them in separating audio files from mixed sounds. They proved that concatenating sets of convolution network with different filters could improve the performance of separation compared with deep neural network (DNN) and fully convolution neural network.

Jen-Cheng el al.[12] proposed audio visual model for speech enhancement using CNN that complete its process by analyzing visual and audio stream which processed using separated CNN models then fed into AUDIO-VISUAL Autoencoder model. Their model could outperform speech enhancement models that use audio-only and existing audio-visual enhancement models.

Jinuk el al. [13] proposed model to separate audio sounds of two speakers. They used CNN followed by LSTM and recurrent neural network (RNN) then a fully connected layer is applied on the concatenation of forward LSTM-RNN and backward LSTM-RNN. They achieved better signal-distortion-ratio (SDR) and signal interference ratio (SIR) compared with deep clustering method that used in sound separation.

Yi Luo and Nima Mesgarani [15] proposed time-domain audio separation model to obtain target speaker sound with acceptable performance. Their proposed autoencoder is able to outperform LSTM that analyze log power magnitude spectrogram to complete the process of separation.

Jitong Chena and DeLiang Wang [16] proposed speech separation model based on LSTM that suitable to temporal dynamics of speech. They used Short-time objective intelligibility (STOI) to measure the performance of their model which outperformed DNN model as LSTM could capture long term context of input speech and has low latency in completing speech separation process.

Aaron et al. [17] proposed dilated CNN model called wavenet that captures the main characteristics of multiple speakers. their model has been trained on thousands of samples. They trained wavenet on different datasets as text to speech and music datasets. It could generate musical fragment that very near to realistic and has very high quality. In WaveNet, a set of dilated convolution layers are stacked together to generate a very large receptive fields with a few layers. Recently, wavenets are used in time domain. It can be used in songs separation from background noise and enhancing speech quality.

Dario et al. [18] proposed speech de-noising technique that able to improve speech signals quality with background-noise. Their proposed model depends on dilated convolution. Their proposed model used different dilation factor in each layer changes in the following order 1, 2, ..., 256, 512. Their pattern is repeated three times. Their model has capability to de-noise an entire audio file in one-shot. They generated their datasets from two sources: speech data from the Voice Bank corpus and environmental sounds from Diverse Environments Multichannel Acoustic Noise Database (DEMAND). Their Perceptual tests proved that their model's outperform the Wiener filtering results. They proved that their system able to learn multi-scale hierarchical representations from raw audio instead of magnitude spectrograms.

Yi Luo and Nima Mesgarani [19] proposed Time-domain Audio Separation Network (TasNet). Mainly, their network structure used in speech separation and also can be used in speech enhancement. It consists of four processes: a preprocessing normalization, an encoder to retrieve the mixture weight, a separation and a decoder to construct waveform. The first process in their network is used to predict the weights that relative to each source in the mixture weight. The weight detection is defined as masks that predict the contribution of each speaker in the mixture weight as T-F masks that are used in short time Fourier transform (STFT) models. In the Separation Network, a deep LSTM network is used to estimate the source masks. They used WSJ0-2mix dataset to evaluate their system on two-speaker speech separation problem. The mixtures relative to their dataset are generated randomly speakers from Wall Street Journal (WSJ0) dataset. Their proposed TasNet system could outperforms systems that use a T-F representation as an input to complete the process of learning. Comparing with STFT system, their experiments proved that the TasNet system was six times faster and could achieve very high quality speech separation performance and also it can be used in speech enhancement.

Craig Macartney and Tillman Weyde [20] proposed speech enhancement model that used Wave-U-Net architecture for speech enhancement. Their experiments proved that the Wave-U-Net could improve several metrics as Perceptual Evaluation of Speech Quality (PESQ), the rating of background distortion (CBAK), the rating of speech distortion (CSIG), segmental signal-to-noise ratio (SSNR) and the predicting rating overall quality (COVL) over the state-of-the-art. They showed that reducing number of hidden layers is more suitable for speech enhancement.

Yong Xu et al. [21] proposed a speech enhancement regression model depending on deep neural networks (DNNs) that consists of multiple-dense layers. Their proposed model could achieve better performance in case of different objective measures. They held a subjective evaluation measure to check the quality of DNN model with 10 listeners, about 76.35% of the listener prefer using DNN results to complete the process of enhancement.

Andreas Jansson et al. [22] proposed a novel architecture of the U-Net architecture that used to complete the source separation task. They benefit from skip connections paths in building encoder-decoders of u-net. They held quantitative evaluation and subjective assessment to evaluate the performance of their model. Their experiments proved that their proposed architecture able to achieves state-of-the-art performance.

III. Proposed Model

In this paper, we proposed a model based on dilation convolution neural network that able to analysis and discover local pattern of the target speaker in STFT. The main objective of the proposed model is suppressing other human's sounds except target speakers. It mainly enhances the quality of speech resulted from separation models [23].

Fig. 1 depicts the structure of the proposed model. It consists of three main phases: Pre-Processing phase, Autoencoder phase and Retrieving Audio phase. It takes low quality audios resulted from separation models as an input and scaled complex ratio mask of the cleared audios with same size of the input using (1,2) as output.

$$Mask_{(t,f)} = \frac{Target\ sound_{(t,f)}}{Mixed\ Sound_{(t,f)}} \tag{1}$$

$$Scaled\ Mask_{(t,f)} = 10\frac{1-e^{-0.1*Mask}}{1+e^{-0.1*Mask}} \tag{2}$$

where, t and f are considered as indexes of time and frequency.

In the Pre-Processing phase, audio files are converted to STFT to generate 304X256X2 real and imaginary numbers that taken as an input to the Autoencoder phase and saved as numbers to make the process of training faster.

Autoencoder phase consists of two modules, Dilated MultiActivation Encoder and Dilated MultiActivation decoder as in Fig. 2. Dilated MultiActivation Encoder module consists of six blocks with different dilation factors and filters to analysis and track relation between audio patterns. Encoder module able to cover more information from patterns depending on dilation factor without pooling which reduces the size of the matrix that are taken as input to the next convolution layer, this will cause missing some audio features and maybe has bad effect on the model performance. Each block consists of three layers, first one is CNN with dilation factor that different in each block to expand the covered area of the input features, second one takes the resulted features from the first layer as an input then apply CNN with the same dilation factor, number of filters of the first layer and uses activation function called activation 1 and finally, the third layer takes the resulted features of first layer as input then applies CNN with activation function named activation 2 where the filter size in second and third layers is s3*S4 as in Fig. 3. Each block has different activations, RELU and element wise multiplication of resulted data from nonlinear output of activation1 and activation2 as experiments results [18] proved that the non-linearity activation work better than RELU activation in dealing with audio signals. Finally, the element wise multiplication of the second and third layers is taken as input to the next block. In the Dilated MultiActivation decoder, the six blocks of the encoder are reversed to build the decoder module. There are different number of filters in each block with kernel size s1*s2 starting with reduction factor changes according to this range [N1, N2, N3, N4, N5, N6] and dilation factor changes according to the following range [D1, D2, D3, D4, D5, D6]. We used dilated convolutional layer in our proposed model as it could expand the covered area in the input audio without pooling. We controlled the covered area in the input by changing the value of the dilation factor. Dilated autoencoder support us in dealing with speech and discover relation between background sounds and target sounds without needing to increase kernel parameters or reducing size of the input features.

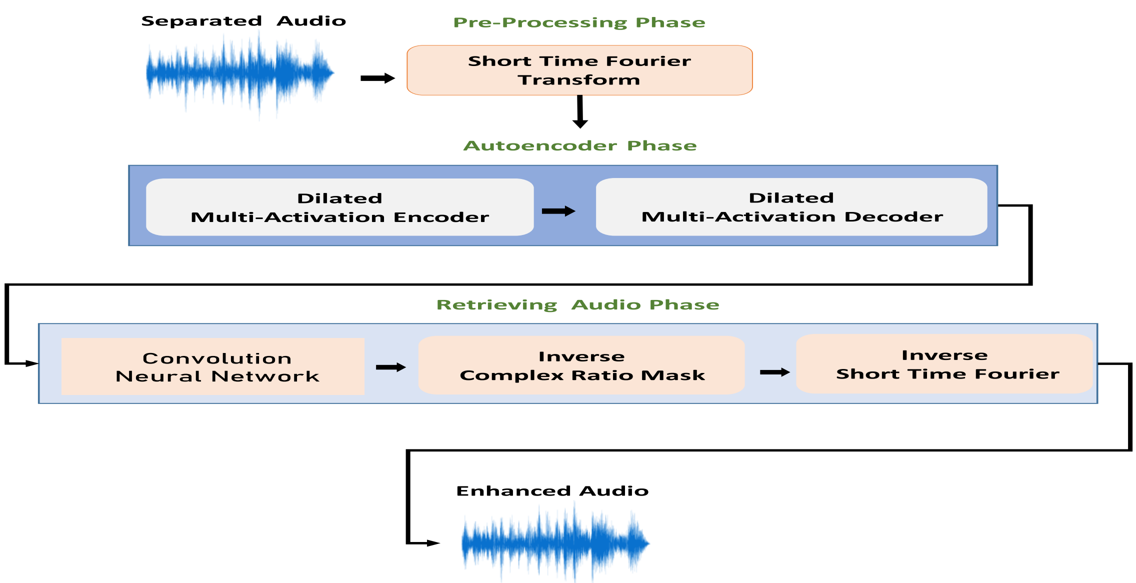


Fig. 1. The proposed model of speech enhancement

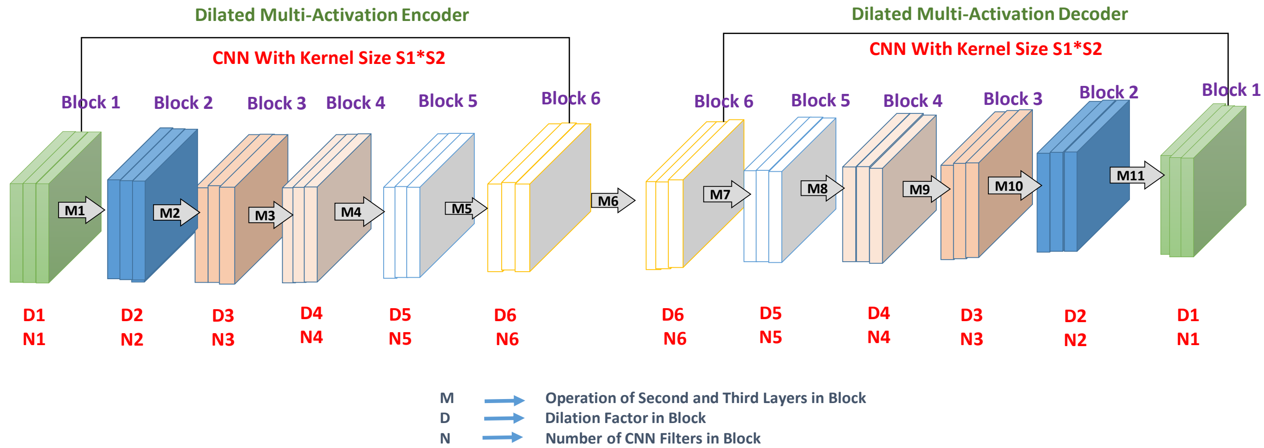


Fig. 2. Dilated multi-activation autoencoder

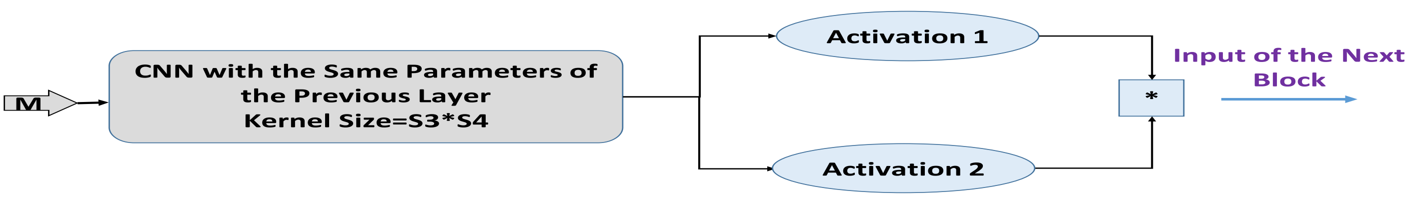


Fig. 3. Second and third layers of autoencoder's blocks

In the retrieving audio phase, CNN layer is applied on the resulted features from decoder, then inverse complex ratio mask is calculated on the resulted features from CNN [13] after rescaling it by using (3) and finally inverse short time Fourier transform (ISTFT) is used to retrieve the enhanced version of separated audios.

$$S(Resulted\ Mask) = \frac{1}{0.1}\log\left(\frac{10-Resulted\ Mask}{10+Resulted\ Mask}\right) \quad (3)$$

where, S is the scaled complex ratio mask.

IV. Experiments and Results

We introduce a small-scale audio dataset containing speech audio with no interfering background signals taking from separated audio of our previous work [23]. The audio segments have length between 3 and 10 seconds The enhanced audio dataset which is used to train our network contains 31250 audio files resulted from separation process [23] that are splitted into 28126 for training and 3124 for testing. Fig. 4 shows Mixed signal of two speakers in time domain and corresponding spectrogram and amplitude spectrum. It also shows magnitude spectrum and spectrogram of speaker 1's sound and speaker 2's sound before mixing them

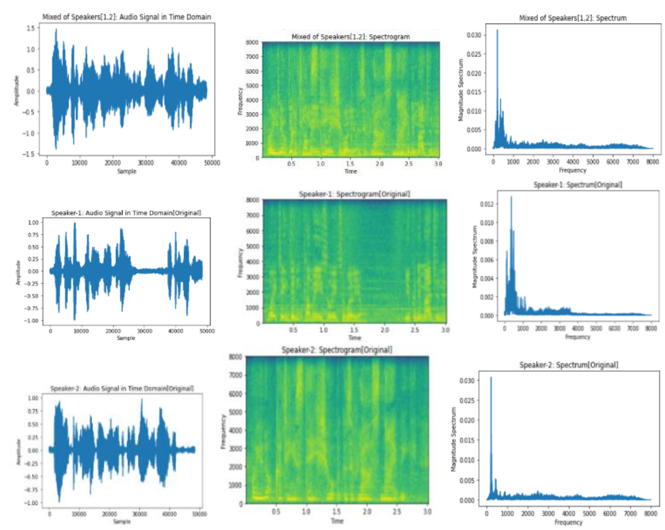


Fig. 4. Original signal in time domain, spectrogram and magnitude spectrum [mixed signal, speaker 1 and speaker2]

We implemented Dilated MultiActivation AutoEncoder in keras. In case of training, we use a batch size of 2 samples and Adam optimizer for 14062 batches with a learning rate $1X10^{-5}$. Tables I and II indicate Hyper-parameters that used to complete the process of training. The nonlinearity activation function is a combination of Tanh and Sigmoid. The number of filters in the retrieving phase is two. Number of filters in the autoencider's blocks starting from 256 to 4 with decreasing factor equal half and dilation factor starting from 64 to 1 with decreasing factor half. We used STFT as a pre-processioning phase on the input audio where STFT is calculated using a Hann window with length 25ms, hop length of 10ms, and FFT size of 512.

TABLE I. TRAINING HYPER-PARAMETERS OF DILATED MULTIACTIVATIO AUTOENCODER MODEL

	Value
Learning Rate	$1X10^{-5}$
Optimizer	Adam
S1*S2	15*15
S3*S4	3*3
Activation 1	Tanh
Activation 2	Sigmoid
CNN Filters of Retrieving audio phase	Number of filters[2]

TABLE II. TRAINING HYPER-PARAMETERS OF MULTI-ACTIVATION ENCODER AND DECODER BLOCKS

	Number of Filters	**Dilation Factor**
Block 1	N1 =256	D1=64
Block 2	N2 =128	D2=32
Block 3	N3 =64	D3=16
Block 4	N4 =32	D4=4
Block 5	N5 =8	D5=2
Block 6	N6 =4	D6 =1

The proposed model takes STFT of separated audios resulted from separation model [23] as input and its output is enhanced version of the target speaker by removing background sounds of other speakers. Fig. 5 shows spectrograms and spectrums of two speaker resulted from separation model [23] before enhancement process where the separation models take mixed signal as input and generated separated sound relative to each speaker by using facial embedding of each speaker as a guide to complete the separation.

To test the performance of our model, we held four experiments, three from literature papers and the last one depending on our proposed model. All experiments use the same dataset resulted from the audio separation models.

In the first experiment, advanced version from DNN [21] is trained to reduce background sounds. It consists of a pipeline of dense layers [24]. The pipeline consists of three dense layers followed by dropout and the previous architecture is repeated two times. Fig. 6 indicates the magnitude spectrum and spectrogram of the enhanced audio after taking the low quality speeches resulted from separation models as input.

In the second experiment, we train u-net [22] to suppress background sounds. Firstly, three convolution layers with filters [128,128,1] is applied on the STFT values to prepare U-net input [25]. It is considered as deep convolutional autoencoder with skip connection between encoder and decoder parts. They adopted their network to de-noise voices and we used the same architecture to enhance audio signal resulted from separation model. After training, u-net could have better spectrogram and magnitude spectrum comparing with results of separation model as in Fig. 7. In the third experiment, we trained Wavenet [18] architecture to enhance speech of the target speaker. Fig. 8 shows the enhanced magnitude spectrum and spectrogram of the target speakers after applying wavenet architecture on separated audio it has better results comparing with DNN as it depends mainly on dilated convolutional neural network. In the last experiment, our proposed multi-activation AuteEncoder is trained to complete the process of enhancement, Fig. 9 shows the enhanced magnitude spectrum and spectrogram resulted from our proposed model that is very near to original spectrogram and magnitude spectrum.

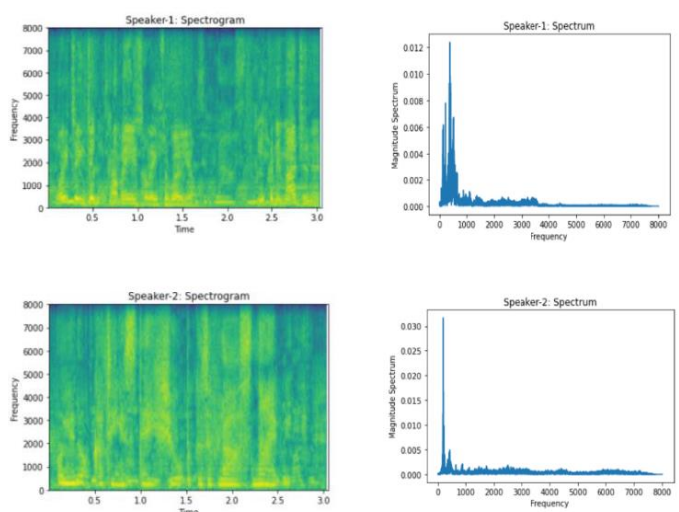


Fig. 5. Speech separation of two speakers [spectrogram and spectrum]

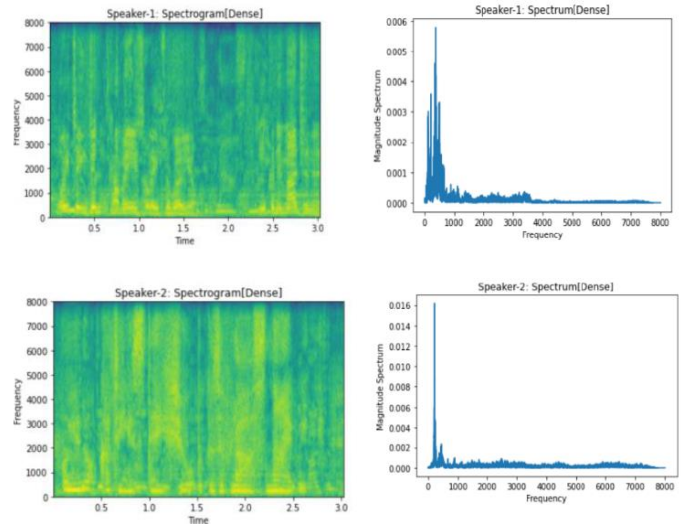


Fig. 6. Speech enhancement using DNN [spectrogram and spectrum]

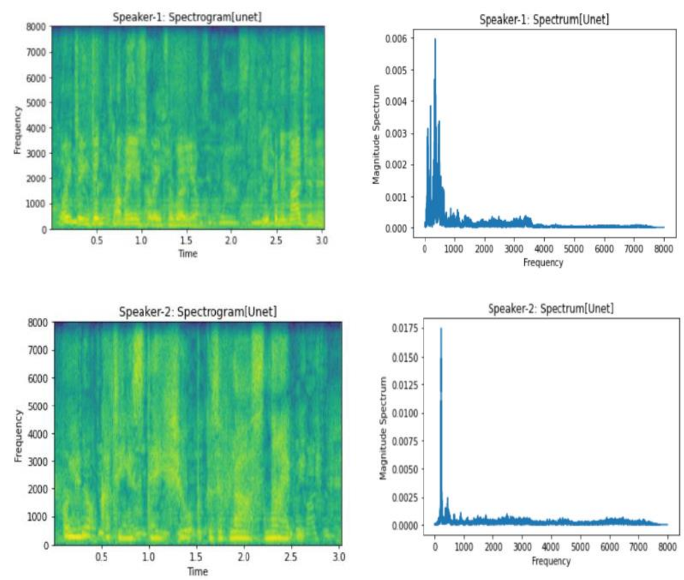


Fig. 7. Speech enhancement using U-net [spectrogram and spectrum]

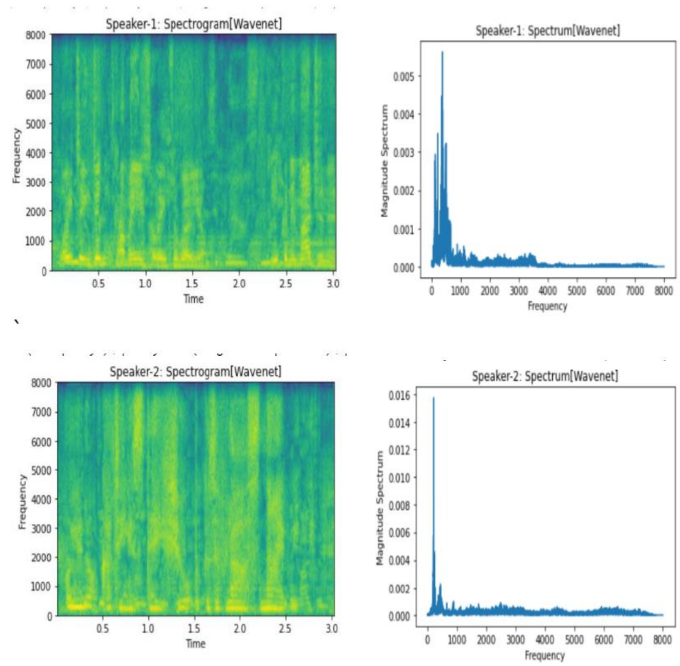


Fig. 8. Speech enhancement using wavenet [spectrogram and spectrum]

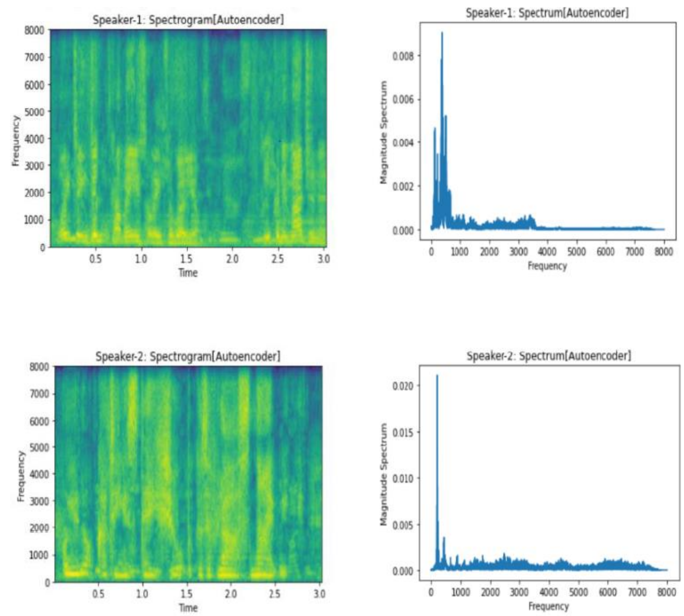


Fig. 9. Speech enhancement using autoencoder [spectrogram and spectrum]

Fig. 5-9 show that the proposed model has best spectrum and spectrogram compared with all experiments that is very near to the original version as it is mainly depends on dilated convolution with different dilation factor in each block to analysis input signal in different level and it does not use pooling in its architecture so there is no missing information during the training process

Fig. 10 illustrates the average testing loss during the process of learning. It is clear that dilated multi-activation Autoencoder experiment has best testing loss comparing with DNN, U-NET and Wavenet.

DNN has loss equal to 0.0991 after the second epoch, U-net [experiment 2]. has loss equal to 0.099 that is very near to DNN Wavenet [experiment 3] has loss equal to 0.095 that less than DNN and U-NET. Dilated multi-activation autoencoder has the minimum loss compared with DNN, U-NET and Wavenet that equal to 0.89.

The number of trainable parameters relative to each experiments is summarized in Fig. 11. It is cleared that the dilated autoencoder has an acceptable number of parameters that very near to Wavenet, u-net and less than DNN. Although DNN consists of high number of dense layers and has very high parameters, it has low performance comparing with our proposed model. Wavenet has the minimum number of parameters but its performance is lower than our proposed model.

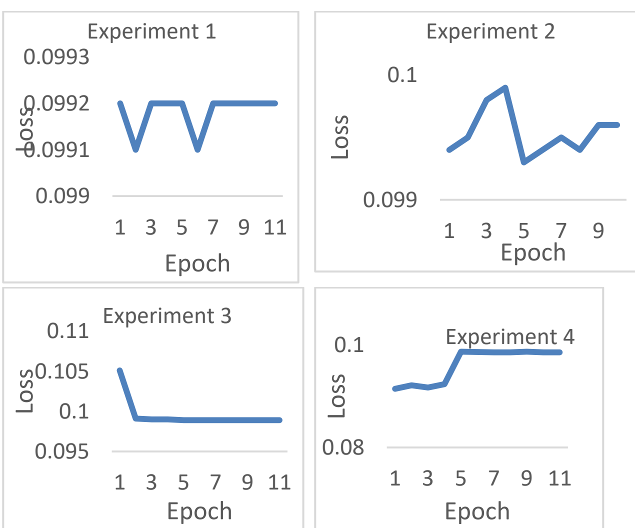


Fig. 10. Average testing loss of enhancement models

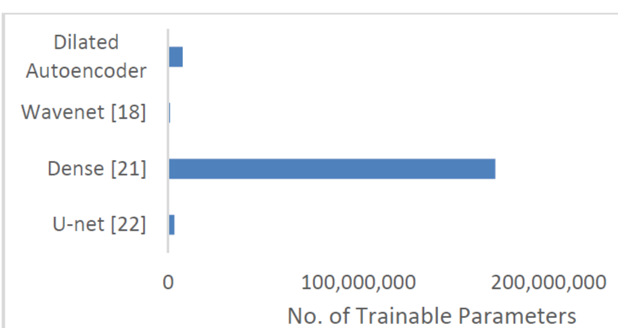


Fig. 11. Trainable parameters of enhancement models

To evaluate objective performance of the proposed model, supporter Python Speech Enhancement Performance Measures (Quality and Intelligibility) project [26] is used. Four speech objective evaluation metrics are calculated to measure speech quality [Segmental Sign al-to-Noise Ratio (SNRseg), PESQ, bark spectral distortion (BSD) and STOI].

Table III and Fig. 12 and 13 indicated that the proposed autoencoder outperform most of enhancement models in SNRseg, STOI and BSD.

It improved SNRseg of separation audio with 33.9%, STOI with 1.3% and reduced BSD with 97%, it is cleared that our proposed model has the best SNRseg, STOI and BSD comparing with audio resulted from separation and other enhancing models. Table III proved that audio files resulted from separation models has high noise and also high distortion comparing with enhanced version resulted from our proposed model that trained on 31250 audio files divided into 28126 for training and 3124 for testing.

TABLE III. OBJECTIVE EVALUATION MEASURE OF SPEECH ENHANCEMENT MODELS

Model/Measure	SNRseg	Stoi	Pesq	BSD
Separation model[23]	3.63	0.80	2.17	4,329.90
U-net[22]	2.87	0.80	2.17	218.90
DNN [21]	2.81	0.80	2.13	231.69
WaveNet[18]	2.71	0.80	2.15	212.62
Dilated AutoEncoder	4.86	0.81	1.95	128.89

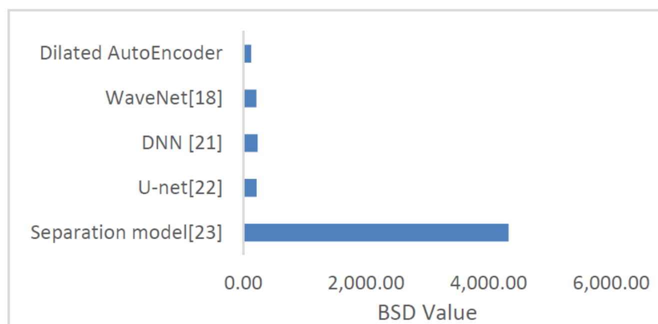


Fig. 12. Bark spectral distortion

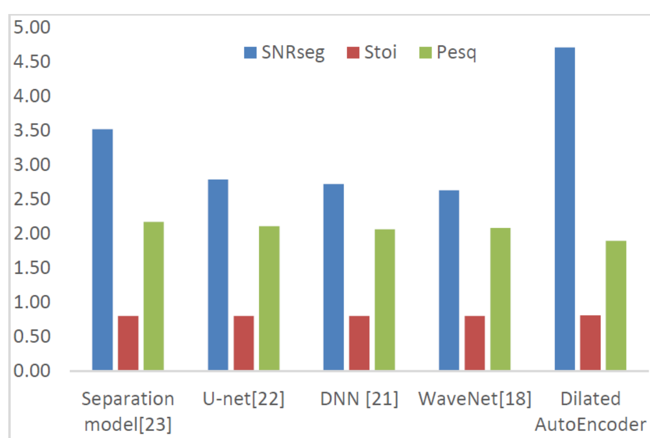


Fig. 13. Objective evaluation measures [SNRseg, Stoi, Pesq]

V. Conclusion

We proposed dilated multi-activation autoencoder to enhance the performance of audio sounds resulted from separation models. It consists of two main modules, dilated multi-Activation encoder and dilated multi-Activation decoder where dilated multi-Activation encoder module has six convolutional neural network blocks with activation functions and different dilation factors. To build the structure of the decoder, the six blocks of the encoder are arranged in reverse order. Dataset consist of 31250 files splitted into training and testing sets where training set consists of 28126 files and testing set consists of 3124. The proposed model improved SNRseg of separated audios with 33.9%, STOI with 1.3% and reduced BSD with 97%. In the future, we will try to improve the performance of the proposed model by adding face embedding features relative to target speaker.

References

[1] Amol Chaudhari; S. B. Dhonde,'A review on speech enhancement technique', Proceedings in 2015 International Conference on Pervasive Computing (ICPC), Pune, India, 8-11 January 2015

[2] Serim Park and Jin Won Lee,' A Fully Convolutional Neural Network for Speech Enhancement', Proceedings in Interspeech, Dublin, Ireland August 2017.

[3] DeLiang Wang and Jitong Chen.' Supervised Speech Separation Based on Deep Learning: An Overview,IEEE/ACM Transactions on Audio, Speech, and Language Processing , Volume: 26, Issue: 10, October 2018.

[4] J. Li, L. Deng, R. Haeb-Umbach, and Y. Gong, Robust Automatic Speech Recognition: A Bridge to Practical Applications, 1st ed. Academic Press, 2015.

[5] B. Li, Y. Tsao, and K. C. Sim, "An investigation of spectral restoration algorithms for deep neural networks based noise robust speech recognition," in Proc. INTERSPEECH, 2013, pp. 3002–3006

[6] T. Venema, Compression for Clinicians, 2nd ed. Thomson Delmar Learning, 2006, chapter. 7.

[7] H. Levitt, "Noise reduction in hearing aids: an overview," J. Rehab. Res. Dev., 2001, vol. 38, no. 1, pp.111–121.

[8] A. Chern, Y. H. Lai, Y.-P. Chang, Yu Tsao, R. Y. Chang, and H.-W. Chang, "A smartphone-based multi-functional hearing assistive system to facilitate speech recognition in the classroom," IEEE Access, 2017.

[9] A. El-Solh, A. Cuhadar, and R. A. Goubran. "Evaluation of speech enhancement techniques for speaker identification in noisy environments," in Proc. ISMW, 2007, pp. 235–239.

[10] J. Ortega-Garcia and J. Gonzalez-Rodriguez, "Overview of speech enhancement techniques for automatic speaker recognition," in Proc. ICSLP, vol. 2, 1996, pp. 929–932.

[11] Emad M. Grais, Hagen Wierstorf, Dominic Ward, and Mark D. Plumbley, 'MULTI-RESOLUTION FULLY CONVOLUTIONAL NEURAL NETWORKS FOR MONAURAL AUDIO SOURCE SEPARATION', Proceedings in 13th International conference on Latent Variable Analysis and Independent Component Analysis, Grenoble, France, 21-23 February 2017

[12] Jen-Cheng Hou, Syu-Siang Wang, Ying-Hui Lai, Yu Tsao 'Audio-Visual Speech Enhancement Using Multimodal Deep Convolutional Neural Networks, IEEE Transactions on Emerging Topics in Computational Intelligence, Volume 2, Number 1, February 2018

[13] Jinuk Park, Jaeseok Kim, Heejin Choi, Minsoo Hahn,'Convolutional Recurrent Neural Network Based Deep Clustering for 2-speaker Separation', Proceeding in the 2018 2nd International Conference on Mechatronics Systems and Control Engineering, Amsterdam, Netherlands, February 2018

[14] Emad M. Grais and Mark D. Plumbley,' SINGLE CHANNEL AUDIO SOURCE SEPARATION USING CONVOLUTIONAL DENOISING, Proceedings in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada, 14-16 November 2017

[15] Yi Luo and Nima Mesgarani,' Real-time Single-channel Dereverberation and Separation with Time-domain Audio Separation Network', Proceeding in Interspeech, , Hyderabad ,India, 2-6 September 2018.

[16] Jitong Che, and DeLiang Wang,' Long Short-Term Memory for Speaker Generalization in Supervised Speech Separation', The Journal of the **Acoustical Society of America Volume 41, Issue 4705, 2017**

[17] A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu,' WaveNet: A Generative Model for Raw Audio', arXiv:1609.03499v2 [cs.SD,] 19 Sep 2016

[18] Dario Rethage, Jordi Pons, and Xavier Serra, A WAVENET FOR SPEECH DENOISING, Proceedings in IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, AB, Canada, April 15-20, 2018.

[19] Yi Luo and Nima Mesgarani,' Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation', IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 27 Issue 8, Page 1256-1266 August 2019

[20] Craig Macartney and Tillman Weyde,' Improved Speech Enhancement with the Wave-U-Net', arXiv:1811.11307v1 [cs.SD] 27 Nov 2018

[21] Yong Xu, Jun Du, Li-Rong Dai, and Chin-Hui Lee, Fellow,' An Experimental Study on Speech Enhancement Based on Deep Neural Networks', IEEE SIGNAL PROCESSING LETTERS, VOL. 21, NO. 1, JANUARY 2014

[22] Andreas Jansson, Eric Humphrey, Nicola Montecchio , Rachel Bittner , Aparna Kumar ,and Tillman Weyde,' SINGING VOICE SEPARATION WITH DEEP U-NET CONVOLUTIONAL NETWORKS', Proceedings in the 18th ISMIR Conference, Suzhou, China, October 23-27, 2017

[23] Ghada Dahy, Mohammed A.A. Refaey, Reda Alkhoribi, M. Shoman.' A speech separation system in video sequence using dilated inception network and U-Net, Egyptian Informatics Journal, Vol 23, Issue 4, December 2022.

[24] https://github.com/boozyguo/ClearWave last accessed: 5/10/2022

[25] https://github.com/vbelz/Speech-enhancement last accessed:10/9/2022

[26] https://github.com/schmiph2/pysepm last accessed: 1/10/2022

A Machine Learning Ensemble Classifier for Prediction of Brain Strokes

Samaa A.Mostafa[1], Doaa S.Elzanfaly[2], Ahmed E.Yakoub[3]

Department of Information Systems-Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt[1, 3]
Department of Information Systems-Faculty of Informatics and Computer Science, The British University in Egypt[2]
Computers and Artificial Intelligence, Helwan University, Cairo, Egypt[2]

***Abstract*—Brain Strokes are considered one of the deadliest brain diseases due to their sudden occurrence, so predicting their occurrence and treating the factors may reduce their risk. This paper aimed to propose a brain stroke prediction model using machine learning classifiers and a stacking ensemble classifier. The smote technique was employed for data balancing, and the standardization technique was for data scaling. The classifiers' best parameters were chosen using the hyperparameter tuning technique. The proposed stacking prediction model was created by combining Random Forest (RF), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), and Naive Bayes (NB) as base classifiers, and meta learner was chosen to be Random Forest. The performance of the proposed stacking model has been evaluated using Accuracy, Precision, Recall, and F1 score. In addition, the Matthews Correlation Coefficient (MCC) has been also used for more reliable evaluation when having an unbalanced dataset, which is the case in most medical datasets. The results demonstrate that the proposed stacking model outperforms the standalone classifiers by achieving an accuracy of 97% and an MCC value of 94%.**

Keywords*—*Stroke disease; prediction model; ensemble methods; stacking classifier

I. Introduction

Stroke is considered one of the riskiest and deadliest diseases affecting humans, as it suddenly strikes the brain. This occurs when the blood flow to the brain is interrupted. Consequently, the brain's ability to receive oxygen and nutrients is compromised, which results in brain cell death within minutes [1]. It is the second leading cause of death globally after ischemic heart disease, as reported by the World Health Organization (WHO) [2]. Ischemic and hemorrhagic strokes are the two primary types. Ischemic stroke happens when a blockage decreases or interrupts blood flow to brain cells, killing the cells within minutes and leading to death. In contrast, Hemorrhagic stroke occurs when weak blood vessels are severely damaged as a result of hypertension, high cholesterol, and other risk factors [3]. Strokes are caused by a variety of risk factors, including medical factors such as high blood pressure, heart disease, diabetes, high cholesterol, and atrial fibrillation, as well as bad habit factors such as smoking, obesity, unhealthy foods, and lack of physical inactivity. The word "FAST" can be used to recognize the main stroke symptoms [4]. The abbreviated FAST stands for four words. F stands for facial laughter perception that they cannot smile or that their mouth or eyes have closed. A refers to the individual with a stroke who might not be capable of raising both arms and maintaining them up. S stands for speech that the person is unable to speak or hard to understand. T is the time at which the patient needs to visit the hospital right away.

The early prediction of brain stroke occurrence to deal with their risk factors is considered a lifesaving matter. Machine learning and AI techniques can be used to determine the likelihood of a stroke occurring in light of their significant advances in predicting different diseases. Different classification algorithms have been used for predicting strokes with reasonable results [5] [6] [7]. The ensemble method is widely used in medical applications due to its accuracy in predicting different diseases [8] [9] [10]. These methods integrate the prediction outcomes of various classification models to enhance the overall performance. They fall into three main categories: Bagging, Boosting, and Stacking. In bagging, several base classification models are sequentially trained, and then use the majority voting to integrate the prediction outcomes [11]. Whereas, in boosting, several base models are trained sequentially to correct the previous models' errors sequentially [12]. In stacking, the classification task is completed in two stages: the first involves training multiple base models on the entire dataset, and the second involves using a meta-learner classifier to train on the first layer's prediction results to provide the final prediction [13]. The base models in bagging and stacking must be homogenous, but in stacking it could be heterogeneous. Rather than relying solely on the output of a single model, these techniques guarantee the delivery of more accurate and trustworthy results from multiple models [14] [15]. Few studies have used ensemble methods for developing brain stroke prediction models, despite the value of using a stacking ensemble classifier to build predictive models with trustworthy outcomes in a variety of fields, including medical and natural phenomena [16] [17] [18].

The main contribution of this paper is to propose a stroke prediction model using a stacking classifier with multiple base model classifiers to enhance the prediction process. Level one classifiers in the stacking model are the Random Forest, K-Nearest Neighbors, Logistic Regression, Support Vector Machine, and Naive Bayes. In level two, the random forest classifier serves as a meta-learner, combining the prediction results of level one to provide the final prediction. To fulfill this aim, the following tasks have been completed:

1) As with most medical datasets, the used dataset is unbalanced. To balance the dataset, this study conducted the SMOTE technique [19].

2) The standard scalar technique has been used to put the data values on the same scale.

3) The model has been constructed using cross-validation with cv = 10.

4) Hyperparameter tuning was employed on the base classifiers to pick the best parameters for each classifier.

5) In addition to the classical evaluation metrics, the MCC (Matthews Correlation Coefficient) value [20] has been used to evaluate the proposed model as it is more realistic for unbalanced datasets.

This paper is structured as follows: Section II provides the literature reviews that have been done on using machine learning classifiers to generate a stroke prediction model. Section III describes the methodology and proposed model used to construct the prediction process. Section IV offers insights into the assessment of the study's findings. Section V displays the discussion of the study. Finally, Section VI reports the conclusion of the work and aspects of future works.

II. Literature Survey

A. *Building Stroke Prediction Model by using Classification Algorithms*

Most works in the area of brain stroke prediction, using machine learning techniques, are building their models based on standalone classifiers.

Singh and Choudhary [21] built a neural network model for stroke prediction. They used a dataset from the cardiovascular health study (CHS). They applied the Principal Component Analysis (PCA) to minimize the dimensionality of the features and then the decision tree algorithm to choose the most relevant features. The number of instances they have used to build the predictive model is small enough to ensure the accuracy of the results.

Nwosu et al. [22] presented a prediction model for brain strokes by using various machine-learning classifiers. The prediction model was built by using three classifiers Neural network (multi-layer perception), Decision tree, and Random Forest. They achieved 75% accuracy when using the neural network classifier. The main purpose of any medical prediction model is to increase the model's accuracy, but in this study, their results are not sufficient to be trustworthy.

Almadani and Alshammari [23] proposed a stroke prediction model using J48, Jrip, and neural networks (multilayer recognition). The model was built using datasets from the data management department of King Abdulaziz Medical City, Saudi Arabia. Comparing the accuracy of the algorithms, they found that the J84 algorithm achieved a higher accuracy prediction of about 95.25% using principal component analysis (PCA).

Jeena and Kumar [24] developed a stroke prediction model that predicts the probability of developing a stroke based on various risk factors. Model age, atrial fibrillation, gait symptoms, visual impairment, etc. Predictive models were created using support vector machine classification with various kernel functions such as linear, quadratic, RBF, and polynomial. The most accurate function was the linear kernel function, which achieved 91% accuracy. The main drawback here is that the database size is not large enough to make the prediction results more reliable and consists of 350 cases. However, this study did not consider the unbalancing in the stroke dataset, resulting in inaccurate results.

Mahesh and Srikanth [25] wanted to develop a stroke prediction model using decision trees, naive Bayes, and artificial neural network classification algorithms for machine learning. Their study highlights the impact of modifiable and non-modifiable risk factors for stroke. The data set has some risk factors, such as high blood pressure, smoking, and other factors. They utilize the AUC (area under the curve) and ROC (receiver operating characteristics) to measure the total performance of predictive models. The higher the AUC result, the better the prognosis. Their results show that the three algorithms provide acceptable accuracy in the prediction process. A web application was used as the user interface to provide stroke risk alerts. The AUC_ROC score alone cannot be considered a measure of a predictive model.

Sailasya and Kumari [26] trained their prediction model for stroke using Logistic regression, Decision tree, Random Forest, K-nearest neighbors, Support vector machine, and Nave Bayes, six machine learning classifiers. They used a dataset containing risk factors for strokes. They also developed an HTML page as a user interface to get the values of stroke parameters from the user and provide him with the prediction result. They evaluated the overall performance by using the F1 score, accuracy, precision, and recall. The outcome demonstrates that the Nave Bayes classifier has achieved the highest accuracy of 82% compared with the other used classifiers. The achieved accuracy is not accurate enough to predict such a critical medical condition.

Sudha et al. [27] used three machine learning algorithms: decision trees, naive Bayes, and neural networks to build a stroke prediction model. They used a series of data consisting of the patient's troops. The dimensional reduction process is done using a PCA. The decision tree algorithm achieved the highest accuracy of 94%.

Tazin et al. [28] implemented a stroke prediction model using the machine learning techniques of the decision tree, random forest, logistic regression, and voting classifier. The prediction model was built using a stroke dataset that included risk factors. They evaluated the classifiers by using the confusion matrix. The random forest classifier has the highest accuracy of 96% among all classifiers.

Cheon et al. [29] performed a study to decide the ability to predict patients with strokes and the ability of death. They constructed their prediction model using data from the Korean Centers for Disease Control and Prevention (KCDC). Deep neural networks were utilized in the model's construction. They reduced the dataset dimension by using PCA (principal component analysis). They evaluated their model by the confusion matrix. Their area under the curve (AUC) was at its highest, at about 83.5%.

1) Monteiro et al. [30] built a model to predict stroke functional diagnosis. A dataset consisting of 541 patients was used. Popular algorithms such as logistic regression, decision tree, support vector machine, random forest, and XGBoost were used to construct the prediction model. The area under the curve (AUC), which is greater than 90%, was used to assess the final performance of the models.

Amini et al. [31] conducted decision tree and k-nearest neighbors' classifiers in their stroke prediction model. A stroke dataset with various risk factors was used in their study. The evaluation step reveals that the decision tree algorithm outperformed the KNN algorithm in terms of accuracy, achieving a score of 95.42%.

Ali et al. [32] extended their prediction model of strokes by using distributed machine learning algorithms with the aid of a popular platform in big data called Apache Spark. The prediction model was built using a Decision Tree, Support Vector Machine, Random Forest, and Logistic Regression classifiers. A healthcare stroke dataset was used in their work. They evaluated the model's performance using accuracy, precision, recall, and the f1-measure. Out of all classifiers, Random Forest has achieved the highest accuracy of 90%.

Islam et al. [33] executed a cloud-based mobile application that helps provide the user with a warning about the probability of having a stroke. Building a prediction model using classifiers from machine learning, such as Logistic Regression, Decision Tree, K-Nearest Neighbors, and Random Forest, is the fundamental principle on which the web application is based. A dataset consisting of stroke risk factors was used. They evaluated their classifiers by using accuracy, precision, and f1-score. The highest accuracy of 96% was attained by random forest across all performance metrics.

Akter et al. [34] proposed a stroke prediction model with acceptable accuracy. Popular machine learning algorithms like Random Forest, Support Vector Machine, and Decision Tree were developed using their model. The confusion matrix was used to evaluate their prediction model, and the results show that the Random Forest classifier had the highest accuracy (95.30%).

My main criticism of the above works is that the evaluation process is based on traditional measures that do not consider that the datasets are unbalanced by nature (like most medical datasets). Furthermore, few research works have been proposed in the area of stroke prediction using ensemble classifiers.

B. Stroke Prediction Model by using Ensemble Classifiers

Govindarajan et al. [35] used homogenous ensemble classifiers and conventional machine learning algorithms to create their prediction model, including artificial neural networks, Support Vector Machines, boosting, bagging, and random forests. Their work has been done on 507 stroke patients. They evaluated their work using accuracy, precision, recall, sensitivity, specificity, and standard deviation. The neural network classifier has achieved the highest accuracy of 95.3.

Rado et al. [36] built an ensemble model using the homogeneous ensemble method Random Forest (Bagging), Adaptive Boosting, and the heterogeneous ensemble method Stacking and compared their results. They evaluated the model's performance by using accuracy, Mean Squared Error (MSE), precision, and F-measure and compared it with standalone classifiers. Their results show that the ensemble classifiers have attained better accuracy than standalone classifiers. With an accuracy of 87.58%, the stacking classifier is the most accurate.

III. Proposed Model

The proposed ensemble model for predicting whether an individual will have a brain stroke or not is based on many risk factors such as age, gender, heart disease, marital condition, and other factors. The model uses multiple algorithms as shown in Fig. 1. The dataset called Stroke Prediction Dataset has been used [37]. The dataset was firstly loaded, and then data preprocessing techniques were applied such as Simple Imputer for handling null values, Label Encoder for converting categorical values into numerical values, standardization technique for making data on the same scale, and SMOTE technique for making data more effective to build the model. Following that, machine learning algorithms such as K nearest neighbors, Gaussian naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest were used after tuning them by using the hyperparameter tuning concept to find the best hyperparameters for each algorithm. These algorithms are the base learners in level one for building the stacking model. The Random Forest was then employed as a meta-learner in level two of the stacking model, which generate the final prediction by using the predictions from the base learners in level one as input. Finally, the proposed prediction model was evaluated by measuring the accuracy, precision, recall, F1 score, and MCC (Matthews Correlation Coefficient) for realistic evaluation.

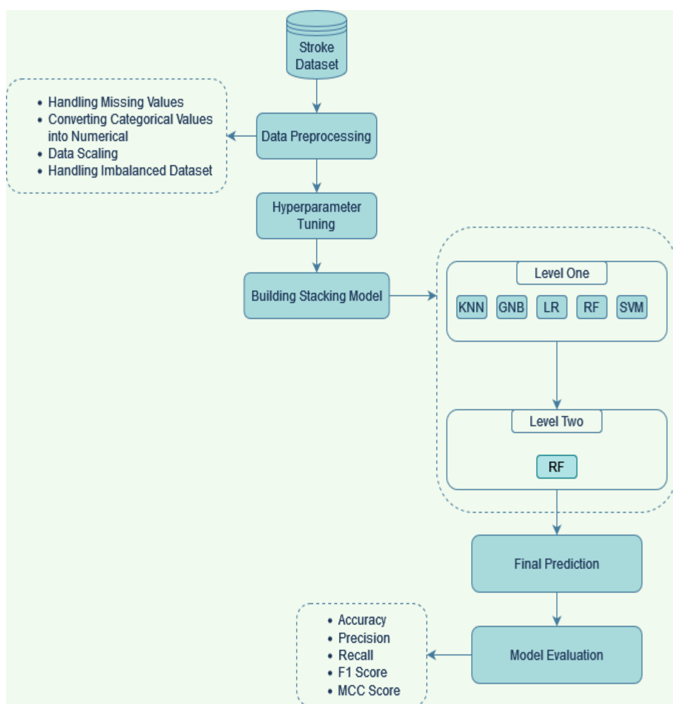


Fig. 1. Stroke prediction model.

A. Stroke Dataset

The prediction model was built by using a stroke prediction dataset from Kaggle that has been presented in Fig. 2. This dataset consists of 5110 rows and 12 columns. The features include ID, gender, age, hypertension, heart disease, ever married, work type, residence type, avg. glucose level, BMI, and smoking status as shown in Table I. The target column is stroke. The identifier column was deleted during the experiment because it does not give any information, it is just the number of the patient.

B. Data Preprocessing

1) Handling missing values: There is an important step before building the predictive model, data preprocessing should be done in which any noise removed, duplication, or incomplete information and handle any missing data. These issues may lead the model to produce incorrect results or affect the overall model quality. In this stroke dataset, there are no duplicated rows. But there are 201 missing values in the BMI column as shown in Fig. 3. Those missing values were filled by using the data column's mean.

2) Converting categorical values into numerical values: The next step was to convert categorical values into numerical ones. The dataset consists of five features with type strings; namely gender, ever married, work type, residence type, and smoking status. The label encoding technique has been employed to convert those features into numerical values.

3) Data scaling: After that, the Standardization technique was used to make the data values in the same range because the input data values fall in different scales. The Standard Scalar function was applied which makes the data values between zero and one, and it is also working with the standard deviation and the data point mean.

4) Handling imbalanced dataset: Class imbalance of datasets is a communal problem in machine learning. Imbalance data can affect the accuracy of machine learning models negatively. This problem occurs when the target class has observations not equal in distribution. That is, there is a high number of instances for a one-class label but an exceptionally small number of observations for the other class exists. In the dataset, the target class of stroke is imbalanced because class "0" which is the number of occurrences of patients who do not have a stroke exceeds class "1" which is the number of patients who have a stroke. As shown in Fig. 4, the total number of instances for classes "0" and "1" is 4861 (about 4.9%) and 249 (about 95.1%), respectively.

To manage this issue, Synthetic Minority Oversampling Technique (SMOTE) which is an oversampling technique is applied. Oversampling requires increasing the number of instances in the minority class by duplicating the records of the minority class to make the instances in the minority class equal to the instances in the minority class. SMOTE technique is a modified version of oversampling in which it is not just duplicating the records in the minority class because it will not add any latest information, it uses the concept of k nearest neighbor to randomly select the neighbor instances and create a synthetic instance. It easily works by selecting examples that are close to the feature space, drawing a line between the examples in the feature space, and drawing a new sample at a point along that line. After applying it, the dataset became balanced with the number of instances of 0 equals 1 in the target class as shown in Fig. 5.

TABLE I. DATASET ATTRIBUTES AND THEIR DESCRIPTION

Attribute Number	Attribute Name	Description
1	id	A unique identifier number for the patient
2	gender	Refers to the gender of the patient
3	age	The age of the patient
4	hypertension	refers to if the patient suffering from hypertension or not
5	heart_disease	refers to whether the patient is suffering from any heart disease or not
6	ever_married	refers to if the patient is married or not
7	work_type	refers to the different types of work
8	Residence_type	refers to the type of the patient's residence
9	avg_glucose_le vel	refers to the level of blood sugar
10	bmi	refers to the body mass index of the patient
11	smoking_status	refers to the patient's smoking status
12	stroke	refers to if the patient had a stroke or not

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
–	–	–	–	–	–	–	–	–	–	–	–
64908	Male	79	0	1	Yes	Private	Urban	57.08	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	162.96	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	73.5	26.1	formerly smoked	0
67855	Female	40	0	0	Yes	Private	Rural	95.04	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	85.37	33	never smoked	0
19584	Female	20	0	0	No	Private	Urban	84.62	19.7	smokes	0

Fig. 2. Stroke prediction dataset.

```
gender                 0
age                    0
hypertension           0
heart_disease          0
ever_married           0
work_type              0
Residence_type         0
avg_glucose_level      0
bmi                  201
smoking_status         0
stroke                 0
```

Fig. 3. Total number of missing values in each column.

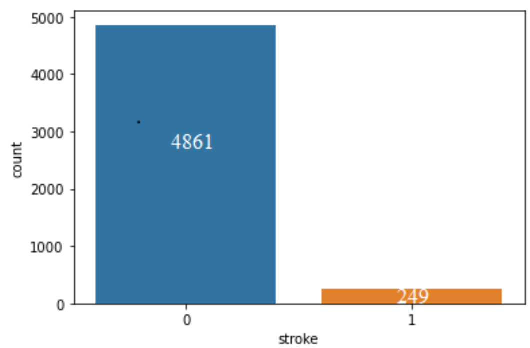


Fig. 4. Stroke proportion before SMOTE.

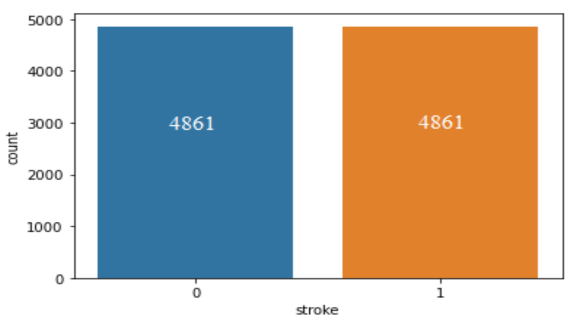


Fig. 5. Stroke proportion after SMOTE.

C. Hyperparameter Tuning

Hyperparameter tuning is the process of finding the optimal hyperparameters for the classifier. It tests several combinations of the parameter values and finds the optimal values that maximize the accuracy of the prediction model. In that work, the grid search technique was used to find the best parameters for using classifiers as presented in Table II.

D. Building Stacking Model

1) Base classifiers for the stacking model: After the step of data preprocessing and hyperparameter tuning, the stacking model was started to build for stroke prediction. The first step is to train multiple heterogeneous algorithms on the dataset, this step is called stacking level one. The second step is to build a meta-model that helps in combining the base learners' predictions with the final prediction. Five popular machine learning classification algorithms were trained on the dataset at level one, which was as follows:

- K-nearest neighbors
- Random Forest
- Gaussian Naïve Bayes
- Logistic Regression
- Support Vector Machine

a) K nearest neighbors: K-nearest neighbors are the most commonly used algorithm for both classification and regression problems in machine learning. It is also known as KNN or K-NN. The basic idea of KNN is to group data points falling in near to each other in the same class. It classifies the new data point based on a similarity measure. It uses Euclidean distance to determine the nearest neighbor class to a data point that needs to be classified.

b) Random Forest: Random Forest is the most commonly used algorithm for classification and regression problems in machine learning based on Bagging ensemble learning. In ensemble learning, it combines the results of the prediction of multiple base classifiers to build one robustness model with higher performance. A random forest consists of several decision trees that were trained individually on a random data sample and subset features. Decision trees produce several results, and these results are combined using the majority voting in the case of the classification problem and the mean average in the case of the regression problem to produce the final result. The greater number of trees in the random forest leads to the highest prediction performance. This algorithm prevents the issue of overfitting and enhances the model's accuracy.

c) Gaussian Naïve Bayes: This classification algorithm employs the Bayes theorem. It assumes that all the input features or attributes are independent of each other. Bayes' theorem finds the probability of an occurrence of an event given the probability of another event that has already occurred before.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

TABLE II. HYPERPARAMETER TUNING FOR BASE CLASSIFIERS

	Best Parameter values with grid search cv	**Best score by selecting the best parameters**
K-Nearest neighbors	neighbors=1 weights='uniform'	0.947
Random Forest	n_estimators= 2000 max_depth= 50 criterion=entropy bootstrap= False random_state=1	0.966
Gaussian Naïve Bayes	var_smoothing= 0.0533	0.776
Logistic Regression	C=0.00294 max_iter=100 penalty=L1 solver=saga	0.792
Support Vector Machine	C=1000 Kernel=rbf Degree=1	0.928

d) Logistic Regression: Logistic regression is an effective algorithm for binary classification problems. It uses some independent features to predict a categorical or discrete dependent variable, such as 0 or 1, male or female, yes or no, and so on. It uses the Sigmoid function that gives the probability values between 0 and 1 instead of giving the output values 0 and 1 by mapping the predicted values to probabilities.

e) Support Vector Machine: The main concept of SVM is to create the best fit line or decision boundary that can split the classes, so it can easily classify the new data point in the correct class in the future. The best fit line or decision boundary is called a hyperplane.

2) Stacking meta learner: Stacking is one of the most efficient machine-learning techniques. It is a widely used ensemble technique because it improves the model performance and solves complex problems. It is used to combine the predictions from multiple models by using a meta-model. In stacking, the dataset is divided into two sets, the first one is the training set and the second one is the test set. This training set is divided into a training set that is trained by heterogeneous base learners to create the first-level models and a validation set that is used by the models to make the predictions of level one, which are used as new features for the second-level meta-learner. This meta-learner is trained on this new training data, which consists of the first-level predictions, and uses the test set to make the final prediction. The major point is to construct a meta-model that is trained with the first-level outcomes. This step helps in providing an accurate final prediction. Fig. 6 shows the sequence of the stacking model. The base model classifiers were K-nearest neighbor, Random Forest, Gaussian Naive Bayes, Logistic Regression, and Support Vector Machine. In the second level, the Random Forest has been chosen to be the meta learner which has been trained on the outputs of the base learners, as it is the most suitable algorithm for providing an accurate result. It has the best accuracy of 96% compared to other stand-alone classifiers.

3) K-Fold Cross-Validation: While building the model, the k-fold cross-validation is conducted to divide the dataset into K collections with equal sizes, Where K represents an integer number. Those collections are called folds. Making iterations equal to the number of folds. At each iteration:

- Take k-1 folds for training and k-fold for validating.
- Changing the folds of training and validating.
- Calculate the accuracy of each iteration.
- Take the average of all accuracies.

It usually gives more accurate results for the model as it trains the model multiple times by changing the training and testing data slot at each iteration. Fig. 7 is an example of cross-validation with five folds.

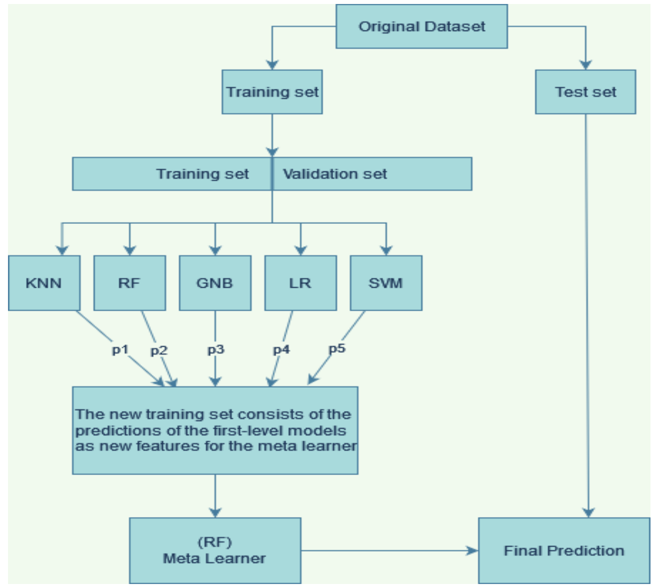


Fig. 6. Sequence of stacking model.

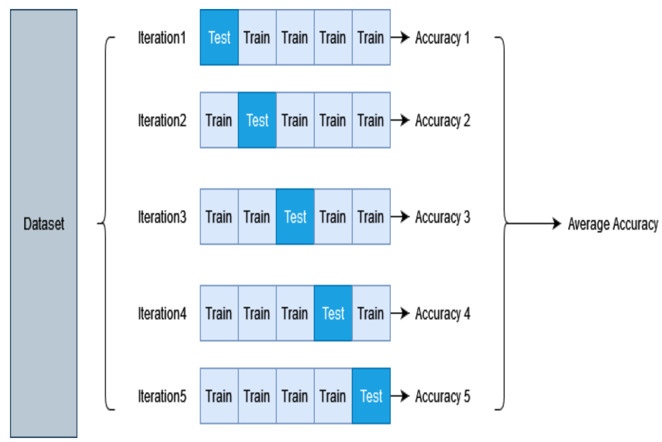


Fig. 7. Example of cross-validation with k=5.

E. Final Prediction

The Random Forest classifier has been used in the second level of the stacking model, which is a meta-learner that combines the results of the first level. This learner trained on those results and provided the final prediction result.

F. Model Evaluation

The prediction model was evaluated by using various evaluation metrics such as accuracy, precision, recall, f1-score, and MCC (Matthews Correlation Coefficient). This paper considered the MCC value for the classifiers because it is an effective measure for binary classification and unbalanced datasets, as in the used stroke dataset. It calculates the correlation between the actual and the predicted values. If that correlation value is higher, that means that the prediction is better. It considered all the confusion matrix values. When the value of MCC is close to one then it means that the model well predicted the actual and the predicted values.

Accuracy = TP/TP+TN+FB+FN

Precision = TP/TP+FP

Recall = TP/TP+FN

F1-score = 2(recall*precision)/(recall+precision)

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

True Positive (TP) Predicted Positive and they are Positive

True Negative (TN) Predicted Negative and they are Negative

False Positive (FP) Predicted Positive but they are Negative

False Negative (FN) Predicted Negative but they are Positive

Positive here means that patient has a stroke (1) and negative means that the patient does not have a stroke (0).

IV. Results

After building the prediction model, the classification algorithms were compared with the accuracy, precision, recall, f1 score, and MCC measures as shown in Table III and Fig. 8:

- All features were used.
- SMOTE was applied to balance the data.
- Standard scalar was applied to make the data values on the same scale.
- Cross-validation was used to build the prediction model with cv=10.
- Hyperparameter Tuning was applied to the base classifiers to choose the optimal parameters for each classifier as shown in Table II.
- Hyperparameter Tuning is the process of choosing the best parameters for the classifier to increase the classifier's performance.
- Built the stacking model with KNN, RF, NB, LR, and SVC in level one and RF as meta learner in level two.
- Model evaluation was done by using accuracy, precision, recall, f1 score, and MCC measures.

TABLE III. Comparison between the Algorithms with Hyperparameter Tuning

	Accuracy	Precision	Recall	F1 Score	MCC
KNN	0.95	0.91	0.99	0.95	0.89
Random Forest	0.96	0.99	0.97	0.96	0.92
Naïve Bayes	0.78	0.79	0.83	0.78	0.55
Logistic Regression	0.79	0.79	0.85	0.80	0.58
SVC	0.93	0.94	0.98	0.94	0.87
Stacking	0.97	0.99	0.97	0.97	0.94

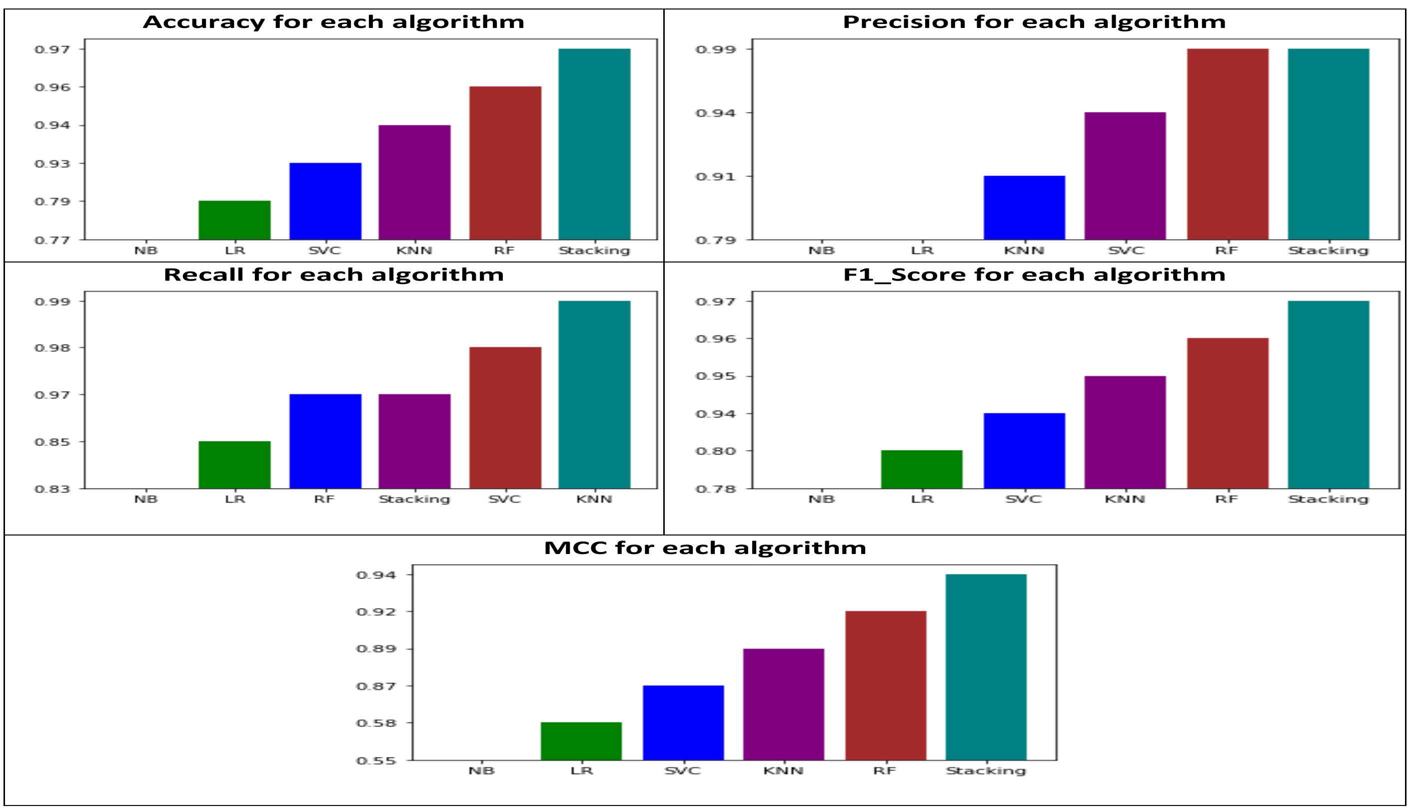


Fig. 8. Comparison between the base classifiers and the stacking model.

From III, the stacking algorithm has achieved the highest accuracy compared with the other standalone classifiers as it achieved the highest accuracy about 97%, and this result shows the efficiency of the ensemble methods. It also achieved the highest MCC value of 94% which means that the stacking model provides an accurate prediction as it well predicts the actual and the predicted values.

Table IV shows the difference between using SMOTE technique to balance the data and without it in comparison with the model accuracy.

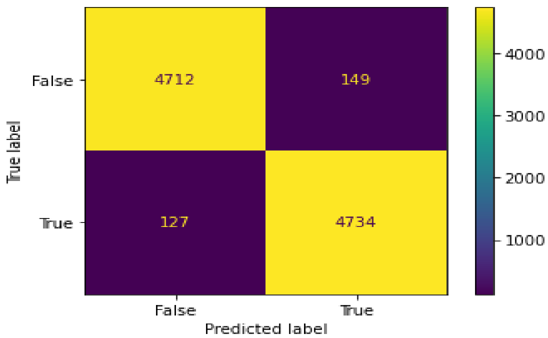


Fig. 9. Stacking confusion matrix.

TABLE IV. ALGORITHMS ACCURACY BETWEEN USING SMOTE AND WITHOUT USING SMOTE

	Accuracy without SMOTE	Accuracy with SMOTE
K-NN	0.91	0.95
Random Forest	0.95	0.96
Gaussian Naïve Bayes	0.87	0.78
Logistic Regression	0.95	0.79
Support Vector Machine	0.95	0.94
Stacking	0.95	0.97

V. DISCUSSION

Over an effort to enhance a brain stroke prediction model a combination of classification algorithms and the stacking ensemble classifier was used to create the prediction model. The outcomes of this research supported the advantages of employing the ensemble method when establishing predictive models.

The prediction model procedures have included the following steps: data preprocessing, in which dataset issues are addressed using various techniques such as column's average for filling in missing values, Label Encoder to convert categorical features into numerical features, Data Scaling to align the data values, and Smote technique to balance this medical dataset. Then, to improve the accuracy of each classifier, hyperparameter tuning was used to determine the most appropriate hyperparameter. The stacking model's level one consists of the Random Forest, K-Nearest Neighbors, Logistic Regression, Support Vector Machine, and Naive Bayes while level two is mainly composed of the Random Forest classifier, which serves as a meta-learner by combining the level one prediction results to generate the final prediction.

Comparing the proposed stacking prediction model with other standalone classifiers, it demonstrated higher classification performance as shown in Table III and Fig. 9. It obtained the best accuracy result of approximately 97%, demonstrating the efficiency of the ensemble methods. As well, it achieved the maximum MCC value of 94%.

VI. CONCLUSION AND FUTURE WORK

This paper intended to demonstrate a stacking ensemble classifier-based prediction of brain stroke disease. Traditional classifiers served as the stacking model's base models. To enhance the final prediction result, the output of those classifiers was combined using the stacking meta-learner.

According to the experimental findings, using a stacking ensemble classifier can significantly increase prediction accuracy as it achieved about 97% and give the highest MCC measure of about 94%, which ensures that the prediction is correct. It also demonstrated that the MCC value is more trustworthy than the conventional measures in the evaluation of the two-class confusion matrix. Using some risk factors, this model aids in accurately predicting whether someone will suffer a brain stroke or not. The future scope of this study will include using other combinations of the base model classifiers in the stacking model. As well, it may extend to utilizing other effective attributes for building the prediction model or employing deep learning algorithms.

REFERENCES

[1] "About Stroke", www.stroke.org, 2022. [Online]. Available: https://www.stroke.org/en/about-stroke.

[2] "The top 10 causes of death", Who. int, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[3] CDC, “About stroke,” Centers for Disease Control and Prevention, 06-May-2022. [Online]. Available: https://www.cdc.gov/stroke/about.htm.

[4] "Stroke",nhs.uk,2022.[Online].Available:https://www.nhs.uk/conditions/stroke/.

[5] A. Roy, A. Kumar, K. Singh, and D. Shashank, “Stroke Prediction using Decision Trees in Artificial Intelligence. Stroke Prediction Using Decision Trees in Artificial Intelligence,” IJARIIT, vol. 4, pp. 1636–1642, 2018.

[6] B. Khalid and N. Abdelwahab, “A model for predicting Ischemic stroke using Data Mining algorithms,” *IJISET*, vol. 2, no. 11, 2015.

[7] M. Rajora, M. Rathod, and N. S. Naik, “Stroke prediction using machine learning in a distributed environment,” in Distributed Computing and Internet Technology, Cham: Springer International Publishing, 2021, pp. 238–252.

[8] I. D. Mienye, Y. Sun, and Z. Wang, “An improved ensemble learning approach for the prediction of heart disease risk,” Inform. Med. Unlocked, vol. 20, no. 100402, p. 100402, 2020.

[9] K. Shilpa and T. Adilakshmi, “Applying ensemble techniques of machine learning to predict heart disease,” in Proceedings of the International Conference on Cognitive and Intelligent Computing, Singapore: Springer Nature Singapore, 2022, pp. 775–783.

[10] Z. Asghari Varzaneh, M. Shanbehzadeh, and H. Kazemi-Arpanahi, “Prediction of successful aging using ensemble machine learning algorithms,” BMC Med. Inform. Decis. Mak., vol. 22, no. 1, p. 258, 2022.

[11] IBM Cloud Education, “What is bagging?” Ibm.com, 11-May-2021. [Online]. Available: https://www.ibm.com/cloud/learn/bagging.

[12] IBM Cloud Education, “What is boosting?” Ibm.com, 26-May-2021. [Online]. Available: https://www.ibm.com/cloud/learn/boosting.

[13] J. Brownlee, “Stacking ensemble machine learning with python,” Machinelearningmastery.com, 09-Apr-2020. [Online]. Available:

https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/.

[14] R. Rosly, M. Makhtar, M. Khalid Awang, M. Isa Awang, M. Nordin Abdul Rahman, and H. Mahdin, “Comprehensive study on ensemble classification for medical applications,” Int. j. eng. technol., vol. 7, no. 2.14, p. 186, 2018.

[15] S. Džeroski, P. Panov, and B. Ženko, “Machine Learning, Ensemble Methods in,” in Encyclopedia of Complexity and Systems Science, New York, NY: Springer New York, 2009, pp. 5317–5325.

[16] B. Pavlyshenko, “Using stacking approaches for machine learning models,” in 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018.

[17] A. Gupta, V. Jain, and A. Singh, “Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications,” New Gener. Comput., pp. 1–21, 2021.

[18] J. Gu, S. Liu, Z. Zhou, S. R. Chalov, and Q. Zhuang, “A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China,” Water (Basel), vol. 14, no. 3, p. 492, 2022.

[19] J. Brownlee, “SMOTE for imbalanced classification with python,” Machinelearningmastery.com, 16-Jan-2020. [Online]. Available: https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/.

[20] 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.

[21] Singh, M. Sheetal, and Prakash Choudhary. "Stroke prediction using artificial intelligence." 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON). IEEE, 2017.

[22] C. S. Nwosu, S. Dev, P. Bhardwaj, B. Veeravalli, and D. John, “Predicting Stroke from Electronic Health Records,” Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, vol. 2019, pp. 5704–5707, Jul. 2019, DOI: https://doi.org/10.1109/EMBC.2019.8857234.

[23] Almadani, Ohoud, and Riyad Alshammari. "Prediction of Stroke using Data Mining Classification Techniques." International Journal of Advanced Computer Science and Applications (IJACSA) (2018).

[24] R. S. Jeena and S. Kumar, “Stroke prediction using SVM,” IEEE Xplore, Dec. 01, 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7988020.

[25] M. Kunder Akash and S. Srikanth, “Prediction of Stroke Using Machine Learning,” ResearchGate, 2020.

[26] G. Sailasya and G. L. A. Kumari, “Analyzing the performance of stroke prediction using ML classification algorithms,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, 2021.

[27] A. Sudha, P. Gayathri, and N. Jaisankar, "Effective Analysis and Predictive Model of Stroke Disease using Classification Methods", International Journal of Computer Applications, vol. 43, no. 14, pp. 26-31, 2012. Available: 10.5120/6172-8599.

[28] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. Monirujjaman Khan, “Stroke disease detection and prediction using robust learning approaches,” Journal of Healthcare Engineering, 26-Nov-2021. [Online]. Available: https://www.hindawi.com/journals/jhe/2021/7633381/.

[29] S. Cheon, J. Kim, and J. Lim, “The use of deep learning to predict stroke patient mortality,” Int. J. Environ. Res. Public Health, vol. 16, no. 11, p. 1876, 2019.

[30] Monteiro, M., Fonseca, A. C., Freitas, A. T., Pinho e Melo, T., Francisco, A. P., Ferro, J. M. and Oliveira, A. L., "Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients", 2018.

[31] L. Amini, R. Azarpazhouh, M. T. Farzadfar, S. A. Mousavi, F. Jazaieri, F. Khorvash, R. Norouzi, and N. Toghianfar, “Prediction and control of stroke by data mining,” International Journal of Preventive Medicine, 2013.

[32] Ali, Abdelmgeid A., “Stroke Prediction using Distributed Machine Learning Based on Apache Spark,” Stroke 28(15), pp. 89-97, 2019.

[33] M. M. Islam, S. Akter, M. Rokunojjaman, J. H. Rony, A. Amin, and S. Kar, “Stroke Prediction Analysis using Machine Learning Classifiers and Feature Technique,” International Journal of Electronics and Communications Systems, vol. 1, no. 2, pp. 17-22, Dec. 2021.

[34] B. Akter, A. Rajbongshi, S. Sazzad, R. Shakil, J. Biswas, and U. Sara, "A Machine Learning Approach to Detect the Brain Stroke Disease", 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022. DOI: https://doi.org/10.1109 /icssit53264.2022.9716345.

[35] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, “Classification of stroke disease using machine learning algorithms,” Neural Comput. Appl., vol. 32, no. 3, pp. 817–828, 2020.

[36] O. Rado, M. Al Fanah, and E. Taktek, “Ensemble of Multiple Classification Algorithms to Predict Stroke Dataset,” Advances in Intelligent Systems and Computing, vol. 998, pp. 93–98, 2019, DOI: https://doi.org/10.1007/978-3-030-22868-2_7.

[37] Fedesoriano, “Stroke prediction dataset,” Kaggle, 26-Jan-2021. [Online]. Available: https://www.kaggle.com/fedesoriano/stroke-prediction-dataset.

Feedforward Deep Learning Optimizer-based RNA-Seq Women's Cancers Detection with a Hybrid Classification Models for Biomarker Discovery

Waleed Mahmoud Ead[1], Marwa Abouelkhir Abdelazim[2], Mona Mohamed Nasr[3]
Information Systems Department-Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt[1, 2]
Information Systems Department-Faculty of Computers and Artificial Intelligence, Helwan University, Egypt[3]

Abstract*—Women's cancers, signified by breast adenocarcinoma and non-small-cell lung cancers, are a significant threat to women's health. Across the globe, the leading cause of death for women is a group of tumors referred to as "female-oriented cancers". The most recent researches in the classification of molecular tumors is the analysis of women's cancers using RNA-Seq data for precision cancer diagnoses. Furthermore, discovering the different genes' patterns behaviors will lead to predict the cancer-specific biomarkers to early diagnosis and detection of cancer-specific in women. An overfit model will be resulted due to the high-dimensional data of RNA-Seq from a small samples of data. In this work, we propose a filter-based selection approach for a deep learning-based classification model. In addition, hybrid classification models have been proposed to be compared with the new modified deep learning model. The Experiments' analysis showed that the proposed filter-based selection approach for a deep learning-based classification model performed better than the other hybrid models in terms of performance evaluation metrics, with an accuracy of 96.7% for RNA-Seq breast adenocarcinoma data and 95.5% for RNA-Seq non-small-cell lung cancer data.

Keywords—Women's cancers; RNA-Seq; deep learning; molecular tumor; hybrid classification models

I. INTRODUCTION

Today, cancer is the number two mortality globally and the number one mortality in both developed and developing nations. Twenty million new cases of cancer are diagnosed each year, and ten million people die from it. Women account for nine million of these cases and 4.4 million deaths globally each year [1]. According to the World Health Organization, breast cancer will be the most prevalent and deadly cancer in women in 2020, accounting for more than two million new cases and 684,996 deaths annually. Lung cancer is the third most frequent type of cancer in women overall, with roughly 607, 465 deaths and 770, 828 cases every year [1].

Determining the presence of cancer, making a primary diagnosis, and identifying new, more effective treatment options could all aid in reducing mortality and morbidity rates. Genetics, individual lifestyle, body shape, age, menopause status, family history, smoking, and history of exposure to carcinogens or viruses, hormone therapy, chemicals, and other airborne particles are all associated with the occurrence and frequency of women's cancers [2]. One of the most crucial methods to investigate genetic correlations in medical investigations is transcriptomics by next-generation RNA sequencing (RNA-seq). Large data sets generated by NGS technologies offer a thorough perspective of the human genome [3]. Numerous molecular structures are adopted by nucleic acids, and these designs are crucial for the storage, processing, and transmission of genetic information [4]. DNA molecules are translated to mRNA for the synthesis of proteins. Proteins are the primary factors in the most fundamental cellular processes. The process via which a fragment of DNA is read and then transformed into an addiction to a protein has excessive awareness in several therapeutic analyses in addition to biological ones [5]. The significant aim of cancer disease research is to recognize the genes that cause normal cells to mutate into cancer [6].

Researchers now have access to an unprecedented amount of tumor genomic and transcriptome data thanks to developments in next-generation sequencing methods. A molecule that is reliably tested and assessed as a marker of healthy biological processes, harmful biological processes, or pharmacologic reactions to therapeutic intervention is referred to as a biomarker [7]. Next-generation sequencing (NGS) technologies are used by RNA-Seq. As a crucial tool, RNA sequencing has been used in many aspects of cancer research and therapy, including the identification of biomarkers and the characterization of cancer heterogeneity and evolution, drug resistance, the cancer immune microenvironment, and immunotherapy, among others [8].

It is critical to develop biomarkers for disease progression and potential therapy response, as this will enable personalized care, enhance clinical outcomes, and accomplish the objective of precision oncology [9]. This kind of technology is gathering data from cells and tissues about variations in gene expression [10]. Depending on whether targeted-exome or whole-exome sequencing is utilized, the potential of RNA sequencing resides in the ability to combine the twin characteristics of discovery and quantification in a single high-throughput sequencing, allowing for the simultaneous investigation of thousands of genes [11]. Finding the set of genes that are associated with and highly expressed in many types of tumor cells is one of the difficult issues in the field of cancer classification [12]. Massive gene data sets with few samples are frequently used to represent gene expression data [13]. The large number of ambiguous and redundant features in gene data has been highlighted as adding to the classifiers' complexity challenges.

There are two significant problems with the RNA-Seq gene expression datasets [13]. Due to the high dimensionality of the RNA-Seq datasets, the datasets are extremely complicated and noisy [14]. In these datasets, only a small number of samples were gathered, even though each sample measures the levels of expression of countless thousands of genes [14]. As a result, the learning model will be overfit due to the dimensionality problem and the problems with such a large dataset. Using gene expression analysis, researchers may categorise malignancies, forecast clinical outcomes, and identify biomarkers connected to the disease. The current main hurdle in the cancer diagnosis problem is thought to be the differentiation of normal from malignant tissues, as well as the selection of the few informative genes [15].

In this work proposed RNA-Seq gene expression classification models that are optimised for deep learning and combined with PDA, SVMRadial, GaussprPoly, NB, RF, NN, and the Glmboost Method. The NCBI GEO accessions GSE19804 and GSE70947 were used to download the RNA-Seq gene expression profiles. And extensive packages that make RNA-Seq analysis possible when using Bioconductor and R programming. It has five modules: feature mapper, preprocessing gene expression, dimension transformer, feature selector, deep learning approach, machine predictors with hyper-parameters, and prediction biomarkers with performance evaluation including accuracy, sensitivity, specificity, precision, the F1 score, and the area under the curve (AUC) score.

The contribution of this work can be summarized as follows:

- Using supervised learning and a deep learning algorithm known as a feedforward neural network with hyper-parameters for model optimization,
- For the selection approach, we introduce filter-based selection for dimensionality reduction methods for selection informative genes by applying the FCBF algorithm.
- To determine the dependencies of genes and identify the optimal subset of genes, the enhanced gene selector and feed-forward neural network classifier combined the statistical results of pertinent genes using the symmetrical uncertainty (SU) assessment.
- We adopted further classification. A model that achieves robust classification with little CPU consumption while maintaining accuracy under test conditions is based on hybrid learning models with feedforward neural networks.

The rest of this paper is organized as follows. Section II discussed some of related works, Section III discussed the materials and methods used in this work, and Section IV illustrates the proposed approach, while Section V discusses our results. Finally, Section VI provides conclusion.

II. Related Work

Different related works have been proposed in the era of detection and the diagnosis of human cancer, and in this section reviews the most recent studies on the use of deep learning and machine learning in the field of malignant tumor gene expression data. Studies on biomarker gene documentation will also be tested. Researchers will be able to evaluate and appraise their suggested analytical methodologies using data on cancer gene expression from the resources they have identified.

Zhang et al. [16], proposed a SVM classifiers based on various features selection to forecast lymphatic metastasis in a range of malignancies. Such classifiers were applied to identify differentially expressed signatures in lymph node metastatic and non-metastatic cancer groups. These SVM classifiers were found to be successful, with an overall accuracy of 81.25% on various profiles with light biomarker sets (seven biomarkers on average). They also contrasted these SVM classifiers with two other benchmark classifiers based on comparable profiles (Random Forest, KNN, and K-Nearest Neighbor, RF).

Han et al. [17], argue that Rao's score statistic is arithmetically appropriate to associate several mechanisms through a typical set of weight factors to yield a biased universal indicator. Next, the weightiness slash statistics measure the purposeful influences of various alteration categories on the target population. Finding cancer-associated genes with mutations that cause the cancer phenotype during cancer genome sequencing is a significant issue for this paper.

Simsek et al. [18], proposed the machine learning classification model for the classification of leukemia subtypes using the gene expression data set from 72 patient records and 7129 gene regions. In the research, machine learning classification techniques such as support vector machines (SVM), linear discriminant analysis (LD), ensemble classifiers (EC), and K-nearest neighbor (KNN) were used. It is evident from experience that the SVM model outperforms the other algorithms despite the fact that the collective test data and collective training data were shared and combined to create a fresh training dataset. Results show that these machine learning models can be helpful in determining the leukemia subtype.

Das et al.'s [19], grouping and classification approach by gene-gene similarity matrix is permeated by the suggested study. The feature selection strategy in this attempt is SVM-RFE. Based on the gene-gene similarity matrix, specific traits are further clustered into several groupings. These pairwise correlation-based clusters use reduction into a smaller set of features for further processing by the neural network, which is required to categorise the various types of cancer.

Yin et al. [20], The CNN-Cox model, which combines a unique CNN framework with prognosis-related feature selection cascaded Wx and has the advantage of fewer computing requests while using light training parameters, has been established as a short and effective survival analysis model. The Cancer Genome Atlas cohort's seven cancer type datasets, including those for head and neck squamous cell carcinoma, bladder carcinoma, brain low-grade glioma, kidney renal cell carcinoma, skin cutaneous melanoma, lung squamous cell carcinoma, and lung adenocarcinoma (LUAD), show that the CNN-Cox model achieved reliable higher C-index values and better survival prediction performance. They demonstrated the use of protein-protein interaction network analysis to identify potential prognostic genes and further investigated the biological roles of 13 core genes, whose high expression is significantly associated with poor survival in LUAD patients.

Houssein et al. [21], proposed an algorithm to choose the most relevant and instructive genes from cancer microarray datasets. The first goal of this study is to select the most predictive genes, and the second goal is to extract the most accurate gene expression datasets with the least amount of difficulty. The most informative genes are chosen from a tiny, filtered dataset that is collected from the IG feature subset evaluator after filtering out irrelevant and noisy genes and getting their relationships from the datasets using the BMO method with the SVM classifier. To calculate this suggested model's efficiency, four benchmark microarray datasets—namely, Leukemia1, Leukemia2, Lymphoma, and SRBCT—were used.

Vaiyapuri et al. [22], proposed a new Red Fox optimizer for deep learning-supported microarray gene expression classification (RFODL-MGEC).The current RFODL-MGEC model aims to improve classification performance by selecting appropriate features. The RFODL-MGEC model employs a novel feature selection technique based on the red fox optimizer (RFO) with the goal of creating an ideal subset of characteristics. A bidirectional cascaded deep neural network (BCDNN) created for data classification is also part of the RFODL-MGEC model.

Shen et al. [23], report that DCGN, a deep learning method, has been proposed for cancer multi-classification tasks; this model is recommended since it can handle high-dimensional sparse gene expression data better than previous models that have been put out. The DCGN performs well on all five of the examined datasets when it comes to classification evaluation factors like accuracy and precision, especially on the BLCA-TCGA and BLCA-CIT datasets.

Rezaee et al. [24], proposed a hybrid method that assigns rank to the five key genes in the microarray data based on soft ensemble and stacking auto-encoders. The least number of genes needed for final classification was found by combining the three soft wrapper techniques with classification using the k-NN algorithm.

At the conclusion of this section, In this work, we are interested in diseases that affect women, and therefore we strive to provide the latest technologies that help in the early detection of these diseases in order to speed up the treatment process and help doctors take accurate measurements and develop medicines suitable for each disease as Targeted therapies are determined according to the biomarkers for women's carcinoma that were discovered through powerful learning models suitable to deal with high-dimension RNA-Seq gene expression with hyper-parameters to optimizing the model, so this model in our study gave the best results with performance evaluation of classification models on test datasets for women's cancer.

III. Materials and Methods Background

A. *Women's Cancers RNA-Seq Gene Expression Datasets*

Biotechnology National Center Information is a key resource for multi-omics research, including genetic data, and it facilitates the advancement of science and health by making biomedical and genomic information accessible. For example, genome, transcriptome, epigenetic, and proteome information are applied to methodology issues in bioinformatics. Fresh genetic structures (i.e., RNA, DNA, ChIP sequence, whole exome sequencing, protein chips, and amino acid structures) are among the most abundant public raw data in omics and are easily accessible via the following group sequencing tools. The RNA-Seq gene expression profiles used in our investigation were downloaded from the Gene Expression Omnibus (GEO) database, a free public database that included various genes (https://www.ncbi.nlm.nih.gov/geo/). Under the accession numbers GSE19804 and GSE70947, the dataset was downloaded. The explanation for each kind of tumor is given in Table I.

B. *Methods and Materials*

In this section, we examine the various methodologies used for the proposed model.

1) Feature mapping: For each row in the gene expression dataset, feature mapping translates the Entrez Gene id to the gene symbol and gene name before using the merging procedure to connect those annotations [25].

2) Preprocessing: Combinations of normalization-transformation and the imputation method. Normalization is a crucial step in the interpretation of RNA-Seq data since normalization- transformation combinations are regulated by preprocessing. To enable samples to be evaluated on the same scale, systematic deviations must be identified and corrected [26]. These systematic changes may result from both within-sample variations such as gene length and sequence composition as well as between-sample variations such as library size (sequencing depth) and the presence of majority fragments. Additionally, for data compatibility, transformations are used. In order to handle missing values in gene expression, imputation is utilized.

TABLE I. WOMEN'S CANCERS RNA-SEQ GENE EXPRESSION DATASETS

Accession number	**Dataset name**	**Number of features (genes)**	**Number of samples**	**Summary**
GSE19804	non-small-cell lung cancer (NSCLC)	54,675 features	120 samples	Even though smoking is the main risk factor for lung cancer, in Taiwan, just 7% of female lung cancer patients had ever smoked, a significantly lower percentage than among Caucasian females. This study provides a thorough examination of the molecular profile of female lung cancer in Taiwan that is not caused by smoking.
GSE70947	breast adenocarcinoma	62,976 features	296 samples	Through accelerating angiogenesis and tissue remodeling in the tumor microenvironment, chronic inflammation aids in the growth and invasion of breast tumors. The intricate interaction between estrogen, which promotes the growth of 70% of breast cancers, and inflammation.

3) The dimension transformer. After the RNA-Seq reads have been mapped to a reference genome or transcriptome, the number of reads mapped to the reference genome can be tallied to determine the abundance of the transcripts. For the approaches to be used, it is crucial that the count values be raw sequencing read counts [27].

4) Feature selector: The high dimensionality of the dataset is one of the main issues with machine learning [24]. The weighting features reduce processing time and redundant data, boosting algorithm performance because the analysis of several features uses a lot of memory and results in overfitting [20]. The method of eliminating all unnecessary and irrelevant genes while also identifying the most informative genes [28]. Finding the group of genes that are associated and highly expressed in many types of tumor cells is one of the other difficult issues in the field of cancer categorization [29]. Gene expression data is frequently characterized by an enormous amount of gene data and a small number of samples. It has become clear that the abundance of confusing and duplicated features restricted in the gene data adds to the classifiers' difficulty. To improve the accuracy of predictive models, this research proposed FCBF filter-based dimensionality reduction approach as a selection method. Gene prioritization, often known as the finding of biomarkers, is another name for the feature selection method.

5) Supervised methods: The cross-validation concept: A method for minimizing bias in the estimation of prediction accuracy is cross-validation [6]. When a classification system is over fitted to a certain dataset, bias might result because the algorithm learns the classification "by heart" but struggles to generalize it to new, untested samples. In a nutshell, the dataset is deterministically divided into a number of training and test sets for cross-validation [18]. Each training set is used to build the model, which is then tested on the test set. Over these fits, the accuracy metrics are averaged. N fits are used in leave-one-out cross-validation, with N training sets of size N-1 and N test sets of size 1. As a result, we employed 10-fold cross-validation to define training control in this study's two data sets, splitting the data randomly into a test set (30% of the dataset) and a train set (70% of the dataset).

a) Learning models: Classification and regression are two examples of supervised learning tasks that attempt to anticipate the intended output based on the input data [30]. For instance, a classification algorithm trained on a dataset of correctly classified genes using supervised learning will learn to recognize diseases. In this study, supervised machine learning algorithms with tuning parameters for the gene expression of women's tumors were incorporated in seven learning models to predict biomarkers.

- Neural Network (NN).
- Support Vector Machines with Radial Basis Function Kernel (SVMRadial).
- Penalized Discriminant Analysis (PDA).
- Naive Bayes (NB).
- Random Forest (RF).
- Gaussian Process with Polynomial Kernel (GaussprPoly).
- Boosted Generalized Linear Model (Glmboost).

b) Feedforward neural network algorithm (FNN): An input, multi-hidden, and output hierarchy shape is the Multilayer Perceptron Architecture's most noticeable feature at first glance (Fig. 1). If input data is provided, the output result is computed directly along the subsequent layers of a multilayer perceptron. This type of neural network operating process is referred to as feedforward [31]. A number is obtained as the current output of each neuron in the middle-hidden layer by multiplying the vector-format output results from the previous layer by a weight vector plus a bias value in the current layer, then feeding the biased weighted sum into a nonlinear function (such as a sigmoid, hyperbolic tangent, or rectified linear unit (ReLU), etc.). The feature layer is a new numeric vector made up of enormous neuron outputs in the same hidden layer.

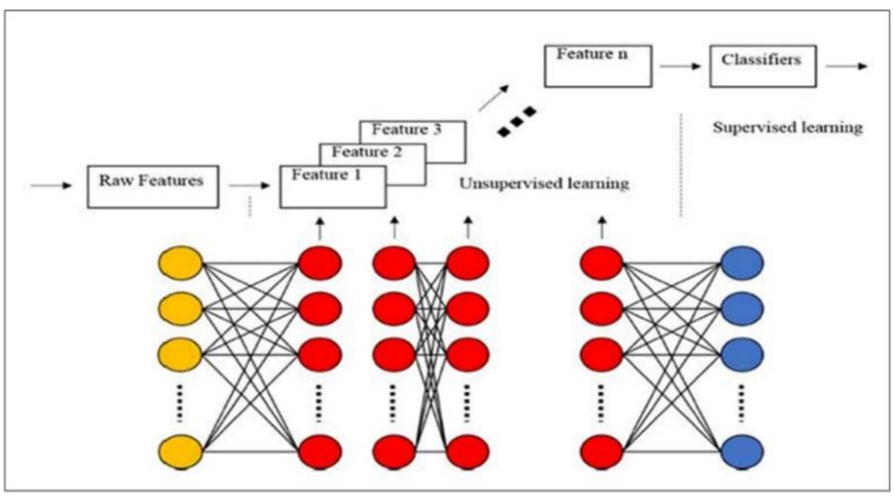


Fig. 1. Multilayer perceptrons architecture.

Evaluation Metrics for Classification Models on test dataset: Several metrics are used to evaluate machine learning approaches. The optimal models are designated using these metrics [30]. To systematically determine the detection effect, the metrics are often used concurrently in our proposed approach as follows:

- Accuracy: A test's accuracy is determined by how well it can distinguish between cancer and healthy instances. Accuracy = TP+TN/TP+TN+FP+FN.
- Sensitivity: A test's sensitivity is how well it can identify cancer instances (true positive rate). Sensitivity = TP / TP + FN.
- Specificity: A test's specificity is how well it can identify healthy instances (false positive rate). Specificity = TN / TN + FP.
- Precision is the ratio of the number of true positive findings to the number of positive results the classifier anticipated. Precision equals TP / TP + FP.
- The F1 score is a direct reflection of the model's performance and is used to evaluate test accuracy. The F1 score can vary from 0 to 1, and the objective is to reach as near to 1 as possible.
- Receiver operating characteristic curve (ROC) / area under curve (AUC) score: The performance of the classification model at every threshold is shown graphically by the ROC curve. The entire region below the ROC curve in two dimensions is known as the AUC. Sensitivity and specificity, two crucial parameters, are produced by this curve.

IV. Proposed Approach

As shown in Fig 2, we employed the RNA-Seq features as the inputs for deep learning-based classification together with other well-known techniques such as PDA, SVMRadial, GaussprPoly, NB, RF, and glmboost to predict biomarkers for women's malignancies.

A. Proposed Approach for Women's Cancers Classification

1) Feature mapper module: By using an organism-level package (an "org" package) that employs a central gene identification (such as the Entrez Gene id) and provides mappings between this identifier and other types of identifiers, annotations can be provided in packages curated by Bioconductor (e.g., GenBank or Uniport accession number, etc.). The number of reads mapped to the reference genome can be tallied to determine the abundance of the transcripts once the RNASeq reads have been mapped to a reference genome or transcriptome. For the approaches to be used, it is crucial that the count values be raw sequencing read counts.

2) A preprocessing module: The normalization-transformation combinations are controlled by using transformations for data compatibility for two reasons: Making non-numeric features into numeric features. Since a string cannot be multiplied using a matrix, it must be converted to a representation that is practically numerical. Resizing inputs to a consistent size. For instance, feed-forward neural networks require input data to be a constant size since they have a certain number of input nodes.

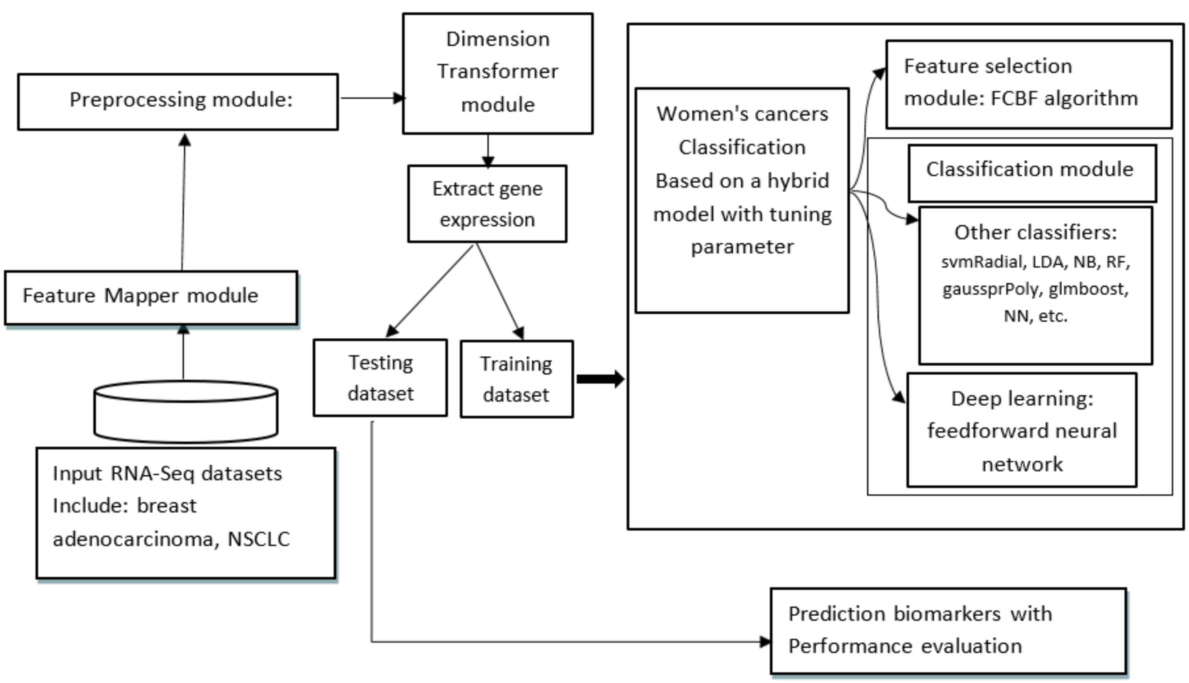


Fig. 2. The pipeline of RNA-Seq female cancer data-based machine learning and deep learning workflow development

NAs in this data collection are another problem. That's typical. I'll explain. Without showing log transformations, the expression data have a very wide range, with several outliers in the higher range. However, as log-transformation frequently produces data with negative infinity (-Inf) values or NAs, which are brought on by all the 0-values in the data since log2 (0) == -Inf, this frequently creates a new issue. As a result, set their NAS value to zero.

Methods used for normalization and transformation: deseq-rlog available in MLSeq package in R programming: Deseq median ratio approach is used for normalization. The normalized data is transformed using a regularized logarithmic formula.

z-Score Normalization (zero-mean Normalization)

3) FCBF-PCA Feature selector module: subsequently applied PCA using the FCBF-reduced datasets. While feature selection shrinks the dataset by deleting useless characteristics, dimensionality reduction uses feature extraction to reshape and simplify the data. Use Bioconductor's FCBF (Fast Correlation Based Filter for Feature Selection) to filter highly correlated genes. A multivariate gene selection technique called a fast correlation-based filter (FCBF) begins with a complete collection of characteristics (genes). It determines the optimum subset by calculating the dependencies of the genes using the symmetrical uncertainty (SU) measurement. An effective computer approach called FCBF is used to discriminate between redundant and irrelevant features.

It assesses each property individually, finds the main correlations, and heuristically eliminates superfluous features. When there are no features, it stops due to an internal halting requirement. Implementing FCBF for GSE70947 and GSE19804 datasets seems like 0.1 as a threshold that is reasonable for this both two datasets. After running FCBF for GSE70947, we went from 62976 features to now a lean set of twenty-four features/genes. As can see, EZH2.1 has the strongest correlation to the target class with an SU value of 0.41, and then comes COL10A1.1 with an SU value of 0.40, and so on. As demonstrated, Table II shows the best informative genes and FCBF for GSE19804, we went from 54,675 features to now a lean set of seventy-nine features/genes. As can see, COL10A1 has the strongest correlation to the target class with an SU value of 0.72, and then comes PROM2 with an SU value of 0.68, and so on. As demonstrated, Table III shows the best informative genes

Subsequently, run the FCBF algorithm using a heat map plot to illustrate the gene correlation of the 79 genes for GSE19804 gene expression: Observing the heat map in Fig. 3, we can see the genes are not either positively or negatively correlated with each other as they appear in a lighter color (blue = negative correlation, red = positive correlation). However, there are some that are quite correlated with each other. However, they are quite correlated with each other. For example, HBM, LOC101927069, and PITPNM2.1 are fairly correlated with each other, and H2AFV, KIAA0101, and COL11A1 are quite correlated as well. As shown in Fig. 4, gene correlation of the 24 genes for GSE70947 gene expression includes EZH2.1, COL10A1.1, and LOC100132724, as well as SDPR, LincRNA.chr2.120459730.120511405_R, and KCNA4.1.

4) Deep learning for classification module: The H2O package serves as the foundation for the deep learning technique, and it uses multi-layer neural networks that have been trained using stochastic gradient descent search to forecast the results of diagnoses. To achieve the best classification outcomes for the neural network setup, H2O enables users to conduct hyper parameter grid searches on several deep learning models. Rectifier or Tanh are often the activation functions.

A single hidden layer site (100 or 200 neurons), two discrete layer locations (10, 20 or 50 neurons each), three discrete layers with 30 neurons each, and four discrete layers with 25 neurons each are predefined for assortments. The feed-in dropout ratio options are available in steps of 0.1 from 0 to 0.9.

Typically, there are zero or two total training samples per iteration, where 0 represents one epoch and 2 represents two. The H2O package chooses the automatic value with caution. The maximum number of epochs (iterations) to run the entire dataset is set at 500. Momentum starts out at a value of 0 or 0.5. (Default zero, without hyper-parameter grid search.).

TABLE II. LIST OF SOME OF THE BEST INFORMATIVE GENES FOR GSE70947

gene symbol	SU values
EZH2.1	0.4083618
COL10A1.1	0.4042125
LOC100132724	0.3069758
lincRNA.chr2	0.2469364
MS4A1.1	0.1901143
PTPN1.1	0.1774414
COL1A1.1	0.1766899
TNKS.1	0.1616661
N4BP2L1	0.1605699
BAX.7	0.1580289

TABLE III. LIST OF SOME OF THE BEST INFORMATIVE GENES FOR GSE19804

gene symbol	SU values
COL10A1	0.72990832
PROM2	0.68648556
SH3GL3.1	0.68648556
GOLM1.1	0.68648556
RTKN2	0.64935173
CA4	0.58091266
DPP6.2	0.54387741
FLJ30901....SCUBE1	0.53771981
HS6ST2.2	0.53771981
CNTN6	0.53265203

Full image size

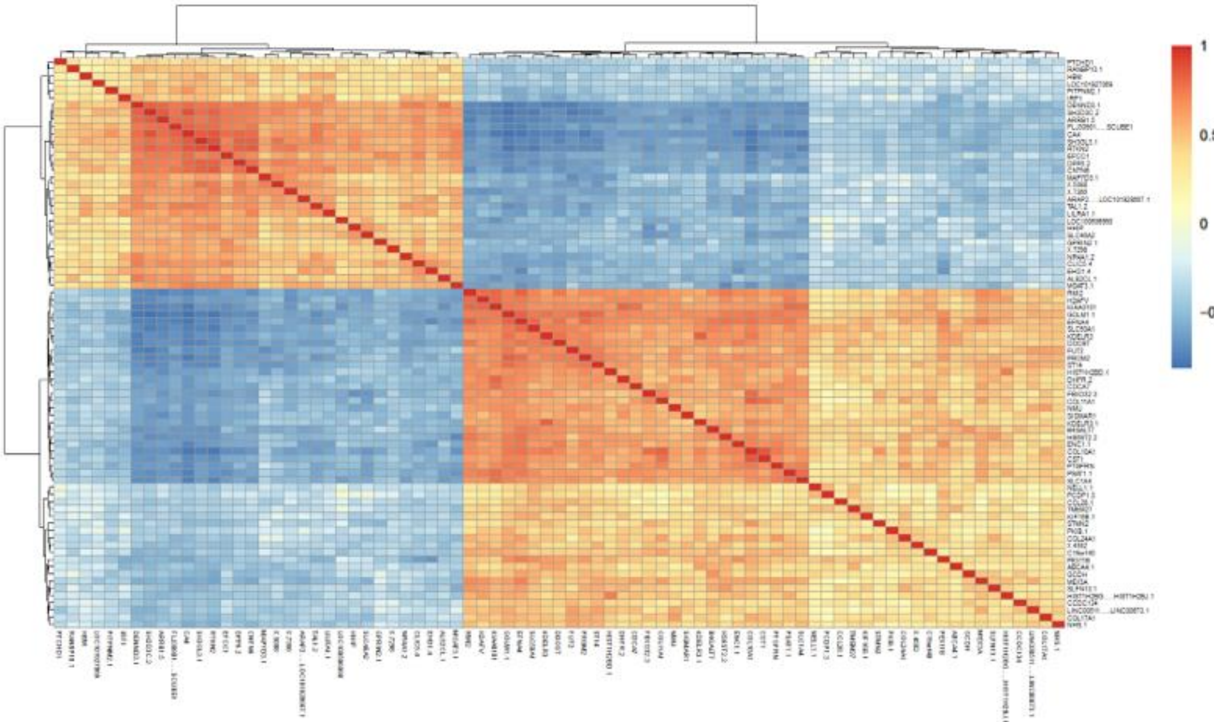


Fig. 3. Heat map for GSE19804 gene expression.

Full image size

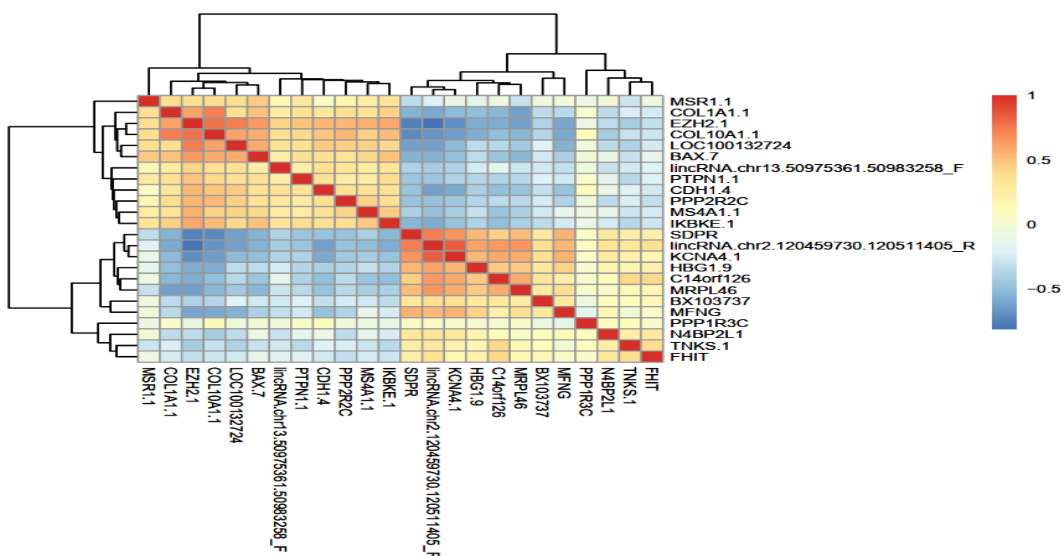


Fig. 4. Heat map for GSE70947 gene expression.

The momentum quickens the iterations for a quicker concourse and dampens the oscillation to achieve the optimal spot. 0.5 Or 0.99 is the adaptive learning rate decay factor (). (Default 0.99, devoid of hyper-parameter grid search) while simultaneously doing quantile regression, the quantile rate (quantile alpha rate in H2O) is set between 0 and 1. Contrary to linear regression, which tests the answer variable's provisional mean, quantile regression tests the provisional quantile. Between 0 and 1 is set as the threshold between quadratic and linear loss (Huber alpha rate in H2O) (default 0.9). In order to make it easier to search on entire combinations of the hyper-parameters, the "random discrete" technique is abandoned.

The most extreme number of models for each run is set at 100 as part of the automatic ML training. If the misclassification values do not increase by 0.01 after five iterations, the training phases come to an end. Score duty cycle, which refers to how frequently validation metrics are computed, is set to 0.025 H2O, which means that no more than 2.5% of the total training time will be spent to create the validation measurements.

Following grid search, the final hyper-parameters for the DL model are listed as follows for the Women's Cancers RNA-Seq dataset: "Rectifier" activation function, four hidden layers with 25 neurons each, insert dropout ratio zero, defaulting training samples each iteration per H2O (value of -2), epoch rate of 430.9, momentum beginning value zero, value of 0.99, quantile regression rate one, and a Huber -value of zero.

Additionally, additional hyper-parameters with an L1 regularization rate of 2.5e-5 and an L2 regularization rate of 2.6e-5 are included. Along with the eleven existing machine learning algorithms that were previously used in H2O for classification, these new DL algorithms are PDA, SVMRadial, GaussprPoly, NB, RF, NN, and glmboost. Based on the data and the size of the sample, to prevent overfitting, N-fold cross-validation with a default N of 10 is feasible. With training data that has been cross-validated 10 times.

To obtain average metrics, we randomly repeated this process ten times. As module number six is illustrated in section VI, bar graphs are used to inform classification metrics such as accuracy, F1 score, area under the curve (AUC) score, precision, sensitivity (SEN), and specificity (SPEC).

V. EXPERIMENT RESULTS AND EVALUATION METRICS WITH TEST DATASET

The proposed architecture was developed using R studio with Bioconductor, with dependencies on the following packages: h2o, dplyr, tidyr, GEOquery, ggplot2, FCBF, pheatmap, devtools, ggbiplot, factoextra, ROCR, limma, psych, caret, foreach, DESeq2, MLSeq, affy, genefilter, hgu133a.db, AnnotationDbi, org, M3C, matrixTests, impute, and gbm The implementation was carried out on a computer server with a core CPU (Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50 GHz, 32 GB RAM) and 64-bit operating system, but it may also work on less powerful machines.

The results indicate that the feedforward neural network algorithm is statistically superior in the metric when compared to other algorithms, with an area under the curve score of 0.982 for RNA-Seq breast adenocarcinoma data and 0.980 for RNA-Seq NSCLC cancer data. Whole evaluation metrics are applied to test datasets, as discussed in Table IV.

So, all evaluation metrics achieved higher rates in breast adenocarcinoma data than in NSCLC cancer data, as plotted in Fig 5. By utilizing a feedforward neural network, this suggested model can assist in the early detection and diagnosis of malignancies in women and, consequently, aid in the formulation of preliminary treatment methods to improve survival. Finally, NSCLC and breast adenocarcinoma cancer may be affected by the top ten possible hub-gene biomarker discoveries.

Biomarkers of lung cancer identified by RNA-non-small-cell sequencing as stated in table V, the top seven choices for differentially expressed genes were found to be shared by all methods. All algorithms found COL10A1 as a common factor, indicating that this gene may be important in NSCLC.

As stated in Table V, the top seven choices for differentially expressed genes were found to be shared by all methods. All algorithms found COL10A1 as a common factor, indicating that this gene may be important in NSCLC. As stated in Table VI, the top four choices for differentially expressed genes were found to be shared by all methods.

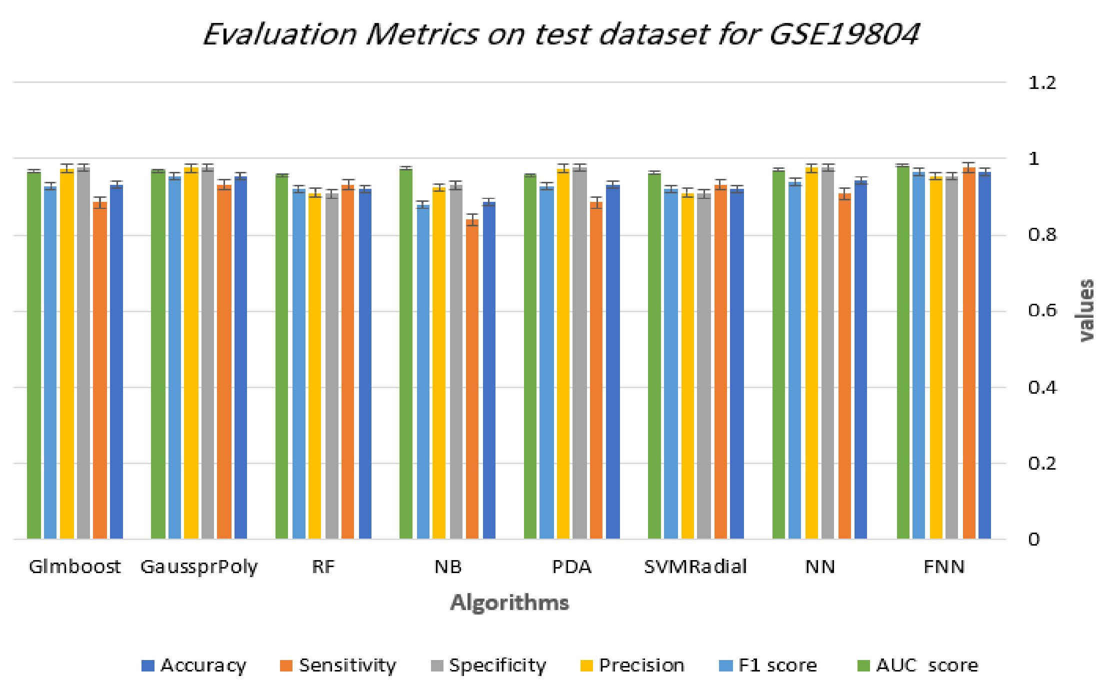


Fig. 5. Evaluation metrics for GSE70947.

TABLE IV. PERFORMANCE EVALUATION OF CLASSIFICATION MODELS ON TEST DATASET

Women's cancers RNA-Seq gene expression Classification using Deep learning with a hybrid Model with Tuning Parameters for Non-Small-Cell Lung Cancer							
Test Datasets	***Algorithm***	***Accuracy***	***Sensitivity***	***Specificity***	***Precision***	***F1 score***	***AUC score***
GSE19804	FNN	0.957	0.971	0.946	0.956	0.965	0.980
	NN	0.9201	0.9316	0.9094	0.9101	0.921	0.9650
	SVMRadial	0.883	0.653	0.927	0.8542	0.891	0.885
	PDA	0.8889	0.7778	1.0000	1.0000	0.875	0.8923
	NB	0.905	0.663	0.920	0.9284	0.907	0.910
	RF	0.892	0.568	0.946	0.8542	0.891	0.877
	GaussprPoly	0.878	0.560	0.939	0.8441	0.882	0.881
	Glmboost	0.906	0.600	0.945	0.9286	0.907	0.911
Women's cancers RNA-Seq gene expression Classification using Deep learning with a hybrid Model with Tuning Parameters for Breast Adenocarcinoma Cancer							
GSE70947	FNN	0.9659	**0.9772**	0.9545	0.9555	0.9662	0.982
	NN	0.9432	0.9091	0.9773	0.9756	0.9412	0.971
	SVMRadial	0.9205	0.9318	0.9091	0.9111	0.9213	0.964
	PDA	0.9318	0.8864	0.9773	0.9750	0.9286	0.957
	NB	0.8864	0.8409	0.9318	0.9250	0.8810	0.975
	RF	0.9205	0.9318	0.9091	0.9111	0.9213	0.957
	GaussprPoly	0.9545	0.9318	0.9773	0.9762	0.9535	0.969
	Glmboost	0.9318	0.8864	0.9773	0.9750	0.9286	0.968

TABLE V. THE BIOMARKERS FOR NON-SMALL-CELL LUNG CANCER RNA-SEQ

SYMBOL	**GENE name**
COL10A1	Collagen Type X Alpha 1 Chain
SH3GL3.1	SH3 Domain Containing GRB2 Like 3, Endophilin A3
GOLM1.1	golgi membrane protein 1
RTKN2	rhotekin 2
EFNA4	ephrin A4
FUT2	fucosyltransferase 2
CLIC5.4	chloride intracellular channel 5

TABLE VI. THE BIOMARKERS FOR BREAST ADENOCARCINOMA RNA-SEQ

SYMBOL	**GENE name**
EZH2.1	enhancer of zeste 2 polycomb repressive complex 2 subunit
COL10A1.1	collagen type X alpha 1 chain
COL1A1.1	collagen type I alpha 1 chain
CDH1.4	cadherin 1

VI. CONCLUSION

Women's cancers are a group of illnesses exhibiting abnormal cell proliferation that have the potential to attack or spread to various body areas. Due to improvements in efficiency and accuracy, RNA-Seq has previously greatly increased the analysis of human genetics and helped to better understand the nature of cancer disorders. In order to classify two different types of cancer, non-small-cell lung cancer and breast adenocarcinoma, this paper introduced an intelligent framework based on a feedforward neural network with an optimization model and applied other integrated learning models suitable for gene expression data for women's cancers. The five modules that made up the suggested strategy were: A core gene identification (such as the Entrez Gene id) is used in the first module, "Feature mapping," which applies an organism-level package (an "org" package) and which contains mappings between this identifier and other types of identifiers (e.g., GenBank or Uniport accession number, etc.). Preprocessing is covered in the second module. The deseq-rlog approach and z-score Normalization and transformation are two techniques used for normalization and transformation. The dimension transformer is the third module. After the RNA-Seq reads have been mapped to a reference genome or transcriptome, the number of reads mapped to the reference genome can be tallied to determine the abundance of the transcripts. For the approaches to be used, it is crucial that the count values be raw sequencing read counts.

The feature selector module, is the fourth module. The fast correlation-based filter (FCBF) was chosen as the method for feature selection in this framework. A deep learning technique, machine predictors with hyper-parameters, and prediction biomarkers with hyper-parameters comprise the final module. Accuracy, sensitivity, specificity, precision, the F1 score, and the area under the curve (AUC) score are among the performance evaluation metrics. Given that the results show that the feedforward neural network approach has an area under the curve score of 0.982 for RNA-Seq breast adenocarcinoma data and 0.980 for RNA-Seq NSCLC cancer data, it is statistically considerably superior to other algorithms in the measure.

References

[1] Saint-Ghislain, Mathilde, Chloé Levenbruck, and Audrey Bellesoeur. "Adverse events of targeted therapies approved for women's cancers." International Journal of Women's Dermatology (2021).

[2] Naz, Faiza, et al. "The role of long non-coding RNAs (lncRNAs) in female oriented cancers." Cancers 13.23 (2021): 6102.

[3] Hamzeh, Osama, and Luis Rueda. "A gene-disease-based machine learning approach to identify prostate cancer biomarkers." Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. 2019.

[4] Moffitt, Jeffrey R., Emma Lundberg, and Holger Heyn. "The emerging landscape of spatial profiling technologies." Nature Reviews Genetics (2022): 1-19.

[5] Vimal, Divya, and Khadija Banu. "Developmental Genetics." Genetics Fundamentals Notes. Springer, Singapore, 2022. 955-1027.

[6] Joseph, M., Madhavi Devaraj, and Larry A. Vea. "Cancer classification of gene expression data using machine learning models." 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, 2018.

[7] Xie, Ying, et al. "Early lung cancer diagnostic biomarker discovery by machine learning methods." Translational oncology 14.1 (2021): 100907.

[8] Hong, Mingye, et al. "RNA sequencing: new technologies and applications in cancer research." Journal of hematology & oncology 13.1 (2020): 1-16.

[9] Kaya, S. Irem, et al. "Recent achievements and challenges on nanomaterial based electrochemical biosensors for the detection of colon and lung cancer biomarkers." Sensors and Actuators B: Chemical 351 (2022): 130856.

[10] Mattath, Mohamed Nabeel, et al. "Nucleic Acid Architectonics for pH-Responsive DNA Systems and Devices." ACS omega 7.4 (2022): 3167-3176.

[11] Salmen, Fredrik, et al. "High-throughput total RNA sequencing in single cells using VASA-seq." Nature Biotechnology (2022): 1-14.

[12] Boussios, Stergios, et al. "BRCA mutations in ovarian and prostate cancer: Bench to bedside." Cancers 14.16 (2022): 3888.

[13] Alhenawi, Esra'A., et al. "Feature selection methods on gene expression microarray data for cancer classification: A systematic review." Computers in Biology and Medicine 140 (2022): 105051.

[14] Gunasundari, B., and S. Arun. "Ensemble Classifier with Hybrid Feature Transformation for High Dimensional Data in Healthcare." 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2022.

[15] Kong, JungHo, et al. "Network-based machine learning approach to predict immunotherapy response in cancer patients." Nature communications 13.1 (2022): 1-15.

[16] (Zhang, Shihua, et al. "Prediction of lymph-node metastasis in cancers using differentially expressed mRNA and non-coding RNA signatures." Frontiers in Cell and Developmental Biology 9 (2021): 605977.) 17. (Han, Y., Yang, J., Qian, X., Cheng, W. C., Liu, S.

[17] H., Hua, X., ... & Lu, Y. (2019). DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies. Nucleic acids research.) 18. Simsek, Ebru, Hasan.

[18] Badem, and Ibrahim Taner Okumus. "Leukemia Sub-Type Classification by Using Machine Learning Techniques on Gene Expression." Proceedings of Sixth International Congress on Information and Communication Technology. Springer, Singapore, 2022.

[19] Das, Ananya, and Subhashis Chatterjee. "Cancer Classification Based on an Integrated Clustering and Classification Model Using Gene Expression Data." *International Conference on Artificial Intelligence and Sustainable Engineering*. Springer, Singapore, 2022.

[20] Yin, Qingyan, et al. "A convolutional neural network model for survival prediction based on prognosis-related cascaded Wx feature selection." Laboratory Investigation 102.10 (2022): 1064-1074. 21.

[21] Houssein, Essam H., et al. "A hybrid barnacle mating optimizer algorithm with support vector machines for gene selection of microarray cancer classification." IEEE Access 9 (2021): 64895-64905. 22.

[22] (Vaiyapuri, Thavavel, et al. "Red Fox Optimizer with Data-Science-Enabled Microarray Gene Expression Classification Model." Applied Sciences 12.9 (2022): 4172.).

[23] Shen, Jiquan, et al. "Deep learning approach for cancer subtype classification using highdimensional gene expression data." BMC bioinformatics 23.1 (2022): 1-17.

[24] Rezaee, Khosro, et al. "Deep learning-based microarray cancer classification and ensemble gene selection approach." IET Systems Biology 16.3-4 (2022): 120-131.

[25] Kolberg, Liis, et al. "gprofiler2--an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler." F1000Research 9 (2020).

[26] Zhao, Yingdong, et al. "TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models' repository." Journal of translational medicine 19.1 (2021): 1-15.

[27] Bu, Dechao, et al. "KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis." Nucleic acids research 49.W1 (2021): W317-W325.

[28] Almugren, Nada, and Hala Alshamlan. "A survey on hybrid feature selection methods in microarray gene expression data for cancer classification." IEEE access 7 (2019): 78533-78548.

[29] Azadifar, Saeid, et al. "Graph-based relevancy-redundancy gene selection method for cancer diagnosis." Computers in Biology and Medicine 147 (2022): 105766.

[30] Liu, Hongyu, and Bo Lang. "Machine learning and deep learning methods for intrusion detection systems: A survey." applied sciences 9.20 (2019): 4396.

[31] Haldorai, Anandakumar, and Arulmurugan Ramu. "Canonical correlation analysis based hyper basis feedforward neural network classification for urban sustainability." Neural Processing Letters 53.4 (2021): 2385-2401.

Transfer Learning for Closed Domain Question Answering in COVID-19

Nur Rachmawati, Evi Yulianti
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

***Abstract*—COVID-19 has been a popular issue around 2019 until today. Recently, there has been a lot of research being conducted to utilize a big amount of data discussing about COVID-19. In this work, we conduct a closed domain question answering (CDQA) task in COVID-19 using transfer learning technique. The transfer learning technique is adopted because a large benchmark for question answering about COVID-19 is still unavailable. Therefore, rich knowledge learned from a large benchmark of open domain QA are utilized using transfer learning to improve the performance of our CDQA system. We use retriever-reader framework for our CDQA system, and propose to use Sequential Dependence Model (SDM) as our retriever component to enhance the effectiveness of the system. Our result shows that the use of SDM retriever can improve the F-1 score of the state-of-the-art baseline CDQA system using BM25 and TF-IDF+cosine similarity retriever by 3,26% and 32,62%, respectively. The optimal parameter settings for our CDQA system are found to be as follows: using 20 top-ranked documents as the retriever's output, five sentences as the passage length, and BERT-Large-Uncased model as the reader. In this optimal parameter setting, SDM retriever can improve the F-1 score of the state-of-the-art baseline CDQA system using BM25 by 5,06 % and TF-IDF+cosine similarity retriever by 24,94 %. Our last experiment then confirms the merit of using transfer learning, since our best-performing model (double fine-tune SQuAD and COVID-QA) is shown to gain eight times higher accuracy than the baseline method without using transfer learning. Further fine-tuning the transfer learning model using closed domain dataset (COVID-QA) can increase the accuracy of the transfer learning model that only fine-tuning with SQuAD by 27, 26%.**

Keywords*—*COVID-19; closed domain question answering; sequential dependence model; transfer learning; BERT

I. Introduction

Corona virus disease 2019, known as COVID-19, has been a popular topic in recent period, from 2019 until today. COVID-19 can cause severe respiratory disease in humans [1]. This disease initially came from Wuhan city in China, and was detected for the first time on December 29, 2019 [2]–[4].

In this COVID-19 pandemic, obtaining information about COVID-19 is really necessary. A specific question answering (QA) system on COVID-19, referred to as closed domain question answering (CDQA) system, is beneficial to provide direct answers to many questions that may arise related with COVID-19 issue. This system will be useful for health practitioners as well as public in general to have deeper knowledge on COVID-19. This motivates us to implement a question answering system on specific COVID-19 topic in this study.

To build an effective CDQA system in COVID-19 using state-of-the-art deep learning approach, a large benchmark consisting of pairs of questions and answers together with a huge collection of documents about COVID-19 issue, is needed. A previous research has built CORD-19 dataset which includes a collection of one million articles about COVID-19 [5]. However, a large number of pairs of questions and answers about COVID-19 from that data are still unavailable to train a deep learning-based QA model. A previous work has paid some attention to this matter and put some effort to build 2019 question-answer pairs in COVID-QA dataset [6]. This number, however, is still relatively low to learn a deep learning-based QA model from scratch.

To tackle the above problem, we adopt transfer learning technique to build more effective CDQA system in COVID-19. Transfer learning is commonly used to make use of knowledge extracted from high-resource data to solve the task on the low-resource data. It becomes our intuition to use transfer learning for our case of low resource COVID-19 QA benchmark. Transfer learning approach enables us to exploit rich knowledge learned from an available large benchmark of open domain QA system, i.e., SQuAD dataset [7] to extract answers about COVID-9 topic from low-resource closed domain COVID-19 dataset, i.e., CORD-19 [5].

Some research has used transfer learning method to build CDQA system [8] [9]. Alzubi et al. [8] and Yang et al. [9] both implemented CDQA system using dual architecture (retriever-reader system) and transfer learning principle. Here, retriever component will find subset of relevant documents, and reader component trained on high-resource data will find the span of text from those documents as the answer to the given question. In this work, we aim to improve the accuracy of CDQA system of Alzubi et al. [8] and Yang et al. [9] by enhancing the retriever component of CDQA system. By improving the accuracy of retriever, it will result in the increasing the accuracy of the reader in extracting answers. While Alzubi et al. [8] and Yang et al. [9] used TF-IDF vectorizer with cosine similarity, and BM25, respectively, as their retriever, we propose to use Sequential Dependence Model [10] as the retriever component in the CDQA system. Here, the retriever model of Alzubi et al. [8] and Yang et al. [9] will be changed to the CDQA systems. Our intuition in using SDM is because it optimizes the scoring functions of documents by modelling question term dependencies, which therefore can capture different variations of question terms in the documents. While many previous work has demonstrated the effectiveness of this model for an ad-hoc retrieval [10]–[12], to the best of our knowledge, none of previous studies have investigated the use of SDM for QA system, more specifically dual architecture CDQA system. This becomes a research gap that is fulfilled by this work.

In this paper, there is some following research question (RQ) that has been investigated:

RQ1: Does the use of SDM as retriever in the retriever-reader architecture improve the performance of state-of-the-art transfer-learning-based CDQA systems in COVID-19 using TF-IDF+cosine and BM25 retriever?

RQ2: What are the effects of tuning some parameters (top n retrieved documents, passage length, and reader variations) on the accuracy of CDQA system in COVID-19? What are the optimal parameter settings for our system?

RQ3: To what extent the use of transfer learning technique using SDM as retriever in reader-retriever architecture can improve the CDQA system in COVID-19 that does not use transfer learning technique?

Our contribution in this work are as follows: (1) we propose to use Sequential Dependence Model as retriever in retriever-reader architecture of CDQA system in COVID-19; (2) we conduct empirical evaluation on the effectiveness of our proposed method compared to the state-of-the-art CDQA method used by Alzubi et al. [6] and Yang et al. [9]; we also perform evaluation to show the merit of using transfer learning technique by comparing our method with the method that does not use transfer learning.

The rest of this paper is organized as follows. Section II describes about the related works. Section III explains our research methodology such as the dual architecture of our CDQA system, dataset, and baseline methods. Section IV and V presents our experiment details and results. Section VI discusses about the limitation and the challenges of this work. Finally, Section VII concludes this study and answers all research questions outlined above.

II. Related Work

A. Closed Domain Question Answering (CDQA)

Question Answering is a system that is used to answer the question given by users. The input to the system is a question and a list of documents. The system tries to find the answer by finding the start/end positions where the answer is located within the text. Question answering can be divided into two parts: closed domain and open domain. Closed domain question answering (CDQA) is a question answering system that has ability in answering questions regarding the specific domain by exploiting mainly domain-specific knowledge [8]. An example of this closed domain question answering is COBERT implemented by Alzubi et al. [8]. They used a specific domain of COVID-19 when building such CDQA system. Meanwhile, open domain question answering is a question answering system that has an ability to answer questions about any domains and rely on general ontology and world knowledge [10]. In open domain question answering, they often use Wikipedia as the unique knowledge source when looking for answers [13]. An example of this open domain question answering is DrQa [13]. In this paper, our work focus is on building CDQA system in COVID-19 domain.

B. Machine Reading Task for Question Answering

Machine reading task for question answering has been accelerated in recent years. This technique is about how machines can read and learn from articles that have been given to the QA system. Some dataset has been built for this machine reading text such as SQuAD [7], WikiQA [14], CoQA [15], QACNN [16], etc.

Some previous researches have paid attention about this topic. DrQA has been built for open domain question answering using a retriever-reader architecture. The retriever is a module using bigram hashing and TF-IDF matching that will retrieve relevant articles from Wikipedia. Then, the reader is RNN model that detects answer spans from the relevant documents. In general, DrQA pipeline combines bigram hashing with TF-IDF matching retriever and bidirectional RNN paragraph reader [13].

Some recent methods have utilized BERT for question answering, as a result of the effectiveness of BERT that has been shown in various text processing tasks in previous work [17], [18]. Alzubi et al. [8] proposed COBERT, COVID-19 Question Answering System Using BERT. The system uses a retriever-reader architecture and CORD-19 dataset [5] as the input. They used TF-IDF vectorizer with cosine similarity, implemented using scikit-learn [19] tool, to get the top *N* most relevant documents. These documents are then split into passages to be inputted into the reader. The reader used BERT model that was fine-tuned with SQuAD dataset [7] to identify the answer spans based on the final score. These answers were then ranked based on this score.

Yang et al. [9] proposed end-to-end open domain question answering with BERTserini, a combining BERT with Anserini toolkit. With the Wikipedia article and SQuAD dataset [7] as the input, BERTserini tried to use dual architecture (retriever-reader). In retriever, they used Anserini IR toolkit with BM25 [20]. The retriever will retrieve a set of documents and produce a set of text segments (passages) to be inputted into the reader. They used BERT reader that has been fine-tuned with SQuAD dataset [7] to identify the answer span. The system was shown to outperform the DrQA system [13]. Nogueira et al. [21] also adopted BM25 and BERT model to perform passage / answer retrieval.

Semnani et al. [22] proposed Mindstone, a domain-specific question answering system using Wikipedia and Snowflake text corpora. In the Mindstone pipeline, there are three components: retriever, ranker, and reader. In retriever, they used Anserini based on Lucene version 8.0, and Okapi BM25. Then, Neural RM3 (Relevance based language models) is used as ranker to expand the query and re-rank the documents based on the new score given by the ranker. For reader, they used BERT-base model to rank the retrieved answers. The system outperformed the BERTserini system in terms of EM (Exact Match) dan F1-scores.

In this paper, we use retriever-reader architecture in machine reading task for question answering, similar to Alzubi et al. [8] and Yang et al. [9]. However, we use different retriever using Sequential Dependence Model (SDM) to improve the performance of CDQA system.

C. Transfer Learning

Transfer learning is a method using machine learning / deep learning that is commonly used to improve the performance on low-resource tasks. Transfer learning describes the learning schemes when information in source task is used to achieve some improvement in target task performance [23].

Some previous researches on question answering system used this method. Akdemir et al. [24] proposed transfer learning for Biomedical Question Answering. They used this method because available datasets in biomedical question answering are limited. Therefore, with transfer learning schema, they want to transfer information learned in high-resource tasks which is a similar domain with the source task, into a low-resource target task. They hope by applying transfer learning can improve the performance on their question answering system. Other research that used transfer learning method is Syed et al. [25]. They used transfer learning by using SQuAD dataset [7] to fine-tune the question answering system. Alzubi et al. [8] also used transfer learning method to build closed domain question answering system in COVID-19. They used transfer learning because available dataset in COVID-19 question answering is still limited. In this paper, transfer learning method will be used too. It is the same approach with Alzubi et al. [8] and Yang et al. by utilizing SQuAD [7] dataset to fine-tune the reader component of our CDQA system. However, we propose to replace the retriever component of Alzubi et al. and Yang et al. with SDM (Sequential Dependence Model) that has been shown to achieve satisfactory results in previous work on document ranking [10]–[12].

III. Methodology

A. System Architecture

In this work, we use dual architecture (retriever-reader) of question answering system similar to Alzubi et al. [8] and Yang et al. [9]. The difference is that the architecture does not use TF-IDF vectorizer or BM25 as retriever. We implement SDM (Sequential Dependence Model) as retriever using PyTerrier tool [26]. Fig. 1 illustrates the architecture of our system. Initially, the collection of COVID-19 articles, CORD-19, are preprocessed, and then indexed using PyTerrier. The retriever is implemented using SDM and the reader is implemented using BERT model that was fine-tuned using SQuAD dataset. The flow of the process in our system is as follows: Given a question input, the SDM retriever will retrieve a top-N relevant documents that further will be split into passages. These passages will be inputted into the reader that will extract the answers from the passages and rank them to generate the ranked list of answers.

1) Retriever system: Sequential dependence model (SDM) was proposed by Metzler et al. [10] to optimize the document scoring function by including proximity in the query based on the occurrences of single terms, ordered phrases, and unordered phrases. The SDM method has an assumption that all pairs of sequential terms extracted from the query are dependent [27]. It assumes that the occurrences of adjacent query terms are related.

$$score_{SDM}(Q,D) = \lambda_T \sum_{q \subset Q}^{|Q|-1} f_T(q,D) + \lambda_O \sum_{i=1}^{|Q|-1} f_O(q_i, q_i + 1, D) + \lambda_U \sum_{i=1}^{|Q|-1} f_U(q_i, q_i + 1, D) \quad (1)$$

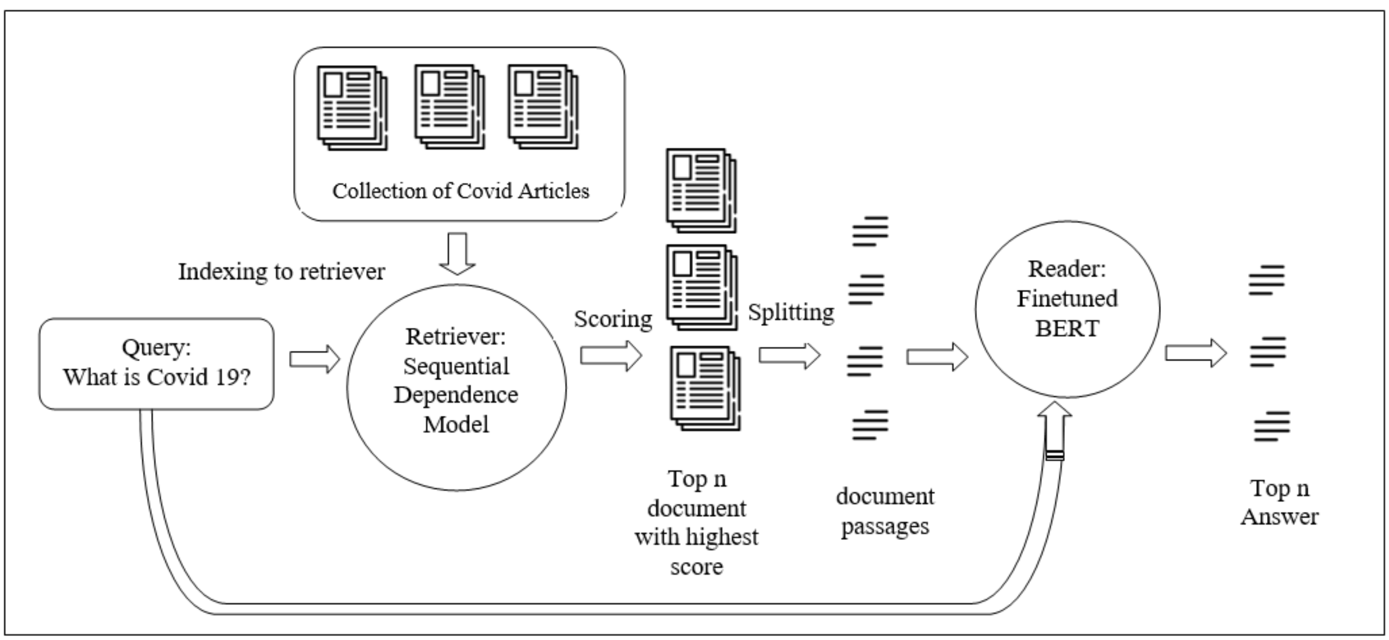


Fig. 1. Architecture System.

The SDM scoring function is described in Eq 1. Given a query Q, the function will calculate the SDM score for each document D in the collection. There are three types of features in SDM: single term features which are standard unigram language model features (fT), exact ordered phrase features which are words appearing in sequence (fO), and unordered window features which require words to be close together, but not necessarily in an exact sequence order (fU) [28]. fT, fO, fU are term frequency for the respective features, and λT , λO, λU are weight of the respective features.

System will rank documents in the collection based on the result of SDM scoring function, and top-n documents will be retrieved. We further split each of these documents into passages and choose eight sentences as the passage length, following the setting used by Alzubi et al. [8].

2) Reader system: The reader system receives a set of passages from top-n documents retrieved by the retriever system. There are three model used as the reader. They are BERT-base-uncased, BERT-large uncased, and BERT-large-word-whole-masking.

- BERT-base-uncased: BERT-base-uncased is variation of BERT, a transformers model pretrained on a large corpus of English language [29], using base and uncased version. Uncased means that it does not make any difference between lowercase and uppercase.
- BERT-large-uncased: BERT-large-uncased is another variant from BERT using large and uncased version. This model was pretrained using higher number of parameters and attention heads than base version of BERT model as a result of using a higher number of encoder layers. BERT-base has 12 encoder layers and 768 hidden layers with 12 attention heads and 110 million parameters. On the other hand, BERT-large has 24 encoder layers and 1024 hidden layers with 16 attention heads with 340 million parameters [11].

The architecture of the single encoder from BERT can be seen from Fig. 2. We can see that the architecture of single encoder consist of input embedding, positional encoding, and N block encoder. In input encoding, word will be converted into vector with some steps: tokenization, numericalization and word embedding generation. After converting into vector, the next part is positional encoding. In this part, it will add positional information. Then, the data will go through N encoder blocks in iterative process. This process will capturing more complex relationships between words in the input sequence.

- BERT-large-uncased-whole-word-masking: BERT-large-uncased-whole-word-masking is another BERT version which was pretrained using new technique, called as ‘whole word masking’ [29]. In this technique, all tokens that are associated with a word will be masked only at once. Meanwhile, for the overall masking rate will remain the same.

The reader component in our CDQA system is each of the BERT model above that is fine-tuned with SQuAD version 1.1 [7]. The CDQA tool is used to fine-tuned the reader. To fine-tune the reader, every document passage will be paired with the query and be transformed into BERT format. Then, the reader will be calculated start logit score and end logit score from every word based model. The logit score itself is the logarithm of the odds from p/(1-p) where p is the probability. Then, the start index is determined based on the start logit value and end index based on end logit value. Once found, predictions are taken based on the start index and end index.

Predictions are taken with several criteria: predictions cannot exceed the maximum answer length, the start index cannot exceed the end index, and some other criteria. Then reader will determine the probability value of each prediction by calculating the softmax function. After that, the results are issued in the form of a tupple by giving three best answer and the order of the answers based on final score. Final score is calculated from retriever score and reader average of start and end logits.

B. Dataset

Table I summarizes all dataset used in this work. In general, there are three datasets utilized in our experiment: SquAD (Stanford Question Answering Dataset) version 1.1, CORD-19 (COVID-19 Open Resource Dataset), and COVID-QA (Question Answering Dataset for COVID-19).

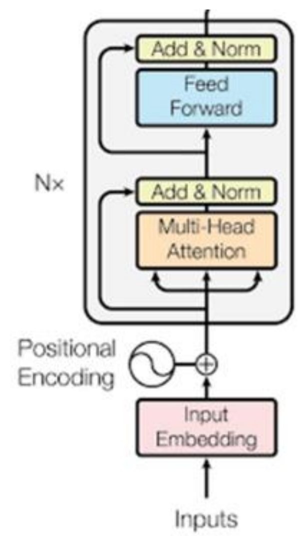


Fig. 2. Architecture from Single Encoder.

TABLE I. STATISTIC DATASET

Dataset	Total Documents	Total Question-Answer Pairs	Topic
SQuAD v.1.1	442	87.599	Open domain
CORD-19	368.618	-	COVID-19 (closed domain)
COVID-QA	147 scientific articles in CORD-19.	2.019	COVID-19 (closed domain)

In order to do transfer learning in our CDQA system, we use SQuAD dataset version 1.1. (SQuAD) as an auxiliary resource. SQuAD will be fine-tune to our reader system in BERT model. SQuAD itself is a reading comprehension dataset. This dataset consists of questions that have been posed by crowd workers from Wikipedia articles. The answer of every question in this dataset can be as a segment of text / span from the corresponding reading passage. This dataset contains over 100.000 question-answer pairs on over 500 articles [7]. This dataset split into training set, development set, and test set. But in this work, we just use training set to fine-tune the reader. So, the squad dataset that we use is 87.599 question-answer pairs on 442 articles. This dataset can be downloaded from https://raw.githubusercontent.com/rajpurkar/SQuAD-explorer/ master/dataset/train-v1.1.json.

CORD-19 dataset use as input in this work. CORD-19 will be indexed by our retriever in order to produce document ranking for every given question to our CDQA system. This dataset is a collection of articles/academic paper in COVID-19, SARS-CoV-2, and related coronaviruses. This dataset was collected by the Semantic Scholar team at the Allen Institute for AI. This dataset consists of over 1,000,000 scholarly articles [5]. But in this research we use 368.618 document that contain the fulltext. This dataset can be downloaded from https://huggingface.co/datasets/cord19.

For evaluation part, COVID-QA dataset [6] will be used because it contains question-answer pairs from the CORD-19 dataset [5] that can provide ground truth answers for the questions. All of COVID-QA dataset will be used to evaluate our CDQA sytem with the baseline. We also use this dataset to evaluate the effectiveness of transfer learning by split this dataset into 70% for fine-tuning and 30% for evaluation. COVID-QA dataset itself consist of 2.019 question-answer pairs. This dataset annotated by volunteer biomedical experts. They selected 147 scientific articles that mostly correlated with COVID-19 from dataset CORD-19 when building this dataset [6]. This dataset can be download in https://github.com/deepset-ai/COVID-QA.

C. Baselines

Three baselines will be used to test the effectiveness of our CDQA system:

- COBERT (TF-IDF+Cosine+BERT) [8]: COBERT [8] is a closed domain question answering that uses TF-IDF Vectorizer with cosine similarity as retriever in order to get top N documents that are related to query.
- BM25+BERT [9]: This baseline uses BM25[9] [20] as retriever to retrieve top N documents that are related to query.
- DLM+BERT: This baseline use Dirichlet Language Model [30], [31] as Retriever to retrieve top n documents that are related to query.

All the above baselines using retriever-reader architecture of CDQA system. They all use BERT model that are finetuned using SQuAD dataset. They only differ in terms of the method used in the retriever component. After the top N documents are split into passages (with the size of passages is eight sentence), the reader then predict / extract answers from the passages and rank them as the results.

IV. Experiment

A. Pre-processing

In the indexing process using PyTerrier, we do some pre-processing in CORD-19 dataset, such as: removing some punctuation, removing stopwords, and lowercasing the text. Then the next process is replacing multiple space into a single space.

B. Hardware

In this work, Google Colab Pro + is used as the machine. This machine use 1 GPU with Tesla T4 type running on CUDA 11.2. Because of technical issue, when doing evaluation part, we move to DGX-1, machine from Tokopedia AI Centre, Universitas Indonesia. It uses 1 GPU with Tesla V100.

C. Finetuning Process

The next process is fine-tuning the reader component based on pretrained language model BERT. Here, BERT-base-uncased, BERT Large Uncased, and BERT-large-uncased-word-whole-masking models are fine-tuned on open domain QA dataset, i.e., SQuAD dataset version 1.1. The hyperparameter that is used to finetune the reader can be seen in Table II.

TABLE II. Hyperparameter to Fine-tune Reader

Hyperparameter	Rincian
Training_batch_size	4
Learning_rate	3e-5
Num_train_epoch	1
Optimizer	adamw
Max_seq_length	384
Max_query_length	64
Gradient_acumulation_step	3
Doc_stride	128

The resulting models are then used to extract and rank answers in closed domain dataset about COVID-19, i.e., CORD-19. By conducting this process, the transfer learning approach is applied by utilizing knowledge learned from large open domain QA dataset to solve the closed domain QA task.

D. Evaluation Metric

In evaluation part, there are two metrics that will be used. They are exact match and F1-score. Exact Match (EM) measures if a resulting answer is exactly similar to the ground truth answer. The score will be 1 if the predicted answer similar with ground truth. Otherwise, the score will be 0.

F1 score is more general metric, that combine together the precision and recall scores. Precision is the ratio that is calculated by the number of the shared words to the total number of words in the prediction. While recall is the ratio that is calculated by the number of shared words to the total number of words in the ground truth. To calculate F1-score, we use the Eq. below (2).

$$F_1 = \frac{2}{recall^{-1} + precision^{-1}} = 2\ x\ \frac{precision - recall}{precision + recall} \quad (2)$$

V. RESULT

A. The Effectiveness of SDM as Retriever in Transfer-Learning-based CDQA System

In the evaluation part, we compare our system with three transfer-learning-based CDQA systems that use retriever-reader architecture as baselines. They are TF-IDF+Cosine+BERT that has been proposed by Alzubi et al. [8], BM25+BERT that has been proposed by Yang et al. [9], and DLM+BERT. Table III shows the comparison of the EM and F1-scores of our method against all baseline methods. The models presented in the table using top 10 document as the retriever output and eight sentences as the passage length.

Our result in Table III show that the use of SDM retriever can improve the F-1 score of the state-of-the-art baseline CDQA system using DLM, BM25 and TF-IDF+Cosine similarity retriever by 3,26% , 3,26% and 32,62%, respectively. Besides, there is improving in EM score when using SDM retriever. SDM retriever can improve EM by 63,06 % for TF-IDF+Cosine similiarity, 2,96% for DLM and BM25. From the result, it turns out that our CDQA system using SDM retriever outperforms all baseline methods. It related with how SDM work. SDM uses various query representations by rewriting each input query in close proximity so that results will be more accurate.

We also explore the answer from two question from all retriever randomly based on Table III in Table IV.

Table IV shows us that one of three answers from SDM retriever and BM25 retriever close to the ground truth both of two questions. Because of that, their F-1 score and EM become one. While, DLM gets the third position, because DLM gets wrong in all answer in first question but true in next question because one of the answer is close to ground truth. While TF-IDF+Cosine similiarity gets the bad result in both of two question because EM Score and F1-score is 0. This cases show us that our SDM retriever is better than TF-IDF+Cosine similiarity practically.

B. The Effect of Tuning the Top N Retrieved Documents, Passage Length, and Reader Variations on the Effectiveness of CDQA System

Table III shows us that SDM becomes the best model among others. Because of that, we try to explore some variation analysis in SDM retriever.

1) Top n document variation: We vary SDM retriever with reader bert based uncased with retriever setting: top five document, top 10 and top 20 document.

Table V shows us that if SDM-BERT-base-uncased sets in top 10 document, it will improve the Top5-SDM-BERT-base-uncased by 10,73% in EM and 3,53% in F-1 score. Then, changing top n document into 20 document can improve Top 5-SDM-BERT-base-uncased by 21,46% in EM and 9,55% in F-1 score. In conclude, setting parameter into top 20 document will make CDQA system become the best result among others.

TABLE III. EVALUATION ON TOP 10 DOCUMENT IN COVID-QA.

Model	EM (%)	F1 Score (%)
Top10-TF-IDF+Cosine-BERT-based-uncased	5.35	17.66
Top10-DLM-BERT-base-uncased	8.42	22.68
Top10-BM25-BERT-base-uncased	8.42	22.68
Top10-SDM-BERT-base-uncased (ours)	**8.67**	**23.42**

TABLE IV. EVALUATION IN TWO QUESTIONS IN TOP 10 DOCUMENT IN COVID-QA

No.	Question			
1.	Question	What is the most common species of Human Coronavirus among adults?		
	Answer (Ground Truth)	HCoV-OC43		
	Method	**3 Answer**	**EM**	**F1-score**
	Top10-TF-IDF+Cosine-BERT-base-uncased	• bats, • obesity • Measles	0	0
	Top10-DLM-BERT-base-uncased	• Bats • bats • Rotaviruses	0	0
	Top10-BM25-BERT-base-uncased	• Rhinovirus • HCoV-OC43 • HCoV-229E	1	1
	Top10-SDM-BERT-base-uncased (ours)	• Rhinovirus • HCoV-OC43 • HCoV-OC43	1	1
2.	Question	What is a natural reservoir of coronavirus?		
	Answer (Ground Truth)	Bats		
	Method	**3 Answer**	**EM**	**F1 score**
	Top10-TF-IDF+Cosine-Bert Based Uncased	• Host Cell Receptor GRP78 • largemouth bass Micropterus salmoides • These reservoirs are located in east Texas	0	0
	Top10-DLM-Bert Based Uncased	• bats • sooty mangabeys • Wild rodents	1	1
	Top10-BM25-Bert Based Uncased	• bats • sooty mangabeys • Chinese horseshoe bats are considered to be the natural reservoirs of SARS-CoV. MERS-CoV	1	1
	Top10-SDM-Bert Based Uncased (ours)	• bats • SARS-CoV • one or more of them may serve as the natural reservoir of SARS-CoV and/or its progenitor virus	1	1

TABLE V. EVALUATION ON TOP N DOCUMENT VARIATION IN COVID-QA

Model	EM (%)	F1 Score (%)
Top5-SDM-BERT-base-uncased	7,83	22,62
Top10-SDM-BERT-base-uncased	8,67	23,42
Top20-SDM-BERT-base-uncased	**9,51**	**24,78**

2) Passage length variation: In Table III, IV and V, we set passage length to eight sentence in every passage based on Alzubi et al. [8] that used eight sentence too. Then, we think about making a deep analysis with some variation of passage length.

Table VI shows us that our CDQA system will get the bad result if the passage length sets under five sentence: one sentence (7,92 for EM and 21,10 for F1-score), three sentence (8,77 for EM and 23,40 for F1-score). It also get the bad result if the passage length sets upper five sentence: eight sentence (8,67 for EM and 23,42 for F1-score), 10 sentence (7,63 for EM and 22,16 for F1-score). So, we can see from Table V that the best passage length is five sentence with 8,77 EM and 23,45 F1-score.

3) Reader variation: We also vary the reader with some models. They are BERT base uncased, BERT large uncased, and BERT large uncased word whole masking in Table VII.

Table VII shows us that the reader with BERT large uncased gets the best result among others. It improves BERT based uncased by 9,68% in EM and 2% in F-1 score. It also improves BERT large uncased whole word masking by 5,55% in EM and 1,01% in F-1 score.

From the above variation, we conclude that the optimal parameters in our CDQA system is retriever with Sequential Dependence Model that uses the top 20 document when retrieved, having passage length five sentence, and having BERT Large Uncased for the reader. Because of that, we try to set all the retriever methods with the optimal parameter in Table VIII.

Table VIII shows us that by optimal parameter in all of the retrievers (top 20 documents as the retriever results, five sentences as the passage length, and BERT-large as the reader model) our CDQA system outperform the baselines by 37,28% in EM and 24,94% in F-1 score for TF-IDF+Cosine similiarity retriever, 5,59% in EM and 4,67% in F-1 score for DLM retriever, and 5,59% in EM and 5,06% in F-1 score for BM25 retriever.

Our CDQA system with SDM retriever with the optimal parameter achieves 10,20 for EM and 24,90 for F1-score. This value is better than SDM retriever in another parameter setting before in Table III, V, VI, and VII.

C. Transfer Learning Versus Non-transfer Learning

We also think that what is the effect of using transfer learning, with the reader retriever (SDM) method for building closed domain question answering, compared to non-transfer learning method. To answer this, we do some work. Different with experiment in Table II, III, V, VI, VII, and VIII that use all COVID-QA for evaluation part. In this experiment, COVID-QA will be split into 70% for fine-tuning and 30% for evaluation. In this experiment, the configuration in the model with the optimal parameter in our closed domain question answering from Table VIII (Top20-SDM-BERT Large Uncased-5 Sentence) will be used. We compare how if the CDQA system use a transfer learning method where the question answering system is fine-tuned with another dataset, such as SQuAD. Then, how if the CDQA system is fine-tuned again with COVID-QA after fine-tuning with SQuAD, and how if the CDQA system is fine-tuned without transfer learning method, that is fine-tuned to COVID-QA only. Table IX will show some experiments about comparing the CDQA system between transfer learning and non-transfer learning.

TABLE VI. EVALUATION ON PASSAGE LENGTH VARIATION IN COVID-QA

Model	EM (%)	F1 Score (%)
Top10-SDM-BERT-base-uncased-1 sentence	7,92	21,10
Top10-SDM-BERT-base-uncased-3 sentence	8,77	23,40
Top10-SDM-BERT-base-uncased-5 sentence	**8,77**	**23,45**
Top10-SDM-BERT-base-uncased-8 sentence	8,67	23,42
Top10-SDM-BERT-base-uncased-10 sentence	7,63	22,16

TABLE VII. EVALUATION ON READER VARIATION IN COVID-QA

Model	EM (%)	F1 Score (%)
Top10-SDM-BERT-base-uncased	8.67	23.42
Top10-SDM-BERT Large Uncased Word Whole Masking	9,01	23,65
Top10-SDM-BERT Large Uncased	**9,51**	**23,89**

TABLE VIII. EVALUATION ON THE OPTIMAL PARAMETER IN COVID-QA

Model	EM (%)	F1 Score (%)
Top20-TF-IDF+Cosine-BERT Large Uncased-5Sentence	7,43	19,93
Top20-DLM-BERT Large Uncased-5Sentence	9,66	23,79
Top20-BM25-BERT Large Uncased-5Sentence	9,66	23,70
Top20-SDM-BERT Large Uncased-5Sentence	**10,20**	**24,90**

TABLE IX. EVALUATION ON TRANSFER LEARNING AND NON-TRANSFER LEARNING METHOD IN 30% COVID-QA

Model	EM (%)	F1 Score (%)
Top 20-SDM-BERT Large Uncased-5 Sentence-Fine-tuned to CovidQA (Non-transfer learning)	0,0	3,20
Top 20-SDM-BERT Large Uncased-5 Sentence-Fine-tuned to SQuAD (transfer learning)	6,44	23.29
Top 20-SDM-BERT Large Uncased-5 Sentence-Double Fine-tuned(SQuAD+COVID-QA) (transfer learning fine-tuned to COVID-QA)	**10,07**	**29,64**

Table IX shows us that models that use transfer learning are better than models that use non-transfer learning method. Models that use transfer learning try to mitigate question answering system that has low resource question-answering dataset by transferring information from auxiliary dataset to improve their performance. It is especially useful for low-resource tasks such as our closed domain question answering in COVID-19. Models that use non-transfer learning, where the reader is fine-tuned to COVID-QA, achieve bad results. It gets 0.0 in EM and 3,20 in F1-score. It looks like this model becomes underfitting because the question-answer pair dataset in COVID-QA train is too small. All of us know that SQuAD has 87599 question-answer pairs, while COVID-QA has only 2019 question-answer pair. Then for fine-tuning, 1413 question-answer pair has been used. The amount of the dataset makes the model unable to learn. If COVID-QA dataset as big as SQuAD may be the model will be better.

Because there is no dataset question-answer pair in CORD-19, our work tries to use transfer learning method. Table IX shows two model that use transfer learning. First is a model that reader is fine-tuned to SQuAD only. The second is a model that reader is fine-tuned to SQuAD and fine-tuned again with COVID-QA. We do the double fine-tuned based on Yang et al. [32]. They said that fine-tuning again question answering system will make improvement to the model.

Table IX shows us that our system that use transfer learning (fine-tune to SQuAD) improves the effectiveness of non-transfer learning method with six times higher accuracy than the baseline method without using transfer learning. Further fine-tuning again the reader model that has fine-tuned before with SQuAD into closed domain dataset (COVID-QA) can increase the accuracy of the model by 27,26 % in F-1 score and improve F-1 score of non-transfer learning method with eight times higher. From this result, we can say that transfer learning method will get better result than non-transfer learning when resource of dataset question-answer pair itself is small.

VI. Discussion

Our experimental results show that transfer learning can give a significant improvement to the CDQA system that does not use transfer learning. The transfer learning model gains eight times higher accuracy than the method without using transfer learning. When the model is further fine-tuned with close domain dataset CDQA, the accuracy increased by 27,26%. This accuracy can further increase when a bigger CDQA dataset is available for fine-tuning. Therefore, creating a bigger CDQA dataset will become a research challenge in this case.

This research also has some limitation in terms of the variation of parameters explored in the experiment. For example, for the number of documents retrieved by retriever, we only experimented until 20 documents retrieved for each question. Then, for the length of passages, we only experimented until 10-sentences length for each passage. This restriction is related with the availability of GPU computing resource that is limited, while our model in general requires high computing resource. From this reason, we limit the variation of parameters in our experiment when fine-tuning the reader component. When a higher GPU is available, it may worth experimenting with more parameter values in our system. We can investigate, for example, whether retrieving more than 20 documents may further increase the performance of CDQA system.

VII. Conclusion

This paper explores the use of transfer learning technique to build close-domain question answering (CDQA) system in COVID-19. We propose to use Sequential Dependence Model (SDM) as retriever in the retriever-reader architecture to improve the accuracy of CDQA system. Our experimental results show that the use of SDM as retriever leads to significant performance of the CDQA system. Our model using SDM as the retriever and BERT-base that was fine-tuned on SQuAD benchmark as the reader can outperform F-1 score of the state-of-the-art baseline by 3,26% for BM25, 3,26% for DLM and 32, 62% for TF-IDF+Cosine similiarity.

The hyperparameter tuning on top n documents, passage length, and reader variations are shown to affect the performance of our CDQA model. We found that the best parameter is achieved by using top 20 documents as the retriever results, five sentences as the passage length, and BERT-large as the reader model. Our CDQA system using this parameter setting results in the best-performing model that achieves higher outperforming in F-1 score to the state-of-the-art baseline by 24,94% for TF-IDF+Cosine similiarity retriever, 4,67% for DLM retriever, and 5,06% for BM25 retriever.

Our results also confirm the merit of using transfer learning to build CDQA system in a condition where a large benchmark for CDQA is unavailable. Our CDQA system using transfer learning technique (i.e., using SDM retriever and BERT-large reader that was fine-tuned on SQuAD benchmark) is significantly more effective than the method that does not use transfer learning. Our best-performing model (double fine-tuning SQuAD and COVID-QA) is shown to gain eight times higher accuracy than the baseline method without using transfer learning. Further fine-tuning the transfer learning model using closed domain dataset can increase the accuracy of the transfer learning model that only fine-tuning with SQuAD by 27,26%.

VIII. Future Work

For future works, we consider developing closed domain question answering with Dense Passage Retrieval to get better result than this work. We also want to modify how if all documents have split into passages before sending it to the retriever.

Acknowledgment

This research was funded by the Directorate of Research and Development, Universitas Indonesia, under Hibah PUTI Pascasarjana 2022 (Grant No: NKB-03/UN2.RST/HKP. 05.00/2022).

References

[1] T. Acter, N. Uddin, J. Das, A. Akhter, T. R. Choudhury, and S. Kim, “Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency,” Science of The Total Environment, vol. 730, p. 138996, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138996.

[2] C. Huang et al., “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,” The Lancet, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[3] N. Zhu et al., “A Novel Coronavirus from Patients with Pneumonia in China, 2019,” N Engl J Med, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[4] P. Zhou et al., “Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin,” Microbiology, preprint, Jan. 2020. doi: 10.1101/2020.01.22.914952.

[5] L. L. Wang et al., “CORD-19: The COVID-19 Open Research Dataset,” 2020, doi: 10.48550/ARXIV.2004.10706.

[6] T. Möller, A. Reina, R. Jayakumar, and M. Pietsch, “COVID-QA: A Question Answering Dataset for COVID-19,” in Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, Online, Jul. 2020. Accessed: Nov. 24, 2022. [Online]. Available: https://aclanthology.org/2020.nlpcovid19-acl.18

[7] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “SQuAD: 100,000+ Questions for Machine Comprehension of Text,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 2016, pp. 2383–2392. doi: 10.18653/v1/D16-1264.

[8] J. A. Alzubi, R. Jain, A. Singh, P. Parwekar, and M. Gupta, “COBERT: COVID-19 Question Answering System Using BERT,” Arab J Sci Eng, Jun. 2021, doi: 10.1007/s13369-021-05810-5.

[9] W. Yang et al., “End-to-End Open-Domain Question Answering with,” in Proceedings of the 2019 Conference of the North, Minneapolis, Minnesota, 2019, pp. 72–77. doi: 10.18653/v1/N19-4013.

[10] D. Metzler and W. B. Croft, “A Markov random field model for term dependencies,” in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’05, Salvador, Brazil, 2005, p. 472. doi: 10.1145/1076034.1076115.

[11] E. Yulianti, R.-C. Chen, F. Scholer, W. B. Croft, and M. Sanderson, “Ranking Documents by Answer-Passage Quality,” in The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor MI USA, Jun. 2018, pp. 335–344. doi: 10.1145/3209978.3210028.

[12] E. Yulianti, R.-C. Chen, F. Scholer, and M. Sanderson, “Using Semantic and Context Features for Answer Summary Extraction,” in Proceedings of the 21st Australasian Document Computing Symposium, Caulfield VIC Australia, Dec. 2016, pp. 81–84. doi: 10.1145/3015022.3015031.

[13] D. Chen, A. Fisch, J. Weston, and A. Bordes, “Reading Wikipedia to Answer Open-Domain Questions,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada, Jul. 2017, pp. 1870–1879. doi: 10.18653/v1/P17-1171.

[14] Y. Yang, W. Yih, and C. Meek, “WikiQA: A Challenge Dataset for Open-Domain Question Answering,” in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015, pp. 2013–2018. doi: 10.18653/v1/D15-1237.

[15] S. Reddy, D. Chen, and C. D. Manning, “CoQA: A Conversational Question Answering Challenge,” 2018, doi: 10.48550/ARXIV.1808.07042.

[16] K. M. Hermann et al., “Teaching Machines to Read and Comprehend,” 2015, doi: 10.48550/ARXIV.1506.03340.

[17] L. F. Simanjuntak, R. Mahendra, and E. Yulianti, “We Know You Are Living in Bali: Location Prediction of Twitter Users Using BERT Language Model,” BDCC, vol. 6, no. 3, p. 77, Jul. 2022, doi: 10.3390/bdcc6030077.

[18] E. Yulianti, A. Kurnia, M. Adriani, and Y. S. Duto, “Normalisation of Indonesian-English Code-Mixed Text and its Effect on Emotion Classification,” IJACSA, vol. 12, no. 11, 2021, doi: 10.14569/IJACSA.2021.0121177.

[19] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research 12 (2011), pp. 2825–2830.

[20] G. Amati, “BM25,” in Encyclopedia of Database Systems, L. Liu and M. T. Özsu, Eds. Boston, MA: Springer US, 2009, pp. 257–260. doi: 10.1007/978-0-387-39940-9_921.

[21] R. Nogueira and K. Cho, “Passage Re-ranking with BERT.” arXiv, Apr. 14, 2020. Accessed: Nov. 24, 2022. [Online]. Available: http://arxiv.org/abs/1901.04085

[22] S. J. Semnani, M. Pandey, and M. Pandey, “Domain-Specific Question Answering at Scale for Conversational Systems,” 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, p. 10.

[23] Akdemir, “Research on Task Discovery for Transfer Learning in Deep Neural Networks,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Online, 2020, pp. 33–41. doi: 10.18653/v1/2020.acl-srw.6.

[24] Akdemir and T. Shibuya, “Transfer Learning for Biomedical Question Answering,” CLEF 2020, 22-25 September 2020, Thessaloniki, Greece, vol. 2696, p. 15.

[25] Z. H. Syed, A. Trabelsi, E. Helbert, V. Bailleau, and C. Muths, “Question Answering Chatbot for Troubleshooting Queries based on Transfer Learning,” Procedia Computer Science, vol. 192, pp. 941–950, 2021, doi: 10.1016/j.procs.2021.08.097.

[26] Macdonald, N. Tonellotto, S. MacAvaney, and I. Ounis, “PyTerrier: Declarative Experimentation in Python from BM25 to Dense Retrieval,” in Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Virtual Event Queensland Australia, Oct. 2021, pp. 4526–4533. doi: 10.1145/3459637.3482013.

[27] S. Huston and W. B. Croft, “A Comparison of Retrieval Models using Term Dependencies,” in Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai China, Nov. 2014, pp. 111–120. doi: 10.1145/2661829.2661894.

[28] Z. J. Zhang et al., “A generic retrieval system for biomedical literatures: USTB at BioASQ2015 Question Answering Task,” p. 7.

[29] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2018, doi: 10.48550/ARXIV.1810.04805.

[30] Zhai and J. Lafferty, “A study of smoothing methods for language models applied to Ad Hoc information retrieval,” in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, Sep. 2001, pp. 334–342. doi: 10.1145/383952.384019.

[31] J. M. Ponte and W. B. Croft, “A language modeling approach to information retrieval,” in Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’98, Melbourne, Australia, 1998, pp. 275–281. doi: 10.1145/290941.291008.

[32] W. Yang, Y. Xie, L. Tan, K. Xiong, M. Li, and J. Lin, “Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering.” arXiv, Apr. 14, 2019. Accessed: Nov. 24, 2022. [Online]. Available: http://arxiv.org/abs/1904.06652

A Cascaded Feature Extraction for Diagnosis of Ovarian Cancer in CT Images

Arathi B[1]*, Shanthini A[2]
Research Scholar, Department of Data Science and Business Systems[1]
Associate Professor, Department of Data Science and Business Systems[2]
SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203[1, 2]

***Abstract*—This paper proposed ovarian cancer detection in the ovarian image using joint feature extraction and an efficient Net model. The noise of the input image is filtered by using Improved NLM (Improved Non-Local Means) filtering. The deep features are extracted using Deep CNN_RSO (Deep Convolutional Neural Network Rat Swarm Optimization), and the low-level texture features are extracted using ILBP (Interpolated Local Binary Pattern or Interpolated LBP). To improve the feature extraction and reduce the error, use a cascading technique for the feature extraction. RSO also helps to efficiently optimize the DCNN features from the images. Finally, the extracted image is classified using the Efficient Net classifier, which performs a global average summary and classification of ovarian cancer (normal and abnormal). The system's performance is implemented on the Cancer Genome Atlas Ovarian Cancer (TCGA-OV) dataset. The system's performance, like sensitivity, specificity, accuracy and error rates, shows better with respect to other techniques.**

Keywords—Ovarian cancer; deep convolutional neural network rat swarm optimization; CT image; joint feature; efficient net; improved non-local means; interpolated local binary pattern

I. Introduction

Ovarian cancer (OC) a disease that affects women's ovaries, is difficult to diagnose early, and has a high fatality rate [1]. OC has the highest mortality rate and is the leading cause of gynaecological deaths. There are two primary forms of invasive ovarian cancer, often referred to as Type-I and Type-II tumours [2]. Low-grade endometrial carcinomas, Low-grade serous carcinomas, clear-cell carcinomas, mucinous carcinomas, and malignant Brenner (transitional) tumours are examples of early-stage, nonaggressive type-I malignancies [3] [4]. Seventy-five percent of epithelial ovarian malignancies are type II tumours, which include high-grade serous carcinomas, high-grade endometriosis carcinomas, undifferentiated carcinomas, and malignant mixed mesodermal tumours. CT scans are routinely used to diagnose ovarian tumours, detect metastases, determine the stage of ovarian cancer, monitor patients following surgery, and evaluate the success of treatment [5].

Some ovarian epithelial tumours do not seem to be malignant when examined in the lab, and they are referred to as borderline epithelial ovarian cancer. Screening tests and exams identify diseases in people with no symptoms, such as cancer [6]. The abdomen and pelvis CT scan is the first-line imaging modality for ovarian cancer staging, therapy options selection and disease response assessment. The imaging surrogate for surgical-pathological FIGO staging is the staging CT, which gives disease distribution and load. This is mostly due to late detection, although OC often recurs after treatment. The remarkable difference in cure between patients with close illness and those with distant disease (15% vs. 25%) eloquently illustrates the need for a non-invasive but feasible diagnostic applicable to a large population expected to be afflicted by OC in its early stages [7].

*Corresponding Author.

Regardless of the fact that OC is only the 6th most frequent female cancer in Singapore, it is still a leading cause of gynaecological cancer mortality. As a result, the risk of cancer increases throughout life [8], leading to some complications, such as the inability of women with OC to have children and a problem after menopause. Therefore, to overcome these problems for women, accurate identification, solid cancer analyses [9], and the treatment procedure need all be improved. Despite all of these efforts, early detection of OC continues as a critical concern [10, 11]. With the use of proteomic analysis in serum, there has been continuing study to detect new cancer indicators. The cause or causative factor for OC is unknown, making early detection of ovarian cancer difficult [12].

The link between OC diagnosis and survival led to thinking and efforts to improve the outcomes obtained at the early recognition phases. This information is used to make better informed decisions while saving money [13, 14]. There are five reasons why a big data strategy can help cancer detection. Pathologists also noticed and examined the numerous gene articulations in this specific OC as well as the amounts of gene articulating in distinct neurotic phases and situations [15], as part of the same inquiry. Both organic and statistical computing may be used to characterize gene articulations for such relationships [16], especially in dynamic conditions when the tumour progresses from the quantitative processing of OC tissue [17, 18]. Different X-ray absorption coefficients in the human body indicate tissue properties in CT scans. CT provides the advantages of quick scan time and accurate localization [19].

The number of lesions and the simplicity with which they may be imaged are more important than the expense of the imaging technique. Radiation is a type of energy that may injure the human body. The resolution of CT scans is very important [20]. According to the statistics and machine learning part, various classification approaches were utilized in the cancer classification process, although it has certain non-trivial job problems. The knowledge about gene articulation is unlike anything else that these approaches have ever dealt with. The supervised learning method uses the training set of categorized data to create a classifier that can be used to classify new data.

1) Motivation: The recognition of CT images is difficult to identify using segmentation or clustering. It may have a greater impact on the FAR and FRR rate at the moment of identification from the knowledge base. Existing systems have a functioning accuracy level of 90 to 95. It may be improved so that the system classifies more precisely and efficiently. Previous techniques were unable to classify the cancer stage. The assessment of the patient's stage aids in the more efficient treatment of the patient. The presented issues of the system motivated to development of a new methodology known as joint feature extraction with an efficient net classifier for ovarian cancer detection.

2) Contribution

- To improve the feature extraction of the CT image, Joint feature extraction is introduced based on the fusion of high-level (deep) and low-level (shallow) features are introduced for performing better image feature extraction.
- The deep features are extracted using Deep CNN_RSO, and the low-level texture features are extracted using ILBP. Both techniques are used to reduce the error rate in feature extraction.
- The efficient Net classifier is used to categorize the ovarian CT image by minimizing the greatest number of features in the data without sacrificing its original properties and decreasing the processing time and memory space.

The structure of the paper: Section II gives the related works of ovarian cancer detection in the CT images, the proposed methodology describes in Section III, Section IV explains about the result and discussion, and Section V concludes the paper.

II. Related Work

Lu et al. [21] presented a technology based on learned and created a non-invasive reduction in the quantity of the main ovarian tumour. Despite best efforts, the five-year survival rate for epithelial ovarian cancer (EOC) is around 35–40 per cent, underlining the use of classification markers for tailored treatment. Here, extracted 657 quantifiable statistical characteristics from the data in this section. 364 EOC patients' pre-operative CT scans at their initial presentation. The "Radiomic Prognostic Vector" is made up of four descriptions (RPV). RPV has a high level of accuracy when it comes to identifying people. The 5% of people with a median survival rate of fewer than two years improves dramatically. Based on well-established prognostic methodologies and has been tested in two separate multi-centre cohorts. In addition, two separate datasets were analyzed genetically, proteomic and transcriptomic. The activation of the stromal phenotype and DNA damage response mechanisms are included.

Lago et al. [22] proposed the Sentinel lymph node technique for the early detection of ovarian cancer (SENTOV). Ovarian cancer in an early stage may be the optimal disease setting for lymph node monitoring. Nonetheless, it appears that the documented experience is limited. The purpose of this study was to evaluate the feasibility and safety of sentinel lymph node biopsy in patients with clinical stages I and II ovarian cancer. Twenty patients in total are being monitored. Sentinel lymph nodes were seen in 14/15 individuals. (93%) pelvic region and each of the twenty para-aortic regions (100%). Due to this, five individuals did not receive an utero-ovarian injection when the average time between injection and beginning of symptoms was reached. The resection of sentinel lymph nodes required 53–15 minutes (range; the average number of sentinel lymphocytes recovered was 30–80). There were 2, 21.50 lymph nodes in the pelvic region (range: 0–5).

Wang et al. [23] suggested non-invasive recurrence forecasting models in HGSOC that aggregate prognostic characteristics from pre-operative computed tomography (CT) images utilising a unique deep learning (DL) technique. Equipment and procedures include 245 HGSOC patients from two hospitals, including a feature-learning cohort (n = 102) and two independent validation cohorts (n = 49 and n = 45). To capture the prognostic biomarkers (DL feature) of HGSOC, 8917 CT images from the feature learning cohort were used to train a unique DL network. A DL-CPH model including the DL feature and Cox proportional hazard (Cox-PH) analysis was created to predict patient recurrence risk and three-year recurrence probabilities.

Castellani et al. [24] developed a standard pre-surgical staging for the identification of ovarian cancer. The most important prognostic factor in ovarian cancer is the stage of the disease at the time of diagnosis. The current staging system (FIGO classification) is based on clinical and histological results. Despite the fact that the clinical stage is the standard method in ovarian cancer, the first patient care relies on an imaging-based which was before staging evaluation to detect disease that is unrespectable or difficult to resect. Radiologists must know the advantages and disadvantages of the many imaging techniques available. For effective staging and treatment planning, a clear understanding of the path of disease dissemination and review regions is essential. CT scans of the chest, abdomen, and pelvis are the current standard of care for pre-surgical staging. This enables a quick assessment of the disease's severity and is cost-effective.

Funston et al. [25] described CA125, an excellent screening test for ovarian cancer in primary care, particularly for women under the age of 50. To avoid diagnostic delays, patients with increased CA125 values should be evaluated for non-ovarian malignancies, especially if ovarian cancer has been ruled out. Our findings allow physicians and patients to predict the risk of ovarian cancer and other malignancies at any CA125 level and age, which can be used to guide individual decisions about the need for additional testing or referral.

Nougaret et al. [26] proposed radio genomics and radionics computational ways to measure tumour heterogeneity to investigate the entire tumour heterogeneity rather than a single biopsy sample. Cancer heterogeneity has been identified at the genetic and histological stages of ovarian cancer and has been linked to poor health outcomes. In advanced ovarian cancer with peritoneal carcinomatosis, traditional magnetic resonance or computed tomography imaging methods do not allow for intra- or inter-tumor heterogeneity. Feature extraction, which involves extracting various variables from pictures and is an element of radionics', has subsequently been suggested to assess progressed ovarian tumour heterogeneity. This brief overview covers the fundamentals of radionics, how to do texture analysis, and how to apply it to ovarian cancer imaging.

Forstner et al. [27] propose a Standardization of imaging techniques for ovarian cancer detection. Early detection is the only way to achieve a high cure rate in females with ovarian cancer. Despite rapidly evolving biomarkers, there is still no viable technique for early detection. Ovarian cancer has a low prevalence, a low specificity, and a high sensitivity. False positive percentages have been one of the screening programmes' drawbacks. Transvaginal sonography and magnetic resonance imaging (MRI) are useful methods for identifying ovarian masses in the hands of professionals. Efforts to standardize procedure and analysis are now underway and are expected to continue. Improve clinical diagnostic capabilities and the usage of cancer risk prediction algorithms. Radiogenomics and Radiomics could provide a plethora of extra information in the diagnosis and treatment of ovarian cancer.

These studies about ovarian cancer detection in CT images show that many issues exist in the detection of ovarian cancer in the CT image, such as increased noise in feature extraction, classification error, low specificity and low sensitivity. To improve the method's performance and reduce the error, here propose a new methodology named ovarian cancer detection in CT image by joint feature extraction and Efficient Net classifier.

III. Proposed Methodology

Ovarian cancer is a type of cancer that affects the ovaries. The cells can enter and harm healthy biological tissue and grow rapidly. Most techniques failed to identify cancer in the CT image due to the error and noise in the output image. Here, the proposed ovarian cancer detection by joint feature extraction and efficient net classifier. In this case, the pre-processing technique is initially done by using Improved NLM filtering to remove unwanted noise in the image. Next, Joint feature extraction is done based on the fusion of high-level and low-level features. Here, the deep features are extracted using Deep CNN_RSO and the low-level texture features are extracted using ILBP. Both the features are cascaded to obtain joint feature extraction for performing ovarian image classification. Finally, the Efficient Net classifier performs a global average pooling and classification of normal and abnormal images. Thus, the suggested Efficient Net classifier can lower the maximum number of data features without altering their original properties and cut processing time and memory space. The performance of the presented approach will be examined with recent existing approaches in terms of performance metrics to prove the performance efficiency. The architecture of the proposed method is shown in Fig. 1.

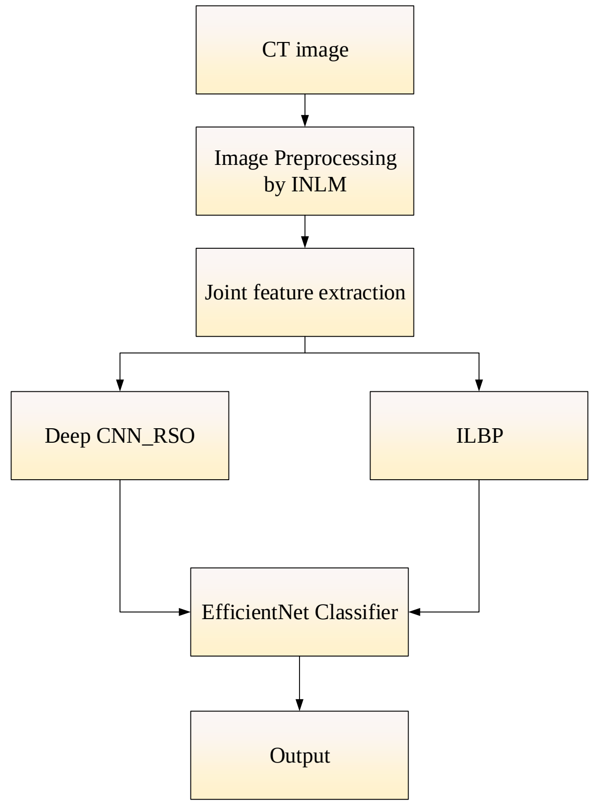


Fig. 1. Architecture of the proposed system.

A. Pre-processing

Image pre-processing is the preparation of images prior to their use in model training and inference. This covers orienting, resizing, and colour adjustments of the CT image. After collecting the CT images of ovarian cancer, upload the dataset picture and transform the RGB image into a grayscale image. Here pre-processing is done by using Improved Non-Local Means (INLM). According to the NLM, the weights are produced under the Gaussian noise assumptions that are used to quantify the closeness between a core area patch in a searching window and nearby patches. Only pixel data is used to construct the weights for the NLM, which quantifies the closeness between a centre area patch and its nearby patches. The variance is used to calculate the neighbourhood similarity measure. This similarity metric does not completely consider the structural data of the picture whenever the noise levels are too high in the case of NLM. There are errors generated in the case of noise in the CT images, so an improved NLM is used for the smooth region to achieve a better effect in the image, to provide good protection for the edge and the adaptive selection of the Gaussian weighting coefficients. The algorithm for INLM is shown below.

$$S = \frac{1}{\sqrt{3P}} \sum_{p=0}^{p=1} \left\| f_k(x,y) - g_k \right\|_2 \quad f_k(x,y) \in Patchwindow \tag{1}$$

$$w(u,v) = \exp\left(-\frac{\sum_{p\in P}\left\| f_k(x+p, y+p) - f_k(x_0+p, y_0+p)\right\|_2^2}{h^2}\right) \quad (2)$$

$$w(x,y,x_0,y_0) = \exp\left(-\frac{\sum_{p\in P}\left\| f_k(x+p, y+p) - f_k(x_0+p, y_0+p)\right\|_2^2}{h}\right) \quad (3)$$

$$w_1(x,y,x_0,y_0)w_{21}(x,y,x_0,y_0)$$

$$g_k(x,y) = \frac{\sum_{u,v\in N} f_k(x+u, y+v)w(u,v)}{\sum_{u,v\in N} w(u,v)} \quad (4)$$

S Shows the general expression for the INLM algorithm, w represents the weight of the CT image and $g_k(x,y)$ shows the grey level representation of the image. The detailed algorithm of the INLM is shown below.

B. Joint Feature Extraction

Fusion of high-level (deep) and low-level (shallow) features is used to do joint feature extraction. Here, Deep CNN RSO is used to extract the deep features, whereas ILBP is employed to extract the low-level texture features. Both the features are cascaded to obtain joint feature extraction for performing ovarian image classification. The architecture of joint feature extraction is shown in Fig. 2.

1) Deep CNN_RSO for deep feature extraction: The high level (deep) features of the ovarian images are extracted by using the CNN (Convolutional neural network). The extracted features are then optimized using the Rat Swarm Optimization (RSO). Fig. 3 shows the DCNN's architecture. In this work, the DCNN consists of five convolutional layers: three max pooling layers and two complete reconnect layers. The activation function of each layer was a Rectified Linear Unit (ReLU). Three max pooling layers with a size of 3*3 pixels and a stride of 2 were employed to reduce the input image size of the subsequent convolutional layer. At the end of the DCNN, two fully linked layers containing a significant number of neurons were applied. The network extracts the deep features, is then optimized.

a) Rat Swarm Optimization: Rats are enormous, medium-sized rodents with long, variable-length tails. Black and brown rats are the two most common species of rats [28]. Male rats are referred to as bucks, while female rats are referred to as does. Rats are naturally smart social beings. They indulge in grooming and jumping, chasing, tumbling, and punching. Rats are territorial and inhabit colonies composed of both males and females. Rats are noted for their aggressive behaviour, which has been known to kill other creatures. This aggressive behaviour is the fundamental motivation for performing this action when pursuing and combating prey. The mathematical expression for determining the optimization value of the deep features of ovarian cancer from a CT picture using RSO is provided below.

The feature of the CT image can be found by,

$$\vec{F} = A.\vec{F}_i(x) + C(.\vec{F}_r(x) - \vec{F}_i(x)) \quad (5)$$

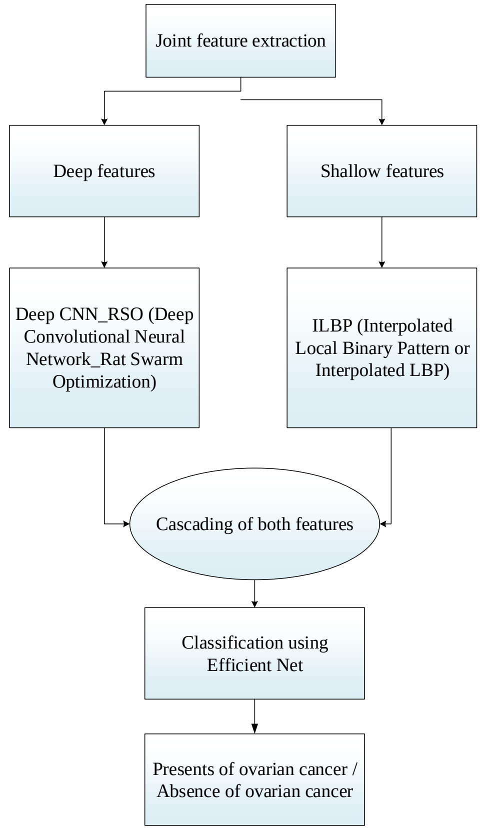


Fig. 2. Joint feature extraction.

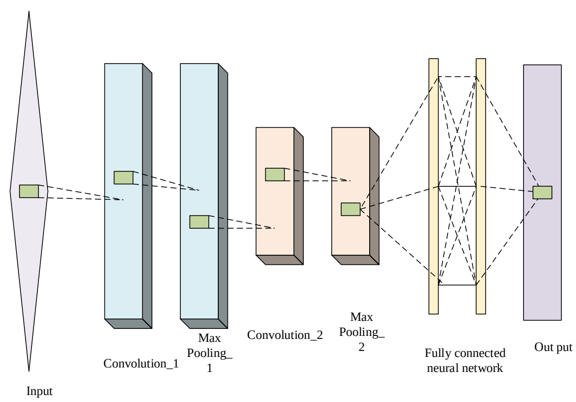


Fig. 3. DCNN's architecture.

Where $\vec{F}$ shows the deep features, $\vec{F}_i(x)$ represents the features present in the CT image, $\vec{F}_r(x)$ shows the best optimal feature, A and C shows the parameters. They can be calculated as,

$$A = R - x \times \left(\frac{R}{M_{itr}}\right) \quad (6)$$

$$x = 0,1,2,\ldots M_{itr} \quad (7)$$

$$C = 2 \cdot rand() \quad (8)$$

R is the random number between the [1, 5] and [0, 2]

The updated features can be obtained by,

$$\overrightarrow{F_{deep}} = \left| \overrightarrow{F_r}(x) - \overrightarrow{F} \right| \tag{9}$$

Here by using the Deep CNN_RSO the deep features of the CT images are extracted. Then, need to extract the shallow feature for that purpose ILBP is used.

2) Interpolated Local Binary Pattern (ILBP) for shallow feature extraction: LBP compares the grey levels of the centre pixel to the grey values of the eight neighbouring pixels in a 33 window and generates 1 or 0 to identify a pixel in the picture. It generates an eight-bit binary image as a result of the difference between grey values, with each grey value being split by its corresponding weight. Finally, these values are added together to produce a decimal value used to designate the centre pixel. Like LBP, the proposed ILBP characterizes each pixel depending on its neighbours. The only variable is that these pixels are chosen in a circular pattern and are evenly spaced in the center pixel to be labeled

Consider the position of the sampling point (X_n, Y_n) is represented as

$$x_n = x_c + R \times \cos\left(\frac{2\pi n}{N}\right),\ y_n = y_c - R \times \sin\left(\frac{2\pi n}{N}\right) \tag{10}$$

(x_c, y_c) Represents the pixel centre and $N(1 \le n \le N)$ shows the surrounding pixel number. N, R Represents the changing values of the CT image. The Generation of the circulation pattern is depicted in Fig. 4. Each phase of the circulation pattern is depicted in Fig. 4(a), (b), (c), and (d).

The grey value point $g(x, y)$ can be estimated as,

$$\begin{aligned} g(x,y) &= g(0,0)(1-x)(1-y) + g(1,0)\,x(1-y) \\ &+ g(0,1)(1-x)y + g(1,1)xy \end{aligned} \tag{11}$$

The weight of each pixel is given as

$$P(x,y) = w_1P_1 + w_2P_2 + w_3P_3 + w_4P_4 \tag{12}$$

$$w_1 = (1-x)(1-y) \tag{13}$$

$$w_2 = x(1-y) \tag{14}$$

$$w_3 = y(1-x) \tag{15}$$

$$w_4 = xy \tag{16}$$

Interpolated LBP can be found by

$$\overrightarrow{F_{shallow}} = \sum_{n=0}^{N-1} f(I_n - I_c)2^n \tag{17}$$

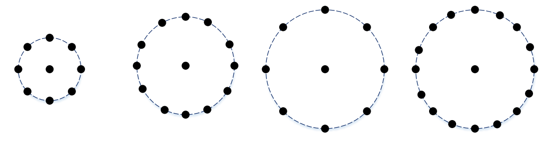


Fig. 4. Generation of circulation pattern.

$\overrightarrow{F_{shallow}}$ Represents the feature extracted by LBP, I_n and I_c Shows the grey value of the neighbouring pixel.

a) Joint feature: The low-level textural characteristics can characterise the edges, direction, and intensity distribution, while the high-level features retrieved from the medical CT picture can indicate the visual characteristics of the lesion's location. CNN RSO gives deep characteristics, whilst the ILBP provides shallow ones. Both CCN_RSO and ILBP output features are cascaded to form feature extraction output. The joint feature gives a better feature extraction output. The joint feature can be obtained by normalizing both the deep and shallow features. The deep feature of the picture is the network’s output, and the deep feature’s dimension is 1024. ILBP is utilised as the image's texture for the shallow texture feature. The ILBP element is 59 inches in length. The high and low level fusion features of an image are generated by normalising and cascading the two features.

$$\overrightarrow{F}_{shallow} = \frac{\overrightarrow{F}_{shallow} - \overrightarrow{F^{\min}}_{shallow}}{\overrightarrow{F^{\max}}_{shallow} - \overrightarrow{F^{\min}}_{shallow}} \tag{18}$$

$$\overrightarrow{F}_{deep} = \frac{\overrightarrow{F}_{deep} - \overrightarrow{F^{\min}}_{deep}}{\overrightarrow{F^{\max}}_{deep} - \overrightarrow{F^{\min}}_{deep}} \tag{19}$$

$$\overrightarrow{F_{eff}} = \left\{\overrightarrow{F_{deep}}, \overrightarrow{F_{shallow}}\right\} \tag{20}$$

$\overrightarrow{F}_{eff}$ represents the joint feature extraction of the deep and shallow features of the ovarian cancer images, $\overrightarrow{F_{shallow}}$ Represents the feature extracted by ILBP and $\overrightarrow{F_{deep}}$ Represents the deep feature extractions.

C. Efficient Net for Classification of CT Images

The classification of the image is used to calculate the neighbourhood similarity metric. This similarity metric does not completely include the structural data of the picture whenever the noise levels are too high. The classifier follows the classification process. Here uses a classifier named Efficient Net classifier. The output of the Efficient Net is given by the global averaged features of the CT image of the ovary. On the image Net, the Efficient Net classification model delivers state-of-the-art performance. The proposed classification of the CT image follows two stages: pre-trained and training stages. The pre-trained stage frees the image weights with respect to the network. The CT image is fully retained in the second stage with an efficient net classifier.

The Efficient Net family of CNN models is based on a revolutionary scaling method. It utilizes a simple compound coefficient that is quite effective. Efficient Net scales each dimension uniformly using a predefined set of scaling factors, unlike previous approaches that scale network parameters such as breadth, depth, and resolution. In practice, increasing subjectivity improves the model's performance, but balancing all network features under available resources considerably improves overall performance. Efficient Net is considerably more compact than other variants with the same Image Net accuracy.

This research applies the Efficient Net CNN model to provide an efficient approach. This form of Efficient Net was chosen because it strikes a reasonable balance between processing resources and precision. These strategies can likewise be used to more resilient variations. Mobile inverted bottleneck convolution (MBConv) is a fundamental component of the Efficient Net model family. The Mobile Net models inspired the development of MBConv. One of the key concepts is the use of depth-wise separable convolutions, which mix depth-wise and point-wise convolution layers in succession. Fig. 5 depicts an efficient architecture for Net classifiers.

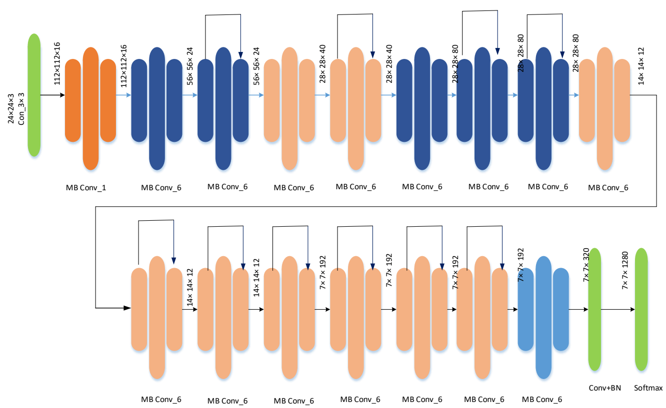


Fig. 5. Architecture of Efficient Net classifier.

The success of the previously proposed strategy for model scaling is strongly dependent on the baseline network. Searches automatically for a CNN model that maximises accuracy and efficiency (in FLOPS). The fundamental network is referred to as Efficient Net, and its essential layout is depicted in Fig. 5. The first thing to observe is that this bass line model is composed of recurring MBConv1, MBConv3, and MBConv6 blocks. These are the various sorts of MB Conv blocks. The second observation is that the number of channels per block has expanded or increased (through a bigger number of filters). The final observation relates to the inverted residual connections that exist between the thin layers of the model. Given the CT scans detection of ovarian cancer, the Efficient Net classifier output is displayed.

IV. Simulation Results

The ovarian cancer detection in the CT image using joint feature extraction and the Efficient Net model is presented using the PYTHON platform with the TCGA-OV dataset. Here the method is analyzed using sensitivity, specificity, accuracy, root mean square error (RMSE) and mean absolute percentage error (MAPE). Also, the ROC curve is obtained for the performance analysis of the system. For the classification of the features, an Efficient Net model is used. This model is analyzed by comparing the classification result using Le Net, Alex net, Google Net and VGG Net using the TCGA-OV dataset.

A. Dataset Description

1) TCGA-OV: TCIA provided a public database of abdominal CT scans used in this study: The Ovarian Cancer (TCGA-OV) dataset from the Cancer Genome Atlas. A group of researchers created the dataset. The Cancer Genome Atlas is a project examining the relationship between cancer characteristics and genetics by supplying clinical photographs. Clinical, genetic, and pathological data are stored in Genomic in TCGA-OV. TCIA stores radiological data, while the Data Commons Data Portal stores data. The TCGA-OV database contains 143 abdominal contrast-enhanced CT images and has two instances. The pelvis was outside the CT scan range, hence it was excluded from this study. The 141 remaining occurrences were included in the current investigation. The 141 samples were randomly divided into 71 training cases.

B. Performance Analysis

Here the performance of the system analyzes with respect to sensitivity, specificity, Accuracy, RMSE and MAPE. Sensitivity is the capacity to sense or respond to simulation. The equation for the sensitivity of the classifier is given as

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{21}$$

Specificity is defined as the condition of being specific in any method. Specificity of the classifier is represented as,

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{22}$$

The system's accuracy is defined as the state of being accurate in a process. The accuracy of the classifier system is given as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

The RMS value is the root mean square value, and the RMSE is the process's root mean square error.

$$RMSE = 1 - \sqrt{\frac{\sum_{i=1}^{m} (T_i - P_i)^2}{N}} \tag{24}$$

The MAPE can find it as

$$MAPE = 1 - \left(\frac{\sum_{i=1}^{m} (T_i - P_i)^2}{N} \right) \times 100 \tag{25}$$

Where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, $i = 0,1,2...,m,\ m = \max value$.

The result obtained by each of the performance metrics is shown below. Fig. 6 shows the sensitivity of the proposed Efficient Net model with respect to other models. Here the graph shows that the sensitivity of the proposed system is 98.8% when using Le Net is 97.5%, Alex Net is 97.4&, Google Net is 97.2%, and VGG Net is 97%. Through this, the sensitivity of the proposed system is higher than that of the other classification methods.

Fig. 7 shows the specificity of the proposed method with respect to the Efficient Net. The figure shows that the specificity of the proposed method is 99.8%, the specificity of the classification method by using Le Net is 98.8%, using Alex Net is 98.6%, using Google Net is 98.4% and using VGG Net is 98.3%. Through this, the specificity of the proposed method using Efficient Net is better than that of the other classification methods.

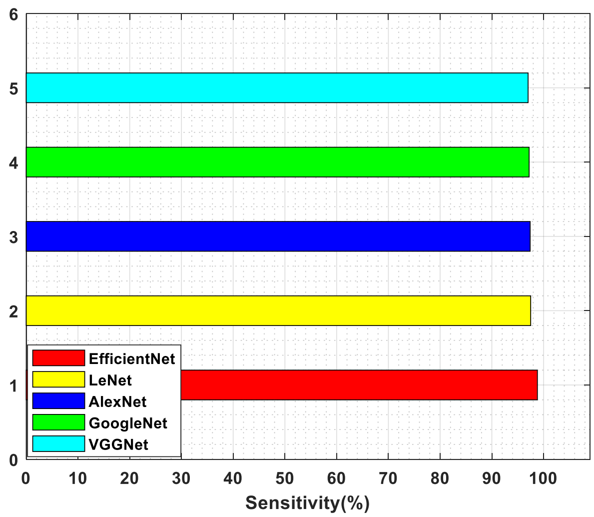


Fig. 6. Sensitivity of the proposed method.

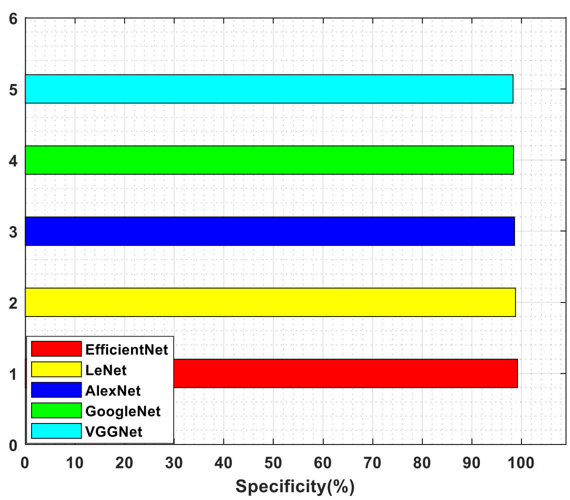


Fig. 7. Specificity of the proposed method.

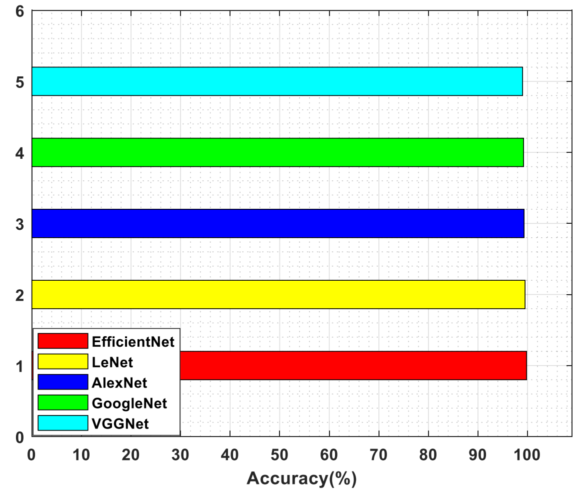


Fig. 8. Accuracy of the proposed method.

Fig. 8 shows how well the proposed method works in comparison to the Efficient Net. The graph shows that the proposed method is 99.8% accurate, the classification method is 99.5% accurate when using Le Net, 99.3% accurate when using Alex Net, 99.2% accurate when using Google Net, and 99% accurate when using VGG Net. Through this, the accuracy of the proposed method using Efficient Net is better than that of the other classification methods.

Fig. 9 shows how the proposed method compares to the Efficient Net in terms of the RMSE. The graph shows that the proposed method has an RMSE of 0.01%, the classification method using Le Net has an RMSE of 0.05%, Alex Net has an RMSE of 0.07%, Google Net has an RMSE of 0.09%, and VGG Net has an RMSE of 0.096%. Because of this, the RMSE of the proposed method using Efficient Net is better than that of the other classification methods.

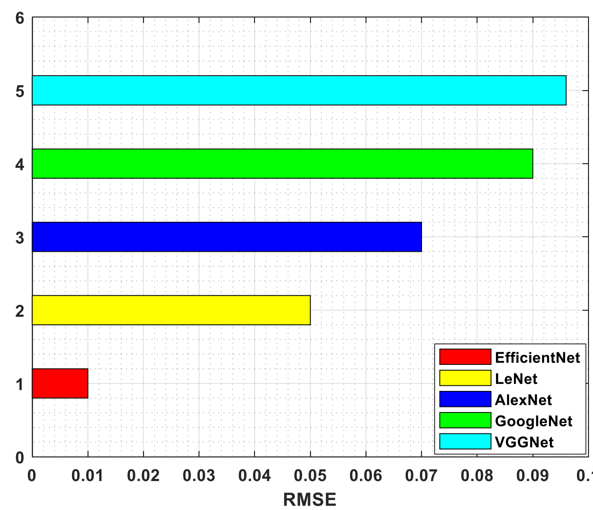


Fig. 9. RMSE of the proposed method.

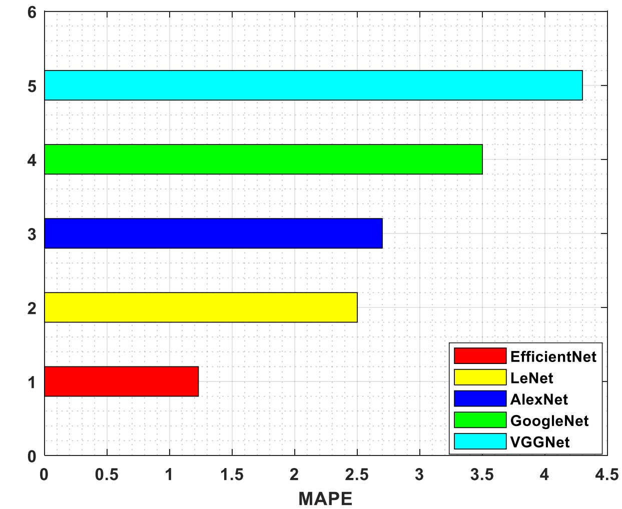


Fig. 10. MAPE of the proposed method.

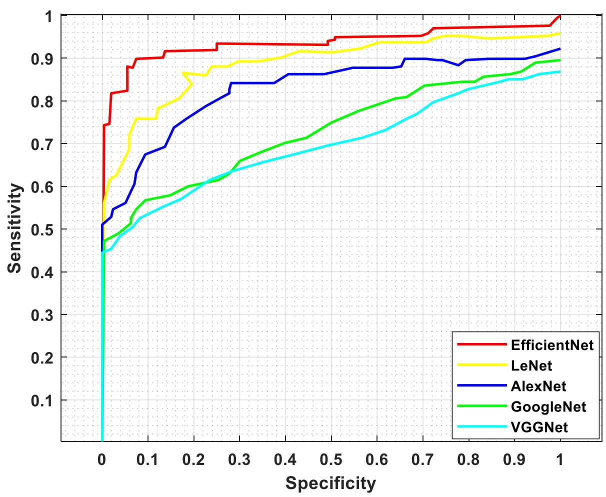


Fig. 11. ROC of the proposed method.

Fig. 10 demonstrates the MAPE of the suggested technique relative to the Efficient Net. The graph reveals that the MAPE of the suggested technique is 1.23 percent, while the MAPE of the classification method utilising Le Net is 2.5 percent, Alex Net is 2.7 percent, Google Net is 3.5 percent, and VGG Net is 4.33 percent. Consequently, the MAPE of the suggested method utilising Efficient Net is superior to that of the other categorization techniques. Fig. 11 illustrates the ROC of the proposed Efficient Net approach versus the Le Net, Alex Net, Google Net, and VGG Net. The curve demonstrates that the suggested Efficient Net model outperforms the alternatives.

V. Conclusion

This paper proposes ovarian cancer detection in CT images by joint feature extraction and an Efficient Net classifier. Here, the deep features are extracted using Deep CNN_RSO (Deep Convolutional Neural Network Rat Swarm Optimization) and the low-level texture features are extracted using ILBP (Interpolated Local Binary Pattern or Interpolated LBP). The cascaded feature extraction is used for the classification of the image. Also, the optimization gives better results to integrating with DCNN. The classification result shows better performance with minimum noise in the extracted images. The joint feature reduces the noise and improves the feature extraction efficiency. Finally, Efficient Net performs global average pooling and classification of ovarian cancer. By using the Efficient Net classifier, the sensitivity, specificity, accuracy, RMSE, and MAPE can be improved to 98.8%, 99.2%, 99.8%, 0.01% and 1.23%. The system's performance is compared using Le Net, Alex Net, Google Net and VGG Net using the TCGA-OV dataset. The outcome indicates that the performance of the proposed system is superior. Future work can improve the classifier's ability to classify tiny cysts in CT images.

References

[1] J. Kemppainen, J. Hynninen, J. Virtanen, M. Seppänen, "PET/CT for evaluation of ovarian cancer," InSeminars in nuclear medicine. Vol. 49, no. 6, pp. 484-492, 2019 Nov 1. WB Saunders.

[2] B. Guo, W. Lian, S. Liu, Y. Cao, J. Liu, "Comparison of diagnostic values between CA125 combined with CA199 and ultrasound combined with CT in ovarian cancer," Oncology letters. Vol. 17, no. 6, pp. 5523-8, 2019 Jun 1.

[3] M. Shetty, "Imaging and differential diagnosis of ovarian cancer," InSeminars in Ultrasound, CT and MRI. Vol. 40, no. 4, 302-318, 2019 Aug 1. WB Saunders.

[4] M. Akazawa, K. Hashimoto, "Artificial intelligence in ovarian cancer diagnosis," Anticancer research. Vol. 40, no. 8, pp. 4795-800, 2020 Aug 1.

[5] S. Nimmagadda, M.F. Penet, "Ovarian cancer targeted theranostics," Frontiers in oncology. Vol. 9, pp. 1537, 2020 Jan 21.

[6] L. Zhang, J. Huang, L. Liu, "Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color ultrasound detecting system," Journal of medical systems. Vol. 43, no. 8, pp. 1-9, 2019 Aug.

[7] A. Gadducci, E. Simonetti, G. Manca, F. Guidoccio, A. Fanucchi, S. Cosio, D. Volterrani, "Positron emission tomography/computed tomography in platinum-sensitive recurrent ovarian cancer: a single-center Italian study," Anticancer Research. Vol. 40, no. 4, pp. 2191-7, 2020 Apr 1.

[8] Z. Momenimovahed, A. Tiznobaik, S. Taheri, H. Salehiniya, "Ovarian cancer in the world: epidemiology and risk factors," International journal of women's health. Vol. 11, pp. 287, 2019.

[9] L. Manganaro, S. Gigli, A. Antonelli, M. Saldari, F. Tomao, C. Marchetti, E. Anastasi, A. Laghi, "Imaging strategy in recurrent ovarian cancer: a practical review," Abdominal Radiology. Vol. 44, no. 3, pp. 1091-102, 2019 Mar.

[10] P. Mohaghegh, A.G. Rockall, "Imaging strategy for early ovarian cancer: characterization of adnexal masses with conventional and advanced imaging techniques," Radiographics. Vol. 32, no. 6, pp. 1751-73, 2012 Oct.

[11] G. Di Lorenzo, G. Ricci, G.M. Severini, F. Romano, S. Biffi, "Imaging and therapy of ovarian cancer: clinical application of nanoparticles and future perspectives," Theranostics. Vol. 8, no. 16, pp. 4279, 2018.

[12] S. Nougaret, C. McCague, H. Tibermacine, H.A. Vargas, S. Rizzo, E. Sala, "Radiomics and radiogenomics in ovarian cancer: a literature review," Abdominal Radiology. Vol. 46, no. 6, pp. 2308-22, 2021 Jun.

[13] S. A. Ahmed, H. Abou-Taleb, A. Yehia, N.A. Abd El Malek, G.S. Siefeldein, D.M. Badary, M.A. Jabir, "The accuracy of multi-detector computed tomography and laparoscopy in the prediction of peritoneal carcinomatosis index score in primary ovarian cancer," Academic Radiology. Vol. 26, pp. 12, pp. 1650-8, 2019 Dec 1.

[14] S. Elsherif, S. Javadi, C. Viswanathan, S. Faria, P. Bhosale, "Low-grade epithelial ovarian cancer: what a radiologist should know," The British Journal of Radiology. Vol. 92, no. 1095, pp. 20180571, 2019 Mar.

[15] B. S. Harrington, Y. He, T. Khan, S. Puttick, P.J. Conroy, T. Kryza, T. Cuda, K.A. Sokolowski, W.C. Brian, K.K. Robbins, B.J. Arachchige,

"Anti-CDCP1 immuno-conjugates for detection and inhibition of ovarian cancer," Theranostics. Vol. 10, no. 5, pp. 2095, 2020.

[16] G. Avesani, M. Arshad, H. Lu, C. Fotopoulou, F. Cannone, R. Melotti, E. Aboagye, A. Rockall, "Radiological assessment of peritoneal cancer index on preoperative CT in ovarian cancer is related to surgical outcome and survival," La radiologia medica. Vol. 125, no. 8, pp. 770-6, 2020 Aug.

[17] S. Rizzo, M. Del Grande, L. Manganaro, A. Papadia, F. Del Grande, "Imaging before cytoreductive surgery in advanced ovarian cancer patients," International Journal of Gynecologic Cancer. Vol. 30, no. 1, 2020 Jan 1.

[18] M. Elezaby, B. Lees, K.E. Maturen, L. Barroilhet, K.B. Wisinski, S. Schrager, L.G. Wilke, E. Sadowski, "BRCA mutation carriers: breast and ovarian cancer screening guidelines and imaging considerations," Radiology. Vol. 291, no. 3, pp. 554-69, 2019 Jun.

[19] A. Meier, H. Veeraraghavan, S. Nougaret, Y. Lakhman, R. Sosa, R.A. Soslow, E.J. Sutton, H. Hricak, E. Sala, H.A. Vargas, "Association between CT-texture-derived tumor heterogeneity, outcomes, and BRCA mutation status in patients with high-grade serous ovarian cancer," Abdominal Radiology. Vol. 44, no. 6, pp. 2040-7, 2019 Jun.

[20] L. Beer, H. Sahin, N.W. Bateman, I. Blazic, H.A. Vargas, H. Veeraraghavan, J. Kirby, B. Fevrier-Sullivan, J.B. Freymann, C.C. Jaffe, J. Brenton, "Integration of proteomics with CT-based qualitative and radiomic features in high-grade serous ovarian cancer patients: an exploratory analysis," European radiology. Vol. 30, no. 8, pp. 4306-16, 2020 Aug.

[21] H. Lu, M. Arshad, A. Thornton, G. Avesani, P. Cunnea, E. Curry, F. Kanavati, J. Liang, K. Nixon, S.T. Williams, M.A. Hassan, "A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic-and molecular-phenotypes of epithelial ovarian cancer," Nature communications. Vol. 10, no. 1, pp. 1-1, 2019 Feb 15.

[22] V. Lago, P. Bello, B. Montero, L. Matute, P. Padilla-Iserte, S. Lopez, T. Marina, M. Agudelo, S. Domingo, "Sentinel lymph node technique in early-stage ovarian cancer (SENTOV): a phase II clinical trial," International Journal of Gynecologic Cancer. Vol. 30, no. 9, 2020 Sep 1.

[23] S. Wang, Z. Liu, Y. Rong, B. Zhou, Y. Bai, W. Wei, M. Wang, Y. Guo, J. Tian, "Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer," Radiotherapy and Oncology. Vol. 132, pp. 171-7, 2019 Mar 1.

[24] F. Castellani, E.C. Nganga, L. Dumas, S. Banerjee, A.G. Rockall, "Imaging in the pre-operative staging of ovarian cancer," Abdominal Radiology. Vol. 44, no. 2, pp. 685-96, 2019 Feb.

[25] G. Funston, W. Hamilton, G. Abel, E.J. Crosbie, B. Rous, F.M. Walter, "The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study," PLoS medicine. Vol. 17, no. 10, pp. e1003295, (2020 Oct 28).

[26] S. Nougaret, M. Tardieu, H.A. Vargas, C. Reinhold, S.V. Perre, N. Bonanno, E. Sala, I. Thomassin-Naggara, "Ovarian cancer: an update on imaging in the era of radiomics," Diagnostic and interventional imaging. Vol. 100, no. 10, pp. 647-55, 2019 Oct 1.

[27] R. Forstner, "Early detection of ovarian cancer," European Radiology. Vol. 30, no. 10, pp. 5370-3, 2020 Oct.

[28] G. Dhiman, M. Garg, A. Nagar, V. Kumar, M. Dehghani, "A novel algorithm for global optimization: rat swarm optimizer," Journal of Ambient Intelligence and Humanized Computing. Vol. 12, no. 8, pp. 8457-82, 2021 Aug.

Big Data and Internet of Things Web Service Management to Support Salt Agriculture Automation

Muhammad Choirul Imam[1], Dedi Trisnawarman[2], Hugeng[3], Hetty Karunia Tunjung Sari[4]
Department of Information System, Universitas Tarumanagara, Jakarta, Indonesia[1, 2]
Department of Electrical Engineering, Universitas Tarumanagara, Jakarta, Indonesia[3]
Department of Mangement, Universitas Tarumanagara, Jakarta, Indonesia[4]

***Abstract*—The integration of the internet of things in the information system application service web platform implemented in the supply chain has given rise to new formats and models, which are important manifestations of industry transformation and improvement. In the context of implementing long-term rural development plans, deep integration of the application of information technology and rural revitalization will act as a trigger that drives productivity and the development of other village business industries. The purpose of this research is to build a web service management model that can be used to manage and help optimize IoT-based salt farming production. The model built consists of software and hardware architectures and interconnections between tools. This research is divided into three stages: the first stage is to identify the data sources needed for big data needs, the second stage is to build a big data microservices model and the IoT model, the third stage is to integrate IoT data with the big data microservices model that has been built. The results of this study are in the form of an IoT device that can be run with big data micro services. The resulting IoT device can be used to automate water distribution based on the salinity value measured using a sensor.**

***Keywords*—Web service; big data; salt; IoT**

I. Introduction

Big data and the Internet of things (IoT) are part of the components of the Industrial Revolution 4.0. Growing rapidly and affecting all fields, including business, technology, health, education, agriculture and increasing benefits for many other fields. IoT is contributing to the growth of large amounts of data, and high speeds, which are characteristics of big data, as described by Gartner.

Indonesia is a country that has more ocean area, so of course it is a country that can produce its own salt and even export it to other countries. In fact, Indonesia has succeeded in producing abundant salt in 2012-2015, with a surplus of up to 2.9 million tons. However, the problem occurred in early 2017 that domestic salt production decreased, so we had to rely on imports for industrial salt needs [1].

Salt is a food ingredient that is consumed by all groups which is obtained through the process of crystallization of seawater or brine from sea salt or by the mining process of rock salt [2]. The influencing factor is that salt production in Indonesia has decreased because it still uses traditional salt processing methods such as dependence on sunlight and less use of the latest technology.

The Internet of Things (IoT) describes the ubiquitous connection between common objects and the Internet. IoT works by deploying thousands of smart devices in a living or industrial environment. This device collects information from the surrounding environment, performs desired processing activities on the data obtained and transmits the processed data via a secure and reliable communication channel. Recent advances in software, hardware and communication systems have significantly improved human lifestyles in terms of time, energy and cost savings [3, 4].

Digital farming is one of the key technologies that has gained support through different research initiatives in the last decade. A fundamental digital farm setup consists of a central management platform, a human-computer interface and a dedicated IoT module, which monitors the animals, their behavior and facilities [5]. All of these building blocks are connected into a common platform, using a communications network that, today, is often based on an IoT architecture combined with cloud- or edge-based functionality. A smart atmosphere is, in the digital farming scenario, a more sophisticated concept, which enables information processing from big data, combined with the current context and conditions, and adequate adaptation according to prior knowledge or decisions generated from machine learning algorithms [6].

This research is in line with the demands of industry 4.0 which is characterized by increased manufacturing digitization with four driving factors: 1) increased data volume, computing power, and connectivity; 2) the emergence of business analysis, capabilities, and intelligence; 3) the occurrence of new forms of interaction between humans and machines; and 4) improvement of digital transfer instructions to the physical world. The uniqueness and advantage of this program is that it integrates the application of Internet of Things (IoT) technology. The purpose of the proposed program is to increase salt production, both in quantity and quality through the application of IoT-based information technology.

Web service is an application of a collection of data (database), software (software) or part of software that can be accessed remotely by various devices with a certain intermediary. The use of web services is able to overcome interoperability problems and integrate different systems. In general, web services can be identified by using a URL like just a normal web (eg: www.webname.com). However, what distinguishes web services from the web in general is the interaction provided by web services [7].

This research resulted in an IoT design that can measure salt salinity, temperature and humidity. The salt salinity measure is then used to open the faucet and automatically circulate water for the next production process. The data generated by the sensors is then combined with other data related to the supply chain of salt commodities using big data architectures and application microservices.

II. Big Data

The term Big Data is often used to describe a variety of different concepts, from the collection and processing of large amounts of data to the various techniques used to process data [8] (Favaretto, 2020). The use of this term is often popularly used in various fields such as: health [9], geography [10], psychology [11], and social life [12]. The term Big Data has been used since the 1990s as technical and marketing jargon by silicon graphics. The use of this term in academia began to emerge in the 2000s on the topic of computer science [13] and statistics/econometrics [14]. Then, in subsequent years, as it develops, the use of big data is increasingly used, especially in conducting data analysis [15].

Weiss and Indurkhya [14] stated that a very large data set and compiled in a centralized data warehouse allows analysis using more complex techniques to analyze more comprehensive data. In theory, big data is able to provide more comprehensive conclusions. However, in practice, there are various problems that arise.

There are various definitions of big data. However, basically, big data has three attributes known as 3Vs, namely volume (large amount), velocity (fast processing), and variety (various data) [16]. Then, along with the larger amount of data, big data attributes are growing with additional attributes to the three previous attributes, such as: veracity, value and variability. Even with various definitions, big data is a large amount of data that comes from various sources [8].

One application of big data technology is in enterprise systems. For example, big data can be applied to manufacturing systems. Big data can help predict manufacturing equipment failures. Potential problems can be found by analyzing structured data (year, equipment make, and model) and unstructured data (log entries, sensor data, error messages, engine temperature, and other factors). With this data, manufacturers can maximize parts and equipment uptime and implement maintenance costs more effectively.

In the village there are several applications of big data in rural areas. For example, big data can be used to automate agriculture. Agricultural automation uses various tools connected to the internet/ internet of things (IoT). The use of various kinds of sensors and tools certainly produces various kinds of data that need to be controlled to automate agriculture [17].

III. Methodelogy

The stages of program implementation can be divided into several parts, namely:

1) Requirements gathering and analysis: Requirements Gathering and Analysis or mapping and analysis to get needs is the initial stage of the prototype model. In this stage the system requirements are defined in detail. In the process, developers and users provide feedback to each other. At this stage the required data sources are also identified, namely:

- Internal Data, The majority of internal data will come from database systems Gadingsari Village. This data includes production data, stock data, company financial data, as well as data from various instruments contained in village business units.

- External Data, this data includes raw material prices, selling prices, weather predictions, and input from related departments. There are various data sources that can be used to support the business units that are owned. These data can be collected from various sources and then analyzed which can produce conclusions that can help village businesses. These data are generated at different times. For example: weather data can be retrieved 7 days in advance while raw material data is sometimes available 1 month later. Of course this time difference can provide imperfect comparisons between data. Data types are also different. The average external data is unstructured, where the data does not have a fixed structure.

2) Design model: The second stage is to build a microservices model and an IoT model. Creating a micro service model design that will provide a brief description of the applications contained in the web service management system that you want to create. IoT model design: namely the IoT model that will be implemented in optimizing salt production. The IoT model will then be run using the proposed web service.

3) Integration: Design of Microservice System Integration and IoT technology in Web Service management.

IV. Result and Discussion

At the initial stage of the design is to build a microservices model and an IoT model, at this stage it is necessary to design a microservices model which will provide a brief description of the applications contained in the web service management system that you want to create. The Webservice Model Design created can be seen in Fig. 1.

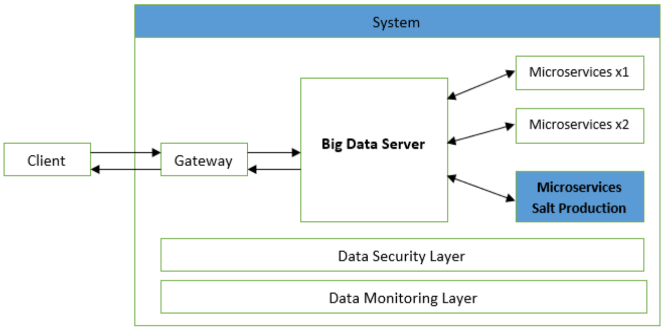


Fig. 1. Salt production webservice model design

A. Application Program Construction

Program construction is a codification process, namely building application programs based on designs that have been made at the application design stage. There are several stages in this activity, namely:

1) The first stage, the application program is to build a data ingestion application, to integrate data from various sources using the Application Programming Interface or API, as shown in Fig. 2. In this section, big data processing is carried out using several layers, such as the data visualization layer, analytical storage , analytical engine, data processing layer.

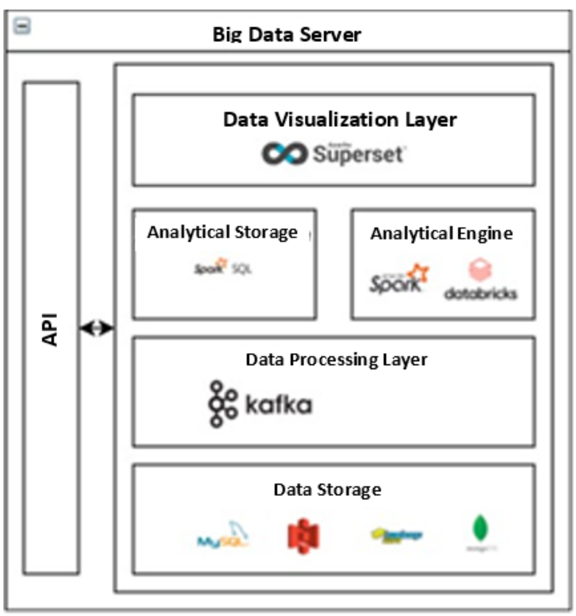


Fig. 2. Big data server architecture design

2) The second stage provides micro service services for users or business actors. At this stage, program construction is carried out to build application services. The target to be achieved is the availability of application services that can be connected to the IoT system for monitoring the salt production process, as shown in Fig. 3.

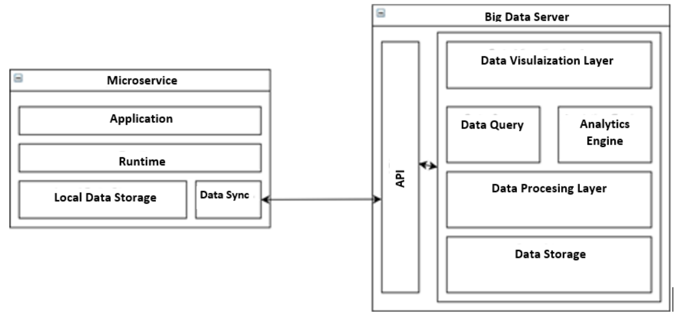


Fig. 3. Big data architecture microservice design

B. IoT Model Design

The conceptual design of the IoT model that will be implemented in optimizing salt production can be seen in Fig. 4, and the IoT design can be seen in Fig 5. The IoT model will then be run using the proposed web service. The picture shows the ESP32, which is a low-cost, low-power system on a chip (SoC) series with dual-mode Wi-Fi and Bluetooth capability! The ESP32 family includes the ESP32-D0WDQ6 (and ESP32-D0WD), ESP32-D2WD, ESP32-S0WD, and ESP32-PICO-D4 system-in-package (SiP) chips. At its heart is a dual-core or single-core Tensilica Xtensa LX6 microprocessor with a clock speed of up to 240 MHz. The ESP32 is highly integrated with a built-in antenna switch, RF balun, power amplifier, low noise receiver amplifier, filter and power management module. Engineered for mobile devices, wearable electronics, and IoT applications, the ESP32 achieves extremely low power consumption through power-saving features including smooth resolution clock gating, multiple power modes, and dynamic power scaling.

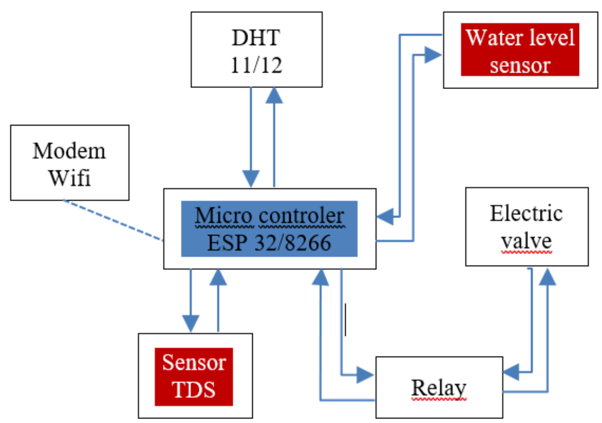


Fig. 4. Conceptual design diagram of the IoT model

Some of the functions of the IoT model design with web services are as follows:

1) The controller is an IoT tool that will automatically control the valve based on sensor conditions. All automation occurs at the controller level to save traffic going out through the modem

2) The controller will be connected to the web service via the tcp protocol. It should be noted that the controller can first connect to the local server or go directly to the web service.

3) The service data will be received. This service also allows the system to send messages to IoT devices.

4) Finally, the management system allows the use of data to and from IoT devices. The system for will also simplify the management of IoT devices

Fig. 5. IoT design

DHT22, shown in Fig. 6, is a digital relative humidity and temperature sensor. The DHT22 sensor uses a capacitor and thermistor to measure the surrounding air and output a signal on the data pin. DHT22 is claimed to have good reading quality, judged by the fast response of the data acquisition process and its minimalist size, and the price is relatively cheap when compared to a thermohygrometer [18].

Fig. 6. DHT22: Temperature and humidity module

Furthermore, the program is run based on program construction using the Python programming language as shown in Fig. 7.

```
from machine import Pin, ADC
import network
import dht
import ujson
import time
from umqtt.simple import MQTTClient

# VOID
# Register All Pin
sensor = dht.DHT22(Pin(12))
pmeter = ADC(Pin(13))
pmeter.atten(ADC.ATTN_11DB) # Set Analog to 3.3V
relay = Pin(15,Pin.OUT)

# Parameter to control Relay
relay_state = 0

# MQTT

# MQTT Server Parameters
MQTT_CLIENT_ID = "device1"
MQTT_BROKER    = "localhost"
MQTT_TOPIC     = "iot/device/1"

print("Connecting to WiFi", end="")
sta_if = network.WLAN(network.STA_IF)
sta_if.active(True)
sta_if.connect('Wokwi-GUEST', '')
while not sta_if.isconnected():
  print(".", end="")
  time.sleep(0.1)
print(" Connected!")

print("Connecting to MQTT server... ", end="")
client = MQTTClient(MQTT_CLIENT_ID, MQTT_BROKER,
port=1883)
client.connect()

print("Connected!")

   # LOOP
while True:

  # Pasive Sensor
  print("Measeuring Data....")
  sensor.measure()
  message = {
    "temp": sensor.temperature(),
    "humidity": sensor.humidity(),
    "pmeter": pmeter.read(),
  }
  # Some Logic Sensor
  relay_state = 0
  if(message["pmeter"] >= 1000):
    print("ook")
    relay_state = 1

  # Active Sensor
  relay.value(relay_state)
  message["relay"] = relay_state

  print(ujson.dumps(message))

  print("Reporting to MQTT topic {}: {}".format(MQTT_TOPIC,
  message))
  client.publish(MQTT_TOPIC, message)

  # Looping
  time.sleep(3)
{
  "version": 1,
  "author": "Choirul Imam",
  "editor": "wokwi",
  "parts": [
    {
      "type": "wokwi-esp32-devkit-v1",
      "id": "esp",
      "top": 0.67,
      "left": -28.67,
      "attrs": { "env": "micropython-20220618-v1.19.1" }
    },
    {
      "type": "wokwi-dht22",
      "id": "dht1",
      "top": 7.13,
      "left": -172.4,
         "attrs": { "temperature": "-11.5" }
    },
    {
      "type": "wokwi-slide-potentiometer",
```

```
    "id": "pot1",
    "top": 186.68,
    "left": -331.47,
    "attrs": { "travelLength": "30" }
   },
   {
    "type": "wokwi-resistor",
    "id": "r1",
    "top": 58.91,
    "left": 169.7,
    "rotate": 90,
    "attrs": { "value": "220" }
   },
   {
"type": "wokwi-resistor",

    "id": "r2",
    "top": 56.91,
    "left": 218.18,
    "rotate": 90,
    "attrs": { "value": "220" }
   },
   { "type": "wokwi-ks2e-m-dc5", "id": "relay1", "top": -32.09,
    "left": 206.97, "attrs": {} },
   {
    "type": "wokwi-led",
    "id": "led1",
    "top": 108.25,
    "left": 219.88,
    "attrs": { "color": "green" }
   },
   {
    "type": "wokwi-led",
    "id": "led2",
    "top": 106.41,
    "left": 177.85,
    "attrs": { "color": "red" }
   }
  ],
  "connections": [
   [ "esp:TX0", "$serialMonitor:RX", "", [] ],
   [ "esp:RX0", "$serialMonitor:TX", "", [] ],
   [ "esp:3V3", "dht1:VCC", "red", [ "v0" ] ],
   [ "pot1:VCC", "esp:3V3", "red", [ "h0" ] ],
   [ "pot1:SIG", "esp:D13", "blue", [ "h0" ] ],
   [ "r1:2", "led2:A", "green", [ "h0" ] ],
   [ "led2:C", "esp:GND.1", "black", [ "v0" ] ],
   [ "led1:C", "esp:GND.1", "black", [ "v0" ] ],
   [ "relay1:NO1", "r1:1", "gray", [ "v0" ] ],
   [ "r2:2", "led1:A", "green", [ "h0" ] ],
   [ "relay1:NC1", "r2:1", "gray", [ "v0" ] ],
   [ "relay1:COIL2", "esp:GND.1", "green", [ "v0" ] ],
   [ "dht1:GND", "esp:GND.2", "black", [ "v0" ] ],
   [ "pot1:GND", "esp:GND.2", "black", [ "v0" ] ],
   [ "dht1:SDA", "esp:D12", "green", [ "v0" ] ],
   [ "esp:3V3", "relay1:P1", "red", [ "v0" ] ],
   [ "esp:D15", "relay1:COIL1", "orange", [ "h0" ] ]
  ]
```


Fig. 7. IoT program construction

Furthermore, the design of the IoT model that has been built with the Python program is run using a server with web service management (Fig. 8).

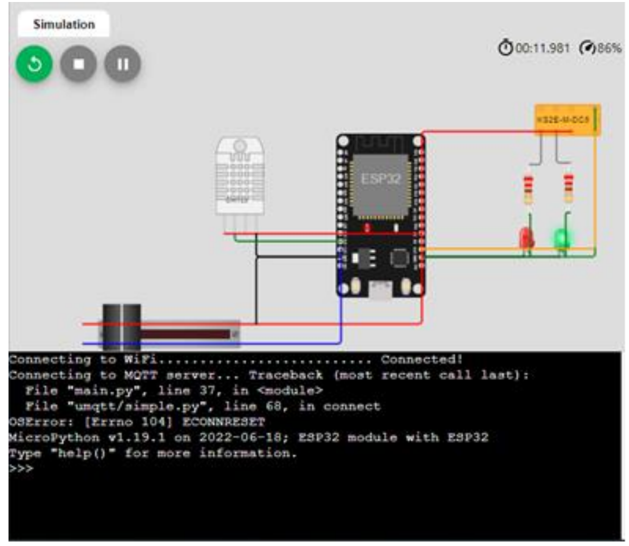


Fig. 8. Running IoT program design with web service management

The program that is run shows the results that the system design is in accordance with the expected goals. Each module of connected devices, such as temperature sensors, humidity, wifi servers can communicate with each other and provide information in the form of the desired report. This proves that the designed web service management model is running as expected.

V. Conclusion

The web service model with IoT has been successfully built and can be run with the Web Service management model. The web service is run using a Public IP server, making it easier to run additional applications. Application programs run well with time optimization, as expected. The test results show that the sensor valve can function automatically based on the water level measurement sensor.

References

[1] Alfiani, L.F., Antinomi Peraturan Perundang-Undangan Mengenai Impor Garam Nasional. Jatiswara, 2021. 36(1), pp.24-37.

[2] Pebrianti, S.A., Kusumah, S.H. and Yunita, N., Identifikasi Permasalahan Kualitas Garam Industri Di PT Niaga Garam Cemerlang Menggunakan Check Sheet, Pareto Chart Dan Fishbone Analysis. Jurnal Fakultas Teknik Kuningan, 2021. 2(3), pp.79-86.

[3] Sisinni, E., Saifullah, A., Han, S., Jennehag, U. and Gidlund, M.,. Industrial internet of things: Challenges, opportunities, and directions. IEEE transactions on industrial informatics, 2018 14(11), pp.4724-4734.

[4] Saeed, N., Alouini, M.S. and Al-Naffouri, T.Y., Toward the internet of underground things: A systematic survey. IEEE Communications Surveys & Tutorials, 2019. 21(4), pp.3443-3466

[5] Jayaraman, P.P., Yavari, A., Georgakopoulos, D., Morshed, A. and Zaslavsky, A., Internet of things platform for smart farming: Experiences and lessons learnt. Sensors, 2016. 16(11), p.1884.

[6] Jukan, A., Masip-Bruin, X. and Amla, N., Smart computing and sensing technologies for animal welfare: A systematic review. ACM Computing Surveys (CSUR), 2017. 50(1), pp.1-27.

[7] Ahmad, F. and Sarkar, A., QaaS (quality as a service) model for web services using big data technologies. Enterprise Information Systems, 2017. 11(9), pp.1352-1373.

[8] Favaretto M, De Clercq E, Schneble CO, Elger BS. What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PloS one. 2020 Feb 25;15(2):e0228987

[9] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems. 2014 Dec;2(1):1-0.

[10] Graham M, Shelton T. Geography and the future of big data, big data and the future of geography. Dialogues in Human geography. 2013 Nov;3(3):255-61.

[11] Kang L, Wu C, Wang B. Principles, approaches and challenges of applying big data in safety psychology research. Frontiers in psychology. 2019 Jul 9;10:1596.

[12] Kang L, Wu C, Wang B. Principles, approaches and challenges of applying big data in safety psychology research. Frontiers in psychology. 2019 Jul 9;10:1596.

[13] Han J, Pei J, Tong H. Data mining: concepts and techniques. Morgan kaufmann; 2022 Jul 2.

[14] Varian HR. Big data: New tricks for econometrics. Journal of Economic Perspectives. 2014 May;28(2):3-28.

[15] Diebold FX. On the origin (s) and development of “big data”: The phenomenon, the term, and the discipline. Pobrano z https://www. sas. upenn. edu/~ fdiebold/papers/paper112/Diebold_Big_Data. pdf. 2019.

[16] Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems. 2022 Jan 1;62:738-52.

[17] Patnaik S, Sen S, Mahmoud MS, editors. Smart Village Technology: Concepts and Developments. Cham: Springer International Publishing; 2020 Feb 7.

[18] Abdulrazzak, I.A., Bierk, H. and Aday, L.A., Humidity and temperature monitoring. Int. J. Eng. Technol, 2018, 7(4), pp.5174-5177.

Recognition of Copy Move Forgeries in Digital Images using Hybrid Optimization and Convolutional Neural Network Algorithm

Anna Gustina Zainal[1], Dr. Chamandeep Kaur[2], Dr. Mohammed Saleh Al Ansari[3], Ricardo Fernando Cosio Borda[4], Dr. A. Nageswaran[5], Rasha M. Abd El-Aziz[6]

Department of Communication, University of Lampung, Indonesia[1]
Lecturer, Department of IT, Jazan University, Saudi Arabia[2]
Associate Professor, College of Engineering-Department of Chemical Engineering, University of Bahrain, Bahrain[3]
Universidad Privada del Norte, Peru[4]
Professor, Department of CSE. Sri Venkateswaraa College of Technology, Vadakkal village, Sriperumbudur- 602105[5]
Department of Computer Science-College of Science and Arts in Qurayyat, Jouf University, Saudi Arabia[6]
Faculty of Computers and Information, Assiut University, Assiut, Egypt[6]

***Abstract*—In the modern day, protecting data against tampering is a significant task. One of the most common forms of information display has been digital photographs. Images may be exploited in a variety of contexts, including the military, security applications, intelligence areas, legal evidence, social media, and journalism. Digital picture forgeries involve altering the original images with strange patterns, which result in variability in the image's characteristics. Among the most challenging forms of image forgeries to identify is Copy Move Forgery (CMF). It occurs by copying a portion or piece of the picture and then inserting it again, but in a different place. When the actual content is unavailable, techniques for detecting fake content have been utilised in image security. This study presents a novel method for Copy Move Forgery Recognition (CMFR), which is mostly based on deep learning (DL) and hybrid optimization. The hybrid Grey Wolf Optimization and African Buffalo Optimization (GWO-ABO) using Convolution Neural Network (CNN) technique i.e., GWO-ABO-CNN is the foundation of the suggested model. The developed model extracts the features of images by convolution layers, and pooling layers; hereafter, the features are matched and detect CMF. The MICC-F220, SATs-130, and MICC-F600 datasets were three publicly accessible datasets to which this methodology has been implemented. To assess the model's efficacy, the outcomes of implementing the GWO-ABO-CNN model were contrasted with those of other approaches.**

Keywords*—Copy move forgery; convolutional neural network; image authentication; deep learning; tampered images

I. Introduction

Today's technological images play a crucial role in a broad range of fields. They are used in a variety of uses in the fields of broadcasting, journalism, medicine, and the army, to name a very few. The computerised image can be seen as a notable resource of information in today's advanced globe due to the advancement in the technology of sophisticated picture, such as sensors, coding, and Computers, as well as the widespread usage of the internet [1]. In addition, advanced image forgery refers to the intentional manipulation of a digitized image in order to change the conceptual interpretation of the contextual perspective contained within. Also, with availability of cutting-edge information structures editing tools like Photoshop, it becomes quite simple to create sophisticated fakes from one or more images. The reliability of photographs plays a crucial role in a variety of fields, such as measurement analysis, criminal probe, surveillance systems, organisational learning, medical imaging, and media broadcasting. Creating phony pictures is a specialty with a lengthy tradition. But in the current technology age, it showed out to be very simple to change the facts talked to by an image without any obvious consequences.

One of the most widely popular techniques for altering digital photographs is copy-move. One of two factors can explain why there are copy regions in an image: first, the proximity of two particles or objects that are identical in size, form, and colouring; one of them may be a copy of the other one. Second, the appearance of duplicate regions in the results is caused by the proximity of a reasonably massive area with one colour and similar in features, such as foundational principles sky, splitter, etc. Copy-move forgeries is created by copying and pasting a region or sector from one spot in an image to some other spot inside the same picture in order to modify or hide one or more objects and create a false vision. Moreover, copy-move forgeries identification is known to be successful when using key point-based analysis. There were some changes made to the image during copy-move forgeries. Moreover, to implement quality impersonating forgeries, techniques including turning, cropping, lightening, reduction, and force and contribute are used. Nowadays, even a non-expert may easily produce convincing forgeries in digital images because of modern digital imaging and robust image manipulation tools. Huge different forgeries have been created in recent decades as a result of digital manipulation, which involves incorporating or removing certain parts from the image. Thus, checking the materials of digital photos or discovering fraudulent areas would be immediately helpful, for example, when images are used as evidence at trial.

Evidently, there is no one single, efficient option to the challenge of detecting digital forgery [2].

Two kinds of techniques and block-based approaches—can be used to recognize the presence of a copy-move in a visual. The fundamental area of assessment consists of examining the decentness of images and identifying signs of modification without the requirement to use image files. Structurally complex transform, resilient to disturbance and geometrical modifications are only a few samples of the main aspects that feature-based techniques for copy-move verification retrieve from copied images. Block-based copy-move identification, in comparison, analyses the properties of each block in the frequency response by splitting the copied images into matching or non-matching blocks [3]. LBP uses the features retrieved from the frames of gray-level image data to pinpoint the fake portions in visuals. Moreover, with SURF, local picture features that are resistant to noisy and geometrical changes are quickly extracted. In terms of how the two methods replace a specific image location with some other picture, copy move manufacture is very similar to picture patching. For instance, copy move forgery uses a portion of the original base picture as its origin instead of using an external image. The comparable image served as both the origin and the endpoint for the modified image. Hereafter, parts of the initial image are duplicated, moved to the perfect location, and then merged in a duplicated move fraud. Typically, this is used to conceal particular nuances or to replicate specific areas of an image [4]. Moreover, to minimise the impression of irregularities between the initial and delayed region, some additional preparation, such as center separation and hiding, is often linked along the edge of the modified region.

Furthermore, the judgement regarding the validity of the data is subsequently made by evaluating the outcomes of several strategies. In single-image forgeries, the copy-move approach is used to substitute a section of the original picture with a section that was previously eliminated. Forgers employ certain comment on these thread modifications on the duplicated portions, including spin, scalability, and reflections, as well as mouldings and mixing, to make forgeries are easier to conceal. The addition of content for one image that comes from another image or photos is called forgeries using various pictures [5]. In addition, there are two distinct streams of computerized photo fraud detection. They are passive tactics and dynamic approaches. Images are extremely simple to change in the technological environment utilising well-known image manipulation software applications like Adobe Illustrator and Gimp, which results in an astounding number of phoney images entering in every day over the Web.

Identification of Digital image forgery is an important research area; however, it seems to be a significant problem. Finding proof of a fake is done by looking for unique characteristics, and traits. The actual image's analytical, architectural, or physiological qualities must be homogeneous for the detection techniques to work, but manipulated images lack this homogenity. Digital image counterfeiting comes in a variety of forms, including picture merging, image reshaping, image editing, image re-sampling, and copy-move forgeries [6]. Several modified over time in a variety of domains, including target detection, ML, and computation imagery, were successfully solved using the DL method. As a result, it can be useful in developing patching forgery investigative techniques.

Among the most successful and effective techniques applied in a variety of applications for image processing during the past ten years is the Artificial Neural Network. In addition, Convolutional Neural Network (CNN) is commonly employed for Copy-Move Forgery Detection in recent years. However, CNN method implemented on challenging datasets has attained less accuracy and high testing time in CMF recognition [7]. To address this issue, the current work proposes an effective hybrid optimization with CNN for recognizing CMFD from digital images, which can achieve great result at a very cheap computational effort. CMFD's objective is to identify areas that are comparable to certain other areas of the image. Architectural or post-processing procedures are typically performed upon altered areas during the tampering procedure to make the forgeries realistic and undetectable. The key piece of evidence in CMFD is the strong resemblance between the modified areas and the origin. Here, the GWO-ABO with CNN technique serves as the framework's main tenet.

The rest of the section is organized as follows: Section II describes the recent literatures related to CMF detection, Section III explains the materials and methods in which the dataset and proposed methodology are detailed, Section IV describes the result and discussion, and Section V concludes the paper.

II. Related Works

Although there are many other styles of digital picture scams, it can be quite difficult to spot copy-move forgeries. In order to identify copy-move fraud, the study utilized a new robust approach which is based on the Speeded up Robust Feature descriptor, Approximate Nearest Neighbour for feature representation, and Simple Linear Iterative Clustering for categorising the actual image into large pixel blocks. The portions that are in dispute are identified by swapping out paired key frames for comparable large pixel blocks, followed by the merging of adjacent blocks based on similar LCF [8]. On a variety of samples, including MICCF220, MICC-F600, MICC-F2000, and CoMoFoD, the paper measured a time complexity of 3.84 secs with 91.95% positioning accuracy. However, because the process takes longer than the alternative way, it cannot be used in the present circumstances.

The author in [9] utilised robust feature enhancement acceleration, and the SVM is used to recognise the particular object. There were some changes made to the image during copy-move forgeries. To create effective portrayal forgery, techniques including rotating, magnification, lightening, reduction, and force and contribute are used in this paper. The dominant features from the source images are chosen in this case by the SURF extracting features. The SVM classifies these input photos using detection and recognition in order to retrieve the paired local features. Here, the outcomes demonstrate that forging images are taken from a collection of test images. When there is additional uncertainty in the data set, SVM doesn't really function very well.

The paper [10] used a reliable technique for identifying Copy-Move forgeries in digital photographs. The procedure begins by removing an image's SIFT characteristics, which are resistant to modifications in lighting, spin, scale, etc. Characteristics are then compared to one another in order to check for any potential picture forgeries due to the resemblance between the inserted region and the duplicate data region. The utilised system works well with various post-image analysis methods and is resistant to complex image analysis because of the high robustness of SIFT features extracted. However, to increase the robustness over inadequate SNR and small-size tampering regions, more research is still required.

The paper [11] utilized a technique for spotting fake parts in digital photographs. The steps in this procedure are as follows: (1) transform the colour image into gray, (2) break the image representation into overlaying blocks of pixels, (3) use DCT to identify and extract feature, (4) aggregate blocks using the K-means approach, and (5) use radix order to correlate features. Here the method was appeal four different photos and on the foundation of the experimental work. Diverse methods, such as DCT and DWT, are typically utilized to extract the characteristics from the digital photos in unrelated domains, such as feature recognition and image fraud prevention.

A reliable copy-move picture fraud prevention method used Gaussian-Hermite Moments in the paper. The method separates the image as an input into overlaying, fixed-size chunks, and then extracts the GHM values for every chunks. GHM is a powerful tool that may be used to recover picture characteristics that are rotation-, translation-, and scale-invariant. Through lexicographically ordering all the characteristics', related chunks are matched. The outcomes of the experiments demonstrate how precisely the suggested technique can find the duplicate forged portions in a forged image [12]. Here, the used technique performs better than previous relevant strategies, according to experimental findings, both at the image and pixel levels. However, the method can necessitate additional time, rendering it inappropriate for the next presentation.

The outcome of applying the lighting estimation technique reveals that the method is only marginally improved by excluding specific observations, such as geometry depending on shadowing and upgrade and improve characteristics on selected input images [13]. The technique to identify image modifications is approved by the publication in consideration of the abnormal illumination. By calculating surfaces and lighting data from the centre points, the counterfeit in images is identified. The process produces acceptable results for many source images. Here, the procedure is not just restricted to human faces and the identification of image regions with the same brightness. It can be used for photographs with any kind of item included in the scene. However, the method needed to be improved in order to estimate sections with the same brightness and to automatically choose spots.

The method developed in this work is a reliable one for spotting digital picture forgeries and is sufficiently robust to withstand attempts at image alteration. The very first process in the work is to generate the Rgb values into YCbCr space. Next, the Hilbert-Huang Transform includes are obtained from the chrominance-red element Cr. Finally, three distinct classification techniques, KNN, SVM and ANN—have been evaluated and contrasted for the task of classifying images as accurate or fake. Structural-Similarity is used to measure the precision of copy - move and to validate the findings. The suggested technique may identify pixels in the context of several images post-processing assaults, according to the findings of the robustness assessment; nevertheless, these cyber-attacks gradually degrade it [14]. However, since these techniques necessitate a large amount of storage for the training examples, the procedure might be slowed significantly.

The paper [15] employed a blind verification approach based on bundle segmentation and visual scrambling to discover repeated sections for copy-move forgeries. DCT is employed to find the DCT coefficient matrices for each level sufficient source image in suspected images. The visual hashing characteristic matrices and matrices are treated in a systematic manner. Additionally, to increase analysis is an essential part; a package grouping technique is utilized in place of conventional textual order techniques in this work. The perceptive hashing vectors in every package and its neighbouring packages, comparable chunks can be distinguished. A computerisation that builds a feature representation to describe an image block using perceptive hash characters can withstand certain common attacks like adding white Gaussian noise and GB. Additionally, the suggested approach is vulnerable to some sophisticated assaults such sector manipulation and scalability. However, it is more expensive than other ways, consequently not everyone can benefit from it.

One of the most prevalent popular forms of digital image forgeries, merging, was identified using a technology in the study [16] . Here, the VGG-16 CNNs are the foundation of the algorithm. The network architecture in used to receive visual features as information and determines whether an update is authentic or fake. During the learning phase, the work chooses portions on the edges of the inserted splice and from the original image portions. The potential for using the strategy when the JPEG technique repeatedly compresses deformed photos over a small range. As compared to other techniques, the reported findings show great generalization ability for a set of photos with intentional imperfections. However, Dynamic graphic images cannot be handled by the JPEG image standard, and stacked images are not supported.

The forensic evidence of analyser sensors was addressed in the paper using a ML-based method. Numerous diagnostic approaches, including digital image authenticating, related to Google, and security mechanisms, are crucial for digital image assessment because of the rising accessibility and effectiveness of image altering applications. In this work [17], a CNN-based method for identifying scanner framework has been used. The test results demonstrated that originating scanner authentication may be done with extreme precision. A system, which brings together the advantages of ResNet and GoogleNet while still being compact. A dependability map that shows the altered areas in a digital image is also generated by the article. However, a deeper network has a significant downside in that it typically takes weeks to train, rendering it unworkable in practical systems.

In the paper [18], two methods for recognizing image combining are supported. Both techniques are used to retrieve input image features by using overlapped chunks. The first method is used to extract LBP or LTP components based on the image's chrominance's gray-levels, while the other approach retrieves ELTP characteristics from the luminance site's fast Fourier transformation. These methodologies' outcomes have been summarized in a favourable manner. By reaching an accurateness of 88.62% on data compression from the CASIAv1.0 collection, the FFT-ELTP approach works reasonably well. However, all of the strategies that are being discussed entail intricate modifications like the DCT and FFT, which makes the technique more difficult. It may be possible to eliminate the requirement for such sophisticated activities with the effort in a similar approach. An additional subject for investigation in the investigation is the distribution of the forgeries in the photograph.

The paper [19] utilized, a key point-based image evidence collection procedure based on the Helmert conversion and super pixel recognition technique has been suggested. The method seeks to gather forensic data while detecting copy-move forgery photos. The activities or tasks make up the suggested approach's process. First, use a SIFT approach to identify the key details and related attributes. The resemblance between key points will then be determined based on the descriptors to produce matched pairings. In order to determine the precise counterfeit zones, the Helmert conversion to group the matched pairs according to geographical distances and geometry restrictions. In respect to province approaches, the experimental outcomes from evaluating multiple samples show that the suggested approach achieves excellent highly precise rates. However, the used method is not resistant to symmetrical, recurrent, and smoother processes for regional manipulation.

III. Materials and Methods

A. Dataset Description

The most difficult or well-known databases in the evaluation of CMF recognition methods are MICC-F220 [20], SATs-130 [21], and MICC-F600 [22]. Table I provides detailed information of various assessment datasets. The MICC-F600 database challenged attacks have been organised into four tiers with various forgery attacks. The MICC-F600 and MICC-F220 databases had been used to effectively test the presented technique. While the SATs-130 database has only 96 images, which is too little and collide with DL method's nature. Large databases are required for the training phase of DL methods in order to accurately extract feature maps and construct system behaviour.

TABLE I. Dataset Specification

Sl. no	Database	Image Composition			Tampered region size	Images size (Pixels)
		Total	*Forged*	*Original*		
1	MICC-F600	600	152	448	The size of the forged region changes from image-to-image.	800×532 to 3888×2592
2	SATs-130	96	48	48	The size of the forged area changes from image-to-image.	1024×683 to 3264×2448
3	MICC-F220	220	110	110	1.2% of the entire image seems to be the forged area.	722×480 to 800×600

B. Proposed Methodology

The suggested model has four steps, and Fig. 1 depicts the proposed method's workflow. Pre-processing seems to be the initial step. Two things are being accomplished by this technique. The first objective is to uniformly scale all of the input images, and the second objective is to transform the images into tensors. The extraction of features seems to be the second stage. To retrieve the features from the raw images, this step is obtained. Feature matching seems to be the third stage, which is obtained to demonstrate the presence of forgeries. The final stage is post processing. First, potential erroneous matches are removed at this stage. The CMF in digital images is afterwards identified.

1) Pre-processing by GWO-ABO: GWO appears to be a bionic optimization method. It mimics the cooperative, well delineated working relationships found in grey wolf hunting behaviour. Grey wolves often live in packs of five to twelve people and have a rigid, dominant hierarchy based on wolf leadership skills. The three phases of the GW pack's predatory process: hunting, encircling, and attacking. The most prominent wolf in the pack, also known as wolf, typically served as its leader. The GWO terminology for the second and third tiers of leadership wolves is wolf and wolf, respectively. These second- and third-ranking auxiliary wolves aid the main wolf in making hunting decisions. All other following wolves are classified as wolves, and they pursue and engage in combat with the prey alongside these powerful wolves.

For hunting, the enclosing of prey approach is used. For iteration i, the mathematical framework for this method is shown in below Eqns. (1) and (2).

$$\vec{F} = \left|\vec{Y} \times \overrightarrow{Q_p}(i) - \vec{Q}(i)\right| \tag{1}$$

$$\vec{Q}(i+1) = \overrightarrow{Q_p}(i) - \vec{D}.\vec{F} \tag{2}$$

Here, $\vec{D}$ and $\vec{Y}$ are coefficient vectors, which is described as $\vec{D} = 2\vec{d}.\overrightarrow{v_1} - \vec{d}$ and $\vec{Y} = 2.\overrightarrow{v_2}$. Where, the random vectors $\overrightarrow{v_1}, \overrightarrow{v_2}\ \varepsilon\ (0,1)$ and $\vec{d} = d_1\left(1 - {}^{i}/_{maxi}\right)$, linearly decreases from d_1 to zero; d_1 value was set as 2 in actual GWO. Moreover, $maxi$ represents maximum number of iterations. The GWO's hunting process has been headed by three finest solutions i.e., $\propto$, β and γ wolves. Thus, these 3 leading solution's positions have been saved in the pack and the remaining ω wolves update their positions predicated on them.

This position updating technique's mathematical model is represented in Eq. (3).

$$\vec{Q}(i+1) = {(\overrightarrow{Q_1} + \overrightarrow{Q_2} + \overrightarrow{Q_3})}\big/{3} \quad (3)$$

Where, $\overrightarrow{Q_1}, \overrightarrow{Q_2}, and\ \overrightarrow{Q_3}$ is computed by Eqn. (4)

$$\begin{aligned} \overrightarrow{Q_1} &= \overrightarrow{Q_\propto}(i) - \overrightarrow{D_1}.\overrightarrow{F_\propto} \\ \overrightarrow{Q_2} &= \overrightarrow{Q_\beta}(i) - \overrightarrow{D_1}.\overrightarrow{F_\beta} \\ \overrightarrow{Q_3} &= \overrightarrow{Q_\gamma}(i) - \overrightarrow{D_1}.\overrightarrow{F_\gamma} \end{aligned} \quad (4)$$

Here, $\overrightarrow{F_\propto}$, $\overrightarrow{F_\beta}$, and $\overrightarrow{F_\gamma}$ are computed by Eqn. (5)

$$\begin{aligned} \overrightarrow{F_\propto} &= \left|\overrightarrow{Y_1} \times \overrightarrow{Q_\propto}(i) - \vec{Q}\right| \\ \overrightarrow{F_\beta} &= \left|\overrightarrow{Y_2} \times \overrightarrow{Q_\beta}(i) - \vec{Q}\right| \\ \overrightarrow{F_\gamma} &= \left|\overrightarrow{Y_3} \times \overrightarrow{Q_\gamma}(i) - \vec{Q}\right| \end{aligned} \quad (5)$$

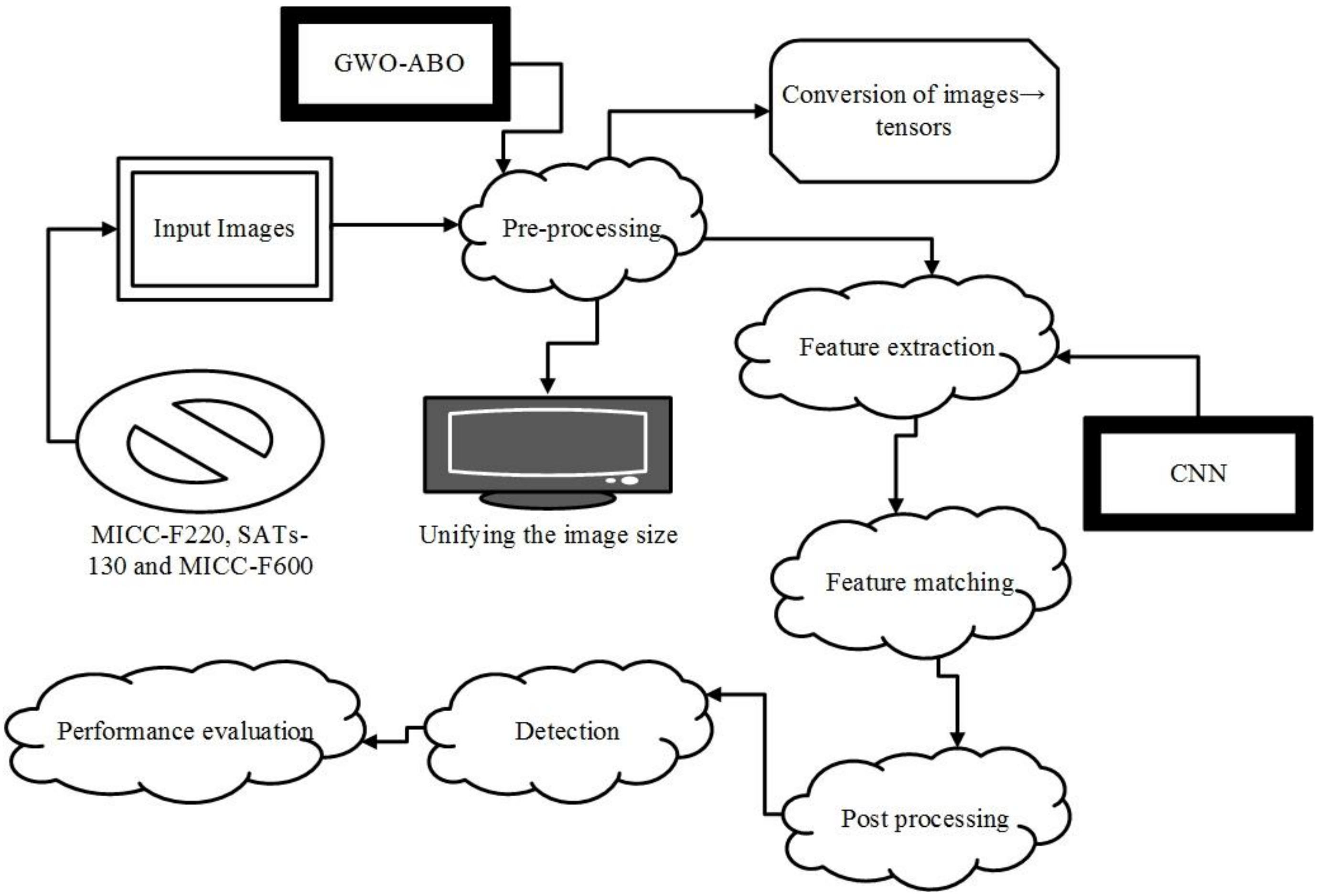


Fig. 1. Presented methods workflow

African buffalos' three principal characteristics are extensive memory capacity, cooperative cum communicative ability in both good and bad times, and extreme intellect gives origin to a democratic character. There are two sounds: '*maaa*' and '*waaa*', which are indicated by P_k and S_k, respectively. The buffalo's movement is determined by Eq. (6).

$$P_k + 1 = P_k + V_1\left(\vec{Q}(i+1) - S_k\right) + V_2(d_u.k - S_k) \quad (6)$$

Here, the input image is denoted as S_k, P_k is the image size, k represents the iteration, the learning factors are represented as V_1 and V_2, the wolf's best fitness is represented as $\vec{Q}(i+1)$ and $d_u.k$ is buffalo's best in each iteration. The image → tensors process is done by Eq. (7).

$$S_k + 1 = {(S_k + P_k)}\big/{\pm 0.5} \quad (7)$$

2) Feature extraction: Three convolutional layers, typically preceded by a pooling layer, make up the suggested feature extraction method. A max pooling layer comes after a convolutional layer that combines 32 filters in total. In complement to the ultimate combo of a convolutional layer with 128 filters and the final max pooling layer, a convolutional layer with 64 filters as well as a max pooling layer are also used. A feature map, which reflects the input image, has been produced by this process. The layers in extracting features from input images are shown in Fig. 2.

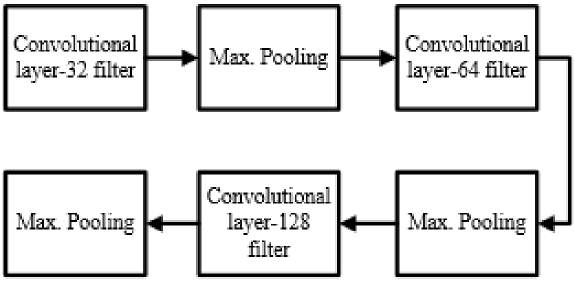


Fig. 2. Layers in feature extraction

The convolutional and pooling layer sequencing creates a feature vector out of the feature map. The CNN structure is shown in Fig. 3. The input for the following step is this feature vector.

a) Convolutional layer: A series of 2D digital filters make up the extracting features layer known as the convolutional layer. A convolution layer was used to minimise the data variables and extract the important features. The convolution layer includes scale invariance, interpretation invariance, and rotation invariance. Both the over-fitting problem is lessened and the generalisation idea is added to the fundamental framework. Eq. (8) illustrates the input of the convolutional layer like a collection of GWO-ABO pre-processed images.

$$h_k^l = f\left(\sum_{j\epsilon Y_k} h_k^{m-1} * R_{ki}^l + Y_k^m\right) \quad (8)$$

Here, h_k is the given map cluster; R_{kj}^m is the convolution kernel, which has been employed for joining the i^{th} input feature map with k^{th} output feature map; Y_k^m defines the consistent bias to the i^{th} input feature map, the activation function is denoted as f.

$$\delta_k^m = \delta_k^{m+1} C_k^{m+1} \times (s^m) = \beta_k^{m+1} up(\delta_k^{m+1}) \times (s^m) \quad (9)$$

Here, $m+1$ describes the pooling layer; the convolution kernel is denoted as C. The error function's partial derivatives convolution kernel and cost is described in Eq. (10) and (11)

$$\partial G / \partial R_{ki}^m = \sum_{v,n} (\delta_k^m) v, n (A_k^{m-1}) v, n \quad (10)$$

$$\partial G / \partial Y_k^m = \sum_{v,n} (\delta_k^m) \mathrm{v}, \mathrm{n} \quad (11)$$

Where, $(A_k^{m-1}) v, n$ is h_k^{m-1} patch for every convolution and $B_{kj}^m\,(v, n)$ is the patch centre. Lower random weights are used to initialize these filter's values. Then, during the training phase, these weights' values have been modified. The characteristics from the input images should be extracted by the filters that have been used. A feature map with an input image's filtered copies depth is produced by this procedure. A pixel's updated value (P_n) is equal to the sum of its former surrounding pixels' (p) values multiplied by the filtration elements (g) that have been used. The calculation is shown in Eq (12).

$$P_n = \sum_{j\in w} P_j \times g_j \quad (12)$$

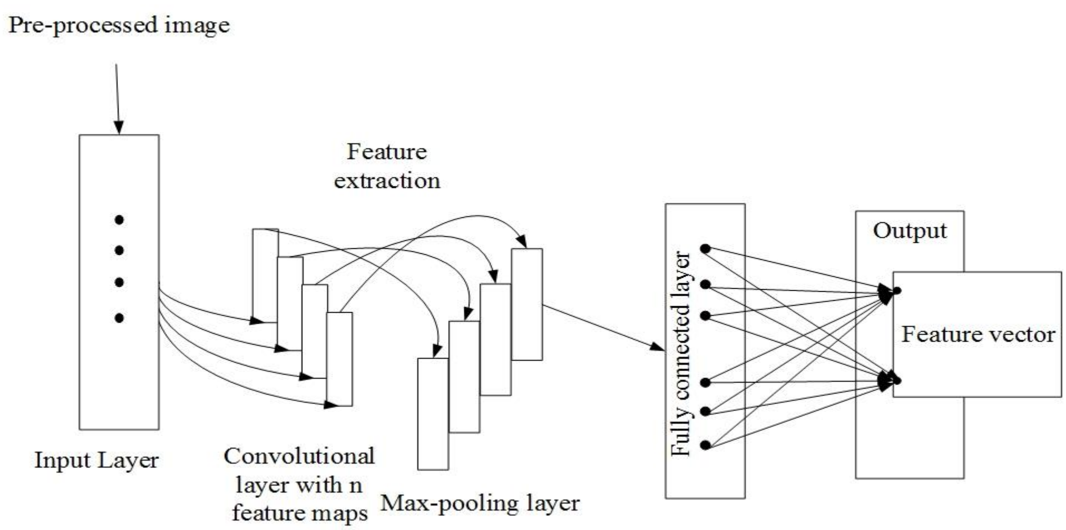


Fig. 3. CNN structure

b) Pooling layer: The convolutional layer's feature map's dimensions are reduced using the pooling layer, a form of feature reduction technique. Windows are created using the split input feature map. Every window is divided into its highest value from its contained values by the max pooling method. In contrast, the mean pooling divides the windows into the average value of the window's contained values. The suggested approach has implemented the max pooling layers.

3) Feature matching: To detect forgeries after obtaining the feature matrix, relevance searching has been carried out. In order to expedite the matching process, the feature matrix has been first lexicographically processed for this reason. Following that, the Euclidean distance used to compare vector similarities is provided. By using the provided Eq. (13) to contrast the vectors distances with a predefined threshold δ, the matching vectors have been identified.

$$f_k^i = (f_1, f_2, \dots, f_{10}), \sqrt{\sum_{k=1}^{10}\left(f_i - f_j\right)^2} \leq \delta \quad (13)$$

Where, δ = 1.5. The Euclidean distance between the matched blocks is used to check similarities in the following stage of matching; it must be larger than threshold β to prevent false matches.

4) Post processing: First, potential erroneous matches have been removed in this stage. The shift vector between identical blocks has been computed for this reason. Shift vectors have been produced from the suspect pairs' top left coordinates. Additionally, it is assessed whether the quantity of blocks with the similar shift vector surpasses a predefined threshold value ($\gamma = 32$). If this criterion is met, it is demonstrated that CMF has finished handling the relevant blocks. The CMF on digital photos is finally discovered.

IV. Results and Discussion

The assessment of the suggested algorithm's findings was reported in the results section. A comparative study with earlier methods working with CMF recognition was conducted after findings listing and discussion. A computer system running Windows 10 with an Intel Core i7 8th generation CPU, 4 GB of GPU hardware supporting CUDA, processor-64-bit, and RAM 8 GB was utilised to construct the suggested method. In additional to Keras and TensorFlow, Python 3.5 software tools were used to implement the suggested method's backend toolkits.

A. Evaluation Metrics

The major metrics for evaluating the DL model's detection performance are: F-measure, TPR, accuracy, FNR, FPR, TNR, and TT. These metrics were employed for evaluating the presented GWO-ABO-CNN models performance in CMF recognition.

True negative (T_n) represents the number of real images that were really identified as real images. False negative (F_n) refers to the number of altered images that were mistakenly identified as real images. True positive (T_p) represents the number of manipulated images that were accurately identified and false positive (F_p) represents the number of real images that were mistakenly identified as altered images.

1) Accuracy: Accuracy has been referred to as the percentage of correctly detected image pixels. Absolute pixel precision is another phrase for this. Despite being the most fundamental performance indicator, anytime there is a class disagreement, it may lead to erroneous image detection findings. When one recognised category outperforms another, there is a category disparity. In this scenario, the dominant class's superior accuracy will outweigh the other group's poorer accuracy, leading to biased findings. If there has been no group discrepancy, the accuracy measure was advised for evaluating detection results using images.

Accuracy is described in Eq. (14),

$$Accuracy = {T_p + T_n}\Big/{T_p + F_p + T_n + F_n} \quad (14)$$

2) True positive rate (TPR): TPR, or the proportion of images which have been accurately identified as CMF, relates to the capacity to positively detect an instance of forgery. Moreover, sensitivity is described as the proportions of predicted forgery among those images are real.

TPR is described in Eqn. (15),

$$TPR = {T_p}\Big/{T_p + F_n} \quad (15)$$

3) True negative rate (TNR): The chance that an image which does not have any forgery will have a negative test outcome is known as TNR, and it is the proportion of forgery which is appropriately detected as CMF. Moreover, specificity is described as the proportions of predicted negative outcomes among those images are tested as original.

TNR is described in Eq. (16),

$$TNR = {T_n}\Big/{T_n + F_p} \quad (16)$$

4) False positive rate (FPR): The FPR is determined by dividing the total quantity of forgery images by the number of forgery images that were mistakenly classified as original images (false positives).

TNR is described in Eqn. (17),

$$FPR = {F_p}\Big/{T_n + F_p} \quad (17)$$

5) False negative rate (FNR): The likelihood that a true positive would be overlooked by the sample is measured by the FNR, also known as the miss rate. The FNR seems to be the percentage of positive test results that result in failed tests. The FNR has been computed as the difference between the number of altered images that were incorrectly identified as true images and the total number of altered images that were incorrectly identified as altered images as well as the number of altered images that were correctly identified as altered images.

FNR is described in Eq. (18),

$$FNR = {F_n}/{T_p + F_n} \tag{18}$$

6) F-measure: The effectiveness of the procedures is demonstrated through a pixel-based assessment that involves determining whether or not each individual pixel is false. The proposed as well as considered approaches are evaluated in terms of performance using the F-measure. Its improved performance in identifying the CMF in digital images is shown by a higher F-measure outcome.

F-measure is described in Eq. (19),

$$F - measure = 2 \times {T_p}/{2 \times (F_p + F_n + T_p)} \tag{19}$$

7) Testing time (TT): Testing Time (TT) has been further employed to assess the presented model and compare it to other models. TT seems to be the average amount of time required to evaluate the images for the specified number of runs (k). Additionally, as this step has only been completed offline once, Learning Time has been not taken into account.

B. Results

The research introduced a novel DL (CNN) method for the procedure of detecting image forgery. The suggested model combines a CNN model with a hybrid optimization of GWO and ABO. The suggested DL model has been tested using the MICC-F600, MICC-F220, and SATs-130 databases, among other datasets. The goal of the computational experiments has always been to develop the best model possible in terms of complexity and TT. Moreover, the suggested method had been tested using the k-fold cross-validation method. This framework starts with a learning phase that will be repeated (k) times to get a diversity of the images being looked at and provide accurate estimation by thoroughly looking through the datasets. The dataset is further divided into (k) groups (folds) with almost the similar dimension in this approach.

The suggested model employed (k-1) the residual categories for testing and the classes for training. For both testing and training there are (k) iterations. The suggested approach made advantage of a method for 5-fold cross-validation. Accordingly, for each of the five rounds, 30% of the dataset's images were utilised for testing while the remaining 70% were randomly chosen for training. Each iteration will employ a different 30% of the images instead of the previous 30% of images for assessment.

TABLE II. PRESENTED MODEL PERFORMANCE @ 25 EPOCHS

Metrics	**Datasets**		
	MICC-F600	***SATs-130***	***MICC-F220***
Accuracy (%)	95	92	97
TPR (%)	90.4	84	100
F-measure (%)	93.98	88.79	91.55
FNR (%)	28.5	6.7	0
TT (sec)	1.23	5.02	1.24
TNR (%)	85.19	77.5	98.2
FPR (%)	5.7	10.4	10.8

Five trials at training epochs of 25, 75, 50, and 100 make up the test. At each epoch, the TPR, FNR, F-measure, TNR, FPR, and accuracy are recorded. In the computational experiments, a recording of the TT has been also taken into account. The simulation outcomes for every epoch for all databases are presented in Tables II to V.

TABLE III. PRESENTED MODEL PERFORMANCE @ 50 EPOCHS

Metrics	**Datasets**		
	MICC-F600	***SATs-130***	***MICC-F220***
Accuracy (%)	97.33	93.04	91.45
TPR (%)	97.5	93.97	100
F-measure (%)	93.4	88	82
FNR (%)	34.4	1.6	0
TT (sec)	1.31	4.25	1.25
TNR (%)	87	93	97
FPR (%)	10.1	5.6	4.3

It is evident that the implementation of the suggested model grows as the number of epoch’s increases. Additionally, how the suggested technique is implemented varies from one database to another. Therefore, the suggested model cannot be implemented due to two limits. The first constraint seems to be the number of training epochs, and the other is the databases construction, based on which the suggested model has been used.

TABLE IV. PRESENTED MODEL PERFORMANCE @ 75 EPOCHS

Metrics	**Datasets**		
	MICC-F600	***SATs-130***	***MICC-F220***
Accuracy (%)	94.4	97.8	98.7
TPR (%)	100	97.5	100
F-measure (%)	98.2	98.2	99.05
FNR (%)	9.7	0	0
TT (sec)	1.23	4.14	1.27
TNR (%)	97.6	96.25	98.54
FPR (%)	1.7	5.4	3.2

The effectiveness metric variable findings for the various four databases at epochs 25, 50, 75, and 100 are shown in Tables II, III, IV, and V, correspondingly. The findings shown in the tables show that after 100 epochs, F-measure, TPR, accuracy, and TNR increase by over 100%. In contrast, FNR and FPR decreased almost to zero at the same time that the number of epochs reached 100. Additionally, as the number of epochs increased, the TT decreased to a minimum.

TABLE V. PRESENTED MODEL PERFORMANCE @ 100 EPOCHS

Metrics	Datasets		
	MICC-F600	***SATs-130***	***MICC-F220***
Accuracy (%)	98.2	100	100
TPR (%)	98.6	99	100
F-measure (%)	99.9	100	100
FNR (%)	2.2	0	0
TT (sec)	1.24	1.32	1.132
TNR (%)	100	98.05	100
FPR (%)	1.04	2.84	0

It is clear from the previous findings that the number of training components has a significant impact on the performance of the entire model. Additionally, it is noted that 100 epochs yield the greatest results. The suggested model included a number of identifiers that might affect the results in terms of accuracy, F-measure, or TT. These variables include the volume of data utilized for training (which is connected to the utilized databases dimension), the volume of data employed as inputs to the network, the volume of hidden prototype layers, and the volume of the chosen epochs. The Identifiers for how much data are employed for training, how much data is employed as input, and how much data is employed for the network's hidden layers all remain constant. The number of the chosen epochs seems to be the only identifier left that might affect testing time.

TABLE VI. PRESENTED MODEL'S OPTIMAL PERFORMANCE

Datasets	Metrics		
	Accuracy (%)	***F-measure (%)***	***TT (sec)***
MICC-F600	98.2	99.9	1.24
SATs-130	100	100	1.32
MICC-F220	100	100	1.132

It was discovered that the number of epochs chosen has a significant effect on how long the method takes to reach the optimal state feature map. The method can obtain the best performance by taking the time necessary to produce the appropriate feature map. The optimum feature map would be retrieved by choosing a precise number of epochs. Along with the number of epochs, interpolation will occur. In addition, the model will perform worse if fewer epochs than the chosen number are used, also failed to obtain the optimal feature map. As a result, if the other identifiers have been kept constant, the number of chosen epochs affects the TT. The optimal performance of proposed model in CMF detection on different databases is shown in Table VI.

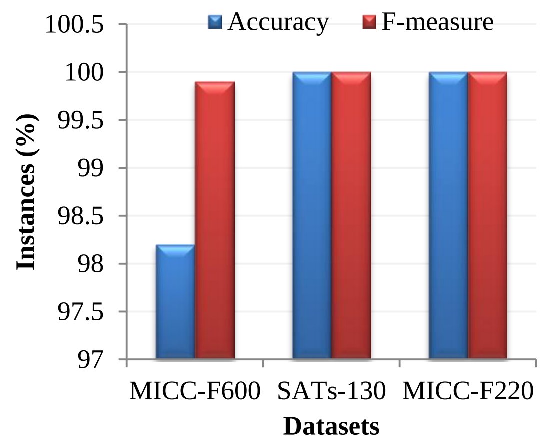


Fig. 4. Different dataset's optimal accuracy and F-measure of detection

The best accuracy and F-measure representation for each database is shown in Fig. 4. According to the simulation findings, the suggested algorithm is a valid and practical method for CMF detection. Moreover, the presented model's MICC-F220 dataset's accuracy was compared with other existing models like Speeded Up-Robust Feature (SURF) [8], Spatial features-based Image CMF detection (IC-MFDs) [23], CNN [24], and Convolutional Long-Short Term Memory (ConvLSTM) [25].

TABLE VII. MICC-F220 DATABASE ACCURACY COMPARISON

Method	Accuracy (%)
SURF	91.95
IC-MFDs	98.44
CNN	100
ConvLSTM	100
Presented (GWO-ABO-CNN)	100

The MICC-F220 dataset accuracy comparison of different models is shown in Table VII and Fig. 5. The comparison outcome indicated that the CMF detection performed by CNN model has attained 100% accuracy.

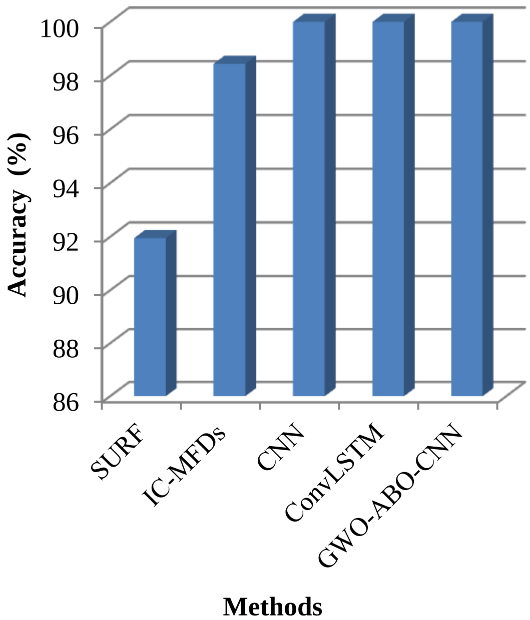


Fig. 5. Comparison of accuracy

The accuracy attained by the model SURF was 91.95%, IC-MFDs was 98.44%, CNN was 100%, ConvLSTM was 100%, and the presented GWO-ABO-CNN was 100%. The models with higher accuracy are reliable for CMF detection. The main problem in the CMF recognition is less accuracy on challenging database. From the current research results, it is evident that the presented GWO-ABO-CNN method detects CMF in challenging datasets like SATs130, MICC-F220, and MICC-F600 with 100% accuracy. Thus, the presented method is reliable for CMF recognition in digital images.

V. Conclusion

This research presents a deep learning system for CMF detection predicated on hybrid optimization GWO-ABO with CNN (GWO-ABO-CNN). The main goal of this study is to construct and enhance the DL classification framework for identifying real and forged classes in alleged digital image forgeries. The suggested approach anticipates creating a new paradigm that will offer improved performance, demand less testing time, and incur little computational expense. Four layers—pre-processing, feature extraction, feature matching, and post-processing (detection)—are addressed in the creation of the proposed forgery detection method. GWO-ABO has been used to unify the picture size and turn the images into tensors in the pre-processing layer. The suggested method is predicated on a novel invention that builds a serial series of convolutional layers and a pooling layer to speed up the detection process in the feature extraction layer. Several challenging datasets, including the SATs130, MICC-F220, and MICC-F600 datasets, have been employed in the evaluation process. The test findings revealed that 100 training epochs had produced the best accuracy.

References

[1] P. Jain, “Shaina shainawadhwa79@gmail.com Adesh College of Engineering and Technology, Faridkot, Punjab,” p. 6, 2019.

[2] A. H. Saber, M. Khan, and B. Mejbel, “A Survey on Image Forgery Detection Using Different Forensic Approaches,” Adv. Sci. Technol. Eng. Syst. J., vol. 5, pp. 361–370, Jan. 2020, doi: 10.25046/aj050347.

[3] W. D. Ferreira, C. B. R. Ferreira, G. da Cruz Júnior, and F. Soares, “A review of digital image forensics,” Comput. Electr. Eng., vol. 85, p. 106685, Jul. 2020, doi: 10.1016/j.compeleceng.2020.106685.

[4] Vivek Singh, Neelesh Kumar Jain, and Jaypee University, guna(M.P.), “Digital Image Forensics in Multimedia Security: A Review,” Int. J. Eng. Res., vol. V4, no. 05, p. IJERTV4IS051057, May 2015, doi: 10.17577/IJERTV4IS051057.

[5] A. Ferreira, T. Carvalho, F. Andaló, and A. Rocha, “Counteracting the contemporaneous proliferation of digital forgeries and fake news,” An. Acad. Bras. Ciênc., vol. 91, no. suppl 1, p. e20180149, 2019, doi: 10.1590/0001-3765201820180149.

[6] M. A. Elaskily et al., “A novel deep learning framework for copy-moveforgery detection in images,” Multimed. Tools Appl., vol. 79, no. 27–28, pp. 19167–19192, Jul. 2020, doi: 10.1007/s11042-020-08751-7.

[7] H. Chen, X. Yang, and Y. Lyu, “Copy-Move Forgery Detection Based on Keypoint Clustering and Similar Neighborhood Search Algorithm,” IEEE Access, vol. 8, pp. 36863–36875, 2020, doi: 10.1109/ACCESS.2020.2974804.

[8] A. Badr, A. Youssif, and M. Wafi, “A Robust Copy-Move Forgery Detection In Digital Image Forensics Using SURF,” in 2020 8th International Symposium on Digital Forensics and Security (ISDFS), Beirut, Lebanon, Jun. 2020, pp. 1–6. doi: 10.1109/ISDFS49300.2020.9116433.

[9] S. Dhivya, J. Sangeetha, and B. Sudhakar, “Copy-move forgery detection using SURF feature extraction and SVM supervised learning technique,” Soft Comput., vol. 24, no. 19, pp. 14429–14440, Oct. 2020, doi: 10.1007/s00500-020-04795-x.

[10] H. Huang, W. Guo, and Y. Zhang, “Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm,” in 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Wuhan, China, Dec. 2008, pp. 272–276. doi: 10.1109/PACIIA.2008.240.

[11] A. Parveen, Z. H. Khan, and S. N. Ahmad, “Block-based copy–move image forgery detection using DCT,” Iran J. Comput. Sci., vol. 2, no. 2, pp. 89–99, Jun. 2019, doi: 10.1007/s42044-019-00029-y.

[12] K. B. Meena and V. Tyagi, “A copy-move image forgery detection technique based on Gaussian-Hermite moments,” Multimed. Tools Appl., vol. 78, no. 23, pp. 33505–33526, Dec. 2019, doi: 10.1007/s11042-019-08082-2.

[13] M. Kumar, A. Rani, and S. Srivastava, “Image Forensics Based on Lighting Estimation,” Int. J. Image Graph., vol. 19, no. 03, p. 1950014, Jul. 2019, doi: 10.1142/S0219467819500141.

[14] H. Kasban and S. Nassar, “An efficient approach for forgery detection in digital images using Hilbert–Huang transform,” Appl. Soft Comput., vol. 97, p. 106728, Dec. 2020, doi: 10.1016/j.asoc.2020.106728.

[15] H. Wang and H. Wang, “Perceptual Hashing-Based Image Copy-Move Forgery Detection,” Secur. Commun. Netw., vol. 2018, pp. 1–11, 2018, doi: 10.1155/2018/6853696.

[16] A. Kuznetsov, “Digital image forgery detection using deep learning approach,” J. Phys. Conf. Ser., vol. 1368, no. 3, p. 032028, Nov. 2019, doi: 10.1088/1742-6596/1368/3/032028.

[17] R. Shao and E. J. Delp, “Forensic Scanner Identification Using Machine Learning,” in 2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), Albuquerque, NM, USA, Mar. 2020, pp. 1–4. doi: 10.1109/SSIAI49293.2020.9094618.

[18] N. Kanwal, A. Girdhar, L. Kaur, and J. S. Bhullar, “Detection of Digital Image Forgery using Fast Fourier Transform and Local Features,” in 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, United Kingdom, Apr. 2019, pp. 262–267. doi: 10.1109/ICACTM.2019.8776709.

[19] H.-Y. Huang and A.-J. Ciou, "Copy-move forgery detection for image forensics using the superpixel segmentation and the Helmert transformation," EURASIP J. Image Video Process., vol. 2019, no. 1, p. 68, Dec. 2019, doi: 10.1186/s13640-019-0469-9.

[20] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-Based Forensic Method for Copy–Move Attack Detection and Transformation Recovery," IEEE Trans. Inf. Forensics Secur., vol. 6, no. 3, pp. 1099–1110, Sep. 2011, doi: 10.1109/TIFS.2011.2129512.

[21] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An Evaluation of Popular Copy-Move Forgery Detection Approaches," IEEE Trans. Inf. Forensics Secur., vol. 7, no. 6, pp. 1841–1854, Dec. 2012, doi: 10.1109/TIFS.2012.2218597.

[22] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, L. Del Tongo, and G. Serra, "Copy-move forgery detection and localization by means of robust clustering with J-Linkage," Signal Process. Image Commun., vol. 28, no. 6, pp. 659–669, Jul. 2013, doi: 10.1016/j.image.2013.03.006.

[23] I. T. Ahmed, B. T. Hammad, and N. Jamil, "Image Copy-Move Forgery Detection Algorithms Based on Spatial Feature Domain," in 2021 IEEE 17th International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, Mar. 2021, pp. 92–96. doi: 10.1109/CSPA52141.2021.9377272.

[24] M. A. Elaskily et al., "A novel deep learning framework for copy-moveforgery detection in images," Multimed. Tools Appl., vol. 79, no. 27–28, pp. 19167–19192, Jul. 2020, doi: 10.1007/s11042-020-08751-7.

[25] M. A. Elaskily, M. H. Alkinani, A. Sedik, and M. M. Dessouky, "Deep learning based algorithm (ConvLSTM) for Copy Move Forgery Detection," J. Intell. Fuzzy Syst., vol. 40, no. 3, pp. 4385–4405, Mar. 2021, doi: 10.3233/JIFS-201192.

Event Detection and Classification Using Deep Compressed Convolutional Neural Network

K. Swapnika[1]*, D. Vasumathi[2]
PhD Scholar[1], Professor[2]
Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, Telangana 500085, India[1, 2]

Abstract—Recently, the number of different kinds of events on social media platforms show a tremendous increase in each second. Hence, event detection holds a very important role in the current scenario. However, event detection is challenging in information technology (IT). Several machine learning-based approaches are established for the event detection process, but it generates a high error and makes various information loss, affecting the system's performance. Thus, the proposed work introduces a new detection strategy based on a deep learning architecture. In this, both text and image data are utilized for event detection. The different procedures for image and text databases are pre-processing, extraction and classification. The text data is pre-processed using four methods: lower case filter, tokenization, stemming, and stop word filter. An adaptive median filter (AMF) is utilized for pre-processing the image data. After the pre-processing stage, feature extraction is performed for text and image-based data in which most useful features are extracted. Finally, the varied events are detected and classified using the proposed Deep Compressed Convolutional Neural Network (DCCNN). The entire work is implemented using the PYTHON platform. The efficiency of the proposed model is measured by evaluating the performance metrics such as accuracy, recall, precision and F-measure. The simulation validation exhibits that the proposed classification method attains an improved accuracy of 97.1%, obtained precision is about 95.06%, recall value is 91.69%, and f-measure is 93.35%. The efficacy of the proposed deep learning method is proved by comparing the attained results with various state-of-the-art techniques.

Keywords—Event detection; erosion; dilation; deep learning; deep compressed convolutional neural network; hashing; median filter

I. Introduction

In digital societies, the demand to express oneself and communicate with others is becoming more prominent. In many sectors of smart societies, this is being accomplished through digital platforms such as social media, which have increasingly become easy and inexpensive sensors compared to physical sensors [1-3]. The term "big data" is used to demonstrate the trend of data explosion. Many research and surveys try to describe what big data means. Each work put forward a variety of viewpoints in terms of background, opportunities, applications, and challenges. Digital data, created by many digital devices, rises at breakneck speed [4-6]. The McKinsey Global Institute (MGI) offered a detailed analysis of big data in terms of three various parameters such as innovation, competition and productivity. Different ML techniques show better results in big data processing applications from the last decades. Also, the ML approaches provide improved detection results in road traffic applications [7] as the size and number of data generated increase daily, handling these data using traditional learning methods is very difficult.

Data generated by social media users is massive in volume, rapidly expanding, diverse in nature, and variable quality. These are the most important elements that define big data. Because of its immense popularity, the subject of event detection has recently drawn the attention of researchers. The big data age has arrived due to the abundance of data in virtually every area of our society [8-10]. It is a hot topic that affects many facets of our lives. Twitter has been used as a valuable data source in various situations [11, 12]. There has been a growing body of research on using Twitter as a sensor for traffic monitoring, flow forecasting, congestion prediction, and incident detection in recent years. These techniques have a lot of promise in this field. Event recognition from unstructured, quickly developing tweets is a difficult task from data mining. Many existing works developed different approaches for twitter-based event detection [13-15]. The volume, pace, variety, and authenticity of Twitter data are all hallmarks of big data. As a result, managing and analyzing Twitter data for event detection is a huge difficulty. Deep learning uses a vast amount of data to learn how features behave during training and predicts the class of data that has never been seen before. The proposed approach will be processed on both text and image datasets to analyze and improvise the performance.

Event detection is a significant area in fundamental language processing that mainly focuses on automatic event prediction [16]. Generally, events are the particular actions performed in a certain period. The events are mainly represented in varied lengths of expressions from the collections of text documents. The certain event detection is based on the features, and the particular data are extracted from the text [17]. The detection and classification of events in texts and images play an important role in several applications. An effective event detection helps to analyze which event has occurred. Various techniques have been developed for event classification in the past few years. Some of the classification approaches are failed to classify the exact events because of the high computational complexity. Mainly, most traditional approaches are not suitable for high data dimensionality. Also, the classification performance model is affected due to the large over fitting issue.

*Corresponding Author.

Moreover, the existing works are troubled to process two different data types like image or text. Hence, event detection with text and image data is difficult in image processing. Thus, an effective methodology for accurate event detection and classification in image processing is necessary. Nowadays, deep learning approaches play a significant role in event classification and also it is highly appropriate for big data applications [18-20]. For this reason, the proposed work concentrates on deep learning techniques for event detection and classification using text and image-based data.

A. Motivation

Event detection using big data plays a very important role in a wide variety of applications. Many researchers use a wide variety of strategies to detect big data events. This work introduces a new deep learning-based technique for the efficient detection and classification of events. The prevalence of big data creates several challenges for the usual feature selection task. Scalability and stability are two issues in feature selection for big data analytics. A minimum amount of work can be done to detect events using Twitter. Big data techniques are critical since they allow for the system's scalability in data analytics and control.

DL can improve the accuracy and efficiency of the event detection process, which would be extremely beneficial. Unsupervised feature learning approaches and deep learning have been widely deployed for image and text applications. These strategies have shown a lot of promise in these domains for autonomously expressing the feature space using unlabelled data to improve the accuracy of subsequent classification tasks.

B. Contribution

The main contributions of the work are summarized below.

- To achieve an efficient event detection process using a compressed deep learning architecture, named deep compressed convolutional neural network.
- To implement the event detection approach for both image and text data, also improvise the overall performance.
- To utilize efficient text and image data pre-processing approaches to get accurate results for the event detection process using big data.
- To evaluate the performance of the event detection using statistical measures such as accuracy, recall and f1-score, and compare it with the existing performance.

C. Organization

The concept is briefly explained in Section I. Some of the recent related work is explained in the Section II. The proposed event detection approach is elaborated in Section III. The implementation results and the dataset description are provided in Section IV. Finally, the conclusion and future scope are provided in Section V.

II. Related Works

Some of the existing works carried out on event detection using different approaches are discussed below.

Cherichi et al. [21] developed a real-time event detection methodology using big data analytics in microblogs. In this existing work, a new metric is developed to enhance the outcome of microblogs searches. Here, the system is processed by executing five modules: lexical analysis, morphological analysis, syntactic analysis, extraction and interpretation module. Initially, the tweets are pre-processed by performing tokenization, splitting the sentence and normalization. After the completion of the pre-processing stage, feature extraction is enabled. The feature extraction is performed with the assists of Natural Language Processing (NLP) by integrating the author's relevance, content relevance, and the relevance of the tweet. Finally, the event information is interpreted in the last stage. The established technique depends upon the framework of temporal marker classes and the contextual searching approach. A knowledge management system is constructed for measuring the developed approach. This existing work cannot obtain better results because of the high correlation between event information in tweets.

Yadav et al. [22] developed signal energy transformations for analyzing various events in real-time. A synchrophasor data is utilized for detecting multiple events in power systems. The events are detected using the Teager-Kaiser energy operator (TKEO) approach. The mentioned technique is also responsible for temporal localization. This approach is invulnerable to large amplitude oscillation in the data, and it provides the extent of time resolution. The events are classified by introducing an energy similarity measure (ESM) to differentiate the events with improved accuracy performance. This developed model categorizes different events with unequal severity and reduced time gaps. For real time, this approach utilizes only 10 samples of data windows. Hence, this technique is appropriate for sensing the events in power systems with better reliability. But, it faces increased computational complexity, which can disrupt the system's entire performance.

Wang et al. [23] established an extended R-FCN model for event detection based on audio. The developed framework involves two phases for event prediction. Initially, the convolutional kernels in the time axis are slide to identify the presence of audio events. Then, the proposals which feasibly involves audio events are produced by utilizing region proposal networks (RPN). In the next phase, the domain: frequency and time are combined with categorizing the proposals, and the boundaries are refined. This developed approach utilizes spectrogram features from the audio signals as the input. This existing work uses ResNet50, ResNet101 and enhanced ResNet50 methods to make real-time outcomes. This approach helps to make the system process easier. The accurate position of the audio event is detected at first, and it cannot consider the event class of audio in the initial phase. Therefore, it generates a binary classification issue. Thus, the process is turned into the second phase in which the proposal is classified, and the boundaries are refined. The developed approach is better for event detection, but it is computationally expensive.

Zhang et al. [24] suggested a deep learning-based approach for detecting events with images and text from social media. This existing study develops an approach named collective action events from social media (CASM). The main intention of this model is to detect the collective action events utilizing the data from social media. Both image and text data are used for the event detection process. For processing the image data, a convolutional neural network (CNN) is utilized, and for text data, a recurrent neural network (RNN) with long short-term memory (LSTM) is used. This study uses a two-stage classification model for social media posts event prediction. The developed CASM is executed on Chinese social media data and detects numerous action events from 2010 to 2017. This deep learning approach provides better classification results in event detection for image and text-based data.

Chen et al. [25] developed a semi-supervised deep learning approaches for traffic event detection using multi-modal data. For traffic event detection, the suggested work uses big data applications. The input data is encoded into numerical vector forms using feature learning components in multi-modal data. The multiples of data are processed simultaneously in a neural network and combine the features of each varied form of input data. The experimental results represent that the suggested technique is not suited for different classes of events. Also, this approach takes high learning time in the training process.

Alomari *et al.* [26] introduced a new approach for road event detection. This paper also put forward a new and efficient big data tool named Iktishaf. It is developed with the help of Apache Spark to detect events related to traffic data. This existing study develops several classifiers using three ML technologies to detect eight different kinds of events. The classifiers were tested against a variety of criteria and other sources. Text pre-processing, event detection, and feature space are all improved with Iktishaf Stemmer. The pre-processing phase minimizes the grammatical errors and mistakes in an Arabic dialect text. Also, this stage diminishes the noises and typos in the input data. Tokenizer and normalizer are performed to improve the data quality in the initial stage. The features are extracted from the pre-processed data using the term frequency-inverse document frequency (TF-IDF). In the end, the event is classified by using the tweets filtering component (TFC). This study faces reduced accuracy, scalability, functionality, and data management issues. Thus, developing an effective technique for event detection is important for obtaining better outcomes.

A. Research Gap Analysis

As surveying several existing works, event detection in image processing is critical for various techniques. Many of the previous works are not applicable for multi-modal applications in big data. In some cases, the traditional approaches fail to perform in large data dimensionality and make the system performance a minimum. The popularity of big data creates several challenges for the usual feature selection task. Scalability and stability are two issues in feature selection for big data analytics. A minimum amount of work can be done to detect events using Twitter. The techniques for detecting events must be exact. The retrieved data must be accurate in order to make better decisions. But many of the existing systems provide less accuracy for event detection procedures. Another limitation of the existing methods is difficulty in the run time and computational complexity. Also, these works utilized only one type of data. The previous techniques did not provide better results for event detection using text and image data. The effective event detection technique has not yet been developed to support multiple input data. So it is better to use multiple types of data to show the method's efficiency. Big data techniques are critical since they allow for the system's scalability in data analytics and administration. Numerous works use big data to detect events. However, several of the studies focused solely on text data. As a result, this research considers both image and text data.

III. Proposed Methodology

Event detection using big data holds a very important role in several applications. This work introduces event detection by the Deep compressed Convolutional Neural network (DCCNN). Here, utilizing both text and image data for processing the event detection approach. Different stages include pre-processing, feature extraction, and classification in-text data processing. The workflow of the proposed approach is shown in Fig. 1.

The different pre-processing steps are tokenization, lower case filter, stop word filter and stemming. Then, feature extraction is done. For the text analysis to be perfect, feature extraction needs to be done, making the process easier. It helps to accurately deduct the data from the vast dataset without losing important information. The speed of the processing technique is increased due to the feature extraction stage. Here this process is achieved with the help of glove and feature hashing approaches. Finally, the classification of text data is carried out with the help of DCCNN. Next is image data processing. Initially, the image data is forwarded for the pre-processing stage to remove noise and other redundant information.

An adaptive median filter is used for noise removal. Then, morphological operations such as erosion and dilation are applied for removing redundant data from the image. Then, it is sent to the DCCNN. The proposed DCCNN can perform extraction, selection and classification of image data. This technique is proposed to learn hierarchical features of data from an entire feature extraction technique. This combination makes DCNN input feature reduction (selection) with complexity reduction. While applying compression on DCNN will lead to feature reduction. In this way, the architecture reduction is achieved. This results in lower latency, as well as learning time, becomes faster.

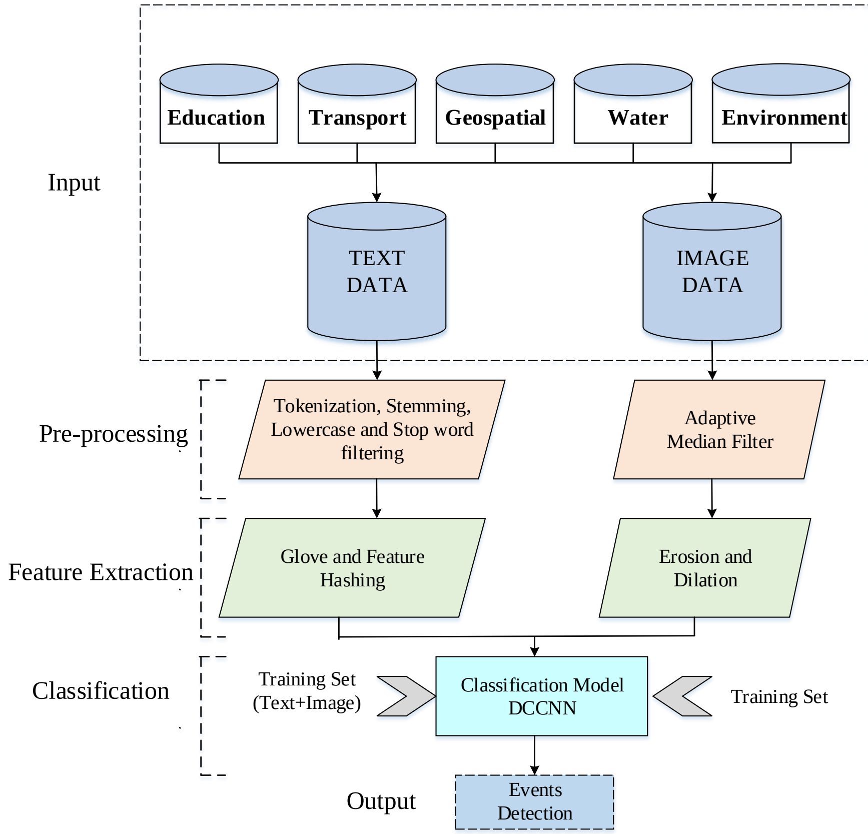


Fig. 1. Workflow of the proposed methodology.

B. *Text Data Processing*

1) Pre-processing: The initial step considered here is the pre-processing. The data usually contains the presence of noise, redundant data and some other irregularities. Hence it should be pre-processed before extraction. The collected data contains several redundant information as well as noise. Here, four well-known text data pre-processing approaches are utilized: tokenization, lower case filter, stop word filter and stemming. Applying the tokenization procedure, a sentence or a paragraph can be split into smaller units. The stemming process can be applied to effectively reduce words into their stem by chopping off the ends of words and often by removing derivational affixes. Stop words usually indicates the words like 'a' 'an' 'the' 'in' etc. and are removed in the pre-processing stage.

2) Feature extraction: The feature extraction (FE) process aims to remove the redundant information from the text data. Here, it is accomplished with the help of glove and feature hashing approaches.

The glove is the method used to represent the words in their vector form. The words are mapped into space to find their vector representation. In this technique, the distance between words is related to semantic similarity. The glove can be explained by considering a matrix of word-word co-occurrence counts(Y). The entries (Y_{ji}) of the matrix tabulate how many times a word i occurred in the context of the word j. The number of words that appear in a context j is denoted by Y_j and is given by the following expression.

$$Y_j = \sum_l Y_{jl} \tag{1}$$

The probability of word i which appears in the context of the word j is given by,

$$\boldsymbol{P_{ji} = P(i / j) = Y_{ji} / Y_j} \tag{2}$$

Feature hashing (Fh) is considered the simplest technique for FE purposes, and it is also termed a hashing trick. The reduced feature space can be accomplished by simply applying a hashing function in this approach. A second hash function is used here to indicate the sign of the feature set obtained, and it can also be possible to remove the collision of feature vectors. The hashing function applied for FE is shown below.

$$\vec{F}_i^{(l)} = \sum_{F \in A_l; h(F=i)} \chi(F^l) \sum_{i=1}^{n(l)F^l} x_i^{(k,A)} \tag{3}$$

Where $\vec{F}_j^{\,l}$ indicates the reduced feature vector, and the l^{th} word is represented using F^l. The context of l^{th} word is indicated using A_l. The hashing function is represented using h and the second hashing function is χ. The parameter $n(l)F^l$ is the number of times a word appears. The weighting of the word A based on the word l is represented as $x_i^{(k,A)}$.

C. Image Data Pre-processing

1) Pre-processing: Usually, the image data consist of different irrelevant information along with noises. Hence it needs to be pre-processed to achieve better classification results. In previous studies, several techniques are used to pre-process image data in the event detection mechanism. The proposed work is accomplished with the help of an adaptive median filter (AMF). The AMF approach removes the noise or other redundant information from the image dataset. AMF is considered a very efficient filter for noise elimination as well as it preserves the edges of the image. The AMF mainly works on two stages. In the first stage, AMF identifies the median value for the kernel. In the second stage, AMF checks whether the current pixel is noised or not. If noise is detected, each pixel's value is replaced with the median values of its neighbourhood, as stated in the equation below.

$$Q(x, y) = median\, p(i, j) : (i, j) \in \alpha \quad (4)$$

Where (x, y) represents the image's location, and its surrounding neighbourhood points are denoted using α.

2) Extraction: The feature extraction stage is an important stage in which the most discriminative information is extracted. This stage aids to improve the classification accuracy of the proposed model. In the proposed work, the image features are extracted with the help of a technique named erosion and dilation. These erosion and dilation are the two basic morphological operators that aim to extract relevant attributes from the image database. It can be achieved by probing the image with another set, known as structuring element (SE) or kernel. This operator provides varied outcomes when executed to binary or gray-scale images. In an image, the dilation process includes pixels to the boundaries of objects, while erosion eliminates the pixels on the boundaries of objects. The number of pixels included or eliminated from the image object is based on the element's shape and size utilized to process the image. Each pixel of an image is minimized by performing the process of erosion. The erosion operators acquire a pair of data as the input, and the SE mentions the accurate effect of the erosion performance on the input pre-processed image. Generally, the dilation process is applied to fill the gaps in the most important pixels, which is useful for classification. The erosion process is utilized to ignore the unessential features of pixels in an image. The extraction of erosion and dilation related features will enhance the classifier's performance. Let $Er_B(A)$ denotes an erosion of set A by the element B, and the expression is given below.

$$Er_B(A) = \{A | B_A \subseteq A\} \quad (5)$$

The dual operator of erosion is termed dilation. The dilation operation can be indicated as $Di_B(A)$, which is caused by the element B on the set A.

$$Di_B(A) = \{A | B_A \cap A \neq 0\} \quad (6)$$

In this process, the features are extracted using morphological operations with the aid of erosion and dilation. This erosion and dilation based features are fed to the input of the classification stage.

D. Classification

The events are classified with the help of a deep learning architecture termed as deep compressed convolutional neural network (DCCNN). DCCNN can process both image and text data. The architecture of DCCNN is illustrated in the following Fig. 2.

Fig. 2 represents the architecture of the proposed DCCNN. The basic CNN method is considered to be efficient neural network architecture, and it utilizes the spatial properties of the input. In the training process, the input image features are passed throughout the DCCNN structure, and then the weights are updated in the backpropagation. The architecture contains different layers such as the input, convolutional, pooling, compression, and output layers. In the input layer, the image-based features are subjected as the input, and it transfers the input data to the convolutional layers. The convolutional layers can extract different kinds of features from the input images. Moreover, the convolutional layers extract features' large dimensionality from the images. This high dimensionality of features is minimized in the pooling layer. Then, it again transferred to convolutional and pooling layers. In the proposed work, the compression layer is included in the deep CNN technique in order to compress both the features of image and text data. In the compression layer, the features from text data are given as the input. This compression layer provides a compressed output, and it helps to improve the classification accuracy. During the process of training, the image and text-based data are updated. The initial layer of the proposed DCCNN architecture is the convolutional layer, and it contains filters to extract the input features. Initially, the image data is applied at the convolution layer of DCCNN. Hence the applied image data is further extracted with the help of this convolution layer. Let $l = (l_0, l_1, \ldots, l_{n-1}) \in R^n$ be an input vector having n components. The convolutional layer gives an output, $m = (m_0, m_1, \ldots, m_{n-1}) \in R^n$:

$$m_i = \sum_{j \in N_i} s_j l_j \quad (7)$$

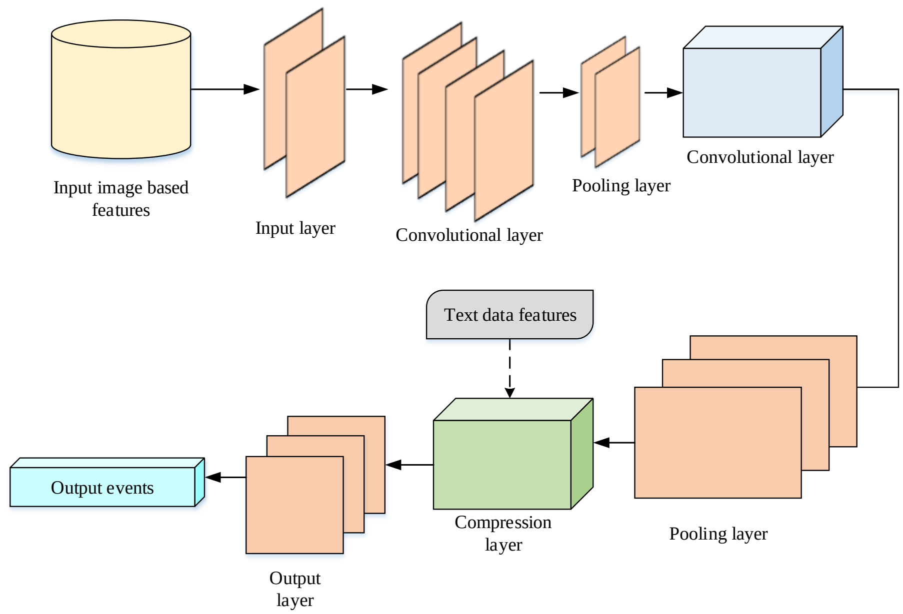


Fig. 2. Architecture of DCCNN.

Where weight is represented using s_j, and N_i represents indices in the local receptive field at l_i. The weight s_j is said to have a constant input of 1. Hence, each output m_i is a weighted sum of neighbours $\sum_{j \in N_i} s_j l_j$ plus constant. Equation (8) is the convolution of l_i using the filtering kernel s_j, and rewriting this equation using convolutional operator $*$ gives,

$$m = l * s, \tag{8}$$

Where, $s = (s_0, s_1, \ldots, s_{c-1}) \in R^c$. The convolutional layers use different weight parameters to process input data, and the resulted output is a concatenated vector. Then it is passed to the next layer (i.e., pooling layer). The pooling layer can select the suitable features from the image dataset. Hence this layer can also be termed the feature selection layer of DCCNN. It can have the ability to generate in variance to small transformations, and it divides the disjointed regions having a size of $S * S$. If the input feature size is applied to the pooling layer is $P * P$, and the following expression gives the resulting output.

$$E_{pooling} = \frac{P}{S} \tag{9}$$

After the pooling/feature selection layer, the compression layer is placed. It is the most important layer in this architecture. The selected image features from the pooling layer are passed to the compression layer, and the processed text features are applied directly to the compression layer. This layer can combine the text and image features related to different events and compress the size of features. The useful features of the compression layer can reduce the computational complexity of the proposed event detection process. This layer forwarded the compressed output to the output layer. This layer usually comprises of softmax classifier. The following equation represents the output of the softmax classifier.

$$S_i = \frac{e^{l_k}}{\sum_{j=1}^{m} e^{y_j}}, k = 1, \ldots\ldots\ldots, m \quad y = y_1, \ldots., y_m \tag{10}$$

Where S_i indicate the outcome of the softmax classifier. l_k denotes the factor of the input vector l, and m indicates the total number of neurons present in the output layer. This layer is responsible for suitably classifying the events. The output contains both image and text data of the same event. The classification stage of the proposed work is depicted in Fig. 3.

The event detection is achieved with the help of image and text data. The proposed model attains the training accuracy of 80% and obtains the testing accuracy of 20%. The data is forwarded for pre-processing and extraction. Further processing and selection are accomplished with the help of the DCCNN network. It is considered one of the best neural networks for the classification process. Here the convolution layer is utilized for extraction, and the selection is made with the help of the pooling layer. Here a new compression layer is introduced to achieve the efficient result and reduce complexity. The text data is applied directly to the compression layer. Finally, the softmax layer obtains the result, containing a single event's compressed and combined output.

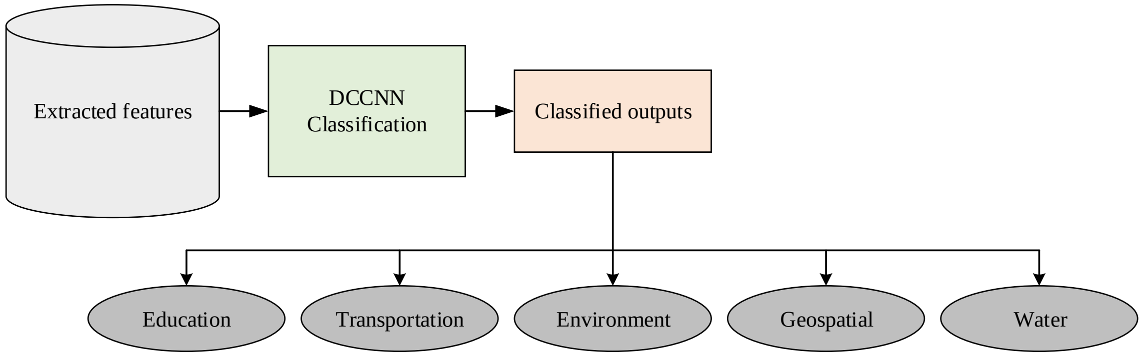


Fig. 3. Classification stage of the proposed method.

IV. Result and Discussion

The proposed DCCNN network is evaluated based on image and text data, and it is implemented with the help of the PYTHON platform. Here seven text data and one image data are utilized. The image data is created using the images available on the Internet. The dataset explanation and the evaluation metrics computation are explained in the following sections.

A. Dataset Description

The proposed methodology uses five text databases and one image database. The image dataset is created by using the images available on the Internet. Various kinds of images are available on the Internet, and the images suitable for event detection are collected for the proposed event detection process. The proposed work collects the Internet images related to the environment, transport, geospatial, water and education. In that, education dataset belongs to 22 images, and the environment consists of 76 images, geospatial have 100 images, transport has 250 images, and water contains 300 images. Among these, 80% of data is used for training and 20% for testing. The text data for the event detection process is attained from the data.world dataset. This dataset contains 127708 different datasets for varied applications like transportation, environment, statistics, atmosphere etc. This dataset has obtained great attention for event prediction in different fields in the past few years. The proposed work gathers the text data from education, transportation, water, geospatial and environment using data.world dataset. The data from both text and image datasets are employed to experiment with event detection in the proposed study. The text data is explained in Table I.

B. Evaluation Metrics

The overall performances of the proposed technique are assessed with the help of various parameters such as accuracy, precision, recall and F-measure. Accuracy is the value close to the true value given by the following expression.

$$Acc = \frac{Tr_p + Tr_n}{Tr_p + Tr_n + Fa_p + Fa_n} \quad (11)$$

Precision is an important parameter, and it represents the ratio between numbers of positive samples to the total samples. It is computed using the following expression.

$$precision = \frac{Tr_p}{Tr_p + Fa_n} \quad (12)$$

Recall indicates the ratio between positively classified positive samples and the total number of positive samples can be computed using the following expression.

$$recall = \frac{Tr_p}{Tr_p + Fa_n} \quad (13)$$

F-Measure generates a single score that accounts for both accuracy and recall concerns in a single number. The expression is given below.

$$F\ measure = 2 \times \left[\frac{precision \times recall}{precision + recall} \right] \quad (14)$$

These equations Tr_n and Tr_p indicates the number of true negative and true positive values. The parameters Fa_n and Fa_p represent the false negative and false positive values.

TABLE I. Text Dataset Description

Name of text dataset	Details
Education – (https://data.world/cityofchicago/performance-metrics-city-colleges-of-chicago-course-s):	This is based on the records obtained in the college of Chicago university and describes the course's success rate
Transportation – (https://data.world/buffalony/y93c-u65y):	This dataset comprises the details about the traffic count data.
Environment – (https://data.world/dublin-city	The annual traffic counts conducted at 33 locations across the city cordon created by the Royal and Grand Canals from 2008 to 2012 at 15-minute intervals
Geospatial – (https://data.world/dcopendata/03fcb333c5a3441ab9823116b4359c4a-4	It includes 311 Service Requests in the last 30 days for illegal dumping on public property.
Water – (https://data.world/city-of-ny/bkwf-xfky	The data tables summarize the turbidity values, coliform, fluoride and chlorine found at sides in distribution.

C. Evaluation Results

The proposed approach is analyzed on four performance parameters: accuracy, precision, recall, and f measure. The overall performance of the proposed approach is explained in the following Table II.

Table II indicates the entire evaluation result of the proposed DCCNN based event detection approach. As Table II indicates, the proposed approach is assessed on four statistical measurements: accuracy, precision, recall, and f-measure. Accuracy value generally refers to the obtained value being close to the true value. The obtained classification accuracy of the proposed work is 97.41%. Improved accuracy in the classification approach exhibits that the proposed approach is highly effective for the event detection process. The proposed DCCNN approach attains the precision value of 95.06%, recall is 91.69%, and the obtained F1-score is 93.35, respectively. The simulation results prove that the proposed model is highly applicable for event detection mechanisms. The class-wise comparison of different performances is indicated by the following Table III.

Table III shows the class-wise comparison of the proposed approach. The proposed approach is assessed based on five different events: education, transportation, environment, geospatial, and water. According to the experimental results, the education database has obtained higher accuracy of 99.35%. The proposed framework attains the accuracy for transportation, environment, geospatial, and water are 96.77%, 98.70%, 98.06% and 94.19%, respectively. Compared with other events, an achieved precision performance of the DCCNN approach is minimum for the education database. The transportation database obtains 97.87% of improved precision value. The precision values obtained for education, water, geospatial and environment are 83.33%, 91.80%, 90.47%, and 95.00%. The education database attains 100% of recall performance, and similarly, the environmental and geospatial database obtains the same recall performance of 95%. Using the transportation and water database, the DCCNN methods obtain the recall value of 92.5% and 93.335%, respectively. The environmental database achieves 95% of the F1 score value, and the attained f1-score of education, transportation, geospatial and water databases is 90.90%, 94.845, 92.685 and 92.56%. The graphical representation of the accuracy comparison is shown in Fig. 4.

Fig. 4 is used to illustrate the accuracy performance of the proposed technique, along with a comparison of the existing approach. In Fig. 4, the horizontal axis denotes different existing approaches along with the proposed approach, and the vertical axis indicates the performance results, expressed in percentage. The existing methods used for the comparative analysis are Simple CNN, ResNet50, and DNN. The simple CNN approach is the deep learning approach; the existing CNN approach also improves event detection outcomes. But it faces over fitting problem, and also it suffers from a class imbalance issue, which affects the overall performance. The training process of the ResNet model is not as much better. Also, the DNN model is highly complex to perform event detection in big data. In order to avoid such issues, the proposed work uses the DCCNN model for perfect event detection. The accuracy obtained for the proposed approach is 97.41%. Simple CNN is better thanResNet50 and DNN for accuracy in the existing methods.

TABLE II. OVERALL PERFORMANCE

Sl. no.	Parameter	Accuracy (%)
1	Accuracy	97.41
2	Precision	91.69
3	Recall	95.06
4	F1-score	93.35

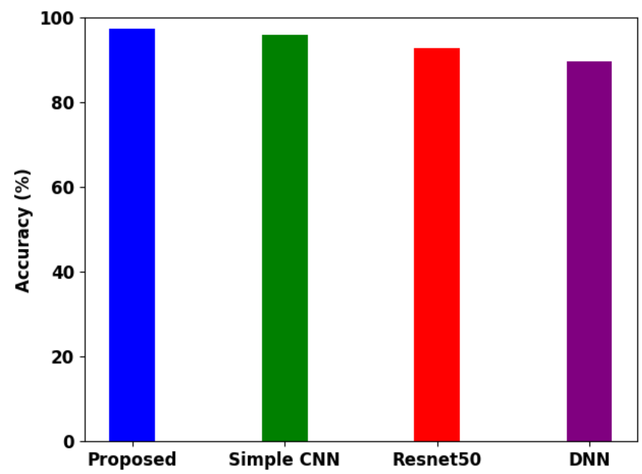


Fig. 4. Accuracy performance comparison.

TABLE III. PERFORMANCE OF DIFFERENT EVENTS

Parameters (%)	Events				
	Education	Transportation	Environment	Geospatial	Water
Accuracy	99.35	96.77	98.70	98.06	94.19
Precision	83.33	97.87	95.00	90.47	91.80
Recall	100	92.00	95.00	95.00	93.33
F1 score	90.90	94.84	95.00	92.68	92.56

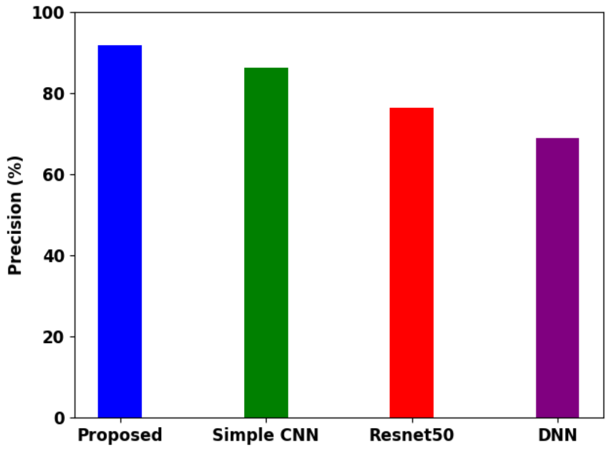


Fig. 5. Precision performance comparison.

Fig. 5 represents the performance comparison in terms of precision. The precision value obtained for the proposed DCCNN based event detection methodology is 91.69%. The Simple CNN approach has a nearly close value to the proposed method in the comparison results. The DNN approach shows the least performance than other methods indicated in Fig. 5. The following Fig. 6 indicates the recall performance comparison.

Fig. 6 illustrates the recall performance comparison of the proposed DCCNN approach with the existing three approaches. The DCCNN method has a recall value of about 95.06%. All other three methods have almost similar performance, among that, Simple CNN shows better performance, and DNN shows the least performance. The following Fig. 7 indicates the f1-score performance comparison.

The performance comparison in terms of f1-score is illustrated in Fig. 7. The corresponding result for the proposed technique is 93.35%. It indicates that the proposed approach is better than all other existing methods. In short, the overall comparison result shows that the proposed approach is better for the event detection process. The graphical representation of class wise performance comparison is shown in Fig. 8.

Fig. 8 represents the performance comparison of different events such as education, environment, geospatial, transport and water. The accuracy of education, transportation, geospatial, environment, and water databases is 99.35%, 96.77%, 98.70%, 98.06% and 94.19%, respectively. It means that this education event showed better performance in terms of accuracy. The precision value for the corresponding events are 83.33%, 97.87%, 95%, 90.47% and 91.80% respectively. Likewise the recall values are 100%, 92%, 95%, 95.5% and 93.33% respectively. The f1-score is also better and the values are 90.90%, 94.84%, 95%, 92.68% and 92.56% respectively. The overall results show that the proposed DCCNN based event detection approach is better than other methods. Table IV represent the overall performance comparison of the proposed approach with previously existing techniques.

The proposed event detection model is compared with previous techniques like simple CNN, ResNet50, DNN. These existing approaches involve drawbacks like over fitting issues, imbalance classification, computational complexity, high processing time etc. Also, many of the conventional approaches are difficult to process in different modalities of data. The proposed work introduces the DCCNN model for event detection and classification by considering these issues in the existing works. This approach provides enhanced event classification results regarding the accuracy, recall, precision and F-measure. The above table reveals that the proposed DCCNN approach is highly effective than the state-of-the-art techniques.

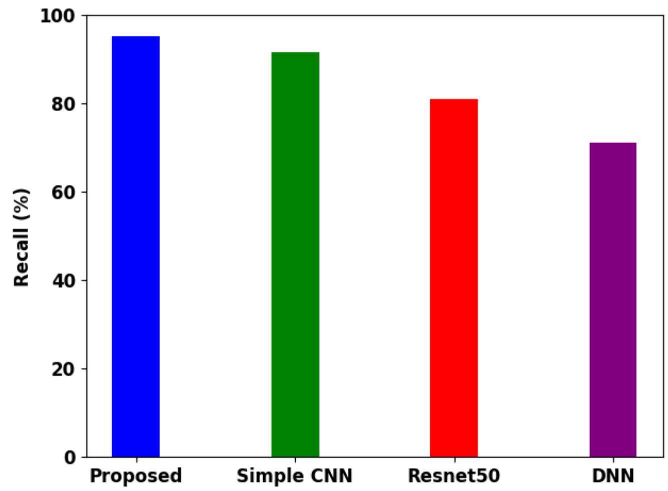


Fig. 6. Recall performance comparison.

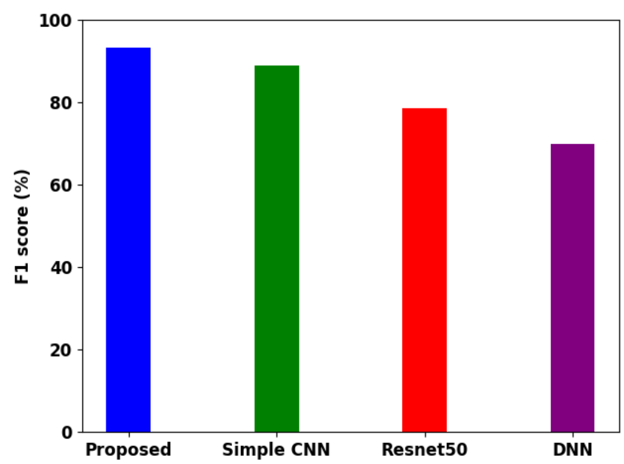
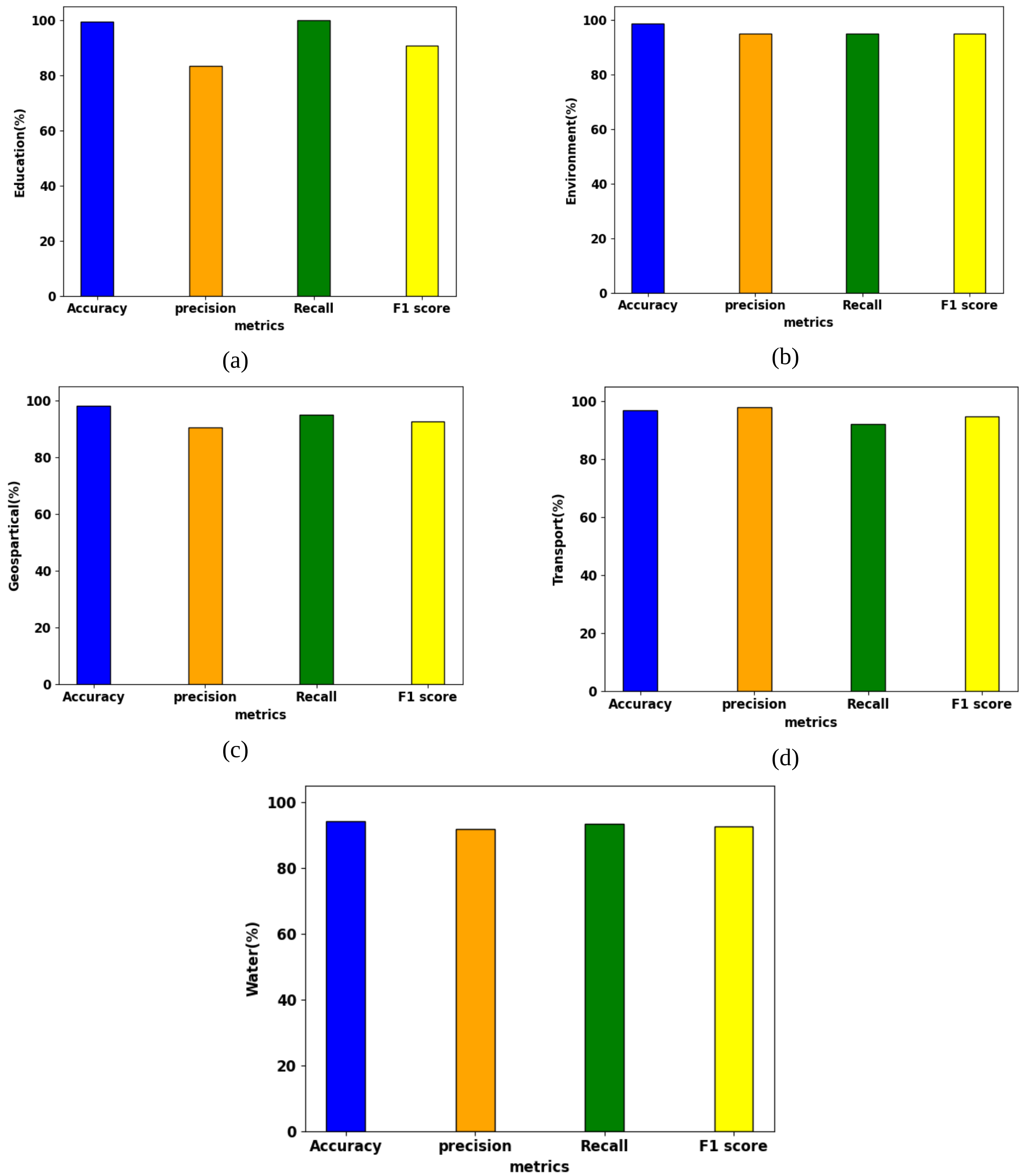


Fig. 7. F1-score performance comparison.



(a) (b) (c) (d) (e)

Fig. 8. Performance comparison of different events (a) education (b) environment (c) geospatial (d) transport (e) water.

TABLE IV. OVERALL PERFORMANCE COMPARISON OF PROPOSED CLASSIFICATION APPROACH WITH EXISTING APPROACHES

Methods	Accuracy (%)	Recall (%)	Precision (%)	F-measure (%)
Simple CNN	95.87	91.53	86.32	88.85
ResNet50	92.77	80.93	76.31	78.55
DNN	89.67	71.06	68.97	70
Proposed DCCNN	97.41	95.06	91.69	93.35

V. Conclusion and Future Scope

Event detection is considered a very important procedure in various fields. Due to its diverse application, much research work exists in this field. Most event detection strategies utilize only one particular type of data (i.e. text or image). Many of the previous traditional approaches are not appropriate for multi-modal input data for event detection. Hence this work put forward a new event detection procedure with text and image data. The proposed event detection framework involves pre-processing, feature extraction and classification. These stages are performed for the input data from the image and text dataset. For pre-processing, the text input data, approaches like tokenization, stop word filter and lower case filter are utilized. Similarly, the image-based input data is pre-processed by using the AMF method. Then, the essential features that smoothen the classification process are extracted in the feature extraction stage. The events from the text and image dataset are classified using the DCCNN model in the last stage. The proposed work is implemented with the help of the PYTHON platform. The evaluation parameters considered are accuracy, precision, recall and f1-score, and the corresponding values obtained are 97.41%, 95.06%/ 91.69%, and 93.35%, respectively. The results show that the proposed approach performs well. In future, we will extend the work with more datasets and improvise the overall performance. Also, the computational complexity of the system will be evaluated, and more parameters to evaluate the system's performance will be included. Moreover, hybrid optimization algorithms will be introduced to enhance the efficacy of the classification approach.

References

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A.H. Byers, "Big data: The next frontier for innovation, competition, and productivity". McKinsey Global Institute 2011.

[2] S. Wunnava, X. Qin, T. Kakar, C. Sen, E.A. Rundensteiner and X. Kong, "Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding". Drug safety vol. 42, no. 1, pp. 113-122, 2019.

[3] Tao, P. Yang and H. Feng, "Utilization of text mining as a big data analysis tool for food science and nutrition". Comprehensive Reviews in Food Science and Food Safety vol. 19, no. 2, pp. 875-894, 2020.

[4] B.K. Daniel, "Big Data and data science: A critical review of issues for educational research". British Journal of Educational Technology vol. 50, no. 1, pp. 101-113, 2019.

[5] H. Sayyadi, M. Hurst and A. Maykov, "Event detection and tracking in social streams". In Third international AAAI conference on weblogs and social media 2009.

[6] J. Wan, P. Zheng, H. Si, N.N. Xiong, W. Zhang and A.V. Vasilakos, "An artificial intelligence driven multi-feature extraction scheme for big data detection". IEEE Access vol. 7, pp. 80122-80132, 2019.

[7] Alomari, I. Katib, A. Albeshri, T. Yigitcanlar and R. Mehmood, "Iktishaf+: A big data tool with automatic labeling for road traffic social sensing and event detection using distributed machine learning". Sensors vol. 21 no. 9, pp. 2993, 2021.

[8] Z. Saeed, R.A. Abbasi, O. Maqbool, A. Sadaf, I. Razzak, A. Daud, N.R. Aljohani and G. Xu, "What's happening around the world? a survey and framework on event detection techniques on twitter". Journal of Grid Computing vol. 17, no. 2, pp. 279-312, 2019.

[9] A.N. Reiz, M.A. de la Hoz and M.S. García, "Big data analysis and machine learning in intensive care units". MedicinaIntensiva (English Edition) vol. 43, no. 7, pp. 416-426, 2019.

[10] Bolón-Canedo, N. Sánchez-Maroño and A. Alonso-Betanzos, "Recent advances and emerging challenges of feature selection in the context of big data". Knowledge-Based Systems vol. 86, pp. 33-45, 2015.

[11] N. Alsaedi, P. Burnap and O. Rana, "Can we predict a riot? Disruptive event detection using Twitter". ACM Transactions on Internet Technology (TOIT) vol. 17, no. 2, pp. 1-26, 2017.

[12] Z. Shah and A.G. Dunn, "Event detection on Twitter by mapping unexpected changes in streaming data into a spatiotemporal lattice". IEEE transactions on big data 2019.

[13] A.H. Hossny, L. Mitchell, N. Lothian and G. Osborne, "Feature selection methods for event detection in Twitter: a text mining approach". Social Network Analysis and Mining vol. 10, no. 1, pp. 1-15, 2020.

[14] A.H. Aliwy, "Tokenization as pre-processing for arabic tagging system". International Journal of Information and Education Technology vol. 2, no. 4, pp. 348, 2012.

[15] Y. Yılmaz and A.O. Hero, "Multi-modal event detection in Twitter hashtag networks". Journal of Signal Processing Systems vol. 90, no. 2, pp. 185-200, 2018.

[16] Z. Subecz, "Event detection and classification in hungarian natural texts". European Scientific Journal 2019.

[17] R. Sprugnoli & S. Tonelli, "Novel event detection and classification for historical texts". Computational Linguistics, vol. 45, no. 2, pp. 229-265, 2019.

[18] T. Nguyen & R. Grishman, "Graph convolutional networks with argument-aware pooling for event detection". In Proceedings of the AAAI Conference on Artificial Intelligence vol. 32, no. 1, 2018.

[19] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu and F.E. Alsaadi, "A survey of deep neural network architectures and their applications". Neurocomputing vol. 234, pp. 11-26, 2017.

[20] L. Yao, C. Mao & Y. Luo, "Clinical text classification with rule-based features and knowledge-guided convolutional neural networks". BMC medical informatics and decision making, vol. 19, no. 3, pp. 31-39, 2019.

[21] S. Cherichi and R. Faiz, "Big data analysis for event detection in microblogs". In Recent Developments in Intelligent Information and Database Systems, Springer, Cham, pp. 309-319, 2016.

[22] R. Yadav, A.K. Pradhan and I. Kamwa, "Real-time multiple event detection and classification in power system using signal energy transformations". IEEE Transactions on Industrial Informatics vol. 15, no. 3, pp. 1521-1531, 2018.

[23] K. Wang, L. Yang and B. Yang, "Audio event detection and classification using extended R-FCN approach". In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), pp. 128-132, 2017.

[24] H. Zhang and J. Pan, "Casm: A deep-learning approach for identifying collective action events with text and image data from social media". Sociological Methodology vol. 49, no. 1, pp. 1-57, 2019.

[25] Q. Chen, W. Wang, K. Huang, S. De and F. Coenen, "Multi-modal generative adversarial networks for traffic event detection in smart cities". Expert Systems with Applications vol. 177, pp. 114939, 2021.

[26] Alomari, I. Katib and R. Mehmood, "Iktishaf: A big data road-traffic event detection tool using Twitter and spark machine learning". Mobile Networks and Applications pp. 1-16, 2020.

Estimation of Transmission Rate and Recovery Rate of SIR Pandemic Model Using Kalman Filter

Wahyu Sukestyastama Putra*, Afrig Aminuddin, Ibnu Hadi Purwanto, Rakhma Shafrida Kurnia, Ika Asti Astuti
Faculty of Computer Science, Universitas Amikom, Yogyakarta, Sleman, Indonesia

***Abstract*—COVID-19 is a global pandemic that significantly impacts all aspects. The number of victims who died makes this disease so terrible. Various policies continue to be pursued to reduce the spread and impact of COVID-19. The spread of a disease can be modeled in differential equation modeling. This differential equation modeling is known as the SIR Model. A differential equation can be expressed in a state-space model. The state-space model is a model that is widely used to design a modern control system. This research carried out the transmission rate and recovery rate estimates in the SIR pandemic model. Estimation of the transmission rate and recovery rate in this study poses a challenge to the value of the number of people confirmed as infected. The experimental result shows that the transmission and recovery rates can be estimated using the data for the infected and recovered persons. Estimates of infected and recovered people were conducted using the Kalman Filter.**

Keywords—Kalman filter; pandemic; SIR model

I. Introduction

COVID-19 is a global pandemic disease that is currently challenging for various countries worldwide. The current effort is to prevent an explosion of infected cases, considering the limited medical facilities available. This condition will be more difficult if not controlled because the growth in cases of infection is exponential [1]. The spread of a disease can be modeled in differential equation modeling. This differential equation modeling is known as the SIR Model. A differential equation can be expressed in a state-space model. The state-space model is a model that is widely used to design a modern control system. Modeling the distribution of COVID-19 to the SIR state-space model is interesting because this model can control the spread of measured diseases. The advantage of controlling using state-space modeling is that the model supports MIMO (Multiple Input and Multiple Output) cases. Two essential parameters must be known in the SIR model: transmission and recovery rate.

Transmission and recovery rates are the points to observe a process of controlling a disease outbreak in this model. The system model can be divided into input, output, and process. In this study, the input in the model is a suspect person (a person who is likely to be infected). Meanwhile, state-space modeling is carried out in the process stage with the state of the infected person and recovered person. At the same time, the output is an active case of an infected person. The output becomes a dynamic system because it adds a migration factor to the control system model.

According to previous research, the SIR model is used to predict and control a pandemic [2]–[9]. However, the pandemic SIR pandemic model has a problem because the transmission rate and recovery rate in the SIR model are constant [10]–[16]. Using these constants in predicting the number of infected and recovered is unsuitable for a condition that does not restrict population movement [17]–[22]. In this study, the authors are interested in identifying the transmission and recovery rates of the SIR model in real time. The updated results of the transmission rate and recovery rate values are then used to predict the number of states in the SIR model. State estimation needs to be done because detection bias of infected people can occur in a pandemic control.

Kalman filter is an algorithm that can be used to estimate the state of the SIR model. The Kalman filter is a two-step process that involves prediction and measurement updates. In the prediction step, the Kalman filter uses the current state of the system and the system's dynamics to predict the state at the next time step. In the measurement update step, the Kalman filter uses the predicted state and the new measurement to compute a more accurate estimate of the system's state.

The rest of this paper is organized as follows. Section II presents the related works of the SIR model implementation in various pandemics. Section III describes the proposed methods used in this study. Section IV presents the result and discussion. Finally, Section V concludes this study.

II. Related Works

Research on SIR models was conducted in 2010, and an H1N1 disease distribution model was designed using the SIR Model [2]. G. Yang [3] also carried out the control design using the SIR model. Based on these two studies, it can be concluded that the pandemic can be modeled into a control system. Research related to control systems using the SIR Model continues to grow. H. Weiss [4] formulated several policy examples to prevent the spread of disease based on the SIR model. W. Chen [5] modeled the Ebola distribution model using the SIR Model. Using the SIR model, W. Huang and G. Provan [6] designed several forecasting filters. The use of the SIR model to form state space was carried out in [7] and [8] forecasted the spread of influenza using the SIR state-space model approach.

Meanwhile, related to the COVID-19 pandemic, Chen et al. [10] added an undetectable infected person factor to the SIR model for the COVID-19 disease. The model uses two types of infected people: detachable infected persons, people with symptoms, and Asymptomatic infected persons. The SIR

*Corresponding Author.

equation turns into a matrix with the order of 3×3. The dataset was taken from the National Health Commission of the People's Republic of China (NHC) daily Outbreak Notification. The experimental result showed that the method could predict infected and recovered persons with an error of 3%.

G. Fabricius and A. Maltz [11] have identified the threshold for disease spread in the SIR model with local and global contacts. The study modeled a SIR model into the Stochastic and deterministic models. The research built space detection parameters in a region to avoid the exponential spread of disease. In comparison, [12] and [13] carried out mathematical modeling of the transmission and control of COVID-19. The method used in [12] is the stochastic transmission method, while in [13], the logistic equation is used to describe and interpret the SIR epidemic model. This study indicates that research [12] made early predictions of locations with the potential to transmit disease based on data models that had occurred in Wuhan. Whereas [13] shows that identifying the transmission rate in the SIR model can be done using the logistics equation model.

C. Tsay et al. [14] use the SEAIR model. The SEAIR model is a modification of the SEIR (Susceptible-Exposed-Infected-Recovered) model. The research built an optimal control for the model by using the parameters of social distancing, quarantining, rate of testing, recovery rate, death rate, and initial exposure. In comparison, [18] used a modified generalized Lotka-Volterra (gLV) model. It developed a control for the model using immigration, infection, recovery, dead, and control parameters. At the same time, the form of the augmented model is to add a birth factor. While the state in the model used is estimated using the Extended Kalman Filter method. In comparison, [17] uses the SEIQR (Susceptible-Exposed-Infected- Quarantine-Recovered) model. It used a fixed control parameter on the migration parameter (M). The main finding of this study was that researchers were able to develop a mathematical model that could be used to observe the dynamics of COVID-19.

A. Abuhasel et al. [19] use the SIR model to predict cases of the spread of COVID-19 in the Kingdom of Saudi Arabia (KSA). At the same time, the ARIMA model is used to predict prevalence cases. The data used in this research is daily case data in Saudi Arabia (KSA). The results show that the SIR model can predict the development of infection cases and shows that the policies taken by the government are appropriate. At the same time, the ARIMA model shows that this model is an estimation model with current and past data with a high correlation and showed a small error. In comparison, [20] used the SIR model to predict the distribution of disease in a community. This prediction helps determine what anticipation needs to be done to control the spread of COVID-19. The data used in this study are data from various countries such as China, South Korea, India, Australia, the USA, and Italy.

III. Proposed Method

The SIR Pandemic model is simple. The model consists of S (susceptible person), I (Infected person), and R (recovered person). A susceptible person is a person who has the potential to be infected with the disease. Meanwhile, infected people are people who have been infected with the disease. The rate of change from susceptible persons to infected persons is known as the transmission rate. Meanwhile, the rate of change from an infected person to a recovered person is called the recovery rate. In this study, an estimate of the transmission rate and recovery rate of the COVID-19 epidemic was carried out. The dataset used in this study is the confirmation data for COVID-19 based on the public database.

To estimate the transmission rate and recovery rate, the SIR epidemic model is first carried out in the form of a differential equation which is written as follows,

$$\frac{\mathrm{dS(t)}}{\mathrm{dt}} = \frac{-\beta(\mathrm{t})\mathrm{S(t)X(t)}}{\mathrm{n}}, \tag{1}$$

$$\frac{\mathrm{dX(t)}}{\mathrm{dt}} = \frac{\beta(\mathrm{t})\mathrm{S(t)X(t)}}{\mathrm{n}} - \gamma(\mathrm{t})\mathrm{X(t)} \tag{2}$$

$$\frac{\mathrm{dR(t)}}{\mathrm{dt}} = \gamma(\mathrm{t})\mathrm{X(t)} \tag{3}$$

where $\frac{dS(t)}{dt}$ is the rate of change of susceptible persons, $\frac{dX(t)}{dt}$ is the rate of change of confirmed infected persons, and $\frac{dR(t)}{dt}$ is the rate of change of people recovering from infection. A susceptible person is a person who has the potential to be infected.

The differential equation (1) to (3) is then changed to the form of the differential equation into:

$$S(t+1) - S(t) = \frac{-\beta(t)S(t)X(t)}{n} \tag{4}$$

$$X(t+1) - X(t) = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t) \tag{5}$$

$$R(t+1) - R(t) = \gamma(t)X(t). \tag{6}$$

where n is the known population in an area. The differential equation (4) to (6) is then converted into the state-space model equation with the assumption that $n = S(t)$ becomes:

$$\begin{bmatrix} S(t+1) \\ X(t+1) \\ R(t+1) \end{bmatrix} = \begin{bmatrix} 1 & -\beta & 0 \\ 0 & 1+\beta-\gamma & 0 \\ 0 & \gamma & 1 \end{bmatrix} \begin{bmatrix} S(t) \\ X(t) \\ R(t) \end{bmatrix} \tag{7}$$

with:

$S(t)$= Susceptible person.

$X(t)$= Infected person.

$R(t)$= Recovered person.

$\beta(t)$= Transmission rate

$\gamma(t)$= Recovery rate

In (7), the infected persons and the recovered person do not depend on the state of susceptible person. Thus (7) can be converted into:

$$\begin{bmatrix} X(t+1) \\ R(t+1) \end{bmatrix} = \begin{bmatrix} 1+\beta-\gamma & 0 \\ \gamma & 1 \end{bmatrix} \begin{bmatrix} X(t) \\ R(t) \end{bmatrix} \tag{8}$$

with the output equation:

$$y = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X(t) \\ R(t) \end{bmatrix} \tag{9}$$

Estimates of X (t) (infected person) and R (t) (recovered person) were performed using the Kalman Filter method. The steps of the Kalman filter algorithm are presented in Fig. 1.

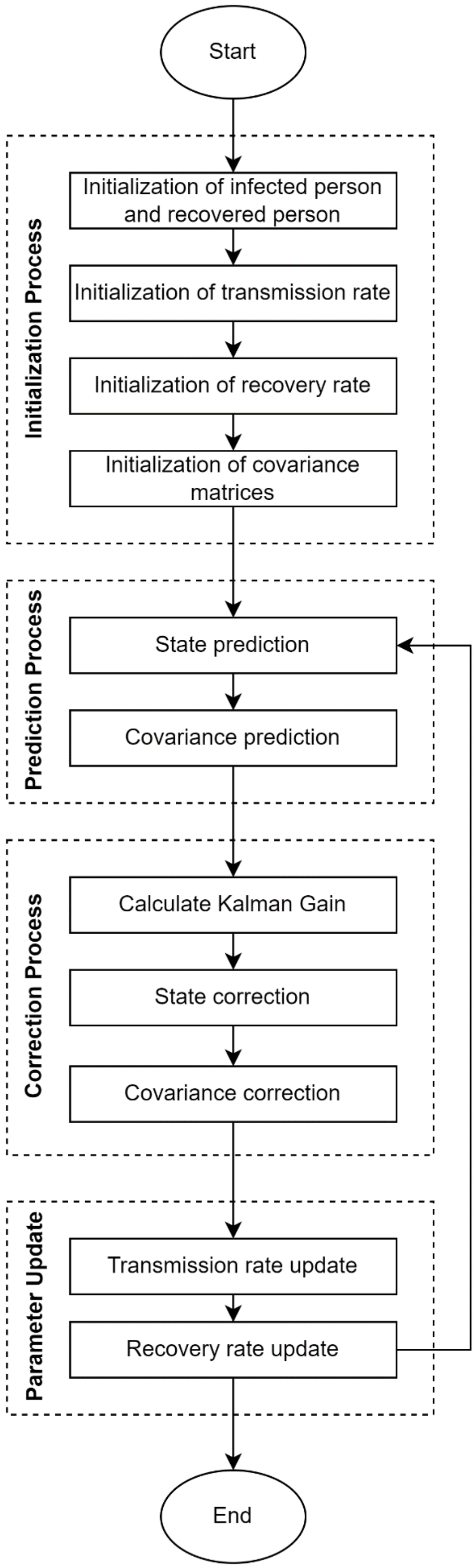


Fig. 1. The flow diagram of the proposed method.

Step 1: Initialization Process

1) Initialization of $X_{k-1}(t)$ and $R_{k-1}(t)$

2) Initialization of $\beta(t)$ using:

$$\beta(t) = \frac{(X(t)-X(t-1))+(R(t)-R(t-1))}{X(t-1)} \quad (10)$$

3) Initialization of $\gamma(t)$ using:

$$\gamma(t) = \frac{R(t)-R(t-1)}{X(t-1)} \quad (11)$$

4) Initialization of covariance matrices

Step 2: Prediction Process

1) State prediction using:

$$\begin{bmatrix} \hat{X}_k(t) \\ \hat{R}_k(t) \end{bmatrix} = \begin{bmatrix} 1+\beta(t)-\gamma(t) & 0 \\ \gamma(t) & 1 \end{bmatrix} \begin{bmatrix} X_{k-1}(t) \\ R_{k-1}(t) \end{bmatrix} \quad (12)$$

2) Covariance prediction using:

$$\hat{P}_k(t) = AP_{k-1}(t)A^T + Q, \quad (13)$$

with Q is a covariance noise and A is transition matrices written as:

$$A = \begin{bmatrix} 1+\beta(t)-\gamma(t) & 0 \\ \gamma(t) & 1 \end{bmatrix} \quad (14)$$

Step 3: Correction Process

1) Calculate Kalman Gain using:

$$K_k = \frac{\hat{P}_k(t)H^T}{H\hat{P}_k(t)H^T+R} \quad (15)$$

which R is measurement noise, and H represents measurement matrices written as:

$$H = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \quad (16)$$

2) State Correction using:

$$\begin{bmatrix} X_k(t) \\ R_k(t) \end{bmatrix} = \begin{bmatrix} \hat{X}_k(t) \\ \hat{R}_k(t) \end{bmatrix} + K_K(z_k - \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{X}_k(t) \\ \hat{R}_k(t) \end{bmatrix}) \quad (17)$$

which z_k is the measurement data

3) Covariance correction using:

$$P_k = (I - K_k \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix})\hat{P}_k(t). \quad (18)$$

Step 4: Parameter Update

1) Transmission rate $\beta(t)$ update using:

$$\beta(t) = \frac{(X_k(t)-X_k(t-1))+(R_k(t)-R_k(t-1))}{X_k(t-1)} \quad (19)$$

2) Recovery rate $\gamma(t)$ update using:

$$\gamma(t) = \frac{R_k(t)-R_k(t-1)}{X_k(t-1)} \quad (20)$$

Step 5: Go to Step 2

IV. Experimental Results

Estimating the number of infected and recovered people is a topic to do because early detection of COVID-19 infection is less massive. This detection is increasingly difficult with the community's attitude, who still think that being infected with COVID-19 is a bad thing that has the impact of being excluded by the community. Awareness of the dangers of COVID-19 has not been good because it was found that people had not implemented health protocols in an orderly manner. With both conditions, matters are complicated because the movement of people who ignore health protocols increases the potential for infection and becomes an infection agent for others. This condition is amplified by the lack of mass testing for COVID-19 and undetectable infected person in the early pandemic.

In this study, the transmission rate and recovery rate are estimated using the SIR pandemic model. The estimation of the transmission rate and recovery rate in this study poses a challenge to the value of the number of people confirmed as infected. The number of people confirmed as infected is essential in determining the value of the transmission and recovery rate. An estimate of the number of infected people and the number of people recovered was carried out to overcome this problem. Estimation is done using the Kalman Filter.

The results of estimating the infected and recovered person using the Kalman Filter are depicted in Fig. 2 and Fig. 3. The estimation results show that the estimates and the confirmed data have the same trend. An error covariance matrix tunning in the estimation algorithm is performed to obtain the slightest difference between the estimated recovered person and the confirmed recovered person. The difference between the estimate and the confirmed recovered person is used as a reference.

The difference between the estimates and the confirmed recovered person is then used to determine the estimation error of the proposed algorithm. The recovered person estimation error is depicted in Fig. 4. Fig. 4 shows that a fairly large estimation error occurs in the first 15 days of the estimation. This error occurs because of the low number of cases of confirmed recovered, so the percentage of estimation error is large. After the 20th day, the estimated error of the recovered person is decreased to less than 5%.

The error of the recovered person is then used as the basis for determining the estimated error of the infected person. Given the estimated error value, an upper and lower limit of the estimate can be made. This limit is then used as the basis for the validation of the data for the confirmed infected person. If the data for the confirmed infected person are within the estimated range, then the data has good validity. However, the data has dubious validity if it is outside the estimated range. The estimated range of the infected person is depicted in Fig. 5. In Fig. 5, the lower limit of the estimate is depicted using the yellow line, and the blue line represents the upper limit of the estimate. Meanwhile, the confirmed infected person is depicted using the orange line. From the comparative data between the confirmed infected person and the estimated range of the infected person in Fig. 5, it can be seen that the position of the confirmed infected person is not always within the estimation range. Thus there are several data of confirmed infected persons who are in doubt.

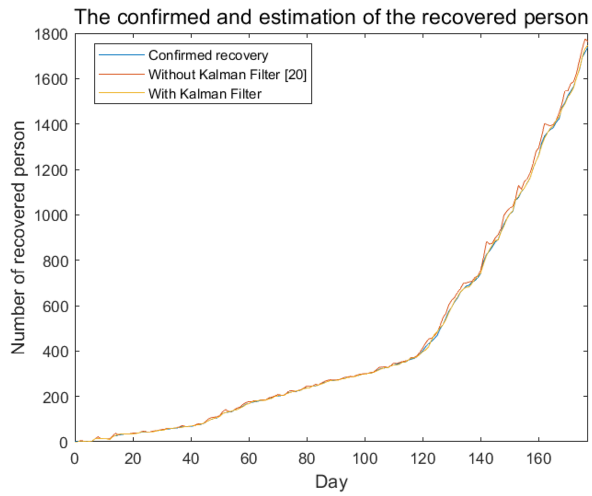


Fig. 2. The confirmed and estimation of the recovered person.

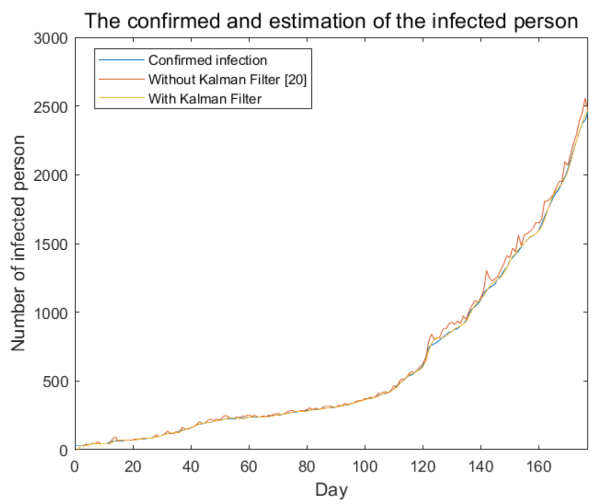


Fig. 3. The confirmed and estimation of the infected person.

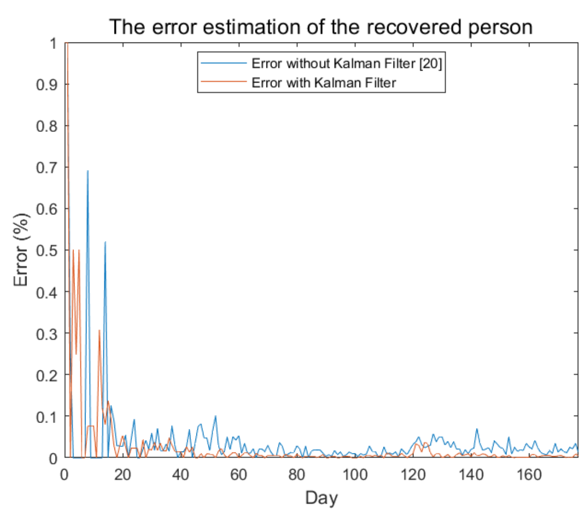


Fig. 4. The error estimation of the recovered person.

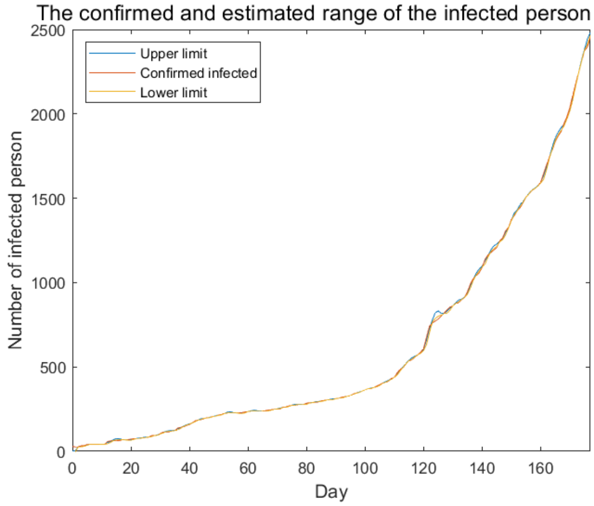


Fig. 5. The confirmed and estimated range of the infected person.

Fig. 6 shows data for confirmed infected persons outside the estimated limits. In Fig. 6, it can be seen that for the estimated day 0 to day 100, the confirmed infected person is still within the estimated limit, and if it exceeds the estimated limit, it only has a slight difference, namely a maximum of 2. However, after the 100th day, the confirmed infected person is almost always outside the estimated range. So the number of confirmed infected persons has questionable validity. On the other hand, the estimated recovered person has a reasonably good estimation error. Thus the transmission rate and recovery rate are predicted using estimated data.

Estimated infected people and estimated recovered people are used to estimate the transition rate and recovery rate. The results of the transmission rate estimation are depicted in Fig. 7. Fig. 7 shows that the transmission rate values from day 0 to day 40 experience convergence characterized by decreasing oscillations. After convergence, the transmission rate oscillates with a value above zero. This value indicates that the transmission process is still occurring. Meanwhile, the transmission rate is said to stop if the transmission rate is zero.

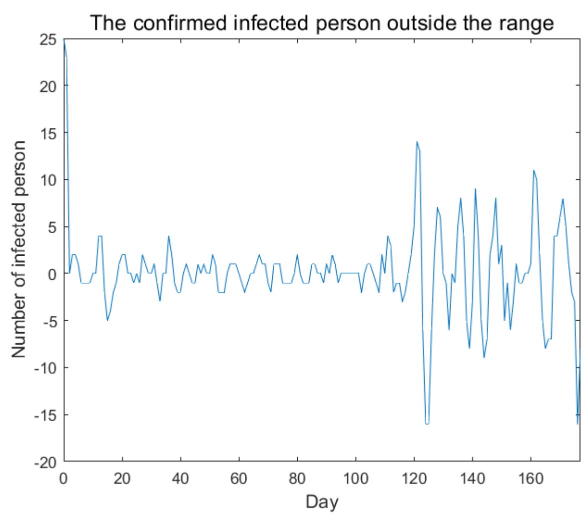


Fig. 6. The confirmed infected person outside the range.

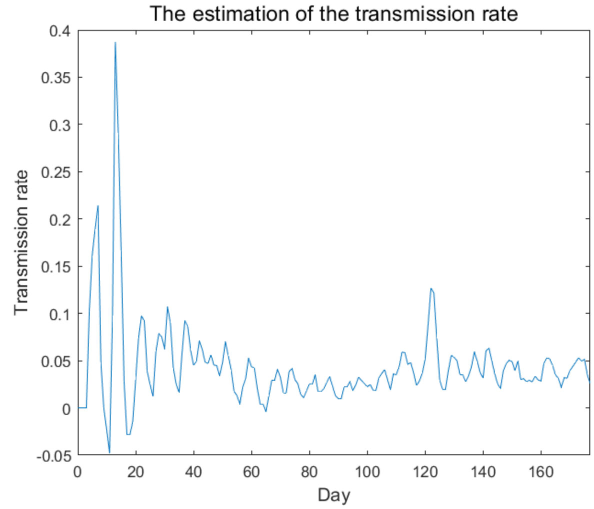


Fig. 7. The estimation of the transmission rate.

The results of the recovery rate estimation are depicted in Fig. 8. Fig. 8 shows that the recovery rate values from day 0 to day 40 experience a convergence like what happened to the transmission rate marked by reduced estimation oscillations. After convergence, the recovery rate also experiences an oscillation like what happens in the transmission rate. However, the oscillation amplitude at the recovery rate is smaller than at the transmission rate. The recovery rate of the oscillation value is also above zero. It means that the recovery process of an infected person is also in progress.

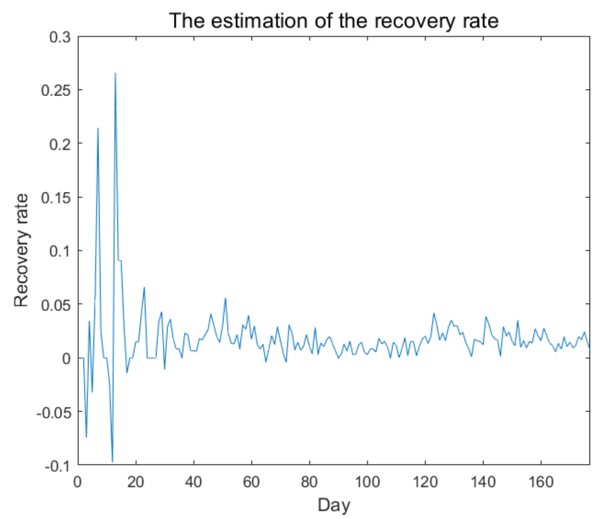


Fig. 8. The estimation of the recovery rate.

V. Conclusions

The research shows that the transmission rate and recovery rate can be estimated using the estimated data for the infected and recovered persons. Estimates of infected and recovered people are carried out using the Kalman Filter. Estimates of the infected and recovered people are carried out to address the data of the confirmed infected persons whose validity is doubtful. In further research, developing a more comprehensive pandemic model is necessary. In pandemic modeling, SIR has not included factors for the asymptomatic infected person. Thus, research on developing pandemic models needs to be carried out.

ACKNOWLEDGMENT

This research funded by Research Department of Universitas Amikom Yogyakarta – Indonesia with contract Number 063/KONTRAK-LEMLIT/AMIKOM/ VII/2020, July, 6th 2020.

REFERENCES

[1] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 6, p. 2022, 2022, doi: 10.14569/IJACSA.2022.0130665.

[2] Y. Liu, "Investigation of Prediction and Establishment of SIR Model for H1N1 Epidemic Disease," in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, 2010, pp. 1–4, doi: 10.1109/ICBBE.2010.5517654.

[3] G. Yang, "Optimal control for an SIR epidemic model," in 2011 Chinese Control and Decision Conference (CCDC), 2011, pp. 515–518, doi: 10.1109/CCDC.2011.5968235.

[4] H. Weiss Sir Ronald Ross, "The SIR model and the Foundations of Public Health," Mater. MATemàtics Vol., vol. 17, no. 3, pp. 1887–1097, 2013.

[5] W. Chen, "A Mathematical Model of Ebola Virus Based on SIR Model," in 2015 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration, 2015, pp. 213–216, doi: 10.1109/ICIICII.2015.135.

[6] W. Huang and G. Provan, "An improved state filter algorithm for SIR epidemic forecasting," in Proceedings of the Twenty-second European Conference on Artificial Intelligence, 2016, pp. 524–532.

[7] E. Sebastian, P. Victor, and T. Nadu, "a State Space Approach for Sir Epidemic Model," Int. J. Differ. Equations, vol. 12, no. 1, pp. 79–87, 2017.

[8] D. Osthus, K. S. Hickmann, P. C. Caragea, D. Higdon, and S. Y. Del Valle, "Forecasting seasonal influenza with a state-space SIR model.," Ann. Appl. Stat., vol. 11, no. 1, pp. 202–224, Mar. 2017, doi: 10.1214/16-AOAS1000.

[9] F. F. Abdulloh, M. Rahardi, A. Aminuddin, S. D. Anggita, and A. Y. A. Nugraha, "Observation of Imbalance Tracer Study Data for Graduates Employability Prediction in Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 8, pp. 169–174, Oct. 2022, doi: 10.14569/IJACSA.2022.0130820.

[10] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons," pp. 1–10, 2020.

[11] G. Fabricius and A. Maltz, "Exploring the threshold of epidemic spreading for a stochastic SIR model with local and global contacts," Phys. A Stat. Mech. its Appl., vol. 540, p. 123208, 2020, doi: https://doi.org/10.1016/j.physa.2019.123208.

[12] J. Kucharski et al., "Early dynamics of transmission and control of COVID-19: a mathematical modelling study.," Lancet. Infect. Dis., Mar. 2020, doi: 10.1016/S1473-3099(20)30144-4.

[13] M. De La Sen and A. Ibeas, "On a Sir Epidemic Model for the COVID-19 Pandemic and the Logistic Equation," Discret. Dyn. Nat. Soc., vol. 2020, 2020, doi: 10.1155/2020/1382870.

[14] Tsay, F. Lejarza, M. A. Stadtherr, and M. Baldea, "Modeling, state estimation, and optimal control for the US COVID-19 outbreak," Sci. Rep., vol. 10, no. 1, p. 10711, 2020, doi: 10.1038/s41598-020-67459-8.

[15] Aminuddin, "Android Assets Protection Using RSA and AES Cryptography to Prevent App Piracy," 2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020, pp. 461–465, Nov. 2020, doi: 10.1109/ICOIACT50329.2020.9331988.

[16] Q. Waseem, W. Isni Sofiah Wan Din, A. Aminuddin, M. Hussain Mohammed, and R. F. Alfa Aziza, "Software-Defined Networking (SDN): A Review," 2022 5th Int. Conf. Inf. Commun. Technol., pp. 30–35, Aug. 2022, doi: 10.1109/ICOIACT55506.2022.9972067.

[17] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. K. Kar, "A model based study on the dynamics of COVID-19: Prediction and control," Chaos, Solitons & Fractals, vol. 136, p. 109889, 2020, doi: https://doi.org/10.1016/j.chaos.2020.109889.

[18] B. Younes and Z. Hasan, "COVID-19: Modeling, prediction, and control," Appl. Sci., vol. 10, no. 11, pp. 1–14, 2020, doi: 10.3390/app10113666.

[19] K. A. Abuhasel, M. Khadr, and M. M. Alquraish, "Analyzing and forecasting COVID-19 pandemic in the Kingdom of Saudi Arabia using ARIMA and SIR models," Comput. Intell., no. September, pp. 1–14, 2020, doi: 10.1111/coin.12407.

[20] Cooper, A. Mondal, and C. G. Antonopoulos, "A SIR model assumption for the spread of COVID-19 in different communities," Chaos, Solitons and Fractals, vol. 139, p. 110057, 2020, doi: 10.1016/j.chaos.2020.110057.

[21] Cucus, L. B. Aji, A.-F. Bin Mubarak Ali, A. Aminuddin, and L. D. Farida, "Selection of Prospective Workers Using Profile Matching Algorithm on Crowdsourcing Platform," 2022 5th Int. Conf. Inf. Commun. Technol., pp. 122–126, Aug. 2022, doi: 10.1109/ICOIACT55506.2022.9972155.

[22] Geetha Mohan, A.-F. Mubarak Ali, M. Ariff Ameedeen, B. Lathakumary Vijayan, A. Aminuddin, and W. Widayani, "Predictive Models Using Supervised Neural Network for Pollutant Removal Efficiency in Petrochemical Wastewater Treatment," 2022 5th Int. Conf. Inf. Commun. Technol., pp. 116–121, Aug. 2022, doi: 10.1109/ICOIACT55506.2022.9971929.

Tracking The Sensitivity of The Learning Models Toward Exact and Near Duplicates

Menna Ibrahim Gabr[1], Yehia Helmy[2], Doaa S. Elzanfaly[3]
Business Information Systems (BIS)-Dept.-Faculty of Commerce and Business Administration, Helwan University, Egypt[1, 2]
Faculty of Information and Artificial Intelligence, Helwan University, Egypt[3]
Faculty of Computer and Informatics, British University in Egypt, Egypt[3]

Abstract*—Most real-world datasets contaminated by quality issues have a severe effect on the analysis results. Duplication is one of the main quality issues that hinder these results. Different studies have tackled the duplication issue from different perspectives. However, revealing the sensitivity of supervised and unsupervised learning models under the existence of different types of duplicates, deterministic and probabilistic, is not broadly addressed. Furthermore, a simple metric is used to estimate the ratio of both types of duplicates regardless of the probability by which the record is considered duplicate. In this paper, the sensitivity of five classifiers and four clustering algorithms toward deterministic and probabilistic duplicates with different ratios (0% - 15%) is tracked. Five evaluation metrics are used to accurately track the changes in the sensitivity of each learning model, MCC, F1-Score, Accuracy, Average Silhouette Coefficient, and DUNN Index. Also, a metric to measure the ratio of probabilistic duplicates within a dataset is introduced. The results revealed the effectiveness of the proposed metric that reflects the ratio of probabilistic duplicates within the dataset. All learning models, classification, and clustering models are differently sensitive to the existence of duplicates. RF and Kmeans are positively affected by the existence of duplicates which means that their performce increase as the percentage of duplicates increases. Furthermore, the rest of classifiers and clustering algorithms are sensitive toward duplicates existence, especially within high percentage that negatively affect their performance.

Keywords*—Deduplication; deterministic duplicates; probabilistic duplicates; supervised learning models; unsupervised learning models; evaluation metrices

I. INTRODUCTION

Data quality has been an active research area that affects different domains. Many data quality dimensions have been addressed through the literature[1]–[5], Data Duplication has been considered as one of the most intriguing dimensions. Data duplication is defined as multiple representation of the same real world object or a measure of undesirable duplicates within a certain field, record or dataset[6]. The duplication can be found in two different types, the Deterministic and the Probabilistic duplications[7], [8]. The Deterministic (exact) duplication, where two records or more are identical and Probabilistic (near/fuzzy) duplication, where multiple records are nonidentical and refer to the same real world entity[9], [10].

Duplicates can occur due to two main causes. The intra source duplicates, and the inter source duplicates[11]. The intra source duplicates occurs when a single data object can be entered many times into the same database. Whereas the duplicates that appear while merging multiple data source are called inter source duplicates[11]. The process of removing the intra source duplicates is called deduplication[12] which is the main scope of this paper. Whereas removing duplicates from inter source duplicates is called Record Linkage[13].

Data quality dimensions are assessed to evaluate by how much the data is qualified for a task at hand. Dimensions are measured either objectively or subjectively. Subjective measurements are based on consumers' opinions like questionnaires, and surveys. Whereas Objective measurements are used to give a simple ratio between the undesirable outcomes and the total[14]. For example, to calculate the percentage of duplicates in the dataset, the following equation can be used (number of duplicate records/Total number of records). This simple metric can be perfectly used within exact duplicates, however, in the case of near duplicates the probability of these duplicate records should be considered to reflect the true percentage of duplicates within a dataset.

The research effort in duplication area is diverse. However, little work has focused on clarifying the effect of both data duplicate types on the analysis results. In the domain of android malware [15], has addressed the sensitivity of the supervised and unsupervised learning models due to the existence of near duplicates. Thus, in this paper an initiative is taken to clarify the effect of both types of duplicates on classification and clustering learning models.

This paper investigates the impact of the deterministic and the probabilistic duplicates on the results of descriptive (clustering task), and predictive (classification task) data analytics. Five classification models namely, Decision Tree (DT), Support Vector Machine (SVM), Naïve Bayes (NB), Linear Discriminant Analysis (LDA) and Random Forest (RF) are used to clarify the effect of the deterministic duplicates with different ratios. The sensitivity of the five classifiers has been evaluated using three evaluation metrics, Accuracy[16], Matthews Correlation Coefficient (MCC)[17] and F1-Score[18]. While the impact of probabilistic duplicates is investigated through four clustering algorithms namely, The Partition Around Medoids (PAM), Clustering for Large Application (CLARA), K-means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Furthermore, two evaluation indices namely the Average Silhouette Coefficient (ASC)[19] and DUNN Index (DI)[20] are used to track the sensitivity of the four clustering algorithms.

Since that the simple metric doesn't consider the probability of the near duplicates, thus a metric is introduced to estimate the true percentage of near duplicates within a dataset.

The remainder of this paper is organized as follows, Section II, reviews the state of the art and presents the related studies. Where the used clustering evaluation indices are described in Section III. The experimental framework and results are discussed in Section IV. Section V concluded the paper.

II. Related Work

The field of duplication and its treatment methods is immense in the literature. It is addressed from different perspectives and domains, started from an overview of the duplication and techniques[21]–[23], improving the detection techniques[24], [25], evaluating the impact of the duplicates[26], and proposing new frameworks[27], [28] and methods[29], [30] to effectively enhance the detection process. However, little research addresses the impact of duplicates on the analysis results. Some studies are presented below.

In [15], they examined the effect of duplication on the supervised and unsupervised learning in the domain of android malware detection. From their perspective, duplication in this domain means that data samples (e.g., the APK, the DEX code, etc.) appear many times within a corpus. They declared that duplication has limited impact on the supervised classification model, however, it has a significant effect on the unsupervised clustering model.

In [31], two of the benchmark datasets are cleaned from near duplicate images. Deep leaning models are tested against the datasets before and after removing near duplicates. The results revealed a decrease in the classification accuracy by 9% to 14% which means that the duplicates' existence can give more inflated results.

Furthermore in[32], the author examined the impact of code duplication while evaluating machine learning models. He declared that code duplication has a negative effect on the machine learning models' performance which sometimes inflated by 100%. So, he recommended removing any exact or near duplicates to have more reliable and accurate results.

In the image processing field[33], the near images between training and test sets are removed to improve the quality of machine learning results. Four classifiers, RF, DT, SGD, and perceptron classifiers are used, and their performance is recorded. There is a slight decrease in the accuracy of four classifiers after removing near images which means that duplicates can give deceptive performance.

The authors[10], introduced a new technique to detect the near duplicates. Their technique doesn't depend on columns to detect duplicates, but on the metadata that describes the datasets. Their experiments declared around 95% accuracy rate.

While [34] propose a new record linkage deduplication framework through six steps that detects and visualizes duplicates in the datasets.

Within another study[35] the natural language preprocessing and machine learning are used to detect the duplicates with 90% for area under the curve.

The probabilistic duplicates detection approaches, such as similarity-based derivation, and decision-based derivation, are presented in[36]. To effectively detect duplicates they examine the adaptation of search space reduction, like using Blocking and Sorted Neighborhood methods which effectively reduces the record pair comparisons.

A new duplication detection framework is proposed[37]. It depends on metric functional dependencies (MFDs) to enhance the detection accuracy. Their experimental results on three real datasets show an improvement of 25% and 34% in precision and recall respectively.

Most of the studies have focused on introducing the duplicates detection techniques and reducing the search space of the records comparison. However, few studies investigated the effect of duplicates on the analytical results. Most of them tackled the problem of probabilistic (near) duplicates[38]–[43] due to its complexity than the exact ones[44].

III. Clustering Evaluation Indices

A. Average Silhouette Coefficient

Average Silhouette Coefficient is a measure of the separation distance between clusters, (1). It is a graphical display of how well each data point is clustered. The silhouette coefficient ranges from −1 to +1, higher values that are closer to 1, indicate more coherent clusters[45].

$$s = \frac{b-a}{\max\{a,b\}} \tag{1}$$

Where (a) is the mean of the intra-cluster distance, the average dissimilarity of data points in the same cluster. And (b) is the mean value of the nearest-cluster distance, the cluster with the smallest average dissimilarity.

B. DUNN Index

The Dunn Index quantifies the ratio between the smallest distance between cases in different clusters and the largest distance within a cluster, (2). A high DI means better clustering since observations in each cluster are closer together, while clusters themselves are further away from each other[46].

$$D = \min_{i=1,\dots,k}\left\{\min_{j=i+1,\dots,k}\left[\frac{d(C_i, C_j)}{\max\limits_{l=1,\dots,k} diam(C_l)}\right]\right\} \tag{2}$$

Where k is the number of clusters, Ci is the ith cluster, d (Ci, Cj) is the distance between cluster Ci and Cj, and diam (Ci, Cj) is the diameter between the two clusters[46].

IV. Experimental Design

The main target is clarifying the effect of the presence and absence of duplication on data analytics. The two types of duplication are investigated, the deterministic and the probabilistic duplication. The effect of the deterministic duplicates is examined through five different classifiers: DT and RF follow decision tree manner, SVM follows linear algorithms that separates between classes with a hyperplane, NB works using Probabilistic technique, and LDA works on discriminating between classes by maximizing distance and minimizing scale between them. The five classifiers are tested against a dataset with different ratios of deterministic duplicates (5%, 10% and 15%). For more accurate tracking of the changes in the performance of each classifier, three evaluation metrics are used, Accuracy, F1-Score and MCC.

Due to the lack of the labelled benchmark datasets within probabilistic duplicates, unsupervised learning is applicable in this case. The sensitivity of four clustering algorithms toward probabilistic duplicates is evaluated through different ratios of probabilistic duplicates (Zero, 3.06% and 5.54%). The PAM, CLARA and K-means are partitioning based clustering. Where DBSCAN is a density-based clustering. The performance of the four clustering algorithms is validated using Average Silhouette Coefficient and DUNN Index. The full description of the used datasets, experimental steps, and results are presented below.

A. Datasets

For a classification task, a synthetic dataset obtained from UCI Machine Learning Repository [47] is used. Whereas two benchmark datasets[48] are used in clustering task. Table I shows more description about these datasets.

TABLE I. DESCRIPTION OF THE DATASETS

Dataset Name	Size	Features	Data Type	Dataset Type
Dry Beans	13,611	17	Numeric	Synthetic
DBLP	2616	5	Mixed	Benchmark
ACM	2294	5	Mixed	Benchmark

B. Experimental Design

A total of 84 experiments were conducted to measure the effect of different types of duplicates on the results of descriptive and predictive data analytics. The detailed experimental steps are presented below.

1) Experimental steps for deterministic duplicates: The sensitivity of five classifiers toward the presence and absence of deterministic duplicates is investigated through 60 experiments, divided into four groups. Group 1, zero duplicates, where the classifiers are tested against the original dataset to report the baseline performance. Group 2, the 5% duplicates, where 5% of the deterministic duplicates are inserted into the original dataset then the sensitivity of the classifiers are measured using the three-evaluation metrics, Accuracy, F1-Score and MCC. Group 3, the 10% duplicates, and Group 4, the 15% duplicates, follow the same structure as Group 2 except that the percentage of the deterministic duplicates is changed before testing the sensitivity of the classifiers toward the duplicates. The 60 experiments, 15 for each Group (5 classifiers x 1 dataset x 3 evaluation metrics), are conducted to clearly report the changes of each classifier's performance. The upper part of Fig. 1 shows a general description of the four groups of experiments.

2) Experimental steps for probabilistic duplicates: In this section the sensitivity of four clustering algorithms toward the presence and absence of probabilistic duplicates is measured through 24 experiments. The experimental steps have been divided into 3 groups, 8 experiments in each group (4 clustering algorithms x 1 dataset x 2 evaluation indices). Group A, the DBLP dataset has been integrated with ACM dataset to be one dataset with 0% of probabilistic duplicates. The four clustering algorithms are tested against the integrated dataset and have been evaluated using the Average Silhouette Coefficient and DUNN Index. In Group B, the percentage of the duplicates increased to 3.06%, then the sensitivity of the four clustering algorithms was tested and evaluated. In Group C, the ratio of probabilistic duplicates increased to 5.54%. Then the performance of the four clustering algorithms is measured and reported using the evaluation indices. In the three groups of experiments, Group A, B, and C, the percentage of near duplicates is measured using our proposed metric. The lower part of Fig.1 represents a general description of the three groups of experiments.

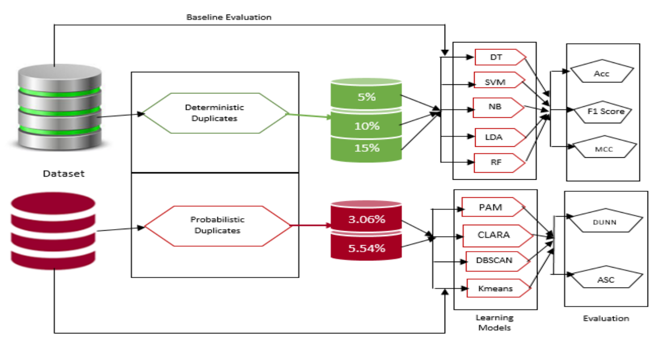


Fig. 1. Framework of all duplication experiments.

C. Experimental Results

The experimental results are addressed in two sections, the first section includes the results of the effect of deterministic duplicates on predictive analytics. Whereas the second section interprets the effect of probabilistic duplicates on descriptive analytics.

1) Results of deterministic duplicates effect: This section takes place into four subsections, the first subsection presents the baseline performance. The three other subsections show the experimental results of Group 2, 3 and 4 when having different ratios of deterministic duplicates.

a) The Zero Duplicates Experiment: The five classifiers are tested against the original dataset (zero duplicates) and their performance is reported and evaluated using three evaluation metrics as shown in Table II. The best classifier performance is recorded for RF followed by SVM, DT, LDA and NB respectively when evaluated by MCC, Accuracy and F1-Score measures.

b) The 5% Duplicates Experiment: In this group of experiments, 5% of deterministic duplicates are added to the original dataset. This duplicate percentage is measured using the simple equation, (3), that measures the ratio of duplicates.

$$Duplication = \frac{Total\ Number\ of\ Duplicate\ Rows}{Total\ Number\ of\ Rows} \quad (3)$$

The sensitivity of the five classifiers toward these duplicates is tested and recorded by the evaluation metrics. It is obviously clear that the performance of the five classifiers increased across all evaluation metrics compared with the baseline performance, Table III. 5% of duplicates positively affected the classifiers, which means that the five classifiers are positively sensitive to the existence of deterministic duplicates in this case. Higher performance means higher sensitivity. In this case, SVM reported higher sensitivity across all metrics, then comes RF, followed by DT and LDA. Whereas NB has the lowest sensitivity compared to other classifiers.

However, the performance of the five classifiers increased, this performance doesn't reflect the reality, true performance as in Table II, thus it can't be trusted or even used in making decisions.

c) The 10% Duplicates Experiment: Following the same steps in Group 2 experiment, the original dataset is contaminated by 10% of deterministic duplicates calculated by the same Eq., (3). As shown in Table IV, the performance of the five classifiers is still increasing after adding the 10% of duplicates compared with their performances in the previous two Groups of experiments, Table (II and III). The results declared that the five classifiers are still positively sensitive to the presence of deterministic duplicates. Furthermore, the performance of the five classifiers is still unreliable and can't be trusted. SVM and RF reported higher positive sensitivity in this case. NB has the lowest positive sensitivity, where DT and LDA have middle sensitivity toward 10% of exact duplicates.

d) The 15% Duplicates Experiment: The sensitivity of the five classifiers is measured and evaluated after adding 15% of deterministic duplicates into the original dataset. Table V shows that the performance of all classifiers except RF drastically decreased under the baseline performance. This indicates that DT, SVM, NB, and LDA are negatively sensitive to the existence of deterministic duplicates with large percentages more than 10%. However, the performance of RF is still increasing through the three-evaluation metrics (MCC, F1-Score, and Accuracy) compared with its performance across all previous experiments. In the case of negative sensitivity, NB has the highest sensitivity, then DT and SVM. Whereas LDA has the lowest negative sensitivity in this case.

Following are some observations after executing the four Groups of experiments:

- The performance of the RF classifier is positively affected by the presence of deterministic duplicates under small and large ratios, Fig. 2(a).
- As shown in Fig. 2(b), (c), (d), and (e), the DT, SVM, LDA and NB are positively sensitive to deterministic duplicates with certain percentages of duplicates ranging from 5% to 10%.
- However, they are negatively sensitive when having deterministic duplicates of more than 10%, Fig. 2(b), (c), (d), and (e).
- In general, the presence of deterministic duplicates limited to 10% has a positive effect on a classification task. As this percentage increases the effect of the deterministic duplicates differs based on the classifier used.
- Neither the positive nor the negative classifiers' sensitivity are good results. The classifiers in both cases are giving deceptive performance that doesn't reflect the reality. Hence, any decision taken based on a dataset contaminated by deterministic duplicates with any ratio is a completely wrong decision which can have negative implications on any business.

TABLE II. THE BASELINE PERFORMANCE

Dataset	Metrics	DT	SVM	NB	LDA	RF
Dry Beans	MCC	89.51%	91.00%	87.45%	87.86%	90.72%
	F1-Score	91.30%	92.50%	89.50%	89.70%	92.30%
	Acc.	91.31%	92.53%	89.54%	89.72%	92.31%

TABLE III. THE SENSITIVITY OF THE CLASSIFIERS WITH 5% OF DETERMINISTIC DUPLICATES

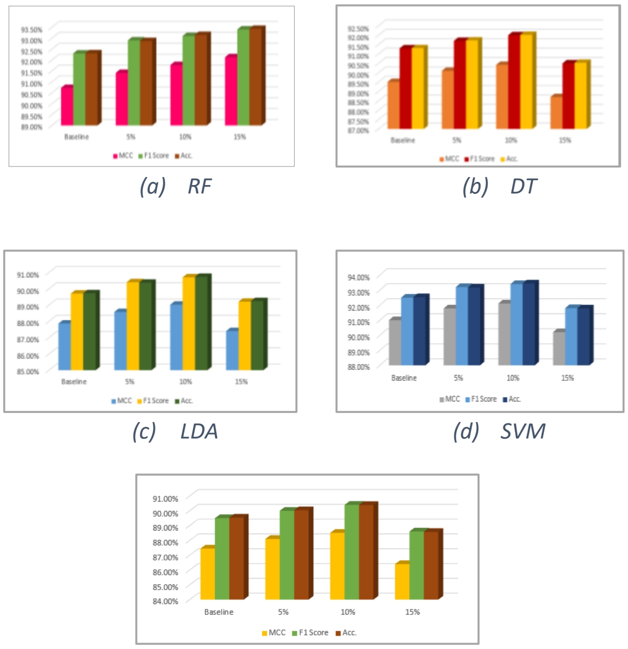
Dataset	Metrics	DT	SVM	NB	LDA	RF
Dry Beans	MCC	90.10%	91.78%	88.10%	88.57%	91.41%
	F1-Score	91.70%	93.20%	90.00%	90.40%	92.90%
	Acc.	91.72%	93.17%	90.04%	90.37%	92.87%

TABLE IV. THE SENSITIVITY OF THE CLASSIFIERS WITH 10% OF DETERMINISTIC DUPLICATES

Dataset	Metrics	DT	SVM	NB	LDA	RF
Dry Beans	MCC	90.42%	92.11%	88.51%	89.02%	91.77%
	F1-Score	92.00%	93.40%	90.40%	90.70%	93.10%
	Acc.	92.02%	93.44%	90.39%	90.73%	93.15%

TABLE V. THE SENSITIVITY OF THE CLASSIFIERS WITH 15% OF DETERMINISTIC DUPLICATES

Dataset	Metrics	DT	SVM	NB	LDA	RF
Dry Beans	MCC	88.70%	90.20%	86.40%	87.40%	92.13%
	F1-Score	90.50%	91.80%	88.60%	89.20%	93.40%
	Acc.	90.53%	91.78%	88.57%	89.24%	93.43%



(a) RF (b) DT (c) LDA (d) SVM (e) NB

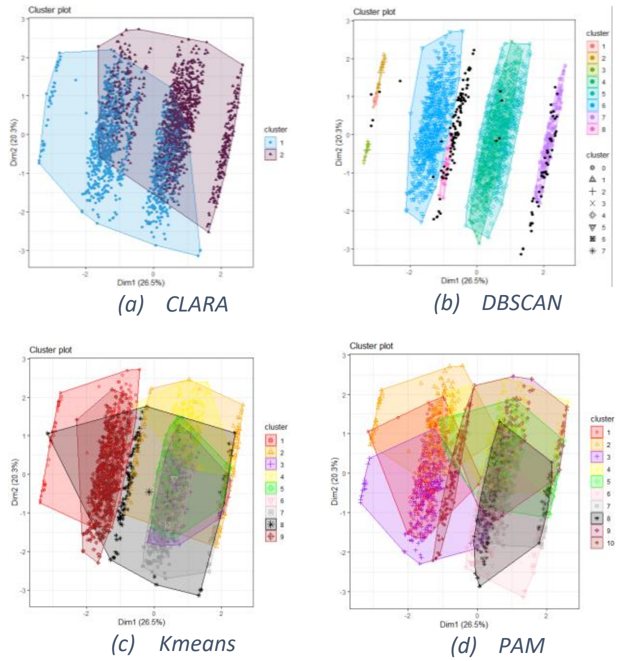
Fig. 2. Performance of the five classifiers with different ratios of deterministic duplicates.

2) Results of probabilistic duplicates effect: The experimental results on the effect of probabilistic duplicates are presented into three subsections. In the three subsections, the performance of the clustering algorithms with different ratios of probabilistic duplicates ranged from 0% to 5.54% is reported and evaluated using the Average Silhouette Coefficient and DUNN Index, both distance functions, Manhattan, and Euclidean functions, are used to calculate DUNN Index.

a) Group A Experiments: In this experimental group, the ACM and DBLP datasets are merged to have one cleaned integrated dataset. Then the four clustering algorithms are applied, and their quality is measured. Table VI shows the baseline performance of all clustering algorithms with zero duplicates. Generally, the performance of all clustering models is low, this is due to the dimensionality reduction problem. As stated in the literature [49]–[51], reducing the dimensions of the dataset improves the clustering results, however this is not the scope of this paper but showing the effect of near duplicates on the performance of the clustering models within different ratios.

TABLE VI. THE SENSITIVITY OF THE CLUSTERING ALGORITHM WITHOUT PROBABILISTIC DUPLICATES

Evaluation Indices	PAM	CLARA	Kmeans	DBSCAN
DUNN- Manhattan	0.018	0.017	0.016	0.096
DUNN- Euclidean	0.029	0.026	0.021	0.149
AVG.silhouette Coefficient	0.2	0.26	0.159	0.076



(a) CLARA (b) DBSCAN (c) Kmeans (d) PAM

Fig. 3. Performance of the four clustering algorithms with zero probabilistic duplicates.

The evaluation done by DUNN index using the two distance functions shows that DBSCAN has the higher performance, followed by PAM then, CLARA whereas Kmeans reported the lowest performance. However, this is not the case when evaluating the quality of the four algorithms with Average Silhouette Coefficient. CLARA reported highest value, followed by PAM then Kmeans whereas DBSCAN has the lowest performance in this case. The baseline performance of the four clustering algorithms is depicted in Fig. 3.

The clusters created using DBSACN and CLARA are clearly depicted, Fig. 3(a) and (b). Whereas the clusters created by PAM and Kmeans, as shown in Fig. 3(c) and (d), overlapped, and cannot be detected clearly.

b) Group B Experiments: In this group of experiments, the integrated dataset is contaminated by 3.06% of probabilistic duplicates. The following example is given to explain how this percentage is calculated.

A dataset is given with the following:

Total Number of Records = 2,667 Unique Records = 2,388

Duplicate Records = 279 Probabilistic Match

The Data Uniqueness of the given dataset using the old metric is:

$$Uniqueness = \frac{\text{Total Unique Rows}}{Total\ Number\ of\ Records} = 2388/2667$$

$$= 0.895$$

So, the data duplication will be:

$$Duplication = 1 - Uniqueness$$
$$= 1 - \frac{Total\ Unique\ Rows}{Total\ Rows}$$
$$= \frac{Total\ Number\ of\ Duplicate\ Rows}{Total\ Number\ of\ Rows}$$
$$= 1\text{-}0.895 = 0.105 = 10.5\%$$

However, the total number of duplicate records in the above equation refers to the deterministic duplicates and it does not consider the probability or the similarity by which the record is considered duplicates. To consider the probabilistic duplicate records, a new metric is proposed.

$$Data\ Duplication = \frac{1}{N}\sum_{i=1}^{N} P_i \qquad (4)$$

Where N is the total number of records, i is the record number, and Pi is the probability that I record has a duplicate. So, if the i record is unique then Pi = 0, if it has a deterministic duplicate, Pi = 1, and if it is probably duplicate, then Pi is the similarity measure between this record and its linkage.

Data Duplication = (0.2* 162 + 0.4*105+ 0.6*12) / 2667 = 81.6/2667 = 0.0306.

This means that the uniqueness of the integrated dataset = (1-0.0306) = 0.969.

Thus the 10.5% of duplicates measured using old metric doesn't reflect the similarity of the probabilistic records, whereas the proposed metric considers the probability of probabilistic duplicates while measuring, thus the true percentage of near duplicates in this case is 3.06%.

Fig. 4. shows different effects of near duplicates on clustering algorithms as the ratio of duplicates increased. Generally, DBSCAN is a robust algorithm as its performance doesn't change from its baseline. Whereas Kmeans is positively affected by such increased ratio. PAM and CLARA give different performances depending on the evaluation indices used. A detailed description of the sensitivity for each clustering algorithm is presented in Table VII.

The following observations can be derived from this group of experiment:

- DBSCAN has robust performance when evaluated by DUNN index. While it is negatively affected by the increase of the duplicates when Average Silhouette Coefficient used, which means that its performance degraded from its baseline performance. It decreased by 0.022%.
- The performance of Kmeans is increased by around 0.006% when evaluated by DUNN index. Kmeans has a positive sensitivity in this case. But this is not the case while using Average Silhouette Coefficient, Kmeans is negatively affected by near duplicates, its performance decreased by 0.012%.
- PAM shows different sensitivity, robust and negatively affected by near duplicates, within DUNN index using both distances Manhattan and Euclidean respectively. Also, Average Silhouette Coefficient shows negative sensitivity for PAM.
- CLARA has different sensitivity, positive, robust, and negative sensitivity, when evaluated by Manhattan, Euclidean, and Average Silhouette Coefficient respectively.

c) Group C Experiments: The ratio of probabilistic duplicates increased to 5.54% within the integrated dataset using the new metric, (4) within the given dataset:

Total Number of Records = 2,947 Unique Records = 2,389

Duplicate Records = 558 Probabilistic Match

Thus, Data Duplication = (0.2* 324 + 0.4*210+ 0.6*24)/2,947 = 163.2/2,947 = 0.0554

The four clustering algorithms are tested against the integrated dataset and their sensitivity is assessed using the evaluation indices. In Fig. 5, it is obviously clear that the performance of the four clustering algorithms except Kmeans is low. This means that the clustering algorithm, CLARA, PAM, and DBSCAN are negatively affected by the new ratio of probabilistic duplicates, Fig. 5(a), (b), and (d). However, Kmeans is positively affected by such increase in the duplicate's ratio, Fig. 5(c).

TABLE VII. THE SENSITIVITY OF THE CLUSTERING ALGORITHM WITH 3.06% OF PROBABILISTIC DUPLICATES

Evaluation Indices	PAM	CLARA	Kmeans	DBSCAN
DUNN- Manhattan	0.018	0.022	0.021	0.096
DUNN- Euclidean	0.023	0.026	0.028	0.149
AVG.silhouette Coefficient	0.19	0.2	0.147	0.054

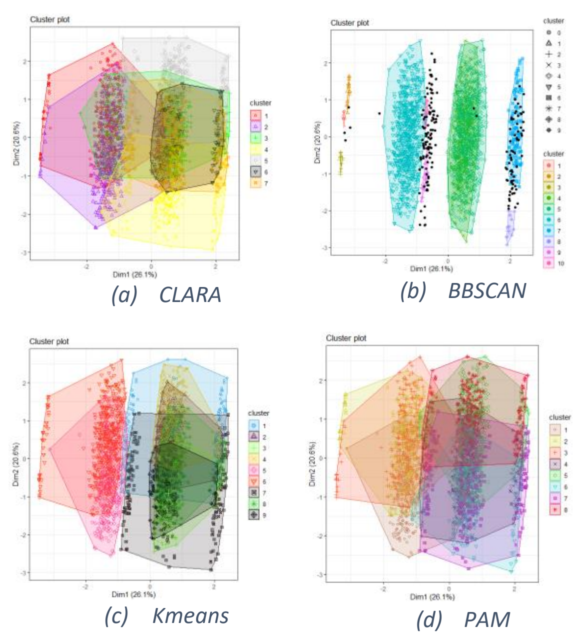


Fig. 4. Performance of the four clustering algorithms with 3.06% of probabilistic duplicates.

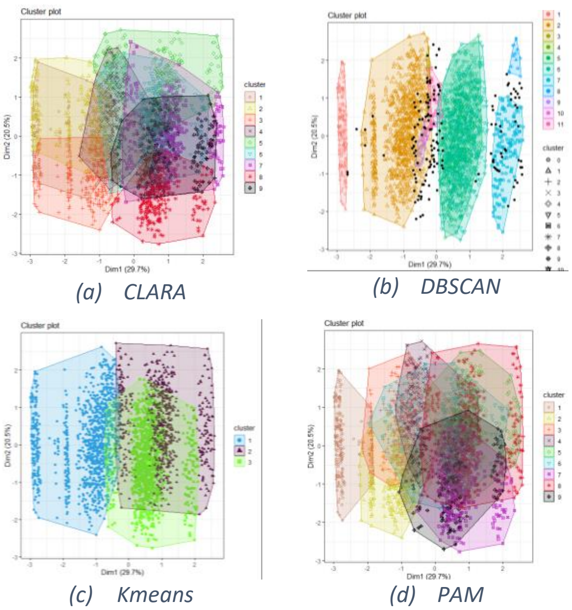


Fig. 5. Performance of the four clustering algorithms with 5.54% of probabilistic duplicates.

TABLE VIII. THE SENSITIVITY OF THE CLUSTERING ALGORITHM WITH 5.54% OF PROBABILISTIC DUPLICATES

Evaluation Indices	PAM	CLARA	Kmeans	DBSCAN
DUNN- Manhattan	0.011	0.012	0.023	0.045
DUNN- Euclidean	0.018	0.017	0.035	0.073
AVG. silhouette Coefficient	0.2	0.21	0.196	0.008

Based on the quantitative results presented in Table VIII, we can conclude that DBSCAN is highly sensitive toward a high ratio of duplicates, PAM came in the second rank of sensitivity, whereas CLARA has the lowest sensitivity compared with other algorithms. In this case the performance of Kmeans increases as ratio of probabilistic duplicates increased too. Thus, Kmeans is the only algorithm that has a positive sensitivity toward near duplicates.

Fig. 6 and Fig. 7 show the sensitivity observation of the four clustering algorithms after conducting all groups of Experiments, Group A (zero duplicates), Group B (3.06% duplicates), and Group C (5.54% duplicates).

1) Sensitivity measured by DUNN index

a) PAM Algorithm has a negative sensitivity toward the existence of probabilistic duplicates which means that its performance decreases as the percentage of the duplicates increases Fig. 6(a).

b) Fig. 6(b) shows that CLARA has a negative sensitivity only in the case of a high percentage of probabilistic duplicates, 5.54%. The performance of CLARA remains the same or slightly increases (by 0.005%) when the percentage of duplicates increased from zero to 3.06%.

c) DBSCAN has the same performance as CLARA, its performance is negatively affected by high ratios of the probabilistic duplicates, Fig. 6(c). Its performance remains the same when having zero and 3.06% of these duplicates.

d) On the other flip of the coin, comes Kmeans. Kmeans has a positive sensitivity to the existence of probabilistic duplicates with different ratios. Fig. 6(d) shows that the performance of Kmeans increases as the ratio of probabilistic duplicates increases too.

2) Sensitivity measured by average silhouette coefficient

a) As depicted in Fig. 7(a), the performance of PAM algorithm decreased from its baseline when 3.06% of duplicates were inserted into the dataset. Then it started to increase again as a larger ratio of probabilistic duplicates inserted. This indicates that PAM has no clear performance.

b) CLARA is a negatively sensitive clustering algorithm to the presence of near duplicates, Fig. 7(b). Its performance decreases when probabilistic duplicates exist. However, the performance of CLARA during different ratios of duplicates (3.06% and 5.54%) is almost the same, only 0.01% difference.

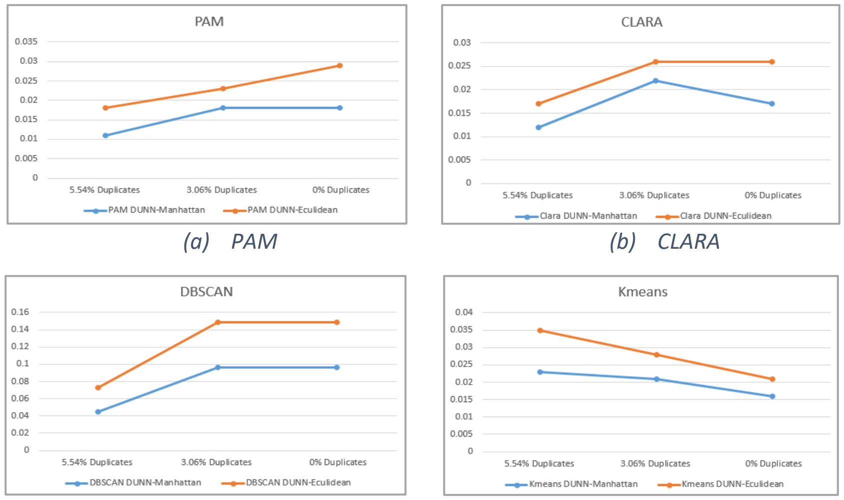
c) In Fig. 7(c), the performance of DBSCAN algorithm is negatively affected by the existence of probabilistic duplicates. The performance of DBSCAN clearly decreased when the ratio of the duplicates increased. Thus, we can conclude that DBSCAN is the highly sensitive algorithm toward probabilistic duplicates.

d) Fig. 7(d) declared that generally Kmeans is considered as a positively sensitive algorithm. Compared with the baseline, its performance increased (by 0.037%) as high ratios of probabilistic duplicates increased too.

Neither the positive nor the negative sensitivity of the clustering models toward probabilistic duplicates is good performance. The same as concluded from deterministic experiments, any decision taken in these cases will cause severe harm to any business either financially wise or management wise.

The experimental results agreed with what mentioned in different studies [15], [31]–[33], that the existence of both types of duplicates is somehow has an effect either positive or negative on the performance of the learning models.

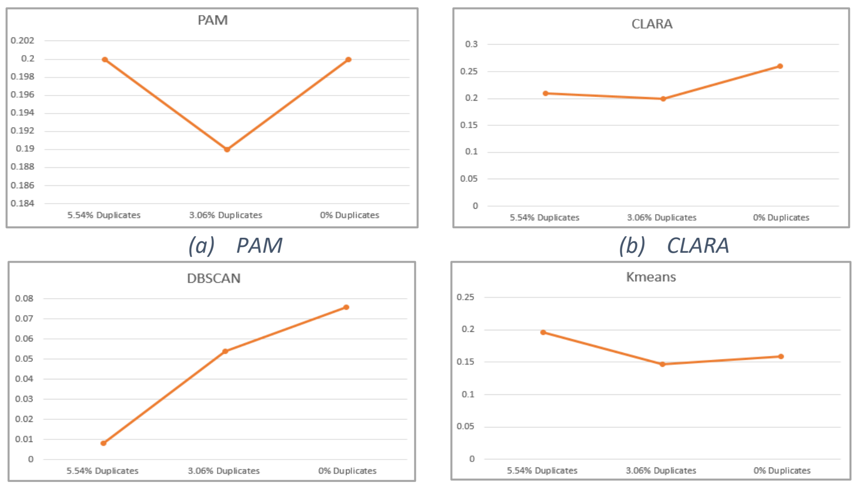
Based on the quantitative results for each experiment, the existence of duplicates has a significant effect on the performance of supervised learning models, whereas it has a moderate impact on the performance of unsupervised learning models, which is the opposite of what mentioned in [15]. Fig. 8 shows the sensitivity of the learning models toward duplicates.



(a) PAM *(b) CLARA*

(c) DBSCAN *(d) Kmeans*

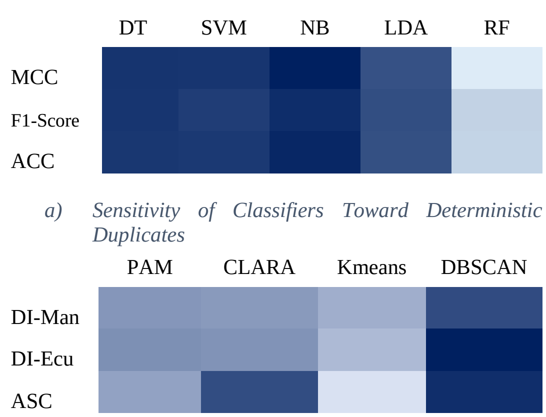
Fig. 6. The sensitivity of the four clustering algorithms using DUNN index.



(a) PAM *(b) CLARA*

(c) DBSCAN *(d) Kmeans*

Fig. 7. The sensitivity of the four clustering algorithms using average silhouette coefficient.



a) Sensitivity of Classifiers Toward Deterministic Duplicates

b) Sensitivity of Clustering Models Toward Probabilistic Duplicates

Fig. 8. The sensitivity of all learning models toward different types of duplicates.

In Fig. 8. the Dunn index using Manhattan function is referred to as (DI-Man), whereas Dunn index using Euclidean function as (DI-Ecu). The Average silhouette Coefficient is referred to as (ASC). Furthermore, the accuracy is shortened in (ACC).

The difference in the average performance from the baseline performance is calculated and depicted in Fig. 8. It is clearly noted that DT, SVM, NB, and LDA are highly sensitive toward the existence of deterministic duplicates. Whereas RF is positively sensitive in this case. Thus, generally deterministic duplicates have a severe influence on most of the classifiers.

However, most of the clustering models are moderately sensitive toward the existence of probabilistic duplicates. Only DBSCAN is extremely sensitive toward near duplicates. Thus, the existence of probabilistic duplicate has a moderate influence on the unsupervised learning models.

The existence of both types of duplicates can cause significant harm to the analysis results and then the whole business. Thus, it is highly recommended to remove any duplicates from the dataset before putting it through the processing phase.

V. CONCLUSION

Cleaning the data from quality issues like missing values, inconsistencies, duplication, etc. is an essential step if accurate decision is needed. Thus, in this paper the sensitivity of supervised (DT, SVM, RF, NB, and LDA) and unsupervised (DBSCAN, Kmeans, CLARA, and PAM) learning models toward the existence of two types of duplicates, probabilistic and deterministic, with different ratios (0%-15%) is investigated. The results of these models are validated using five evaluation metrics, MCC, F1-Score, Accuracy, Average Silhouette Coefficient, and DUNN Index. Three datasets are used through 84 experiments. The experimental results can be concluded as follows. First, both types of duplicates have an influence on the sensitivity of the learning models, which differs based on the learning model itself. Second, small percentages of deterministic duplicates have a positive impact on the five classifiers. This declares that the performance of five classifiers increased when exact duplicates exist with small ratios. Third, RF is the only algorithm that has a positive sensitivity toward exact duplicates with ratios more than 10%, whereas negative sensitivity is reported for the rest of classifiers under high ratio of exact duplicates. Fourth, Kmeans clustering algorithm has a positive sensitivity when having near duplicates with either small or large ratios. Fifth, generally the rest of clustering algorithms are negatively sensitive toward near duplicates especially with high percentages. Sixth, the proposed duplicate metric proved its effectiveness in measuring the true percentage of near duplicates within a dataset.

REFERENCES

[1] N. Micic, D. Neagu, F. Campean, and E. H. Zadeh, “Towards a Data Quality Framework for Heterogeneous Data,” in Proceedings - IEEE International Conference on Internet of Things, IEEE Green Computing and Communications, IEEE Cyber, Physical and Social Computing, IEEE Smart Data, iThings-GreenCom-CPSCom-SmartData, 2018, vol. Janua, pp. 155–162.

[2] J. Alipour and M. Ahmadi, “Dimensions and assessment methods of data quality in health information systems,” Acta Medica Mediterr., vol. 33, no. March, pp. 313–320, 2017.

[3] Y. Kodra, M. Posada, D. Paz, A. Coi, M. Santoro, F. Bianchi, F. Ahmed, Y. R. Rubinstein, and J. Weinbach, “Data Quality in Rare Diseases Registries,” in Rare Diseases Epidemiology: Update and Overview, Springer Cham, 2017, pp. 149–164.

[4] T. Romanian and E. Journal, “Data Quality Dimensions to Ensure Optimal Data Quality,” Rom. Econ. J., no. 63, pp. 89–103, 2017.

[5] M. Rehman, “Role of FCBF Feature Selection in Educational Data Mining,” Mehran Univ. Res. J. Eng. Technol., vol. 39, no. 4, pp. 772–778, 2020.

[6] M. I. Gabr, Y. M. Helmy, and D. S. Elzanfaly, “Data Quality Dimensions, Metrics, and Improvement Techniques,” Futur. Comput. Informatics J., vol. 6, no. 1, pp. 25–44, 2021.

[7] T. Avoundjian, J. C. Dombrowski, M. R. Golden, J. P. Hughes, B. L. Guthrie, J. Baseman, and M. Sadinle, “Comparing Methods for Record Linkage for Public Health Action: Matching Algorithm Validation Study,” JMIR Public Heal. Surveill., vol. 6, no. 2, pp. 1–12, 2020.

[8] J. A. Oostema, A. Nickles, and M. J. Reeves, “A Comparison of Probabilistic and Deterministic Match Strategies for Linking Prehospital and in-Hospital Stroke Registry Data,” J. Stroke Cerebrovasc. Dis., vol. 29, no. 10, 2020.

[9] Q. Chen, J. Zobel, and K. Verspoor, “Benchmarks for measurement of duplicate detection methods in nucleotide databases,” Database, no. January, 2017.

[10] M. Chevallier, N. Rogovschi, and F. Boufar, “Detecting Near Duplicate Dataset with Machine Learning,” Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 14, no. July, pp. 1–12, 2022.

[11] F. Naumann and M. Herschel, “An Introduction to Duplicate Detection.” 2010, pp. 1–87.

[12] W. Xia, D. Feng, H. Jiang, Y. Zhang, V. Chang, and X. Zou, “Accelerating content-defined-chunking based data deduplication by exploiting parallelism,” Futur. Gener. Comput. Syst., vol. 98, pp. 406–418, 2019.

[13] A. Gkoulalas-Divanis, D. Vatsalan, D. Karapiperis, and M. Kantarcioglu, “Modern Privacy-Preserving Record Linkage Techniques: An Overview,” IEEE Trans. Inf. Forensics Secur., 2021.

[14] R. Vaziri, M. Mohsenzadeh, and J. Habibi, “TBDQ: A pragmatic task-based method to data quality assessment and improvement,” PLoS One, vol. 11, pp. 1–30, 2016.

[15] Y. Zhao, L. Li, H. Wang, H. Cai, T. F. Bissyandé, J. Klein, and J. Grundy, “On the Impact of Sample Duplication in Machine-Learning-Based Android Malware Detection,” ACM Trans. Softw. Eng. Methodol., vol. 30, no. 1, pp. 1–36, 2021.

[16] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, “Evaluation Measures for Models Assessment over Imbalanced Data Sets,” J. Inf. Eng. Appl., vol. 3, no. 10, pp. 27–39, 2013.

[17] D. Chicco, “The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen ’ s Kappa and Brier Score in Binary Classification Assessment,” IEEE Access, vol. 9, no. Mcc, pp. 78368–78381, 2021.

[18] A. Tharwat, “Classication Assessment Methods,” Appl. Comput. Informatics, 2020.

[19] H. Řezanková, “Different approaches to the silhouette coefficient calculation in cluster evaluation,” 21st Int. Sci. Conf. AMSE, no. August, pp. 1–10, 2018.

[20] N. Saini, S. Saha, and P. Bhattacharyya, “Automatic Scientific Document Clustering Using Self-organized Multi-objective Differential Evolution,” Cognit. Comput., vol. 11, pp. 271–293, 2019.

[21] G. Papadakis, D. Skoutas, E. Thanos, and T. Palpanas, “Blocking and Filtering Techniques for Entity Resolution: A Survey,” ACM Comput. Surv., vol. 53, no. 2, 2020.

[22] B. H. Li, Y. Liu, A. M. Zhang, W. H. Wang, and S. Wan, “A Survey on Blocking Technology of Entity Resolution,” J. Comput. Sci. Technol., vol. 35, no. 61772268, pp. 769–793, 2020.

[23] V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, and K. Stefanidis, "End-to-End Entity Resolution for Big Data: A Survey," arXiv Prepr. arXiv1905.06397, vol. 1, no. 1, 2019.

[24] Z. A. Omar, M. A. A. Bakar, Z. H. Zamzuri, and N. M. Ariff, "Duplicate Detection Using Unsupervised Random Forests A Preliminary Analysis," 2022 3rd Int. Conf. Artif. Intell. Data Sci. (AiDAS). IEEE, 2022.

[25] A. Hindle and C. Onuczko, "Preventing duplicate bug reports by continuously querying bug reports," Empir. Softw. Eng., vol. 24, pp. 902–936, 2019.

[26] W. Elouataoui, I. El Alaoui, S. El Mendili, and Y. Gahi, "An End-to-End Big Data Deduplication Framework based on Online Continuous Learning," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. September, pp. 281–291, 2022.

[27] Y. Aassem, I. Hafidi, and N. Aboutabit, "Enhanced Duplicate Count Strategy: Towards New Algorithms to Improve Duplicate Detection," ACM Int. Conf. Proceeding Ser., pp. 1–7, 2020.

[28] F. Panse and J. Schildgen, "Similarity-driven Schema Transformation for Test Data Generation," EDBT, pp. 408–413, 2022.

[29] K. Loannis, T. Papenbrock, and F. Naumann., "MDedup: Duplicate detection with matching dependencies," Proc. VLDB Endow. 13.5, pp. 712–725., 2020.

[30] S. Chavhan, P. Patil, and G. Patle, "Implementation of Improved Inline Deduplication Scheme for Distributed Cloud Storage," in 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020, no. Icces, pp. 1406–1410.

[31] B. Barz and J. Denzler, "Do We Train on Test Data? Purging CIFAR of Near-Duplicates," J. Imaging, vol. 6, no. 41, 2020.

[32] M. Allamanis, "The adverse effects of code duplication in machine learning models of code," Onward! 2019 - Proc. 2019 ACM SIGPLAN Int. Symp. New Ideas, New Paradig. Reflections Program. Software, co-located with SPLASH 2019, pp. 143–153, 2019.

[33] C. Geier and A. Science, "Training on test data: Removing near duplicates in Fashion-MNIST.," arXiv Prepr. arXiv1906.08255 (2019)., no. 2019, pp. 1–4.

[34] O. Azeroual, M. Jha, A. Nikiforova, K. Sha, M. Alsmirat, and S. Jha, "A Record Linkage-Based Data Deduplication Framework with DataCleaner Extension," Multimodal Technol. Interact., vol. 6, 2022.

[35] F. Halawa, M. Abdul, and R. Mohammed, "Applying Machine Learning for Duplicate Detection, Throttling and Prioritization of Equipment Commissioning Audits at Fulfillment Network," in IISE Annual Conference and Expo 2022, 2022, p. 2022.

[36] F. Panse, M. Van Keulen, A. De Keijzer, and N. Ritter, "Duplicate detection in probabilistic data," Proc. - Int. Conf. Data Eng., no. Section II, pp. 179–182, 2010.

[37] Y. Huang and F. Chiang, "Refining Duplicate Detection for Improved Data Quality," TDDL/MDQual/Futurity@ TPDL, p. { 1–10}, 2017.

[38] A. Biloshchytskyi, A. Kuchansky, S. Biloshchytska, and A. Dubnytska, "Conceptual model of automatic system of near duplicates detection in electronic documents," 2017 14th Int. Conf. Exp. Des. Appl. CAD Syst. Microelectron. CADSM 2017 - Proc., vol. 4, no. 1, pp. 381–384, 2017.

[39] D. V Luciv, D. V Koznov, G. A. Chernishev, and A. N. Terekhov, "Detecting Near Duplicates in Software Documentation 1," Program. Comput. Softw., vol. 44, no. 5, pp. 335–343, 2018.

[40] Y. He and J. Gao, "Detecting Short Near-Duplicates with Semantic Relations," Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, vol. 2018-Novem, pp. 122–125, 2019.

[41] M. Fröbe, J. Bevendorff, J. H. Reimer, M. Potthast, and M. Hagen, "Sampling Bias Due to Near-Duplicates in Learning to Rank," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020., 2020, pp. 1997–2000.

[42] H. Matatov, M. Naaman, and O. Amir, "Dataset and Case Studies for Visual Near-Duplicates Detection in the Context of Social Media," arXiv Prepr. arXiv2203.07167 (2022)., 2022.

[43] K. K. Thyagharajan and G. Kalaiarasi, "A Review on Near-Duplicate Detection of Images using Computer Vision Techniques," Arch. Comput. Methods Eng., vol. 28, no. 0123456789, pp. 897–916, 2021.

[44] E. Provencal and L. Laperrière, "Detection of exact and near duplicates in phased-array ultrasound weld scan," Procedia Manuf., vol. 54, no. 2019, pp. 263–268, 2020.

[45] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient," Proceeding - 2nd Int. Conf. Technol. Policy Electr. Power Energy, ICT-PEP 2020, pp. 258–262, 2020.

[46] P. Rathore, Z. Ghafoori, J. C. Bezdek, M. Palaniswami, and C. Leckie, "Approximating Dunn's cluster validity indices for partitions of big data," IEEE Trans. Cybern., vol. 49, pp. 1629–1641, 2019.

[47] Dua and D. and Graff, "UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.," 2019.

[48] "https://dbs.uni-leipzig.de/research/projects/object_matching/benchmark_datasets_for_entity_resolution," Last Visited 29/10/2022.

[49] M. Alkhayrat, M. Aljnidi, and K. Aljoumaa, "A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA," J. Big Data, vol. 7, 2020.

[50] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," Inf. Fusion, vol. 59, no. May 2019, pp. 44–58, 2020.

[51] S. Sun, J. Zhu, Y. Ma, and X. Zhou, "Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis," Genome Biol., vol. 20, pp. 1–21, 2019.

A Hybrid Optimization Approach with Deep Learning Technique for the Classification of Dental Caries

Dr. Riddhi Chawla[1], Dr. Konda Hari Krishna[2], Dr. Araddhana Arvind Deshmukh[3]
K.V.Daya Sagar[4], Dr. Mohammed Saleh Al Ansari[5], Ahmed I. Taloba[6]
Dental School, Akfa University, Tashkent, Uzbekistan[1]
Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation[2]
Vaddeswaram, Guntur District, Andhra Pradesh, India[2]
Head and Associate Professor, Department of Artificial Intelligence and Data Science[3]
Marathwada Mitra Mandal's College of Engineering, Affiliated to Savitribai Phule Pune University[3]
Associate Professor, Department of Electronics and Computer Engineering[4]
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India[4]
Associate Professor, College of Engineering-Department of Chemical Engineering, University of Bahrain, Bahrain[5]
Department of Computer Science-College of Science and Arts in Qurayyat, Jouf University, Saudi Arabia[6]
Department of Information System-Faculty of Computers and Information, Assiut University, Assiut, Egypt[6]

***Abstract*—Due to the wealth of data available from different radiographic images, detecting dental caries has traditionally been a difficult undertaking. Numerous techniques have been developed to enhance image quality for quicker caries detection. For the investigation of medical images, deep learning has emerged as the preferred methodology. This study provides a thorough examination of the application of deep learning to object detection, segmentation, and classification. It also examines the literature on deep learning-based segmentation and identification techniques for dental images. To identify dental caries, several techniques have been used to date. However, these techniques are inefficient, inaccurate, and unable to handle a sizable amount of datasets. There is a need for a way that can get around these issues since the prior methods failed to do so. In the domains of medicine and radiology, deep convolutional neural networks (CNN) have produced amazing results in predicting and diagnosing diseases. This new field of healthcare research is developing quickly. The current study's objective was to assess the effectiveness of deep CNN algorithms for dental caries detection and diagnosis on radiographic images. The Convolutional Neural Network (CNN) method, which is based on artificial intelligence, is used in this study to introduce hybrid optimal deep learning, which offers superior performance.**

Keywords—Dental caries; deep learning; convolutional neural network

I. Introduction

Detecting dental caries was an exciting task due to the information gleaned from various radiographic images [1]. One of the most frequent chronic illnesses worldwide is dental caries. Especially large, obvious cavities on afflicted teeth respond well to such treatment methods. The effectiveness of conventional approaches has been hindered by the complicated visual features of dental caries images, which include concealed or difficult-to-access lesions [2]. The majority of common caries detection techniques focus on visual examination of the teeth, which also seems to be effective for both big, clearly apparent carious lesions and those that are only faintly visible but may be seen in hand mirrors. Dental radiography is used to find lesions that are hard to see or that are otherwise concealed from view [3]. The adoption of new instruments is advantageous since early diagnosis of dental caries lesions is a key therapeutic factor. The area of health care with the biggest growth is dentistry treatments.

By doing this, the danger, therapy, and identification of oral disorders are reduced [4]. Bitewing radiographs are frequently used by dentists to help them find tooth cavities. They concentrate on both the radiological data and the medical background of their patients. Detecting dental caries is a difficult process, and even skilled dentists may fail to spot carious lesions on bite-wing radiographs [5]. Dental caries detection methods have also typically used visual-tactile methods [6]. Visual-tactile techniques have poor sensitivity, particularly when applied to the surfaces of the proximal posterior teeth. Although radiographic techniques entail the utilization of ionizing radiation, they have a high sensitivity [7]. It might be challenging to detect cavities in certain teeth early on. Numerous carious lesions are discovered at a later stage [8]. The great majority of its screenings, however, are based on data analysis, even though dental radiography is often recognized as one of the most accurate screening methods for finding dental caries. Additionally, there is little research on identifying and diagnosing dental caries [9]. The effectiveness of deep learning techniques for spotting dental cavities in bite-wing radiographs. Blob identification on bite-wing radiographs helps us specify where and how to look for dental caries.

Blob detection is a statistical method for identifying particular regions in radiographic or digital images. Blobs are areas that stand out greatly from their surroundings. Blobs are also areas that contrast with their environment in brightness or darkness. Since their locations are derivative functions, blob detectors fall within the category of variable approaches. Blobs include data about different areas of interest that may be utilized in subsequent image processing with an increase in popularity in the processing of medical imaging [9]. Dentists frequently employ biting radiography to help identify tooth cavities. But this is a challenging procedure. Dentists use their clinical knowledge and the people's clinical background to validate their radiographic caries results. Even seasoned dentists tend to overlook cavities. Dental caries is brought on by an inflammation of the calcified tissue on the teeth. Through early discovery and treatment, they are easily preventable [10].

Effective and prompt treatment may result from the development of a reliable framework for the identification and categorization of dental caries [11]. Cavities are one of the most prevalent illnesses of the mouth that impact individuals of all ages, from infants to elders. It is a condition that weakens the structure of the teeth. Caries is mostly brought on by bacteria that turn the glucose and carbohydrate in food into acid. Because this acid melts and damages the minerals in the enamel, it causes significant damage to the tooth. Dental caries is a chronic condition that most patients experience a delayed onset of. It may impact the dentine tissue, which lies underneath the cementum and enamel on the root, as well as the enamel and cementum's exterior coverings. Infant and toddler caries is the term used to describe cavities in young children's primary teeth [12].

Dental hard tissues undergo periodic demineralization and remineralization as a result of the multifactorial, dynamic, biofilm-mediated illness known as dental caries. Including both permanent and primary teeth, caries can develop at any age and harm the tooth crown as well as the exposed root portions over time. Caries' start and development are impacted by the imbalance of protective and pathogenic factors. The categorization of individuals and groups into carious lesions risks identified is supported by the interplay of these variables, allowing for a more individualized approach to management [13]. Dental caries is a treatable, unequally distributed illness that imposes high costs on both the economy and the quality of life. The worldwide drop in cavities over the past several decades is attributed to the regular utilization of toothpaste. To find caries, experienced testers often conduct a thorough visual examination of clean teeth. Even though they are still frequently utilized, pointed dental probes offer little additional diagnostic value and can harm some teeth.

The detection and treatment of dental caries are changing [14]. Dentists presently use optical, tactical, and radiological data to identify somewhat severe alterations in the oral hard tissues. Instead of treating the illness itself, the clinical therapy of dental caries has focused on treating its symptoms by implanting solutions. Utilizing cutting-edge equipment, dentists would be able to identify developing dental caries before the clinically noticeable white spot. Dental caries is a continuing phenomenon that can be stopped at any stage—from the beginning to the end. When chemicals from bacterial respiration spread into enamel and dentine and break down the mineral, a dental cavity develops as a fungal illness. As a consequence of the bacteria's metabolism of fructose and glucose, organic acids are created [15].

But it has been known for more than a generation that bacteria digest meals produce acids that dissolve tooth minerals and cause dental disease. Dental caries is a contagious bacterial condition mostly brought on by the germs mentioned above feasting on the carbohydrates that people put in their mouths. In the dentistry industry, detecting dental caries is still referred to as the finding of demineralized areas, particularly cavities. Instead of responding medically before cavitation occurs and while the nutrition loss mechanism is still repairable, or at the very least may be implicated, the practising dentist "fixes" cavities by drilling and filling. Dental caries is not the actual disease; rather, it is symptomatic of an existing and past illness. As a result, it is crucial to document the beginning phase of parasite infection or the phases of lesion growth that are not cavitated. [16].

The key contributions of this study are:

- Initially, radiographic images from a large number of patients are collected, and the datasets are analyzed in the framework.
- Moreover, the collected images contain noise, which is filtered using the Wiener filter.
- The filtered images are subjected to segmentation, in which an enhanced Bat Whale optimization model has been used.
- The GLCM is used in the feature extraction to extract the features, and the Convolutional Neural Network is used for classification.
- The BWO-CNN model continues to perform best in terms of accuracy and provides independent advice to the medical professional in the classification of dental images.

The remainder of the essay is divided into the following sections. In Section 2, a summary of prior research on dental caries is provided. Section 3 discusses the proposed classification and framework. Section 4 presents the findings and a discussion of both the proposed and current approaches. The article comes to a close in Section 5 with a few more observations.

II. Related Works

Some of the literature related to dental carries are summarized as follows:

Deep learning for earlier dental caries identification in bitewing radiographs was proposed by Shinae Lee [17]. In this work, they evaluated whether a convolutional neural network model for caries identification on bitewing radiographs employing a U-shaped deep CNN may enhance physicians' effectiveness. The study conformed with all applicable standards of ethics. 304 bitewing radiographs altogether were utilized to train the convolutional neural network model, and 50 radiographs were employed to assess its efficacy. On the entire test dataset, the convolutional neural network model performed pretty accurately in terms of accuracy, recall, and F1-score, with scores of 63.29%, 65.02%, and 64.14 % respectively. The overall classification performance of all three physicians improved significantly when utilizing the convolutional neural network model findings as the data source to identify caries, as seen by an enhanced sensitivity ratio. Thus, physicians may benefit from using the deep learning model to effectively identify dental caries. Despite matter how much the diagnostics performance of the CNN model is improved, there will inevitably be some false-positive mistakes.

Prerna Singh and Priti Sehgal employed the optimal CNN-LSTM for the categorization and preparation of the G.V. Black dental caries [18]. A common health issue for many individuals worldwide is dental caries, a type of oral illness. Inflammation of the calcified tissue of the teeth leads to dental caries. Early detection and treatment make it simple to avoid and cure them. One approach that is well-liked around the globe is the G.V. Black Categorization for dental caries. Depending on the location of caries, it divides the condition into six classifications. To identify and diagnose dental caries on periapical dental images, a novel deep convolution layer network using an LSTM model is proposed in this research. This study's primary goal is to identify dental cavities and categorize them according to the G.V. Black Classification. Deep convolutional neural networks receive pre-processed periapical dental images as input. The Dragonfly optimization technique was used to optimize the suggested algorithm, which provided an accuracy of 96%. Examining and contrasting the suggested model with current state-of-the-art deep learning models is done through experiments. The suggested ideal CNN-LSTM model exhibits the best efficiency and aids in the categorization of dental images as a recommendation for the medical specialist.

The Logit-Based Artificial Bee Colony Optimization Approach was utilized by M. Sornam and M. Prabhakaran to classify dental caries using a back propagation neural network [19]. Caries, a bacterial infection that may cause oral discomfort, is a problem that can greatly reduce people's capacity to carry out their daily activities. Dental practitioners are working hard to discover a suitable strategy to minimize the miscategorization of dental caries phases and probable incorrect diagnoses. Utilizing dental X-ray images as the numerical input produced through a GLCM, a texture feature extraction process, this method is used to strengthen the back-propagation algorithm for an appropriate training and testing process, thus further achieving the best categorization accuracy. But the obtained accuracy is not higher when compared to the other optimization techniques.

For teledentistry, Devesh Saini et al. [20] presented a convolutional neural network for the identification of dental cavities. The primary cause of tooth loss is dental caries, a bacterial illness that progresses over time. This happens as a result of bad dental hygiene, which also causes several dental illnesses. This study attempts to identify dental cavities utilizing digital colour images at a preliminary phase, allowing for simple and efficient treatments. On a categorical collection with caries-containing and non-caries-containing images, training, validation, and testing have been done. Vgg16, Vgg19, Inception v3, and Resnet50 models are employed to attain classification performance, with Inception v3 achieving the greatest efficiency among them with training and validation accuracy of 99.89% and 98.95%, respectively, with the least loss when comparing to Vgg16 CNN models.

Hongbing Yu et al. [21] proposed a new method for Dental Caries Identification on the Child's First Permanent Molar. Recent studies have revealed a significant incidence and prevalence of caries in children's teeth, particularly in the first permanent molar, which might be very harmful to their overall health. Luckily, early identification and prevention can lessen treatment complexity and safeguard kids' dental health. To develop a new method for completely automated identification of dental caries on a child's first permanent molar, they present a unique caries identification and evaluation (UCDA) architecture in this research. The presented UCDA system primarily consists of a backhaul that is started with ResNet-FPN and two simultaneous task-specific subnetworks for area reduction and area categorization. These subnetworks were motivated by an effective in-network featured pyramid and anchor boxes.

Shashikant Patil et al. [22] proposed an algorithm for the detection of dental caries employing an adaptable neural network architecture. AI approaches have a long-lasting effect on biomedicine and provide broadly recognized results. The study's only goal is to examine the effectiveness of combining the Adaptive Dragonfly algorithm, Neural Network classifier, feature extraction, and categorization of dental images for accurately detecting caries. The suggested caries classification method in this case is made to accurately identify dental cavities. The two key stages of this technique are feature extraction and categorization. To assess performance, the 120 entire picture database is divided into three sets at random. Additionally, this classification of the test scenarios helps to guarantee speed improvement. Comparing the proposed MPCA-ADA model to other existing features allows for the model's performance to be assessed. The potential effectiveness of IP and NN algorithms for the identification and treatment of dental caries is emphasized in the study effort.

III. PROPOSED BWO-BASED CNN

The radiographic images of dental caries are first gathered. The images are then used for training and testing purposes. Images from dental caries radiography go through a pre-processing stage where the wiener filter is employed to eliminate the extra noise. In this work, dental cavities are detected early on using a combined bat and whale optimization technique. Dental caries and their severity are categorized using the proposed BWO-CNN approach. To operate enormous datasets and obtain a greater accuracy value, a convolutional neural network is used. Fig. 1 represents the various stages of the detection of dental caries.

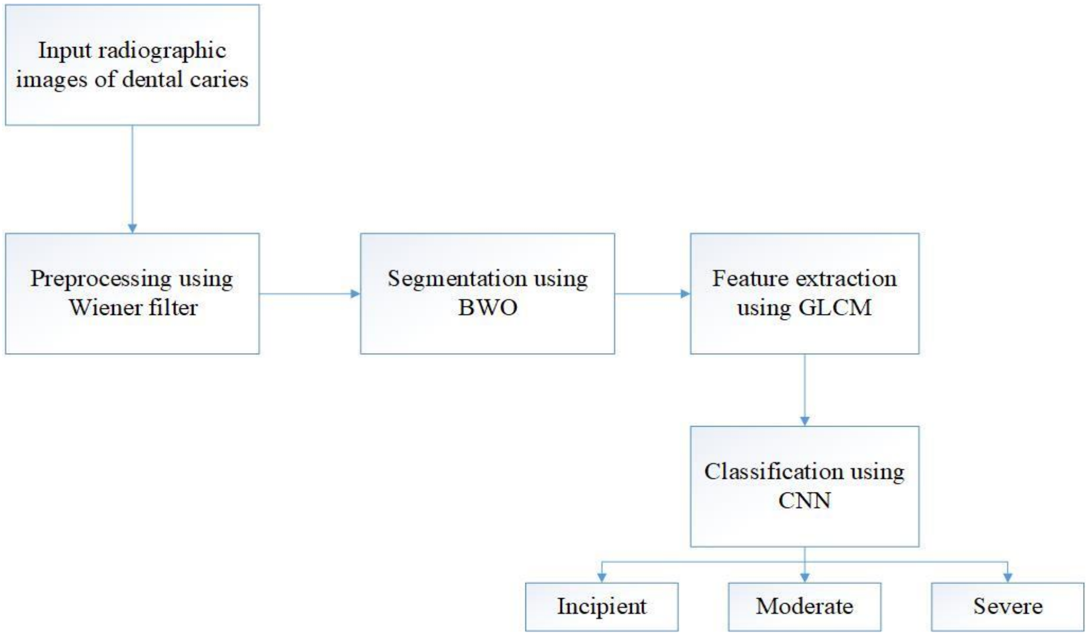


Fig. 1. Proposed BWO-CNN model.

A. Data Collection

This experiment employed data from about 10,000 datasets, including radiographic images of dental cavities. 50 per cent of the images from this set are utilized as training data and 50 per cent are used as testing data. Accordingly, 5000 radiographic images of dental caries are used as training data and 5000 radiographic images of dental caries are used as testing data.

B. Pre-processing

To detect dental caries, pre-processing is the initial step. It is employed to fill in dataset gaps and remove irrelevant data. The radiography images are impacted by independent and abnormal noises, which reduces the sample image analysis rate. The speckle noises, which can be induced by both internal and external sources, have a significant negative impact on radiography images. To minimize the noise in dental caries radiography images, the developed Wiener filter is used. The wiener filter is a linear filter used to conceal visual noise. The BWO-CNN model for identifying dental caries, therefore, makes advantage of the noise-reduced images. Given the wiener function,

$$w(a, b) = \sigma^2[m - p(a, b)] \quad (1)$$

Here σ^2 is the Gaussian noise variance, a and b are the dimensions of the pixel, and m is the feature of noise.

C. Segmentation using BWO

In radiography images, the segmentation procedure is primarily utilized to isolate the afflicted area. The effectiveness of the segmentation process is necessary for the execution of image processing. Image segmentation is frequently used to identify where dental cavities are present as well as any limitations posed by curves and lines in the images. Image segmentation divides the images into sets of pixels and labels the pixels in the samples. The primary objective of image segmentation in medical image processing is to identify dental cavities and produce sufficient data for further identification. Here, an upgraded and optimized Bat and whale algorithm are used to complete the image segmentation.

D. Bat Optimization algorithm

The echolocation of microbats was used to develop a metaheuristic bat algorithm. Bats typically utilize echolocation to find food. The bats typically emit brief pulses, but when they come upon food, both the pulse rate and frequency rise. The precision of the location is improved and the echolocation time is reduced due to the increase in frequencies. Each individual in the bat algorithm has a fixed position $u_i(n)$ and velocity $v_i(n)$in the search space, which are upgraded as the number of iterations rises. The new positions $u_i(n)$ and the velocities $v_i(n)$ can be calculated as follows:

$$u_i(n+1) = u_i(n) + v_i(n) \quad (2)$$

$$v_i(n+1) = v_i(n) + (u_i(n) - q(t)).f_i \quad (3)$$

$$f_i = f_{min} + (f_{max} - f_{min}).\beta \quad (4)$$

Where [0, 1] is the range of the random vector, which has a similar distribution. f min = 0, f max = 1, and q (t) represents the current global optimal solution.

1) Whale optimization algorithm: The whale optimization algorithm was inspired by humpback whales' use of bubble nets for hunting. The humpback whales will randomly locate their prey, begin to envelop it in a spiral-shaped bubble, and then swim up toward the surface. Two steps make up the whale Optimization algorithm. The prey is randomly hunted in the first phase, and then spiral bubble nets are used to enclose and attack the prey in the second phase. To find its prey, the whale must be in the best possible posture. The equation shown below can achieve this:

$$\overrightarrow{Y}(i+1) = \vec{Y}^*(i) - \overrightarrow{A} \cdot \vec{B} \quad (5)$$

$$\vec{B} = |\overrightarrow{C} \cdot \vec{Y}^*(i) - \vec{Y}(i)| \quad (6)$$

$$\vec{D} = 2 \cdot \vec{z} \cdot \vec{r} - \vec{z} \quad (7)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (8)$$

The ideal whale posture is represented by $\vec{Y}^*$ *(i).* The search agent's current location is given by the variables $\vec{Y}$*(i + 1)* and $\vec{B}$ (the range between the whale and its target). The coefficients are $\overrightarrow{C}$ and $\vec{D}$. The value of z decreases from 2 to 0, reaching 0 at the end of the iteration. A random variable, $\vec{r}$, has a value between 0 and 1.

2) Bat-Whale Optimization algorithm: To reliably identify dental caries, the traits of both whales and bats are merged and optimized. The whale's fitness is indicated in eq. (9)

$$y_i(n+1) = y_i(n) + v_i(n) - \overrightarrow{D} \cdot \vec{B} \quad (9)$$

Algorithm: BWO-CNN mechanism

Input: *Radiographic images of dental caries*

Output: *Detection of regions with dental caries*

Load input images

Train the input images P_i in the system, where i = 1 to n

Pre-processing of images

Let J(i) be the input image from the dataset

$w(a, b) = \sigma^2[m - p(a, b)]$ // Wiener filter

Segmenting the affected part // Bat and Whale Optimization

Initially, the affected part is detected using Eqn. (2)

If (initial region is met)

Update the fitness of Bat and Whale to find the next affected region using Eqn. (9)

Else

Find the initial location

Repeat until the stopping condition is reached // until all the affected region is identified

End if

Return

Feature extraction using GLCM

Classification of dental caries //CNN Classifier

E. Feature Extraction

Inexperienced data are transformed into numerical traits called features that can be used to save the information in the original data set. Each patient perceives images differently; these traits are derived from the total number of images captured. The image's dimensions are enhanced during testing, however, to identify dental cavities, the image's dimensions must be decreased. The processing of feature extraction helps to resolve this issue. In feature extraction, GLCM is used. Correlation, energy, homogeneity, contrast, entropy, and other second-order image attributes are examined to eliminate the probabilistic texture feature.

1) Energy: Squares with frequently higher grayscale values and unpredictable concentration values in images are summed to create energy. In Eq. (10), the energy of the input data is determined.

$$E = \sum_x \sum_y \{N(x, y)\}^2 \quad (10)$$

Where the squares in the image with grey levels are labelled as (x, y), and the images are denoted as N.

2) Contrast: The local contrast of an image is measured using features, and it is projected to be low when the mean concentration is even. The complete grayscale information of the main image is then projected, as per Eq. (11),

$$C = \sum_{a=0}^{I_y} a^2 \left\{ \sum_{x=1}^{I_y} \sum_{y=1}^{I_y} M(x, y) \right\} \quad (11)$$

(x, y) represents the square of an image's grayscale, N represents the images, and *I* stands for the images' grayscale.

3) Correlation: Using the correlation characteristics shown in Eq. (12) the numerical connections between the variables and the proportional dependency of grey levels on pixels.

$$C_o = \frac{\sum_x \sum_y (x, y) N(x, y) - \mu_u \mu_v}{\sigma_u \sigma_v} \quad (12)$$

The values of the mean and standard deviation in the images are represented as rows and columns by the letters μ_u, μ_v, $\sigma_{u,}$ and σ_v.

4) Entropy: Entropy is referred to in Eqn. (13), the anticipated high value of the range of grey levels being randomly generated sticks out.

$$En = -\sum_x \sum_y N(x, y) \log(N(x, y)) \quad (13)$$

F. Classification using Convolutional Neural Network (CNN)

To identify dental caries, CNN classifiers are used. It effectively evaluates the radiography images and extracts the necessary characteristics. Four layers are present in convolutional neural network classifiers: the convolutional layer, the max pooling layer, the fully connected layer, and the softmax layer. The CNN ranges the intensity values of the image pixel before the training phase. CNN is the model that trains the quickest. The input radiography image ought to be the same size. Each image in the training set has an equation that is presented in Eq. (14).

$$p(a, b) = \frac{o(a,b) - \mu}{\sigma} \quad (14)$$

1) Convolutional layer: A convolution layer was used to decrease the image of factors and extract the important features. The convolution layer includes rotation invariance, scaling partially invariant, and interpretation invariance. Both the over-fitting problem is lessened and the generalization idea is added to the fundamental framework. The convolutional layer collects a variety of images to analyze how complicated each one is. It is related to the qualities we seek in the given image. It is written as an eqn. (15).

$$f_i^m = x(\sum_{j \in M_i} f_j^{m-1} * p_{ji}^m + x_i^m) \qquad (15)$$

M_i – It stands for an input choice. The output has been specified as an additive bias b. When the sum of maps a and b across map i the kernel is applied to map i.

2) Max pooling layer: In order to reduce fitting and the size of the neurons used in the downsampling layer, this layer is used. The pooling layer reduces the rate of calculation, the size of the feature map, the training time, and regulates overfitting. Overfitting is calculated using an equation and is defined as 50% on test data and 100% on the training dataset.

$$x_{mab} = max_{(s,t) \in} f_{mst} \qquad (16)$$

Map, f_{mst} is the component at (s, t) in the pooling region pst, and it represents the immediate area.

3) Fully connected layer: In the field of classifying images, a Fully Connected Layer has been applied. The FC layers are placed after the Convolutional layers. The FC layer makes it easier to map a picture between the input and output. Fully linked layers make up the network's upper tiers. The output of the max pooling layer serves as the input for the fully connected layer.

4) Softmax layer: The Softmax layer converts the values into a normalized ratio distribution. The output is provided to the classifier as an input. The dental caries structure is implemented in the Softmax layer and the Softmax classifier is the standard contribution classifier. It is depicted in Eq. (17).

$$\sigma(\vec{X})_n = \frac{e^{x_n}}{\sum_{i=1}^{n} e^{x_i}} \qquad (17)$$

The whole process for detecting dental caries is shown in Fig. 2. The loading of input images starts the flowchart. The next step after loading the images is to train the loaded images. The Wiener filter is used during pre-processing to get rid of any unwanted noises from each image. The segmentation method uses the images that have had the noise removed. The segmentation process uses the suggested BWO-CNN. Using Eq. (2), the impacted region's position is initially determined. If the first region is met, use equation (9) to locate the subsequently impacted region; otherwise, return to the previous step and repeat the calculation of equation (2) until the initial region is fulfilled. Repeat Eq. (9) after identifying the subsequently impacted region until the stopping condition is satisfied. The process moves on to feature extraction if the halting condition is satisfied. GLCM is used for feature extraction. Following that, a Convolutional Neural Network is used to classify dental caries.

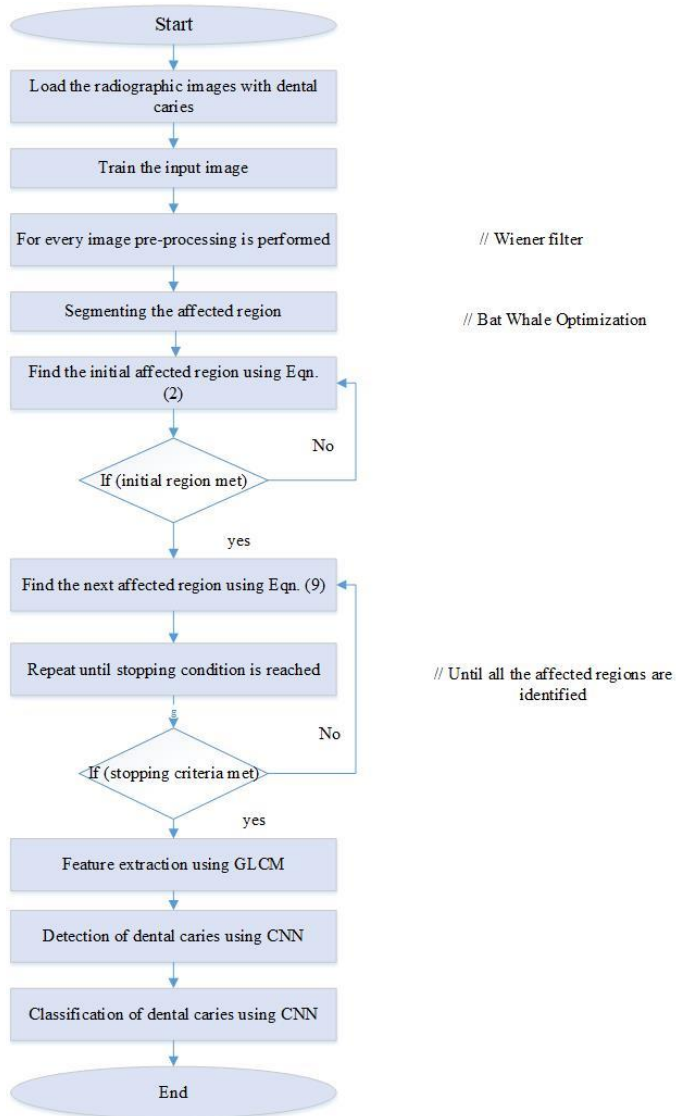


Fig. 2. Flow diagram of the BWO-CNN model.

IV. Results and Discussion

The existing methods consume a large time for processing and they cannot process a large number of data sets. More precise accuracy cannot be achieved with the existing methods. The employed BWO-CNN method overcame these limitations. The proposed strategy has been tested utilizing dental caries radiography datasets. The combined innovative bat and whale optimization-based convolutional neural network is employed to identify dental caries. Performance metrics are used to evaluate how well the given approach is presented.

A. Accuracy

Measured by accuracy, the system model's performance across all classes is evaluated. In general, it is the idea that all observations will be correctly predicted. In Eq. (18), accuracy is expressed.

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \quad (18)$$

B. Precision

The number of correct positive assessments that deviate from the overall positive evaluations is used to measure precision. By using an Eq. (19), you can compute the precise identification of dental caries,

$$P = \frac{T_{Pos}}{T_{Pos} + F_{Pos}} \quad (19)$$

C. Recall

The relationship between the number of genuine positives correctly labelled as positives and the overall number of positive samples is known as the recall. It describes the proportion of predictions that were correct in the Eq. (20) based detection of dental caries,

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \quad (20)$$

D. F1-Score

The F1-Score calculation is the combination of recall and precision. The F1-Score shown in Eq. (21) is calculated using precision and recall.

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (21)$$

Table I shows that the training accuracy of CNN and BWO-CNN is 99.2% and 99.55% respectively. The testing accuracy of CNN and BWO-CNN is 97.3% and 98.6% respectively. It is represented in Fig. 3.

TABLE I. PERFORMANCE EVALUATION BASED ON BWO-CNN

	CNN (%)	BWO-CNN (%)
Training accuracy	99.2	99.5
Testing accuracy	97.3	98.6

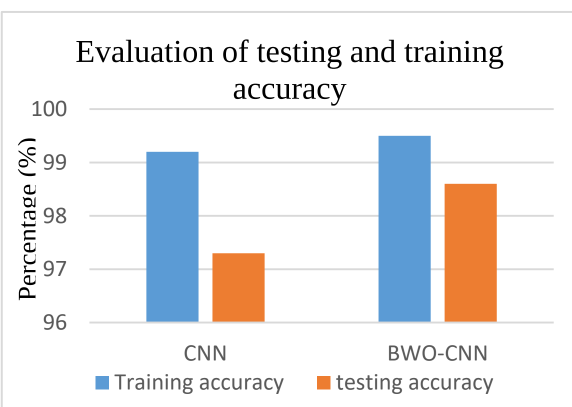


Fig. 3. Evaluation of testing and training accuracy.

TABLE II. ACCURACY COMPARISON OF VARIOUS METHODS

Method	Accuracy (%)
U-shaped deep CNN	63.29
CNN-LSTM	96
CNN	98.95
BWO-CNN	99.12

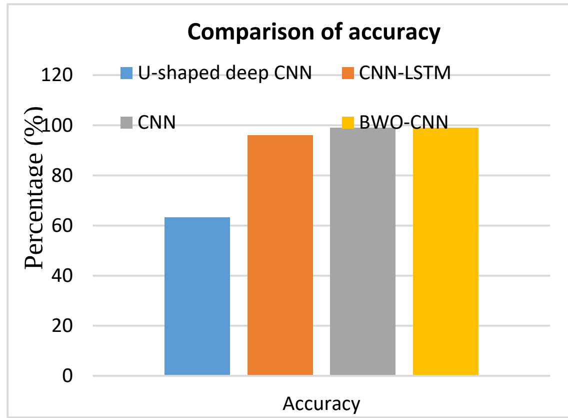


Fig. 4. Comparison of accuracy.

The projected technique Combined novel Bat and Whale Optimization-based Convolutional Neural Network achieves higher accuracy when compared to the existing dental caries detecting methods such as U-shapes CNN, CNN-LSTM, CNN, and BWO-CNN which are tabularized in Table II. Fig. 4 represents the Comparison of accuracy between BWO-CNN and other methods.

The projected technique Combined novel Bat and Whale Optimization-based Convolutional Neural Network achieves higher recall when compared to the existing dental caries detecting method like U-shaped CNN and BWO-CNN which are tabularized in Table III. Fig. 5 represents the Comparison of recall between BWO-CNN and U-shaped CNN.

TABLE III. COMPARISON OF RECALL

Method	Recall (%)
U-shaped deep CNN	65.02
BWO-CNN	92

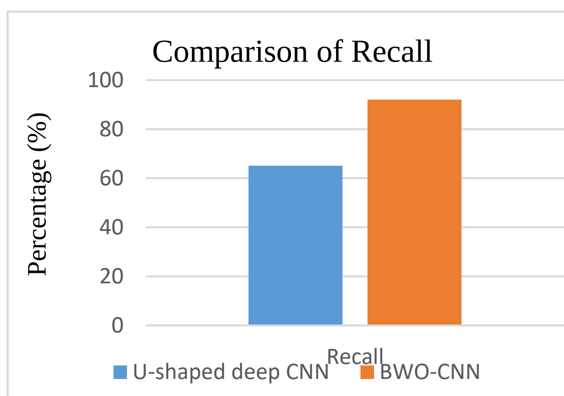


Fig. 5. Comparison of recall.

TABLE IV. COMPARISON OF F1-SCORE

Method	F1-score (%)
U-shaped deep CNN	64.14
BWO-CNN	91

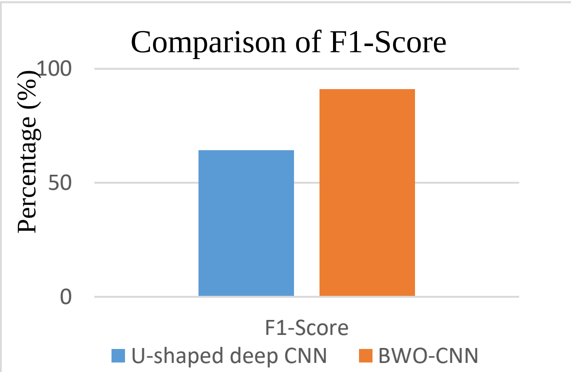


Fig. 6. Comparison of F1-score.

The proposed technique Combined novel Bat and Whale Optimization-based Convolutional Neural Network achieves a higher F1-score when compared to the existing dental caries detecting method like U-shaped CNN and BWO-CNN which are tabularized in Table IV. Fig. 6 represents the Comparison of recall between BWO-CNN and U-shaped CNN.

V. CONCLUSION

Although image processing in medical technology is expanding quickly nowadays, there are some situations where the complexity of the images makes it challenging to classify and identify diseases. The suggested approach is utilized to categorize, find, and segment the diseased or damaged portion. It focuses mostly on finding dental cavities. Individuals' information is first provided through radiographic images of patients with various types of dental caries, which are gathered. The unwanted effects or noise in the radiography images are then eliminated during the pre-processing step using a wiener filter. GLCM was used for feature extraction. Additionally, the damaged part is segmented using the suggested BWO-CNN method, and dental caries is classified using a CNN. Using BWO-CNN, high categorization and accuracy are also attained. The prediction accuracy of this model was 99.12% as a consequence. Even though the proposed method is effective, the performance can be increased by introducing LSTM. Therefore, in the future, a lot of research is being focused on creating a lightweight CNN-LSTM model that can shorten training time and is resistant to weight initiation effects.

REFERENCES

[1] Majanga and S. Viriri, "A Survey of Dental Caries Segmentation and Detection Techniques," Sci. World J., vol. 2022, p. e8415705, Apr. 2022, DOI: 10.1155/2022/8415705.

[2] V. Majanga and S. Viriri, "Automatic Blob Detection for Dental Caries," Appl. Sci., vol. 11, no. 19, Art. no. 19, Jan. 2021, doi: 10.3390/app11199232.

[3] "Socioeconomic Inequality and Caries: A Systematic Review and Meta-Analysis - F. Schwendicke, C.E. Dörfer, P. Schlattmann, L. Foster Page, W.M. Thomson, S. Paris, 2015." https://journals.sagepub.com/doi/abs/10.1177/0022034514557546 (accessed Nov. 14, 2022).

[4] M. M. Srivastava, P. Kumar, L. Pradhan, and S. Varadarajan, "Detection of Tooth caries in Bitewing Radiographs using Deep Learning." arXiv, Nov. 23, 2017. doi: 10.48550/arXiv.1711.07312.

[5] S. Valizadeh, M. Goodini, S. Ehsani, H. Mohseni, F. Azimi, and H. Bakhshandeh, "Designing of a Computer Software for Detection of Approximal Caries in Posterior Teeth," Iran. J. Radiol., vol. 12, no. 4, p. e16242, Aug. 2015, doi: 10.5812/iranjradiol.12(2)2015.16242.

[6] H. Nokhbatolfoghahaie, M. Alikhasi, N. Chiniforush, F. Khoei, N. Safavi, and B. Yaghoub Zadeh, "Evaluation of Accuracy of DIAGNOdent in Diagnosis of Primary and Secondary Caries in Comparison to Conventional Methods," J. Lasers Med. Sci., vol. 4, no. 4, pp. 159–167, 2013.

[7] U. K. Vural and S. Gökalp, "Diagnostic methods for dental caries used by private dental practitioners in Ankara," Niger. J. Clin. Pract., vol. 20, no. 3, Art. no. 3, Mar. 2017, doi: 10.4314/njcp.v20i3.

[8] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," J. Dent., vol. 77, pp. 106–111, Oct. 2018, doi: 10.1016/j.jdent.2018.07.015.

[9] K. T. M. Han and B. Uyyanonvara, "A Survey of Blob Detection Algorithms for Biomedical Images," in 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), Mar. 2016, pp. 57–60. doi: 10.1109/ICTEmSys.2016.7467122.

[10] P. Singh and P. Sehgal, "G.V Black dental caries classification and preparation technique using optimal CNN-LSTM classifier," Multimed. Tools Appl., vol. 80, no. 4, pp. 5255–5272, Feb. 2021, doi: 10.1007/s11042-020-09891-6.

[11] P. Singh and P. Sehgal, "Automated caries detection based on Radon transformation and DCT," in 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Jul. 2017, pp. 1–6. doi: 10.1109/ICCCNT.2017.8204030.

[12] "8.pdf." Accessed: Nov. 14, 2022. [Online]. Available: https://khorraman.ir/sites/default/files/articles/Dentol/8.pdf

[13] N. B. Pitts et al., "Dental caries," Nat. Rev. Dis. Primer, vol. 3, no. 1, p. 17030, Dec. 2017, doi: 10.1038/nrdp.2017.30.

[14] *T. Zero, "DENTAL CARIES PROCESS," Dent. Clin. North Am., vol. 43, no. 4, pp. 635–664, Oct. 1999, doi: 10.1016/S0011-8532(22)00818-7.*

[15] J. Featherstone, "Dental caries: a dynamic disease process," Aust. Dent. J., vol. 53, no. 3, pp. 286–291, 2008, doi: 10.1111/j.1834-7819.2008.00064.x.

[16] O. Fejerskov, "Concepts of dental caries and their consequences for understanding the disease," Community Dent. Oral Epidemiol., vol. 25, no. 1, pp. 5–12, 1997, doi: 10.1111/j.1600-0528.1997.tb00894.x.

[17] S. Lee, S. Oh, J. Jo, S. Kang, Y. Shin, and J. Park, "Deep learning for early dental caries detection in bitewing radiographs," Sci. Rep., vol. 11, no. 1, p. 16807, Dec. 2021, doi: 10.1038/s41598-021-96368-7.

[18] P. Singh and P. Sehgal, "G.V Black dental caries classification and preparation technique using optimal CNN-LSTM classifier," Multimed. Tools Appl., vol. 80, no. 4, pp. 5255–5272, Feb. 2021, doi: 10.1007/s11042-020-09891-6.

[19] M. Sornam and M. Prabhakaran, "Logit-Based Artificial Bee Colony Optimization (LB-ABC) Approach for Dental Caries Classification Using a Back Propagation Neural Network," in Integrated Intelligent Computing, Communication and Security, vol. 771, A. N. Krishna, K. C. Srikantaiah, and C. Naveena, Eds. Singapore: Springer Singapore, 2019, pp. 79–91. doi: 10.1007/978-981-10-8797-4_9.

[20] D. Saini, R. Jain, and A. Thakur, "Dental Caries early detection using Convolutional Neural Network for Tele dentistry," in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, Mar. 2021, pp. 958–963. doi: 10.1109/ICACCS51430.2021.9442001.

[21] H. Yu, Z. Lin, Y. Liu, J. Su, B. Chen, and G. Lu, "A New Technique for Diagnosis of Dental Caries on the Children's First Permanent Molar," IEEE Access, vol. 8, pp. 185776–185785, 2020, doi: 10.1109/ACCESS.2020.3029454.

[22] S. Patil, V. Kulkarni, and A. Bhise, "Algorithmic analysis for dental caries detection using an adaptive neural network architecture," Heliyon, vol. 5, no. 5, p. e01579, May 2019, doi: 10.1016/j.heliyon.2019.e01579.

FDeep: A Fog-based Intrusion Detection System for Smart Home using Deep Learning

Tahani Gazdar
Cyber Security Department-CCSE
University of Jeddah, Jeddah, Saudi Arabia

***Abstract*—Smart Home is an application of the Internet of Things (IoT) that connects smart appliances and the Internet. The emergence of Smart Home has caused many security and privacy risks that can lead to fatal damages to the user and his property. Unfortunately, Intrusion detection systems designed for conventional networks have shown their inefficiency when deployed in Smart Home environments for many reasons that rely basically on the resources-constrained devices and their inherent intermittent connectivity. So, an intrusion detection system designed for IoT and particularly Smart Home is mandatory. On the other hand, Deep learning shows its potential in enhancing the performance of Intrusion Detection Systems. According to recent studies, Deep learning-based intrusion detection systems are deployed either on the devices or in the Cloud. However, Deep learning models are greedy in terms of resources which makes it challenging to deploy them on Smart Home devices. Besides, in the IoT architecture, the IoT layer is far from the Cloud layer which may cause additional latency and jitter. To overcome these challenges, a new intrusion detection system for Smart Home deployed in the Fog Layer is proposed, it is called FDeep. FDeep will inspect the traffic using a Deep Learning model. To select the most accurate model, three Deep Learning models are trained using an IoT dataset named TON/IIOT, also the proposed models are compared to an existing one. The obtained results show that the long short-term memory model combined with the convolutional neural networks outperforms the other three models. It has the best detection accuracy compared to other Deep Learning models.**

Keywords—Fog computing; smart home; deep learning; IDS; classification

I. Introduction

The IoT environment is a system of smart devices that are interconnected and enabled with sensing and data transmitting capabilities [1]. IoT technology is used in Smart Home to enhance safety and convenience. Nevertheless, providing privacy and security in Smart Home environments is the key challenge facing their deployment. Many attacks threaten a Smart Home and may expose the data or properties to risks. One of the popular attacks is Denial of Service (DoS) on client-server applications. Eavesdropping is another common attack where an attacker intercepts data packets flowing in the network.

An intrusion detection system (IDS) analyzes the data packets to detect potential intrusion, and then generates alerts in case of any suspicious behaviour [2]. Particularly, anomaly-based IDSs are characterized by their efficiency in detecting zero-day and new attacks. This kind of IDS is based on Machine learning (ML) models that can automatically learn from experience and analyze patterns based on collected data. Deep Learning (DL) is a subcategory of ML, it can understand patterns using many layers of processing however it needs a large amount of data [3][4]. This represents one of the main challenges facing deploying IDS based on DL in IoT, in addition to the limited computation capabilities of the IoT devices. More important, most IDSs proposed in the literature are deployed on IoT devices which is not practical. DL models cannot run efficiently on resources constrained IoT devices because of the huge amount of data to analyze in addition to the complexity of DL models that needs high computation resources [2]. Further, many approaches implement the DL models in the Cloud, however, a centralized intrusion detection approach is not adequate for IoT because of their large scale and the caused latency. A distributed approach would be more appropriate. Besides, the efficiency of DL models degrades over time, so they need to be re-trained regularly with new data to enhance the detection capabilities mainly for zero-day and unknown attacks [5]. Most approaches re-train the models either on the Cloud or on desktops, then replace the old model with the new one. In these approaches, a huge amount of data packets is required. In the context of IoT and particularly Smart Home, the collected data may contain sensitive data related to the end user which may violate his privacy. Thus, an efficient, distributed, and accurate IDS for Smart Home is required to provide strong intrusion detection capabilities [2].

In this paper, new Fog-based IDS architecture dedicated to Smart Home named FDeep is proposed. The objective is to leverage Fog computing in IoT to provide a distributed, efficient, and high computation capacity IDS for Smart Home. An IoT application particularly a Smart Home consists of three main layers: the Edge layer and the Cloud layer and a Network layer. The Edge layer consists of a set of Smart devices responsible for sensing their environment, collecting data, and transmitting it through a Network layer to the Cloud Layer. This data is processed and analyzed in the Cloud layer to provide a wide range of applications and services to the end user. The two layers communicate which each other through gateways and routers. Generally, the Edge layer is too far from the Cloud layer, which might cause additional latency and jitter and requires high bandwidth consumption. Here, the paradigm of Fog computing comes to overcome this limitation by extending the Cloud and making it closer to the user. Fog computing provides computing, networking, and storage services to the Edge and Cloud layers in a distributed environment [6]. Unlike many approaches where DL models are implemented either in the Cloud or the Edge layers, in

FDeep the proposed DL model will be implemented in the Fog layer which represents the glue between the Cloud and Edge Layers. Besides, to maintain the DL model and avoid its degradation in terms of detection, periodic training of the DL model will be triggered using the data newly collected from the real network in the Fog layer.

The key contributions of the paper are summarized as follows:

- In FDeep, the DL model will be implemented in the Fog layer to guarantee low latency and reduce bandwidth consumption in addition to a fast inspection and detection of the attacks occurring in the IoT layer.
- An accurate DL model aiming to classify the attacks into seven types is proposed.
- An IoT dataset is used to train our models and select the most accurate one to be deployed in FDeep.
- The DL model to be deployed in FDeep will be periodically re-trained and updated to avoid its degradation. To do so, the data packets are collected from the IoT layer of the real network and fed to the model to re-train it.

The remainder of the paper is structured as follows. In Section II, the most recent approaches interested in DL-based IDS for IoT and its applications are discussed. Section III presents the architecture of FDeep in detail. In Section IV, the used dataset is described. In Section V, the experiments set up are presented and the obtained results are analyzed. Section VI concludes the paper and presents future work.

II. Related Work

DL-based Intrusion detection systems are widely adopted to inspect network traffic and detect intrusion and misbehaviour [2][7]. In [8], a DDoS detection system based on Deep Learning DL is proposed. The authors evaluated the performance of several DL approaches with ML techniques for DDoS attack detection. The obtained results point out the potential of Deep Learning in enhancing the accuracy of detection of DDoS attacks. A behavioural model is proposed in [9] to detect malicious traffic generated by compromised IoT devices using Autoencoder (AE). It is a kind of Deep Learning that learns unsupervised data. AEs are widely used in IoT security. In their study, AEs were used to extract features derived from cyber systems. In [10] a Deep Learning based intrusion detection system for IoT has been proposed. The authors used an SMO model to enhance the convergence time and feature extraction. Besides, they used the SDPN classifier to distinguish benign traffic from malicious one. The authors pretend that their model can handle datasets with redundant values and uncertain or missing data. However, the proposed model detects a few attacks: DoS, U2R, probe, and R2L. Unfortunately, all the above-discussed approaches don't mention how to implement the proposed DL models.

The authors proposed in [5] an attack detection system for IoT based on Deep Learning. They have implemented the proposed framework on Fog nodes. The authors have adopted an LSTM model to analyze and inspect the data packets collected from IoT devices to detect potential attacks. The maintenance of the model is performed in the Cloud layer to update the DL model and avoid its degradation in terms of detection capabilities. The limitation of the proposed architecture is the fact that the administrator must regularly check the accuracy of the model and maintain it manually. Additionally, the authors used dataset KDD CUP 99 which is not an IoT-specific dataset. However, according to [2], a model trained for a specific network or application may not be efficient for another network or application. In IoT, it is even recommended to design each kind of device with its own DL/ML model given the diversity of their sensing capabilities [2]. In [11], an LSTM model has been used to detect attacks in Fog-to-Things Communications. They proposed a distributed framework to detect attacks. However, their approach detects only attacks on Fog nodes and disregards the risk from IoT devices.

In [12], the authors proposed an intrusion detection model to detect attacks in a Fog computing environment. They have integrated CNN with LSTM to obtain an integrated and more accurate model. The limitation of this approach is the use of NSL-KDD which is not an IoT dataset. According to [13], many research studies are interested in intrusion detection systems for Fog-based IoT applications, however, these approaches detect only intrusions in Fog nodes. In [14], the authors proposed an IDS based on Ensemble learning for Fog-to-things environments. To enhance anomaly detection accuracy, the authors combine many classifiers to build two levels of classifications. The first layer detects the potential anomalies while the second layer classifies the detected anomalies. Similarly, DL models for attack detection in Fog-assisted IoT are proposed in [15] and [16]. Again, a non-IoT-specific dataset is used. Consequently, the accuracy of the model may be negatively affected once it is deployed in Fog-to-things environments.

To overcome the limitations of the above-discussed models we propose in this paper to implement the DL model in the Fog layer. Additionally, an IoT-specific dataset called TON/IIoT is used, it contains records collected from real IoT devices, and also it contains records about seven attacks.

III. Proposed IDS Architecture

A. Fog Computing and IoT

Recently, Cloud computing is the most popular computing paradigm. However, the spreading of IoT creates several challenges for Cloud computing. Particularly, most IoT applications require low latency which is not easily provided by Cloud computing since it has a centralized architecture that ill-suites the huge scale and the distribution of IoT. Moreover, the huge scale of IoT that relies basically on the high number of Smart devices increases the amount of data generated in the network. Sending this high amount of data to the Cloud will certainly need a high network bandwidth in addition to privacy concerns. Data generated in edge devices should be processed locally as most as possible. However, edge devices are usually resource-constrained devices unable to perform complex tasks and run complex protocols and models. Moreover, the intermittent connectivity due to the geographical distribution of edge devices makes them unable to benefit from uninterruptible Cloud services. Thus, an intermediate layer between the Cloud and the IoT devices is unavoidable [6].

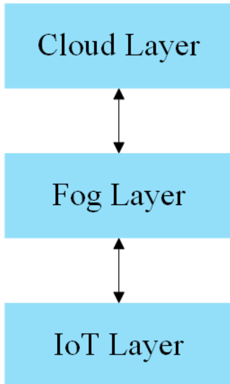


Fig. 1. 3-tier fog architecture.

Fog computing comes to bridge this gap and provide distributed networking, storage, control, and computing, capabilities between the Edge Layer and the Cloud layer [6]. Typically, the Fog layer is connected to the edge layer, which means that it is closer to the user than the Cloud layer. It is characterized by its flexibility to provide computation services to a large scale of end users in its proximity. Hence, it fosters decreasing the latency and accelerating decision-making. As depicted in Fig. 1, Fog computing has three-tier architecture: IoT layer, Fog layer, and Cloud Layer.

The lower layer is the IoT layer which is composed of the IoT devices such as smart vehicles smartphones, sensors, Cameras, smart appliances, etc. These IoT devices are heterogeneous in terms of computing capabilities, vendors, firmware, etc. They are geographically distributed, and their role is to collect data like temperature, fire, images, etc, and send them to the upper layer. The Fog layer consists of many Fog nodes responsible for providing computing, storage, and control services to the Cloud layer and IoT layer. A Fog node may be a router, a virtualized Fog server, or a simple data centre. Typically, the data is routed from the IoT layer to the Fog layer using smart gateways. The Cloud layer is the upper layer that consists of many servers characterized by their high storage and computational capabilities.

B. FDeep Architecture

As depicted in Fig. 2, the architecture of FDeep consists of 3 layers: IoT layer, Fog layer, and Cloud Layer. In the following, the role of each layer and the communications between them are detailed.

1) Data collection: All packets exchanged between the Smart devices in the IoT layer or to/from the Fog Layer pass through gateways. A network sniffer such as sflow or tcpdump, is deployed on these gateways to capture all network traffic, and the output will be saved in .pcap files. Then, the records are preprocessed to extract the useful features (timestamp, protocol, IP addresses, etc) and create CSV files containing pertinent data that will be inspected in the Fog layer to detect potential intrusions in the Smart Home. These data will be also used to re-train the DL model to improve its detection capabilities.

2) Data analysis and intrusion detection: The objective of FDeep is to detect potential intrusions in a Smart Home environment. To this end, a DL model is used to catch attacks and classify them. Three DL algorithms will be trained later and the best one in terms of detection accuracy will be deployed in FDeep.

In the current paper, the aim is to leverage Fog computing to build an efficient IDS by deploying the DL model on the Fog layer, doing so has many benefits. First, Fog computing is usually close to the user which reduces the latency and makes it best suited the real-time applications in IoT contexts. This feature is interesting in our context because rapid intrusion detection is required. The data will be processed close to the smart devices from which it is originating, no need to transmit it to the Cloud which reduces bandwidth consumption. Secondly, the OpenFog Security Group[1] has defined two main security goals of Fog computing. The first goal is the intrinsic security of Fog computing in terms of responsiveness, availability, fault tolerance, and trust. Fog computing can provide security services for the IoT layer since most of them are resources constrained such as identity verification, and end-point protection.

In the current study, the aim is to use a DL classifier to detect and classify intrusion in a Smart Home based on data received from the IoT layer, and since most IoT devices have limited resources, it seems interesting to implement the DL model in the Fog layer instead. It will provide a scalable, distributed and high computation environment for FDeep.

3) Updating the DL model: Generally speaking, DL models degrade over time because new applications may be deployed, and attacks may emanate. That is why maintaining the DL model is mandatory [5]. Hence, periodic training of the DL model will be triggered in the Fog Layer using the traffic newly captured from the IoT layer. Doing so has two main benefits, first, it will reduce the load on the Cloud layer in terms of computation, and also reduce the latency of transmitting the collected data from the IoT Layer to the Cloud. Secondly, it will make the model customized to the set of edge devices under the control of the Fog nodes. The edge devices deployed in the IoT layer may have different sensing capabilities and are vulnerable to different attacks, so it is interesting to have a DL model trained on the data received from the target devices [2].

The different models obtained in the Fog layer may be later shared with the Cloud layer to create a more generic model in a way similar to Federated Learning [17].

[1] https://opcfoundation.org/markets-collaboration/openfog/

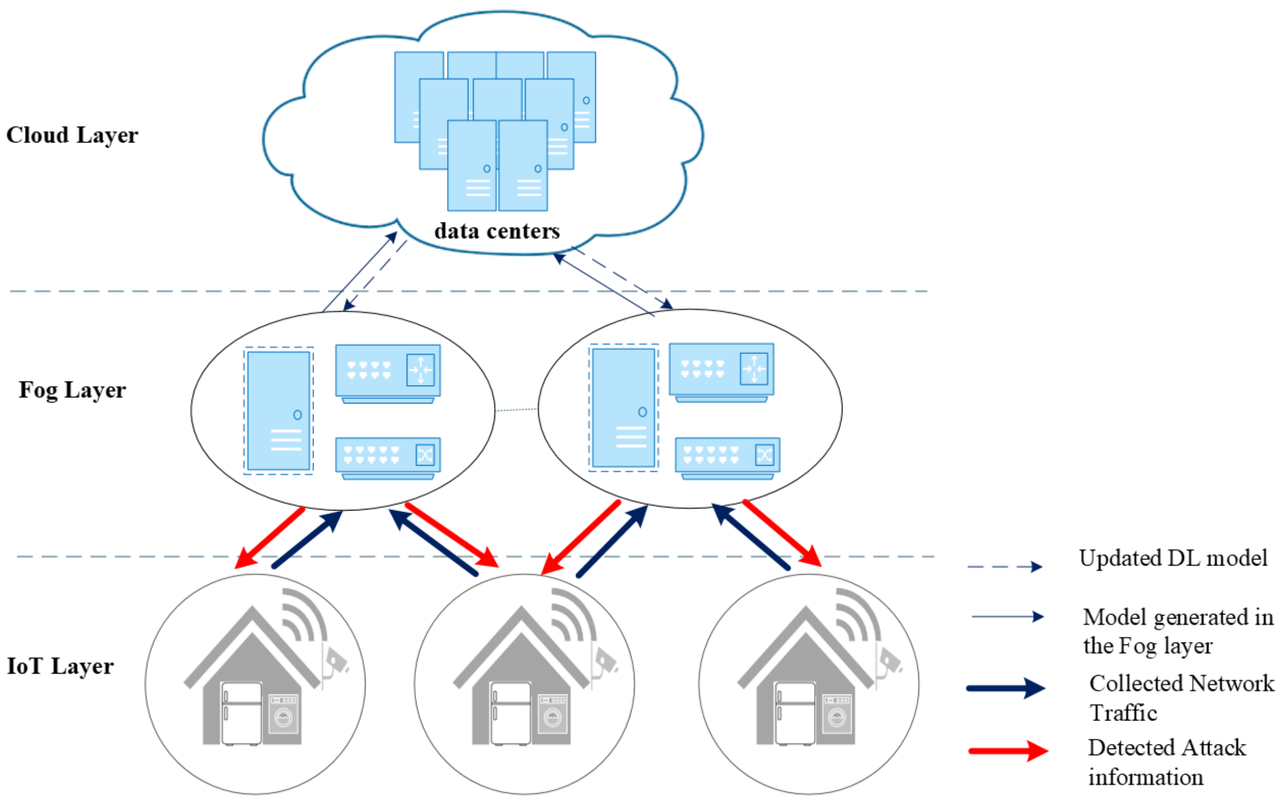


Fig. 2. FDeep architecture.

C. DL Model

To select the best DL model to deploy in FDeep, the performances of three DL algorithms are compared in terms of accuracy, precision, recall, and F1 score.

1) LSTM: Long Short-Term Memory Network (LSTM) is capable of learning order dependence in sequence prediction problems. It has feedback connections, that make it can process the entire sequence of data [18]. The state of an LSTM model is maintained over time by a memory cell named 'cell state'. This characteristic of LSTM makes it very suitable for environments where data should be processed sequentially like network traffic analysis, particularly, IoT systems. In intrusion detection, the major feature of LSTM is its ability to differentiate between benign and malicious traffic by only inspecting a small number of network data packets which makes it efficient for real-time analysis of the traffic [7].

2) CNN: CNN (Convolutional neural network) is among the most commonly used deep learning algorithm. The advantage of CNN is its ability to handle large datasets also its computational efficiency. CNN is also known for its efficiency in feature extraction. CNN consists of an input layer, a hidden layer, and an output layer. The hidden layer consists of many convolution layers, pooling layers, and a fully connected layer. First, the convolution layer extracts the features or information from the data using filters. The pooling layer is responsible for parameter reduction. Unlike conventional feature extraction algorithms, CNN can automatically learn the best features. Afterwards, the convolution maps are combined to form a unique vector known as CNN code [2]. Then, the classification layer receives the CNN code from the preceding levels and merges its characteristics to categorize the data [2].

3) CNN-LSTM: CNN-LSTM is a hybrid model that combines CNN and LSTM [19]. It is initially designed for visual time series predictions and textual generation from image sequences. In this model, the CNN is the features extractor, then the output is fed to LSTM which is the classifier. CNN-LSTM has been used also for intrusion detection [5].

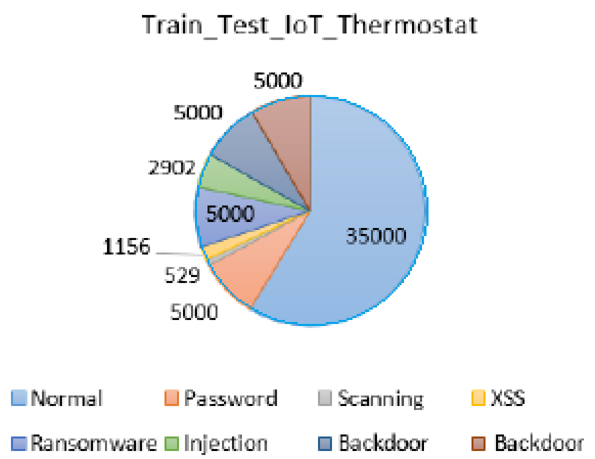
IV. Dataset

A new IoT dataset named TON/IIOT [20] is used. It is a recent dataset that consists of data collected from a real testbed of IoT devices like a Thermostat, smart fridge, weather sensor, motion light, etc. The dataset consists of many CSV files, each file corresponds to one device, and it contains data, features, and attacks that rely on the sensing capabilities of that device. The CSV files contain records that correspond either to benign or malicious traffic. The malicious records correspond to the following attacks: backdoor, DoS, DDoS, jamming, ransomware, scanning, password, XSS, cracking, injection, and man-in-the-middle [20]. In the current study, the results for only two devices are shown: the weather sensor and the Thermostat. Table I represents the set of features in the CSV file of the Thermostat.

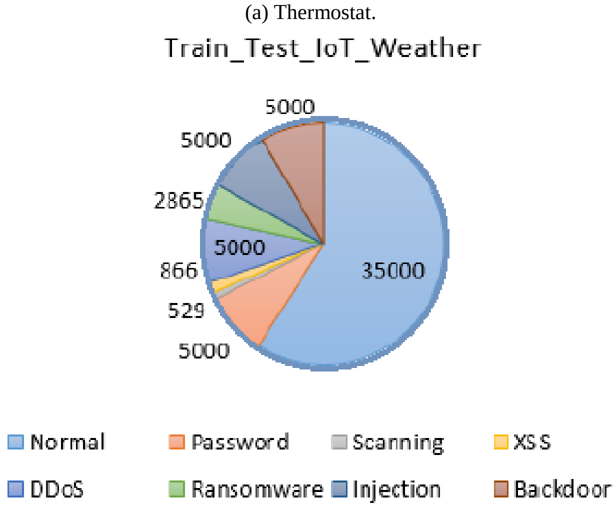
TABLE I. FEATURES IN THE THERMOSTAT CSV FILE

Feature	Description
Ts	Timestamp of sensor reading data
Date	Date of logging sensor's telemetry data
Time	Time of logging sensor's telemetry data
current_temperature	Current Temperature reading of a Thermostat sensor
thermostat_status	Status of a Thermostat sensor is either on or off
label	Identify normal and attack records, where '0' indicates normal and '1' indicates attacks
Type	A tag with normal or attack sub-classes, such as DoS, DDoS and backdoor attacks

The CVS file of the Thermostat contains 52774 where 17774 correspond to malicious traffic and the rest represent normal traffic. Similarly, the CSV file of the IoT Weather Sensor contains 54260 records whereas 19260 records correspond to attacks. The distribution of records among classes is shown in Fig. 3.



(a) Thermostat.



(b) IoT Weather sensor.

Fig. 3. Records distribution per attack [2].

V. EXPERIMENTS AND RESULTS

A. Experiments Set up

Google Colaboratory platform is used in all the experiments since it provides a panoply of Python libraries that support Deep Learning. The dataset is split into a training set that contains 75% of the total records, and 25% of the records are used in testing. DL models are implemented using Keras libraries. Softmax is used activation function since it is a multiclass classification problem and Adam is an optimizer to compute individual learning rates and enhance the accuracy of the models. The batch size is equal to 256, the total epochs number is equal to 300, and the LSTM output is equal to 100.

Since the attacks are classified into seven classes (attacks), we have opted for Categorical Cross Entropy as a loss function. The loss function is a way to evaluate to which extent the algorithm is modelling the dataset [21]. In the Categorical Cross Entropy, one category value encoded in binary is assigned to the output label. If it is in integer form, the 'keras.utils to_categorical' method is used to convert it into categorical encoding using.

B. Results Analysis

We have used the accuracy, precision, recall, and F1-score to evaluate the performance of the DL models[2]. The accuracy is the ratio of correctly classified inputs to the total number of inputs:

$$\text{Accuracy}=\frac{\#\text{ correctly classified input}}{\#\text{ of inputs}} \quad (1)$$

The precision is the ratio of correctly predicted positive observations to the total number of positive observations. It is computed as follows:

$$\text{Precision}=\frac{\#\text{True Positives}}{\#\text{True Positives}+\#\text{False Positives}} \quad (2)$$

Recall evaluates the proportion of malicious input correctly identified. Its mathematical equation is as follows:

$$\text{Recall}=\frac{\#\text{True Positives}}{\#\text{True Positives}+\#\text{False Negatives}} \quad (3)$$

F1 score shows if the model has correctly classified malicious input while minimizing false positives and false negatives rates:

$$\text{F1}=2*\frac{\text{Precision*Recall}}{\text{Precision+Recall}} \quad (4)$$

Fig. 4 and 5 represent the accuracy of LSTM, CNN, CNN-LSTM, and the LSTM model proposed in [19] for the IoT Weather sensor and the Thermostat. CNN-LSTM achieves the highest accuracy for both devices. It is about 98% for the IoT whether sensor and 75% for the Thermostat. The difference in performance between the devices is due to the unbalanced dataset for the Thermostat. Besides, as depicted in Fig. 4 and 5, the LSTM model proposed in [20] has the lowest accuracy among all models, it is about 85% for the IoT weather sensor and about 67% for the Thermostat. The accuracy of CNN-LSTM is close to LSTM which is about 97% for the IoT weather sensor.

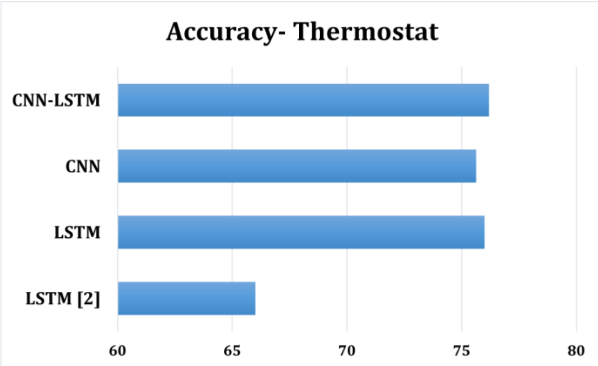


Fig. 4. Average accuracy–thermostat.

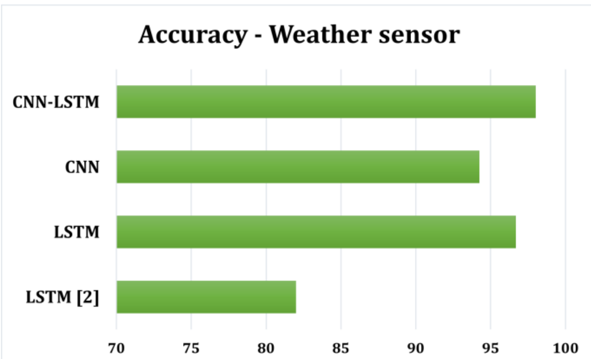


Fig. 5. Average accuracy-IoT weather sensor.

Fig. 6 and 7 plot the loss values for CNN-LSTM for the Thermostat and the IoT weather sensor respectively as a function of the epochs. The loss function is used to compute the distance between the current outputs of the algorithm training) and the expected output (testing). As shown in Fig. 6 and 7, as the training progresses, the value of the loss continuously decreases.

Fig. 7 shows that the loss decreases inversely to the epochs for the IoT weather sensor; it stabilizes and reaches a minimum of 0.05 after the first 100 epochs. Similarly, in Fig. 8 and 9, the accuracies of CNN-LSTM in the training and testing are compared for the Thermostat and IoT weather sensor respectively as a function of the epochs. As depicted in Fig. 8 and 9, the accuracies of the model in training and testing are close for both devices. It increases as the number of epochs increases to achieve the best accuracy at 300.

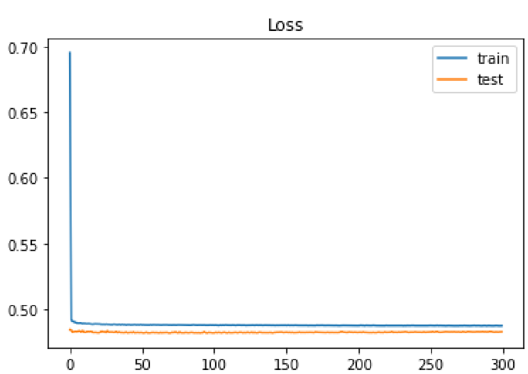


Fig. 6. CNN-LSTM loss values-thermostat.

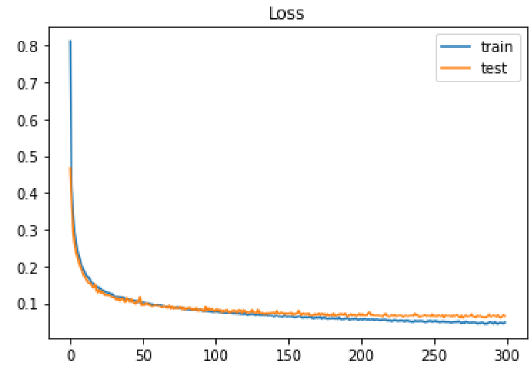


Fig. 7. CNN-LSTM loss values-IoT weather sensor.

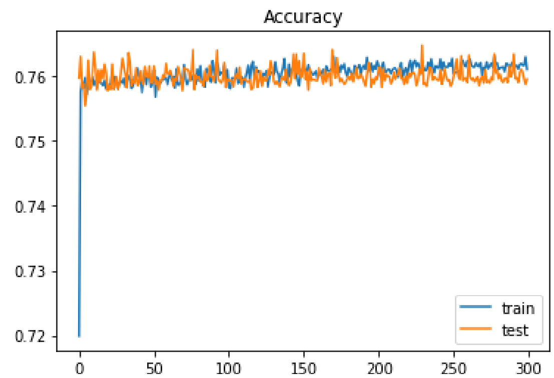


Fig. 8. Accuracy performance CNN-LSTM-thermostat.

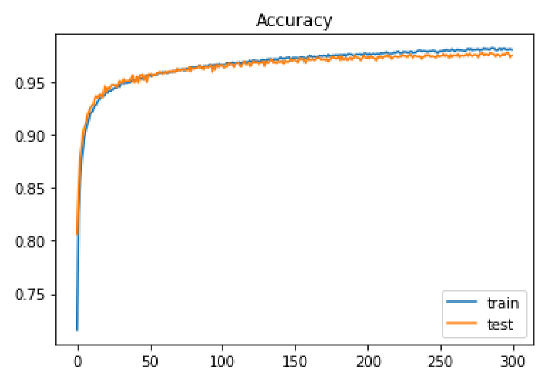


Fig. 9. Accuracy performance CNN-LSTM-IoT weather sensor.

Tables II and III show the values of accuracy, precision, recall and F1 score of the different DL models in multi-class classification for the Thermostat and the IoT weather sensor, respectively. Additionally, the performance of the proposed models is compared to the performance of the LSTM model proposed in [20] for the Thermostat and the IoT weather sensor. The results show that CNN-LSTM outperforms all the other DL models for both devices. It reaches the highest accuracy of 76.19% for the Thermostat and 98% for the IoT weather sensor. Additionally, the LSTM model proposed in [20] has the worst performance compared to other DL models.

TABLE II. PERFORMANCE OF DIFFERENT MODELS- IOT WEATHER SENSOR

	Accuracy (%)	Precision (%)	F1-score (%)	Recall (%)
LSTM[20]	82	82	80	81
LSTM	96.69	96.87	96.69	96.52
CNN	94.25	94.67	94.24	93.82
CNN-LSTM	98.02	98.02	98	97.97

TABLE III. PERFORMANCE OF THE DIFFERENT MODELS FOR IOT WEATHER SENSOR IN MULTICLASSIFICATION

	LSTM			CNN			CNN-LSTM		
	Precision (%)	Recall (%)	F1 score (%)	Precision (%)	Recall (%)	F1 score (%)	Precision (%)	Recall (%)	F1 score (%)
Backdoor	94	93	94	94	76	84	94	76	84
DDoS	92	88	90	78	90	83	78	90	83
Injection	94	95	94	96	89	92	96	89	92
Normal	100	100	100	100	100	100	100	100	100
Password	88	93	91	85	85	85	85	85	85
Ransomware	92	89	91	75	91	82	75	91	82
Scanning	99	96	98	99	96	97	99	96	97
XSS	87	88	87	89	77	83	89	77	83

It is also obvious from Tables II and III that CNN-LSTM has the best performance in terms of precision, F1-score and recall for both devices. Particularly, the performance of CNN-LSTM for the IoT weather sensor is better than the Thermostat. The outperformance of CNN-LSTM is due to the proven efficiency of LSTM in the classification in addition to the effective feature extraction performed by CNN.

For multi-class classification, every CSV file of a device has seven attacks in addition to the normal or 'benign class'. Only the performance of the IoT Weather Sensor is presented in detail for multi-class classification because of space limitations. Table IV shows the precision, recall, and F1 score for each DL model.

As depicted in Table IV, CNN-LSTM outperforms LSTM in terms of overall performance for 5 classes (among 8), however, the performance of CNN-LSTM is close to that of CNN. The unbalance of the dataset has a negative impact also on the performance of all DL models.

TABLE IV. PERFORMANCE OF DIFFERENT MODELS - THE THERMOSTAT

	Accuracy (%)	Precision (%)	F1-score (%)	Recall (%)
LSTM [20]	66	45	54	67
LSTM	76	100	79.4	66.28
CNN	75.63	100	79.31	66.17
CNN-LSTM	76.19	100	79.31	66.17

VI. CONCLUSION

In this paper, a new distributed IDS system based on Fog-computing to detect intrusions in a Smart Home is proposed. The architecture consists of three layers: IoT layer, Fog layer, and Cloud layer. The IDS detects the intrusion by inspecting the traffic collected from the IoT devices in the IoT layer. A DL model implemented in the Fog layer has been used to inspect the data packets. The implementation of the DL model in the Fog layer provides a high computation environment for the IDS, reduced latency, and avoids implementing the DL model in the end-user devices. To maintain and update the DL model, the model is re-trained periodically in the Fog layer using the data collected from the IoT layer. Many experiments have been conducted to evaluate the performance of many DL models to select the appropriate and efficient one to be implemented in the IDS. In the training, an IoT dataset named TON/IIoT is used, it contains the data collected from a real testbed composed of many IoT devices. The results show that CNN-LSTM has the best performance in terms of accuracy, precision, recall, and F1-score compared to other DL models and the model proposed in [20]. In future work, the performance of the architecture in terms of response time and robustness will be evaluated.

REFERENCES

[1] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," 2018 10th International Conference on Cyber Conflict (CyCon), 2018, pp. 371-390, doi: 10.23919/CYCON.2018.8405026.

[2] Asharf, Javed, Nour Moustafa, Hasnat Khurshid, Essam Debie, Waqas Haider, and Abdul Wahab. 2020. "A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions" Electronics 9, no. 7: 1177. https://doi.org/10.3390/electronics9071177.

[3] T. A. Abdullah, W. Ali, S. Malebary, and A. A. Ahmed, "A review of cyber security challenges, attacks and solutions for internet of things based smart home," Int. J. Comput. Sci. Netw. Secur., vol. 19, no. 9, p. 139,2019.

[4] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," 2018 10th International Conference on Cyber Conflict (CyCon), 2018, pp. 371-390, doi: 10.23919/CYCON.2018.8405026.

[5] A. Samy, H. Yu and H. Zhang, "Fog-Based Attack Detection Framework for Internet of Things Using Deep Learning," in IEEE

Access, vol. 8, pp. 74571-74585, 2020, doi: 10.1109/ACCESS.2020.2988854.

[6] M. De Donno, K. Tange and N. Dragoni, "Foundations and Evolution of Modern Computing Paradigms: Cloud, IoT, Edge, and Fog," in IEEE Access, vol. 7, pp. 150936-150948, 2019, doi: 10.1109/ACCESS.2019.2947652.

[7] B. Farhan, A. Jasim, Survey of Intrusion Detection Using Deep Learning in the Internet of Things. Iraqi Journal for Computer Science and Mathematics. 2021, 10.52866/ijcsm.2022.01.01.009.

[8] Susilo, Bambang, and Riri Fitri Sari. 2020. "Intrusion Detection in IoT Networks Using Deep Learning Algorithm" Information 11, no. 5: 279. https://doi.org/10.3390/info11050279.

[9] Rezvy S, Luo Y, Petridis M, Lasebae A, Zebin T (2019) An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks. In: 2019 53rd annual conference on information sciences and systems (CISS), IEEE, pp 1–6.

[10] Y. Otoum, D. Liu, A. Nayak, DL-IDS: a deep learning–based intrusion detection framework for securing IoT. Transactions on Emerging Telecommunications Technologies, Volume 33, Issue 3, March 2022 pp e3803 https://doi.org/10.1002/ett.3803.

[11] A. Diro and N. Chilamkurti, "Leveraging LSTM Networks for Attack Detection in Fog-to-Things Communications," in IEEE Communications Magazine, vol. 56, no. 9, pp. 124-130, Sept. 2018, doi: 10.1109/MCOM.2018.1701270.

[12] Kalaivani K., , and Chinnadurai M. "Ensemble Deep Learning Intrusion Detection Model for Fog Computing Environments," International Journal of Software Innovation (IJSI) 10, no.1: 1-14, 2022. http://doi.org/10.4018/IJSI.303587.

[13] Cristiano Antonio de Souza, Carlos Becker Westphall, Renato Bobsin Machado, Leandro Loffi, Carla Merkle Westphall, Guilherme Arthur Geronimo, Intrusion detection and prevention in fog based IoT environments: A systematic literature review, Computer Networks, Volume 214, 2022, 109154, ISSN 1389-1286.

[14] Illy, Poulmanogo & Kaddoum, Georges & Miranda, Christian & Kaur, Kuljeet & Garg, Sahil, Securing Fog-to-Things Environment Using Intrusion Detection System Based On Ensemble Learning, IEEE Wireless Communications and Networking Conference (WCNC), 2019. 10.1109/WCNC.2019.8885534.

[15] Sahar, Nausheen, Mishra, Ratnesh, Kalam, Sidra, Deep Learning Approach-Based Network Intrusion Detection System for Fog-Assisted IoT, Proceedings of International Conference on Big Data, Machine Learning and their Applications, Springer Singapore, pages 39--50, 2021, doi: //doi.org/10.1007/978-981-15-8377-3_4.

[16] A. Abeshu and N. Chilamkurti, "Deep Learning: The Frontier for Distributed Attack Detection in Fog-to-Things Computing," in IEEE Communications Magazine, vol. 56, no. 2, pp. 169-175, Feb. 2018, doi: 10.1109/MCOM.2018.1700332.

[17] Zhang, K., Song, X., Zhang, C. et al. Challenges and future directions of secure federated learning: a survey. Front. Comput. Sci. 16, 165817 (2022). https://doi.org/10.1007/s11704-021-0598-z.

[18] R.H. Hwang, M.C. Peng, V.L. Nguyen, Y.L. Chang, An LSTM-Based Deep Learning Approach for Classifying Malicious Traffic at the Packet Level. Appl. Sci. 2019, 9, 3414. https://doi.org/10.3390/app9163414.

[19] S.Rezvy, M.Petridis, A.Lasebae, T.Zebin. (2019). Intrusion Detection and Classification with Autoencoded Deep Neural Network. In: Lanet, JL., Toma, C. (eds) Innovative Security Solutions for Information Technology and Communications. SECITC 2018. Lecture Notes in Computer Science(), vol 11359. Springer, Cham. https://doi.org/10.1007/978-3-030-12942-2.

[20] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood and A. Anwar, "TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems," in IEEE Access, vol. 8, pp. 165130-165150, 2020, doi: 10.1109/ACCESS.2020.3022862.

[21] https://www.analyticsvidhya.com/blog/2022/06/understanding-loss-function-in-deep-learning/ (consulted in 25-11-2022).

A Novel Approach to Cashew Nut Detection in Packaging and Quality Inspection Lines

Van-Hung Pham, Ngoc-Khoat Nguyen*,Van-Minh Pham
Faculty of Electrical Engineering, Hanoi University of Industry, Hanoi, Vietnam[1, 3]
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam[2]

***Abstract*—YOLO standing for *You Only Look Once* is one of the most famous algorithms in computer vision used for detecting objects in a real-time environment. The newest version of this algorithm, namely YOLO with the seventh version or YOLOv7, is proposed in the present study for cashew nut detection (good, broken and not peeled) in packaging and quality inspection lines. Furthermore, this algorithm using an efficient convolutional neural network (CNN) to be able to successfully detect and identify unsatisfactory cashew nuts, such as chipped or burnt cashews. In order to deal with the quality inspection process, a new dataset called CASHEW dataset has been built at first by collecting cashew images in environments with different brightness and camera angles to ensure the model's effectiveness. The quality inspection of cashew nuts is tested with a huge number of YOLOv7 models and their effectiveness will also be evaluated. The experimental results show that all models are able to obtain high accuracy. Among them, the YOLOv7-tiny model employs the least number of parameters, i.e. 6.2M but has many output parameters with higher accuracy than that of some other YOLO models. As a result, the proposed approach should clearly be one of the most feasible solutions for the cashew's quality inspection.**

Keywords—Cashew; CNN; cashew detection; YOLOv7; computer vision

I. Introduction

It is undeniable cashews (anacardium occidentale) are among the most widely consumed nuts globally, owing to their nutritional value. As of 2016, the world's cashew nut production reached 4.89 million tons, with Vietnam being the leading producer [1]. Cashews are high in protein and polyunsaturated fats, high in carbohydrates and fats, and contain abundant calcium, iron, and phosphorus [2]. Often, they are served as snacks or incorporated into confectioneries or food preparations to increase their value and are generally consumed in raw, roasted, salted or flavoured forms. Consequently, cashews provide various health benefits, such as cancer prevention, cardiovascular protection, nerve protection, antioxidant action, and vitamin content [1].

The cashew nut processing process includes many stages such as roasting, shell cracking and removal, peeling of kernel skin, grading, and packaging [3]. The main issue affecting product cost in the cashew nut processing industry is the yield and quality of cashew nuts after processing. Therefore, many automation technologies have been applied in the cashew nut production process. However, cashew nuts are often graded by vibrators, removing unsatisfactory particles, followed by a manual inspection and hand-picking. In general, skilled workers grade cashews based on their size, shape, and color. Such a manual inspection method is time-consuming, tedious, laborious and less productive.

Therefore, several studies have applied a computer vision approach to assist people with classification, defect detection, quality inspection, and grading of fruits [4-9], vegetables [10-12], grains [13-15], and other food products [16, 17]. For example, Cervantes-Jilaja et al. [13] proposed a computer vision-based method to detect and identify visual defects in chestnuts using external features such as shape, color, size, and texture. In [18], the authors used invertible neural networks to locate and segment damaged soybean seeds in a fast and effective manner. Several review papers on the classification and quality evaluation of fruits and vegetables based on a computer vision approach can be found in [12, 19].

This paper aims to propose a novel algorithm for detecting and classifying cashew nuts. A practical system is also built to verify the studied method in steads of using a manual system. The main contributions of this study focus on:

1) Proposing a new algorithm based on an efficient convolutional neural network (CNN) to detect and recognize unsatisfactory cashew nuts, i.e. chipped or burnt cashew ones.

2) Designing a simulation model representing practical packaging and inspection lines systems consists of conveyor and camera (see Fig. 1). This model is employed to construct a dataset of cashew images.

3) Using the newest and best version of YOLO (v7-tiny) to effectively train the models and evaluate quality of these models.

The rest of this paper is organized as follows. First, Section II presents a number of related works which motivate this study. Then, Section III presents the materials and methods for cashew nut detection and recognition. Next, Section IV shows the evaluation results, while the last section provides the study's conclusions and future work.

II. Related Works

There have been some computer vision-based studies regarding the identification and grading of cashew nuts in the literature. The simplest methods of classifying cashew kernels are based on their color or texture and a single layer neural network [20, 21]. Aran et al. [22] extended this work by testing external features like color, texture, shape and size. In addition, they analyzed the impact of some image preprocessing methods on cashew nut classification. Finally, they evaluated

*Corresponding author: Ngoc-Khoat Nguyen

several classifiers in terms of accuracy and performance. The results show that Back Propagation Neural Network (BPNN) was the most optimal among the classifiers. However, these studies will not be able to classify splits, which occur when kernels split lengthwise naturally. In [23], a novel machine vision method for classifying whole and split cashews (split-up and split-down) is proposed using object shadow combined with surface grayscale-intensity profile.

In addition, methods based on convolutional neural networks (CNNs) are also applied to classify cashew nuts. For instance, Shivaranjani et al. [15] introduced CashNet-15: an optimized cashew nut grading system based on deep CNN and data augmentation. However, this method can only classify wholes and others (scorched whole, splits, butts, pieces). In [24], four deep CNNs combined with image processing were used to classify cashew kernels into five categories based on their adulteration with butts and pieces. Therefore, the above studies demonstrate the effectiveness of machine vision in the classification and grading of cashew nuts. To the best of our knowledge, no study has been conducted on the cashew detection problem, i.e. being able to identify the location and type of cashews (good, broken, and shelled) as they move along the packaging and product inspection lines. Solving this problem can help remove bad nuts (broken or shelled) before packaging and automate product quality inspection.

In recent years, CNN-based object detection has made significant progress and applied to many practical problems. There are two main approaches. The first is a two-stage target detection method such as Faster Region-based CNN (Faster R-CNN) [25], which identifies proposed regions first and then classify them by region. The other method employs a single neural network to detect the type and location of objects simultaneously in an image, such as Single shot multi-box detector (SSD) [26] and YOLO [27-30]. Due to its much faster execution speed than Faster R-CNN, the YOLO model is continuously improved in both speed and accuracy of detection, with many versions such as YOLOv1, YOLOv2, YOLOv3, YOLOv4, YOLOv5 [27-30] and YOLOv7 [31]. Several studies have been conducted using Faster R-CNN [14, 32, 33] or YOLO [34-38] to detect fruits.

However, these approaches have not been applied to the cashew detection problem. This paper proposes an algorithm based on a convolutional neural network (CNN) to detect and identify unsatisfactory cashew nuts, such as chipped or burnt cashews. First, we construct a dataset containing cashew images obtained from cameras mounted on packaging and inspection lines, and then we label the images. Then, we use the YOLOv7 (version 7) to train the model on this dataset. Finally, we conduct a quantitative evaluation of the proposed model's effectiveness on the test dataset and a qualitative assessment from the videos obtained by cameras mounted on the cashew packing line.

III. Materials and Methods

A. YOLOv7

YOLO (You Only Look Once) is a well-known CNN architecture used for general object detection problems because it balances quality requirements and speed. This architecture not only detects the presence of an object but also determines the object's position in an image. YOLO has three official versions YOLOv1, YOLOv2, and YOLOv3. Then there are many improved versions e.g. YOLOv4, YOLOv5, YOLOX [39], YOLOR [40], and the latest is YOLOv7. The newest version, YOLOv7, significantly enhanced speed and accuracy over previous versions. Therefore, in this paper, we choose this version for cashew detection.

YOLOv7 has applied several improvements to increase speed and accuracy. The computational block in the YOLOv7 backbone is called E-ELAN, which stands for Extended Efficient Layer Aggregation Network. Through the E-ELAN architecture of YOLOv7, the network can continuously improve its learning ability by using "expand, shuffle, merge cardinality" without losing its gradient path.

Model scaling allows the generation of models that meet the requirements of different applications by adjusting key attributes. It is possible to optimize a model by scaling it in terms of width (number of channels), depth (number of stages), and resolution (input image size). With traditional concatenation-based architectures (for example, ResNet or PlainNet), scaling factors cannot be analyzed independently and must be analyzed together. Model depth scaling, for example, causes a change in the ratio between the input and output channel of a transition layer, resulting in a decrease in hardware usage. As a result, YOLOv7 introduces a compound model scaling method for concatenation-based models. Compound scaling maintains the model's properties at its design level and thus maintains its optimal structure. Scaling compound models involves two steps: scaling the depth factor of a computational block requires a change in its output channel, and width factor scaling requires a similar change in its transition layer.

RepConv achieves excellent performance in VGG architectures but is significantly less accurate when applied directly to ResNets and DenseNets. YOLOv7 implements re-parameterized convolution using RepConv without identity connection (RepConvN). Re-parameterized convolution avoids an identity connection when replacing residual or concatenation convolution with a re-parameterized one.

A YOLO architecture contains a backbone, a neck, and a head. The head responsible for the final output is called the lead head, and the head used to assist training in the middle layers is named the auxiliary head. In addition, and to enhance the deep network training, a Label Assigner mechanism was introduced that considers network prediction results together with the ground truth and then assigns soft labels. Compared to traditional label assignment that uses ground truth as a basis for generating hard labels based on given rules, reliable soft labels utilize calculation and optimization methods that also consider the quality and distribution of prediction outputs along with the ground truth.

B. Dataset

To collect the dataset, we built a simulation model for the actual packaging and quality inspection lines, as shown in Fig. 1.

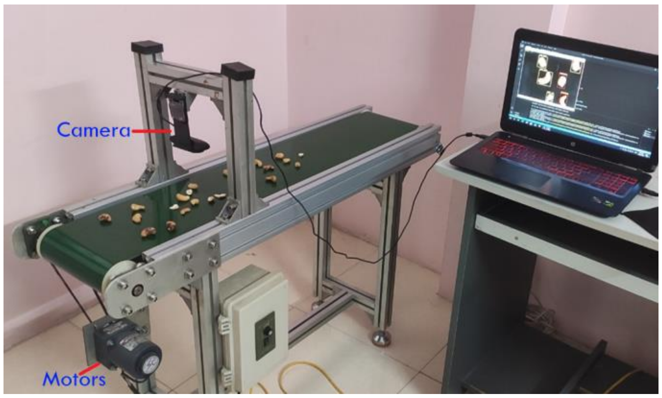


Fig. 1. Cashew nut identification system model

A surveillance camera is securely mounted on a rack and viewed at right angles to the conveyor belt. The camera is connected to a computer via a USB port to capture images of cashews moving on the conveyor. To ensure the robustness of the model in practical applications, we changed the light intensity of the laboratory to collect images of cashew nuts at different brightness levels (strong, normal, and weak). Furthermore, the cashews are placed on the conveyor with varying rotation angles and distances from the camera's position. In this study, we only consider three types of cashew nuts that need to be detected and classified: broken cashews, unshelled cashews, and good cashews (not broken and peeled), as shown in Fig. 2.

Fig. 2. Types of cashews to detect and classify: good cashews (left), broken cashews (middle) and not peeled cashews (right)

After collecting the images, the next step is to annotate them manually using the graphical image annotation tool LabelImg [41]. The annotation classes include 0-good, 1-broken, and 2-Not peeled. As a result, a new dataset named CASHEW dataset was constructed. It contains 312 images and 5115 instances, of which 80% (248 images) were selected randomly as the training dataset, 10% (32 images) as the validation dataset, and the remaining 10% (32 images) as the test dataset. The test dataset was only used to evaluate the model performance after the training, as shown in Table I.

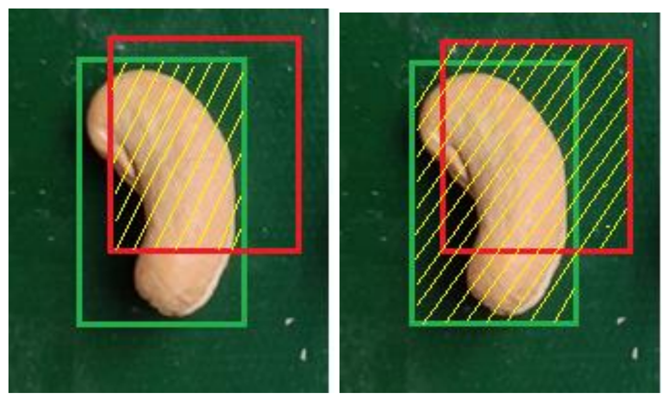
TABLE I. NUMBER OF ANNOTATED IMAGES FOR EACH CASHEW TYPE

CASHEW dataset	*Good*	*Broken*	*Not peeled*	*Number of images*
Train	*1387*	*1569*	*1163*	*248*
Test	*316*	*139*	*89*	*32*
Val	*154*	*164*	*134*	*32*

C. Metrics for Performance Evaluation

To evaluate the effectiveness of a model on the test dataset, we use the following metrics: precision (P), recall (R), average precision (AP), and mean average precision (mAP). Before defining these metrics, we need to define an intersection over union (IoU). The IoU is defined as the percentage overlap between the ground truth (G) and detection boxes (D) and is calculated as:

$$IoU = \frac{Intersection}{Union} = \frac{G \cap D}{G \cup D} \tag{1}$$



Intersection Union

Fig. 3. Examples of intersection and union. ground truth (green) and detection (red)

An example describing the intersection and union with ground truth (green) and detection (red) is shown in Fig. 3. For quantifying accuracy, confusion matrix criteria are calculated, such as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP should have the same ground truth class as detection, and its IoU should be greater than 50%. In contrast, if the detection belongs to the same class as the ground truth and its IoU is less than 50%, it is considered FP. If a ground truth box exists, but the model does not make any detection, this situation is categorized as FN. Lastly, there is no ground truth or detection for the background, so it is classified as TN.

Precision refers to the percentage of correct positive predictions among all detections, while Recall represents the percentage of true positives among all ground truths, as shown in (2) and (3).

$$P = \frac{TP}{TP + FP} = \frac{TP}{\text{All detection}} \quad (2)$$

$$R = \frac{TP}{TP + FN} = \frac{TP}{\text{All ground truths}} \quad (3)$$

The average precision (AP) is the area under the Precision multiplies Recall curve for each class. This can be calculated using (4). After this, the mean average precision (mAP), presenting the overall performance of the model in the detection of possible classes, could be estimated by calculating the mean of all classes' AP.

$$AP = \sum_{i=1}^{n} p(i).\Delta r(i) \quad (4)$$

Where n indicates the total number of detections, i refers to the rank of detection in the list of sorted detections, p(i) represents the precision of the sub-list from the first to the ith detection, and $r(i)$ represents the change in Recall from $(i - 1)th$ to ith detection.

IV. Results and Discussions

A. Dataset Training of YOLOv7

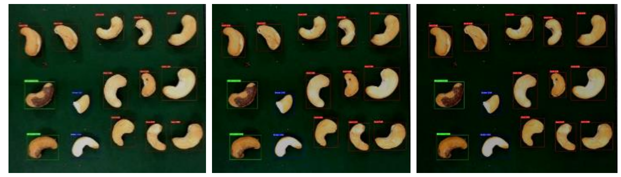
We next conduct the training on the Google Collab Pro platform with the dataset described above. YOLOv7 is trained on the Pytorch platform. First, we select the desired model and edit the configuration file to match the requirements. Next, based on the Train and Val datasets, use pre-trained weights to conduct training process. The number of layers to be classified is 3, including 0-good, 1-broken, and 2-Not peeled. The stochastic gradient descent (SGD) was adopted with an initial learning rate of 0.001. The hyper-parameters used in training the YOLO models are shown in Table II.

TABLE II. Hyper-Parameters Used in Training YOLOv7 Models

Models	Initial Learning Rate	Momentum	Decay	Batch Size
YOLOv7-tiny	0.01	0.937	0.0005	16
YOLOv7	0.01	0.937	0.0005	8
YOLOv7X, W6, E6	0.01	0.937	0.0005	4
YOLOv7D6, E6E	0.01	0.937	0.0005	2

B. Results of Evaluation

Some results of detecting cashews while moving on a conveyor belt with the YOLOv7 model are shown in Fig. 4. It is found that, due to the reflection of light, the accurate detection of cashew nuts is good, chipped or left peeled easily confused. We have tested the performance of the proposed model under different ambient light conditions. The results in Fig. 4 show the same case with three other lighting conditions (strong, normal, and weak). YOLOv7 model could still detect cashews, whether they were unshelled or broken, even though the particles were traveling at less than 0.6 cm/s on the conveyor.



Strong light Normal light Weak light

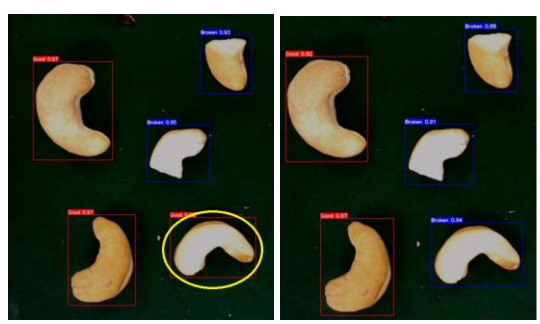
Fig. 4. Results of the YOLOv7 model for detecting cashew nuts as they move on a conveyor at a speed of 0.6 cm/s. The algorithm can accurately detect not_peeled (green), broken (blue), and good seeds (red)

Next, we perform a quantitative evaluation of the model's effectiveness on the test dataset. The evaluation results are given in Table III below. The YOLOv7tiny, YOLOv7, and YOLOv7X models with 6.2M, 36.9M, and 35.7M parameters, respectively, were trained with an input image of size 640 × 640. YOLOv7tiny model has the least number of parameters but has higher F1-score, Precision, Recall, and mAP values than YOLOv7. Even YOLOv7tiny has higher F1-score and Recall parameters than YOLOv7X.

The YOLOv7W6, YOLOv7E6, YOLOv7D6, and YOLOv7E6E models had 70.4M, 97.2M, 154.7M, 151.7M, respectively, and were all trained with 1280×1280 input images. Model YOLOv7D6 has the highest number of parameters, 154.7M; however, many parameters, like Precision and mAP, have the smallest value (blue) among all models.

Comparing all seven models shows that YOLOv7W6 has the largest Recall and mAP values, while YOLOv7E6E model has the largest Precision and F1-score values. On the other hand, we also find that the YOLOv7tiny model with the least number of parameters can be installed on low-profile embedded computers but can still achieve high evaluation indicators. Therefore, the YOLOv7tiny model can be applied to detect and identify actual cashew nuts.

When considering the influence of the ambient light intensity, we found that the light intensity can affect the model's detection results. In Fig. 5, the YOLOv7E6E model incorrectly detected one case (highlighted in yellow), while the YOLOv7tiny model with much fewer parameters correctly detected all the cases.



(a) Detection result of YOLOv7E6E (b) Detection result of YOLOv7tiny

Fig. 5. Comparing YOLOv7tiny and YOLOv7E6E cashew detection results in weak light conditions

TABLE IV. ENT RESULTS ON THE TEST DATASET

Model	#Param	Image size	F1-score	Precision	Recall	AP			mAP
						Good	*Not peeled*	*Broken*	
YOLOv7tiny	6.2M	640	0.88	0.88	0.88	0.952	0.942	0.824	0.906
YOLOv7	36.9M	640	0.86	0.862	0.856	0.934	0.938	0.827	0.900
YOLOv7X	71.3M	640	0.84	0.881	0.801	0.960	0.949	0.814	0.908
YOLOv7W6	70.4M	1280	0.89	0.873	0.903	0.950	0.960	0.822	0.911
YOLOv7E6	97.2M	1280	0.87	0.875	0.866	0.940	0.905	0.864	0.904
YOLOv7D6	154.7M	1280	0.86	0.856	0.869	0.835	0.958	0.790	0.895
YOLOv7E6E	151.7M	1280	0.90	0.903	0.888	0.951	0.949	0.827	0.909

V. CONCLUSIONS AND FUTURE WORK

This study has presented an algorithm to detect and identify good, broken, or missing cashews in the packaging line and check product quality. The CASHEW dataset was developed to test the model's effectiveness by collecting cashew images in different lighting conditions and camera angles. The YOLOv7 convolutional neural network model with many different versions was employed. The evaluation results revealed that the model can recognize cashew nuts with an average accuracy (mAP) of about 90%. Besides, the YOLOv7tiny model has the least number of parameters (6.2M) but has many parameters with higher accuracy than that of some other YOLO ones. In this paper, we have only applied the YOLOv7 model without any improvement. Therefore, it is possible to improve the YOLOv7 model, test and compare it with some other neural network models, and integrate the proposed algorithm on embedded systems with low configurations for the actual cashew packaging and quality inspection lines.

ACKNOWLEDGMENT

This work was supported and financed by Hanoi University of Industry, Project Granted number 15-2021-RD/HĐ – ĐHCN.

REFERENCES

[1] N. N. Oliveira, C. G. Mothé, M. G. Mothé, and L. G. de Oliveira, "Cashew nut and cashew apple: a scientific and technological monitoring worldwide review," *Journal of Food Science and Technology*, vol. 57, no. 1, pp. 12-21, 2020.

[2] M. Aremu, I. Ogunlade, and A. Olonisakin, "Fatty acid and amino acid composition of protein concentrate from cashew nut (Anarcadium occidentale) grown in Nasarawa State, Nigeria," *Pakistan Journal of Nutrition*, vol. 6, no. 5, pp. 419-423, 2007.

[3] A. Berry and S. Sargent, "Cashew apple and nut (Anacardium occidentale L.)," in *postharvest biology and technology of tropical and subtropical fruits*: Elsevier, 2011, pp. 414-423e.

[4] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors*, vol. 20, no. 7, p. 2145, 2020.

[5] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and electronics in agriculture*, vol. 157, pp. 417-426, 2019.

[6] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sensing*, vol. 13, no. 9, p. 1619, 2021.

[7] S. K. Behera, A. K. Rath, A. Mahapatra, and P. K. Sethy, "Identification, classification & grading of fruits using machine learning & computer intelligence: a review," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-11, 2020.

[8] S. Naik and B. Patel, "Machine vision based fruit classification and grading-a review," *International Journal of Computer Applications*, vol. 170, no. 9, pp. 22-34, 2017.

[9] C. C. Ukwuoma, Q. Zhiguang, M. B. Bin Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, "Recent Advancements in Fruit Detection and Classification Using Deep Learning Techniques," *Mathematical Problems in Engineering*, vol. 2022, 2022.

[10] X. Jin, Y. Sun, J. Che, M. Bagavathiannan, J. Yu, and Y. Chen, "A Novel Deep Learning-Based Method for Detection of Weeds in Vegetables," *Pest Management Science*, 2022.

[11] M. K. R. Asif, M. A. Rahman, and M. H. Hena, "CNN based disease detection approach on potato leaves," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020: IEEE, pp. 428-432.

[12] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 3, pp. 243-257, 2021.

[13] C. Cervantes-Jilaja *et al.*, "Optimal Selection and Identification of Defects in Chestnuts Processing, through Computer Vision, Taking Advantage of its Inherent Characteristics," in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2019: IEEE, pp. 513-520.

[14] S. Parvathi and S. T. Selvi, "Detection of maturity stages of coconuts in complex background using Faster R-CNN model," *biosystems engineering*, vol. 202, pp. 119-132, 2021.

[15] A. Sivaranjani, S. Senthilrani, B. Ashokumar, and A. S. Murugan, "CashNet-15: an optimized cashew nut grading using deep CNN and data augmentation," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019: IEEE, pp. 1-5.

[16] N. V. Ganganagowder and P. Kamath, "Intelligent classification models for food products basis on morphological, colour and texture features," *Acta Agronómica*, vol. 66, no. 4, pp. 486-494, 2017.

[17] K. T. Islam, S. Wijewickrema, M. Pervez, and S. O'Leary, "An exploration of deep transfer learning for food image classification," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018: IEEE, pp. 1-5.

[18] Z. Huang *et al.*, "Fast location and segmentation of high-throughput damaged soybean seeds with invertible neural networks," *Journal of the Science of Food and Agriculture*, 2022.

[19] K. Hameed, D. Chai, and A. Rassau, "A comprehensive review of fruit and vegetable classification techniques," *Image and Vision Computing*, vol. 80, pp. 24-44, 2018.

[20] V. Narendra and K. Hareesha, "Cashew kernels classification using colour features," *International Journal of Machine Intelligence*, vol. 3, no. 2, pp. 52-57, 2011.

[21] V. Narendra and K. Hareesha, "Cashew kernels classification using texture features," *International Journal of Machine Intelligence*, vol. 3, no. 2, pp. 45-51, 2011.

[22] M. Aran, A. G. Nath, and A. Shyna, "Automated cashew kernel grading using machine vision," in *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, 2016: IEEE, pp. 1-5.

[23] S. Sunoj, C. Igathinathane, and S. Jenicka, "Cashews whole and splits classification using a novel machine vision approach," *Postharvest Biology and Technology,* vol. 138, pp. 19-30, 2018.

[24] S. K. Vidyarthi, S. K. Singh, R. Tiwari, H. W. Xiao, and R. Rai, "Classification of first quality fancy cashew kernels using four deep convolutional neural network models," *Journal of Food Process Engineering,* vol. 43, no. 12, p. e13552, 2020.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems,* vol. 28, 2015.

[26] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016: Springer, pp. 21-37.

[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934,* 2020.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[29] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* 2018.

[30] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263-7271.

[31] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696,* 2022.

[32] S. Wan and S. Goudos, "Faster R-CNN for multi-class fruit detection using a robotic vision system," *Computer Networks,* vol. 168, p. 107036, 2020.

[33] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017: IEEE, pp. 3626-3633.

[34] R. Gai, N. Chen, and H. Yuan, "A detection algorithm for cherry fruits based on the improved YOLO-v4 model," *Neural Computing and Applications,* pp. 1-12, 2021.

[35] A. Koirala, K. Walsh, Z. Wang, and C. McCarthy, "Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'," *Precision Agriculture,* vol. 20, no. 6, pp. 1107-1135, 2019.

[36] H. Mirhaji, M. Soleymani, A. Asakereh, and S. A. Mehdizadeh, "Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions," *Computers and Electronics in Agriculture,* vol. 191, p. 106533, 2021.

[37] D. Wang and D. He, "Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning," *Biosystems Engineering,* vol. 210, pp. 271-281, 2021.

[38] J. Yao, J. Qi, J. Zhang, H. Shao, J. Yang, and X. Li, "A real-time detection algorithm for Kiwifruit defects based on YOLOv5," *Electronics,* vol. 10, no. 14, p. 1711, 2021.

[39] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430,* 2021.

[40] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "You only learn one representation: Unified network for multiple tasks," *arXiv preprint arXiv:2105.04206,* 2021.

[41] D. Tzutalin, "LabelImg," *GitHub Repository,* vol. 6, 2015.

Improvement Classification Approach in Tomato Leaf Disease using Modified Visual Geometry Group (VGG)-InceptionV3

Jiraporn Thomkaew, Sarun Intakosum
Automation Innovation Intelligence and Data Science Research Unit-Department of Computer Science-
School of Science, King Mongkut's Institute of Technology, Ladkrabang, Bangkok, Thailand

***Abstract*—This paper presents a new method for optimizing tomato leaf disease classification using Modified Visual Geometry Group (VGG)-InceptionV3. Improved performance of VGG-16 model as a base model with InceptionV3 block reduced the number of convolution layers of VGG-16 from 16 to 10 layers, and added an InceptionV3 block that was improved by adding convolution layer from 3 to 4 layers to increase the accuracy of tomato leaf disease classification and reduce the number of parameters and computation time of the model. The experiments were performed on tomato leaves from the PlantVillage dataset of 10 classes, consisting of nine classes of diseased leaves and one class of healthy leaves. The results showed that the proposed method was able to reduce the number of parameters and computation time with and accuracy of tomato leaf disease classification was 99.27%. Additionally, the proposed approach was compared with state-of-the-art Convolutional Neural Network (CNN) models such as VGG16, InceptionV3, DenseNet121, MobileNetV2, and RestNet50. Comparative results showed that the proposed method had the highest accuracy in the tomato leaf disease classification and required a smaller number of parameters and computational time.**

Keywords*—*Modified VGG-InceptionV3; InceptionV3; VGG-16; tomato leaf disease classification

I. Introduction

The rapidly changing climate will have a detrimental effect on agricultural crops. If farmers do not deal with such impacts in time, the result could be widespread damage, causing the production to not meet the standards and market demands. Therefore, there are many researchers from various fields of knowledge trying to develop new research and technology to help solve problems for farmers to be able to recognize and solve plant diseases quickly and accurately. One of the popular technologies for plant disease identification is image processing. Computer image classification is becoming increasingly popular and applied to a wide variety of applications. Developing a computer that can see images like humans and training computers to learn images, be able to analyze and recognize images, and be able to classify images like humans. Computer image processing can be applied to machine learning to allow computers to process large numbers of images accurately and quickly. Detecting and identifying plant diseases requires computer-aided image classification because machine learning can classify images accurately and quickly.

Computer vision is a part of Artificial Intelligence technology that seeks to train computer algorithms to automatically analyze, classify and detect images using neural networks and deep learning. It can detect and recognize images in both still images, video images, as well as real-time images. Computer vision technology is widely used in current research such as medical imaging diagnosis [1], [2], [3], [4] sign language classification [5] facial emotion recognition [6], and includes the detection and identification of various plant diseases such as the identification of corn leaf disease [7], pepper bell [8], rice diseases [9], mango leaf disease [10] etc. In addition, research on real-time plant disease detection and identification [11], [12], [13] was developed by developing a model that can be used on resource-constrained mobile devices. It can be used more easily and conveniently to including the development to use on mobile devices and work without an internet connection [14]. Currently developing or improving the model of the Convolutional Neural Network (CNN) focuses on reducing the number of parameters and computational costs of the model. Because the previous CNN model was a large, multi-layered deep network, and the huge number of parameters used for computation also required a lot of machine resources, it was not suitable for use on mobile devices with resource constraints. A lightweight Convolutional Neural Network is a small CNN model developed to support deployment on mobile or resource-constrained devices. Lightweight CNN reduces the number of model layers and the number of compute parameters of the model to reduce computation time but to still have better performance. Describe lightweight CNN models [15] such as MobileNet, MobileNetV2, SqueezeNet, ShuffleNet, and PeleeNet. These models use a smaller number of parameters and computation time and can perform faster while providing cost-effectiveness. The accuracy is still high compared to the base and state-of-the-art CNN model.

A lightweight CNN model is presented in this paper. The aim is to optimize tomato leaf disease classification with lightweight models that can reduce the number of model parameters and reduce computational time for the model to be deployed on mobile devices. However, the accuracy of disease classification was still high. We present Modified VGG-InceptiopV3. By using the base model of VGG-16 from the original VGG-16 which has a total of 16 layers and approximately 138.4 million parameters. This is considered a large network, so it takes a long time to process and requires a lot of computing costs, in this work the number of layers has been reduced from 16 to 10 layers and the InceptionV3 block has been improved by adding a convolution layer for getting the input image from 3 to 4 convolution layers to make the model wider and faster in parallel, which can reduce the number of parameters and computation time and resizing the convolution filter size from 3 x 3 to 1 x 3 and 3 x 1, then add an modified InceptionV3 block to the VGG model. The proposed model was implemented for the disease classification of tomato leaves from the PlantVillage dataset of 10 classes for nine classes of diseased leaves and one class of healthy leaves. The proposed method used 1,767,652 parameters and compared the accuracy of tomato leaf disease classification and compared the results with state-of-the-art CNN models such as VGG16, InceptionV3, DenseNet121, MobileNetV2, and ResNet50. The proposed method uses the least parameters but had the highest accuracy.

The remainder of this paper is as follows: "Related Work" is presented in the second part, "Proposed Methods" is presented in the third part, presents the, "Results and Discussion" is presented in the fourth part, and the "Conclusion" is presented in the fifth part.

II. Related Work

This section presents articles related to tomato leaf disease classification using the CNN model and various improvements to CNN models. The research without the CNN model for disease classification of tomato leaves required manual extraction of tomato leaf features [16] such as color features, shape features, texture features, etc. But using the CNN model, the image features are extracted automatically.

Zhou and Zhou et al. [17] presents a hybrid deep learning model for tomato leaf disease identification by improving residual dense network to Restructure Deep Residual Dense Network. The develop model would combines deep residual networks and dense networks. Original residual dense blocks extract image features through dense-connected convolution layers, where RDBs are directly connected to all subsequent layers, where input images are aggregated into Res-Dense-Block (RDB) batch normalized and are then added after convolution in the RDB. A LeakyReLU activity function was added. Then, in the residual layer, concatenate tensor combines an improved RDB block with the input layer. Experiments were performed on a tomato leaf dataset from the AI CHALLENGER dataset of 13,185 images. The results showed that the proposed method was 95% accurate in tomato leaf disease classification and had the highest accuracy compared to other CNN models such as Deep CNN, ResNet50, and DenseNet121 with 93.21%, 88.49%, 91.96% accurate, respectively.

Djimeli - Tsajio and Thierry et al. [18] presents an automated method for the detection and identification of tomato leaf disease using a neural network. The model of the proposed method uses transfer learning from ResNet101 and ResNet152 models to extract the features of tomato leaf images. The disease of tomato leaves was then classified by Multi-Layer Perceptron (MLP). Transfer learning from pre-trained ResNet101 and ResNet152 models to achieve shape features, color features, and other features. Features are separated into multiple levels, giving different feature vector sizes. ResNet101 and ResNet152 were then adjusted and combined image features based on the variance and mean deviation of features, mean of two features, and concatenation of two features. Then, use the MLP to classify the features vector that is assigned several 400 neurons in the hidden layer and 10 in the output layer. The results showed that the classification from the features vector with the combination of features from ResNet101 and ResNet152 models had an accuracy of 98.3% with the highest accuracy being 98.9%.

Thangaraj and Anandamurugan et al. [19] presented a model to identify tomato leaf disease using a Modified-Xception model deep neural network. The proposed method uses transfer learning for transfer learning weights and parameters. The process of the model is divided into two parts: feature extraction and classification. The feature extraction process consists of convolution layers and pooling layers using the Rectified linear unit (ReLu) activation function, using max-pooling to reduce the size of the feature map obtained from the convolution layers. The classification process uses Global Average Pooling (GAP) instead of the fully connected layer in the original networks. GAP calculates the mean value of each feature map and sends it to the output layer. In the output layer, use the Softmax function for tomato leaf disease classification. After replacing the output layer with GAP, fine-tuning was used to update the weights obtained from training the model with the tomato leaf dataset. The model training process was compared using various optimizers such as Adaptive Moment Estimation (Adam), Stochastic Gradient Descent (SGD), and Root-Mean-Square Propagation (RMSprop), and experiment with tomato leaf datasets from the PlantVillage dataset. The results showed that using Adam and RMSprop optimizers had better accuracy than SGD optimizer. The Adam optimizer had the highest accuracy of 99.55%, followed by RMSprop of 99.01%, SGD of 81.77%, respectively.

Kaur and Harnal et al. [20] presented a Modified InceptionResNet-V2 (MIR-V2) model to identify tomato leaf disease. Improved from the basic model, the InceptionResNet-V2 (IR-V2) model, which is a collaboration of InceptionV1 and ResNetV2, is used to transfer learning from the InceptionResNet-V2 model to increase the accuracy of tomato leaf disease identification and improve the model with four different InceptionResNet-V2 types. Changing the internal Maxpool layer of InceptionResNet-V2 as follows: 1) 3 - Max pool layer InceptionResNet - V2 (3MPL - IR - V2), 2) 3 - Max pool layer with skip connection (3MPL - SC), 3) 2 - Max pool - layer InceptionResNet - V2 (2MPL - IR - V2) and 4) 2 - Max pool layer with skip connection (2MPL - SC). Then, experiment with the tomato leaf dataset and determine the accuracy of the modified MIR-V2 model and basic IR-V2 to obtain the highest accuracy in disease identification of tomato leaves. The results showed that the 2MPL-SC modification of the Maxpool layer of InceptionResNet-V2 with the highest accuracy was 98.92% compared to the remaining three improvements.

Moussafir and Chaibi et al. [21] presents a model to optimize tomato leaf disease identification using a CNN model in combination with a genetic algorithm. In the first step, tomato leaf disease identification results were compared from the 10 classes PlantVillage dataset. Experiment with CNN models such as VGG16, ResNet50, EfcientNetB0, EfcientNetB1, EfcientNetB2, EfcientNetB3, and EfcientNetB4 using transfer learning and fine-tuning techniques. Two of the models with the highest accuracy in tomato leaf disease classification were obtained. The two most accurate models are then adapted to the genetic algorithm. The genetic algorithm selects the weights generated during the CNN learning process to create a new weight matrix and then considers it the weighted average. The CNN model for disease identification of tomato leaves showed that the two most accurate models, ResNet50 and EfcientNetB0, had 96.6% and 95.6% accuracy. When a genetic algorithm was used to create a new weight matrix from the ResNet50 and EfcientNetB0 models, the accuracy of tomato leaf disease identification was increased to 98.1%.

Thakur and Sheorey et al. [22] presents a model for the identification of plant disease with leaf imagery. Convolutional Neural Network named 'VGG-ICNN'. The presented model is a collaboration of VGG-16 and GoogleNet InceptionV7 block models, aiming to reduce the size of the network while maintaining high data classification performance. The proposed model consists of two convolutions (Conv) layers using filter size 64, followed by a max-pooling layer, followed by two convolutions (Conv) layers using filter size 128, and followed by the max-pooling layer like the base model of VGG-16. Three Inceptionv7 blocks were then added to the end of the VGG-16 max-pooling layer. The InceptionV7 block uses a filter size of 1024. The Global Average Pooling (GAP) layer is used instead of the flattening layer to reduce the number of parameters for model training followed by a fully connected layer using the SoftMax activation function for data classification. The proposed model can reduce the number of parameters to approximately six million parameters, which is comparable to lightweight CNN models. The model was evaluated by experimenting with five datasets, including small datasets and large datasets: PlantVillage dataset, Embrapa dataset, Apple dataset, Maize dataset, and Rice dataset. The results showed that the proposed method when experimented with all five datasets, the PlantVillage dataset had the highest accuracy of 99.16%, followed by the Rice dataset, Apple dataset, Embrapa dataset, and Maize dataset with 96.67%, 94.24%, 93.66%, and 91.36% respectively.

Tuncer, A. [23] presents a new model for the detection and identification of plant leaf diseases. The proposed method is a hybrid deep neural network with Inception architecture and depthwise separable convolutions intended to reduce the number of parameters and computational costs but still has high data classification efficiency. The Inception architecture can extract image features in parallel and can combine features, while depthwise separable convolution separates the features to independently convoluted the spatial convolution of the input image and combine them as output from the convolution layer. The proposed model replaced 1×1 and 3×3, 1×1, and 5×5 convolutions with 3×3 depthwise and 1×1 pointwise convolution, and 5×5 depthwise and 1×1 pointwise convolution in the Inception block. The model consists of two blocks of Modified Inception architecture, four depthwise separable convolution layers, and four pooling layers, followed by a fully connected layer and Softmax classifier. Experiment with the PlantVillage dataset of 30 classes. The results showed that the proposed method was able to achieve a plant disease classification accuracy of 99.27% and a 75% reduction in the number of parameters from the base CNN model. Like Hassan, M.S. et al. [24] convolution layer was replaced with a depth separable convolution to reduce the number of parameters and computational costs of the model.

Lee and Lin et al. [25] presents CNN architecture for potato leaf disease detection. The proposed model consists of one convolution layer with filter size 64, two convolution layers with filter size 128, three convolution layers with filter size 256, after the convolution layer, followed by a Pooling layer and a Dropout layer. There is a filter size equal to the convolution layer and finally, the Fully Connected layer that uses the Softmax function. In the experimental process to prevent overfitting, the number of cycles in the training model was increased to maintain the model's accuracy. The model presented was experimented with the potato leaf dataset from the PlantVillage dataset and compared the performance of the models presented with the VGG-16 and VGG-19 models. The results showed that the proposed method had the highest accuracy of 99.53%, followed by VGG-16 and VGG-19 with an accuracy of 98.15% and 48.55%, respectively, and the proposed method could reduce the number of parameters used by approximately 99.39%.

Jiang and Chen et al. [13] presents a model for real-time detection of apple leaf disease, the INAR-SSD model, developed the VGGNet model in combination with Inception in the process of extracting image features. Then use the Rainbow concatenation method to improve the integration of feature maps. Pooling and deconvolution will work together to create a feature map between layers. The proposed model adds two layers of Inception modules to the VGG-16 model network to improve the feature extraction capabilities of multi-scale images. Thus, the proposed model can detect diseases of various sizes on the same plant leaf. The model architecture consists of the first seven layers retaining the VGG-16 architecture, then the eighth and ninth layers are replaced by two modules of Inception, while the tenth to thirteenth layers retains the VGG-16 architecture. The Fully Connected layers were replaced by three convolutions: 1x1, and the last layer was followed by the Softmax layer. The experiment was performed with a dataset of apple leaves obtained from laboratories and field photographs. The results showed that the proposed method has an accuracy of 97.14% in apple leaf disease identification and provides the highest accuracy compared to other CNN models such as AlexNet, GoogLeNet, InceptionV3, ResNet-101, ResNet-50, ResNet-34, ResNet-18, and VGGNet-16, it shows that the INAR-SSD model has a detection efficiency of 78.80% mAP and a detection speed of 23.13 FPS.

Nagi and Tripathy [26] presents a model for grape leaf disease identification using a lightweight CNN that reduces the number of parameters for calculation but still maintains high accuracy for grape leaf disease identification. The model presented is developed on the basis of the VGG-16 model which has a total of 16 layers, thus presenting the model by reducing the number of layers of VGG-16 to 6 layers and adjusting the filter size of the convolution layer. The model consists of an input layer that receives input image size 128 x 128 pixels, the first convolution layer using filter size 128, followed by a Maxpooling 2x2 layer, and next, one convolution layer using filter size 64, Maxpooling 2x2, then followed by Flatten and Fully connected layer and in the output layer use Softmax function. An Adam optimizer was used to experiment with grape leaves from the PlantVillage dataset of 3,423 images. The results showed that the proposed method had the highest accuracy in disease identification of grape leaves 98.4%, compared to MobileNet, VGG-16, and AlexNet accuracy values are 98.1%, 96.6%, and 95.7% respectively. Furthermore, the proposed method can reduce the number of parameters to approximately two million parameters.

According to the research article, reducing the number of layers of the VGG-16 model can reduce the number of model parameters and computation time. But it still has higher classification efficiency, and using the Inception block can extract parallel image features, thereby reducing run time and obtaining full image features. With the use of CNN models, the disease of tomato leaves can be identified more accurately.

III. Proposed Methods

In this section, we present the workflow of the proposed method. It contains a detailed dataset used for the experiment and describes in detail the VGG-16 model, the InceptionV3 model, as well as the Modified VGG-Inception which is a new model for disease identification of tomato leaves.

A. Dataset

In this experiment, we used the tomato leaf dataset from the Plant Village dataset [27], accessible from www. kaggle.com. This public dataset for image processing and computer vision contains 38 classes of diseased and healthy plant leaves, 14 different plants. In this work, we selected a dataset of nine classes of diseased tomato leaves and one class of healthy tomato leaves for a total of ten classes including Bacterial spot, Early blight, Late blight, Leaf Mold, Mosaic virus, Septoria leaf spot, Spider mites, Target spot, Yellow leaf curl virus, and Healthy as shown in Fig. 1 The image has an RGB color model with dimensions of 256 x 256 pixels, so it resizes the image to 224 x 224 pixels and removes the image's background before training the model. A total of 18,159 tomato leaf images were included, detailed in Table I, and the dataset was divided into three folders: training 70%, validation 20%, and test 10%.

B. VGG-16

VGG-16, developed by Simonyan, K. et al. [28] from Visual Geometry Group, Department of Engineering Science, University of Oxford, won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2014. The VGG-16 consists of an input layer, a convolution layer, and an output layer, totaling 16 layers. The input layer receives image data size 224 x 224 pixels and is an RGB color image, thus having 3 color channels. The convolution layer consists of 13 Conv2D layers using filter size 64 of 2 Conv2D, 128 of 2 Conv2D, 256 of 3 Conv2D, and 512 of 6 Conv2D using kernel size 3 x 3 all layers of the network, after Conv2D followed by a Maxpooling layer of 5 layers, size 2 x 2 strides equal to 2 all layers of the network and followed by 3 Fully Connected (FC) layers, 2 of which Fully Connected use 4096 classes and the last layer uses 1000 classes, while the output layer uses SoftMax as the activation function. The architecture of the VGG-16 model is shown in Fig. 2. The basis of the VGG-16 model is approximately 138.4 million parameters.

C. InceptionV3

InceptionV3 was developed by Szegedy, C. et al. [29] in 2015 and received first-place certification in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2015. InceptionV3 is the 3rd edition of Inception. Its purpose is to increase the efficiency of image classification but to reduce the computational cost, parameters, and calculation time. The first architecture of Inception is shown in Fig. 3 using the principle of extracting features of parallel images, different filters size, resulting in feature maps of different sizes. This contains a convolutional layer 1x1, convolution layer 1x1, followed by a layer convolution layer 5x5, convolution layer 1x1, followed by a layer convolution layer 3x3 and the Maxpooling layer, followed by the convolution layer 1x1. And later InceptionV3 adjusted the architecture as shown in Fig. 4. InceptionV3 adjusts the filter size to be smaller. Base Inception used convolution layer 5 x 5, change to convolution layers 3x3 two layers. The architecture of InceptionV3 consists of layers convolution layer 1x1, convolution layer 1x1, followed by two convolution layers 3x3, convolution layer 1x1, followed by a convolution layer 3x3, and the Max pooling layer, followed by the convolution layer 1x1. Which the replacement Convolutional layer 5x5 to two convolutional layers 3x3, the number of parameters has been reduced from (5x5) = 25 to (3x3) + (3x3) = 18 parameters.

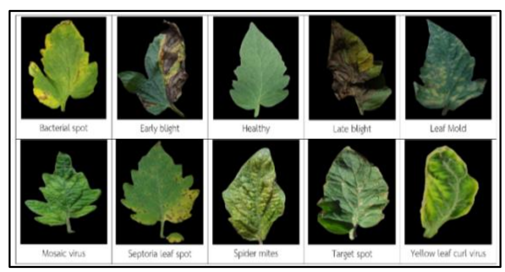


Fig. 1. Sample of tomato leaves from the PlantVillage dataset.

TABLE I. Describes the Number of Classes and Images of the Dataset

Class	Number of Images	Class	Number of Images
Bacterial spot	2,127	Mosaic virus	373
Early blight	1,000	Septoria leaf spot	1,771
Healthy	1,591	Spider mites	1,676
Late blight	1,908	Target spot	1,404
Leaf Mold	952	Yellow leaf curl virus	5,357
Total number of images			**18,159**

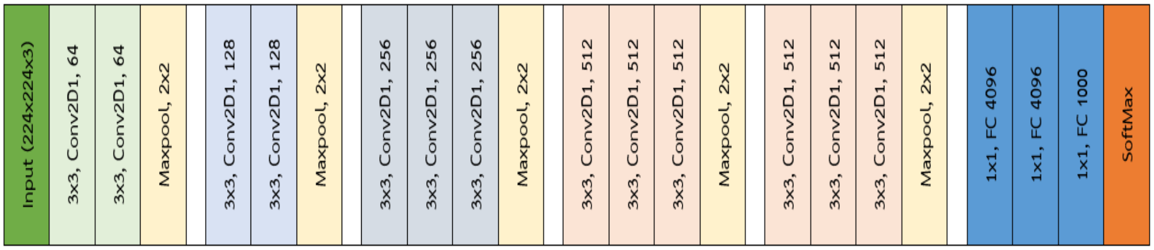


Fig. 2. Base architecture of the VGG-16 model.

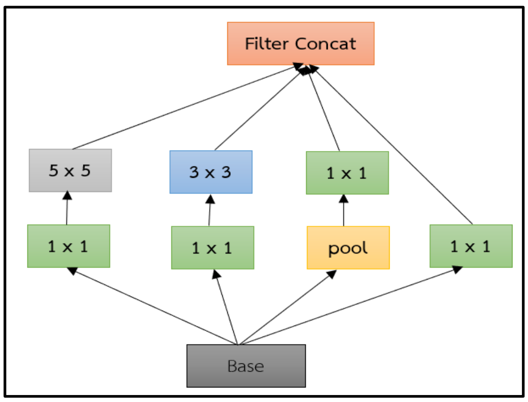


Fig. 3. Base inception architecture.

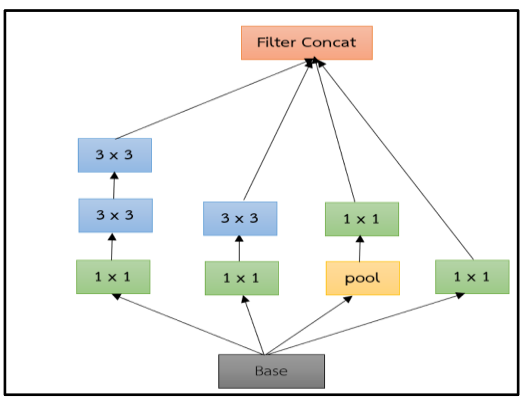


Fig. 4. Base inceptionV3 architecture.

D. The Proposed Modified InceptionV3

The advantage of the InceptionV3 architecture is that the feature extraction of parallel images is distributed simultaneously, providing feature maps of different sizes and reducing the number of parameters and computation time. Therefore, in this work, the architecture of InceptionV3 has been improved by increasing the network width by having the number of layers. The convolution layer has been increased from three layers to four layers, as shown in Fig. 5. The optimized InceptionV3 architecture consists of layers. Convolution layer 1x1, convolution layer 1x1, followed by convolution layer 1x3, and convolution layer 3x1, convolution layer 1x1, followed by convolution layer 1x3 and convolution layer 3x1, convolution layer 1x1, followed by convolution layer 1x3 and convolution layer 3x1, and the Maxpooling layer, followed by the convolution layer 1x1. Extract features in parallel across the four convolution layers and reduces the number of features with the Maxpooling layer which is the replacement of convolution layer 3x3 with convolution layer 1x3 followed by convolution layer 3x1 as convolution separation to be smaller and can reduce computational times.

E. The Proposed Modified VGG-Inception

In this work, we present an improvement on tomato leaf disease classification methods using Modified VGG-Inception. The proposed method was developed using VGG-16 as the base model in combination with the improved InceptionV3 block. The objective was to increase the accuracy of tomato leaf disease classification and reduce computation costs in terms of the number of parameters and processing time. The architecture of the proposed model is shown in Fig. 6. Reduced the number of layers of the VGG-16 model from 16 to 10 layers, and then added a block of modified InceptionV3 between the fourth and fifth layers of the VGG model. The proposed model consists of an input layer that receives image input as a 224 x 224 pixel, RGB color image. Then the first and second layers of the convolution layer use filter size 64, and next, the third and fourth layers use filter size 128, just like the base model of VGG-16. The next layer adds a block of Modified InceptionV3. The fifth to the seventh layer, the convolution layer uses a filter size of 256, followed by Maxpooling 2x2 and BatchNormalization at all layers using a kernel size 3x3 for all of the networks as with the base VGG-16 model. Then the eighth to the tenth layer is followed by a Fully Connected layer with 10 classes in the tomato leaf disease classification using the SoftMax function. Configure the parameters for the experiments of the presented model as in Table II.

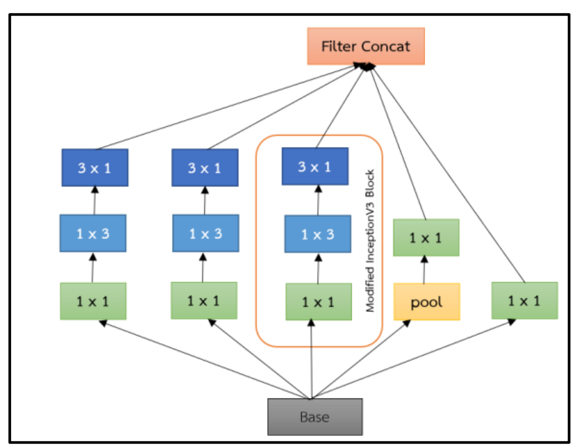


Fig. 5. Modified inceptionV3 architecture.

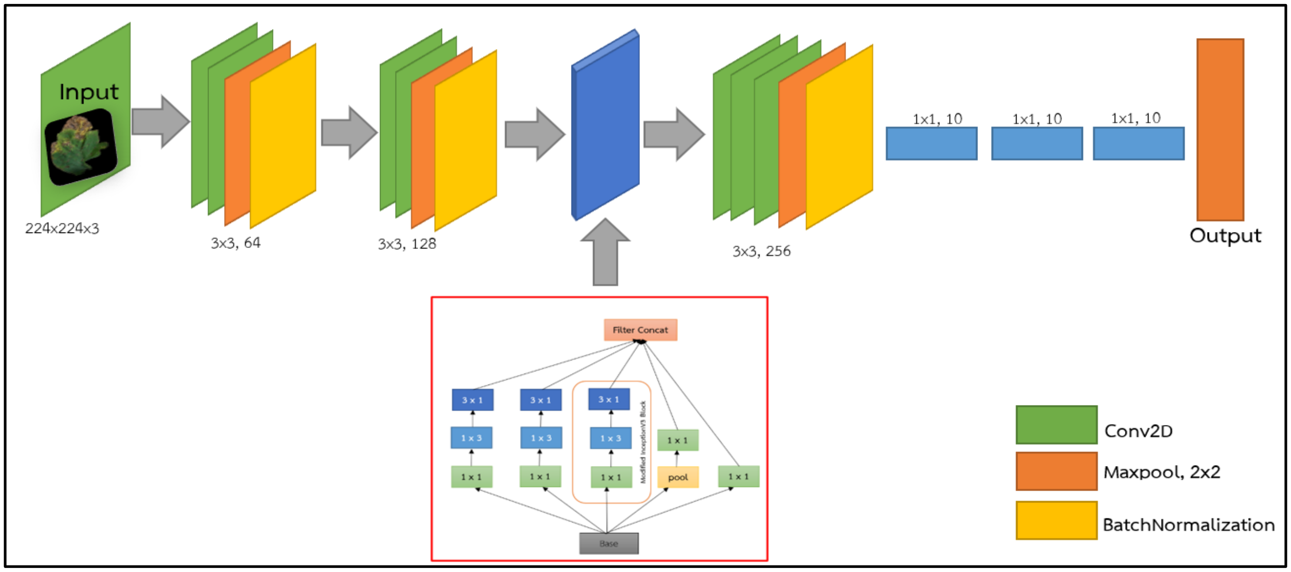


Fig. 6. The proposed modified VGG-InceptionV3 architecture.

TABLE II. CONFIGURING PARAMETERS FOR EXPERIMENTS

Name of the Parameter		Parameter
1.	Optimizer	Adam optimizer
2.	Learning rate	0.001
3.	Batch size	64
4.	Epochs	50
5.	Training data	70%
6.	Validation data	20%
7.	Testing data	10%

IV. RESULTS AND DISCUSSION

This research presents a Modified VGG-InceptionV3 model aiming to improve the limitations of VGG-16, which is a large network and requires high computational costs to a lightweight network by reducing the number of parameters and computational costs, but the model still delivers high disease classification accuracy of tomato leaves and allows this lightweight model to be used on mobile devices or low-resource devices. The presented models were evaluated using the PlantVillage dataset of 18,159 tomato leaf images, nine classes of diseased leaves, and one class of healthy leaves. The image processing procedure resizes the image to 224 x 224 pixels and removes the background of the image. The dataset is divided into a training set of 70%, a validation set of 20%, and a testing set of 10%. Experiment using Adam as an optimizer, set the learning rate to 0.001, batch size to 64, and experiment with 50 epochs.

The proposed model is an improvement based on the VGG-16 model by reducing the number of VGG-16 layers from 16 to 10 layers and improving the InceptionV3 block by increasing the number of convolution layers from 3 layers to 4 layers for a more fully distributed extraction of image features and resized the kernel size of the convolution layer in the InceptionV3 block from 3x3 to 1x3 and 3x1 instead to reduce the number of model parameters. The number of parameters of the Modified InceptionV3 block is shown in Table III. This can reduce the number of parameters by 58.42% from the base InceptionV3 block.

In the process of reducing the number of layers of the base model VGG-16, we experimented with adding a Modified InceptionV3 block to the VGG-16 network and then tried reducing the number of layers of the VGG-16 model layer by layer. The results of the experiment found that by reducing the number of layers of the VGG model, the number of layers with the highest accuracy of tomato leaves classification was obtained using the ten-layer VGG model, as shown in Table IV. In addition, as the number of VGG-16 layers is reduced, the number of parameters and processing time of the model are also reduced.

The results of the experiment showed that the use of the improved InceptionV3 block in combination with the base model VGG-16 reduced the number of model parameters from approximately 138.4 million to approximately 14.7 million. When reducing the number of layers of the VGG-16 model layer by layer until there are ten layers remaining, the highest accuracy for tomato leaf disease classification was 99.27%, and the number of parameters was reduced from approximately 14.7 million to 1,767,652 parameters, with an F1 score of 99.07%. But when reducing the number of layers of the VGG-16 model to nine and eight layers, there was a decrease in the number of model parameters but a decrease in accuracy. The proposed method has accuracy in classifying tomato leaf diseases according to the Confusion Matrix as shown in Fig. 7.

TABLE III. COMPARES THE NUMBER OF PARAMETERS OF THE INCEPTIONV3 BLOCK AND MODIFIED INCEPTIONV3 BLOCK

Base InceptionV3 Architecture			**Modified InceptionV3**		
Layers	***Filter size***	***Parameters***	***Layers***	***Filter size***	***Parameters***
Conv2D1_1	1x1, 32	4,128	Conv2D1_1	1x1, 32	4128
Conv2D1_2	3x3, 32	9,248	Conv2D1_2	1x3, 16	1552
Conv2D1_3	3x3, 32	9,248	Conv2D1_3	3x1, 16	784
Conv2D2_1	1x1, 32	4,128	Conv2D2_1	1x1, 32	4128
Conv2D2_2	3x3, 32	9,248	Conv2D2_2	1x3, 16	1552
MaxPooling	3x3	-	Conv2D2_3	3x1, 16	784
Conv2D2	1x1, 32	4,128	Conv2D3_1	1x1, 16	2064
Conv2D3	1x1, 32	4,128	Conv2D3_2	1x3, 32	1568
			Conv2D3_3	3x1, 32	3104
			MaxPooling	3x3	-
			Conv2D4	1x1, 16	2064
			Conv2D5	1x1, 32	4128
Total Parameters		**44,256**	**Total Parameters**		**25,856**

TABLE IV. COMPARE THE ACCURACY OF VGG-16 WITH THE MODIFIED INCEPTIONV3 BLOCK

Number of Layers (VGG-16 Model)	Accuracy (%)	Loss	Times	Parameter	F1 Score
16	94.70	0.2325	728s	14,715,108	92.44
15	97.70	0.0698	676s	12,355,300	97.07
14	99.12	0.0341	643s	9,995,492	98.85
13	98.86	0.0417	563s	7,648,996	98.41
12	91.44	0.2992	517s	5,289,188	89.32
11	97.10	0.0964	412s	2,929,380	96.10
10	**99.27**	**0.0267**	**386s**	**1,767,652**	**99.07**
9	98.69	0.0471	379s	1,177,572	98.32
8	97.75	0.0236	290s	1,029,988	97.20

class		0	1	2	3	4	5	6	7	8	9	Accuracy (%)
Bacterial spot	0	460	0	0	0	0	0	0	0	0	0	100
Early blight	1	0	231	2	0	0	0	2	0	0	0	98.30
Late blight	2	0	5	407	1	2	0	0	0	1	0	97.84
Leaf Mold	3	0	0	0	223	1	1	0	0	0	0	99.11
Septoria leaf spot	4	1	2	0	0	385	0	0	1	0	0	98.97
Spider mites	5	0	0	0	0	0	364	0	2	4	0	98.38
Target spot	6	0	0	0	0	0	1	313	0	1	0	99.37
Yellow leaf curl virus	7	0	0	0	0	0	0	0	1106	0	0	100
Healthy	8	0	0	0	0	0	0	0	0	353	0	100
Mosaic virus	9	0	0	1	0	0	0	0	1	0	90	97.83

Fig. 7. Confusion matrix.

TABLE V. COMPARING ACCURACY BETWEEN THE PROPOSED MODEL AND STATE-OF-THE-ART CNN

Model	Accuracy (%)	Loss	Times	Parameter
ResNet50	57.31	1.3035	699s	23,608,202
VGG16	86.19	0.4178	875s	14,719,818
Inception-V3	94.22	0.2311	196s	21,823,274
DenseNet121	94.42	0.1644	416s	7,047,754
MobileNetV2	98.59	0.0759	117s	2,270,794
Proposed Model	**99.27**	**0.0267**	**386s**	**1,767,652**

The class that can be classified with the highest accuracy is Bacterial spot, Yellow leaf curl virus, and Healthy with 100% accuracy, followed by Target spot and Leaf Mold with 99.37% and 99.11% accuracy respectively. The least accurate was Mosaic virus and Late blight with 97.83% and 97.84% accuracy respectively. Because of the diseased leaves of the Late blight class, the leaves are almost burnt brown to the extent that they resemble those in the Early blight class, so the model classified the leaves from the Late blight class as the Early blight class and mosaic virus disease, the leaves are pale green to almost yellow, similar to those with the Yellow Leaf Curl Virus class and Late blight class. In addition, to evaluate the effectiveness of the proposed model, we compared tomato leaf disease classifications using the methods presented against state-of-the-art Convolution Neural Network models including VGG-16, Inception-V3, DenseNet121, MobileNetV2, and ResNet50, shown in Table V. Comparative results showed that the proposed models had the highest accuracy in tomato leaf disease classification, followed by MobileNetV2, DenseNet121, Inception-V3, VGG-16, and ResNet50 with 98.59%, 94.42%, 94.22%, 94.22%, 86.19%, and 57.31% accurate, respectively. The proposed model uses fewer parameters than MobileNetV2, a lightweight model suitable for use on mobile devices. Although the proposed method takes longer to process, when comparing accuracy and loss values, the proposed method still gives better results.

V. CONCLUSION

This paper presents a new approach to improving the accuracy of tomato leaf disease classification. Improvement of the Modified VGG-InceptionV3 model based on the VGG-16 model in combination with the InceptionV3 block, reducing the number of layers of the VGG-16 model from 16 to 10 layers and improving the InceptionV3 block by adding the convolution layer from 3 layers to 4 layers. Tomato leaf disease was classified with the model presented by using the tomato leaf disease dataset from the PlantVillage dataset, with a disease identification accuracy of 99.27%. The proposed method can reduce the number of model parameters to 1,767,652 parameters, which is considered a lightweight but still highly accurate model for classifying tomato leaf disease. Compared with state-of-the-art CNN models such as VGG16, Inception-V3, DenseNet121, MobileNetV2, and ResNet50, the proposed method has the highest accuracy in tomato leaf disease classification. Therefore, in the future, the proposed model can be developed for use on small mobile devices and possibly classify more diverse plant leaf diseases to assess the effectiveness of the model.

REFERENCES

[1] S., Deepak, and P. M., Ameer, "Automated categorization of brain tumor from MRI using CNN features and SVM", *Journal of Ambient Intelligence and Humanized Computing 12*, pp. 8357–8369, 2021.

[2] R., Sivanandan and J., Jayakumari, "A new CNN architecture for efficient classification of ultrasound breast tumor images with activation map clustering based prediction validation", *Medical & Biological Engineering & Computing 59*, pp. 957–968, 2021.

[3] H., Che, L.G., Brown, D.J., Foran, J.L., Nosher and l., Hacihaliloglu, "Liver disease classification from ultrasound using multi-scale CNN", *International Journal of Computer Assisted Radiology and Surgery 16*, pp. 1537–1548, 2021.

[4] A.M., Ibraheem, K.H., Rahouma, and H.F.A., Hamed, "3PCNNB-Net: Three parallel CNN branches for breast cancer classification through Histopathological images", *Journal of Medical and Biological Engineering 41*, pp. 494–503, 2021.

[5] A.A., Barbhuiya, R.K.,Karsh and R., Jain, "CNN based feature extraction and classification for sign language", *Multimedia Tools and Applications 80*, pp. 3051–3069, 2021.

[6] R.Z., Cabada, H.R., Rangel, M.L.B., Estrada, ans H.M.C., Lopez, "Hyperparameter optimization in CNN for learning-centered emotion recognition for intelligent tutoring systems", *Soft Computing 24*, pp. 7593–7602, 2020.

[7] J., He, T., Liu, L., Li, Y., Hu and G., Zhou, "MFaster R-CNN for maize leaf diseases detection based on machine vision", *Arabian Journal for Science and Engineering*, 2022. https://doi.org/10.1007/s13369-022-06851-0.

[8] H., Mustafa, M., Umer, U., Hafeez, A., Hameed, A., Sohaib, S., Ullah and H.A., Madni, "Pepper bell leaf disease detection and classification using optimized convolutional neural network", *Multimedia Tools and Applications*, 2022. https://doi.org/10.1007/s11042-022-13737-8.

[9] R., Yakkundimath, G., Saunshi, B., Anami and S., Palaiah, "Classification of rice diseases using Convolutional Neural Network models", *J. Inst. Eng. India Ser. B 103*, pp. 1047–1059, 2022.

[10] M. Prabu and B.J., Chelliah, "Mango leaf disease identification and classification using a CNN architecture optimized by crossover-based levy flight distribution algorithm", *Neural Computing and Applications 34*, pp. 7311–7324, 2022.

[11] J., Liu and X., Wang, "Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model", *Plant Methods 16:83*, 2020. https://doi.org/10.1186/s13007-020-00624-2.

[12] X., Wang, J., Liu and X., Zhu, "Early real-time detection algorithm of tomato diseases and pests in the natural environment", *Plant Methods 17:43*, 2021. https://doi.org/10.1186/s13007-021-00745-2.

[13] 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. 59069-59080, 2019.

[14] A., Shelke and N., Mehendale, "A CNN-based android application for plant leaf classification at remote locations", *Neural Comput & Applic*, 2022. https://doi.org/10.1007/s00521-022-07740-1.

[15] B., Abdelmalek, K., Ahmed, and T.M., Amine, "A survey on lightweight CNN-based object detection algorithms for platforms with limited computational resources", *International Journal of Informatics and Applied Mathematics*, Vol. 2, No. 2, pp. 28-44.

[16] J., Basavaiah and A.A., Anthony, "Tomato leaf disease classification using multiple feature extraction techniques", *Wireless Personal Communications 115*, pp. 633–651, 2020.

[17] C., Zhou, S., Zhou, J., Xing and J., Song, "Tomato leaf disease identification by Restructured Deep Residual Dense Network", *IEEE Access*, vol. 9, pp. 28822-28831, 2019.

[18] A.B., Djimeli-Tsajio, N., Thierry, L.T., Jean-Pierre, T.F., Kapche and P., Nagabhushan, "Improved detection and identification approach in tomato leaf disease using transformation and combination of transfer learning features", *Journal of Plant Diseases and Protection 129*, pp. 665–674, 2022.

[19] R., Thangaraj, S., Anandamurugan and V.K., Kaliappan, "Automated tomato leaf disease classification using transfer learning-based deep convolution neural network", *Journal of Plant Diseases and Protection 128*, pp. 73–86, 2021.

[20] P., Kaur, S., Harnal, V., Gautam, M.P., Singh and S.P., Singh, "A novel transfer deep learning method for detection and classification of plant leaf disease", *Journal of Ambient Intelligence and Humanized Computing*, 2022. https://doi.org/10.1007/s12652-022-04331-9.

[21] M., Moussafr, H., Chaibi, R., Saadane · A., Chehri, A.E., Rharras and G., Jeon, "Design of efficient techniques for tomato leaf disease detection using genetic algorithm-based and deep neural networks", *Plant Soil*, 2022. https://doi.org/10.1007/s11104-022-05513-2.

[22] P.S., Thakur, T., Sheorey and A., Ojha, "VGG-ICNN: A lightweight CNN model for crop disease identification", *Multimedia Tools and Applications*, 2022. https://doi.org/10.1007/s11042-022-13144-z.

[23] A., Tuncer, “Cost-optimized hybrid convolutional neural networks for detection of plant leaf diseases”, *Journal of Ambient Intelligence and Humanized Computing 12*, pp. 8625–8636, 2021.

[24] S.M., Hassan, A.K., Maji, M., Jasinski, Z., Leonowicz and E., Jasinska, “Identification of plant-leaf diseases using CNN and transfer-learning approach”, *Electronics 10(12)*, 1388, 2021. https://doi.org/10.3390/electronics10121388.

[25] T.Y., Lee, I.A., Lin, J.Y., Yu, J., Yang and Y.C., Chang, “High efficiency disease detection for potato leaf with Convolutional Neural Network”, *SN Computer Science 2:297*, 2021. https://doi.org/10.1007/s42979-021-00691-9.

[26] R., Nagi and S.S., Tripathy, “Deep Convolutional Neural Network based disease identification in grapevine leaf images”, *Multimedia Tools and Applications 81*, pp. 24995–25006, 2022.

[27] D.P., Hughes and M., Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics", arXiv:1511.08060, 2015.

[28] K., Simonyan and A., Zisserman, “Very Deep Convolutional Neural Networks for large-scale image recognition”, *Published as a conference paper at ICLR 2015*, pp. 1 - 14, 2015.

[29] C., Szegedy, V., Vanhoucke, S., Ioffe, J., Shlens, and Z., Wojna, “Rethinking the Inception architecture for Computer Vision”, arXiv:1512.00567v3.

Deep Neural Network Training and Testing Datasets for License Plate Recognition

Ishtiaq Rasool Khan[1], Saleh M. Alshomrani[2], Muhammad Murtaza Khan[3], Susanto Rahardja[4]
University of Jeddah, College of Computer Science and Engineering, Jeddah, Saudi Arabia[1, 2, 3]
Northwestern Polytechnical University, School of Marine Science and Technology, Xi'an, China[4]

***Abstract*—Modern society has made tremendous progress towards automation to increase the quality of life and reduce the margin of human error. Intelligent transportation systems are a critical aspect of this evolution. The core technology of these systems is the automatic identification of vehicles' license plates to monitor safety and control violations of traffic rules and other crimes. The research on license plate detection and recognition has gone a long way, from traditional computer vision techniques to features (color, shape, text, etc.) based classification and finally to modern deep learning structures. The deep networks comprising hundreds of layers require enormous amounts of training data. The training dataset should contain plates from different countries; otherwise, the system will be specific to only certain types of plates (from a country or province). There are several datasets collected by researchers containing large numbers of license plates from different countries. This paper provides a detailed survey of such datasets available in the public domain. Sample images from each dataset are shown, and details such as the dataset size, size of images, download link, and country of origin are provided. This survey will be a helpful reference for new researchers in the field for the tasks of training new networks and benchmarking their performances.**

Keywords—License plate recognition; deep neural networks; public datasets

I. Introduction

License plate (LP) recognition has been a well-researched problem in the literature. The earlier works used certain features of the plates, such as height-to-width ratio, color of plate and text, font and color of text, and number and relative locations of alphanumerics, etc. [1-2]. These handcrafted features were generally specific to certain types of plates, and detection accuracy was low. The conventional object detection and recognition methods have recently been replaced by neural networks based deep learning techniques in the computer vision domain. This has resulted in improved accuracy and the ability to recognize license plates in more challenging environments [3-4].

One main issue associated with deep learning techniques is the requirement for large amounts of data for training and testing. Since these techniques are data-dependent, the data's variability and quantity make a difference in the results. Therefore, the dataset used for training a deep neural network (DNN) is important, and a DNN trained on a dataset cannot be fairly compared with another DNN trained on a different dataset. Moreover, the results reported in published literature cannot be reproduced or verified if the dataset is unavailable. Researchers working in this domain realize this issue, and many have started sharing their datasets and algorithms in the public domain. This healthy trend increases confidence in the results presented in the literature, creates benchmarks for comparing new research, and serves as a starting point for new researchers entering the field.

Several surveys have been presented in the existing literature on the techniques used for license plate detection and recognition [5-6]. However, to our knowledge, no comprehensive survey has been carried out of datasets that can be used for training and testing the DNNs for LP detection and recognition. In this paper, we review these datasets and describe their characteristics, such as the sizes of datasets and dimensions of images. We also show some sample images from each dataset.

Most of the existing datasets contain English alphanumeric plates; hence, the networks trained on them cannot perform well on plates with text in other languages. Therefore, this paper also presents a new dataset comprising 2108 license plates from Saudi Arabia. These plates contain bilingual text in Arabic and English, which poses additional challenges to the research community. First, the DNN will need to be trained to ignore Arabic characters, some of which can be easily confused with certain English characters. For example, the English "l" and the first Arabic alphabet "alif" are written similarly; there are many other similarities. Secondly, to write two strings of text on the plate, the size of the characters is generally reduced to smaller than the characters on the license plates with only English characters, which makes them hard to be detected. For the presented bilingual dataset, we provide the ground truth, i.e., the text on the LP and the coordinates of its four corners, along with the images in the dataset, to make it more useful for the research community.

The next section briefly describes several datasets comprising images of vehicles and LPs. The criterion for including a dataset in this survey is that it was published in a journal or a conference paper or used in some scientific work to test or train a DNN. Section III briefly discusses some key observations that can be made from this survey. Finally, the paper is concluded in Section IV.

II. Datasets in Public Domain

A. Application Oriented License Plate (AOLP) dataset

One of the earliest available and widely used datasets is the Application Oriented License Plate (AOLP) dataset containing Taiwanese license plates. Collected by Hsu et al., it comprises three categories named Access Control (AC), Law

Enforcement (LE), and Road Patrol (RP), containing 681, 757, and 611 images, respectively [7]. Sample images from each subgroup are shown in Fig. 1.

At access control locations, such as toll points or entrance and exit of an area under surveillance, the vehicles must stop at a fixed passage or move slowly. The AC subset of the AOLP dataset captures these scenarios. The camera is approximately 5 meters or less away from the car in these images. The camera pan is in the [-30, 30] degrees range, and the tilt is between 0 and 60 degrees, where 0 degrees is assumed parallel to the ground. The ratio of the plate width to image width is between 0.2 and 0.25, and the plates are generally straight with an orientation of fewer than 10 degrees. Images are captured at different illuminations, including outdoor, indoor, day and night times, and various weather conditions. The size of the images is 352 x 240 pixels.

LE images have panning from -40 to 40 degrees, tilting from 20 to 70 degrees, and capturing distances less than 15m. The width of the license plate is 0.1 to 0.2 times that of the image and contains images captured by a roadside camera when the vehicle violated a traffic rule. LE images have a higher resolution of 640 x 480 pixels than AC and RP images in the dataset.

Finally, the RP images have panning from -60 to 60 degrees, tilting from 0 to 50 degrees, and capturing distances less than 15m. The width of the license plate is 0.1 to 0.4 times the image width, and the images were captured either with a handheld or a mounted camera. Road patrol purposes include searching for lost vehicles, scanning for parking violations, and security checking in a restricted area. The image resolution in this dataset is 320 x 240.

B. Caltech Cars

Caltech Cars1999 dataset [8] consists of 126 car images having a resolution of 896 x 592 pixels. The images of parked cars during the daytime are included in this dataset. Another dataset Caltech Cars2001 [9], comprising 526 images at 360 x 240 pixels resolution, has also been made available by the researchers at the same link. The images in the Cars2001 dataset were captured on a highway; therefore, there is more variation in plate sizes and capturing distances and angles in this dataset. Sample images of both datasets are displayed in Fig. 2.

C. Peking University Dataset (PKU)

The Peking University dataset (PKU) is created by the National Engineering Laboratory for Video Technology (NELVT), a research group at Peking University China. The data set was captured by surveillance cameras, in China, during both day and nighttime [10]. Images in the data set are divided into five categories: G1 (810 images, 1082x727 pixels each), G2 (700 images, 1082x727 pixels each), G3 (743 images, 1082x727 pixels each), G4 (572 images, 1600 x 1236 pixels each), and G5 (1152 images, 1600 x 1200 pixels each). All images contain front number plates. The images have good quality, and the plates are easily readable. Some images in the G5 category have multiple cars. In Fig. 3, we have shown some images from G1 and G4 categories.

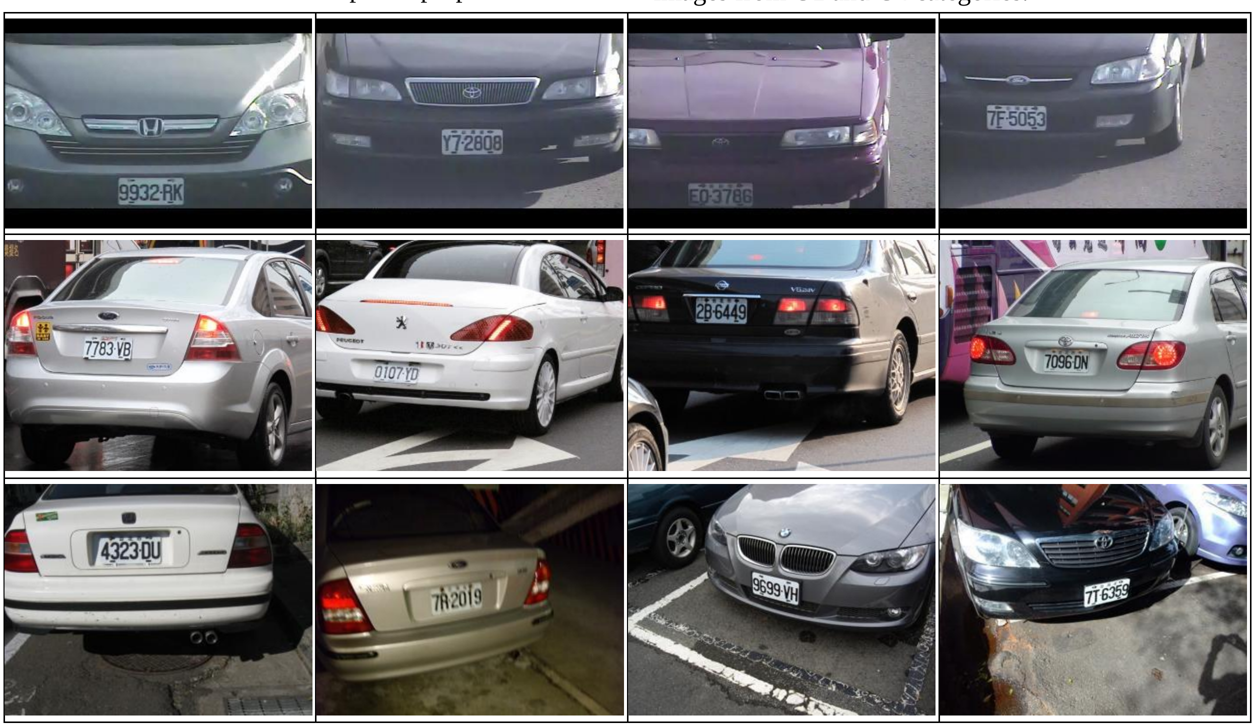


Fig. 1. Sample images from AOLP AC (top row), LE (middle row), and RP (bottom row) datasets Download link: https://github.com/AvLab-CV/AOLP.

Fig. 2. Sample images from Caltech Cars1999 (top row) and Cars2001 (bottom row) datasets Download Link: https://www.vision.caltech.edu/datasets.

Fig. 3. Sample images from PKU dataset category G1 (top row) and category G5 (bottom row) Download link: https://github.com/ofeeler/LPR/tree/master/pku_vehicle_dataset/images.

D. Synthetic Dataset

Bjorklund et al. [11] released a dataset of synthetic images for training CNNs. The synthetic plates were generated considering a wide range of conditions affecting the aspect of real plates. The synthetically generated text was projected on unconstrained backgrounds instead of cars. The authors argue that using license plates on unnatural backgrounds enhances the learning process's ability to account for large text and background variability. The major advantage of generating synthetic plates is that the laborious task of capturing and annotating thousands of plates is avoided. Plated can be generated in any pose, and the template can be changed to different fonts, shapes, and countries. The authors released a training dataset of 20,000 images containing plates and 20,000 images without plates, and a validation dataset of 2500 images with plates and 2500 images without plates. The images represent Taiwanese plates and have a resolution of 768 x 384 pixels. In Fig. 4, we have shown a few sample images containing plates.

E. University of Zagreb Dataset

Researchers at the University of Zagreb produced a database containing 500 images of Croatian license plates. The images contain rear views of cars, trucks, and buses under varying lighting conditions from various angles using a handheld digital camera. The typical sizes of the images are 1024 x 768 and 640 x 480. Some samples of these images are shown in Fig. 5.

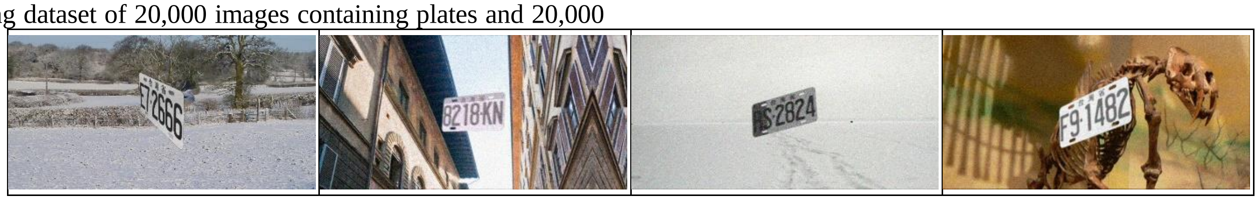


Fig. 4. Sample images from the synthetic dataset by Bjorklund et al. Download link: https://www.kaggle.com/datasets/tbjorklund/annotated-synthetic-license-plates.

Fig. 5. Sample images from the University of Zagreb dataset of car. Download link: http://www.zemris.fer.hr/projects/LicensePlates/english/results.shtml.

F. Chinese Car Parking Data Set (CCPD)

The Chinese Car Parking Data set (CCPD) [12] is a comprehensive data set, 12 GB in size, containing over 300,000 images of Chinese license plates captured with a handheld camera by license plate inspectors across China. The images were acquired between 0730 and 2200 hrs. Each image has only one plate, and the typical resolution of the images is 720 x 1160 pixels. Detailed annotations of the images are provided, with the following information: area ratio of LP to entire image, horizontal tilt, vertical tilt, bounding box's top-left and bottom-right coordinates, exact coordinates of four vertices of LP, 7-letters LP number, brightness of LP region, and blurriness of LP region. The database provides splitting for training and testing subsets. Classification based on other features, such as blur, rotation, tilt, etc., is also available. A subset with challenging images is also part of the classification. The authors updated the dataset in 2020 with new number plates containing eight digits, which are assigned to energy-efficient "green" vehicles. Sample images from CCPD are shown in Fig. 6.

G. National Technical University of Athens Dataset (NTUA)

The National Technical University of Athens released a dataset of the images of Greek license plates and categorized the content as still images and video. At this point, only one video is available, but there are a large number of images in the dataset, and they are divided into several categories based on certain characteristics (color, grayscale, and blurry images) and capture time (day or night). The images with shadows, dirt, and more than one vehicle and those taken from a close distance are stored separately. The images have different resolutions, and we found 1792 x 1312, 800 x 600, and 640 x 480 pixels images in the dataset. There are 716 images in total at this point, but the authors plan to keep adding more. Sample images of this dataset are shown in Fig. 7.

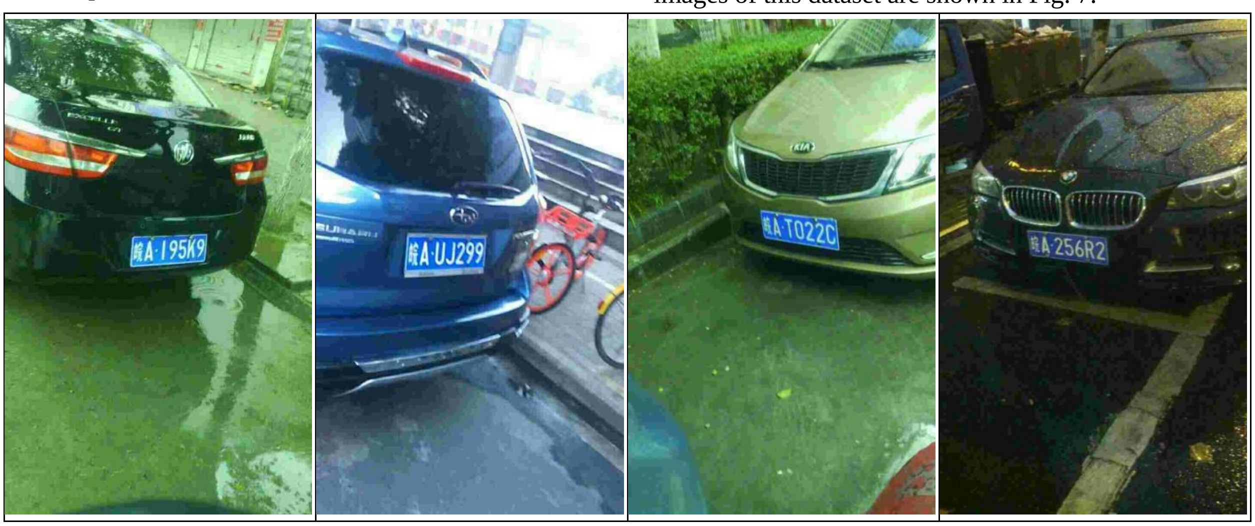


Fig. 6. Sample images from CCPD dataset Download link: https://sites.google.com/site/avlabaolp/download.

Fig. 7. Sample images from National Technical University of Athens' database of Greek plates. Download link: http://www.medialab.ntua.gr/research/LPRdatabase.html.

H. Federal University of Parana Dataset (UFPR-ALPR)

A Brazilian car dataset was released by the Federal University of Parana researchers. The dataset was referred to as UFPR-ALPR by the researchers, and it contains 4500 fully annotated images from 150 vehicles captured in real-world situations [13]. The camera and the vehicle were both moving, so several challenging scenarios were captured. Multiple images of a car taken from different angles are included in the dataset. Images are stored losslessly in PNG format and have a resolution of 1920 x 1080 pixels. Three cameras – GoPro Hero4, Huawei P9 Lite, and iPhone 7 Plus – were used to capture these images, 1500 with each camera, including 900 images of cars with gray LP, 300 images of cars with red LP, and 300 images of motorcycles with gray LP. The images are split into training, validation, and testing sets in a 40:20:40 ratio. Images are annotated with the following information: camera name, type of vehicle (car or motorcycle), manufacturer, model, and manufacturing year of car, position and text of the LP, and position of characters. Some representative images of the UFPR-ALPR dataset are shown in Fig. 8.

I. Kurpiel Dataset

A Brazilian video dataset released by Kurpiel et al. [14] comprises five videos. Under different weather conditions, the authors extracted 4070 plates in 1829 images taken at a resolution of 1920 x 1080 pixels. The plates are quite noisy, and the text is unreadable in many cases. Therefore, this dataset can be used for training and testing LP detection algorithms; however, recognizing text can sometimes be challenging or even impossible. Some samples of the LPs in this dataset are provided in Fig. 9. These are taken from the original paper. The download link is not working now, but we have included this dataset in our review, as interested readers might approach the authors directly and request access.

J. Goncalves Dataset

Another Brazilian database comprising 6,660 images with 8,683 license plate images from 815 different vehicles was accumulated by Goncalves et al. [15]. Images are divided into training (3595 images), testing (2360 images), and validation (705 images) subsets. These images have 1920 x 1080 pixels per image, and the license plate size varies from 5 x 12 pixels to 86 x 196 pixels, with an average of 22 x 57 pixels. These full HD images are stored in PNG format losslessly, one image taking 2.4 MB on average, which makes the total data size quite large. This database can be accessed by submitting a form to the authors and requesting access for non-commercial purposes. Extracted characters from license plates were shared in [16]. There are 101 on-track vehicles captured during the day with a digital camera in Full-HD resolution (1920×1080 pixels). Multiple frames are captured for each vehicle, 19.8 on average, each stored in PNG format (4.08 MB on average), making the total size of the database more than 8.5 GB. This dataset can be used in the license plate character segmentation problems. Some sample images of characters are shown in Fig. 10. The download links of both datasets are provided in the figure caption.

Fig. 8. Sample images from UFPR-ALPR dataset of Brazilian plates. Download link: https://web.inf.ufpr.br/vri/databases/ufpr-alpr/.

Fig. 9. Sample images from the dataset of Kuprial (copied from the original paper) Download link: https://pessoal.dainf.ct.utfpr.edu.br/rminetto/projects/license-plate (not accessible).

Fig. 10. Sample images of characters extracted in the Goncalves dataset. Download link: http://smartsenselab.dcc.ufmg.br/en/dataset/banco-de-dados-sense-alpr/ (vehicles) http://smartsenselab.dcc.ufmg.br/en/dataset/sense-segplate/ (characters).

K. University of California, San Diago Dataset (UCSD-Calit2)

The database shared by the University of California, San Diago (UCSD-Calit2) comprises around 10 hours of video of cars entering and leaving the campus during various hours of the day. The still frames extracted from these videos were hand labeled; thus, the data for 878 cars was annotated. The dataset also comprises 291 car images taken in the car park using a handheld digital camera. The dataset was available at the following link, which is inaccessible now: http://vision.ucsd.edu/belongie-grp/research/carRec. We have provided this link, assuming that the website might be restored in the future or interested readers may want to approach the authors.

L. Spanhel Dataset

Spanhel et al. [17] shared a database comprising relatively low-quality images for the recognition of license plates. The authors captured 9.5 hours long from 8 different bridges on highways under different conditions. They used a boosted soft cascade classifier to detect LPs and tracked them across frames with a Kalman filter. Human observers recognized the plates across a sequence of frames and annotated the whole sequence with the text of the plate. This was an efficient approach to label multiple frames at once instead of doing them one by one. There are 76,412 color images of cropped license plates in this dataset, named ReId dataset. The authors also provided a high dynamic range (HDR) dataset, which is relatively smaller in size and contains 652 images taken at different exposure times. Images were cropped manually to obtain LP images. A few samples of plates taken from the ReId and HDR datasets are shown in Fig. 11. Since the images of cars were captured from different distances and angles, the cropped images of plates have different dimensions.

M. AI Tunisia Hack 2019 Challenge Dataset

A license plate data set was made available in AI Tunisia Hack 2019 challenge. The data comprised 900 images of vehicles taken from the Internet and annotated manually for the coordinates of the bounding box containing the LP. A separate set of 900 images of extracted LPs was also provided in which the annotations mark the text of the LP. Thus, the first dataset can be used to train a network for the detection of LP, whereas the second dataset is specific to the recognition task. Sample images from both datasets are provided in Fig. 12. The images are truncated to show the cars and plates in both datasets; therefore, they have different sizes.

N. Platesmania

License plates from different countries can be downloaded from the Platesmania website, which contains license plates from more than 70 Asian, European, American, and African countries. The data set includes thousands of images. For example, there are 314830, 305643, and 190448 images of vehicles from Germany, France, and the US, respectively. Images are not annotated; hence, the researchers need to mark the bounding boxes and the plate text themselves before they can use them for training a network. A clear picture of the LP is provided with each vehicle image, but it is in image format and not the text, thus not of much use in the training task. However, the picture of the LP is quite clear, and OCR algorithms should be able to extract the text accurately. Some images of cars and their number plates from four different countries are shown in Fig. 13.

Fig. 11. Sample images from the ReId (top row) and HDR datasets (bottom row) of Czech Republic plates by Spanhel et al. Download link: https://medusa.fit.vutbr.cz/traffic/ (not accessible now).

Fig. 12. Sample images from AI Tunisia Hack 2019 challenge. Download link: https://zindi.africa/competitions/ai-hack-tunisia-2-computer-vision-challenge-2

Fig. 13. Sample images of vehicles from Japan, Kyrgyzstan, Mexico, and Russia, from Platesmania. Download link: https://platesmania.com/

O. Stanford Dataset

The cars dataset developed by Stanford University researchers contains approximately 16,185 images of 196 cars, split into 8144 training and 8041 test images [18]. The number plates are not necessarily visible, and the dataset is meant to be used for identification of the car type and not the license plate. However, we have mentioned it here for interested readers, as these details can be important in an intelligent transport system. Images have different resolutions, typically 584 x 328, 634 x 357, or similar. Classes are made at the car make, model, and year level, e.g., the 2012 Tesla Model S or 2012 BMW M3 Coupe. Annotations are provided, including class labels and bounding boxes for all images. Initially, an evaluation server was set up for automatic evaluation and comparison with the provided results. However, it has been decommissioned now, but the manual evaluation can still be done by using the test dataset's ground truth. We have provided some sample images of this dataset in Fig. 14.

P. University of Jeddah License Plate Dataset (UJLP)

The license plates in Saudi Arabia have bilingual text, i.e., the digits and characters are written on top in Arabic and at the bottom in English, and this comes with its unique challenges. First, the recognition systems trained on English characters must discard the Arabic text. Secondly, to fit both texts, the size of the characters is kept small. However, bilingual text offers a unique advantage: separate NNs can be trained on Arabic and English characters, and their results can be combined for higher accuracy. To the best of our knowledge, there has been no notable dataset of Saudi Arabian plates in the public domain. We have developed one consisting of 2108 license plates [19]. To make this dataset, we captured a large number of videos of parked and moving cars from a moving car using our mobile phones. We developed a customized software tool to extract the frames of interest from the captured videos. We also developed a tool for annotation which displays a frame and allows the user to mark the corners of the plate with mouse clicks. The coordinates of these corners are stored in a text file. The extracted images, the tools for extraction and annotation, and the annotation files are shared in the public domain and can be used for free for academic research.

The dataset captured from the traffic moving on the roads offers several challenges, such as variable size and orientation of the plates. Some samples are shown in Fig. 15. There are some very challenging images in the dataset, such as images of the plates partially covered with mud or with poor visibility due to low or bright light. This dataset can be very helpful in developing an LPR system for deployment in police cars.

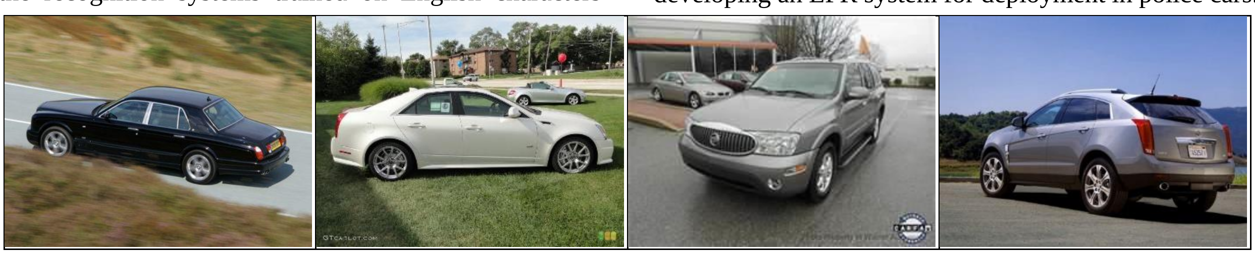


Fig. 14. Sample images of vehicles from Stanford dataset. https://ai.stanford.edu/~jkrause/cars/car_dataset.html.

Fig. 15. Sample images from the UJLP dataset. Download link: https://github.com/ishtiaqrasool/2022-Electronics-Automatic-License-Plate-Recognition-in-Real-World-Traffic-Videos.

III. Discussion

A summary of the different characteristics of the datasets discussed above is presented in Table I. The second column in the table shows the size of the dataset, i.e., the number of images and plates (if available). The largest dataset is provided on the Platesmania website, which provides plates from many countries. However, this dataset was not particularly designed for LP detection and recognition research; hence, the images are not labeled. Among the labeled datasets, CCPD is the largest, with 300,000 images of Chinese plates. The dataset of Czech Republic plates with 76,412 images is the next largest.

While comparing the detection and recognition results reported in different papers, the size of the dataset used for training must be considered; larger datasets would generally lead to better training and accuracy of the DNN. The quality of training images and the difficulty level of testing images are also important factors, and therefore the results reported using different datasets should be compared with caution.

TABLE I. Summary of the Reviewed Datasets

Dataset	# of Images	Image Size	Country
AOLP	AC: 681 LE: 757 RP: 611	AC: 352 x 240 LE: 640 x 480 RP: 320 x 240	Taiwan
Caltech Cars 1999	126	896 x 592	USA
Caltech Cars 2001	526	360 x 240	USA
PKU	G1: 810 G2: 700 G3: 743 G4: 572 G5: 1152	1082 x 727 1082 x 727 1082 x 727 1600 x 1236 1600 x 1200	China
Synthetic	Plates: 20,000 (training) + 2500 (validation) Non-Plates: 20,000 (training) + 2500 (validation)	768 x 384	Taiwan
University of Zagreb	500	1024 x 768, 640 x 480	Croatia
CCPD	300,000	720 x 1160	China
NTU Athens	716	1792 x 1312 800 x 600 640 x 480	Greece
UFPR-ALPR	4500	1920 x 1080	Brazil
Kurpiel et al. [8]	4070 LPs in 1829 images	1920 x 1080	Brazil
Goncalves et al. [9]	8683 LPs in 6600 images	1920 x 1080	Brazil
UCSD-Calit2	878	N/A	N/A
Spanhel et al. [11]	76,412 (ReID) 652 (HDR)	Not fixed	Czech Republic
AI Tunisia Hack 2019	900	Not fixed	Tunisia
Platesmania	Millions of images	Not fixed	Different countries
Stanford [12]	16,185	584 x 328 634 x 357 or close	USA
UJLP	2108	1920 x 1080 720 x 1280	Saudi Arabia

The third column in Table I compares the resolution of images in different datasets. Training a DNN with a large number of layers is a slow process; therefore, the training datasets generally comprise images of small sizes. However, with enhancements in the computational capabilities of GPUs, training images of Full High Definition (FHD) resolution, i.e., 1920 x 1080 pixels are not uncommon, as can be seen in the table. The execution time of a trained DNN to process a test image depends on the hardware platform as well as the image size. These factors should be given due consideration when comparing the efficiency of different methods reported by their respective authors.

IV. Conclusion

A detailed survey of different datasets used for training and testing of deep neural networks for license plate detection and recognition was presented. Different characteristics of these datasets were discussed. The information compiled in this paper can serve as a valuable reference for advancing research in this domain and benchmarking.

Acknowledgment

This work was funded by the Deanship of Scientific Research (DSR), University of Jeddah, under Grant No. UJ-03-18-ICP. The authors would like to acknowledge and thank the technical and financial support provided by DSR.

References

[1] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," IEEE Transactions on circuits and systems for video technology, Vol. 23, No. 2, 2012, pp.311-325.

[2] C. Anagnostopoulos, I. Anagnostopoulos, I. Psoroulas, V. Loumos, E. Kayafas, "License plate recognition from still images and video sequences: A survey", IEEE Transactions on Intelligent Transportation Systems, vol.9, no.3, pp.377-391, 2008.

[3] Y. Gong, L. Deng, S. Tao, X. Lu, P. Wu, Z. Xie, Z. Ma, and M. Xie, "Unified Chinese License Plate detection and recognition with high efficiency," Journal of Visual Communication and Image Representation, 2022, p.103541.

[4] H. Li, P. Wang, and C. Shen, "Toward end-to-end car license plate detection and recognition with deep neural networks," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 3, pp.1126-1136.

[5] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," IEEE Transactions on circuits and systems for video technology, vol. 23, no. 2, 2012, pp.311-325.

[6] M. S. Zandi, R. Rajabi, "Deep learning based framework for Iranian license plate detection and recognition," Multimedia Tools and Applications, vol. 81, no. 11, 2022, pp.15841-15858.

[7] G. S. Hsu, J. C. Chen, and Y. Z. Chung, "Application-oriented license plate recognition," IEEE Trans Veh Technol, vol. 62, no. 2, 2013, doi: 10.1109/TVT.2012.2226218.

[8] Weber and Perona, “Caltech Cars 1999,” CaltechDATA, Apr. 06, 2022. doi: 10.22002/D1.20084.

[9] Philip, Updikeand Perona, “Caltech Cars 2001,” CaltechDATA, Apr. 06, 2022. doi: 10.22002/D1.20085.

[10] Yuan, Y., Zou, W., Zhao, Y., Wang, X., Hu, X., & Komodakis, N. (2016). A robust and efficient approach to license plate detection. IEEE Transactions on Image Processing, 26(3), 1102-1114.

[11] Björklund, Tomas, Attilio Fiandrotti, Mauro Annarumma, Gianluca Francini, and Enrico Magli. "Robust license plate recognition using neural networks trained on synthetic images." Pattern Recognition 93 (2019): 134-146.

[12] Z. Xu, W. Yang, A. Meng, N. Lu, H. Huang, C. Ying, and L. Huang, "Towards end-to-end license plate detection and recognition: A large dataset and baseline." In Proceedings of the European conference on computer vision (ECCV), pp. 255-271. 2018.

[13] R. Laroca, E. Severo, L. Zanlorensi, L. Oliverira, G. Goncalves, W. Schwartz, D. Menotti, "A robust real-time automatic license plate recognition based on the YOLO detector", IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1-10, 2018.

[14] Kurpiel, Francisco Delmar, Rodrigo Minetto, and Bogdan Tomoyuki Nassu. "Convolutional neural networks for license plate detection in images." In 2017 IEEE International Conference on Image Processing (ICIP), pp. 3395-3399. IEEE, 2017.

[15] G. Goncalves, M. Diniz, R. Laroca, D. Menotti, W. Schwartz, "Real-time automatic license plate recognition through deep multi-task networks", IEEE SIBGRAPI Conference on Graphics, Patterns and Images, pp.110-117, 2018.

[16] G. Goncalves, S. da Silva, D. Menotti, W. Schwartz, "Benchmark for license plate character segmentation," Journal of Electronic Imaging, vol. 25, no. 5, pp. 34–53, 2016.

[17] J. Spanhel, J. Sochor, R. Juranek, A. Herout, L. Marsik, P. Zemcik," Holistic recognition of low quality license plates by CNN using track annotated data", IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, 2017.

[18] Krause, Jonathan, Michael Stark, Jia Deng, and Li Fei-Fei. "3d object representations for fine-grained categorization." In Proceedings of the IEEE international conference on computer vision workshops, pp. 554-561. 2013.

[19] I. R. Khan, S. T. A. Ali, A. Siddiq, M. M. Khan, M. U. Ilyas, S. Alshomrani, and S. Rahardja, "Automatic License Plate Recognition in Real-World Traffic Videos Captured in Unconstrained Environment by a Mobile Camera," Electronics, vol. 11, no. 9, 2022, p. 1408.

Prediction of Oil Production through Linear Regression Model and Big Data Tools

Rehab Alharbi[1], Nojood Alageel[2], Maryam Alsayil[3], Rahaf Alharbi[4], and A'aeshah Alhakamy[5]
Faculty of Computers and Information Technology, Master of Artificial Intelligence at University of Tabuk, Saudi Arabia[1,2,3,4,5]
Industrial Innovation & Robotics Center (IIRC), and Faculty of Computers and Information Technology,
Department of Computer Science at University of Tabuk, Saudi Arabia[3]
;

***Abstract*—Fossil fuels, including oil, are the most important sources of energy. They are commonly used in various forms of commercial and industrial consumption. Producing oil is a complex task that requires special management and planning. This can result in a serious problem if the oil well is not operated properly. Oil engineers must have the necessary knowledge about the well's status to perform their duties properly. This study proposes a linear regression method to predicate the oil production value. It takes into account various independent variables, such as the pressure, downhole temperature, and pressure tubing. The proposed method can accurately reach a very close prediction of the actual production value by achieving very interesting results at the end of this study.**

Keywords*—Big data; machine learning; oil production; regression model; features; prediction; PySpark

I. Introduction

The current rapid evolution of science and technology causes more progress in the oil field development, especially with the increasing demand for petroleum resources worldwide. For instance, the prediction of crude oil production represents the capacity of the oil field in the future, and it also directly impacts the future planning for this field [1].

In the petroleum industry, oil production and its future prediction have always been the center of interest to many scientists and researchers. This is because oil plays an impactful and effective role in energy production. Naturally, oil is produced within oil wells. In fact, many factors influence the production of the well including internal geological factors as well as geological location, time, and production equipment which are considered external factors. These influencing factors make it hard for the prediction of oil production to be precise and accurate, especially when using traditional methods like curve analysis and mathematical modeling [2]. Nonetheless, the need to accurately predict oil production is a necessity. As a matter of fact, the efficient accurate prediction can help save manpower, the consumed materials, and the resources required for the extraction process in addition to enhancing the economical aspect of the oilfields [3].

Oil production has an influential role in improving the economy. It is also a part of the political aspects of developing and established countries. Fossil fuel is considered a common industrial energy source due to the fact that it's versatile, easily transported, accessible, and expensive [4], [5]. Moreover, the oil serves various uses in industry, agriculture, energy transformation as well as commercial and residential services as shown in Fig. 1.

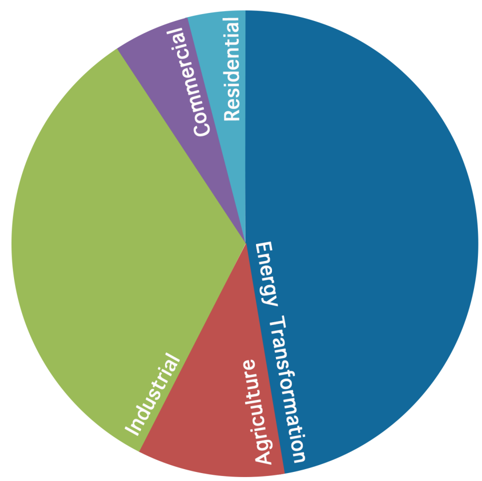


Fig. 1. Uses of fossil fuels and oil in multiple sectors: energy transformation, Agriculture, industrial, commercial, and residential

When discussing oil properties, it's important to mention that oil is found underground in porous rocks where rock strata are stacked above them creating pressure. This pressure is compensated by the pressure of the fluid (oil, gas, or water) which is in the reservoir. When drilling takes place, the fluid rises to the level of the surface because of the reservoir pressure. In addition, this lost fluid pressure is compensated by a displacement of the deeper fluid from underground to fill its place [6]. Consequently, the reservoir pressure decreases. It keeps decreasing gradually until it is no longer possible for the fluid to flow to the surface. In this case, a pump can be used to control the flow of fluid to the surface by controlling the difference between the pressure in the well and the reservoir pressure. Furthermore, the pump mode can be adjusted to have control over the bottom-hole pressure which can be a mode for setting the well's operations in the sense.

It is important to be able to predict the oil production rate or volume since it provides the engineers with useful information. As such, they can, in turn, evaluate and optimize reservoir management-related issues. Unfortunately, it is extremely difficult to perform highly accurate predictions without taking into consideration the subsurface conditions [7]. On the other hand, computing technology and data analytic can help resolve this issue of oil production forecasting while maintaining the complexity of the relationships between the inputs and outputs [8].

Prediction is the process of producing estimations based on observations. Upon analyzing the observable occurrences, it is then possible to project the data to make future predictions for a similar type of occurrence. Often these observable data are too large, and no deductions or predictions can be done by simply observing them. Manual analysis and interpretation are both tiresome and time-consuming in addition to the great error margin [9]. Prediction techniques vary in their type and content as can be seen in Fig. 2, yet some of them are more advanced than others.

In oil and petroleum exploration and development, artificial intelligence (AI) has gained a remarkable spot as it can be the solution for the issues that arise in this field. Recent studies have shown that using data-driven models ensures better prediction results than experience-based prediction models [10]. It is important to mention that AI is involved in the prediction of oil and gas properties, optimization of well layout plans, and prediction of reservoir physical properties and that of oil production among others [11]. Fig. 3 illustrates the various interaction models that can be used for predictions including data-driven methods such as regression, classification, and dimensional reduction.

AI incorporates computational power and human intelligence in order to achieve reliable smart systems that can resolve extremely nonlinear and highly complex problems. Thus, AI paves the way for computers to make data-based decisions [12]. Machine learning (ML), on the other hand, is included within AI. It delivers statistical tools for big data analysis and exploration. ML is actually further subdivided into supervised, unsupervised, and reinforced learning.

Even though AI has been involved in E&P practices before, the recent expansion of digital techniques in the oil field made it increasingly involved in advanced predictive and prescriptive analytics [13].

Moreover, AI provides great potential in solving issues in almost all areas of the oil industry including prediction, classification, and clustering. The data-driven systems can locate and define the relationship between oil production and other data acquired from the field. This is achieved through using sensors and ML models. In reservoir engineering, combining ML techniques with data analytics has various benefits [14]. The studies show that this combination can help predict the bottom-hole pressure, optimize water flooding, and forecast hydrocarbon production.

ML algorithms include Support Vector Machine, K-means, decision tree, Apriori, neural networks, and Naive Bayes algorithm [11]. Linear regression is one of the basic regression techniques for forecasting, predicting, or estimating. In the past years, specifically in 1800, Gauss created a method called the Least Square Method for the purpose of suiting an equation according to linear parameters. A linear regression illustrates the existing relationship between a random factor with a dependent factor, whose prediction or estimation is the center of focus in this study. In both research and statistical studies, linear regression is still frequently used [15]. Linear Regression as can be seen from Fig. 2 is one of the qualitative prediction methods.

Notably, the researchers of this study are highly motivated by the increasing importance of oil and fossil fuels production. Hence, the main objective of the study is to develop a system that can analyze the fed data and transform them into valuable inputs that can be used for oil production prediction in the future. The proposed system is based on Linear Regression which achieves high levels of accuracy. For a better understanding of the presented work, the developed model is evaluated to answer the following research questions:

(R.Q.1) *What factors directly affect the oil production value?*

(R.Q.2) *Can a Linear Regression model effectively predict the production value of oil based on different factors?*

This study is composed of three parts. The first part presents some related works to the prediction of oil production. These studies use different methodologies for the purpose of analyzing the data in the oil fields by presenting suitable models. Then, the methodology is the second part of this study which is explained in detail. This includes the levels of the structure of the dataset, implementations, importing the data, exploratory data analysis, data processing, data scaling, feature selection, and applying machine learning algorithms. The last part is the results part where the researchers explain that the linear regression method suggested by them has proven to be highly successful by having results that are really close to the actual results.

It is important to mention that this study has a high significance in the field of fossil fuels, especially in oil production. As such, it attempts to decrease and reduce the problems that may occur in the prediction process of oil production. The researchers highlight that this study with its simple method, linear regression, can achieve high and real results in comparison to other methods with much more complex features. Such a method can be flexible, practical, and effective for explaining and illustrating the process of successful oil predictions. It can add more knowledge and information for future research for the purpose of enhancing and saving the most significant source of energy for the upcoming generations.

II. Related Work

When considering the topic of oil production, many studies come to light especially those that attempt to accurately predict the rate or volume of oil production. ML models are quite popular in this case. Nonetheless, many studies prefer to use linear regression specifically due to its many advantages.

Emeke and Bello addressed the issue of predicting the oil production volume in an area within Nigeria through linear regression techniques [16]. In their study, the authors have made use of MATLAB in analyzing the data acquired from the oil field from Kwale, Niger Delta. In addition, multivariate linear regression has been chosen as a model for its capability of housing multiple variables. For regression, the input variables are the metered volume, the basic sediment and

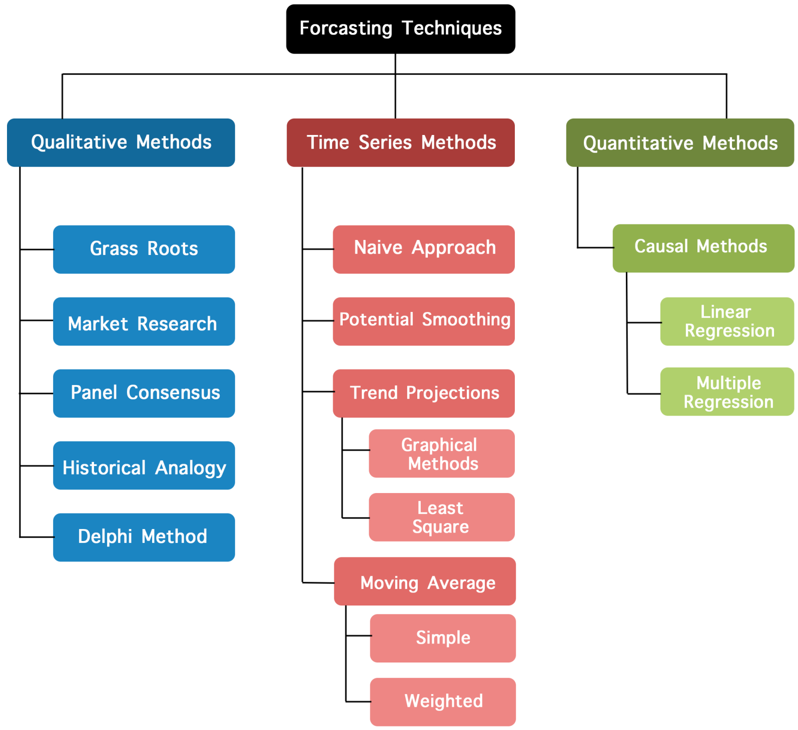


Fig. 2. Prediction traditional methods including qualitative, time series and quantitative methods

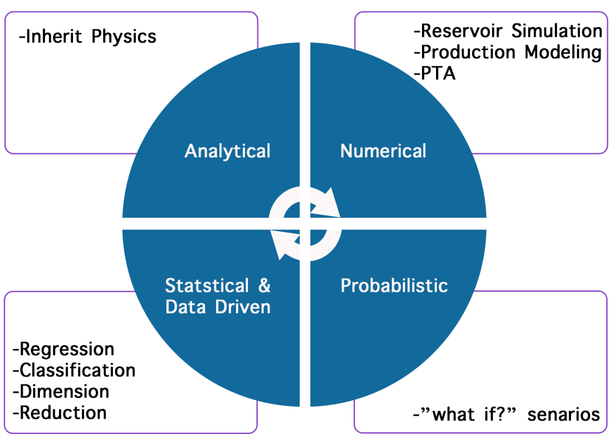


Fig. 3. Prediction tools including data-driven methods such as regression models

water, the volume correction factor, the metered factor, gross standard volume, API @60, and temperature among others. These data are collected over the course of several months, and the entire dataset is split into 70% for training the LR model and 30% for testing it. After that, a couple of models are considered (with different input variables) in contrast to the new generalized model which has been obtained by coefficient comparison and inspection of the lowest root mean square error value RMSE. One of the distinctive features of the new model is the exclusion of some factors like water percentage and basic sediment. As a result, the RMSE value for the new model is much lower than the old one. Furthermore, the residual values are calculated to see the differences between the predicted values and the actual values, and the new generalized model, which are low, indicates a close prediction. In addition, the R-squared value is 0.9889.

In another study, Mgbemena and Chinwuko aim to predict the crude oil production value in Nigeria through implementing multiple linear regression [17]. In this study, the authors attempt to produce a daily, weekly, monthly, and yearly forecast. To do so, the authors initially analyzed many forecasting models before creating the LR model. The past historical data for the oil field are collected from the years 2010 to 2016. For linear regression, the least-squares equation is used. Overall, the performances of six models are compared. Among the graphical forecasting, 3-year simple moving average, 6-year simple moving average, 8-year simple moving average, least squares LR and the exponential smoothing with o.1 or 0.6 as weighing factors, the Least Squares Linear Regression model have achieved the lowest mean squared error MSE value. This means that the lowest mean method makes a good prediction with minimal error. After the comparison, the authors developed their own multiple linear regression models to predict the daily oil production values. This model takes into consideration the number of natural wells which are flowing, the number of gas-lifted weak wells, and the amount unbiased compressors which are running. As a result of comparing the developed system's predicted value and the actual value, the system misses the direct hit point by 43 barrels per day, which is a very satisfactory result.

In their study, Kumar and his colleagues aim to develop

TABLE I. COMPARISON OF LITERATURE REVIEW STUDIES ACCORDING TO YEAR, AIM, AND APPROACH

Paper	Year	Aim	Approach
[16]	2019	To create a system that can accurately predict oil production in the field in Nigeria	Multivariate linear regression model has been developed taking into consideration the following factors: metered volume, gross standard volume, metered factor, and volume correction factor.
[17]	2020	To develop a model that can perform a precise forecast for oil production in Niger Delta Region	A comparison of many models has been performed, and the multiple variables regression model has been finally developed achieving very good results.
[18]	2019	To predict the oil and gas production rates based on Machine Learning	RapidMiner tool is used in addition to linear regression SET ROLE operator and APPLY MODEL operator.

a forecasting model based on machine learning to predict gas and oil production [18]. To execute the study, data has been acquired from sensors and processed, then the CRISP (Cross Industry Standard Process) model has been applied to the processed data. This involves understanding the data, preparing data, modeling, and evaluating. After that, the data is imported to Rapid Miner where correlation analysis is performed to check for existing relationships between them, the variables, and the dependent variable. Next, linear regression is applied based on the identified variables from the correlation analysis. Finally, data visualization techniques are used to visualize and discuss the data. Ultimately, six variables are chosen for input to achieve oil production and output prediction, including average wellhead temperature, average choke size, average gas lift rate, etc. Upon experimentation and testing, bottom-hole pressure has been found to be an important variable in predicting the average oil production rate. Table I presents a comparison between the literature review studies according to the year, aim, and approach.

III. METHODOLOGY

A. Dataset

The Norwegian oil and gas company Equinor has announced starting in June 2018, all of the data from the Norwegian continental shelf are available for research and study. This has given academic institutions, students, and researchers worldwide permission to use crucial scientific information in line with the Equinor Open Data License without the requirement for further written consent.

The Volve production data has been released in the form of an excel file which is made up of two (02) sheets, namely Daily Production Data and Monthly Production Data. Oil production estimation is one of the major duties of production engineers, depending on several operational parameters that include tubing differential pressure, bottom hole pressure, and wellhead pressure. This estimation process can happen as an effective way of applying empirical correlations or nodal analysis. Furthermore, physics is generally included in most of the literature work since it involves a lot of assumptions.

The researchers of this study are inspired to apply a completely different approach based on Data-Driven applications, such that physics is being taught to the computer based on data only. This approach ensures that no complexity or assumptions are being made and that the results are purely generated from the fed data. Both linear and polynomial regression models are going to be used to predict the oil production from the date of the Volve field daily production. In the Norwegian North Sea, a hydrocarbon reservoir termed Volve has been operational for a couple of years between 2005 and 2016. Volve has a 54% recovery which is considered good. Additionally, the Volve dataset is considered the top open-source dataset for exploration and production in terms of completion.

B. Implementation

Apache Spark is a framework for data processing that can swiftly process operations on extremely large data sets and distribute processing operations over several computers, either individually or in conjunction with other distributed computing technologies. It is a super-quick machine learning and large data analytics engine. The Apache Spark community has published a tool, PySpark, to enable Python with Spark. Python programmers may work with RDDs by using PySpark.

1) Spark MLlib: Spark MLlib is used in Apache Spark to execute machine learning. Popular tools and algorithms may be found in MLlib. MLlib is a scalable Machine Learning library in Spark that offers both high-quality algorithms and high performance. Clustering, classification, regression, collaborative filtering, and pattern mining are examples of machine learning algorithms. Additionally, MLlib contains lower-level machine learning primitives like the general gradient descent optimization technique.

The main Machine Learning API of Spark is spark.ml. For creating ML pipelines, the library Spark.ml provides a higher-level API built on top of DataFrames. Below are some utilities for Spark MLlib:

- Pipelines
- Featurization
- ML Algorithms
- Persistence
- Utilities

2) ML algorithms: The core of MLlib is its ML algorithms. These include well-known learning techniques including collaborative filtering, clustering, regression, and classification. The goal of MLlib is to standardize APIs to make it simpler to integrate several algorithms into a single pipeline or workflow. The Pipelines API is one of the core ideas, and the scikit-learn project serves as an inspiration for the pipeline concept.

3) Transformer: An algorithm known as a Transformer may change one DataFrame into another DataFrame. Basically, a Transformer performs the transform () function, which adds one or more columns to one DataFrame to change it into another. For example: A feature transformer may take a DataFrame, read one column (for example, text), map it into another column (for example, feature vectors), and then produce a new DataFrame with the mapped column attached.

A learning model may use a DataFrame as an input, read the column containing feature vectors, predict the label for every feature vector, and then produce a new DataFrame with the predicted labels attached as a column.

4) Estimator: An algorithm which is known as an Estimator may be fitted to a DataFrame to create a Transformer. In terms of technical implementation, an Estimator uses the function fit(), which takes a DataFrame and outputs a Model, a Transformer. For instance, using fit() trains a Logistic Regression Model, which is a Model and subsequently a Transformer, from a learning method like Logistic Regression, which is an Estimator. Both Estimator.fit() and Transformer.transform() are stateless functions. Alternative notions may enable stateful algorithms in the future. Each Transformer or Estimator instance has a unique ID which may be used to provide parameters (discussed below).

5) Featurization: Selection, dimensionality reduction, feature extraction, and transformation are all parts of featurization. Feature extraction is the process of extracting features from raw data. Furthermore, the term "feature transformation" refers to the resizing, updating, or changing of features. In feature selection, a small subset of essential characteristics is chosen from a large pool of features.

6) Pipelines: A pipeline specifies an ML process by connecting several transformers and estimators. Additionally, it offers resources for building, assessing, and fine-tuning ML Pipelines. It is typical in machine learning to run a series of algorithms to analyze and learn from data. Such a process is represented by a pipeline in MLlib, which consists of a series of pipeline stages (Transformers and Estimators) that must be executed in a certain order. In this part, the researchers of this study utilize this simple method as an example. As such, the pipeline example below does the data preparation in the following order:

- Use the String Indexer method to determine the index of the category columns.
- Use OneHot encoding to present the category columns.
- Use the String indexer to locate the column of the output variable "label"
- VectorAssembler is used to assemble category and numerical columns. A transformer called VectorAssembler creates a single vector column from a provided list of columns.

C. Importing Data

The first step is to import the data, including three essential libraries, namely Spark which is used to import product data and to develop DataFrame, in addition to the Seaborn and Matplotlib libraries.

There are several parameters considered in the uploading of the dataset. It contains various data such as the good type being water injection or oil production, as the average downhole pressure, average drill pipe tubing, date of production, oil volume, average downhole temperature, gas volume, average choke size, average well-head temperature, average annulus pressure, average well-head pressure, water volume, and type of flow (production or injection). In addition, the data on oil production is documented based on the daily collection.

D. Exploratory Data Analysis

It is critical to perform exploratory data analysis in any machine learning approach, even though it is also time-consuming. This step is essential since upon encountering the data, it must be understood before continuing with the other steps. There are many processes included in the exploration of data, such as discovering patterns, checking assumptions, and spotting anomalies through visualizations and statistical summaries. Fig. 4 illustrates the distribution of acquired data of the seven wells.

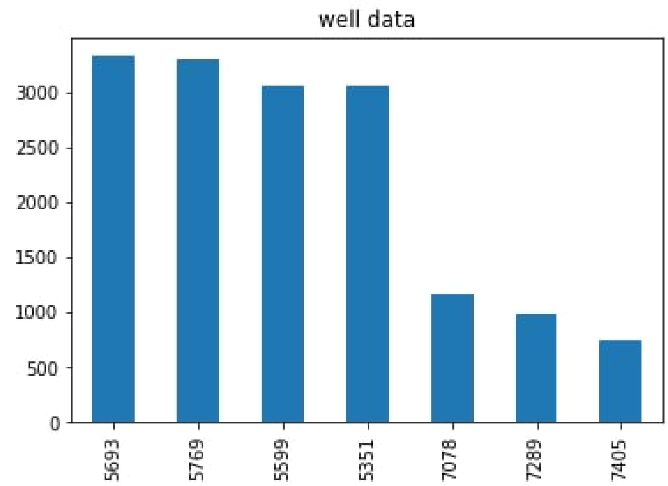


Fig. 4. Distribution of data acquired from seven wells

Fig. 5 shows the ECDF plots of each of the wells. From these plots, it can be observed that 40% of the data points relative to well number 4 have zero bore oil production. For well number 5, the observations show that 20% of their data points have zero bore oil production. These indicate that there is no flow in these wells since zero bore oil production is different from not available NA values. Thus, when the seven wells are considered in this study. Well number 4 and well number 5 data should be eliminated, as they can't be used for training. On the other hand, well number 1 shows empty values in the ECDF plot, and well number 2 has some missing values. Hence, for these two wells, the ECDF of water injection volume is plotted to check if they are injectors.

Fig. 6 shows the water injection plots for wells number 1 and number 2. By analyzing the graph, it becomes clear that well number 1 is an injector well. On the other hand, data of well number 2 indicates that it is an injector in addition to being a producer.

Generally, wells can be either injectors or producers. When fluids are poured into the underground to be placed in porous geological formations such as limestone, deep sandstone, or a shallow soil layer, then the well has termed an injection well. Additionally, the injected fluid can be either water, salt water "brine", wastewater, or chemically treated water. On the other hand, a well specified for extracting oil or gas is called a production well. Fig. 7 shows the difference in well data between the two well types.

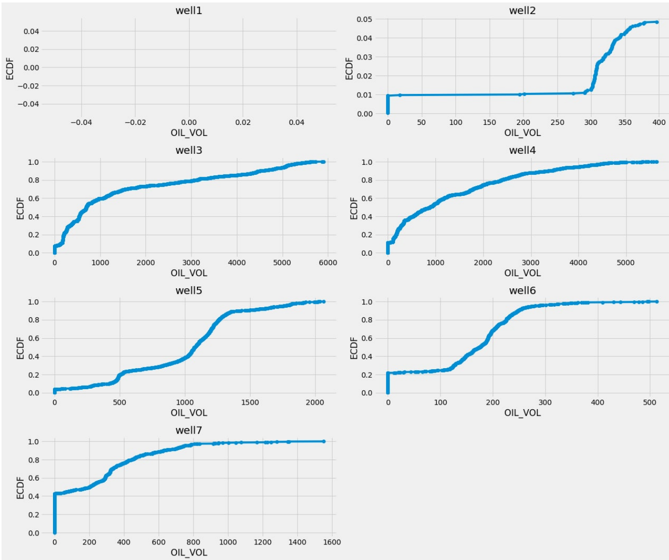


Fig. 5. ECDF plots for each of the seven wells

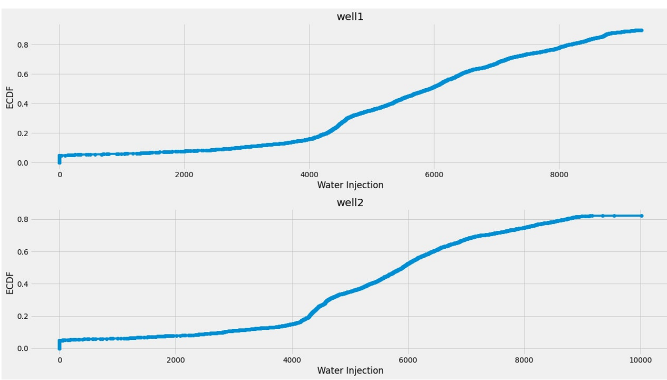


Fig. 6. ECDF plots for well number 1 and 2 referring to water injection data

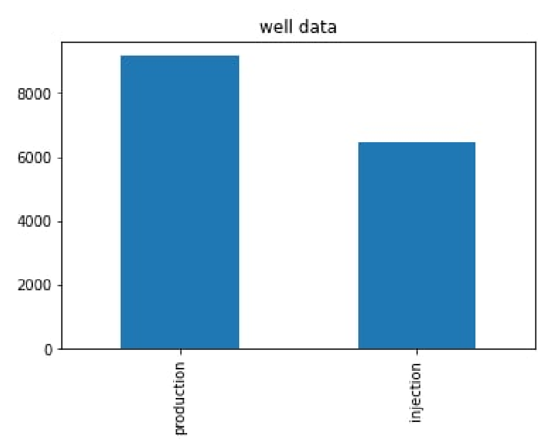


Fig. 7. Well data for a production well in comparison to an injection well

E. Data Pre-Processing

Data preparation is the next step after data analysis. Preprocessing the data involves dealing with string formats inserted in a numerical column, negative values in permeability columns, empty columns, null values, and theoretically incorrect values.

From analyzing the data using a box plot, it is evident that the data is highly skewed depending on the well, and missing data need to be fixed. The forward filling is the technique to be performed to fill in the missing data here. However, usually, in the case of missing data, the mean value is used, yet this cannot be the case in the data of this study since it is skewed. For this reason, the null value is filled with a value just above it.

F. Data Scaling

In order to prepare the data for machine learning approaches, data scaling or normalization should be done. The scaling process is meant to alter the numeric values within the columns into a common scale while keeping the differences in the range of values and maintaining the information. The data is transformed in a way that the features are within a specific range [0, 1].

$$x' = (x - x_{min})/(x_{max} - x_{min}) \tag{1}$$

G. Feature Selection

Feature selection is also called feature engineering since the input features of the data are examined and evaluated for their importance and impact on the results. In this way, the features, that contribute to getting the precise predicted value and are correlated to the anticipated output, are chosen either automatically or manually. The importance of this step is explained when the accuracy of the model decreases when using irrelevant features is noticed. However, using highly correlated features is also tricky since it might lead to reducing the accuracy given that the input data becomes homogenous and lacking variability. It might also lead to data leakage which increases the accuracy beyond acceptance. For this reason, correlation heat maps must be generated. The correlation heat maps are generated to determine the highly correlated features. Fig. 8 shows the variation of oil production volume with respect to downhole pressure and temperature.

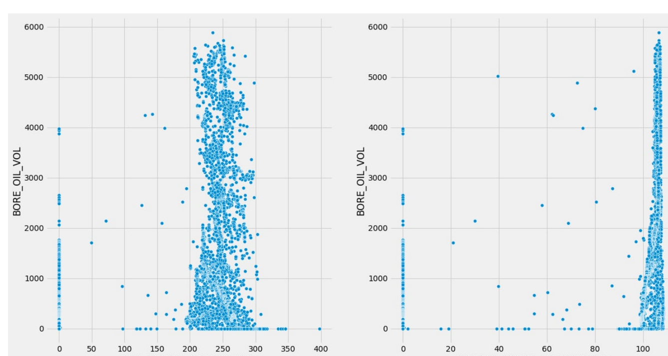


Fig. 8. Variation of oil volume with respect to downhole pressure and temperature

H. Applying Machine Learning Algorithm

Linear regression is one of the many existing regression methods, which are considered among the most important machine learning and statistical techniques. In regression, relationships among the different variables are explored. For instance, applying regression in a company allows exploring how the salary of its employees changes according to their experience, education, their role in the company, etc. In this case, the data of each employee is a representation of one observation, and in this regression a problem, the dependent variable is the salary, whereas the independent variables are the role, education level, years of experience, etc.

In a regression analysis, a chosen phenomenon is considered with a group of observations, where each observation involves at least two features or variables. Then, an attempt of finding the relationship of dependency between the various features or variables occurs based on assumptions and further testing and analysis. Thus, regression not only finds how a phenomenon is influenced by a feature but also how the different features are related to each other. The outputs of regression are often denoted by y, whereas the inputs are denoted by x, such that the independent features can be put in a vector $x = (x_1, ..., x_r)$ with r being the number of inputs.

Forecasting is one of the other uses of regression, where output can be predicted based on a set of predictors. Regression is used in many different fields, including economics, computer science, and the social sciences. The importance of regression increases with the increase of data volume and practical value of data.

Outputs, responses, and dependent variables are the terms used for the dependent features, whereas inputs, regressors, and predictors are the terms used for the independent features. In regression problems, there are often one continuous dependent variable and many continuous, categorical, or discrete independent variables.

Linear Regression is a simple supervised machine learning model. In this problem, the aim is to create a linear relationship between the dependent variable "Oil Production", and the other training independent variables. The linear equation will assign a coefficient to each training feature, and an intercept is added to the equation as well. Input instance –feature vector: $x = (x_0, x_1, \ldots, x_n)$, and the predicted Output: $y = w_0x_0 + w_1x_1 + \ldots w_nx_n + b$, therefore, for parameters to estimate (PtoE) is represented as follow:

$$PtoE = \begin{cases} \hat{w} & := \hat{w_0}, ..., \hat{w_n}\text{:feature } \frac{weights}{model}\text{coefficients} \\ \hat{b} & : \text{constant bias } \frac{term}{intercept} \end{cases} \tag{2}$$

The variation of actual responses $y_i, i = 1, ..., n$ occurs partly due to the dependence of the predictors x_i. However, there's also an additional inherent variance in the output.

The coefficient of determination R^2 resembles the degree of variation of y depending on x, through the particular regression model. If R^2 has a large value, it means that the model is highly capable of explaining the variation of the dependent variable according to the independent variables, thus it is a better fit. The value $R^2 = 1$ corresponds to $SSR = 0$.

The whole workflow of the represented system can be summed up in Fig. 9 Variables such as the downhole temperature, pressure, DP tubing, annulus press, and choke size are considered as an input for our linear regression model, which can develop an output of predicted oil production volume based on these features.

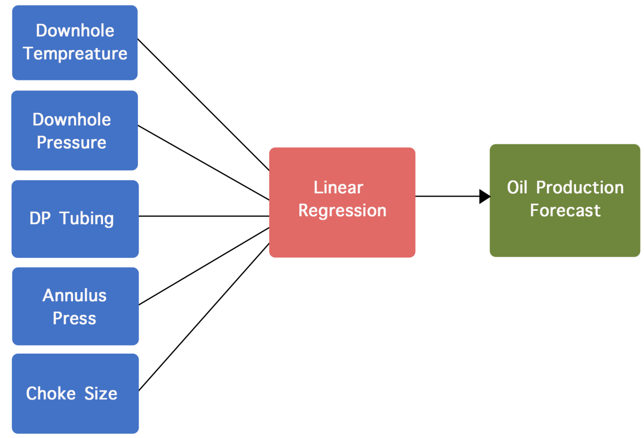


Fig. 9. Flow chart of oil production prediction system based on linear regression model

IV. Results

Scatterplot and Matplotlib libraries are used to show how the model is predicting the oil production in comparison to the actual production volume (Fig. 10).

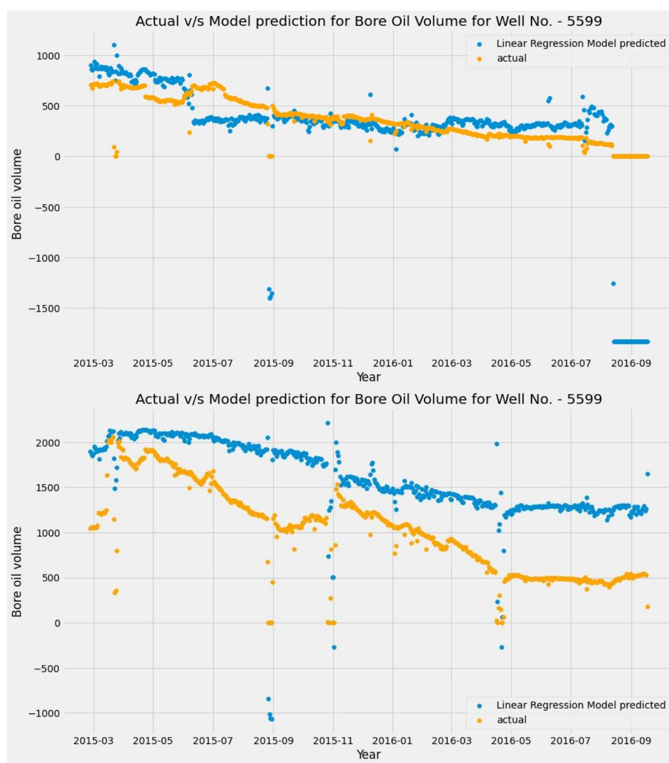


Fig. 10. Scatterplot and Matplotlib libraries using to show how the model is predicting the oil production in comparison to the actual production volume

$$RMSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n} \tag{3}$$

RMSE	MSE
0.30292596179787934	0.2861434684281355

Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n}} \tag{4}$$

Root Mean Squared Error (RMSE) is the square root of Mean Squared error. It measures the standard of residuals.

Based on the above two results, it obvious that the model which is used in this study is correct and accurate because the mean absolute error that has been received after the analyses is 0 (MSE is 0.2 & RMSE is 0.3). This adds more advantages to the method used by the researchers of this study.

The results of this study can be explained by comparing the results of other studies presented in the literature review part. In study [16], the features used in the model are mostly related to volumes such as standard volume and metered volume. However, the model of this study, linear regression, takes into consideration much broader features such as temperature, pressure, and choke size that explains the system more. Moreover, in study [17], eight models have been used to make oil predictions which are very complex. Whereas, in this study, only a linear regression model has been used due to its simplicity and accuracy which have achieved really interesting results. Hence, a good prediction has been obtained in this study through only one mode. Furthermore, the researchers in this study haven't focused on each feature individually and which one is more important, but have favored getting a general model that takes all the features into account contrary to what Kumar and his colleagues have done in study [18].

V. Conclusion

Today, fossil fuels, such as oil, are the most important sources of energy. They are commonly used in various industrial and commercial sectors. However, their production is complex and requires special management and planning.

It is crucial that oil engineers are aware of the status of their wells and are able to perform their duties properly. In this study, they proposed a linear regression method to estimate the oil production value. This method can be used to analyze the various independent variables that affect the oil production process.

The proposed method would be able to accurately predict the oil production value. It can also achieve interesting results by analyzing the data collected during the study.

References

[1] H. Hu, Y. Pu, and X. Guan, "Oil field crude oil production level prediction method based on ahp-pso-bp," in *2020 IEEE 8th International Conference on Information, Communication and Networks (ICICN)*, 2020, pp. 214–218.

[2] W. Liu, W. D. Liu, and J. Gu, "Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network," *Journal of Petroleum Science and Engineering*, vol. 189, p. 107013, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092041052030108X

[3] S. Pan, J. Wang, and W. Zhou, "Prediction on production of oil well with attention-cnn-lstm," *Journal of Physics: Conference Series*, vol. 2030, no. 1, p. 012038, sep 2021. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/2030/1/012038

[4] K. I. Wong and P. K. Wong, "Optimal calibration of variable biofuel blend dual-injection engines using sparse bayesian extreme learning machine and metaheuristic optimization," *Energy Conversion and Management*, vol. 148, pp. 1170–1178, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890417306088

[5] Y. Xing, Z. Zheng, Y. Sun, and M. Agha Alikhani, "A review on machine learning application in biodiesel production studies," *International Journal of Chemical Engineering*, vol. 2021, 2021. [Online]. Available: https://doi.org/10.1155/2021/2154258

[6] A. Davtyan, A. Rodin, I. Muchnik, and A. Romashkin, "Oil production forecast models based on sliding window regression," *Journal of Petroleum Science and Engineering*, vol. 195, p. 107916, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0920410520309712

[7] X. Ma and Z. Liu, "Predicting the oil production using the novel multivariate nonlinear model based on arps decline model and kernel method," *Neural Computing and Applications*, vol. 29, no. 2, pp. 579–591, 2018. [Online]. Available: https://doi.org/10.1007/s00521-016-2721-x

[8] S. D. Mohaghegh, "Subsurface analytics: Contribution of artificial intelligence and machine learning to reservoir engineering, reservoir modeling, and reservoir management," 2020.

[9] M. Rajesh *et al.*, "Price prediction for pre-owned cars using ensemble machine learning techniques," *Recent Trends in Intensive Computing*, vol. 39, p. 178, 2021.

[10] I. Makhotin, D. Koroteev, and E. Burnaev, "Gradient boosting to boost the efficiency of hydraulic fracturing," *Journal of Petroleum Exploration and Production Technology*, vol. 9, no. 3, pp. 1919–1925, 2019.

[11] S. Qin, J. Liu, X. Yang, Y. Li, L. Zhang, and Z. Liu, "Predicting heavy oil production by hybrid data-driven intelligent models," *Mathematical Problems in Engineering*, vol. 2021, 2021.

[12] Z. Tariq, M. S. Aljawad, A. Hasan, M. Murtaza, E. Mohammed, A. El-Husseiny, S. A. Alarifi, M. Mahmoud, and A. Abdulraheem, "A systematic review of data science and machine learning applications to the oil and gas industry," *Journal of Petroleum Exploration and Production Technology*, vol. 11, no. 12, pp. 4339–4374, 2021.

[13] M. Maucec and S. Garni, "Application of automated machine learning for multi-variate prediction of well production," in *SPE Middle East Oil and Gas Show and Conference*. OnePetro, 2019.

[14] C. S. W. Ng, A. Jahanbani Ghahfarokhi, and M. Nait Amar, "Application of nature-inspired algorithms and artificial neural network in waterflooding well control optimization," *Journal of Petroleum Exploration and Production Technology*, vol. 11, no. 7, pp. 3103–3127, 2021.

[15] F. Salim and N. A. Abu, "Used car price estimation: Moving from linear regression towards a new s-curve model," *International Journal of Business and Society*, vol. 22, no. 3, pp. 1174–1187, 2021.

[16] K. B. C. Emeke, "A novel model developed for forecasting oilfield production using multivariate linear regression method," *Journal of Science and Technology Research*, vol. 29, no. 2, pp. 579–591, 2019.

[17] C. Mgbemena and E. a. Chinwuko, "Forecast of crude oil production output in an oil field in the niger delta region of nigeria," *International Journal of Industiral Engineering & Producion Research*, vol. 31, no. 1, 2020. [Online]. Available: http://ijiepr.iust.ac.ir/article-1-891-en.html

[18] A. K. K., R. Ramasree, and M. Faisal, "Performing predictive analysis using machine learning on the information retrieved from production data of oil & gas upstream segment," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, 2019, pp. 0385–0391.

A Comparison of Pathfinding Algorithm for Code Optimization on Grid Maps

Azyan Yusra Kapi[1], Mohd Shahrizal Sunar[2*], Zeyad Abd Algfoor[3]
Computing Sciences Studies-College of Computing-Informatics and Media, Universiti Teknologi MARA,
Johor Branch, Pasir Gudang Campus, 81700 Johor, Malaysia[1]
Department of Emergent Computing-Faculty of Computing, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia[1, 2]
Media and Game Innovation Centre of Excellence-Institute of Human Centered Engineering,
Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia[1, 2]
Department of Computer Science-College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq[3]

***Abstract*—There have been various pathfinding algorithms created and developed over the past few decades to assist in finding the best path between two points. This paper presents a comparison of several algorithms for pathfinding on 2D grid maps. As a result, this study identified Jump Point Search Block Based Jumping (JPS (B)) as a potential algorithm in term of five evaluation metrics including search time. The comparisons pointed out the potential algorithm and code optimization was performed on the selected JPS(B) algorithm, and it was named JPS(BCO). This paper also explores issues regarding the JPS(B) and how to resolve them in order to optimize access to the index pointer. The presented enhance JPS(BCO) is capable to search optimal path quicker than the original JPS(B) as demonstrated by experimental findings. An experiment of different size grid maps is conducted to validate the ability of the proposed algorithm in term of search time. The comparative study with original JPS (B) exhibits the enhancement that has more benefits on grid maps of different size in terms of search time.**

***Keywords*—Comparative; jump point search; optimization; pathfinding; path planning**

I. Introduction

Pathfinding algorithm has become one of the popular techniques to search for a path while avoiding obstacles at the same time. There have been numerous applications of pathfinding such as robotics, virtual reality [1], and commercial games [2]. In the past, pathfinding has traditionally focused on finding the shortest route. However, nowadays, it also focuses on finding the safest, cheapest, or most convenient route to avoid tolls, roads, or other obstacles [3].

Maps and graphs are common methods for representing environments in pathfinding. Most pathfinding algorithms consider the pathfinding environment as a key attribute to determine navigation performance [4]. Navigation meshes, grid maps, and waypoint graphs are three popular pathfinding environments. The superiority of one pathfinding environment over another has been debated for many years. A few advantages and limitations of the most common pathfinding environments used in games are summarized in [5]. Grid maps are the most popular pathfinding environment due to the simplicity and ease of use [6]. Furthermore, it is also relatively fast to generate a grid map and it only includes two categories of cells: passable and block able [7]. The movement of an agent in grid map is limited to eight possible directions. Every vertical and horizontal movement in the grid has one cost unit, while the diagonal movement has a cost of 1.4 units.

Over the past decades, the performance of the pathfinding has been evaluated through various comparisons and analysis. Recently, [8] has compared traditional pathfinding which is A* algorithm and Depth First Search (DFS) with state-of-the-art algorithms, Jump Point Search (JPS) and Subgoal Graph. By testing the four algorithms on eight different grid maps, the author discussed several of the advantages and disadvantages of each algorithm according to the grid maps. As another example, [9] also compared two well-known algorithms: A* and Iterative Deepening A∗ (IDA*). They concluded that when there are no obstacles on the map, IDA* generally performs better than A*. However, when it comes to memory and time usages, IDA* may perform worse if opponent characters are in parallel positions and blocked by obstacles.

On the other hand, [10] conducted an analysis to compare performance of A∗ and Basic Theta∗ algorithm. Results from this study indicate that the A* and Basic Theta* algorithms have both similar completeness and time taken, but the A* has the benefit of searching less nodes, while the Basic Theta* algorithm returns shorter results.

Despite being one of the oldest algorithms, A* is still being favored to be included in the various researches for comparison. Although there are numerous researches in the literature on the comparative analysis of several pathfinding algorithms, however, to date, there have been very few research on the comparative analysis of Jump Point Search algorithm. This paper aims to compare A* and four versions of Jump Point Search using source code which are made available by the author [6]. The total five algorithms which included in this study are A*, JPS, JPS(B), JPS+, and JPS+ (P).

One of the evaluation metrics that is commonly used in comparisons and analyses is the amount of time taken during search. In video games [11], search and rescue (SAR) [12], and unmanned aerial vehicle (UAV) navigation [13], the search time is one of the most critical aspects. Therefore, the key motivation for this study is to reduce the search time.

*Corresponding Author
Ministry of Education Malaysia and Universiti Teknologi MARA (sponsors).

Apart from comparing the pathfinding algorithms, this paper will also discuss code optimizations on the potential algorithm JPS(B) which are derived from the preliminary analysis. The code optimization yields a slight increase in terms of the search time. Thus, the fundamental objective and contribution of the current paper aim to present a comparative analysis on several pathfinding algorithm and a slight code optimization to the potential algorithm. This comparative analysis is important to other researcher for evaluating the performance of pathfinding algorithm and the code optimization performs on it.

In the remainder of the paper, the sections are arranged as follows. In Sections II, reviews of related works on JPS algorithms and descriptions of the JPS are provided. The methodological approach for the experimental setup was explained in Section III. Experimental result is presented in Section IV, while JPS(BCO) optimization and result is discussed in Sections V and VI, respectively. Conclusions and recommendations for future work are presented in Section VII.

II. RELATED WORKS

It was found that most pathfinding algorithms are based on A*, regardless of whether they are single or multi-agent pathfinding. A* algorithm was successful in solving pathfinding problems, and since that, numerous studies have concentrated on improving and optimizing A* algorithm. Path scoring is used by A* algorithm to determine the optimal path from the initial node to the end node [14]. Another prominent pathfinding algorithm is the Jump Point Search (JPS) which is the successor of A* variants introduced by [15]. They affirmed that the JPS accelerated more than A* and JPS has gain many attentions from other researcher after that.

Using JPS, an undirected and eight-connected grid map can be identified and eliminated from many path symmetries through combination of A* search and pruning rules. JPS algorithm utilizes pruned neighbor rules to determine which nodes should be searched while jump points are determined by their location relative to forced neighbor. Later in 2014, enhanced JPS is presented in [6] which includes four varieties of enhancement which includes jumping by block, new pruning rules and adds a pre-processing step prior to searching process.

The existing literature on JPS is extensive and focuses particularly on describing some of the more recent developments and enhancements. A summary of the research on JPS's improvement can be found in Table I.

TABLE I. SUMMARY OF JPS'S ENHANCEMENT

Ref.	Explanation		
	Objectivity	***Description***	***Application***
[16]	Improve waiting steps and movement steps	IJPS combines JPS with Congestion control in two stages: online and offline mode.	Autonomous Ground Vehicle (AGV)
[17]	Decrease search time and secure distance for robot and the barrier	SD-JPS merge the idea of a jump point with the node domain matrix to suggest a different jump point and limitation that satisfy the JPS 's quick search results and achieve the secure distance between the robot and the barriers	Robotic
[18]	In addition to identifying the direction of the next path point relative to the current path point, it uses vector cross product to verify connectivity between the previous and last points of each original point.	Enhancements are presented in this paper to minimize redundant path points and optimize path length by shortening paths. In order for a robot to move, its pose must be adjusted in points, so the vector cross product and the vector dot product are applied.	Home Service Robot
[19]	Using grid signage, the grid environment is pre-processed, and a valuation function is used to determine the optimal path through the grid.	To improve the path efficiency, this algorithm optimizes the open list with the minimum binary tree, and enlarges the weight coefficient to choose the appropriate valuation function.	Radar simulation system
[20]	The InvJPS algorithm attempts to resolve the problem of inventory-driven pathfinding in the literature	When used in inventory-based variants of game maps, InvJPS maintains JPS's optimality guarantees and its symmetry breaking benefits.	Video Games

Based on Table I, it is shown how JPS has been enhanced and applied to several areas, including home service robots and even in logistics for AGV. Most of the improvements were focused on reducing search times, which are considered valuable even for a few seconds. Apart from that, the aim is to minimize or shorten the path length as shown in (Ma et al., 2019). The related works conclude that the trends in optimizing the JPS algorithm focus on faster search times, which is also the motivation for this study.

III. EXPERIMENTAL SETUP

Original source code from [6] can be downloaded from https://bitbucket.org/dharabor/pathfinding. The given source code consists of several pathfinding algorithm in the same program. In Table II, an overview of the five pathfinding algorithms selected for comparison is presented in detail.

Table II shows that all algorithms were unweighted, except for the A* algorithm. A weighted version of A* algorithm is also available as used in [21] for their enhancements. However, for the purpose of simplification and standardization, only the unweighted version of the algorithms has been considered in this study, and the weighted version has been discarded.

While the given source code is in C++, [6] developed the code on Linux 20.04. Computer with Windows 10 operating system were used for the purpose of this study. Therefore, an Oracle VM VirtualBox 6.1.30 and Ubuntu 20.04 platform was used to match the original platform.

TABLE II. OVERVIEW OF PATHFINDING ALGORITHM USED IN THIS STUDY

Algorithm	***Algorithm name (source code)***	***Description***	***Unweighted***	***Weighted***
A*	astar	Traditional algorithm which utilizes an open list and close list to store each visited and unvisited nodes.	✓	✓
JPS	jps	Enhance A* by utilizing a jump point to skip unnecessary nodes by following two set of pruning rules.	✓	X
JPS (B)	jps2	In order to boost the original JPS's performance, it uses block-based jumping	✓	X
JPS+	jps+	This algorithm differs from original JPS as it adds pre-processing to the search method	✓	X
JPS+ (P)	jps2+	This algorithm similar to JPS+ which use pre-processing and also enhance pruning rules	✓	X

The computer has a minimum specification of 3.40GHz Intel Core i7-6700 processor with 4MB of RAM and 8MB of L3 cache. The first test data consists of four benchmark problem sets which were generated by [22]. Maps are taken from commercial games that are standards for 2D benchmarking. Only four maps and problem sets were selected to be included in the study. A description of the four selected benchmark files is presented in Table III.

From Table III, it is apparent that this selection of benchmark file was made to represent different dimension and problem sets complexity. For example, arena file represents smallest dimension which is 49x49 while orz700d contain 3880 problem sets with 1260x1104 dimension.

The comparison analysis relies on a five-evaluation metrics; the descriptions of these metrics can be found in Table IV.

The evaluation metrics used in this analysis are similar to those in [6] as shown in Table IV.

TABLE III. DESCRIPTION OF BENCHMARK FILES USED IN THIS STUDY

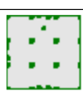
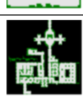
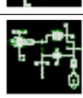
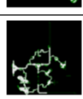
Map's name	***Preview***	***Dimension***	***Number of problem set***	***Maximum length in scenario***
arena		49 x 49	130	51.84
den501d		338 x 320	1170	466.90
brc202d		481 x 530	2550	1019.05
orz700d		1260 x 1104	3880	1551.98

TABLE IV. EVALUATION METRICS USED IN THIS STUDY

Metrics	***Description***
Total of expanded nodes	Since time is usually a hardware and software dependent factor, it is recommended to also determine the total number of expanded nodes in the search.
Total of generated nodes	An indicator of how many nodes are added to an open list after they are generated.
Total of touched nodes	Number of nodes undergoing evaluation, which may result in an update to priority queue.
Search time	Search time is measured in microseconds (wall clock time) which represents how long it takes the algorithm to find solution to the problem. Milliseconds are equal to microseconds multiplied by 1,000.
Memory cost	Metric that measures how much memory in bytes the algorithm uses in finding the end result.

IV. EXPERIMENTAL RESULT ON ANALYSIS OF PATHFINDING ALGORITHMS

Using the experimental procedure described by [23], all experiments were repeated 10 times and the average results calculated for each evaluation metric. A complete disconnect from the internet is kept throughout the experiment to prevent any disruptions caused by unwanted activity. Based on the five-evaluation metrics described previously, the following section discusses the results of the analysis. Results for the expanded nodes (average) are depicted in Table V.

To measure algorithm performance during pathfinding, the average of expanded nodes is an important criterion. In general, as the number of nodes grows, it takes longer time to find the path. Based on Table V, out of five algorithms, there is a tie between JPS (B) and JPS+ (P), whereas JPS+ (P) wins the remaining three benchmark files. Table IV and Table VII compare the result in term of generated nodes and touched nodes, respectively.

TABLE V. COMPARISON RESULT IN TERMS OF EXPANDED NODES

Algorithm	**Expanded nodes (average)**			
	arena	***den501d***	***brc202d***	***orz700d***
A*	31.42	5862.18	15997.63	29104.21
JPS	4.87	196.80	578.16	919.00
JPS (B)	**1.55**	113.43	342.81	479.53
JPS+	3.90	195.86	577.22	918.09
JPS+(P)	**1.55**	**113.43**	**342.81**	**479.53**

TABLE VI. COMPARISON RESULT IN TERMS OF GENERATED NODES

Algorithm	**Generated nodes (average)**			
	arena	***den501d***	***brc202d***	***orz700d***
A*	99.05	6100.47	16294.91	29497.70
JPS	11.11	224.03	606.61	951.14
JPS (B)	10.94	**151.64**	**386.26**	**527.28**
JPS+	**9.46**	222.27	604.85	949.36
JPS+(P)	11.05	151.80	386.45	527.44

TABLE VII. COMPARISON RESULT IN TERM OF TOUCHED NODES

Algorithm	Touched Nodes (Average)			
	arena	***den501d***	***brc202d***	***orz700d***
A*	274.07	48830.13	133876.84	247457.73
JPS	17.05	465.26	1336.50	2177.84
JPS (B)	**13.69**	**334.51**	**1022.47**	**1565.88**
JPS+	14.45	462.73	1333.97	2175.27
JPS+(P)	13.92	335.51	1023.80	1567.10

In Table VI, the average generated nodes that are injected into the open list show JPS+ to be the winner for the arena file, while JPS (B) is the winner for the rest of the files. In Table VII, in term touched nodes, JPS (B) is the most excellent for all four benchmarks file as it exhibits the smallest number of touched nodes in average. In terms of search time, the obtained result is presented in Table VIII.

In Table VIII, in terms of search time in microseconds, JPS+ is the winner for arena, while JPS+ (P) for the rest of three benchmark files. This situation is because the JPS+ and JPS+ (P) is based on pre-processing enhancement, thus, it will speed up the search time.

For the final evaluation metrics, the algorithms were compared in terms of memory in bytes. The comparison result is depicted in Table IX.

Based on Table IX, the lowest memory consumption for arena file is A*, while JPS (B) is the top algorithm for the rest of the files.

As a continuation of the analysis described previously, this study includes further calculations to find the potential and superior algorithm for further code optimization. To identify the superior algorithm among five previously tested algorithms, a scoring method was used.

TABLE VIII. COMPARISON RESULT IN TERMS OF SEARCH TIME

Algorithm	Search Time (microseconds)			
	arena	***den501d***	***brc202d***	***orz700d***
A*	18.86	3510.18	12111.38	24633.38
JPS	7.01	86.00	238.84	414.97
JPS (B)	3.36	69.91	206.85	319.55
JPS+	**2.37**	73.52	220.77	332.63
JPS+(P)	2.57	**57.88**	**171.90**	**268.09**

TABLE IX. COMPARISON RESULT IN TERMS OF MEMORY

Algorithm	Memory (bytes)			
	arena	***den501d***	***brc202d***	***orz700d***
A*	6182968	6551496	7215324	8400304
JPS	6183768	6301944	6721476	7513048
JPS (B)	**6183728**	**6301904**	**6721436**	**7248744**
JPS+	6239296	8224656	10925668	30021816
JPS+(P)	6239256	9017384	12775476	40856424

TABLE X. OVERALL SCORE FOR THE SUPERIOR ALGORITHM

Algorithm	The Superior Algorithm (Overall)				
	arena	***den501d***	***brc202d***	***orz700d***	***Total score***
A*	1	0	0	0	1
JPS	0	0	0	0	0
JPS (B)	**0.5**	**3**	**3**	**3**	**9.5**
JPS+	2	0	0	0	2
JPS+(P)	2	2	2	2	8

For this experiment, each winner algorithm receives a score of 1 and a score of 0.5 for a tie situation. Scoring calculation has been summarized and presented in Table X.

Table X shows that according to five evaluation metrics as explained previously, JPS (B) is the superior and potential algorithm among the five algorithms tested in the given source code. Thus, JPS(B) is the chosen algorithm for further testing and code optimization which will be explained in the following section.

V. JPS(B) CODE OPTIMIZATION

For the implementation of JPS(B), the use of vector is widely utilized in the program's code as highlighted in Fig. 1.

As depicted in Fig. 1, JPS (B) in its implementation spends numerous of its time accessing the vector, thus, impeding, and slow down the overall process to search the path. For example, there are 130 scenarios in arena.scen benchmarking file, and it requires almost 1448 method invocation by the syntax "jp_ids_at(i)" in the loop. Every time the program runs the syntax "at(i)" to access the vector's element, compiler will do a range checking. When the program trying to access an element that does not exist in the vector, it throws an exception. Errors will be more easily found using this checking procedure through the syntax "at(i)". However, as mentioned earlier, the syntax "at(i)" will cause overhead to the overall program in terms of search time.

In this code optimization for JPS(B), the original implementation access values from the "std::vect" class by using syntax "at(i)". The implementation is changed from syntax "at(i)" to syntax "[i]" to access the vector as highlighted in Fig. 2.

The two differ significantly: "at()" checks boundaries while operator "[]" does not. Thus, this will result in a reduction of overhead for the program's code since the syntax for accessing the vector has been changed.

```
uint32_t searchid = problem->get_searchid();
uint32_t id_mask = (1 << 24)-1;
for(uint32_t i = 0; i < jp_ids_.size(); i++)
{
    // bits 0-23 store the id of the jump point
    // bits 24-31 store the direction to the parent
    uint32_t jp_id = jp_ids_.at(i);
    warthog::jps::direction pdir = (warthog::jps::direction)*(((uint8_t*)(&jp_id))+3);

    warthog::search_node* mynode = nodepool_->generate(jp_id & id_mask);
    neighbours_.push_back(mynode);
    if(mynode->get_searchid() != searchid) { mynode->reset(searchid); }

    if((current->get_g() + costs_.at(i)) < mynode->get_g())
    {
        mynode->set_pdir(pdir);
    }
}
```

Fig. 1. Snippet for JPS(B) source file.

```
uint32_t searchid = problem->get_searchid();
uint32_t id_mask = (1 << 24)-1;
for(uint32_t i = 0; i < jp_ids_.size(); i++)
{

    uint32_t jp_id = jp_ids_[i];
    warthog::jps::direction pdir = (warthog::jps::direction)*(((uint8_t*)(&jp_id))+3);
    warthog::search_node* mynode = nodepool_->generate(jp_id & id_mask);
    neighbours_.push_back(mynode);
    if(mynode->get_searchid() != searchid) { mynode->reset(searchid); }


    if((current->get_g() + costs_[i]) < mynode->get_g())
    {
        mynode->set_pdir(pdir);
    }
}
```

Fig. 2. Snippet for JPS (BCO) source file.

VI. Result and Discussion for Code Optimization of JPS (BCO)

Based on the results of previous preliminary tests, it is necessary to gain a good understanding of the JPS (B) algorithm in order to optimize it.

In the given source code implementation, JPS (B) were noticed to make good use of vectors and it can be manipulated to take advantage to shorten the search time. In JPS (B) source code, the way vectors are accessed has been modified, and the optimized code version is named JPS (BCO). The following Table XI is the comparison between JPS (B) and JPS (BCO) in terms of search time. The same four benchmark files were selected as used in previous preliminary testing. The experiment is also repeated 10 times and the average is calculated as shown in Table XI.

Based on Table XI, for the comparison, in terms of expanded, generated, and touched nodes, there are no differences between the JPS (B) and JPS (BCO). This is also the same with memory consumption. It is because the changes made to the vector did not affect any nodes. However, the result was placed in the same table as search time, to demonstrate the differences of overall performance.

TABLE XI. Comparison of JPS (B) and JPS (BCO)

Evaluation metrics	*Algorithm*	*arena*	*den501d*	*brc202d*	*orz700d*
Expanded nodes	JPS (B)	1.554	113.429	342.806	479.533
	JPS (BCO)	1.554	113.429	342.806	479.533
Generated nodes	JPS (B)	10.938	151.638	386.259	527.278
	JPS (BCO)	10.938	151.638	386.259	527.278
Touched nodes	JPS (B)	13.692	334.507	1022.475	1565.875
	JPS (BCO)	13.692	334.507	1022.475	1565.875
Search time (micro seconds)	**JPS (B)**	**4.987**	**98.240**	**273.451**	**430.474**
	JPS (BCO)	**4.021**	**90.212**	**264.906**	**416.608**
Memory (bytes)	JPS (B)	6185888	6304568	6719204	7245720
	JPS (BCO)	6185888	6304568	6719204	7245720

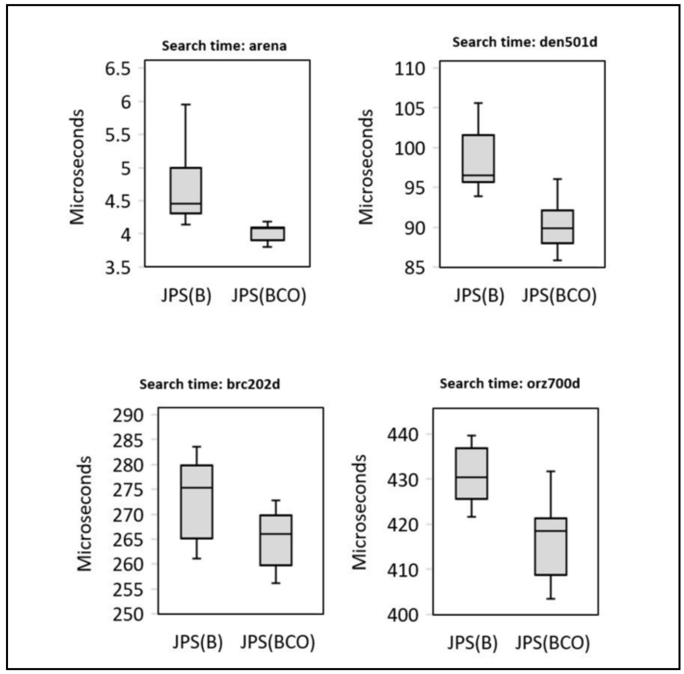


Fig. 3. Boxplot of result of JPS (B) and JPS (BCO).

In terms of search time, the average result is shown in microseconds which is much smaller unit than second and milliseconds. For arena file, the average search time using JPS (BCO) is only a microsecond shorter than JPS(BCO) performance. For the rest of the three benchmark files, each represents 8.0, 8.5 and 13.9 microseconds of faster acceleration in terms of search time. In order to provide a clear understanding of the code optimization results, a comparison of JPS(B) and JPS(BCO) search times is shown in Fig. 3.

From Fig. 3, in this boxplot, average time for JPS(B) take longer time in microseconds than JPS(BCO). In this modification of vector access's syntax, switching from "at()" to operator "[]" was a straightforward fix, but it resulted in a speed boost for the average of search time.

VII. Conclusion

Several pathfinding algorithms were compared including the well-known and traditional algorithm A*. The preliminary comparisons identify JPS(B) as a potential and the superior algorithms among the five tested algorithm. Thus, this paper performs a code optimization to the JPS(B) and called it as JPS(BCO). JPS(BCO) enhance JPS(B) performance by slightly shorten the search time.

In future, several modifications to the code implementation need to be studied in order to improve and enhance the performance of the JPS(B) significantly. The optimization should aim to improve not only the search time, but other evaluation metric such as total expanded nodes. Generally, it is not necessary to record time across platforms or machines for total expanded nodes to prove their performance, since the average is constant regardless of the platform.

In conclusion, this study has provided insight into potential and superior algorithms among JPS family members. As a result, this proposed work presents a comparison of existing algorithms. Apart from that, this study improved the JPS(B) search time through code optimization. The direction of future research can be explored in other areas, such as SAR and UAV navigation for pathfinding optimization.

ACKNOWLEDGMENT

This research was conducted during the main author's study leave under the 2019 Academic Training Scheme for Bumiputera (SLAB) scholarship scheme. A special thanks is extended to the Ministry of Education Malaysia and Universiti Teknologi MARA for the generous funding and support. It is with great appreciation that this research is being conducted with the assistance of postgraduate students from the Media and Game Innovation Centre of Excellence (MaGICX) at Universiti Teknologi Malaysia (UTM).

REFERENCES

[1] N. Nor, M. Sunar, and A. Kapi, "A review of gamification in virtual reality (VR) sport," EAI Endorsed Trans. Creat. Technol., vol. 6, no. 21, 2019.

[2] A. Y. Kapi, M. S. Sunar, and M. N. Zamri, "A review on informed search algorithms for video games pathfinding," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 3, pp. 2756–2764, 2020, doi: 10.30534/ijatcse/2020/42932020.

[3] A. Y. Kapi, M. S. Sunar, and Z. A. Algfoor, "Summary of Pathfinding in Off-Road Environment," in 2020 6th International Conference on Interactive Digital Media (ICIDM), 2020, pp. 1–4, doi: 10.1109/ICIDM51048.2020.9339639.

[4] Z. Abd Algfoor, M. S. Sunar, and H. Kolivand, "A comprehensive study on pathfinding techniques for robotics and video games," Int. J. Comput. Games Technol., vol. 2015, 2015, doi: 10.1155/2015/736138.

[5] S. Rabin and N. R. Sturtevant, "Choosing a Search Space Representation," Game AI Pro 360, vol. 1, no. c, pp. 13–18, 2019, doi: 10.1201/9780429055096-2.

[6] D. Harabor and A. Grastien, "Improving jump point search," in Proceedings of the International Conference on Automated Planning and Scheduling, 2014, vol. 24, pp. 128–135.

[7] A. N. Sabri, N. Haizan, M. Radzi, and H. Hassan, "The State of Art Heuristic Pathfinding in Games," vol. 24, no. 2, pp. 1273–1278, 2018, doi: 10.1166/asl.2018.10731.

[8] X. Wei and D. Lu, "A Comprehensive Study on Pathfinding Algorithm for Static 2D Square Grid," Proc. - 2022 2nd Asia Conf. Inf. Eng. ACIE 2022, pp. 77–80, 2022, doi: 10.1109/ACIE55485.2022.00024.

[9] A. Primanita, R. Effendi, and W. Hidayat, "Comparison of A∗ and Iterative Deepening A∗ algorithms for non-player character in Role Playing Game," ICECOS 2017 - Proceeding 2017 Int. Conf. Electr. Eng. Comput. Sci. Sustain. Cult. Herit. Towar. Smart Environ. Better Futur., pp. 202–205, 2017, doi: 10.1109/ICECOS.2017.8167134.

[10] E. R. Firmansyah, S. U. Masruroh, and F. Fahrianto, "Comparative analysis Of A∗ and basic theta∗ algorithm in android-based pathfmding games," Proc. - 6th Int. Conf. Inf. Commun. Technol. Muslim World, ICT4M 2016, pp. 275–280, 2017, doi: 10.1109/ICT4M.2016.56.

[11] A. Rafiq, T. A. A. Kadir, and S. N. Ihsan, "Pathfinding Algorithms in game development," in IOP Conference Series Materials Science and Engineering, 2020, vol. 769, no. 1, p. 12021.

[12] A. Goyal, P. Mogha, R. Luthra, and N. Sangwan, "Path finding: A* or dijkstra's?," Int. J. IT Eng., vol. 2, no. 1, pp. 1–15, 2014.

[13] Z. Wu, Z. Meng, W. Zhao, and Z. Wu, "Fast-RRT: A RRT-Based Optimal Path Finding Method," Appl. Sci., vol. 11, no. 24, p. 11777, 2021.

[14] J.-Y. Wang and Y.-B. Lin, "Game AI: Simulating Car Racing Game by Applying Pathfinding Algorithms," Int. J. Mach. Learn. Comput., vol. 2, no. 1, pp. 13–18, 2012, doi: 10.7763/ijmlc.2012.v2.82.

[15] D. Harabor and A. Grastien, "Online graph pruning for pathfinding on grid maps," Proc. Natl. Conf. Artif. Intell., vol. 2, pp. 1114–1119, 2011.

[16] Y. Zhang and H. Huang, "Multi-AGVs Pathfinding Based on Improved Jump Point Search in Logistic Center," in Algorithmic Aspects in Information and Management, 2020, pp. 358–368.

[17] X. Zheng, X. Tu, and Q. Yang, "Improved JPS Algorithm Using New Jump Point for Path Planning of Mobile Robot," in 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Aug. 2019, pp. 2463–2468, doi: 10.1109/ICMA.2019.8816410.

[18] L. Ma, X. Gao, Y. Fu, and D. Ma, "An Improved Jump Point Search Algorithm for Home Service Robot Path Planning," Proc. 31st Chinese Control Decis. Conf. CCDC 2019, pp. 2477–2482, 2019, doi: 10.1109/CCDC.2019.8833422.

[19] J. Wang and J. Jiang, "Jump point search plus algorithm based on radar simulation target path planning," Proc. - 2017 Int. Conf. Comput. Technol. Electron. Commun. ICCTEC 2017, pp. 480–483, 2017, doi: 10.1109/ICCTEC.2017.00110.

[20] D. Aversa, S. Sardina, and S. Vassos, "Path planning with Inventory-driven Jump-Point-Search," no. September, 2016, [Online]. Available: http://arxiv.org/abs/1607.00715.

[21] Z. A. Algfoor, M. S. Sunar, and A. Abdullah, "A new weighted pathfinding algorithms to reduce the search time on grid maps," Expert Syst. Appl., vol. 71, pp. 319–331, 2017, doi: 10.1016/j.eswa.2016.12.003.

[22] N. R. Sturtevant, "Benchmarks for grid-based pathfinding," IEEE Trans. Comput. Intell. AI Games, vol. 4, no. 2, pp. 144–148, 2012.

[23] C. McMillan, E. Hart, and K. Chalmers, "Collaborative diffusion on the gpu for path-finding in games," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9028, pp. 418–429, 2015, doi: 10.1007/978-3-319-16549-3_34.

Integrating User Reviews and Issue Reports of Mobile Apps for Change Requests Detection

Laila Al-Safoury, Akram Salah, Soha Makady
Department of Computer Science-Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt

***Abstract*—There is abundance of mobile Apps released continuously on the App store, where developers are required to maintain these Apps to attain user satisfaction. Developers should consider all user feedback, as they are important resources for planning of next App's release. In order to consider user feedback, many platforms host mobile Apps and allow users to submit their opinions, such as: Google Play App store and Github Open-Source Development platform. The automated consolidation of user feedback from such platforms, and transforming it into a list of change requests would result in satisfying users across different platforms, and their analysis helps developer to reduce cost of time and effort to plan for the new release of the mobile App. In this paper, a framework is proposed which integrates user feedback from different sources and analyzes them using a state-of-art user reviews analysis tool to obtain a list of change requests, such list is further examined for similarity to remove duplicates and prioritize the identified change requests. A prototype is designed to implement the proposed framework and applied to AntennaPod. Consequently, the framework experimentation results show that reviews and issue reports can be analyzed almost equally despite the difference of text's nature.**

Keywords*—*User review; feedback analysis; mobile app maintenance; text similarity

I. Introduction

The mobile App stores play an important role in distributing software products from different domains. In 2022 according to Statista website[1], Google Play store offers 3.3 million Apps for Android, while App store roughly includes 2.11 million Apps for iOS. The number of Apps increases over the years as they are widely discovered, purchased, and updated through the mobile App stores (*e.g.*, Google Play store and App store). Recently, researchers have studied the effect of App stores on software engineering practices [1, 2], while others have analyzed the benefit of using user reviews for software engineering [3, 4].

One of the most essential resources for the requirements elicitation activity is user reviews [5], which is offered by App stores allowing users to evaluate the downloaded Apps and to express their opinions [6]. App reviews are textual feedback associated with a star rating that indicates user satisfaction from one to five, where one is the lowest rate and five is the highest rate. However, the analysis of user reviews manually for extracting user needs is a challenging and time-consuming task [7]. As stated by Pagano et al. [8] 23 reviews per day are submitted in non-popular mobile Apps approximately and on average 4,275 reviews per day are received in popular Apps, such as Facebook. Besides, the feedback is usually written as unstructured text which cannot be parsed and analyzed easily.

Automated approaches are required to handle the consolidation of large amount of reviews and to perform review analysis tasks, such as: classifying feedback into maintenance tasks [9] or classifying feedback based on predefined topics [10] or based on user intention [11]. For example, reviews had been classified by Sorbo et al. into four intention categories based on user's intention while expressing their opinion such as information giving (opinion), information seeking (question), feature request (improvement or new feature), and problem discovery (bug report) [11].

In a recent review analysis survey [4], it had been raised that App store reviews can be integrated with other feedbacks available for developers to attain users' needs from more than one source, such as: Github [12, 13] and tweets [14, 15]. Since a mobile App can be hosted across platforms, such integration would demand a lot of manual effort from the App developer a lot of time and effort from App developer. The manual effort entails filtering manually such a large number of users' feedbacks from more than one platform to produce a list of change request which might still include duplicate change requests from different platforms.

In this paper, Google Play App store[2] reviews and Github[3] issue reports are integrated. GitHub is a leading open-source software development platform worldwide, it has more than one million android open-source repositories, four million issue reports, and 69K contributors in 2022[4]. Accordingly, Github is a good candidate as an additional source for integration since there are a reasonable number of mobile Apps exists on both platforms. The proposed framework aims at combining Google Play reviews and Github issue reports for a certain App's version release to obtain a list of change requests which includes bugs and feature requests submitted by users from both platforms and to group similar change requests for producing a list of unique change request. This list guides App's developers in planning for next release to meet further users' satisfaction, while eliminating the manual needed developer effort to analyze each platform's feedbacks separately, and the possible error proneness of such process.

This paper is organized as follows: Section II presents a brief background overview and motivating example which

[1] https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

[2] https://play.google.com/
[3] https://github.com/
[4] https://github.com/search?q=Android

demonstrate the importance of the proposed framework. Section III, discusses the related work of issue reports and user reviews integration and compares between the previous work and the approach proposed. Section IV introduces the proposed framework used for integration users' feedback to obtain a unique list of informative user feedback of two intentional category types. Section V describes a case study for applying the framework's phases using a dataset and it also discusses the evaluation metrics used for evaluating the framework's results. Section VI presents the results of the case study and its discussion. Finally, Section VII concludes the study and explains the future avenues.

II. Background and Motivation Example

In this section, the Google Play store's reviews and Github's issue reports are introduced to illustrate their attributes, also the importance of users' feedbacks integration across different platforms is explained further.

Fig. 1. Example of Google app review.

The App reviews are considered as one of the main framework inputs, therefore, it is important to understand the metadata associated per user review. As illustrated in Fig. 1, an example is provided by Google Play App store where a user named "Jon Dean" submitted a review on 12 May 2022, rated the App by "4 stars", and stated a bug experienced while using Antenna pod App version 2.5.2 (i.e., a podcast manager and player App). As shown in Fig. 1, each review has a date, user name, title, body, rate, number of likes, and a reply.

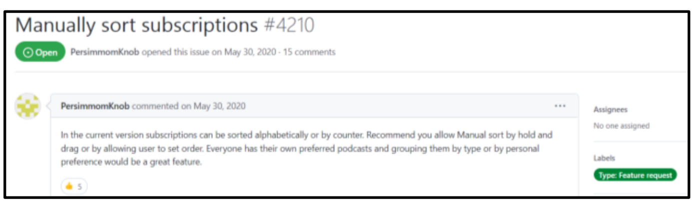


Fig. 2. Example of Github issue report.

In Github platform, for each App's release there are a set of issues submitted by users to be received by App developers to inform them whether there is a bug or a feature request to be considered in the following release and sometimes users might ask a question, when the issue is executed, it became a closed issue else it remains as an open issue. In Fig. 2, a user named "PersimmomKnob" submitted issue #4210 in Github platform, the user expressed a desired feature request regard sorting subscription feature, and it was labeled as feature request (not all issues are labeled). As shown in Fig. 2, each issue has a date, an id., open or closed issue label, user name, title, body, label (type of issue), number of likes, comments, and Assignees (i.e., members who works on this issue to be solved).

TABLE I. AntennaPod Mobile App Issue Reports and Review Examples

Feedback Type	Description
Review	Downloads don't show in the downloads folder. _Podcasts dont seem to download automatically._ I like the design though...
Issue Report	When I add the Feed http://podcast.hr2.de/derTag/podcast.xml it shows all files/episodes as 0 Byte. _They are not automatically downloaded._ I can download manually and listen to every episode. Every time the feed updates the first entry vanishes (even when downloaded) until everything is gone. Recent version (0.9.8.0) from play store on Android 4.1.1.
Issue Report	When the automatic download updates the feeds and a new episode is found, that _episode is not downloaded immediately._ Instead AntennaPod will wait for your custom set update interval (2h in my case). Leaving the house with an undownloaded episode is a bit annoying, easpecially when the download would have only taken a few seconds on your wifi. I did not test if that issue appears when setting a high update interval, but if it does one would receive all new episodes only after a high delay. My personal fix is to regularly check for new podcasts and hitting the update button manully, but I would like AntennaPod to be purely push- and not pull-style.

In Table I, an example of AntennaPod mobile App's version 0.9.8.0 reviews from Google Play[5] and issue reports from Github[6] where users express the same problem regard not being able to download their podcast automatically (i.e., underlined sentences). In contrast to reviews, the issue reports are more detailed and all sentences are focused in describing one problem rather than stating more than one information. The problem is expressed using different keywords, such as "download automatically" is similar to "download immediately" and "Podcasts" is similar to "Episode" as a podcast consists a set of episodes. Accordingly, the automation of integrating reviews and issue reports prevents the challenges of manual integration, such as: time-consuming, error prone and redundancy. Besides, the automation is needed to (i) handle the different usage of phrases which is challenging automatically to group all such reviews and issue reports, which express a bug or a feature request, into one list, and (ii) consider the rate at which the same problem is reported from different platform as a possible priority/importance indicator for that problem.

III. Related Work

A few approaches address integrating user reviews and other resource and most of those researches aims at using one source of data to enrich the other for a specific purpose which fulfill their proposed approach's purpose. Such approaches are discussed below in order to explain their aims and the differences between the proposed framework and their work.

Zhang et al. [12] proposes an approach for tagging the unlabeled Github issue reports using labeled issue reports and user reviews. Github issue report have an associated label, which define the type of this issue report (i.e. bug or feature request), and this label are optionally added which leads into some unlabeled issue reports. On the other hand, user reviews are classified into bug and feature request using a tool then both labeled issue reports and classified reviews are used as input for calculating text similarity between unlabeled issue

[5] https://play.google.com/store/apps/details?id=de.danoeh.antennapod
[6] https://github.com/AntennaPod/AntennaPod/issues

reports and the input. On the contrary, the proposed framework focuses on integrating both reviews and issue reports for filling all issue reports' labels, which helps the developer to filter issue reports according to its label without dealing with redundant issue reports.

Zhang et al. [13] introduces an approach for grouping user reviews as a cluster which addresses bugs and feature requests of a certain mobile App, along with relevant issue reports for enriching the cluster of user feedback. This approach utilizes user feedback clusters for linking each cluster to set of relevant code classes (i.e., affected classes if this group of bugs or feature requests is implemented). The classes are obtained through calculating the text similarity between classes' names and text of reviews and issue reports within each cluster producing a ranked list of classes recommended for each cluster. The paper integrated reviews and issue reports for enriching the reviews' cluster for recommending more accurate classes, which guides the developer on having a list of classes recommended per each cluster, unlike the proposed framework, if an issue report is not similar to any review, then it will be excluded, also if there are a number of similar issue reports that are not similar to any review, they will be excluded.

Yadav et al. [15] proposes a framework to analyze the users' feedback, from Google Play store and Twitter, by embedding their semantics. The framework classifies the feedback into two classes of bug reports and feature requests then it investigates whether the approach can identify the similar feedbacks. This paper is an example for integration using Twitter, where user can express their opinions regard mobile Apps through posting a tweet with a limited 140 character. On the contrary, the proposed framework uses different type of information which Github issue report where users are able to write in more details as shown in Table I and they are not limited to express their opinions briefly using limited short text.

IV. Proposed Framework

In this framework, as shown in Fig. 3, the Google Play App store is used as a source of reviews while Github is used as a source of issue reports where both reviews and issue reports form feedback inputs. There are three main phases which process the input to produce a unique change request list, which are explained as follows:

1) Feedback pre-processing: Each feedback has a text which is divided into sentences and pre-processed by applying Snowball Stemming [16].

2) Feedback classification: Each sentence is parsed using Stanford Typed Dependencies (STD) [17] then mapped to a set of 246 natural language processing (NLP) heuristics[7] to obtain the sentence's structural patterns associated with one of user's categories. Additionally, each sentence is analyzed to get sentiment annotator using Stanford CoreNLP [18] for improving the accuracy of intent classification. Both structural patterns and sentiment features are used as an input for a pre-trained Machine Learning [19, 20]. The output of this phase is a list of change requests classified as bug or feature request.

3) Text similarity calculation: According to previous work [12], cosine similarity measure and BM2F model were used for calculating textual similarity between issue reports, while Jaccard similarity is used for calculating textual similarity between issue reports and reviews because it performs better when data is sparse. The output of this phase is a unique change requests list where similar change requests are grouped.

For applying phase 1 and phase 2, there are a set **of** publicly-available App review mining tools according to the recently published survey [4]**.** SURF (Summarizer of User Reviews Feedback) was used in other research papers and they showed promising results [12, 13]**.** For input processing, SURF is used as a tool which summaries the feedback written by users to assist developers in figuring out user needs and dealing with an abundance number of reviews. The tool works as follows: (i) classifies user's intention determining the type maintenance task required to fulfill user's needs, (ii) gathers sentences of the same topic, (iii) receives user feedback in XML format and also generates the output in XML format which allows integrating them in third party frameworks, and (iv) produces a visualized report for browsing the summaries easily.

Moreover, SURF is constructed based on User Reviews Model (URM) [21, 19] which categorizes each sentence contained in App reviews into two dimensions: (i) the user intention: It is user's goals when writing a review (such as: Information giving, Information seeking, Feature request, Problem discovery, or other), and (ii) the review topics: It finds the most relevant topic(s) belong to this review (such as: App, GUI, Contents, Pricing, Feature or Functionality, Improvement, Updates/Versions, Resources, Security, Download, Model, or Company).

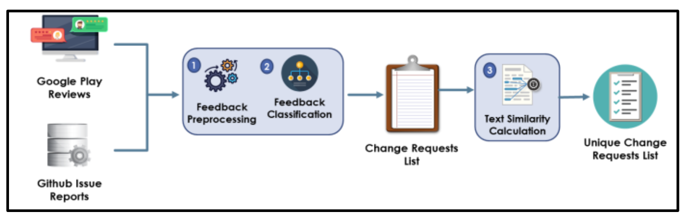


Fig. 3. Proposed framework phases.

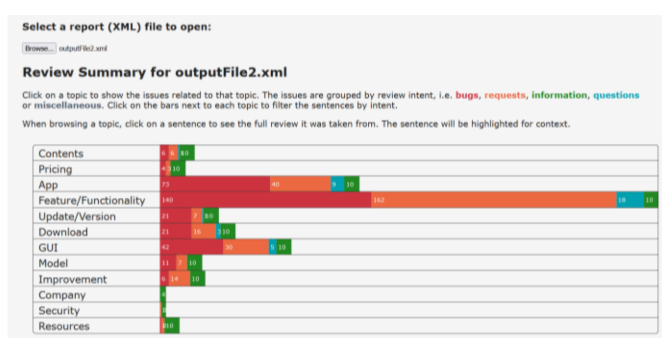


Fig. 4. Example of output report.

[7] https://www.ifi.uzh.ch/seal/people/panichella/Appendix.pdf

Not all of the change request sentences will appear on the output report in case it did not match an intention type or one of the predefined topics. Moreover, a sentence can be dismissed from the output report due to the sentence scoring which is calculated to select 2/3 of total sentences having the highest scores will be included on the report. The rewarding factors of sentence scoring are: (i) sentences classified as feature requests or bug reports, (ii) sentences related to specific topics, (iii) longer sentences, and (iv) sentences containing frequently discussed features. Afterwards, the output report is obtained in XML format which can be easily browsed through visualized utility supported in the SURF replication package[8]. As illustrated in Fig. 4, the topics are listed in rows where each topic has four types of intention categories colored differently, each type has number of sentences extracted from xml input file, and the sentences appear when you click on the number.

The SURF tool produces XML file report which includes each feedback's sentence and its user intention category. In this paper, the sentences are considered as change request, if it is classified as bug or feature request, extracted to form a list of change requests, then the text similarity is applied for reducing the list size by grouping similar sentences into one change request.

V. Framework Casestudy and Evaluation

A case study is applied on the proposed framework using an open-source dataset for implementing and evaluating the phases of the framework. In this section, the details of such case study (section 5.1) and evaluation (section 5.2) are presented to address two research questions: (RQ1) 'Can the framework classify the users' reviews and issues to obtain a list of change requests?'; (RQ2) 'To what extend the framework is capable of classifying all the change requests into bug or feature request accurately?'; (RQ3) 'Does the framework similarity percentage is reasonable and worth to be a part of the proposed framework?'.

A. Experimental Case Study

In phase 1, the dataset used for the experimentation had been publicly shared by [12], it includes top 17 popular android open-source mobile Apps which are available in both Google Play store and Github. The user reviews and the issue reports were collected for each App during a specific period which is different from App to another, where the dataset includes 20,135 issue reports and 43,649 reviews. The framework is applied on "AntennaPod" App which includes 1108 Github issue reports and 2082 Google Play reviews, such data was collected from 3 August 2012 till 9 January 2018. Particularly, App version 0.9.8.0 was selected as input which is gathered from 19 October 2013 to 26 February 2014 (i.e. before version 0.9.8.1 publishing date on 27 Feb 2014) and it has 43 issue reports and 50 reviews.

Additionally, the date attribute in issue xml file had been added to the dataset for obtaining the version of App in which the issue was submitted.

[8] https://github.com/panichella/SURF/tree/SURF-v.1.0

```
<review id="1901">
    <app_version/>
    <user>Drew Miller</user>
    <date>October 20, 2013</date>
    <star_rating>5</star_rating>
    <review_title/>
    <review_text>Best android podcast client</review_text>
</review>
```

```
<Manager Number="211">
    <labels>
        <label>wontfix</label>
    </labels>
    <bugID>Issue #302</bugID>
    <url>
    https://github.com/AntennaPod/AntennaPod/issues/302
    </url>
    <description>
    Downloading of audio file failed if audio URL contains
    spaces. Streaming is fine.
    </description>
    <title>Failed Download</title>
    <date>Oct 26, 2013</date>
</Manager>
```

Fig. 5. Example of dataset review and issue XML files.

As shown in Fig. 5, the dataset stores reviews and issue in form of XML format, where a set of attributes are defined for each review and issue report. In reviews, the xml file contains user name, date, user rating, review title, and review text (body), on the other hand, the issue reports xml file contains label, issue id., issue URL, description (body), title, and date.

In phase 2, the SURF tool requires the input in form of XML file, where attributes should be written in a specific format (as shown in Fig. 6). Accordingly, the attributes of the dataset xml files has to be converted to fit SURF tool input prerequisite specification in attributes naming (e.g. the review title attribute is named as <title> in issue report xml file, it has to be changed into <review_title> for all stored issue reports to fit SURF input file). For all reviews and Github issue reports of a version App, the naming of attributes are changed to match SURF input file, all non-English feedback are removed because the tool only process English text, then both feedbacks are gathered in one file for to be processed by the tool.

```
<reviews>
  <review>
      <date>2014-12-04</date>
      <star_rating>1</star_rating>
      <user>Jmbrowne</user>
      <app_version>2.0.1</app_version>
      <review_title>Painfully slow</review_title>
      <review_text>
           This puzzle would be enjoyable if it was not
           so painfully slow.
      </review_text>
   </review>
   . . .
</reviews>
```

Fig. 6. SURF tool input XML file format [11].

TABLE II. SURF Classified Sentences for Dataset

Source of Classified Sentence	Number of Sentences
Total Sentences (S)	70
Feature Requests in Reviews (R_F)	9 (7)
Bugs in Reviews (R_B)	4 (4)
Feature Requests in Issues (I_F)	19 (14)
Bugs in Issues (I_B)	10 (9)

In Table II, when SURF tool is used to process the input only 70 sentences were classified while the rest did not appear in the output XML report. The SURF tool does not provide a sentence in output report unless it is able to classify its intention type (i.e., according to "predefined patterns" matching) and also if it is able to classify its topic (i.e., based on "topic words dictionary" matching). Additionally, sentence scoring results is reduced by 30% of total sentences according to four factors which are previously mentioned. As shown in Table II, the number of sentences per each type is mentioned and the number of correctly classified types is written between brackets. For example, the tool classified nine review's sentences as feature requests and compared to the manually labeled 19 sentences mentioned in Table IV only 7 sentences were classified correctly. According to the results mentioned in Table II, RQ1 is answered:

Answer to RQ1: the framework is able to classify both reviews and issues using the same approach explained previously in Section IV.

TABLE III. CHANGE REQUESTS SIMILARITY FEATURES

Change Request	Issue	Review
Login Dark Theme	2	0
Auto Delete	1	3
Auto Download	2	1
Pause Frequently	0	2
Stop Playing	1	2
fix widget	0	3
Download Scheduler	0	2
Less Cache	0	2
skip back	1	1
Total	7	16

In phase 3, the list of change requests is studied manually to analyze the repeated requests, the results are depicted on Table III which lists the name of the addressed feature, and the number of issues or reviews mentioned this feature as a bug or a feature request having similar change request. There are nine change requests which had been expressed by more than one user whether in Github or Google Play store platforms.

B. Case Study Evaluation and Results

After executing the framework, the dataset's sentences are manually classified and labeled to be used in evaluating the accuracy of the framework's output. Thus, the input is divided into a set of sentences to be manually classified for obtaining the user intention type for each sentence to be used as a ground truth. For splitting the input into sentences, a natural language toolkit (NLTK) has been utilized in python program which offers NLP methodology implementation.

As shown in Table IV, the total number of sentences in the dataset are 252, part of these sentences belongs to issue report extracted from Github platform (135 sentences) and the other sentences belong to reviews extracted from Google Play store (117 sentences). The manually labeled sentences show that 16% of reviews' sentences are feature requests while 23% of reviews' sentences are bugs, on the other hand, 17% of issue report's sentences are feature requests while 39% of issue report's sentences. Those percentages show that, in such dataset, only half of the total sentences are useful for the developers as they state the changes requested by the App's users.

For evaluating accuracy of the change request list's output sets (i.e. list of sentences), both ground truth's four sets (MI_B, MI_F, MR_B, and MR_F) shown in Table IV and SURF tool's produced sets (I_B, I_F, R_B, and R_F) shown in Table II are used to measure recall and precision which are a well-established metrics reported in Equations 1 and 2.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

TABLE IV. MANUALLY CLASSIFIED SENTENCES FOR DATASET

Source of Sentence	Number of Sentences
Total Sentences (MS)	252
Google Play Reviews (MR)	117
Github Issue Reports (MI)	135
Feature Requests in Reviews (MR_F)	19 (16%)
Bugs in Reviews (MR_B)	27 (23%)
Feature Requests in Issues (MI_F)	23 (17%)
Bugs in Issues (MI_B)	53 (39%)

In particular, TP represents the number of true positives reviews/issue reports classified correctly; False Positive (FP) is the number of reviews/issue reports classified incorrectly by the approach; and False Negative (FN) is the number of correct reviews/issue reports (i.e., manually classified as bug or feature request) that are not classified by the approach.

For applying accuracy measurement, three variables are required to calculate precision and recall as in equation (1) and (2), shown in Table V. Also, F-measure is used as it combines both recall and precision by calculating the harmonic mean of precision and recall to obtain the overall accuracy of the proposed approach, known as the F-measure in Equation 3:

$$\text{F-measure} = 2 * \frac{\text{Precision} * \text{recall}}{\text{Precision} + \text{recall}} \tag{3}$$

After calculating accuracy measurements, it is noticed that precision of bug is better than feature request (i.e. the ability of predicting change requests correctly), as shown in Table VI and Table VII, for both inputs issue reports or reviews. On the other hand, the recall of feature request is better than bug (i.e., the ability of to find all manually labeled relevant cases) in both issue reports and reviews. The SURF tool is able to predict the intentional type as they follow predefined textual patterns, such as: bug patterns "[something] is missing" and "[someone] has an issue", but these patterns are not sufficient for covering all sentences expressing bugs and feature requests to be matched with the dataset. According to the results mentioned in Table VI and Table VII, RQ2 is answered:

TABLE V. THE CALCULATION OF ACCURACY MEASUREMENTS' VARIABLES

Accuracy Measure	TP	FN	FP
Bug Issues	Numbers of issues sentences classified as "bug" and manually labeled as "bug".	Numbers of issues sentences manually labeled as "bug" but not classified as "bug".	Numbers of issues sentences not manually labeled as "bug" but classified as "bug".
Feature Request Issues	Numbers of issues sentences classified as "feature request" and manually labeled as "feature request".	Numbers of issues sentences manually labeled as "feature request" but not classified as "feature request".	Numbers of issues sentences not manually labeled as "feature request" but classified as "feature request".
Bug Reviews	Numbers of reviews sentences classified as "bug" and manually labeled as "bug".	Numbers of reviews sentences manually labeled as "bug" but not classified as "bug".	Numbers of reviews sentences not manually labeled as "bug" but classified as "bug".
Feature Request Reviews	Numbers of reviews sentences classified as "feature request" and manually labeled as "feature request".	Numbers of reviews sentences manually labeled as "feature request" but not classified as "feature request".	Numbers of reviews sentences not manually labeled as "feature request" but classified as "feature request".

TABLE VI. CLASSIFICATION EVALUATION OF SURF RESULTS FOR ISSUE REPORTS

Change Request Type	TP	FN	FP	Precision	Recall	F-measure
Bug	9	44	1	90%	17%	29%
Feature Request	14	9	5	74%	61%	67%

TABLE VII. CLASSIFICATION EVALUATION OF SURF RESULTS FOR REVIEWS

Change Request Type	TP	FN	FP	Precision	Recall	F-measure
Bug	4	23	0	100%	15%	26%
Feature Request	7	12	2	78%	37%	50%

Answer to RQ2: The experiment's precision was better than recall which affect negatively in F-measure, which means that the tool is able to classify most of the sentences correctly but it is not able to detect all relevant sentences in the used dataset. A few of modifications are exerted in Section 6 for enhancing the results of the tool used in phase 3.

The output list is reduced using similarity calculation to reduce the total number of change requests (CR); thus, each sentence S_i text is compared to other sentences. If S_i is similar to one on more sentences, then they will be grouped as one change request, else nothing will change. The evaluation is performed by calculating the percentage of reduction using CR which is the total number of change requests' sentences before reduction (i.e., before grouping) and unique change requests (UCR) which is the total number of unique change requests' sentences, shown in Equation 4.

$$\text{Reduction Percentage}= \frac{\#UCR*100}{\#CR} \quad (4)$$

According to Table III and Equation 4, the total number of repeated change requests is 23 change requests (i.e., 7 issues and 16 reviews), the change requests list will be reduced from 252 into 238 (i.e., reduced by 6%). According to such results mentioned, RQ3 is answered:

Answer to RQ3: The framework's output of the similarity percentage is slightly reduced in phase 3 compared to the original change request list produced from phase 4, this observation is related to the nature of selected App's version dataset which needs to be investigated for other Apps' dataset.

VI. RESULTS ENHANCEMENT AND DISCUSSION

In this section, the impact of adding the title of the change request (*e.g.,* review or issue report) is investigated to the change request body as an input to phase 2. Then, the output of phase 2 is evaluated to check whether the added topic may improve the recall values. Additionally, the patterns of SURF tool were studied to examine their sufficiency for covering all types of sentences belongs to bug or feature request.

Another experimental trial has been performed for enhancing the results, which is adding the title of the change request into the beginning of the body. Because SURF tool only considers the review/issue report's body without including its title, i.e., it is a sentence-based classifier, although the title might contain valuable general information about the bug or feature request itself while the body includes extra explanatory information. Also, including the title might increase the fourth factor of sentence scoring which increase the feature discussed. In Table VIII, the evaluation results slightly changed only in review input, and it was a minor change where an extra bug was classified by SURF correctly (i.e., increasing recall and f-measure) and a feature request where misclassified (i.e., decreasing precision and f-measure).

As shown in Table IX, when the sentences are changed to follow the predefined SURF pattern, such in all examples except the third one (i.e., begin with the pattern by splitting the sentence), the SURF tool was able to classify the sentences correctly. Additionally, the position of the pattern affects the ability of classification, since the tool is not able to detect the pattern when it is embedded in the middle of a long sentence. It is noticed that papers [12, 13], explained in Section II, used SURF tool and both mentioned obtaining high overall accuracy of 91.36% and 95.64% respectively which are stated to be acceptable results. Although, this study uses the same dataset as paper [12] but the accuracy obtained was 77% as not all of the dataset was used, only one version of an App.

TABLE VIII. CLASSIFICATION EVALUATION OF SURF RESULTS FOR REVIEWS INPUT (INCLUDING TITLE)

Change Request Type	TP	FN	FP	Precision	Recall	F-measure
Bug	5	23	0	100%	19%	31%
Feature Request	7	12	2	70%	37%	48%

TABLE IX. ALTERING SENTENCES ACCORDING TO SURF PATTERNS

SURF Pattern	Original Sentence	Altered Sentence
"[something] can be fixed by [something]"	My personal fix is to regularly check for new podcasts and hitting the update button manully, but I would like AntennaPod to be purely push- and not pull-style.	This can be fixed by regularly checking for new podcasts and hitting the update button manully, but I would like AntennaPod to be purely push- and not pull-style.
"[something] should be fixed"	The widget needs fixing since it doesnt always work.	The widget should be fixed since it doesnt always work.
[someone] could/should add/provide/offer/integrate [something]	Congrats for your job in this App, its so amazing and for me almost perfect, I don't know if its hard to do or not, but if you could add the auto delete of the podcast when we finish it will be good.	Congrats for your job in this App, its so amazing and for me almost perfect, I don't know if its hard to do or not. if you could add the auto delete of the podcast when we finish it will be good. Thanks!
	Wish i could schedule when it downloads the podcast.	Wish i could add a schedule when it downloads the podcast.

VII. CONCLUSION AND FUTURE WORK

The users' contribution in expressing their opinions is considered as an important source for evolving mobile Apps and for better release planning. Mobile App can be distributed on many platforms where each platform allows users to submit their feedback for other users and App's developers to ask questions, state information, report a bug, or suggest a modification. Accordingly, the integration of more than one feedback source assists developer to satisfy large number of users which increase App's rating. Because of the large amount of feedback provided by users, an automatic tool for feedback analysis is required. In this paper, the proposed framework analyzed Google Play reviews and Github issues to filter them and to extract change requests out of the integrated feedback. After performing two trials, the accuracy results of reviews and issue reports showed slightly different values which proves that the nature of both textual feedbacks did not affect results after using the same tool. Despite the tool was able to classify the integrated inputs into bugs and feature requests, it was not able to extract all of them. Besides, the similarity can assist on reducing the number of change requests to decrease of the time needed for developer to check the list of change requests. In the future, we are interested in including more Apps feedback from dataset to be able to evaluate whether the results' insights are repeated using other versions or other Apps. Also, the similarity will be automated and examined in larger dataset versions and Apps to evaluate further the effect of reducing the list of change request. Finally, other analysis tools can be investigated for producing better results of the integrated users' feedback classification.

REFERENCES

[1] A. A. Al-Subaihin, F. Sarro, S. Black, L. Capra and M. Harman, "App store effects on software engineering practices," IEEE Transactions on Software Engineering, vol. 47, no. 2, pp. 300-319, 2019.

[2] A. A. Al-Subaihin, "Software Engineering in the Age of App Stores: Feature-Based Analyses to Guide Mobile Software Engineers," PhD diss., UCL (University College London), 2019.

[3] W. Martin, F. Sarro, Y. Jia, Y. Zhang and M. Harman, "A survey of app store analysis for software engineering," IEEE transactions on software engineering, vol. 43, no. 9, pp. 817-847, 2016.

[4] J. Dabrowski, E. Letier, A. Perini and A. Susi, "Analysing app reviews for software engineering: a systematic literature review.," Empirical Software Engineering, vol. 27, no. 2, pp. 1-63, 2022.

[5] E. A. AlOmar, W. Aljedaani, M. Tamjeed, M. W. Mkaouer and Y. N. El-Glaly, "Finding the needle in a haystack: On the automatic identification of accessibility user reviews.," in Proceedings of the 2021 CHI conference on human factors in computing systems, 2021.

[6] S. Lim, A. Henriksson and J. Zdravkov, "Data-driven requirements elicitation: A systematic literature review.," SN Computer Science, vol. 2, no. 1, pp. 1-35, 2021.

[7] Y. Zhang, J. Du, X. Ma, H. Wen and G. Fortino, "Aspect-based sentiment analysis for user reviews," Cognitive Computation, vol. 13, no. 5, pp. 1114-1127, 2021.

[8] P. Dennis and W. Maalej, "User feedback in the appstore: An empirical study," in 21st IEEE international requirements engineering conference, 2013.

[9] A. Al-Hawari, H. Najadat and R. Shatnaw, "Classification of application reviews into software maintenance tasks using data mining techniques," Software Quality Journal, vol. 29, no. 3, pp. 667-703, 2021.

[10] M. v. Vliet, E. C. Groen, F. Dalpiaz and S. Brinkk, "Identifying and classifying user requirements in online feedback via crowdsourcing," in International Working Conference on Requirements Engineering: Foundation for Software Quality, 2020.

[11] A. Di Sorbo, S. Panichella, C. V. Alexandru, C. A. Visaggio and G. Canfora, "SURF: summarizer of user reviews feedback," in 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C), 2017.

[12] T. Zhang, H. Li, Z. Xu, J. Liu, R. Huang and Y. Shen, "Labelling issue reports in mobile apps," vol. 13, no. 6, pp. 528-542, 2019.

[13] T. Zhang, J. Chen, X. Zhan, X. Luo, D. Lo and H. Jiang, "Where2Change: Change Request Localization for App Reviews," IEEE Transactions on Software Engineering, vol. 47, no. 11, pp. 2590-2616, 2019.

[14] P. R. Henao, J. Fischbach, D. Spies, F. Julian and A. Vogelsang, "Transfer Learning for Mining Feature Requests and Bug Reports from Tweets and App Store Reviews.," in 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW), 2021.

[15] A. Yadav, R. Sharma and F. H. Fard, "A semantic-based framework for analyzing app users' feedback," in IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER), 2020.

[16] M. Bounabi, K. El Moutaouakil and K. Satori, "A comparison of text classification methods using different stemming techniques.," International Journal of Computer Applications in Technology, vol. 60, no. 4, pp. 298-306, 2019.

[17] J. Kleenankandy and A. N. KA, "An enhanced Tree-LSTM architecture for sentence semantic modeling using typed dependencies.," Information Processing & Management, vol. 57, no. 6, p. 102362, 2020.

[18] C. D. Manning, M. Surdeanu, J. Bauer, J. . R. Finkel, S. Bethard and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, 2014.

[19] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora and H. C. Gall, "How can I improve my app?

Classifying user reviews for software," in Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, 2016.

[20] S. Panichella, A. . D. Sorbo, E. Guzman, C. A. Visaggio, G. Canfor and H. C. Gall, "Ardoc: App reviews development oriented classifier," in Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering, 2016.

[21] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora and H. C. Gall, "What would users change in my app? summarizing app reviews for recommending software changes," in Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, 2016.

Research on Quantitative Security Protection Technology of Distribution Automation Nodes based on Attack Tree

Yinfeng Han*, Yong Tang, Xiaoping Kang, Hao Jiang, Xiaoyang Song
Ningbo Power Supply Company, State Grid Zhejiang Electric Power Company, Zhejiang, China

***Abstract*—In order to improve the security of distribution automation system nodes and ensure the safe operation of distribution network, a quantitative security protection technology for distribution automation nodes based on attack tree is proposed. This paper analyzes the factors of node risk assessment of the distribution automation system, and through the evaluation and analysis of node vulnerabilities; it discovers the faulty nodes in the distribution automation system in advance. Based on the node vulnerability evaluation results of the distribution automation system, the risk of the distribution automation system is comprehensively evaluated using the attack tree. Establish the distribution network control model under network attack, locate the fault node, and use trusted computing technology to design trusted distribution terminals. When the amount of data is large, more effective symmetric encryption algorithm SM4 is required to achieve node security protection in the distribution network automation system. The experimental results show that the method has high fault node location accuracy, low reliability calculation time, and the distribution automation system network has certain robustness, which fully verifies the application effect of the method.**

Keywords—Attack tree; security protection; risk assessment; fault node location; trusted computing; SM4 encryption algorithm

I. Introduction

A distribution grid failure is an unpredictable situation that can cause the entire system to operate unbalanced. For various reasons, faulty nodes may appear between regions of the distribution network [1,2]. Illegal access to the power grid to steal power is very likely to cause the access point to become a faulty node, which will lead to unpredictable technical problems in the power system. In addition, due to this part of illegal access to unaccounted electricity consumption, from an economic point of view, this problem will seriously affect the income of the power company. Therefore, the research on quantitative security protection of distribution network automation system nodes is urgent [3,4].

Literature [5] proposes a network security protection method of distribution network automation system based on independent security chip, designs a dual protection scheme based on "network layer + application layer", and proposes a one-time password authentication protocol based on the combination of SM2, SM3 and SM4 national secret algorithms and message authentication codes, so as to realize bidirectional identity authentication and business data encryption between distribution master station and distribution terminal. Ensure the integrity and confidentiality of communication data, and solve the network security protection problem of distribution network automation system. Literature [6] proposes an active distribution network security protection method based on trusted computing. Firstly, the information security problems existing in the power generation, transmission, distribution and utilization of the active distribution network are analyzed. Then, referring to the design of power distribution automation system, the overall security and reliability protection scheme is proposed for the regional energy Internet. According to the scheme, a trusted chain is established to ensure the reliability of nodes, network connections and applications. Finally, an example is given to verify the scheme. SM2 and SM4 encryption algorithms are mainly used for data transmission and storage. Considering that the downlink data will have a large number of repeated cryptographic operations. Batch certification method is adopted. By using Merkle hash tree to merge multiple authentication requests, the data processing efficiency is improved. In addition, some scholars have proposed the key technologies of information security protection of active distribution network. Firstly, the characteristic model of active distribution network architecture is analyzed, and the trusted protection environment model of active distribution network is described in combination with the existing typical application mode of active distribution network security protection; Secondly, the embedded trusted distribution terminal is designed for the application scenario of the embedded distribution terminal, and the active distribution network networking model based on the IEC61850 standard is designed for the information flow interaction and data analysis between the various levels of the active distribution network; Thirdly, in view of the complementary coupling and flat management structure requirements of the active distribution network, combined with the supporting role of the active distribution network for the regional energy Internet, the prefecture-level regional energy based on multi-level energy-information routers and energy-information switches is defined. Based on the Internet system model, a trusted in-depth protection system for prefecture-level regional energy Internet security based on Trusted 2.0 technology has been constructed. Finally, combined with the advanced nature of Trusted 3.0 technology, starting from energy nodes, with trusted computing as the basic idea, autonomous controllability as the goal, and safety immunity as the characteristic, the trusted protection scheme is further optimized and adjusted, its safety protection has certain engineering application and reference value.

*Corresponding Author.

In order to improve the security of distribution automation system nodes and ensure the safe operation of distribution network, a quantitative security protection technology for distribution automation nodes based on attack tree is proposed. This paper analyzes the factors of node risk assessment of the distribution automation system, and through the evaluation and analysis of node vulnerabilities; it discovers the faulty nodes in the distribution automation system in advance. Based on the node vulnerability evaluation results of the distribution automation system, the risk of the distribution automation system is comprehensively evaluated using the attack tree. Establish the distribution network control model under network attack, locate the fault node, and use trusted computing technology to design trusted distribution terminals. When the amount of data is large, more effective symmetric encryption algorithm SM4 is required to achieve node security protection in the distribution network automation system. Finally, the experiment proves the effectiveness of the design method.

II. Quantitative Assessment of Node Risk Distribution Network Automation System

A. Risk Assessment Factors of Distribution Network Automation System Nodes

It is very important to take an effective quantitative risk assessment method to objectively and accurately assess the risk degree of the power distribution system to ensure the safe, stable and reliable operation of the power distribution system. Here, for the node risk of the distribution network automation system, the node voltage over-limit risk and the line power flow over-limit risk are mainly considered. Each risk is represented by the probability of occurrence, and the risk of the distribution system is measured by establishing a comprehensive system over-limit risk index.

1) The probability calculation of the node voltage exceeding the limit is shown in formula (1):

$$U_i = Q_r\left(U_i > U_{i\max}\right) = 1 - \left[K\left(U_i\right) - K\left(U_{i\min}\right)\right] \tag{1}$$

In the formula: U_i is the voltage amplitude of node i; $U_{i\max}$ and $U_{i\min}$ are the upper and lower limits of the node voltage; $K\left(U_i\right)$ is the probability distribution function of the voltage amplitude of node i.

2) The probability calculation of line power flow exceeding the limit is shown in formula (2):

$$Q_r\left(F_j\right) = 1 - Y\left(G_j\right) \tag{2}$$

In the formula: F_j is the conveying capacity of branch j, MVA; G_j is the rated conveying capacity of branch j, MVA; $Y\left(G_j\right)$ is the probability distribution function of the conveying capacity of branch j.

3) The output change of the distributed power generation in the distribution network may cause the node voltage and line power flow to have the risk of exceeding the limit. In order to evaluate the various kinds of over-limit risk R_f separately, a comprehensive over-limit risk index is defined here, such as formula (3) as shown:

$$R_f = \frac{1}{M}\sum_{i=1}^{M}\left(P_g - P_f\right) \tag{3}$$

In the formula: M is the number of system nodes or lines; P_g and P_f are the over-limit probabilities of node voltage or branch power flow before and after access to distributed power. This index can be used to measure the comprehensive risk of exceeding the limit of voltage and power flow of the distribution system, which is referred to here as the comprehensive risk of exceeding the limit of node voltage and the comprehensive risk of exceeding the limit of line power flow.

B. Node Vulnerability Assessment of Distribution Network Automation System

When the nodes in the distribution network automation system are disturbed or influenced by the outside world, it is easy to cause successive failures of the components in the system, resulting in large-scale power outages. Through node vulnerability assessment and analysis, the faulty nodes in the distribution network automation system are found in advance, and the improvement of these faulty nodes will reduce the possibility of accidents in the distribution network automation system.

The power distribution network can be simplified as an undirected weighted sparse graph with M nodes and L edges by the complex network definition. The traditional complex network theory has its statistical characteristics, such as characteristic path length, degree, clustering coefficient and betweenness, and can be directly used as an index for risk assessment of distribution network automation system nodes. This paper considers that the structure of distribution network and transmission network is different, so these statistical characteristics need to be improved to meet the structural characteristics of distribution network.

The traditional node degree is defined as the number of nodes connected to the node, and the average degree of the network can be obtained by averaging the degrees of all nodes. The node degree can reflect the importance of the nodes in the network, that is, the node with a larger degree is more important in the network, and the more vulnerable it is. When the distributed power source is connected, the number of system nodes is not changed. Therefore, the node degree of the node connected to the distributed power source should be increased by 1. The calculation formula is:

$$s_i = s_{idp} + 1 \tag{4}$$

In the formula: s_i is the degree of access to the distributed power node; s_{idp} is the degree of the original node of the access to the distributed power node.

Since the distribution network is mostly radial network, which is relatively sparse compared with the transmission network, it is difficult to compare the differences between nodes with the same degree only considering its own node degree. The idea of cohesion is introduced here, and the node degree is calculated considering the influence of the nodes connected with node i:

$$Ms_i = \frac{s_i}{\overline{s_i}} \times \int_i^M D_{iM} \tag{5}$$

In the formula: $\overline{s_i}$ is the average degree of nodes; D_{iM} is the set of all nodes connected to node i.

It is difficult to judge the vulnerability of a node scientifically and comprehensively only by a single node vulnerability index. The node vulnerability index proposed based on the complex network theory mainly analyzes the vulnerability of the node from the structure. The voltage over-limit risk index analyzes the node's vulnerability from the aspect of fault risk. Combining the two, it is proposed that the node's comprehensive vulnerability index can make up for the shortcomings of the two and make a comprehensive judgment.

In this paper, the structural vulnerability index obtained based on the complex network theory is given weight by using the analytic hierarchy process [7,8] (AHP). Through consulting experts, it is equally important to set the node degree and the node intermediate number, and the node active power injection power is more important. The comparison matrix can be obtained by applying the analytic hierarchy process:

$$S = \begin{bmatrix} 1 & 1 & 1/3 \\ 3 & 3 & 1 \\ 1 & 1 & 1/3 \end{bmatrix} \tag{6}$$

The matrix satisfies the requirement of consistency, and the weights of node degree, node betweenness and active power injection power are obtained as 0.2, 0.2 and 0.6, respectively. The structural vulnerability index MS_i of the distribution network node i based on the complex network theory obtained by the AHP is:

$$Mw_i = 0.2 \times Ms_i + 0.2 \times Mq_i + 0.6 \times Mr_i \tag{7}$$

Due to the differences in the units and orders of magnitude of the above indicators, they cannot be directly added. The unit and order of magnitude of the indicators should be normalized first. In this paper, the maximum value of each indicator is taken as the benchmark for normalization, and the data is normalized to [0,1] interval.

Combined with the risk theory, the structural vulnerability index of node i is multiplied by the risk value of the node as the comprehensive vulnerability index of node i, and its expression is:

$$M_i = Ms_i \times M_{w_i} \tag{8}$$

C. *Comprehensive Risk Assessment of Distribution Network Automation System*

Based on the node vulnerability assessment results of distribution network automation system, the attack tree is further used to comprehensively evaluate the risk of distribution network automation system.

1) Definition of attack tree: The attack tree model [9,10] is a method of modeling security threats to the system, which represents each attack against the system in the form of a tree structure. The root node of the tree represents a goal to be achieved by a network attack, and the leaf nodes represent possible means to achieve this attack purpose. Each path from the root node to the leaf node represents a complete attack process to achieve this attack goal. The nodes of the attack tree are divided into two types: AND (AND) nodes or AND (OR) nodes. Among them, and node means: only after all child nodes are implemented, this node can be implemented, and the task will continue to be passed up; or node means: as long as one of the child nodes has been implemented, this node can be implemented, and the task will be up first-level delivery.

2) Stage division of intrusion process: Every complete invasion process will be very complex, and any two kinds of invasion are also very different. Therefore, it is very difficult and unrealistic to model the whole intrusion process. Similarly, it is unrealistic to use a single attack tree to model all intrusions and attacks against the distribution network automation system. The constructed attack tree will be very large, which is not conducive to analysis and maintenance.

Based on the above reasons, this paper considers dividing the intrusion process into several stages, and each stage has a stage target, and models these stages respectively. By modeling in stages, the complexity of constructing the attack tree can be greatly reduced. Specifically, a complete intrusion process can be divided into the following seven stages:

a) Host survey: search for a distribution network automation system node as an attacked target.

b) Vulnerability discovery: Discover the security loopholes on the target distribution network automation system nodes.

c) Target penetration: use the security loopholes of the distribution network automation system nodes to obtain unauthorized access rights.

d) Privilege escalation: Obtain privileged privileges on the nodes of the distribution network automation system.

e) Hidden and hidden: cover up the activity track, in order to prepare for the next invasion of the distribution network automation system node.

f) Grab information: obtain and modify the data and information on the nodes of the distribution network automation system.

g) Springboard attack: Use the controlled node as a springboard to launch attacks on other nodes.

In the following, according to the stages of the attack process, each stage is regarded as the root node of an attack tree, and an attack tree is constructed for each stage of the network attack.

3) Comprehensive risk assessment: Comprehensive risk assessment is an important process of risk management, and the risk is often expressed by the product of the loss caused by the event and the probability of the event, namely:

$$R_h = L_h \times P_h \tag{9}$$

In the formula: L_h is the loss caused by the event; P_h is the probability of the event occurring.

The root node in the attack tree represents the ultimate goal of the attacker, so the ultimate goal of the risk assessment based on the attack tree is to determine the risk value of the root node of the attack tree, the attack path that affects this value, and the attack method most likely to be exploited by the attacker, so that technicians can formulate corresponding defense countermeasures according to the risk assessment results. The specific steps of the comprehensive risk assessment method for distribution network automation system based on attack tree proposed in this paper are as follows:

a) Determine the attack target and establish the attack tree model of the system.

b) Select the appropriate evaluation index, and quantify the index of the attack tree leaf node.

c) Calculate the probability P_h of leaf node (attack event) occurrence.

d) Calculate the probability P_s of attacking the root node of the tree (that is, the attacker successfully achieves the final attack target).

e) Analyze the loss L_t caused by the realization of the attack target of attacking the root node of the tree, and use formula (9) to calculate the risk value of the root node.

f) Analyze the attack sequence and calculate the probability P_{hi} of each attack sequence.

g) Judging and analyzing the attack paths and methods most likely to be exploited by attackers according to the results of comprehensive risk assessment.

III. Quantitative Security Protection of Distribution Automation Nodes

A. *Construction of Distribution Network Control Model under Network Attack*

There are two types of network attacks for multi-node systems: centralized attacks and decentralized attacks. A decentralized attack is to launch an attack on all nodes in the area, which has a wide range of influence, but requires a higher degree of control of the attacker. Concentrated attack is to launch an attack on the nodes in the area in a targeted manner. This method has a small direct impact. However, if the attack is launched against the weak nodes in the area, the target can be destroyed at a lower attack cost. Therefore, this paper mainly considers the impact of centralized attacks.

The ultimate purpose of physical attack is to destroy the power infrastructure. It is a direct effect on physical equipment. With the subjective will of the attacker, it can directly cause the abnormal operation of a large number of physical equipment, and even cause a cascading failure to cause the collapse of the power system, with great destructive power, but the concealment is weak and the attack cost is high. The network attack takes the information space as the entrance, exploits the vulnerability of the distribution network and its nodes to destroy the distribution network automation system, and finally achieves the purpose of causing large-scale power outages. Compared with physical attacks, network attacks are not easy to be detected, have a long incubation period, and have low attack costs, but they can achieve significant damage.

To sum up, network attacks can covertly realize the attacker's intention and destroy the normal operation of the distribution network automation system. In order to study the scenario where the distribution network is attacked by the network, a distribution network control model is established on the basis of the comprehensive risk assessment of the distribution network automation system. The evaluation index of the distribution network includes power supply quality, economy, safety, etc., which is called the controlled quantity in this paper, denoted by G. In general, these controlled quantities are determined by the protection action, dispatch control and user behavior of the power company, which can be described as:

$$G = R\left(e, v, z\right) \tag{10}$$

In the formula: e is protection action; v is scheduling control; z is user behavior.

Formula (10) is a nonlinear equation, and the solution of the equation is related to the input $\left(e, v, z\right)$ and the initial state of the distribution network. In the traditional distribution network, user behavior is reflected in daily life and production activities, and is a random variable that conforms to certain laws. At this time, the distribution network is mainly controlled by dispatching and protection. The dispatching system and protection device control the distribution network according to the state detection quantity to ensure that the controlled quantity G meets the requirements of the stable operation of the distribution network.

Unlike normal scheduling, protection, and user usage behavior, attack behavior is unpredictable. Therefore, malicious control behavior is introduced into the distribution network control model, which can be expressed as:

$$G = R(e, v, z, U) \tag{11}$$

In the formula: U is the aggressive behavior.

Also, the total formula (11) is a nonlinear equation. U is sent by an attacker, and the attack may cause the controlled quantity G to deviate from the safe and stable operation requirements of the distribution network, causing a safety and stability accident. Fig. 1 is a schematic diagram of the distribution network control model under network attack.

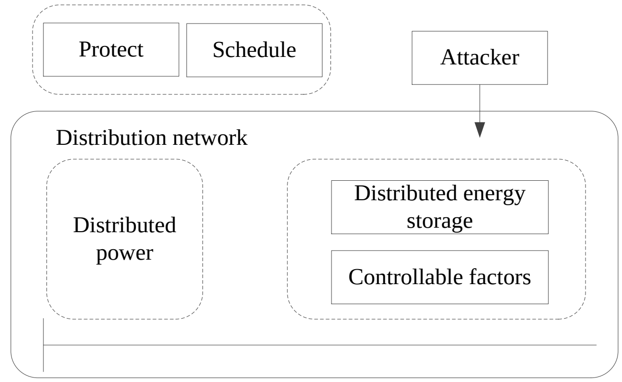


Fig. 1. Schematic diagram of distribution network control model under network attack.

B. Fault Node Location

Based on the simulation analysis of the distribution network control model under network attack, the voltage array of the distribution network nodes is obtained. According to the operation characteristics of the distribution network automation system in each iteration; the presence of fault nodes is detected. The difference between the measured voltage array and the estimated voltage array is:

$$\Delta V = V_\alpha - V_\beta \tag{12}$$

In the formula: V_α is the measured voltage array; V_β is the estimated voltage array. ΔV is zero in the absence of faulty nodes.

The area where the faulty node exists is displayed through the abnormality matrix elements, and the peak value of the elements in the faulty node area A_d shows the faulty node closest to the fault location in the distribution network automation system. The faulty node is defined as follows:

$$\forall x \in A_d : x_i \geq x_k \tag{13}$$

This method can detect the existence of the faulty node and determine the location of the faulty node. The flow chart of the positioning method is shown in Fig. 2.

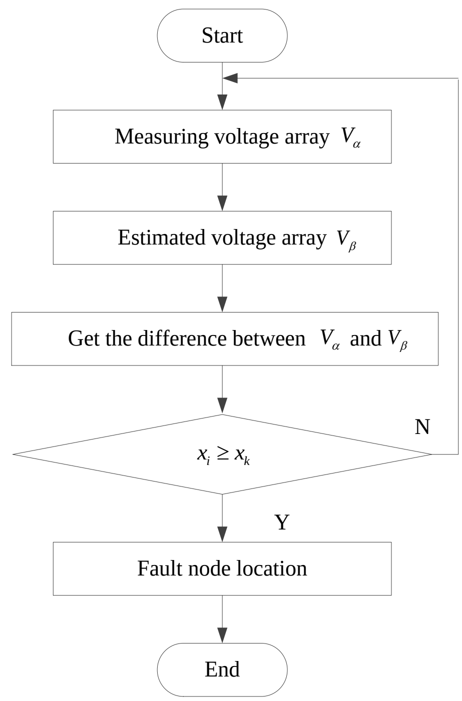


Fig. 2. Flowchart of fault node location.

C. Realization of Quantitative Security Protection for Distribution of Automation Nodes

1) Logical architecture and overall hardware structure of trusted power distribution terminal: In order to protect the nodes of the distribution network automation system, the trusted computing technology [11,12] is comprehensively used to design a trusted distribution terminal based on trusted computing, with access control as the core and security management as the support. The trusted power distribution terminal is based on the trusted platform of the hardware layer as the root of trust, and the trusted cryptographic module provides cryptographic computing services for the trusted computing platform. In order to give full play to the computing function of the trusted cryptographic module, it is also necessary to implement the trusted service management platform module based on the security protocol at the operating system layer. The logical architecture of the trusted power distribution terminal is shown in Fig. 3.

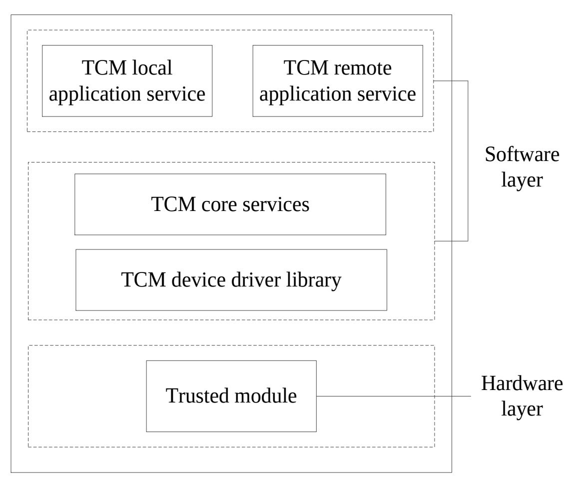


Fig. 3. Logical architecture of trusted power distribution terminal.

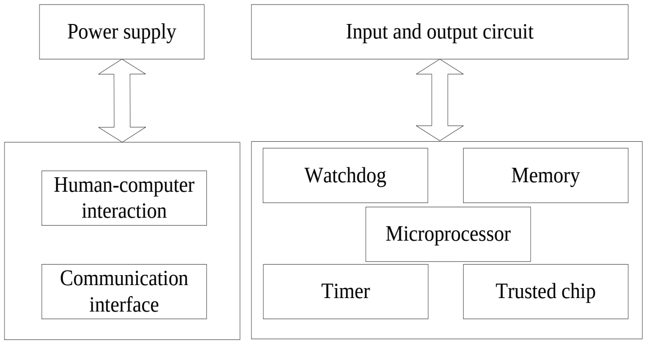


Fig. 4. Overall hardware structure of trusted power distribution terminal.

The overall hardware structure of the trusted power distribution terminal is shown in Fig. 4, which mainly includes microprocessors, human-computer interaction and communication interfaces [13,14]. The microprocessor consists of a watchdog, a memory, a timer and a trusted chip: the trusted chip is the core for realizing the trust and provides the root of trust for the trust chain; the memory is mainly responsible for the storage and memory functions; the timer realizes the timing and computing functions. The watchdog provides system resilience against transient failures [15-17]. The trusted control module is the core component of trusted computing. The root of trust is implanted inside to realize the root of trust control function, so as to combine password and control to realize active control and active defense of the nodes of the entire distribution network automation system. While measuring the mainboard controls, it also provides security computing capability support for the measurement mechanism of the software system.

2) Software design of trusted power distribution terminal: The distribution network automation system node is a device that can operate and control independently in the distribution network automation system [18, 19]. It is the basic object of constructing the distribution network automation system. The logical architecture and overall hardware structure of the trusted power distribution terminal have been introduced above, and the following will focus on the analysis of the trusted power distribution terminal software.

The control function application interface is a bridge connecting the software system and the hardware platform. When the hardware layer is powered on, the trusted control module will measure the motherboard controls, and complete the hardware environment and configuration settings through secure boot, thereby establishing a trusted chain. Provide support and services for the software layer. While processing information flow and energy flow, nodes have different processing capabilities of information, so the specific implementation will be set according to the functional requirements of nodes.

Nodes usually perform denoising, compression and other processing operations on data, and these operations will be recorded in PCR for metric protection. The node usually stores the collected data locally for a period of time, and then submits it to the superior node unit, or submits it to the superior node unit in time; therefore, data information storage is very important. For a small amount of collected data d, this paper adopts the domestic SM2 asymmetric encryption algorithm [20], and the encryption method is expressed as:

$$d_s = \omega_{TCM}\left(d, \theta_{PCR}, \mu_d\right) \tag{14}$$

In the formula: d_s is data encryption; ω_{TCM} is TCM algorithm; θ_{PCR} is the expected value of PCR, only when the expected value can be decrypted; μ_d is the node encryption public key. When the amount of data is large, the more efficient symmetric encryption algorithm SM4 is used, namely:

$$d_s = \omega_{SM4}\left(d, \theta_{PCR}, \mu_{SM4d}\right) \tag{15}$$

In the formula: ω_{SM4} is the SM4 encryption algorithm; μ_{SM4d} is the symmetric key.

When the system needs to use data, it will first obtain the current PCR value of the system and compare it with the expected value. Only when the comparison result is consistent, it will use the key stored by itself to decrypt it, so as to ensure the integrity and confidentiality of the distribution automation system nodes, and thus realize the quantitative security protection of the distribution automation nodes.

IV. EXPERIMENTAL VERIFICATION ANALYSIS

In order to verify the effectiveness and application effect of the distribution automation node quantitative security protection technology based on the attack tree, an experimental study is carried out.

The IEEE33 node standard distribution system with a voltage level of 10kV is selected as the analysis object. For the distribution network with multiple nodes, the method in this paper, the method in the literature [5] and the method in the literature [6] are used to carry out the quantitative safety protection test of the distribution automation nodes. In order to ensure the accuracy of the experimental results, MATLAB software is used to process the experimental results.

1) Fault node location effect: In order to verify the safety protection effect of the method in this paper, compared with the method of literature [5] and literature [6], the fault node location accuracy of the three methods is analyzed. The fault node location accuracy will have a certain impact on the quantitative safety protection of distribution automation nodes. The higher the accuracy of fault node location, the better the quantitative safety protection effect of distribution automation nodes. The specific experimental results are shown in Fig. 5.

2) Time-consuming test of trusted computing: Secondly, it is time-consuming to test the trusted computing of the method in this paper, the method in the literature [5] and the method in the literature [6]. The test results are shown in Table I.

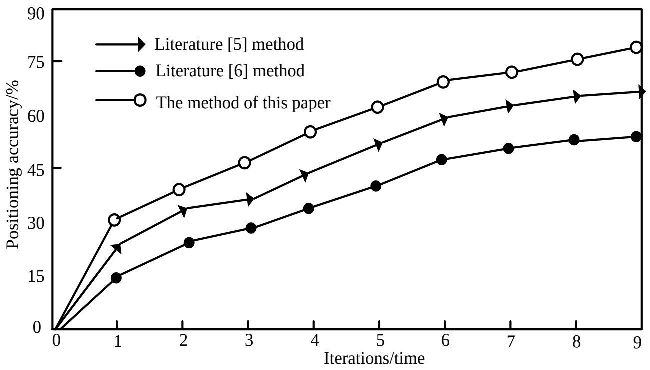


Fig. 5. Comparison results of fault node location accuracy.

TABLE I. TRUSTED COMPUTING TIME-CONSUMING TEST RESULTS/S

Iterations/time	The method of this paper	Literature [5] method	Literature [6] method
1	3.62	5.21	4.32
2	3.70	5.30	4.52
3	3.84	6.26	5.24
4	4.25	6.68	5.97
5	4.29	7.19	6.32
6	4.40	7.24	6.65
7	4.54	7.85	7.57
8	5.62	8.54	7.85
9	5.85	9.25	9.25

3) Network robustness test of distribution network automation system: Finally, the robustness of the distribution network automation system network is tested under the condition of increasing number of fault nodes. The specific test results are shown in Fig. 6.

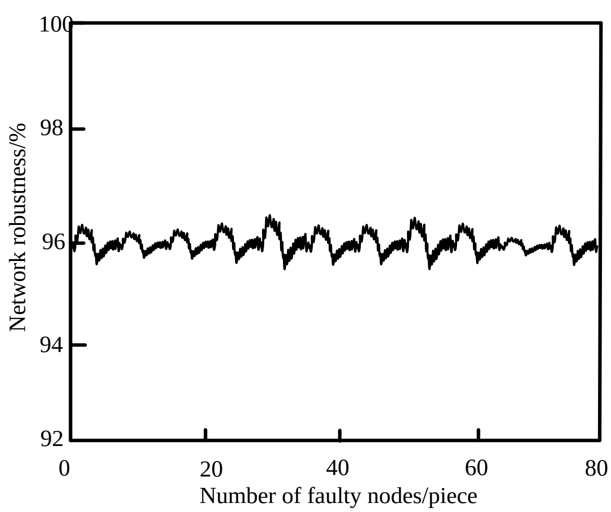


Fig. 6. Network robustness test results.

V. DISCUSSION

The effectiveness of the design method has been experimentally analyzed in the previous section, and the following conclusions can be drawn:

1) It can be seen from Fig. 5 that the node positioning accuracy of the distribution network automation system of this method is the highest among all methods. With the continuous increase of the number of iterations, the fault node positioning accuracy curve continues to grow, and the highest value of the fault node positioning accuracy reaches more than 80%, which shows that the method in this paper has a good security protection effect of distribution automation node quantification. In contrast to the other two methods, although with the continuous increase of the number of iterations, the fault node location accuracy is gradually improved, but there is still a certain gap between the fault node location accuracy and the method in this paper. This is because the method in this paper analyzes the risk assessment factors of the nodes in the distribution automation system, and through the evaluation and analysis of node vulnerabilities, the fault nodes in the distribution automation system are found in advance, so as to improve the positioning accuracy of the fault nodes.

2) According to the data in Table I, the method in this paper, the method in literature [5] and the method in literature [6] have a certain increase in the time-consuming of trusted computing in the process of increasing the number of iterations, but the overall increase in the method in this paper is small, indicating that the trusted computing efficiency of the method in this paper is high, and the minimum trusted computing time is only 3.62s, which shows that the quantitative security protection of the distribution automation nodes of the method in this paper is faster, which is beneficial to improve the security of the power system. This is because the design method took the lead in analyzing various factors and carried out analysis on the basis of grasping the overall factors, thus reducing unnecessary time waste and improving efficiency.

3) According to the network robustness fluctuation test results in Fig. 6, it can be found that after the number of fault nodes increases, the robustness of the distribution network automation system network does not fluctuate greatly, and the overall fluctuation has certain regularity, which indicates that the security protection effect of this method is good, and the network attack cannot be cracked. This is because this method uses trusted computing technology to design trusted distribution terminals. When the amount of data is large, a more effective symmetric encryption algorithm SM4 is required, which can better achieve the node security protection in the distribution automation system.

VI. Conclusion

In order to improve the security of distribution automation system nodes and ensure the safe operation of the distribution network as the research goal, a quantitative security protection technology based on attack tree for distribution automation nodes is proposed. Analyze the factors of node risk assessment of the distribution automation system, evaluate the node vulnerability of the distribution automation system, and comprehensively assess the risk of the distribution automation system using the attack tree. According to the evaluation results, the distribution network control model under network attack is constructed, the fault node is located, and the trusted distributed terminal is designed using trusted computing technology to achieve the node security protection of the distribution network automation system. The research results show that the fault node location accuracy of the method in this paper is high, the maximum fault node location accuracy is more than 80%, the trusted computing time is low, and the minimum trusted computing time is only 3.62 seconds. Moreover, the robustness of the system network fluctuates little, and the overall fluctuation has certain regularity, which shows that the method has certain effectiveness, and has certain application value and advantages in this field. However, as other performance indicators were not analyzed in the analysis process, there may be some deficiencies, which will become the focus of the next study.

Acknowledgments

The study was supported by "Research and Development of Key Technologies for Safety Protection of Power Distribution Automation System Based on Electric Power Industrial Control Safety Shooting Range (Grant No. B311NB220002)".

References

[1] Lin Y J, Jing C, Cao C, Bai K. Distributed Power Supply Access Distribution Network Planning with Timing Characteristics. Computer Simulation, 2022, 39(2):51-55,72.

[2] Xie L, Luo L, Li Y, Zhang Y, Cao Y. A Traveling Wave-Based Fault Location Method Employing VMD-TEO for Distribution Network. IEEE Transactions on Power Delivery, 2020, 35(4):1987-1998.

[3] Bhagavathy S, Pearsall N, Putrus G, Walker S. Performance of UK Distribution Networks with single-phase PV systems under fault. International journal of electrical power and energy systems, 2019, 113(12):713-725.

[4] Chen K, Hu J, Zhang Y, Yu Z, He J. Fault Location in Power Distribution Systems via Deep Graph Convolutional Networks. IEEE Journal on Selected Areas in Communications, 2020, 38(1):119-131.

[5] Ni W D, Wu L H, Wang J F. Cyber security protection and hardware acceleration of distribution automation system based on autonomous security chip. Journal of Electric Power Science and Technology, 2020,35(03):166-172.

[6] Wu J Y, Chen H Q, Zhang L J, Lai Y Y. Research on Information Security Protection of Active Distribution Network Based on Trusted Computing. Guangdong Electric Power, 2020,33(03):79-87.

[7] Shameem M, Khan A A, Hasan M G , Akbar M A . Analytic Hierarchy Process Based Prioritisation and Taxonomy of Success Factors for Scaling Agile Methods in Global Software Development. IET Software, 2020, 14(4):389-401.

[8] Yang L G, Li C, Lu L, Guo T. Evaluation of port emergency logistics systems based on grey analytic hierarchy process. Journal of Intelligent and Fuzzy Systems, 2020, 39(3):4749-4761.

[9] Buldas A, Gadyatskaya O, Lenin A, Mauw S, Trujillo-Rasua R . Attribute evaluation on attack trees with incomplete information. Computers & Security, 2020, 88(1):101630.1-101630.17.

[10] Abdo H, Kaouk M, Flaus J M, Masse F. A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie -combining new version of attack tree with bowtie analysis. Computers & Security, 2018, 72(1):175-195.

[11] Ibrahim F, Hemayed E E. Trusted Cloud Computing Architectures for infrastructure as a service: Survey and systematic literature review. Computers & Security, 2019, 82(5):196-226.

[12] Yoro R E, Ojugo A A. An Intelligent Client-Trusted and Dependable Security Framework to Ease Smartphone Portability on Community Cloud-Computing. Computer Networks, 2019, 6(1):1-7.

[13] Zou X F, Xiao Y X. Modbus telegram security of distribution network based on SM2. Power System Protection and Control, 2018, 46(12):151-157.

[14] Chen W, Xue H, Wang T Y, Cui J S, Wang L N. Olympic:A Data Symmetric Encryption Algorithm Based on Multiple Plaintexts. Journal of Wuhan University(Natural Science Edition), 2021, 67(3):213-220.

[15] Li T, Wang G, Liu Y, Yang Z B, Ren S, Shang W L. Information Security Risk Analysis of Intelligent Terminal in Distribution Network. Smart Power, 2020, 48(9):118-122.

[16] Li Hongxin, Zeng Jiang, Zhang Huaying, et al. Voltage sag analysis of a transmission and distribution network based on a compensation method under a distribution network fault. Power System Protection and Control, 2020, 48 (16): 45-53.

[17] Sun Yangsheng, Tu Qi, Huang Zhenyu, et al. Intelligent transmission technology of fault information in a resilient distribution network based on 5G and IEC61850. Power System Protection and Control, 2022, 50 (21): 108-117.

[18] Miao Renjie, Liu Yulin, Zhang Li, et al. A Fault Location Algorithm in Radial Distribution Networks with Distributed Generators Based on Multi-Agent Technology. Power System and Clean Energy, 2021, 37 (01): 8-15.

[19] Li Zhenxing, Xu Hao, Fu Yuting, et al. Fault Identification Strategy for Distribution Network with DGs Using Current Cosine Similarity. Proceedings of the CSU-EPSA, 2022, 34 (4): 1-10.

[20] Li Jiawei, Wang Xiaojun, He Jinghan, et al Distribution Network Fault Location Based on Graph Attention Network. Power System Technology, 2021, 45 (6): 2113-2121.

Analysis and Detection of Tomatoes Quality using Machine Learning Algorithm and Image Processing

Haichun Zuo
School of Electrical and Computer Engineering,
Nanfang College Guangzhou, Guangzhou
Guangdong, 510970, China

Abstract*—Grading of agricultural products methods based on artificial intelligence is more important. Because these methods have the ability to learn and thus increase the flexibility of the system. In this paper, image processing systems, detection analysis methods, and artificial intelligence are used to grade tomatoes, and the success rate of grading these methods is compared with each other. However, the purpose of this study is to obtain a solution to detect appearance defects and grade and sort the tomato crop and provide an efficient system in this field. A visual dataset is created, to investigate the approach of image processing and machine learning based on a tomato image. Tomato models are placed individually under the camera and samples are classified in a lighting box away from the effects of ambient light. Data sets have been used in three types of first, second, and third quality categories. It should be noted that quality category one has the best quality and quality category two has the medium quality and category three has the worst quality, Also, each data class contains 80 samples. Using tomato appearance such as size, texture, color, shape, etc. Image processing is performed for extract features. Tomato images are pre-processed for optimization. Then, to prepare for classification, the dimensions of the images are reduced by principal component analysis (PCA). Three categories of an artificial neural network, a support vector machine, and a decision tree are compared to show the most efficient support machine. The analysis is examined in two classes and three classes. The support vector machine has the best accuracy compared to other methods so this rate is 99.9% for two classes and 99.79% for three classes.

***Keywords*—Machine learning; image processing; product category; tomato quality rating**

I. INTRODUCTION

Farmers usually separate healthy and damaged tomatoes according to their size and quality [1,2]. The acceptable quality of tomatoes makes separating the healthy from the bad more accessible, and it prevents the spread of rotten tomatoes among healthy ones [3,4]. Damaged tomatoes are either sold at a lower price, or discarded. Sorting tomatoes in the traditional way are still done by old people, which takes a lot of time. Relying on a product defect alone can lead to a fundamental error by relying on human inspection. The desire to carry quality tomatoes with undivided defects has a great impact on customer feedback and satisfaction. In this study, Image processing and machine learning technology are used. In this way, an image of a tomato is taken, and the presence of defects is automatically detected in a computer vision system to distinguish healthy from damaged tomatoes, using the concept of a developed system [5,6].

Every day, intelligent systems with industrial applications are increasing. The most important biological processes in the production of crops are the classification of vegetables and fruits. Still, such processes are done manually in a country like Egypt based on a database [7]. After wheat, tomato is the eighth most used crop. The global production of tomatoes is reported to be about 159 million tons for 144 countries. Fresh fruit revenue in 2011 is about $582 trillion [8]. The performance, quality, and weight of the products are excellent in five countries including the United States, China, India, Turkey, and Egypt. Although very important in the production of previously mentioned tomatoes, there is very little literature on the works. In recent years, tomato grading has become very important, especially with the detection of diseases, due to new market restrictions. Therefore, the need for new technologies in the process of separation and product quality monitoring has increased.

Daily, millions of people use tomatoes and vegetables, that's why it is one of the most popular food products in human life. However, labor costs have increased as the workforce ages, making many farms less profitable. As mentioned, tomato is a popular food product among people, so with the increase in population, tomatoes need to be produced more. An effective solution to control quality and reduce costs is to use a robot instead of human labor to harvest tomatoes. Therefore, most researchers have spent the past few decades building robots to harvest fruits and vegetables [9,10]. Tomato color is the main indicator for the detection of ripening. Different steps are taken to produce tomato fruit. These stages change the tomato's color from green to light pink, then to pink, then to bright red, and finally to red, which categorizes them into different categories [11,12]. The closer the tomato is red, the better its quality. and leads to an increase in prices for companies. The storage time for a quality harvest is about 70 to 75 days in total. The time stages of tomato ripening include 21 to 28 days for green, 15 to 20 days for light pink, 7 to 14 days for pink, 5 to 6 days for bright red, and 2 to 4 days for red [13,14]. Therefore, improving the tomato classification system is an important task for designing a crop robot. In recent years, machine vision and pattern recognition methods have been well received by researchers, especially in the processing or sorting of many intelligent agricultural products. According to EU regulations on processed vegetables and fruits, the main condition is that the tomatoes arrive intact, completely red, fresh, without rotting and cracks in satisfactory condition, and without damage. The sorting and grading method is done manually. As mentioned, the manual method has low accuracy, high cost, and high effort [15,16]. Efficiency sorting governs how the product is delivered through packaging lines and quality standards. Therefore, as stated above, it can be concluded that the importance of this research is the requirement of high accuracy, high speed, and low cost is very important for sorting using machinery tomato detection. The effective solution is the use of mechanical sorting machines for tomatoes. But mechanical sorting machines can only classify tomatoes according to size and weight.

Now, with the advancement of technology and the introduction of vision devices, it covers the limitations of mechanical devices. New generation machines classify tomatoes with high accuracy, high speed, and low cost [17,18]. In 1985, Sarkar et al. reported the first tomato classification based on machine vision. This study is about the classification of tomatoes according to their size, shape, color, and defects [19-21]. Classification of products is an important step in the packaging and storage processes, which image processing is one of the practical tools in the field of post-harvest technologies. Izadi et al., obtain an algorithm for detecting physical defects [22] and degree tomato product classification and providing an efficient system in this field [23-25]. The specific objectives of this research are to develop a computer vision system and develop an efficient image processing algorithm, to create an accurate and accurate detection algorithm, and to create different classifications for classifying tomatoes into different grading categories [26-30]. One of the purposes of this study is to develop an image processing algorithm and develop a computer vision system, which is used immediately after harvesting tomatoes. Therefore, the time of the process of picking and classifying tomatoes is greatly reduced, in addition, due to the high accuracy of the optical device, quality sorting is done.

Given that image processing and machine vision are one of the current technologies in the world, the main task of a researcher in the field of using machine vision in agriculture is to extract the properties of objects in digital images using these technologies [31-33].

The most important sensory feature of fruits and vegetables is the appearance that affects the market value, priority, and consumer choice. Although sorting and grading are done by humans, it is inconsistent, time-consuming, variable, subjective, heavy, expensive, and easily affected by the environment. Hence, a clever fruit grading system is needed. Therefore, the motivation of this study is to provide a method of quality detection and automatic grading for many tomato disorders.

Most agronomic disorders are observed according to roots, stems, leaves, and fruits based on the type and causes of damage. Fruit images are used for diagnostic purposes. The image acquisition method used to collect the data set ensures that the entire fruit is examined. In the first part, general fruits, especially tomatoes, are introduced. The second part presents the general method of grading tomatoes based on fruit, image processing algorithm, classification methods, machine learning systems, neural network methods, and application of appearance features, color recognition, and texture recognition in detail. In the third section, the proposed methods are introduced and the steps related to the method are described from an algorithmic point of view. In the fourth part, the proposed method is implemented on the set of images, and the evaluation criteria are evaluated. Finally, the fifth part will end with a conclusion.

II. Related Works

The most advanced image-based tomato detection method is shown in this section.

Malik et al [49] proposed a Watershed segmentation method was utilized to separate the clustered fruits according to ripe tomato identification algorithms, which were based on the HSV color space. The Otsu segmentation algorithm was used by Arum Sari et al. [50] to multiply the Cb and V channels from the YCbCr and YUV color spaces, respectively, when categorizing tomatoes into 6 separate groups.

Using the color space L*a*b*, Yin et al. [51] segmented ripe tomatoes using K-means clustering, and Indriani et al. [52] divided tomatoes into 5 groups using the HSV color space, Gray Level Co-occurrence Matrix, and K-Nearest Neighbor. To segment and localize ripe tomatoes in a greenhouse, Huang et al. [53] employed the L*a*b* color space. Goel and Sehgal [54] used fuzzy rule-based classification through the RGB color space to distinguish the fruits from the background.

Some studies use edge detection operations like the Canny [55] or the Sobel [56] operators as suggested in the Zhang [57] and Benavides et al. [58] investigations, respectively, to enhance detection and produce better segmentation.

III. Criteria and Data Sets

In this section, in the first stage, the criteria used in this study, including sensitivity, specificity, and accuracy are examined. The data set to be evaluated is then introduced.

A. Criteria

Contrast is one of the most important criteria for determining image quality. The contrast shows the difference between the darkest and lightest parts of an image. When the contrast of the image is low, the difference between the brightness of those points is small. As a result of low image contrast, the image will fade. One way to increase the contrast of a low-quality image is the Histogram Equalization smoothing technique. It changes the values of the gray surfaces of the image to cover the entire range from 0 to 255 per pixel.

1) Sensitivity: Sensitivity is one of the performance evaluation criteria of binary classification tests, and it is also used in artificial intelligence. Sensitivity measures the ratio of positively identified positives, meaning that the classifier can correctly identify the data in classes. The terms "positive" and "negative" do not mean profit, but the existence or non-existence of a condition. For example, if caries is caries, "positive" means "caries" and "negative" means "healthy". In many sensitivity tests, the degree to which positive cases are ignored is called positive, so false negatives are rare.

Sensitivity is considered to be the correct positive ratio (the percentage of tuples that are positively labeled and their class is actually positive). The sensitivity parameter is obtained based on Equation (1) [34].

$$Sensitivity = \frac{TP}{TP+FN} \qquad (1)$$

Where *TP* is the correct number of first-class predictions, and *FN* is the incorrect number of second-class predictions.

2) Specificity: Specificity measures the proportion of negatively identified negatives (for example, the percentage of healthy tomato samples that are correctly identified as not rotten). Characteristic is the category in which real negatives are classified in this way, so false positives are rare. A sensitive test seldom ignores a real positive (for example, no caries is shown despite caries). The specificity parameter is obtained based on Equation (2) [35].

$$Specificity = \frac{TN}{FP+TN} \qquad (2)$$

Where *TN* is a true negative, or healthy samples of tomatoes are properly identified as healthy. The *FN* is right, it's wrong.

3) Accuracy: Accuracy is the most intuitive measure of performance and is simply the ratio of the predicted correct observations to the total observations. It may be thought that our model is the best if we are very careful. In general, accuracy is a valuable measure when there is a symmetric data set where the negative, positive, and false positives are approximately the same. Therefore, to evaluate the performance of your model, you should look at other parameters [36].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} (3)$$

Accuracy is also used as a statistical criterion that a binary classification test correctly identifies or removes a condition. In a classification task, the accuracy for a class is the number of actual positives divided by the total number of elements labeled positively related to the class (i.e. the sum of the actual positives and False positives are items that are incorrectly labeled as belonging to the class). The recall is defined as a positive number divided by the total number of elements that actually belong to the positive class, (i.e. the sum of false positives and false negatives are those that do not have a positive classification label) [37,38].

B. Data Sets

One of the important factors in quality recognition and model evaluation is the use of appropriate data sets. Given that in this study, with the help of image processing and machine learning, the proposed solution to determine the quality of tomatoes is presented. Therefore, it uses data sets in three types of first, second, and third-degree quality categories. It should be noted that quality category one has the best quality and category three has the worst quality, and also each data class contains 80 samples. An example of images related to different types of tomatoes can be seen in Fig. 1. The data set is graded into three categories in terms of quality, which include first grade (best quality), second grade (medium quality, or small healthy-slightly spoiled-immature), and third grade (worst quality). In this study, in order to improve the quality of images and eliminate their potential noise, the data preprocessing method has been used. Methods including applying the median filter [39], calculating the area in an image [40], the optimal local pattern and the matrix of gray surface occurrence [41], the Local Binary Pattern method (LBP) [42], and LAB color channel [43] are used in the step of extracting the shape, color, and texture characteristics. Then, the dimensions are reduced by Principal Component Analysis (PCA) to prepare the vector for classification [44]. Three classification models are used, including the decision tree [45], the support vector machine [46], and the artificial neural network [47]. These classifiers are compared to measure the performance of different classifiers on datasets.

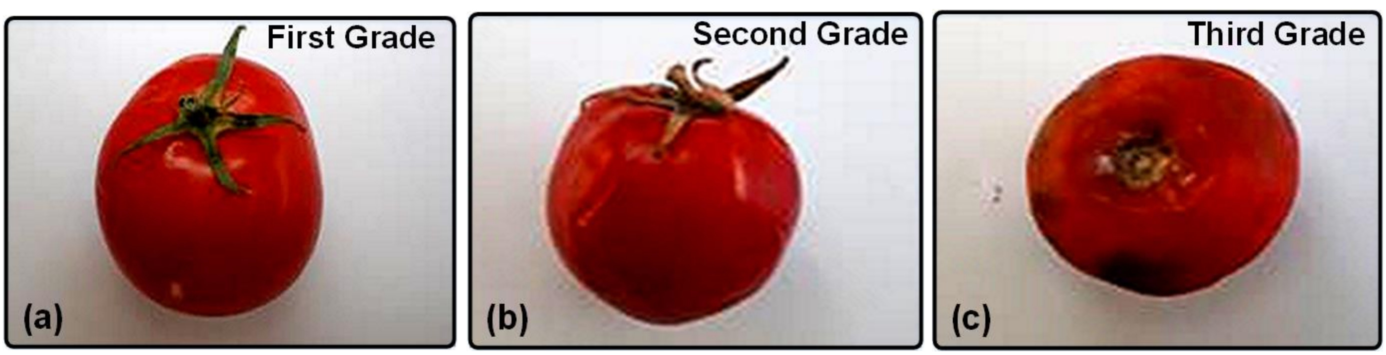


Fig. 1. Tomato sample image of class (a) first grade, (b) second grade, (c) third grade.

IV. Proposed Method

This article, with the help of features related to the shape, texture, and color of tomatoes, offers a suitable solution for quality recognition and analyzes the effect of using image processing methods in quality recognition using machine learning. The purpose of using preprocessing operations on the data set is to improve its quality and then extract the appropriate feature from them to compare with other methods and also by selecting the appropriate features finally with the help of the proposed algorithms the ability of the machine learning model in support vector machine algorithms, the artificial neural network and the decision tree are examined.

With the decreasing trend of the labor force and also the advancement of technology in order to mechanize agricultural operations, there is a possibility to distinguish the quality of agricultural products. Product quality ultimately leads to post-evaluation economic justification for continuing to grow crops. There are several factors in identifying tomatoes in different quality categories. Classification of agricultural products causes, in addition to harvesting at the right time, also causes the supply of products to be done in a timely manner, to reduce their quality and eliminate first-class products and turn them into second-class products. And three and prevent losses due to waste of time. One of the most important processes in the classification and maintenance of products is their grading and sorting, for which the use of image processing techniques is an important step in this area. The use of machine vision systems makes the post-harvest stages fast, easy, cheap, and with high accuracy, and on the other hand, reduces inspection time, quality assurance, and product grading.

For this purpose, the proposed design has been developed according to Fig. 2 in order to determine the quality of tomatoes. Accordingly, in order to increase the quality of images by applying preprocessing operations by changing the dimensions of images to the size of 400 × 300, converting color images to gray, Contrast Enhancement, and applying a median Filter. In the next step, the properties related to texture, shape, and color are extracted, so that in the next step, with the help of Principal Component Analysis (PCA), the most useful features are evaluated according to certain parameters using machine learning. Using the right data set plays an important role in achieving a reliable model. The use of appropriate hardware to some extent causes the image obtained from the cameras to be of good quality. In addition, the application of a series of algorithms and software methods causes images in the data set to have better quality. To improve the quality of images and eliminate their possible noise, preprocessing methods are used. First, the dimensions of the images are changed so that all the images by different dimensions are converted to 400 × 300, then the color images are converted to gray, the contrast is improved and the median filter is applied.

A. Feature Extraction

The median filter is the non-linear filters used to remove noise. To calculate the median in the 3 × 3 neighborhood, we have 9 cells, which after sorting them, we put the value in the middle of the array in output. One of the important features of this filter is maintaining the edges and their position without displacement. Fig. 3 shows a good example of the function of this filter in simple language.

To extract the shape feature, the appearance of the tomato and the space occupied by it has been used. In other words, according to Fig. 4, the path area within the boundary of an area is declared as a scale.

The LBP and Gray Level Co-Occurrence Matrix (GLCM) are some of the items used to extract image texture information. The LBP is one of the methods that can easily produce suitable features for high-precision tissue classification. In the usual LBP method, the histogram is used to extract features. It is a powerful tool for tissue analysis because the local binary model uses both statistical and structural characteristics of the tissue. In the LBP operator, local texture patterns are extracted by comparing the value of the adjacent pixels and the value of the central pixel and are finally represented by binary codes.

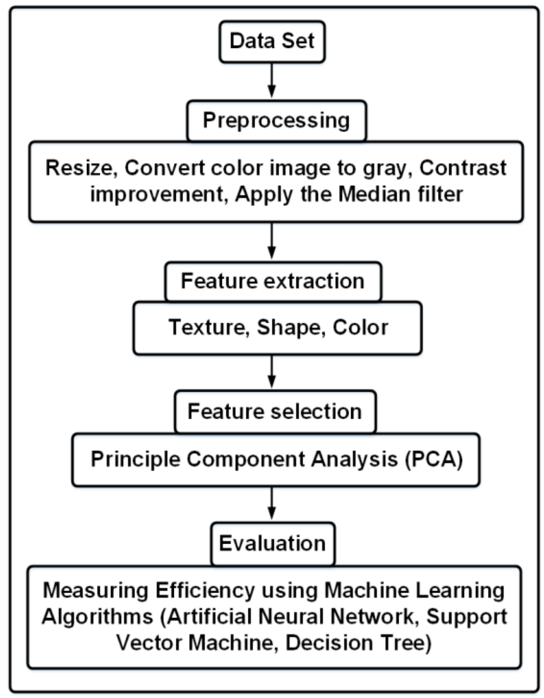


Fig. 2. Proposed method for detection the quality of tomatoes.

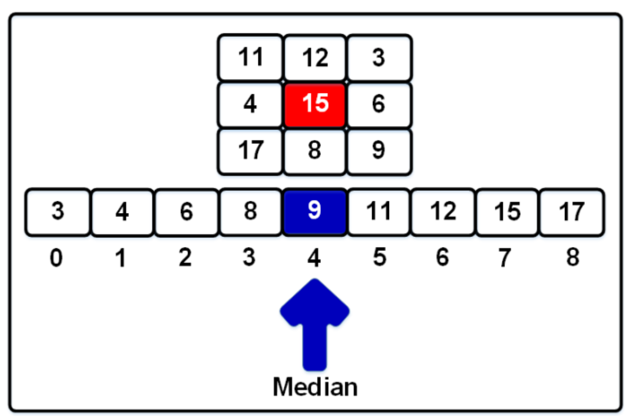


Fig. 3. The example of how to apply the median filter.

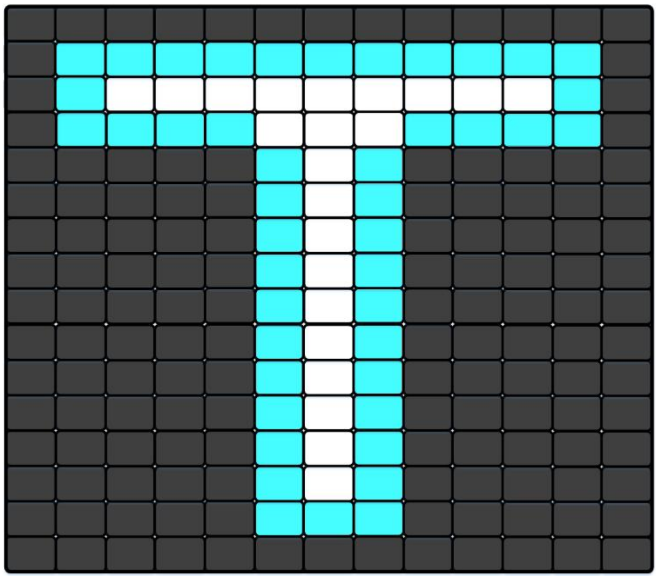


Fig. 4. How to calculate the area in an image.

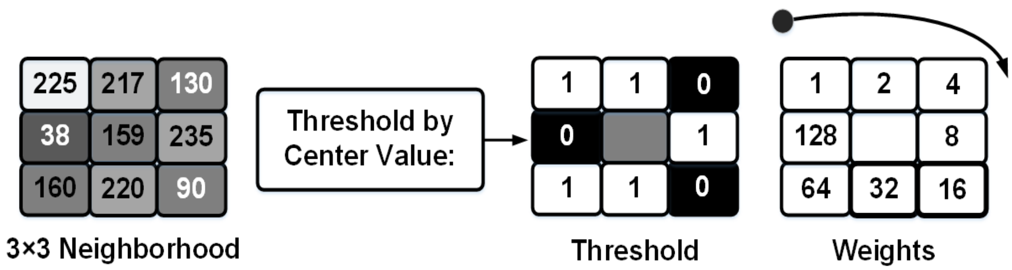


Fig. 5. The LBP basis method.

Local Binary Pattern (LBP) is another development of the local binary patterns. In order to identify the maturity stage of object feature representation, textural information provided by the local binary pattern (LBP) method was utilized [59]. The LBP transforms any local 3 3 picture region into a particular binary pattern. LBP basically contrasts the intensities of a center pixel (ic) and its eight surrounding neighbors (ik; where k = 1, 2,..., 8). If ikic is true, the k-th nearby pixel would receive a value of 1; otherwise, it would receive a value of 0. When an image rotates on the screen, all the neighbors will rotate around the center pixel in one direction. This rotation effect leads to different values for the local binary pattern. To this end, researchers have introduced a new development of the binary local pattern that is independent of rotation (the LBP with constant rotation). The expression independent of rotation here does not refer to spatial changes related to changes in light or different objects. Independent of rotation to eliminate the adverse effect of rotation, a circular bit rotation operator is used to the right to obtain all the binary codes that can be generated with these bits with several repetitions and then the minimum decimal value of the binary pattern (Fig. 5) [48].

In tissue statistical analysis, tissue properties are calculated through statistical distributions of compounds observed for grayscale in a particular position relative to each other. A matrix is also a matrix event, whose number of rows and columns is equal to the number of grayscale levels used in the image. That is, if the number of gray surfaces is up to G, then the dimensions of the matrix of the event in question are equal to one matrix G×G. The matrix P, P(i,j|Δx,Δy) represents the number of iterations of the relationship between two pixels separated by a pixel distance (Δx,Δy), and is defined by the neighborhood relation that one of the pixels has a gray *i* degree and the other has a gray *j* degree, as shown in Fig. 6.

In tissue statistical analysis, tissue properties are calculated through statistical distributions of compounds observed for grayscale in a particular position relative to each other. A matrix is also a matrix event, whose number of rows and columns is equal to the number of grayscale levels used in the image. That is, if the number of gray surfaces is up to G, then the dimensions of the matrix of the event in question are equal to one matrix G×G. The matrix P, P(i,j|Δx,Δy) represents the number of iterations of the relationship between two pixels separated by a pixel distance (Δx,Δy), and is defined by the neighborhood relation that one of the pixels has a gray *i* degree and the other has a gray *j* degree, as shown in Fig. 6.

L, A, B color mode is used to extract the color feature. This color fashion is the most complete color space designated by the International Committee on Lighting and describes all colors visible to the human eye. The three coordinates L, A, B are shown in Fig. 7.

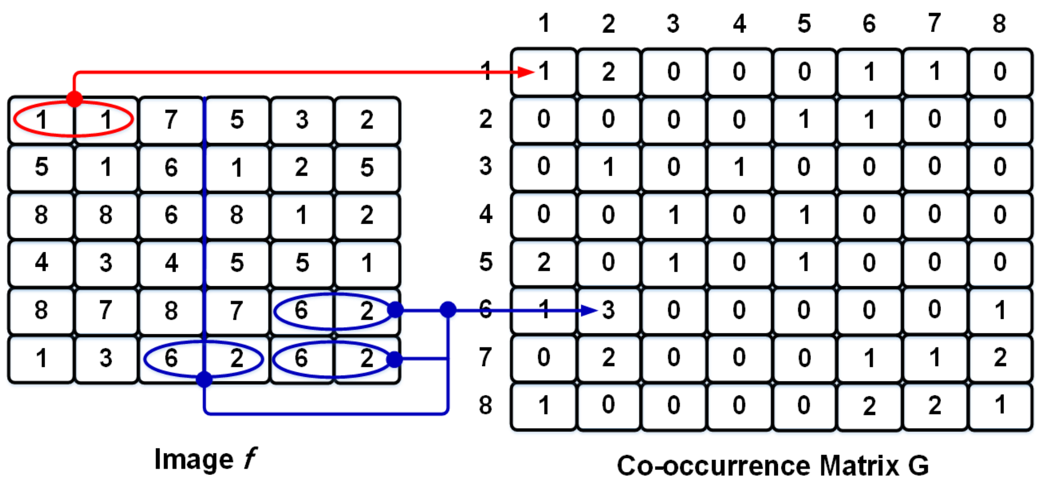


Fig. 6. Gray Level Co-occurrence Matrix (GLCM).

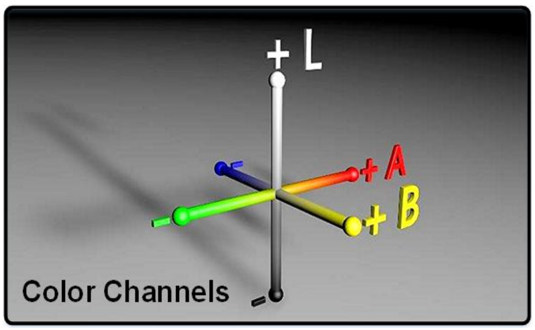


Fig. 7. L,A,B color channels.

The L indicates the intensity of light. The L=0 black and L=100 indicate the scattering of full light. A's position varies between green and red, negative values represent green and positive values represent red. The B Its position varies between blue and yellow, negative values B indicate blue colors and positive values represent yellow colors.

In order to reduce the dependence and noise of bands in images with a large number of bands, the methods of reducing the dimensions of the feature space are used. One of the most well-known methods of reducing the size of the feature space is the principal component analysis method. As the name implies, it can identify key components and help us analyze a series of features that are more valuable instead of examining all the features. In fact, PCA extracts those features that give us more value. After collecting samples and performing preprocessing steps and after extracting and selecting features, machine learning techniques such as support vector machine algorithms, artificial neural networks, decision tree, and k-means clustering are used to evaluate the model.

V. Results and Discussion

What has been presented in the previous sections is a description of the problem of examining the quality of tomatoes using machine learning and image processing. In fact, the background, challenges, and necessities of the study have been analyzed. In this section, by presenting a method based on machine learning along with image processing techniques, the problem of quality recognition of tomatoes is improved, and the proposed method is examined in terms of implementation and evaluation. The proposed steps are then evaluated step by step.

A. Preprocessing and Feature extraction

As mentioned earlier, after applying any type of processing algorithm and pre-processing evaluation, the data set is introduced. For this purpose, preprocessing goals are performed by changing the dimensions of the images to 400 × 300, changing the color channels from three channels to one channel, or in other words, turning the color image to gray, improving the contrast, and applying a median filter. In fact, after these steps, the images in the data set become images like Fig. 8.

According to the data set and based on their differences in different qualitative categories, different types of features are obtained according to Table I.

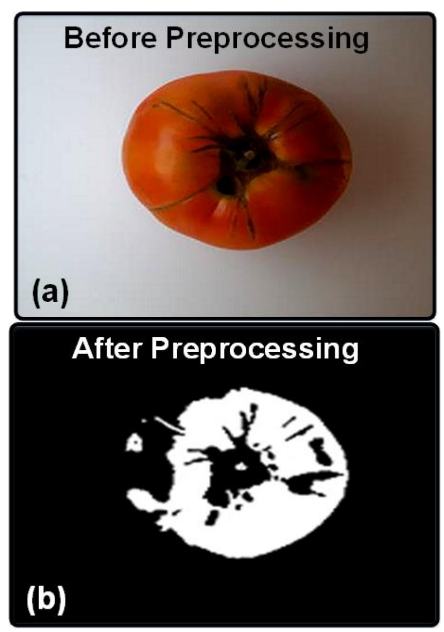


Fig. 8. An example of a tomato image, (a) before preprocessing, (b) after preprocessing.

TABLE I. TYPES OF FEATURE EXTRACTION USED

Feature	Category	Output
Area	Shape	A numerical value
Local Binary Pattern	Texture	A 300 × 400 matrix
Gray Level Co-Occurrence Matrix	Texture	A 255 × 255 matrix
L,A,B color model	Color	A matrix of 300 × 400 × 3

Agricultural products such as tomatoes in the ripe state, as well as in the post-ripening period may be damaged. Based on this, the characteristics of color, shape, and texture cause the quality of tomatoes to be recognizable in different periods, and tomatoes of suitable quality have the best shape, color, and texture. On the other hand, the size of the tomato causes the difference in the first and second quality degrees to be distinguishable. Applying these features together causes the learning model to be able to distinguish each of these quality degrees in the later stages.

B. Feature Selection

As the number of features increases, the dimensions of the feature space also increase. On the other hand, not all of these features are useful, but their presence may even cause the model to be too compatible with the data set, or so-called overfitting. In the PCA method, with the help of feature removal and extraction methods, an attempt is made to separate important features from less important features. And further, reduce the size of the feature by removing minor features. Fig. 9 shows well how the data set is distributed under different PCA components.

PCA has three different components, each of which is well represented in each data class in the data class. A basic component is a normalized linear combination of the main predictions in the data set. Component # 1, component # 2, and component # 3 are the basic components. Component # 1 is a linear combination of the main predictions that contains the largest variance in the data set. This component determines the direction of maximum changes in the data. The higher the range of changes in component 1, the more information there is in this component. Component # 2 and then component # 3 will be important. Fig. 9(a) shows that with the prediction components, the accuracy has reached 89.8% and shows the high accuracy of the diagnosis with only components # 3. Fig. 9(b) is created by the Principal Component Analysis Tool. The principal component coefficients for each variable, the principal component score, are also displayed.

C. Validation

After extracting and selecting the appropriate features, the conditions for learning the model are provided. Accordingly, the evaluation criteria of each of these techniques are expressed as a prediction matrix and the parameters of accuracy, sensitivity, and specificity. It is important to note that in this article we classify two classes for the quality class by merging the first and second class into one class (these two classes differ slightly in size and appearance of products) and the third class in another class, as well as three classes for First, second and third-degree qualitative grading, has been used in three different data classes. The prediction matrix is used to evaluate the performance of the proposed classifier. The predicted label for each sample can have four possible states, including true negative (TN), true positive (TP), false negative (FN), and false positive (FP), as shown in Table II.

The detect components represent the number of samples that are correctly classified, while the main diameter components in the table represent errors. The prediction matrix is the result of evaluating the results of two neural network classes according to Table III. For example, 159 instances of class one has been correctly identified and two instances of data class one has been incorrectly predicted to belong to class two.

Fig. 10 shows a comparative trend of the accuracy of different classifications. Accordingly, the support vector machine has the best accuracy compared to the others, so that even its accuracy in two-class data in the case of merging first- and second-class products is better than three-class classification.

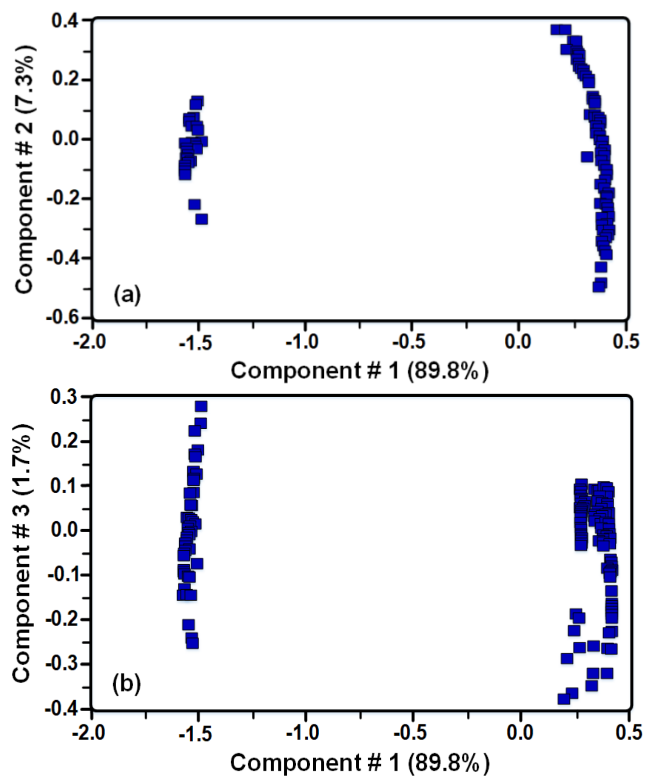


Fig. 9. Accuracy of classification performance of data set (a) component # 1 and component # 2, (b) component # 1 and component # 3.

TABLE II. PREDICTION MATRIX

		Prediction class	
		Class 1	Class 2
Real class	Class 1	TN	FP
	Class 2	FN	TP

TABLE III. NEURAL NETWORK PREDICTION MATRIX

		Prediction class	
		Class 1	Class 2
Real class	Class 1	159	2
	Class 2	1	58

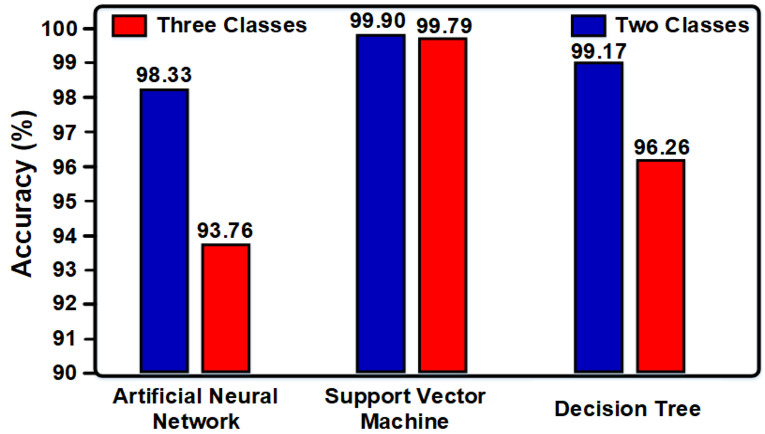


Fig. 10. Performance accuracy of data classification.

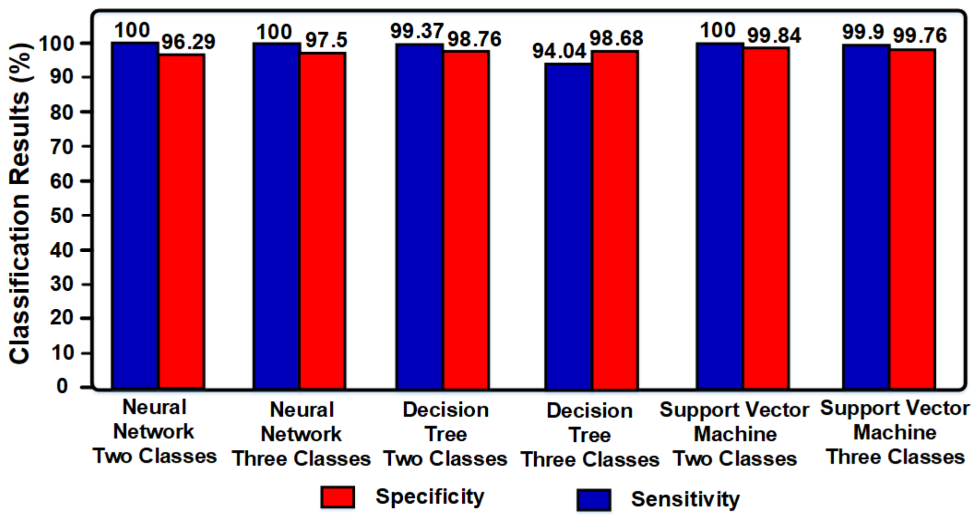


Fig. 11. Comparison of sensitivity and specificity of data classification performance.

According to Fig. 11, after the vector machine, the artificial neural network support performs better in the sensitivity criterion and the decision tree in the specificity criterion. But as a general result, it can be seen that compared to the neural network, the decision tree has the same performance as the support vector machine.

In this section, in the first step, in order to prepare the data set for the next stages of preprocessing operations. In the next step, by extracting the feature with the help of features related to texture, shape, and color and by selecting the features using principal component analysis and support vector machine classifications, artificial neural network, and tree, the decision is applied. The results show that the use of this method has better results compared to other methods and also the integration of the first and second quality grade also improves the accuracy of identification.

VI. Conclusion

The most important sensory feature of fruits and vegetables is the appearance that affects the market value, priority, and consumer choice. Although sorting and grading are done by humans, it is inconsistent, time-consuming, variable, subjective, heavy, expensive, and easily affected by the environment. Hence, a clever fruit grading system is needed. This paper provides an in-depth overview of various methods such as preprocessing, segmentation, feature extraction, and classification that focus on fruit and vegetable quality based on color, texture, size, shape, and defects. Using the appearance characteristics of tomatoes such as color, shape, size, and texture, image processing operations have been performed with MATLAB software to extract the mentioned features. Furthermore, the proposed algorithm can be used for some other types of image data because the proposed LPB feature representation enables to represent the object which can be presented better performance. In the future, a system can be set up that reads images on the rails simultaneously and processes them in video images. Algorithms such as deep learning are introduced that can be more efficient in real-time processing. Thus, the study method can be developed for future work with deep learning algorithms.

Acknowledgment

2021 Guangzhou Science and technology plan project (202102080497).

2021 school level teaching quality and teaching reform project (XJJG2110).

REFERENCES

[1] Pogonyi, A., Pék, Z., Helyes, L., & Lugasi, A. (2005). Effect of grafting on the tomato's yield, quality and main fruit components in spring forcing. Acta Alimentaria, 34(4), 453-462.

[2] Afifah, E. N., Murti, R. H., & Wahyudhi, A. (2021). Evaluation of a promising tomato line (Solanum lycopersicum) derived from mutation breeding. Biodiversitas Journal of Biological Diversity, 22(4).

[3] De Luna, R. G., Dadios, E. P., Bandala, A. A., & Vicerra, R. R. P. (2019). Tomato fruit image dataset for deep transfer learning-based defect detection. In 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM) (pp. 356-361). IEEE.

[4] Sanaeifar, A., ZakiDizaji, H., Jafari, A., & de la Guardia, M. (2017). Early detection of contamination and defect in foodstuffs by electronic nose: A review. TrAC Trends in Analytical Chemistry, 97, 257-271.

[5] da Costa, A. Z., Figueroa, H. E., & Fracarolli, J. A. (2020). Computer vision based detection of external defects on tomatoes using deep learning. Biosystems Engineering, 190, 131-144.

[6] Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. Food Research International, 62, 326-343.

[7] Pfister, S., Vionnet, S., Levova, T., & Humbert, S. (2016). Ecoinvent 3: assessing water use in LCA and facilitating water footprinting. The International Journal of Life Cycle Assessment, 21(9), 1349-1360.

[8] Semary, N. A., Tharwat, A., Elhariri, E., & Hassanien, A. E. (2015). Fruit-based tomato grading system using features fusion and support vector machine. In Intelligent Systems' 2014 (pp. 401-410). Springer, Cham.

[9] Sembiring, A., Budiman, A., & Lestari, Y. D. (2017). Design and control of agricultural robot for tomato plants treatment and harvesting. In Journal of Physics: Conference Series (Vol. 930, No. 1, p. 012019). IOP Publishing.

[10] Oktarina, Y., Dewi, T., Risma, P., & Nawawi, M. (2020). Tomato harvesting arm robot manipulator; a pilot project. In Journal of Physics: Conference Series (Vol. 1500, No. 1, p. 012003). IOP Publishing.

[11] Lili, W., Bo, Z., Jinwei, F., Xiaoan, H., Shu, W., Yashuo, L., ... & Chongfeng, W. (2017). Development of a tomato harvesting robot used in greenhouse. International Journal of Agricultural and Biological Engineering, 10(4), 140-149.

[12] Feng, Q., Zou, W., Fan, P., Zhang, C., & Wang, X. (2018). Design and test of robotic harvesting system for cherry tomato. International Journal of Agricultural and Biological Engineering, 11(1), 96-100.

[13] Syahrir, W. M., Suryanti, A., & Connsynn, C. (2009). Color grading in tomato maturity estimator using image processing technique. In 2009 2nd IEEE International Conference on Computer Science and Information Technology (pp. 276-280). IEEE.

[14] Rupanagudi, S. R., Ranjani, B. S., Nagaraj, P., & Bhat, V. G. (2014). A cost effective tomato maturity grading system using image processing for farmers. In 2014 International Conference on Contemporary Computing and Informatics (IC3I) (pp. 7-12). IEEE.

[15] Arjenaki, O. O., Moghaddam, P. A., & Motlagh, A. M. (2013). Online tomato sorting based on shape, maturity, size, and surface defects using machine vision. Turkish Journal of Agriculture and Forestry, 37(1), 62-68.

[16] Rokunuzzaman, M., & Jayasuriya, H. P. W. (2013). Development of a low cost machine vision system for sorting of tomatoes. Agricultural Engineering International: CIGR Journal, 15(1).

[17] Gould, W. A. (1975). A Preliminary Report on Mass Sorting of Mechanically Harvested Tomatoes.

[18] Duarte, A. C. (1965). The economics of mechanical harvesting of tomatoes in the Salinas Valley, California.

[19] Chen, Y. R., Chao, K., & Kim, M. S. (2002). Machine vision technology for agricultural applications. Computers and electronics in Agriculture, 36(2-3), 173-191.

[20] Mahendran, R., Jayashree, G. C., & Alagusundaram, K. (2012). Application of computer vision technique on sorting and grading of fruits and vegetables. J. Food Process. Technol, 10, 2157-7110.

[21] Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food and bioprocess technology, 4(4), 487-504.

[22] Izadi, H., Kamgar, S., & Raoufat, M. H. (2016). Tomato grading system using machine vision technology and neuro-fuzzy networks (ANFIS). Journal of Agricultural Machinery, 6(1).

[23] Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y. D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. Scientia Horticulturae, 263, 109133.

[24] Ismail, N., & Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. Information Processing in Agriculture, 9(1), 24-37.

[25] Kukreja, V., & Dhiman, P. (2020). A Deep Neural Network based disease detection scheme for Citrus fruits. In 2020 International conference on smart electronics and communication (ICOSEC) (pp. 97-101). IEEE.

[26] Minten, B., Tamru, S., & Reardon, T. (2021). Post-harvest losses in rural-urban value chains: Evidence from Ethiopia. Food Policy, 98, 101860.

[27] Granatstein, D., & Kupferman, E. (2006). Sustainable horticulture in fruit production. In XXVII International Horticultural Congress-IHC2006: International Symposium on Sustainability through Integrated and Organic 767 (pp. 295-308).

[28] Ireri, D., Belal, E., Okinda, C., Makange, N., & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artificial Intelligence in Agriculture, 2, 28-37.

[29] Clement, J., Novas, N., Gazquez, J. A., & Manzano-Agugliaro, F. (2012). High speed intelligent classifier of tomatoes by colour, size and weight. Spanish Journal of Agricultural Research, 10(2), 314-325.

[30] Patel, K. K., Kar, A., Jha, S. N., & Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. Journal of food science and technology, 49(2), 123-141.

[31] Chen, Y. R., Chao, K., & Kim, M. S. (2002). Machine vision technology for agricultural applications. Computers and electronics in Agriculture, 36(2-3), 173-191.

[32] Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food and bioprocess technology, 4(4), 487-504.

[33] Patel, K. K., Kar, A., Jha, S. N., & Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. Journal of food science and technology, 49(2), 123-141.

[34] Hintze, H., Lussi, A., Cuisinier, F., & Nyvad, B. (2015). Additional caries detection methods. In Dental caries: the disease and its clinical management. Oxford: Wiley Blackwell.

[35] Nasiri, A., Taheri-Garavand, A., & Zhang, Y. D. (2019). Image-based deep learning automated sorting of date fruit. Postharvest biology and technology, 153, 133-141.

[36] Stein, R. M. (2002). Benchmarking default prediction models: Pitfalls and remedies in model validation. Moody’s KMV, New York, 20305.

[37] Ebenuwa, S. H., Sharif, M. S., Alazab, M., & Al-Nemrat, A. (2019). Variance ranking attributes selection techniques for binary classification problem in imbalance data. IEEE Access, 7, 24649-24666.

[38] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the operational research society, 54(6), 627-635.

[39] Topno, P., & Murmu, G. (2019). An improved edge detection method based on median filter. In 2019 Devices for Integrated Circuit (DevIC) (pp. 378-381). IEEE.

[40] Freeman, T. G. (1991). Calculating catchment area with divergent flow based on a regular grid. Computers & geosciences, 17(3), 413-422.

[41] Park, Y., & Guldmann, J. M. (2020). Measuring continuous landscape patterns with Gray-Level Co-Occurrence Matrix (GLCM) indices: An alternative to patch metrics?. Ecological Indicators, 109, 105802.

[42] Basavaiah, J., & Arlene Anthony, A. (2020). Tomato leaf disease classification using multiple feature extraction techniques. Wireless Personal Communications, 115(1), 633-651.

[43] Khan, S., & Narvekar, M. (2020). Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment. Journal of King Saud University-Computer and Information Sciences.

[44] Sivakumar, J., Prashanth, J. E. P., Rajesh, N., Reddy, S. M., & Pinjari, O. B. (2020). Principal component analysis approach for comprehensive screening of salt stress-tolerant tomato germplasm at the seedling stage. Journal of biosciences, 45(1), 1-11.

[45] Basavaiah, J., & Arlene Anthony, A. (2020). Tomato leaf disease classification using multiple feature extraction techniques. Wireless Personal Communications, 115(1), 633-651.

[46] Das, D., Singh, M., Mohanty, S. S., & Chakravarty, S. (2020). Leaf disease detection using support vector machine. In 2020 International Conference on Communication and Signal Processing (ICCSP) (pp. 1036-1040). IEEE.

[47] Küçükönder, H., Boyaci, S., & Akyüz, A. (2016). A modeling study with an artificial neural network: developing estimationmodels for the tomato plant leaf area. Turkish Journal of Agriculture and Forestry, 40(2), 203-212.

[48] Tan, L., Lu, J., & Jiang, H. (2021). Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. AgriEngineering, 3(3), 542-558.

[49] Malik, M. H., Zhang, T., Li, H., Zhang, M., Shabbir, S., & Saeed, A. (2018). Mature tomato fruit detection algorithm based on improved HSV and watershed algorithm. IFAC-PapersOnLine, 51(17), 431-436.

[50] Sari, Y. A., Adinugroho, S., Adikara, P. P., & Izzah, A. (2017, November). Multiplication of V and Cb color channel using Otsu thresholding for tomato maturity clustering. In 2017 International Conference on Sustainable Information Engineering and Technology (SIET) (pp. 209-214). IEEE.

[51] Yin, H., Chai, Y., Yang, S. X., & Mittal, G. S. (2009, December). Ripe tomato recognition and localization for a tomato harvesting robotic system. In 2009 International Conference of Soft Computing and Pattern Recognition (pp. 557-562). IEEE.

[52] Indriani, O. R., Kusuma, E. J., Sari, C. A., & Rachmawanto, E. H. (2017, November). Tomatoes classification using K-NN based on GLCM and HSV color space. In 2017 international conference on innovative and creative information technology (ICITech) (pp. 1-6). IEEE.

[53] Huang, L., Yang, S. X., & He, D. (2012). Abscission point extraction for ripe tomato harvesting robots. Intelligent Automation & Soft Computing, 18(6), 751-763.

[54] Goel, N., & Sehgal, P. (2015). Fuzzy classification of pre-harvest tomatoes for ripeness estimation–An approach based on automatic rule learning using decision tree. Applied Soft Computing, 36, 45-56.

[55] Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6), 679-698.

[56] Nausheen, N., Seal, A., Khanna, P., & Halder, S. (2018). A FPGA based implementation of Sobel edge detection. Microprocessors and Microsystems, 56, 84-91.

[57] Zhang, F. (2015, March). Ripe Tomato Recognition with Computer Vision. In 2015 International Industrial Informatics and Computer Engineering Conference (pp. 466-469). Atlantis Press.

[58] Benavides, M., Cantón-Garbín, M., Sánchez-Molina, J. A., & Rodríguez, F. (2020). Automatic tomato and peduncle location system based on computer vision for use in robotized harvesting. Applied Sciences, 10(17), 5887.

[59] Zhuang, J., Hou, C., Tang, Y., He, Y., Guo, Q., Miao, A., ... & Luo, S. (2019). Assessment of external properties for identifying banana fruit maturity stages using optical imaging techniques. Sensors, 19(13), 2910.

Electricity Theft Detection using Machine Learning

Ivan Petrlik[1], Pedro Lezama[2], Ciro Rodriguez[3], Ricardo Inquilla[4]
Julissa Elizabeth Reyna-González[5], Roberto Esparza[6]
Faculty of Industrial and Systems Engineering, National University Federico Villarreal, Lima, Perú[1, 2, 6]
Faculty of Electronic Engineering and Informatics, National University Federico Villarreal, Lima, Perú[3]
Faculty of Systems Engineering and Informatics, National University Mayor de San Marcos, Lima, Perú[3]
Faculty of Engineering, National University of Cañete, Cañete, Perú[4]
Faculty of Industrial and Systems Engineering, National University Hermilio Valdizán, Huánuco, Perú[5]

***Abstract*—This research work dealt with the indiscriminate theft of electric power, reported as a non-technical loss, affecting electric distribution companies and customers, triggering serious consequences including fires and blackouts. The research focused on recommending the best prediction model using Machine Learning in electrical energy theft. The source of the information on the electricity consumption of 42372 consumers was a dataset published in the State Grid Corporation of China. The method used was data imputation, data balancing (oversampling and under sampling), and feature extraction to improve energy theft detection. Five Machine Learning models were tested. As a result, the accuracy indicator of the SVM model was 81%, K-Nearest Neighbors 79%, Random Forest 80%, Logistic Regression 69%, and Naive Bayes 68%. It is concluded that the best performance, with an accuracy of 81%, is obtained by using the SVM model.**

Keywords—Energy theft; non-technical losses; machine learning; support vector machine

I. Introduction

In the world, 70% of electricity consumption is lost and 30% in the Caribbean and South America, of which Peru stands out with 7% according to the Inter-American Development Bank [1]. Electricity losses are categorized into two categories: energy delivered to customers (unpaid energy) and losses generated in transmission and distribution lines, which are inherent to electricity transmission. Likewise, non-technical losses comprise the majority of losses in electricity networks and can account for more than 40% of the total electricity produced [2]. These types of losses are attributed to different sources, the most important and common being the alteration of metering equipment, illegal connections to the electrical grid, and energy theft [3]. Regarding distribution in Peru, annually, electricity theft generates losses of 103 million soles, equivalent to 207 GWh, for the companies providing the service [4]. However, this type of loss not only affects these companies but also the offenders themselves and people in the surrounding area, causing various accidents such as electric shocks, fires, and power outages.

According to [5], as presented in Table I, a division is made into countries, utilities, and society, which are categories represented in non-technical effects or consequences in which electric power has many losses.

The background of the respective research is based on multiple studies that have been conducted in many countries, designing intelligent systems that help to deal with this problem, mainly using Machine Learning techniques, which will be presented below:

In 2018, wide convolutional networks (CNNs) were used for one-dimensional data, and deep convolutional networks (CNNs) were used for two-dimensional data. The one-dimensional data were converted into two-dimensional electricity consumption data [6]. On the other hand, a study was carried out in which SVM was applied, using customer consumption data and the total energy distributed by the supplier, which allowed the calculation of the errors produced by electricity meters [7].

In 2019, a combination of neural networks, employed for the conversion of a one-dimensional dataset into a two-dimensional one, and random forests were used to perform customer classification [8]. In 2020, the k-nearest neighbors algorithm and empirical mode decomposition were used to extract the most important attributes from the dataset and obtain good accuracy in detecting energy theft [9]. Another study used a text convolutional neural network (Text-CNN) to effectively extract periodic features about energy consumption and detect electricity theft [10].

In 2021, several classification algorithms were compared, the main one being lightGBM, a fast algorithm based on decision trees, which achieved an accuracy of 84% [11]. Other algorithms compared are logistic regression, with an accuracy of 71%, stochastic gradient descent, with an accuracy of 65%, and decision tree, with 86%.

TABLE I. THE MAIN CONSEQUENCES OF NON-TECHNICAL ASPECTS OF POWER THEFT

Countries	Utilities	Society
Increased use of scarce natural resources	Negative impact on the economy and finances	Total or partial outages
Increased contamination	Reduction of power plant efficiency	Increase in electricity rates
Increased use of public funds	Reduced capacity to upgrade the power system	Fires

According to [12], they proposed that for feature learning (classified into theft and non-theft), a deep convolutional neural network was used. Smart counters at different epochs provided data that was used for SVM training. The time interval of 15 minutes that the smart meter had to record the data through a source coming from the residential and industrial sector which is comprised of 26530 consumers which is the product of data collection.

In the study of [13], the authors evaluated 23 classifiers using the F1 score as a performance parameter. They used as a basis the data of a Brazilian company oriented to the electric power industry, with 261,489 consumers, with approximately 1400 fields. From the results obtained, they concluded that the classifiers (ensemble methods) are the most appropriate, allowing the identification of non-technical cases of electric power loss. The F1 score of 0.45 is the result of the gradient boosted three and an accuracy of 66.50% (actual field inspections) with respect to the rotation forest.

II. Materials and Methods

The procedure that was applied as a solution for electricity theft detection encompassed in the respective workflow is basically made up of five parts: data set acquisition, preprocessing, data balancing, feature extraction, classification, and acquired data set, shown in Fig. 1.

According to Fig. 1, the parts of the workflow will be detailed as follows:

A. Dataset Acquisition

The method of data collection was done through smart meters. The data comes from the daily consumption of electric energy belonging to the State Grid Corporation of China (http://www.sgcc.com.cn/), which was founded on December 29, 2002, and which supplies more than 1.1 billion inhabitants, covering 88% of the national territory. The description of the dataset used is presented in Table II.

According to Table II, we have the temporality range of the data, which comprises from January 1, 2014 to October 31, 2016 (approximately 147 weeks). The file size is 167 MB (175,194,613 bytes) in csv format , with respect to the data structure of the dataset is divided into customers who steal electricity amounting to 3615 (8.55%) and normal customers who consume electricity amounting to 38757 (91.5%) of which add up the total amount of records in 42372 (total customers).

B. Preprocessing

Usually, the electricity consumption represented by the dataset is constituted in some cases by erroneous and missing values and this is caused by problems in smart meters, storage with many problems, unreliability in transmission of metering data and others [20]. For the recovery of missing values in the content of the dataset of the respective research is the interpolation method [21] which is represented through the following formula 1:

$$f(x_i) = \begin{cases} \frac{(x_{i+1} + x_{i-1})}{2} & if\ x_i \in \text{NaN},\ x_{i-1}\ and\ x_{i+1} \notin \text{NaN} \\ 0 & if\ x_i \in \text{NaN},\ x_{i-1}\ or\ x_{i+1} \in \text{NaN} \\ x_i & if\ x_i \notin \text{NaN}. \end{cases} \quad (1)$$

Where:

x_i : Attribute of electricity consumption data

NaN: Non-numeric value

Next, the technique for the recovery of missing data was applied, using the average electricity consumption of each customer for which missing values were substituted. In addition, outliers were found, very different from the rest, which were restored using the equation of the three sigma rule, shown in the respective formula 2

$$f(x_i) = \begin{cases} avg(x) + 2 * std(x) & if\ x_i > avg(x) + 2 * std(x) \\ x_i & otherwise \end{cases} \quad (2)$$

where:

std(x) : typical deviation

avg(x): mean value of x

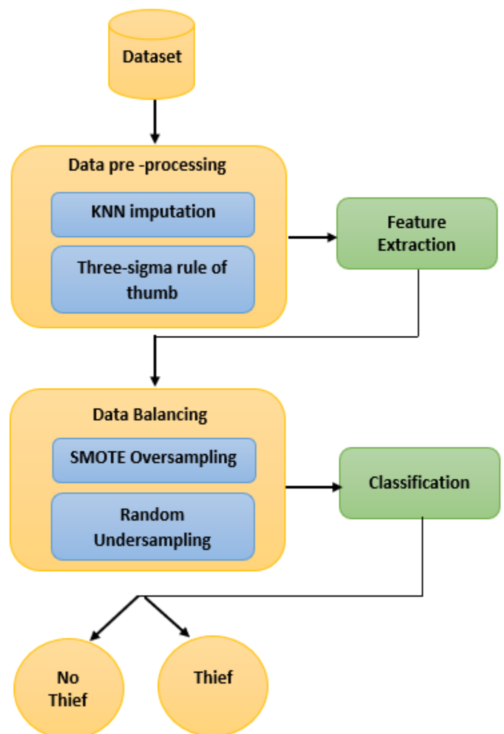


Fig. 1. Arch Workflow.

TABLE II. Description Del Dataset Utilizado

Description	Value
Temporal range of data	01/01/2014 – 31/10/2016
Dataset file size	167 MB (175,194,613 bytes)
Normal customers consuming electricity	38757 (91.5%)
Customers stealing electricity	3615 (8.55%)
Total customers	42372
Cases with missing data	Approximately 25%.

C. Feature Extraction

In order to classify the consumers, characteristics were extracted from their electricity consumption records. The characteristics used were the following: mean, standard deviation, peak to peak, skewness, median absolute deviation, entropy, and kurtosis.

D. Data Balancing

The dataset being used has been found to be imbalanced, with a greater amount of data representing people who are not stealing electricity compared to those who are stealing, which complicates the classification process. To balance the data, techniques such as oversampling and undersampling can be applied. For oversampling, a technique called SMOTE was often used, in which new instances are synthesized from other instances using the k-Nearest Neighbors technique [14]. However, it is suggested to use a subsampling technique in conjunction with the SMOTE technique [15]. For this research work on the dataset, the random subsampling technique was applied, dividing the data into disjoint training and test sets that are randomly partitioned several times [22].

E. Classification

According to [23], the classification process allows to obtain different classes, but based on a grouping of outputs through one or more input variables. In the research, a set of algorithms were applied for this purpose, each of which will be detailed below:

The SVM algorithm is designed to find the optimal separating hyperplane between classes based on support vectors (extremes of the class distributions). The training data are separated into classes using boundaries, which results in the maximization of the distance between the various data sets and the boundary [16].

The training dataset, consisting of n cases represented by {x_i, y_i}, i = 1...n, where y_(i)∈{1,-1}, is used to form a classifier for accurate generalization. A hyperplane is defined as:

$$w * x_i + b = 0 \tag{3}$$

Where there is a normal vector denoted by w and a point x, where both are in the hyperplane and b is the bias. And each point in the sample must satisfy:

$$y_i(w * x_i + b) > 1 \tag{4}$$

The k-Nearest Neighbors algorithm is a supervised classification algorithm that classifies or predicts based on proximity, which is calculated using various distance metrics [17]. In this study, we will use the Manhattan distance, defined as:

$$Manhattan\ Distace = d(x, y) = \sum_{i=1}^{m} |x_i - y_i| \tag{5}$$

Random forest is a supervised classification and regression algorithm that performs well on classification problems. It builds a set of decision trees and bases the final output on majority voting in classification problems. The decision tree algorithm for regression and classification is constructed by evaluating questions and node splits, which contribute to the further reduction of Gini impurities when answering [18].

Logistic regression is a classification algorithm that aims to predict or explain the values of a qualitative target variable as a function of a set of qualitative or quantitative explanatory variables. It is an extension of linear regression that uses the logit function for qualitative classification [19]. The logit function is defined as:

$$logistic(\eta) = \frac{1}{1 + \exp(-\eta)} \tag{6}$$

Following the calculation of the conditional probabilities of which one event occurs with respect to the other, is the concept of Bayes Theorem of which naive bayes is a classification algorithm which is defined by the following respective formula:

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)} \tag{7}$$

III. Results

The results were obtained based on the preprocessing of the data, totaling 33,009 instances, of which 20% were used for testing and 80% for training to predict the respective model to be compared. The processed dataset is shown in Table III.

TABLE III. PROCESSED DATASET

Characteristic	Value
Total Instances	33009
Train Instances	26407
Test Instances	6602

A. Support Vector Machine Algorithm

After experimenting with different kernels to determine the optimal kernel for classification using the Support Vector Machine, the results of this experiment are shown in Table IV. The RBF (Radial basis function) kernel was chosen.

TABLE IV. COMPARISON BETWEEN THE ACCURACIES OBTAINED USING DIFFERENT KERNELS

Kernel	Accuracy
Linear	75%
Polynomial	67%
RBF	80%
Sigmoid	72%

The parameters chosen were "gamma": 0.5 and "C": 100, obtaining an accuracy of 81%. Fig. 2 and Table V show the results through the classification report and the confusion matrix as follows.

B. K-Nearest Neighbors Algorithm

Using the Manhattan metric, the best number of neighbors for this classification was 5, as shown in Fig. 3, obtaining an accuracy of 79%. The results obtained in the confusion matrix and ranking report using these two parameters are defined in the following graphs (as seen in Fig. 4 and Table VI).

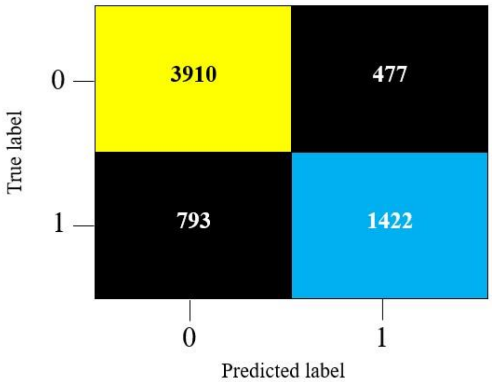


Fig. 2. Confusion matrix of the SVM based detection model.

TABLE V. SVM ALGORITHM CLASSIFICATION RESULT (DETECTION MODEL)

	Precision	Recall	F1 – score	Support
0	0.83	0.89	0.86	4387
1	0.75	0.64	0.69	2215
Accuracy			0.81	6602
Macro avg	0.79	0.77	0.78	6602
Weighted avg	0.80	0.81	0.80	6602

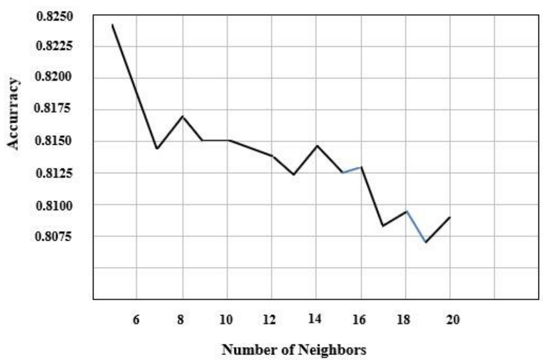


Fig. 3. Accuracies obtained using the manhattan metric and different numbers of neighbors.

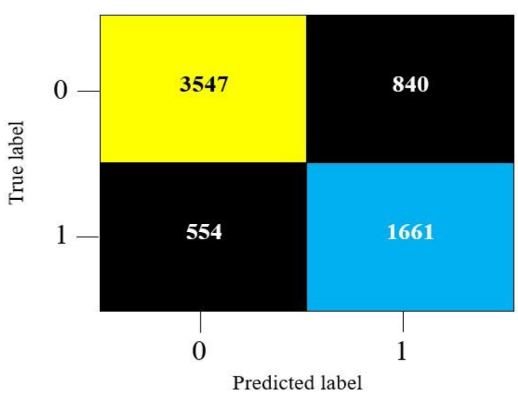


Fig. 4. Confusion matrix of the k-nearest neighbors based detection model.

TABLE VI. CLASSIFICATION RESULT OF THE K-NEAREST NEIGHBORS ALGORITHM (DETECTION MODEL)

	Precision	Recall	F1 – score	Support
0	0.86	0.81	0.84	4387
1	0.66	0.75	0.70	2215
Accuracy			0.79	6602
Macro avg	0.76	0.78	0.77	6602
Weighted avg	0.80	0.79	0.79	6602

C. Random Forest Algorithm

The parameters chosen were 'max_depth': 20, 'n_estimators': 100, 'max_figures': 'auto', 'criterion'='entropy', achieving an accuracy of 80%. The following classification report and confusion matrix are determined by the following graphs (Fig. 5 and Table VII).

The following classification report and confusion matrix were generated using the parameters 'max_depth': 20, 'n_estimators': 100, 'max_figures': 'auto', 'criterion': 'entropy', which resulted in an accuracy of 80%. The results are shown in Fig. 5 and Table VII.

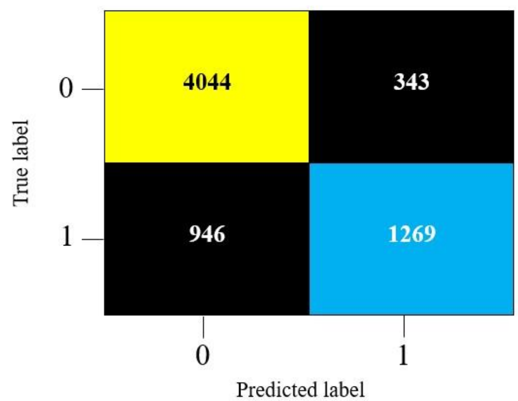


Fig. 5. Confusion matrix of the random forest based detection model.

TABLE VII. RANDOM FOREST ALGORITHM CLASSIFICATION RESULT (DETECTION MODEL)

	Precision	Recall	F1 – score	Support
0	0.81	0.92	0.86	4387
1	0.79	0.57	0.66	2215
Accuracy			0.80	6602
Macro avg	0.80	0.75	0.76	6602
Weighted avg	0.80	0.80	0.80	6602

D. Logistic Regression Algorithm

This model was trained using 1000 iterations and the inverse of the regularization strength 'C' as 10, obtaining an accuracy of 69%. The classification report and the confusion matrix results were obtained, which are determined through the following graphs (Fig. 6 and Table VIII):

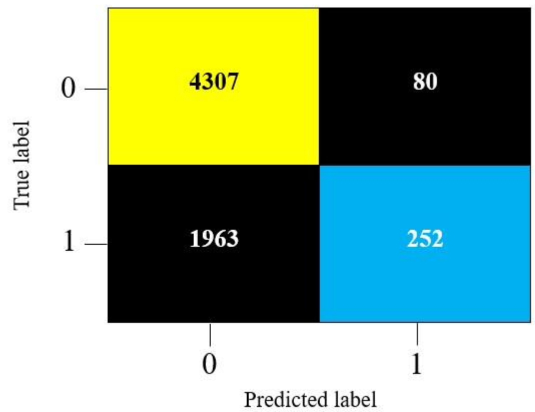


Fig. 6. Confusion matrix of the logistic regression based detection model.

TABLE VIII. LOGISTIC REGRESSION ALGORITHM CLASSIFICATION RESULT (DETECTION MODEL)

	Precision	Recall	F1 – score	Support
0	0.69	0.98	0.81	4387
1	0.76	0.11	0.20	2215
Accuracy			0.69	6602
Macro avg	0.72	0.55	0.50	6602
Weighted avg	0.71	0.69	0.60	6602

E. *Naive Bayes Algorithm*

The default parameters for classification were used for this algorithm, obtaining the following results (see Fig. 7 and Table IX).

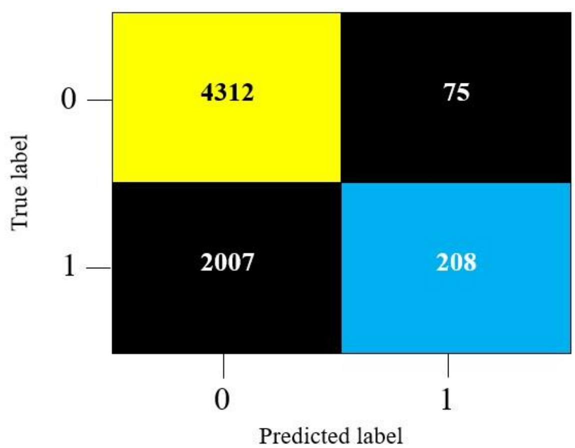


Fig. 7. Confusion matrix of the naive bayes based detection model.

TABLE IX. CLASSIFICATION REPORT OF THE NAIVE BAYES BASED DETECTION MODEL

	Precision	Recall	F1 – score	Support
0	0.68	0.98	0.81	4387
1	0.73	0.09	0.17	2215
Accuracy			0.68	6602
Macro avg	0.71	0.54	0.49	6602
Weighted avg	0.70	0.68	0.59	6602

Table X shows a consolidation of the results obtained from the percentage values of the accuracy indicator of all the proposed models.

TABLE X. LIST OF RESULTS OF THE ACCURACY VALUES OF THE PROPOSED MODELS

Machine Learning Models	Accuracy
Support Vector Machine	81%
Random Forest	80%
K-Nearest Neighbors	79%
Logistic Regression	69%
Naive Bayes	68%

As shown in Table X, the SVM model has a higher accuracy indicator score of 81%.

IV. DISCUSSION

The research of [6] focused on making a comparison of CNN, SVM, LR, RUSBoost models in order to know who has the best prediction. The accuracy result of the SVM model was 0.772, contrasting with our research that also developed a comparison of models such as SVM, RF, KNN, LR and NB, having the best accuracy results of 0.81 for SVM and 0.80 for RF. If we compare the SVM model results of both researches, there is an improvement of 0.038 (3.8%) in favor of the present research. Likewise, the research of [8], also makes a comparison of models such as CNN-RF, CNN-GBDT, CNN-SVM, CNN, SVM, RF, LR and GBDT, the SVM model has an accuracy of 0.77, compared with the present research, achieving an improvement of 0.04 (4%). Next, we have another research by [10], which proposes a new model (TextCNN) for electricity theft detection and also makes a comparison with traditional machine learning models (LR, SVM), the SVM model has an accuracy of 0.70, compared with the present research, achieving an improvement of 0.11 (11%). The SVM model has been compared for all research, however, this model compared with the research [6], which uses the CNN model, results in an accuracy of 0.92 (92%) and the research [10], whose model is Text-CNN whose accuracy value is 0.90 (90%), although it is true that both have better performance, however more computing power is needed when identifying consumers who steal electricity.

V. CONCLUSIONS

This research proposed an electricity theft detection model based on Support Vector Machine using electricity consumption information obtained from the State Grid Corporation of China, achieving a maximum detection accuracy of 81%.

The models have limitations because it was not possible to correctly classify about 25% of the electricity theft cases, which may be due to the lack of data on electricity thieves compared to those who did not steal electricity. However, we attempted to solve this problem using data balancing techniques (oversampling and under sampling).

The experiments conducted show that the system using SVM performs better than most of the other prediction systems tested, such as Logistic Regression, Random Forest, and K-Nearest Neighbors, while the Naive Bayes model does not correctly fit this problem.

REFERENCES

[1] R. Jiménez, T. Serebrisky, and J. Mercado, Power Lost: Sizing Electricity Losses in Transmission and Distribution Systems in Latin America and the Caribbean. Inter-American Development Bank, 2014. doi: 10.18235/0001046.

[2] P. Glauner, P. Valtchev, C. Glaeser, N. Dahringer, R. State, and D. Duarte, “Non-Technical Losses in the 21st Century: Causes, Economic Effects, Detection and Perspectives,” 2018. [Online]. Available: https://www.researchgate.net/publication/325297875.

[3] B. K. Hammerschmitt et al., “Non-Technical Losses Review and Possible Methodology Solutions,” Proceedings - 2020 6th International Conference on Electric Power and Energy Conversion Systems, EPECS 2020, pp. 64–68, Oct. 2020, doi: 10.1109/EPECS48981.2020.9304525.

[4] “Hurto de Energía - enel.pe.” https://www.enel.pe/es/ayuda/hurto-de-energia.html (accessed Nov. 17, 2021).

[5] de S. Savian, J. C. M. Siluk, T. B. Garlet, F. M. do Nascimento, J. R. Pinheiro, and Z. Vale, “Non-technical losses: A systematic contemporary article review,” Renewable and Sustainable Energy Reviews, vol. 147, p. 111205, Sep. 2021, doi: 10.1016/J.RSER.2021.111205.

[6] Z. Zheng, Y. Yang, X. Niu, H. N. Dai, and Y. Zhou, “Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids,” IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1606–1615, Apr. 2018, doi: 10.1109/TII.2017.2785963.

[7] S. C. Yip, W. N. Tan, C. K. Tan, M. T. Gan, and K. S. Wong, “An anomaly detection framework for identifying energy theft and defective meters in smart grids,” International Journal of Electrical Power & Energy Systems, vol. 101, pp. 189–203, Oct. 2018, doi: 10.1016/J.IJEPES.2018.03.025.

[8] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, “Electricity Theft Detection in Power Grids with Deep Learning and Random Forests,” Journal of Electrical and Computer Engineering, vol. 2019, 2019, doi: 10.1155/2019/4136874.

[9] S. Aziz, T. Aslam, S. Zohaib, H. Naqvi, and M. U. Khan, “Electricity Theft Detection using Empirical Mode Decomposition and K-Nearest Neighbors,” 2020.

[10] X. Feng et al., “A novel electricity theft detection scheme based on text convolutional neural networks,” Energies (Basel), vol. 13, no. 21, Nov. 2020, doi: 10.3390/en13215758.

[11] S. V. Oprea and A. Bâra, “Machine learning classification algorithms and anomaly detection in conventional meters and Tunisian electricity consumption large datasets,” Computers and Electrical Engineering, vol. 94, Sep. 2021, doi: 10.1016/j.compeleceng.2021.107329.

[12] U. Haq, J. Huang, H. Xu, K. Li, and F. Ahmad, “A hybrid approach based on deep learning and support vector machine for the detection of electricity theft in power grids,” Energy Reports, vol. 7, pp. 349–356, Nov. 2021, doi: 10.1016/J.EGYR.2021.08.038.

[13] R. M. R. Barros, E. G. da Costa, and J. F. Araujo, “Evaluation of classifiers for non-technical loss identification in electric power systems,” International Journal of Electrical Power and Energy Systems, vol. 132, Nov. 2021, doi: 10.1016/J.IJEPES.2021.107173.

[14] He and Y. Ma, “Imbalanced learning: Foundations, algorithms, and applications,” Imbalanced Learning: Foundations, Algorithms, and Applications, pp. 1–210, Jan. 2013, doi: 10.1002/9781118646106.

[15] N. v. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal Of Artificial Intelligence Research, vol. 16, pp. 321–357, Jun. 2011, doi: 10.1613/jair.953.

[16] Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, “Predicting Student Success Using Big Data and Machine Learning Algorithms,” International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 12, pp. 236–251, Jun. 2022, doi: 10.3991/IJET.V17I12.30259.

[17] “What is the k-nearest neighbors algorithm? | IBM.” https://www.ibm.com/topics/knn (accessed Apr. 26, 2022).

[18] R. Sujatha, S. L. Aarthy, J. M. Chatterjee, A. Alaboudi, and N. Z. Jhanjhi, “A Machine Learning Way to Classify Autism Spectrum Disorder,” International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 6, pp. 182–200, 2021, doi: 10.3991/IJET.V16I06.19559.

[19] T. Hamim, F. Benabbou, and N. Sael, “Survey of Machine Learning Techniques for Student Profile Modelling,” International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 4, pp. 136–151, 2021, doi: 10.3991/IJET.V16I04.18643.

[20] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, “Recovering missing data via matrix completion in electricity distribution systems,” in Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on, 2016.

[21] Zheng, Z.; Yang, Y.; Niu, X.; Dai, H.N.; Zhou, Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Trans. Ind. Informat. 2017, 14, 1606–1615. [CrossRef]

[22] L. Rokach, Ensemble learning: Pattern classification using ensemble methods (second edition). World Scientific Publishing Company, p.2010, 2019.

[23] R. D. K. Hiran, Machine learning: Master supervised and unsupervised learning algorithms with real examples. New Delhi, India: BPB Publications, 2021.

Machine Learning for Securing Traffic in Computer Networks

Ahmed BaniMustafa[1], Mahmoud Baklizi[2], Khalaf Khatatneh[3]
(Senior Member, IEEE)-Data Science and Artificial Intelligence Department, Isra University, Amman, Jordan[1]
Computer Science Department, Isra University, Amman, Jordan[2, 3]

***Abstract*—Computer network attacks are among the most significant and common threats against computer-wired and wireless communications. Intrusion detection technology is used to secure computer networks by monitoring network traffic and identifying attacks. In this paper, we investigate and evaluate the application of four machine learning classification algorithms for identifying attacks that target computer networks: DDoS, Brute Force Web, and SQL Injection attacks, in addition to Benign Traffic. A public dataset of 80 features was used to build four machine learning models using Random Forest, Logistic Regression, CN2, and Neural Networks. The constructed models were evaluated based on 10-fold cross-validation using Classification Accuracy (CA), Area under the Curve (AUC), F1, Recall, Specificity, and Sensitivity metrics in addition to Confusion Matrix, Calibration, Lift, and ROC plots. The Random Forest model achieved 98% in the CA score and 99% in the AUC score, while the Logistic regression achieved 90% in the CA score and 98% in the AUC score.**

Keywords*—*Machine learning; data mining, cyber security; computer networks; intrusion detection

I. Introduction

Computer networks are pivotal in today's world. They connect people, machines, and systems on various scales. Cisco has estimated that 29 billion devices will be connected through computer networks by 2023 [1].

However, network communication security is subject to being hampered by several novel attacks, which necessitate more effective monitoring, detection, identification, and prevention of network attacks which remains an unsolved challenge in the cyber world. This is due to the variety, complexity, and ever-growing sophistication of the technology utilized in these attacks.

Machine learning has successfully solved similar problems in various domains and applications [2, 3] using several supervised and unsupervised learning algorithms [4-6]. This success can be inspired to solve problems in network security by detecting and identifying various types of intrusion attacks that enable monitoring and prevention of such attacks.

This work investigates and evaluates the use of machine learning technology for detecting and identifying four types of network traffic: DDoS attacks, SQL Injection, Brute Force Web attacks, and Benign Traffic. Four machine learning techniques were applied in this study. These include Random Forest (RF), Logistic Regression (LR), CN2 Rules Inducer, and Neural Networks (NN). The constructed models were then evaluated based on Classification Accuracy (CA), Area Under the Curve (AUC) [7], Precision, Recall, and F1 metrics [8]. A confusion matrix was also created for each constructed model [9]. The Calibration, Lift, and Receiver Operating Characteristic (ROC) curves were used to compare and confirm the validity and robustness of the created models [10].

The dataset in this work was sampled from a publicly simulated big dataset which was created by the Communications Security Establishment (CSE) and the Canadian Institute of Cybersecurity (CIC) for Intrusion Detection Systems (IDS). The dataset was published under the name CSE-CIC-IDS2018 [11, 12]. The original dataset comprises 16 million samples described using 80 features related to the traffic flow on the network [13, 14]. The dataset was acquired using the Amazon Amazon Web Services (AWS) Command Line Tool (CLI) [12] and underwent intensive processing involving intensive filtering, sampling, integration, and randomization procedures. It was then explored to examine its quality, distribution, and potential. The sampled data covers three types of attacks: (1) Distributed Denial of Service (DDoS) Low Orbit Ion Cannon (LOIC) Datagram Protocol (UDP), which is referred to as DDOS LOIC-UDP attacks; (2) Structured Query Language (SQL) Injection attacks; (3) Brute-Force Web attacks in addition to Benign Traffic.

Research Problem: Network attacks are difficult to detect and identify using normal network hardware and software tools. Differentiating normal Traffic from network attacks is a complex task that might create threats to network security and interrupt services.

Research Question: Can machine learning classification algorithms detect and identify network attacks based on the numerous characteristics of network traffic flow?

Research Objectives: Investigate and evaluate the use of four machine learning algorithms for detecting and identifying common types of network attacks. The resulting models can be embedded in the network firewall, proxies, routers, and other security tools, suites, and solutions.

Research Contribution:

1) Creating successful machine learning models that can be used for detecting and identifying types of network attacks.

2) Identifying significant traffic features can be used as predictors for identifying and detecting network attacks.

Section II provides a theoretical framework for the conducted research, while Section III reviews the related work. Section IV describes the dataset, while Section V describes the research methodology applied in the study. Section VI describes the research results, Section VII discusses the obtained results, and Section VIII draws a conclusion and comments on the limitation and the future work which can extend this study.

II. Related Work

This section reviews five of the most relevant works on intrusion detection using machine learning.

A study was reported in [15] which aimed at investigating the use of A Convolutional Neural Network (CNN) and Recurrent Neural Networks for detecting denial service (DoS) attacks using the CSE-CIC-IDS2018 and Knowledge Discovery in Databases (KDD) Cup 1999 data. The Convolutional Neural Network (CNN) model detected the Denial of Service (DoS) attacks using the CSE-CIC-IDS2018 dataset with a CA score of 91.5% and 99% using the KDD Cup 99 dataset. On the other hand, the Recurrent Neural Network (RNN) model could classify the CSE-CIC-IDS2018 with a CA score of 65% and a CA score of 93% based on the KDD Cup 99 dataset.

In [16], Andercut reported applying the K-Nearest Neighbor (KNN) algorithm in identifying attacks from benign Traffic with a CA score of 99% using 9/10 of the files in the CSE-CIC-IDS2018 dataset. However, the model scored a CA score of 72% using one of the files in the dataset.

The Iterative Dichotomiser 3 (ID3) and Naïve Bayes algorithms were used in [17] to detect four attack classes and a normal traffic class using the KDD Cup 99 dataset. The study reported a CA score between 97% and 99% in classifying Probe, DoSm R2L, and Normal classes and 94% in predicting attacks that belong to the U2R class.

A study reported using K-Nearest Neighbor Classifier (KNN), Naïve Bayes, Adaboost with Decision Tree, Support Vector Machine, and Random Forest using CSE-CIC-IDS2018 in identifying portrayal botnet attacks with a reported accuracy of 99% [18].

A study in [30] reported using KNN, Random Forest, and Logistic Regression to identify botnet and infiltration attacks using the CSE-CIC-IDS2018 dataset. The applied algorithms identified the attacks with a CA score of 90%.

The related work analysis shows that most of the surveyed studies were conducted to identify a subset of the attacks in the CSE-CIC-IDS2018 and KDD Cup 99 datasets with a CA score ranging from 65% to 99%. However, most studies that reported high CA scores aimed at identifying only one or two types of attacks.

Furthermore, the performance of the applied machine learning techniques scored better when they were applied to the KDD Cup 99 dataset, which contains much fewer samples and much fewer features than the CSE-CIC-IDS2018 dataset. On the other hand, most of the reported studies used the CA metric to evaluate the performance of their reported models, while some used other metrics only to confirm the CA score.

III. Theoretical Framework

Here we provide a brief overview of the concepts related to network intrusion attacks and the machine learning techniques applied in this study.

A. Network Attacks

Network attacks are ″a set of malicious activities that disrupt, deny, degrade, or destroy information and services in a computer network.″ The attacks are usually carried out by sending streams of data that intrudes to affect the availability, integrity, privacy, and secrecy of the services and data communicated through the targeted network [19]. Four types of computer network traffic are investigated in this work: Benign, DDos-LOIC-UDP, Brute-Force-Web, and SQL Injection attacks.

- Brute-Force Web: attacks that aim at cracking vulnerable computer networks that depend on weak user credentials, which consist of weak usernames and passwords [11, 20, 21].
- Distributed Denial of Service (DDoS): attacks designed to target network or web servers with limited bandwidth by overwhelming them with requests from tens or hundreds of distributed URL addresses [11, 22, 23]. DDOS LOIC-UDP stands for Distributed Denial of Service (DDoS) Low Orbit Ion Cannon (LOIC) Datagram Protocol (UDP).
- SQL Injection: stands for Structured Query Language (SQL) attacks. It works by injecting code into databases to gain unauthorized access to data by executing unsolicited queries or damaging the integrity of your database [11, 24-27].
- Benign Traffic: generated to realistically simulate the normal and usual behavior of human users in a typical and normal computer network [11, 16, 19].

B. Machine Learning

- Machine learning is an artificial intelligence field that aims to mimic the human learning process by creating algorithms fed with data [28]. Machine learning techniques are divided into two categories: supervised and unsupervised. While supervised learning aims to find patterns in pre-labeled data, unsupervised learning depends on finding patterns based on data self-labeling [2, 3, 5].
- Random Forest: An ensembled supervised learning technique for regression, classification, and feature ranking. The Random Forest algorithm builds multiple trees created based on a recursive partitioning approach that divides the feature space into several regions that encapsulate a set observation with relative values of responses. The random-forest algorithm is applied to many variables with hard-to-analyze relationships

[29-31]. Random Forest were applied by recent studies in detecting intrusion attacks [32-35].

- Logistic Regression: A supervised learning algorithm that works by assuming a non-linear relationship between features using a logit function that is used to predict the probability of data belonging to a predefined class which is assigned a value between 0 and 1 and then labeled with the closest value (0 or 1). Unlike linear regression, Logistic Regression can be applied to numerical and nominal data [36]. Examples of recent studies that reported the successful use of Logistic Regression in intrusion detection are available in [6, 34, 37].
- CN2: A rules induction algorithm that aims at inducing simple and understandable rules in the form of (if-then-statements). The CN2 algorithm is used for solving classification problems and is known for its ability to handle noise in data [38]. A recent comparative study reported using CN2 in network security [39].
- Neural Networks: A supervised machine learning technique that simulates thinking in the brain. It consists of interconnected neurons organized into a set of layers that input, output, and one or more hidden layers. The algorithm can achieve learning by example by adjusting the triggering weights assigned to the neurons using a sigmoid function and then adjusting through backpropagation. Neural Networks can solve regression and classification problems [40, 41]. Several example studies reported that the neural networks technique was successful in detecting network intrusion attacks [18, 42, 43].

IV. Dataset

The original dataset in this research is a simulated public dataset published by the Communications Security Establishment (CSE) and the Canadian Institute of Cybersecurity (CIC) for Intrusion Detection Systems (IDS) under the name CSE-CIC-IDS2018 dataset. The dataset was collected over a sixteen-day period which extends between 14/02/2018 and 20/03/2018.

The data was labeled with the date of its recording and then stored in a Comma Separated Values (CSV) file corresponding to that date. The dataset can be described as big data as it consists of 16 million records distributed over 80 features which cover: attack type, time stamp, protocols, and number, in addition to 76 other features which are related to the typical flow and Traffic of data in a computer communication network such as flow duration, packet number, bytes number, size of the packet, etc. More details about the original dataset can be found in [13, 14].

The dataset was sampled as a representative sample of the original data. The sampled dataset consists of 1,359 records. It covers three types of network attacks: (1) DDOS LOIC-UDP attacks; (2) SQL Injection attacks; (3) Brute-Force Web attacks in addition to Benign Traffic.

V. Methodology

The method applied in this study consists of five stages that correspond to the typical phases of popular data mining process models [2, 44], which cover: (1) Data acquisition, (2) Data preprocessing, (3) Data exploration, (4) Model construction; and (5) Model evaluation. Fig. 1 illustrates the applied research method and its involved five stages.

A. *Data Acquisition*

The data acquisition procedures involve obtaining the dataset by accessing, downloading, and storing it in the proper file format. The dataset is stored on Amazon AWS and was obtained using the AWS CLI tool [12].

B. *Data Preprocessing*

Due to the vast number of records in the dataset, which consists of 16 million records, the dataset must be treated as big data. The data needs intensive preprocessing procedures, which involve data filtering, randomization, sampling, and combining to construct a representative and informative dataset which can be used to construct useful models that can achieve excellent performance while avoiding both under-fitting and over-fitting through achieving balance in class distribution [5].

C. *Data Exploration*

Data exploration aims at prospecting the dataset and examining its quality, distribution, and potential toward achieving the desired machined learning objectives.

The data exploration procedures involve investigating and visualizing the dataset distribution to confirm the sufficiency of the dataset in each predicted class and to ensure the balance of samples in each class.

The relationship between variables is also visualized and examined using projection plots to uncover relationships between the data set features. The data exploration also involves assessing the importance of dataset features and their contribution toward enhancing the power of the classifier's prediction. Applied variable importance algorithms include the popular information gain (IG) and Gini algorithms.

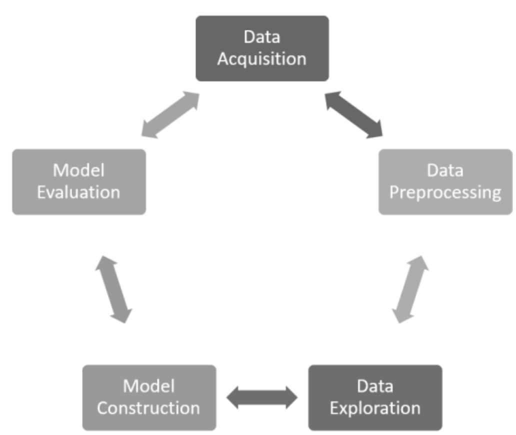


Fig. 1. The five stages of the applied research method.

D. Model Construction

The model construction phase will involve building four prediction models using four classification algorithms which include (1) Random Forest; (2) Logistic Regression; (3) CN2; and (4) Neural Networks.

E. Model Evaluation

The model evaluation phase involves scoring the performance of all the constructed classification models using Classification Accuracy (CA), Area Under the Curve (AUC), and F1 metrics [7, 45]. The equation for calculating the classification accuracy (CA), Precision, Recall, and F1 are described by Equations 1, 2, 3, and 4.

$$\text{Classification Accuracy (CA)}= TP+TN/ N \quad (1)$$

$$\text{Precision}= TP/(TP+FP) \quad (2)$$

$$\text{Recall}= TP/(TP+FN) \quad (3)$$

$$F1= 2TP/2TP+FP+FN \quad (4)$$

Where TP represents the number of samples classified as belonging to the assigned class, TN represents the number of samples classified as not belonging to the assigned class. N is the total number of samples.

In addition to constructing a confusion matrix [9] for each successful model, the performance of all models is visualized and confirmed using the Lift, Calibration, and ROC curves [10].

VI. Results

A. Data Acquisition Results

The CSE-CIC-IDS2018 dataset [13] was downloaded using the AWS CLI tool from the Amazon AWS website [12]. The dataset is managed by the Canadian Institute of Cybersecurity (CIC) [14]. The data was exported and stored in 10 CSV files. Each file was labeled with a name that represents the data acquisition date.

B. Data Preprocessing Results

The data processing procedures involved combining the dataset into one file and then filtering the dataset to contain the dataset records that correspond only to the three classes of attacks considered in this study in addition to a benign class: and excluding all other records.

The resulting dataset was filtered once again and then spitted into four datasets, each representing only one attack classification to the benign class. Each of the four datasets was stored in a separate CSV file which was then loaded and sampled using a stratified random method that involved dividing the dataset into several homogeneous groups. The size of each sample size is a maximum number of 400 records. This number was set to tackle the dataset's size complexity due to the Computer’s limited power in this experiment.

The four resulting datasets were then randomized and then combined in a single dataset that was stored again in another CSV file which contains a total of 1359 records, where 400 records represent each of the DDOS LOIC-UDP, Brute Force Web, and Benign classes, and 159 records represent the entire dataset recorded for the SQL Injection attack class.

C. Data Exploration Results

The data exploration involved examining the dataset quality, its trends and distribution, and finding the correlation and association between the dataset features.

The sampled dataset consists of 1,359 samples distributed over four classes. Each class contains 400 samples, except the SQL Injection, which contains only 159 classes, representing 100% of the samples recorded in the original dataset. The sampled dataset has n and use and no significant outliers. Fig. 2 illustrates the distribution of the dataset over the four assigned classes.

The dataset trends were analyzed using the data projection method, which involves projecting the data samples over four selected dimensions. The projection results show a considerable influence on the time feature and the packet lengths, as shown in Fig. 3. This result was consistent with the variable importance calculated based on the Information Gain and Gini.

The variables were also ranked using Gini and Information Gain (IG) algorithms. Both algorithms agreed on ranking the top four important features, including Fwd Pkt Len Max, Tot Len Fwd Pkts, Subflow Fwd Byts, and Fwd Header Len.

The Gini algorithm ranked the Timestamp feature as number nine, while information ranked it as a number. On the other hand, while Gini ranked the number of the forwarded bytes sent in the initial window as number seven, the Information Gain algorithm ranked it as number nine. Analyzing the variable's importance ranking results was useful in prospecting the dataset's potential in predicting the attack classes. Table I ranks the top-ten features based on Gini and Information Gain variable importance ranking algorithms.

TABLE I. Top-Ten Variable Importance Ranking

#	GINI Ranked Feature	Score	Information Gain Ranked Feature	Score
1	Fwd Pkt Len Max	0.796	Fwd Pkt Len Max	1.56
2	TotLen Fwd Pkts	0.72	TotLen Fwd Pkts	1.43
3	Subflow Fwd Byts	0.72	Subflow Fwd Byts	1.43
4	Fwd Header Len	0.72	Fwd Header Len	1.36
5	**Fwd Pkt Len Mean**	0.70	*Timestamp*	1.33
6	<u>Fwd Seg Size Avg</u>	0.70	**Fwd Pkt Len Mean**	1.30
7	***Fwd Win Byts***	0.68	<u>Fwd Seg Size Avg</u>	1.30
8	Pkt Len Max	0.68	Pkt Len Max	1.29
9	*Timestamp*	0.66	***Fwd Win Byts***	1.26
10	Tot Fwd Pkts	0.62	Tot Fwd Pkts	1.22

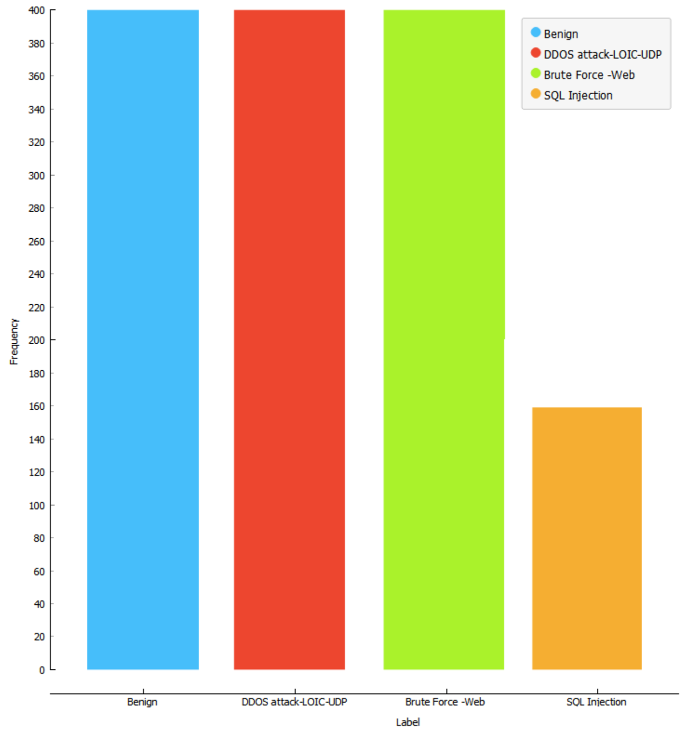


Fig. 2. Distribution of dataset records over the selected four classes.

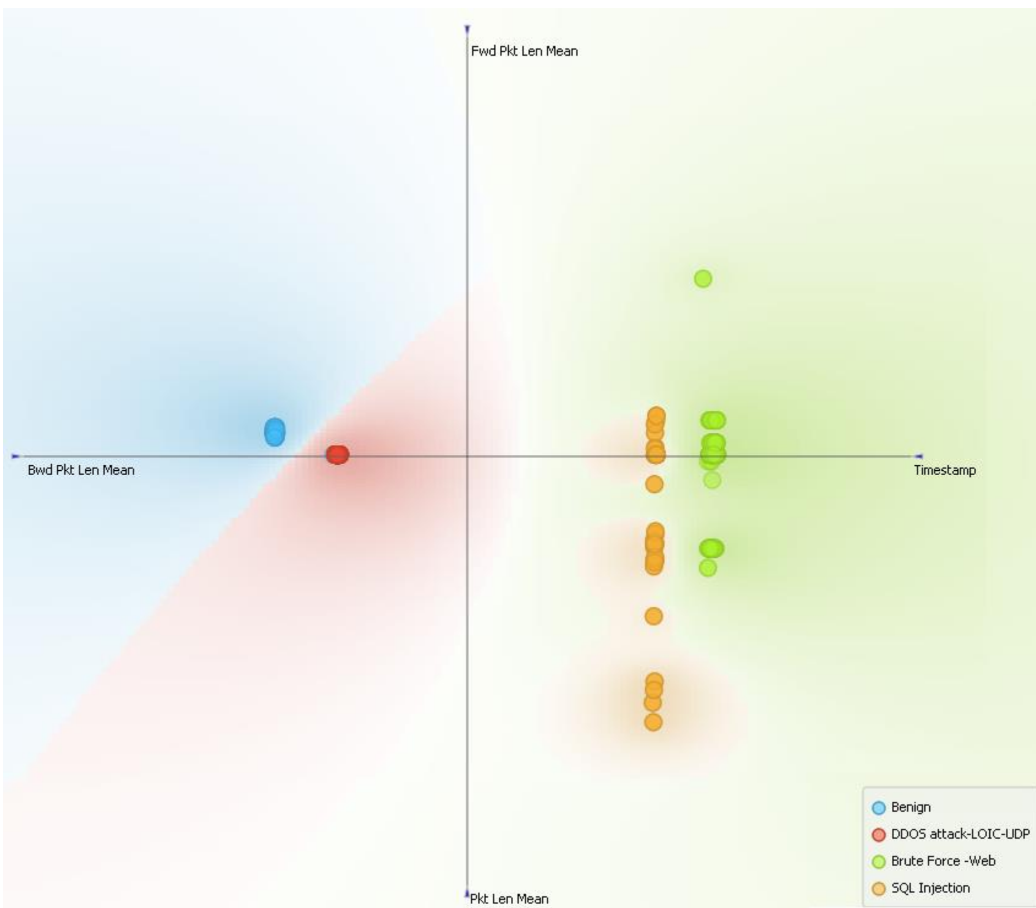


Fig. 3. An example of samples projection over an asset of four features.

D. Model Construction Results

Four models have been constructed in this study using four classification algorithms which include: Neural Networks, CN2, Logistic Regression, and Random Forest. The constructions of the Logistic Regression and Random Forest were the fastest, while the construction of the neural networks was the slowest, and the CN2 model was the second slowest.

E. Model Evaluation Results

The four created models were evaluated using three performance metrics: Classification Accuracy (CA), Area under the Curve (AUC), and F1.

The Random Forest model performed best with a CA score of 98%, an AUC score of 99%, and an F1 score of 98%. The Logistic Regression model scored the second-best performance, scoring 90% in the CA metrics, 98% in the AUC metrics, and 91% in the F1 metrics.

The CN2 achieved satisfactory results with a CA performance of 70%, an AUC performance of 94%, and an F1 performance of 63%, while the neural network model failed to achieve satisfactory results. Table II shows a comparison between the performances of the four constructed models.

The confusion matrices of the three successful models confirm the validity of the models for predicting the classification of all classes except the CN2 model, which failed to predict the Benign class. At the same time, it performed excellently in predicting the three other classes.

The confusion matrix of the CN2 model is illustrated in Table III. It shows that it performed well in predicting all attacks but poorly in predicting benign Traffic. On the other hand, its confusion matrix is illustrated in Table IV; despite the excellent performance of the Logistic Regression model in predicting Benign Traffic, Brute-Force Web, SQL Injection, and DDOS attacks. Shows its inferior performance in predicting Brute-Force Web attacks. The confusion matrix of the Random Forest model shown in Table V was the best, as the model performed excellently in predicting all classes.

TABLE II. CLASSIFICATION MODELS PERFORMANCE

Model	AUC	CA	Precision	Recall	F1
Random Forest	0.99	0.98	0.98	0.98	0.98
Logistic Regression	0.98	0.90	0.94	0.90	0.91
CN2 Rule Inducer	0.94	0.70	0.82	0.71	0.63
Neural Network	0.50	0.27	0.27	0.28	0.27

TABLE III. CN2 MODEL CONFUSION MATRIX

		Predicted				
		Benign	**DDOS attack-LOIC-UDP**	**Brute Force -Web**	**SQL Injection**	**Sum**
Actual	**Benign**	**4**	1	**49**	346	**400**
	DDOS attack-LOIC-UDP	1	**399**	0	0	**400**
	Brute Force -Web	0	0	**298**	2	**400**
	SQL Injection	0	0	4	**159**	**159**
	Sum	**5**	**400**	**447**	**507**	**1359**

TABLE IV. LOGISTIC REGRESSION CONFUSION MATRIX

		Predicted				
		Benign	**DDOS attack-LOIC-UDP**	**Brute Force -Web**	**SQL Injection**	**Sum**
Actual	**Benign**	**399**	0	1	0	**400**
	DDOS attack-LOIC-UDP	0	**399**	0	0	**400**
	Brute Force -Web	0	0	**275**	125	**400**
	SQL Injection	0	0	4	**155**	**159**
	Sum	**399**	**399**	**281**	**280**	**1359**

TABLE V. RANDOM FOREST CONFUSION MATRIX

		Predicted				
		Benign	**DDOS attack-LOIC-UDP**	**Brute Force -Web**	**SQL Injection**	**Sum**
Actual	**Benign**	**398**	0	1	0	**400**
	DDOS attack-LOIC-UDP	0	**400**	1	0	**400**
	Brute Force -Web	0	0	**385**	15	**400**
	SQL Injection	0	0	3	**156**	**159**
	Sum	**399**	**399**	**281**	**280**	**1359**

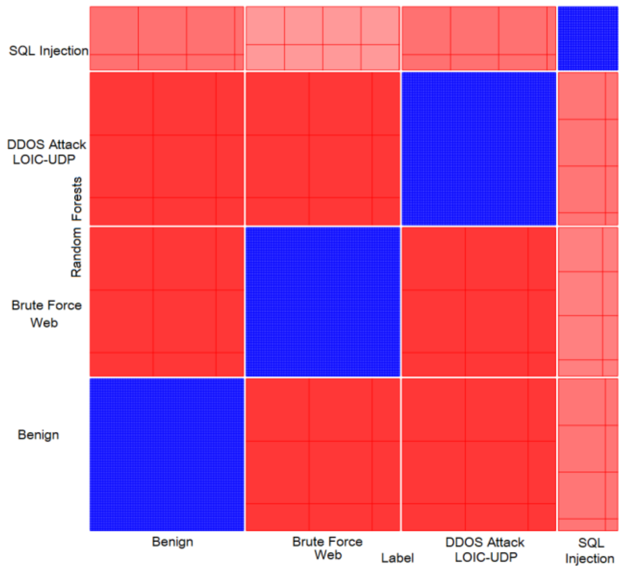


Fig. 4. An example sieve plot that shows the performance of the random.

A sieve plot was constructed for the Random Forest model to visualize the model performance, shown in Fig. 4. The correctly classified samples are represented by blue rectangles. In contrast, the wrongly classified samples are shown in red colors. The size of the rectangles corresponds to the size of the classified samples.

The Calibration curve, shown in Fig. 5 also confirms the validity and robustness of the constructed models. The closest the curve to the logistic function curve is, the better. While the Random Forest and Logistic Regression both show excellent performance, the performance of the CN2 model is relatively modest. In contrast, the performance of the Neural Network model was inferior as it appears as a straight flat line in the plot.

The Lift curve shows the relation between the predicted positive samples and those positive. The Lift curve in Fig. 6 confirms the validity of the Random Forest and Logistic Regression models. They both have excellent curves, and while the CN2 model shows a fluctuating curve, the neural network model performance was poor. This is demonstrated by their excellent lift curves showing the relationship between the predicted positive samples and those that are positive. While the CN2 model shows a fluctuating curve, the neural network model performance was inferior.

The ROC curve in Fig. 7 confirms the robustness and consistency of the performance of the constructed models, where the Random Forest model achieves the best performance. The region under its curve covers a large portion of the chart, followed by the Logistic Regression and CN2 model. However, the performance of the neural networks model was poor as its ROC curves pass through the baseline of the middle region, which indicates that its performance matches the performance achieved by a random model.

The region under the curve of the Random Forest model is the largest in the ROC chart, followed by the Logistic Regression and the CN2 models, respectively. The curve of the neural networks model was found to be poor as its performance curve passes through the baseline of the ROC, which matches the performance of a random stochastic model.

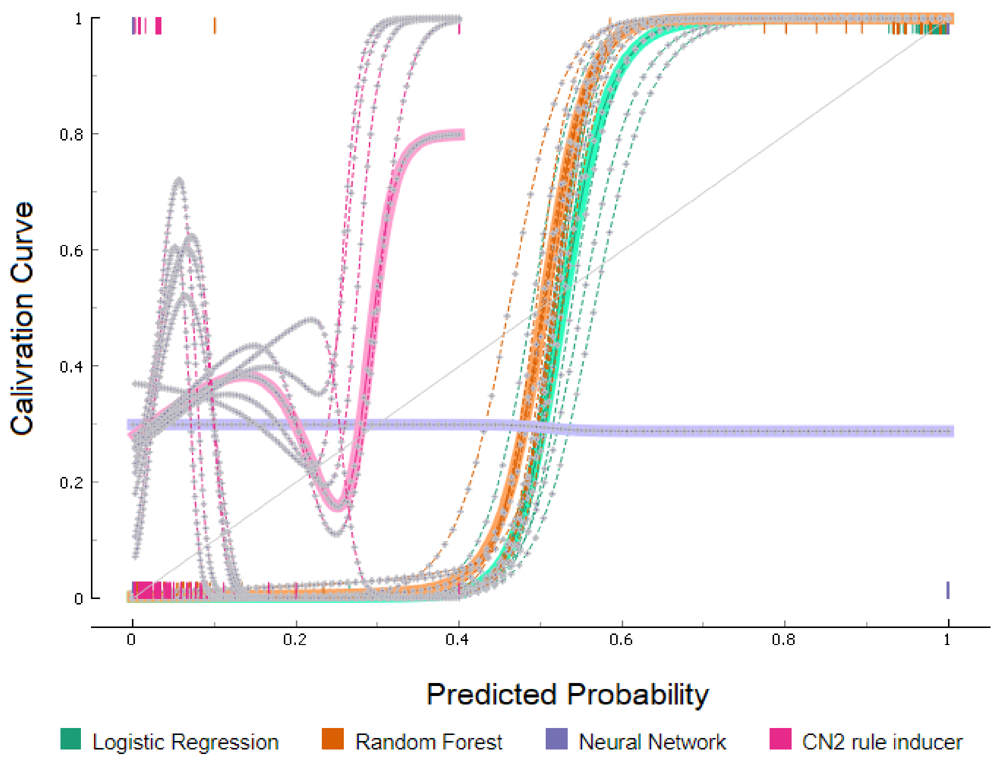


Fig. 5. Calibration curve shows the performance of the four constructed classification models.

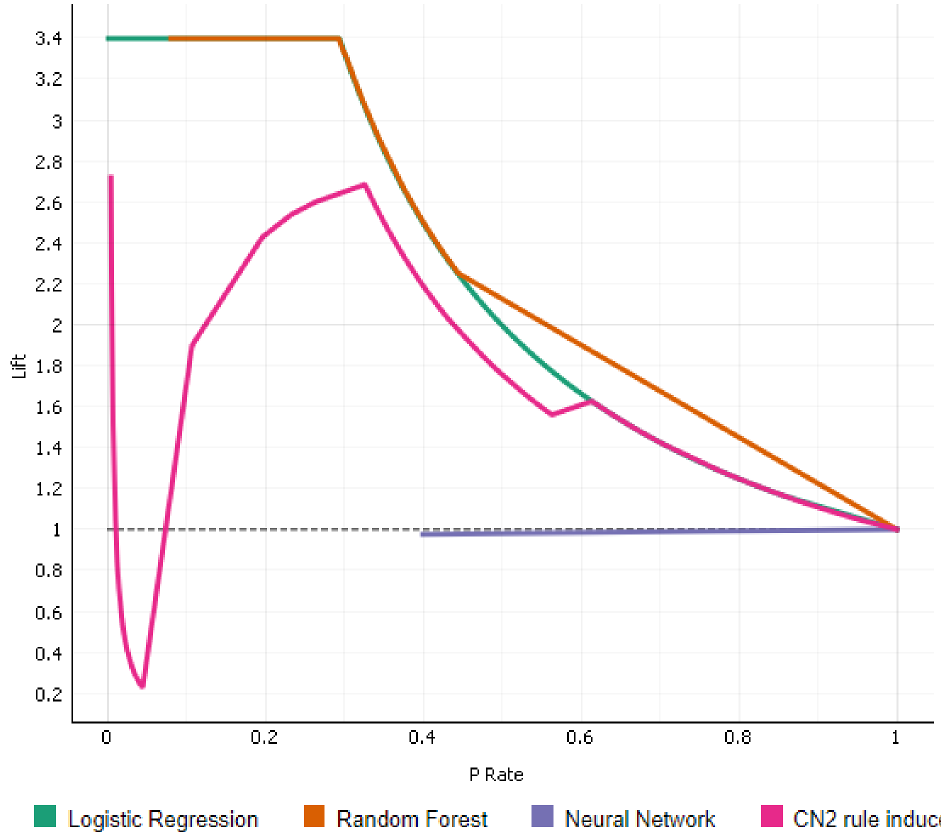


Fig. 6. A lift curve that shows the performance of the four constructed classification models.

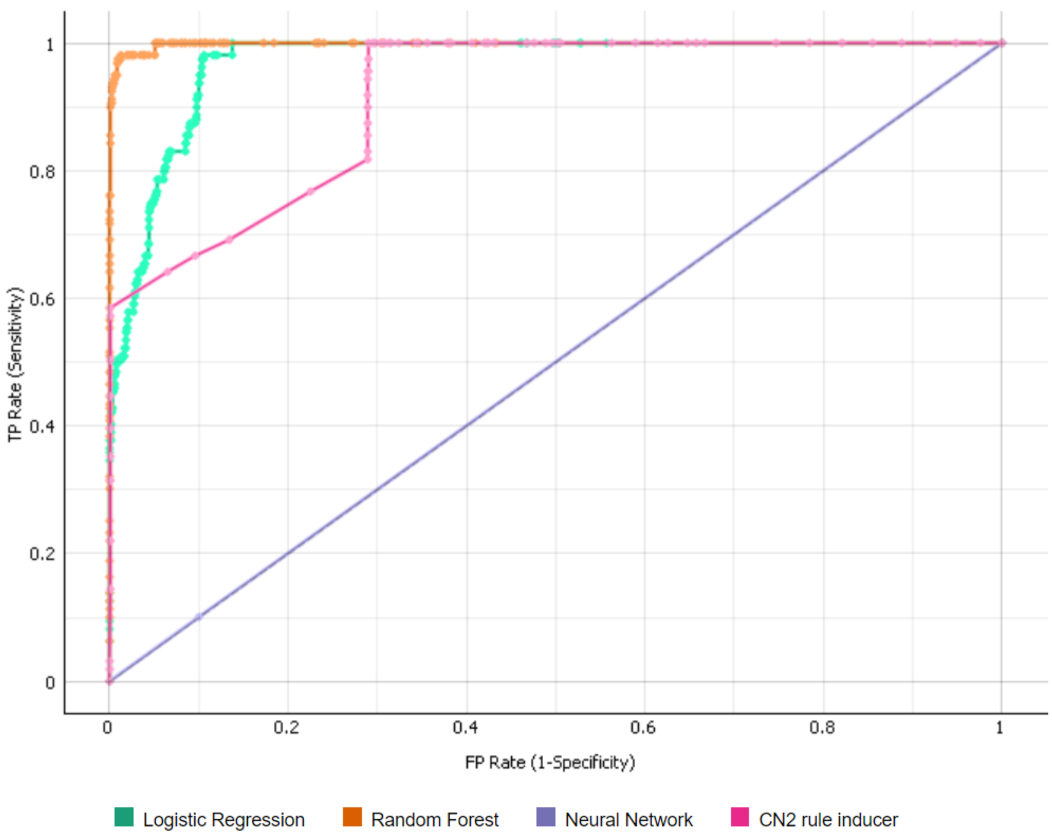


Fig. 7. The ROC curve of the constructed models.

VII. Discussion

The results of this study confirmed the validity of the applied machine learning algorithms for detecting and identifying three classes of network attacks and a benign class depending on several features related to computer traffic.

Furthermore, the variable importance ranking results conducted using GINI and Information Gain successfully identified Fwd Pkt Len Max, Tot Len Fwd Pkts, Subflow Fwd Byts, and Fwd Header Len. As the most prominent features for identifying network attacks. This result is consistent with the data exploration stage findings and the network security domain knowledge discussed in the literature [18, 46].

The Random Forest classifiers scored a CA score of 98%, while the Logistic Regression model scored a CA score of 90%. The Random Forest model achieved an AUC of 99%, while the Logistic Regression model achieved 98%. On the other hand, the CN2 model scored a modest CA performance of 70%, while the Neural Network model failed with a CA score of only 27%. The robustness of the Random Forest and Logistic Regression algorithms in creating the two most successful models was confirmed by the excellent performance shown in the confusion matrix for the two models and the results of the Lift and Calibration curves.

When comparing the results of this study with the results of other studies reported in the literature, in this respect, the results reported in this study outperformed all the excellent results reported in the surveyed literature. While most of the surveyed studies which achieved high CA Scores focused on identifying one or two types of network attacks. For example, the surveyed study reported in [15] aimed at detecting denial of service (DoS) attacks, while the study reported in [16] aimed at distinguishing benign Traffic from network attacks. The study reported in [30] aimed to distinguish botnets from infiltration attacks, while the study reported in [18] aimed to identify portrayal from botnet attacks. In comparison, the Random Forest model that was created in this work was successful in identifying four classes of network traffic: (1) Brute-Force Attacks; (2) Distributed Denial of Service Attacks (DDoS); (3) SQL Injection Attacks; and (4) Benign Traffic with an excellent 99% CA accuracy and 98 AUC.

In addition, the result of this study is quite significant when considering the type of dataset used in this study. This study was applied using a sample of the CSE-CIC-IDS2018 dataset, which is more challenging than the KDD Cup 99 that was used in some of the studies as reported in [15, 17] since CSE-CIC-IDS2018 is larger and it also has more features when compared to the KDD Cup 99 dataset.

However, the limitations of this study come from its dependence on using simulated data rather than real-world ones. In addition, the other important constraint in this study was the limited computational power of the standard PCs, which caused some issues related to the long execution time while experimenting.

VIII. Conclusion

This work has successfully investigated and evaluated the use of four machine learning algorithms: Random Forest, Logistic Regression, CN2, and Neural Networks for detecting and identifying three types of network traffic attacks: DDoS, Brute-Force Web, and SQL Injection, in addition to Benign Traffic.

The study provided a positive answer to the research question and successfully achieved the research objectives. Four machine learning algorithms were applied to a randomized sample of the CSE-CIC-IDS2018 dataset: Random Forest, Logistic Regression, CN2, and Neural Networks. Random Forest scored a CA score of 98%, while Logistic Regression scored 90%. Random Forest scored 99% in the AUC metric, while Logistic Regression scored 98%.

The results obtained in this study contribute towards solving some of the most important problems in computer network traffic and cyber security. These contributions can be summarized as (1) Creating successful machine learning models that can be deployed to detect and identify three types of network attacks based on network traffic data; (2) Identifying significant network traffic features that can be used as predictors for providing fast, accurate and reliable detection and identification of network attacks. These contributions can contribute towards developing solutions that help for preventing, monitor, and mitigate harmful computer network attacks, which affect the reliability, efficiency, and availability of the services provided by computer networks.

The practical use of the successful results that are reported in this study implies deploying the two most successful models reported in this study: Random and Logistic Regression, by embedding them in a network router, security suit, proxy, or firewall system in order to detect, identify or filter network traffic that creates a potential threat to the network.

Future work could be conducted to cover more machine learning techniques and more types of network attacks. In addition, a more real-world dataset can also be used rather than depending only on a simulated one. In addition, more powerful computers can also be used to conduct the same or a comprehensive study on the entire dataset rather than on a sample of it.

References

[1] U. Cisco, "Cisco annual internet report (2018–2023) white paper," CISCO Public, vol. 10, 2020.

[2] A. Banimustafa and N. Hardy, "A Scientific Knowledge Discovery and Data Mining Process Model for Metabolomics," IEEE Access, vol. 8, pp. 209964-210005, 2020.

[3] A. BaniMustafa, "Predicting Software Effort Estimation Using Machine Learning Techniques," in 2018 8th International Conference on Computer Science and Information Technology (CSIT), Amman, 2018, pp. 249-256: IEEE.

[4] A. H. BaniMustafa and N. W. Hardy, "A strategy for selecting data mining techniques in metabolomics," in Plant Metabolomics, N. Hardy and R. Hall, Eds.: Springer, 2011, pp. 317-333.

[5] A. BaniMustafa, "Enhancing learning from imbalanced classes via data preprocessing: A data-driven application in metabolomics data mining," ISeCure, vol. 11, no. 3, pp. 79-89, 2019.

[6] E. Y. Güven, S. Gülgün, C. Manav, B. Bakır, and Z. G. J. E. Aydın, "Multiple Classification of Cyber Attacks Using Machine Learning," vol. 22, no. 2, pp. 313-320, 2022.

[7] C. X. Ling, J. Huang, and H. Zhang, "AUC: A Better Measure than Accuracy in Comparing Learning Algorithms," in Advances in Artificial Intelligence, Berlin, Heidelberg, 2003, pp. 329-341: Springer Berlin Heidelberg.

[8] M. J. Zaki, W. Meira Jr, and W. Meira, Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, 2014.

[9] R. Susmaga, "Confusion Matrix Visualization," in Intelligent Information Processing and Web Mining, Berlin, Heidelberg, 2004, pp. 107-116: Springer Berlin Heidelberg.

[10] M. Vuk and T. Curk, "ROC Curve, Lift Chart and Calibration Plot," Metodolo ški zvezki, vol. 3, no. 1, pp. 89-108, 2006.

[11] A. Ferriyan, A. H. Thamrin, K. Takeda, and J. Murai, "Generating Network Intrusion Detection Dataset Based on Real and Encrypted Synthetic Attack Traffic," Applied Sciences, vol. 11, no. 17, 2021.

[12] T. C. I. f. C. (CIC). (2018, 15/10/2022). A Realistic Cyber Defense Dataset (CSE-CIC-IDS2018). Available: https://registry.opendata.aws/cse-cic-ids2018.

[13] C. I. f. C. (CIC). (2018, 15/10/2022). CSE-CIC-IDS2018 on AWS: A collaborative project between the Communications Security Establishment (CSE) & the Canadian Institute for Cybersecurity (CIC). Available: https://www.unb.ca/cic/datasets/ids-2018.html.

[14] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," presented at The International Conference on Information Systems Security and Privacy (ICISSp), Funchal, Madeira, Portugal, 2018.

[15] CNN-based network intrusion detection against denial-of-service attacks, 9, 2020.

[16] M. Andrecut, "Attack vs Benign Network Intrusion Traffic Classification," arXiv preprint arXiv:.07323, 2022.

[17] D. M. Farid, J. Darmont, N. Harbi, H. H. Nguyen, and M. Z. Rahman, "Adaptive network intrusion detection learning: attribute selection and classification," in International Conference on computer systems Engineering (ICCSE 2009), Bangkok, Thailand, 2009, p. TH60000: World Academy of Science, Engineering and Technology (WASET).

[18] V. Kanimozhi and T. P. Jacob, "Artificial Intelligence outflanks all other machine learning classifiers in Network Intrusion Detection System on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing," ICT Express, vol. 7, no. 3, pp. 366-370, 2021/09/01/ 2021.

[19] A. A. Ghorbani, W. Lu, and M. Tavallaee, Network intrusion detection and prevention: concepts and techniques. Springer Science & Business Media, 2009.

[20] R. Hofstede, M. Jonker, A. Sperotto, A. Pras, and s. management, "Flow-based web application brute-force attack and compromise detection," Journal of network, vol. 25, no. 4, pp. 735-758, 2017.

[21] C. Adams, G.-V. Jourdan, J.-P. Levac, and F. Prevost, "Lightweight protection against brute force login attacks on web applications," in 2010 Eighth International Conference on Privacy, Security and Trust, 2010, pp. 181-188: IEEE.

[22] C. Douligeris and A. Mitrokotsa, "DDoS attacks and defense mechanisms: classification and state-of-the-art," Computer networks, vol. 44, no. 5, pp. 643-666, 2004.

[23] A. Chadd, "DDoS attacks: past, present and future," Network Security, vol. 2018, no. 7, pp. 13-15, 2018.

[24] L. K. Shar and H. B. K. Tan, "Defeating SQL Injection," Computer, vol. 46, no. 3, pp. 69-77, 2013.

[25] W. G. Halfond, J. Viegas, and A. Orso, "A classification of SQL-injection attacks and countermeasures," in Proceedings of the IEEE international symposium on secure software engineering, 2006, vol. 1, pp. 13-15: IEEE.

[26] M. Baklizi et al., "A Technical Review of SQL Injection Tools and Methods: A Case Study of SQLMap," vol. 10, no. 3, pp. 75-85, 2022.

[27] A. Almomani et al., "An enhanced online phishing e-mail detection framework based on evolving connectionist system," vol. 9, no. 3, pp. 169-175, 2013.

[28] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, 1959.

[29] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

[30] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," in Ensemble machine learning: Springer, 2012, pp. 157-175.

[31] L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.

[32] M. Choubisa, R. Doshi, N. Khatri, and K. K. Hiran, "A Simple and Robust Approach of Random Forest for Intrusion Detection System in Cyber Security," in 2022 International Conference on IoT and Blockchain Technology (ICIBT), 2022, pp. 1-5: IEEE.

[33] T. Wu, H. Fan, H. Zhu, C. You, H. Zhou, and X. J. E. J. o. A. i. S. P. Huang, "Intrusion detection system combined enhanced random forest with SMOTE algorithm," vol. 2022, no. 1, pp. 1-20, 2022.

[34] N. S. Bhati and M. Khari, "An Ensemble Model for Network Intrusion Detection Using AdaBoost, Random Forest and Logistic Regression," in Applications of Artificial Intelligence and Machine Learning: Springer, 2022, pp. 777-789.

[35] B. Yogesh, G. S. J. T. J. o. C. Reddy, and M. Education, "Intrusion Detection System using Random Forest Approach," vol. 13, no. 2, pp. 725-733, 2022.

[36] R. E. Wright, "Logistic regression," in Reading and understanding multivariate statistics.Washington, DC, US: American Psychological Association, 1995, pp. 217-244.

[37] P. Kanimozhi, T. J. C. Aruldoss Albert Victoire, C. Practice, and Experience, "Oppositional tunicate fuzzy C-means algorithm and logistic regression for intrusion detection on cloud," vol. 34, no. 4, p. e6624, 2022.

[38] P. Clark and T. Niblett, "The CN2 induction algorithm," Machine learning, vol. 3, no. 4, pp. 261-283, 1989.

[39] N. Kumar, U. Kumar, and Applications, "Comparative analysis of CN2 rule induction with other classification algorithms for network security," Multimedia Tools, vol. 81, no. 26, pp. 37119-37135, 2022.

[40] J. A. Anderson, An introduction to neural networks. MIT press, 1995.

[41] C. M. Bishop, "Neural networks and their applications," Review of scientific instruments, vol. 65, no. 6, pp. 1803-1832, 1994.

[42] J. Wei, Y. Chen, Y. Lai, Y. Wang, and Z. J. I. C. L. Zhang, "Domain adversarial neural network-based intrusion detection system for in-vehicle network variant attacks," vol. 26, no. 11, pp. 2547-2551, 2022.

[43] A. Rosay, K. Riou, F. Carlier, and P. J. A. o. T. Leroux, "Multi-layer perceptron for network intrusion detection," vol. 77, no. 5, pp. 371-394, 2022.

[44] A. BaniMustafa, "A Knowledge Discovery and Data Mining Process Model for Metabolomics," PhD, Computer Science Dept., University of Wales, Aberystwyth Aberystwyth, 2012.

[45] J. Novakovic et al., "Evaluation of Classification Models in Machine Learning," vol. 7, pp. 39-46, 2017.

[46] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets, and challenges," Cybersecurity, vol. 2, no. 1, 2019.

Fish Detection in Seagrass Ecosystem using Masked-Otsu in HSV Color Space

Sri Dianing Asri[1], Indra Jaya[2], Agus Buono[3], Sony Hartono Wijaya[4]
Computer Science Department, IPB University, Bogor, Indonesia[1, 3, 4]
Marine Science and Technology Department, IPB University, Bogor, Indonesia[2]

***Abstract*—Seagrass ecosystems are coastal ecosystems with high species diversity, especially fish. Fish diversity determines the abundance of communities based on the number of species. Detection of fish directly (in-situ) and conventionally by catching them requires more energy, costs, and relatively needs time. Therefore a computer vision method is needed that can detect fish well using underwater images. The fish detection model used Masked-Otsu Thresholding, HSV color space with closing techniques in morphological operations. The dataset is in the form of 130 underwater images, divided into 80% training data and 20% testing data. The test results showed a model accuracy value of 0.92, Precision value of 0.84, Sensitivity value of 0.93, and F1 Score of 0.88. With these results, the model could detect fish in the seagrass ecosystem.**

Keywords—Fish Detection; HSV color space; masked-otsu thresholding; morphological operation; seagrass ecosystem

I. Introduction

Seagrass ecosystems are ecosystems in coastal areas with a high diversity of species, especially fish. These varieties are variations in how species interact with each other in their environment. Diversity indices are useful for determining abundance in a community based on the number of species in a location, such as the research of R. Machrizal et al. on the diversity of macrozoobenthos species[1], fish assemblages[2], seahorses species[3], sea cucumber[4]. Detection and classification of fish in seagrass ecosystems are done manually by fishing, photographing, recognizing, and recording in the ledger. This technique is quite good but takes a long time and costs a relatively high amount. Therefore, computer vision techniques and digital image processing will be very helpful.

The underwater environment is more complex than the above-water surface. Sunlight entering the water will undergo absorption and dispersal of light by water media and floating particles in water. The light absorption and dispersal effect results in color distortion, low contrast, and blur in the resulting underwater image[5]. In addition, the difference in attenuation causes the underwater image to be bluish, greenish, or other colors. In addition, the image contains foreground and background caused by the placement of the camera, which is difficult to get close to the observation object. That observation object will be smaller than the background image. Caused bias in the object detection process. It is challenging to make underwater observations using both underwater images and videos.

One solution to separate the background to get the desired object is to use an image segmentation technique based on RoI (Region of Interest). One of the RoI segmentation techniques is the Otsu Thresholding method. The otsu method[6] aims to automatically segment the histogram value of the grayish image in two different regions by entering a threshold value. After the background has been successfully separated, this value is used to detect objects.

Fish detection has been carried out [5][7][8][9] using fish data taken on land, namely by taking photos of fish manually or taking data in an aquarium that has good lighting. In addition, research on fish detection using underwater video was carried out by [10][11]. Retrieving datasets in a limited environment (aquarium) results in a simpler background and foreground. Data collection in the original environment, such as in the seagrass ecosystem, has a more complex background, such as the color of fish objects similar to seagrass leaves or water.

This study used primary data from UTS (Underwater Televisual System) equipment on Beralas Pasir Island, Bintan Regency, Riau Islands, Indonesia. In this study, underwater image data is used, with the area of fish observation objects, seagrass leaves, and transects. Analysis of the underwater image dataset used in this study, it can see that fish objects swimming between seagrass leaves have a color that is almost similar to the color of water or seagrass leaves, that the shape of the fish tends to be smaller (narrow object area) compared to the image background, this makes it difficult to detect it, therefore to remove the background of the underwater image and obtain the detection object, then otsu thresholding is used. The use of HSV color space and morphological operation to obtain a range of HSV values used for area detection of fish objects. This study aims to create a fish detection model associated with seagrass beds on Beralas Pasir Island, Bintan Regency, Riau Islands, Indonesia, using Masked-Otsu in HVS color space. By determining research questions as follows:

R1: How to segmentation of fish objects with an underwater image background using the masked-otsu algorithm.

R2: How to detect fish by applying a range of HSV values and object contours through morphology operation.

II. Related Works

Research on fish detection in seagrass ecosystems has been widely practiced. Most of this research is still done manually, namely by catching fish, being photographed, recognized, and recorded in books. The results have been good, but it requires a lot of time and cost [12][13][14][15][16].

Recently, the use of machine learning and deep learning for fish detection has been widely carried out, and extensive resources and longer time consumption have become considerations in this technique. Meanwhile, digital image processing techniques such as segmentation that can separate the background and foreground with less time consumption can be further developed.

Petrelis Nikos [17] in 2021, utilizing image processing and deep learning techniques to estimate the length, height, and area of fish through fish morphological features and CNN methods, fish objects are visible in shape, and the system built has estimation errors of 1.9% and 13.2%. In 2021, Shoffan Saifullah et al. [7] detected fish through morphological operation and K-Mean segmentation, using photos of fish taken on land so that fish objects were visible, obtaining an SSMI distribution value of 0.9994. Heningtyas research et al. in 2020 [8], using the Expectation Maximization (EM) segmentation algorithm, photos of fish objects taken from an aquarium with good lighting obtained an accuracy of 89.14. Kartika et al. in 2016 [9] classified Koi fish using Naïve Bayes and SVM using K-Fold Cross Validation in HSV space; The Koi Fish image was taken on land, and the study results had a success rate of 0.968. Anggraeny research et al. in 2020 [18] detected fish with the Histogram of Oriented Gradients (HOG) algorithm and AdaBoost-SVM, the accuracy of which was achieved at 0.848. In this Studi, we propose a masked-otsu thresholding segmentation technique to separate objects from the background. The object area is divided into 3, namely the fish, seagrass leaf, and transect areas. After the object is separated from the background, fish detection uses upper and lower HSV color space values and morphological operations with the closing technique. The contribution of this study is to detect fish in complex dynamic environments, namely in seagrass ecosystems, by separating the background of underwater imagery to obtain fish objects.

III. Research Methodology

Broadly speaking, this research consists of several stages, namely: Starting from collecting datasets in the form of underwater images with a size of 620x480 pixels as many as 130 RGB image data, Fig. 1 shows the system flowchart, and Fig. 2 shows the original underwater image.

The original underwater image is cropped at the top to remove the color of the water so that the cropping image is 620x330 pixels. This image is later used to detect fish in a seagrass frame. Furthermore, this dataset will be divided into 2, namely Data Training as much as 80% and Data Testing as much as 20%.

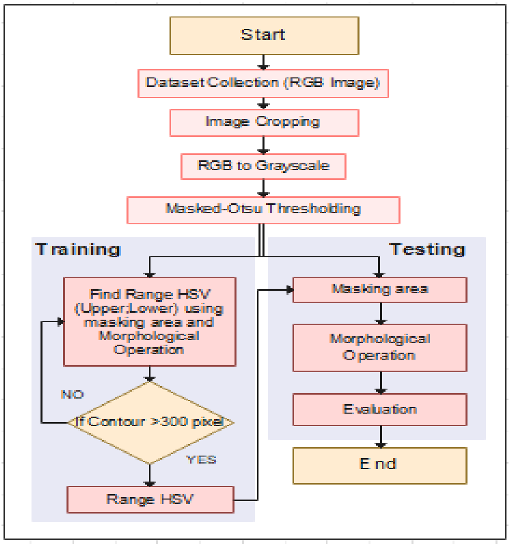


Fig. 1. Research phases.

Fig. 2. Original underwater image.

The third step converted the RGB image to Grayscale. Then, the grayscale image was converted again into a binary image using the Otsu Threshold Method to separate objects (fish) from the background; a threshold value of 120/255 or 0.47 was used.

In the training stage, a search for HSV value range values is performed to detect fish areas, using masking areas and morphological operations. The morphological operation used is a closing operation to close gaps or remove small holes between contours so that it can soften a large object without changing the object significantly.

The equation of the closing operation with the notation A (original image) and B (structuring element) is as follows:

$$A \bullet \boldsymbol{B} = (\boldsymbol{A} \oplus B) \ominus B \qquad (1)$$

Dimension search for detecting objects using the HSV value range is by setting the Upper and Lower values based on the reference object (fish) specified in the thresholding process. The testing stage will be implemented value at range determination is done by paying attention to the area of the object to be detected; if the range is too wide, then unwanted objects will appear; on the other hand, if the range is too narrow, cannot detect the object.

In the testing phase, the masked-otsu threshold result image, given masking, namely the area that is included in the HSV value range that has been obtained previously set as the object dimension. Then the morphological process of closing the area object is carried out; if the contour value is> 300 pixels, it will be marked as fish. The last step is to evaluate by looking for Accuracy, Precision, and Recall values from fish detection models in the seagrass ecosystem. Accuracy, Precision, and Recall using the confusion matrix are as follows:

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \tag{2}$$

$$Precison = \frac{TP}{TP+FP} \tag{3}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{4}$$

$$F1\,Score = 2x\frac{Precision\ x\ Sensitivity}{Precision + Sensitivity} \tag{5}$$

Where TP (True Positive) is the number of fish objects successfully detected, TN (True Negative) is the number of non-fish objects (seagrass and transects) detected as non-fish objects, FN (False Negative) if the fish object is not successfully detected as fish. Finally, FP (False Positive) refers to the object that is not fish (seagrass or transect) and is recognized as fish.

IV. Result and Discussion

The data set used in this study is in the form of an underwater image with a size of 640x480 pixels, as many as 130 RGB image data, after cropping the original image's dimensions to 640x330 pixels. Each image is divided into three sample areas: fish, seagrass, and transect. After the dataset is formed, the next step is to cut the top of the image to remove the color of the water so that later this image is used to detect fish in the seagrass frame. Fig. 3 shows the change of the original image to a threshold image with masked-otsu in RGB space.

The dataset is divided into 80% for the training stage, 104 underwater images, 20% for the testing stage, or 26 images. All of these images will be thresholds, then at the image training stage, as many as 104 are used to train the model so that it can detect fish objects; the HSV value range is obtained as follows:

Lower = [74,83,190]

Upper = [88,99,230]

This value is implemented in the data test, after which masking the area for the object's dimensions is included in the HSV value range. Then a closing operation is carried out to close the gap between the contours so that the corresponding object contour is obtained, namely the contour value > 300 pixels marked as fish. While if the contour value <= 300 will be discarded. Fig. 4 shows the fish detection process starts by changing the RGB image from implementing Otsu thresholding to HSV color space, masking area, and operation morphology until the fish object is successfully detected.

At the training stage, the accuracy level is calculated using Equation (2) with 104 underwater image data covering three area objects: Fish, Seagrass, and Transect. The total area used is 314 areas. Based on the experiments, an accuracy of 0.91 or 91% was obtained, which means that the model can correctly predict the area of objects (fish, seagrass, and transects) (Positive).

Meanwhile, at the testing stage, a dataset of 26 underwater images was used. Fig. 5. Show a confusion matrix for testing data.

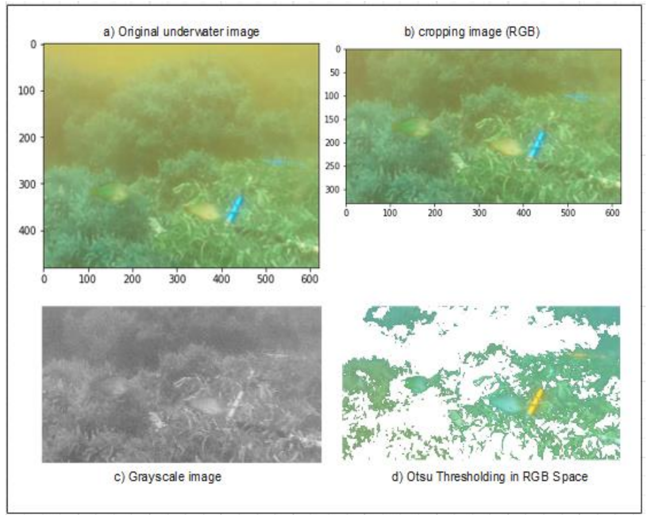


Fig. 3. Change of original Image to threshold image in RGB space.

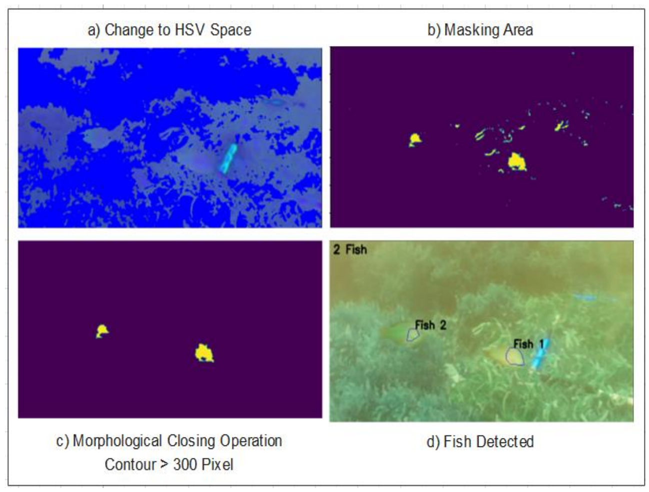


Fig. 4. Fish detection process.

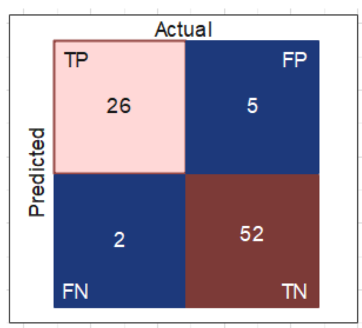


Fig. 5. Confusion Matrix on testing data.

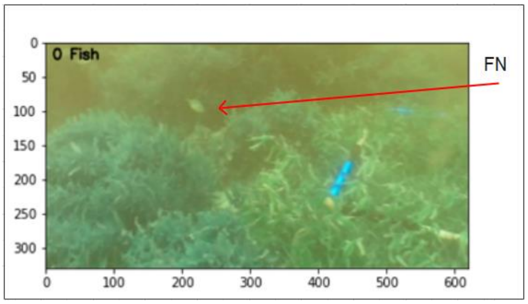


Fig. 6. False Negative (FN) on testing data.

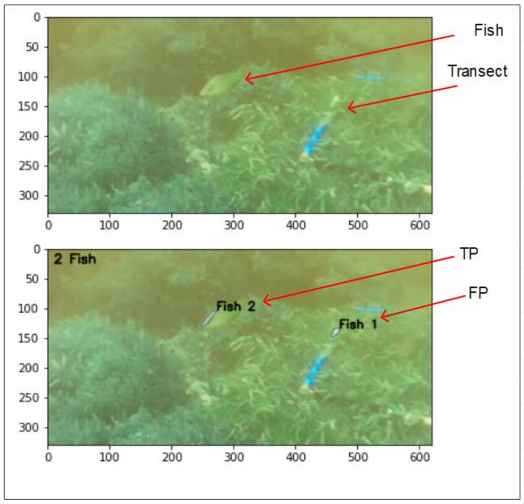


Fig. 7. False Positive (FP) on testing data.

The FN error appears when the model cannot detect the fish object as a fish, as in Fig. 6. Meanwhile, FP occurs because the model states a non-fish object as a fish, where the non-fish object can be seagrass or transect. Fig. 7 shows False Positive.

Based on Table I of the data testing results, there is an increase in the percentage of data testing accuracy values compared to the results of training data experiments by 1%. This increase in accuracy percentage is not significant, so the model created is already optimal. The detection error rate is 0.08 or 8%. This detection error is caused by the presence of FN and FP on the model. The error arises because the area of the fish to be detected is narrow, blurry, and unclear; besides that, the shape and color of the fish are almost the same as the seagrass leaf. Even with ordinary eyes, this fish is difficult to detect.

In addition to the Accuracy value, Table I also shows a Precision value of 0.84 or 84%. This Precision value signifies the reliability of a reliable model for repeated data testing. The sensitivity value is obtained at 0.93 or 93%, which means that the model can detect or give a True value to positive data. Then last is the F1 Score. The F1 Score is 0.88 or 88%, meaning that the model is capable and reliable in detecting fish well.

TABLE I. DETAIL RESULTS OF TESTING DATA

Accuracy	Precision	Sensitivity	F1 Score
0.92	0.84	0.93	0.88

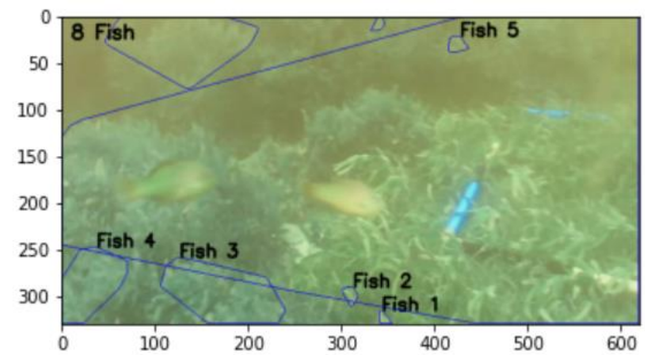


Fig. 8. Fish Detection Model Without HSV.

This study also tried to test fish detection directly, namely from masked Otsu Thresholding directly carried out the morphological process of the operation without using the HSV color space range value. The results are shown in Fig. 8. The picture shows that the detection results could not be better because the error rate that occurs is high, and the detection results are widened.

V. CONCLUSION

Based on the test results on the testing data, an accuracy value of 0.92 was better than the accuracy of the training data of 0.91. It is not significant, meaning that the model built is optimal for detecting fish. In addition to accuracy, the model gets a Precision value of 0.84 which means it can be relied upon for repeated data testing. The Sensitivity value is 0.93, which means that the model can detect or give a correct value to positive data. Then last is the F1 Score. The F1 Score is worth 0.88, meaning that the model is capable and reliable in detecting fish well.

The test results also showed a misdetection of 8%. The error arises because the area of the fish to be detected needs to be narrower, opaque, and unclear. Besides that, the shape and color of the fish are almost the same as the seagrass leaf. Even with ordinary eyes, this fish is difficult to detect. This cause the resulting underwater image to have low contrast and blur. The next study is to create a fish detection model that can capture small fish dimensions with a fish shape or color almost the same as the color of water and seagrass leaf based on video using the GMM (Gaussian Mixture Model) algorithm and morphological Operation. Using datasets in the form of underwater videos considering fish objects will look moving and can be distinguished from the image background. GMM was chosen because this algorithm effectively separates the background and foreground on the input frame sequence compared to Otsu Thresholding.

ACKNOWLEDGMENT

This research was supported by Penelitian Disertasi Doktor (PDD), Contract No. 3792 /IT3. L1/PT.01.03/P/B/2022. Seagrass ecosystem data set used in this paper is obtained from DDRG-LIPI research grant, Contract Number: B-1201/IPK.02/KS/III2018 and Number: 238/KS/00/00/2018 to Prof. Dr. Indra Jaya/FPIK-IPB. The authors also would like to thank Muhammad Iqbal for preparing the data set.

REFERENCES

[1] R. Machrizal, Khairul, and R. H. Dimenta, “Keanekaragaman makrozoobentos pada ekosistem lamun di Perairan Natal Sumatera Utara,” Gorontalo Fish. J., vol. 3, no. 1, pp. 56–67, 2020.

[2] Q. Q. Ren et al., “Fish Assemblages in Subtidal Seagrass Meadows Surrounding the West Sand, South China Sea,” Front. Environ. Sci., vol. 9, no. February, pp. 1–12, 2022, doi: 10.3389/fenvs.2021.765702.

[3] R. Ambo-Rappe, Y. A. La Nafie, and ..., “Seagrass habitat characteristics of seahorses in Selayar Island, South Sulawesi, Indonesia,” Aquac., 2021, [Online]. Available: https://search.proquest.com/openview/57907c68b11eaa761a1e94c190ecbc23/1?pq-origsite=gscholar&cbl=2046424.

[4] G. D. Manuputty, M. M. Pattinasarany, G. V. Limmon, and A. Luturmas, “Diversity and abundance of sea cucumber (Holothuroidea) in seagrass ecosystem at Suli Village, Maluku, Indonesia,” IOP Conf. Ser. Earth Environ. Sci., vol. 339, no. 1, 2019, doi: 10.1088/1755-1315/339/1/012032.

[5] Y. Zhang, F. Yang, and W. He, “An approach for underwater image enhancement based on color correction and dehazing,” Int. J. Adv. Robot. Syst., vol. 17, no. 5, pp. 1–10, 2020, doi: 10.1177/1729881420961643.

[6] N. Otsu et al., “Otsu_1979_otsu_method,” IEEE Trans. Syst. Man. Cybern., vol. C, no. 1, pp. 62–66, 1979.

[7] S. Saifullah, A. P. Suryotomo, and B. Yuwono, “Fish detection using morphological approach based-on k-means segmentation,” Compiler, vol. 10, no. 1, pp. 1–9, 2021, doi: 10.28989/compiler.v10i1.946.

[8] Y. Heningtyas, R. Andrian, A. Junaidi, and S. Susiyani, “Goldfish (Carassius auratus) Segmentation Using Expectation Maximization (EM) Method,” J. Appl. Informatics Comput., vol. 4, no. 2, pp. 107–115, 2020, doi: 10.30871/jaic.v4i2.2387.

[9] D. S. Y. Kartika and D. Herumurti, “Koi Fish Classification based on HSV Color Space,” Proc. 2016 Int. Conf. Inf. Communiation Technol. Syst. ICTS 2016, pp. 96–100, 2017, doi: 10.1109/ICTS.2016.7910280.

[10] A. Jalal, “Fish detection and species classification in underwater environments using deep learning with temporal information,” Ecol. Inform., vol. 57, 2020, doi: 10.1016/j.ecoinf.2020.101088.

[11] A. Salman et al., “Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system,” ICES J. Mar. Sci., vol. 77, no. 4, pp. 1295–1307, 2020, doi: 10.1093/icesjms/fsz025.

[12] M. Jalaludin, I. N. Octaviyani, A. N. Praninda Putri, W. Octaviyani, and I. Aldiansyah, “Padang Lamun Sebagai Ekosistem Penunjang Kehidupan Biota Laut Di Pulau Pramuka, Kepulauan Seribu, Indonesia,” J. Geogr. Gea, vol. 20, no. 1, pp. 44–53, 2020, doi: 10.17509/gea.v20i1.22749.

[13] A. Syukur, L. Zulkifli, A. Al Idrus, and B. N. Hidayati, “Species diversity of seagrass-associated bivalves as an ecological parameter to support seagrass conservation along with the Coastal Waters of South Lombok, Indonesia,” Biodiversitas J. Biol. Divers., vol. 22, no. 11, 2021, doi: 10.13057/biodiv/d221152.

[14] R. Tahapary, H. Tuaputty, S. Liline, T. S. Kurnia, and M. T. Kubangun, “Keanekaragaman Jenis dan Kepadatan Ikan di Padang Lamun Pantai Desa Akoon Kecamatan Nusalaut Kabupaten Maluku Tengah,” no. 2004, pp. 140–146, 2007.

[15] S. Fakhri A, I. Riyantini, D. J. P, and H. Hamdani, “Korelasi Kelimpahan Ikan Baronang (Siganus Spp.) Dengan Ekosistem Padang Lamun Di Perairan Pulau Pramuka Taman Nasional Kepulauan Seribu,” J. Perikan. Kelaut., vol. VII, no. 1, pp. 165–171, 2016.

[16] S. Prisilia, W. Adi, and A. Febrianto, “Fish Communities Structure On Seagrass Ecosystems In Puding Beach Bangka Selatan . Keywords : Fish , Seagrass , Fish Relationship with Seagrass,” pp. 35–44, 2018.

[17] N. Petrellis, “Measurement of fish morphological features through image processing and deep learning techniques,” Appl. Sci., vol. 11, no. 10, 2021, doi: 10.3390/app11104416.

[18] F. T. Anggraeny, B. Rahmat, and S. P. Pratama, “Deteksi Ikan Dengan Menggunakan Algoritma Histogram of Oriented Gradients,” Inform. Mulawarman J. Ilm. Ilmu Komput., vol. 15, no. 2, p. 114, 2020, doi: 10.30872/jim.v15i2.4648.

Prioritizing the Factors Affecting the Application of Industry 4.0 Technology in Electrical Appliance Manufacturing using a Fuzzy Analytical Network Process Approach

Apiwat Krommuang, Atchari Krommuang
King Mongkut's Institute of Technology Ladkrabang Business School
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

Abstract—**The fourth industrial revolution is a technological advancement that is posing new challenges in manufacturing and services. Industries must adopt innovations to create value-added for their products and services to gain a competitive advantage and increase production efficiency. Therefore, this research aims to study the factors that influence the application of Industry 4.0 technology for managing electrical appliance production by focusing on five major factors: the internet of things, cloud manufacturing, big data analytics, additive manufacturing, and cyber-physical systems, which can be further subdivided into 23 sub-factors. The fuzzy analytic network process (FANP) technique is used to prioritize the factors to develop criteria for selecting appropriate applications of Industry 4.0 technology in manufacturing. Besides, a questionnaire based on the FANP approach is used to collect data from 82 electrical appliance manufacturers to calculate the weight of each factor. Consequently, the Internet of Things is ranked first, followed by big data analytics and additive manufacturing. While the results have indicated the importance of sub-factors as data-driven, data collection, tracking, monitoring, and automation, respectively. The benefit of this research is that manufacturers of electrical appliances can use this research as a criterion for implementing Industry 4.0 technology for long-term effectiveness.**

Keywords—Industry 4.0; big data analytics; internet of things; cloud manufacturing; additive manufacturing; cyber-physical systems; fuzzy analytic network process

I. Introduction

The manufacturing industry is currently undergoing a transformation in innovation and technology, which is pushing successful output. Connectivity and data analytics advancements are a part that can help the manufacturing industry create value-added and competitive advantages for its products and services. Therefore, integrated production processes are critical in the industrial revolution to enhance production levels and make them more efficient by incorporating Industry 4.0 concepts such as Internet technology, communication technology, and automated production. Besides, the manufacturing system is combined with precision machine tools and devices that may be connected to the internet for efficient production control and data analysis. Various industries have adapted to the Industry 4.0 transformation using new technologies, which will create new business needs and opportunities [1, 2], including machine-to-machine communication for automated control [3, 4]. Therefore, smart factories must have sensors connected to devices or machines to control their operations effectively, resulting in real-time data management and exchange [5, 6]. Additionally, these devices or machines in the manufacturing process can help to create effective preventive maintenance plans for machines and equipment in production [7, 8]. Technological advances enable the detection and processing of communication capabilities with devices and services, resulting in higher standards and a more efficient manufacturing operating environment [1, 7, 9]. Consequently, Industry 4.0 technology is a necessary technology for increasing production efficiency and creating sustainability.

This study presents the challenges of sustainable manufacturing in Industry 4.0, intending to study the impact of various factors in the application of Industry 4.0 in the manufacturing of electrical appliances, as well as the significance of these factors in forecasting decision-making's tendency in the future application of Industry 4.0 technology. The fuzzy analytical network process (FANP) approach in the multiple criteria decision-making principle (MCDM) is used in this research because FANP can simplify complex problems and allow dependence and feedback in the hierarchy. Thus, the FANP approach is applied for prioritizing factors to analyze and compare their relative importance, which impacts the decision-making for applying Industry 4.0 technology in the electrical appliance industry. Finally, the study's findings will assist stakeholders in making informed decisions in terms of the implementation of Industry 4.0 technology in their manufacturing processes to increase their businesses' competitiveness and performance in the long run.

In the first section, we review Industry 4.0 technology and its applications to determine the factors that affect the application of Industry 4.0 technology in manufacturing. In the following section, we describe the research methodology used in this study, as well as related theories such as the fuzzy set and fuzzy analytical network process, which are used for prioritizing factors influencing decision makers in electrical appliance manufacturers in terms of Industry 4.0 technology application. The study's results have shown the importance of factors and sub-factors including their ranking by FANP in

Section IV. Section V discusses the findings and management implications. Finally, Section VI concludes with a summary of the study's findings, limitations, and future research.

II. Literature Review

Building a competitive advantage requires the development of systems [10], whereas the evolution of Industry 4.0 enables more transparent resource utilization through efficient practices in both production and innovation [11, 12]. Furthermore, the government has recognized the critical importance of Industry 4.0, information technology, and the environment in driving the economy, issuing a slew of regulations to support and safeguard them [13]. By incorporating intelligent mechanisms and automation in the industrial environment, Industry 4.0 enables businesses to improve production processes [5]. The challenges of Industry 4.0 encompass not only the technological dimension, but also the organizational, social, and national economic dimensions. Thus, organizations and governments must adapt to Industry 4.0 to achieve operational success [14]. The fourth industrial revolution is the result of rapid technological advancement [15], as evidenced by the interconnection of intelligent manufacturing systems used in Industry 4.0. The goal is to create human-machine interaction by automating and resilient production, parts and product tracking, and machine-to-machine communication. This will result in increased productivity and lower costs. To enable on-demand production and more efficient supply and demand, the intelligent manufacturing system will connect product design, analytics, manufacturing processes, stock and supply chain systems, product customization, delivery systems, and customers [16]. While, the Internet and technology are future-oriented, and they improve the human-machine interaction paradigm to provide added value to processes [14, 17, 18]. Information technology is a critical component of the economy's driving force [19], enabling the digitization of production processes, the relationships between physical and controllable objects, and the improvement of integration between the physical and digital worlds [20]. Manufacturing technologies are being developed to communicate and deliver big data, allowing businesses to improve forecasting and productivity [21, 22]. Additionally, the technology incorporates processing capabilities to improve manufacturing processes, organizational interactions, human resource utilization, waste reduction, environmental impact reduction, energy savings, as well as the industrial value chain, all of which contribute to sustainability [5] From the literature review of Industry 4.0, this study has identified five major factors affecting the application of Industry 4.0 technology as follows: Internet of Things (IoT), Cloud Manufacturing (CMg), Big Data Analytics (BDA), Additive Manufacturing (AM), and Cyber-Physical Systems (CPS), as shown in Table I.

TABLE I. Previous Research on the Application of Industry 4.0 Technology

Factors	Sub-Factors	Reference
Internet of Things (IoT)	Monitoring	[30], [48], [49], [50], [51], [55]
	Tracking	[20], [32], [33]
	Information	[20], [54]
	Automation	[20], [37], [47], [54]
Cloud Manufacturing (CMg)	Security	[31], [38], [39], [45], [52]
	Resources as service	[34], [40], [41], [43], [46]
	Orchestration	[31], [38], [40]
	Resource management	[38], [41], [42], [43], [46], [57]
	Utilization	[40], [41], [44], [46]
Big Data Analytics (BDA)	Predictive	[14], [18], [20], [51], [52]
	Data Collection	[32], [33], [51], [55]
	Data-Driven	[20], [30], [32], [33], [51], [52], [53], [55]
	Accurate	[20], [32], [51], [52], [54]
	Predictive	[14], [18], [20], [51], [52]
Additive Manufacturing (AM)	Prototype	[7], [23], [33], [60],[61],[62],[63]
	Flexibility	[33], [62], [63]
	Optimization	[1], [19], [52], [56]
	Reduction	[20], [22], [23], [27],[28], [33]
Cyber-Physical System (CPS)	Collaborates	[7], [30], [33], [40], [51], [52], [64]
	Detects Production System	[20], [33], [36]
	Decision Making	[20], [33], [51], [54], [59]
	Visibility and Traceability	[30], [32], [52]
	Synchronization	[8], [9], [20], [52]
	Intelligent	[20], [36], [37], [47], [65], [66], [67]

A. Internet of Things

The Internet of Things (IoT) is a collection of intelligent devices and sensors that can communicate with one another via the Internet [7, 23]. It can identify specific physical and virtual connections for communication, configuration, management, and data collection [14, 17, 18]. Besides, IoT can accurately monitor and measure a wide range of processes, such as real-time tracking, predictive maintenance, production volume management, and sales data [21]. Consequently, the collected data is routed to a central processing system for analysis and decision-making, resulting in solutions in automated manufacturing processes, lower production costs, and higher technological standards [20, 22]. IoT also has access to product lifecycle data, allowing products to be traced from the factory to the customer [23]. Furthermore, it can improve product cycle visibility, reduce resource consumption, increase productivity [24, 25, 26] provide maintenance traceability, monitor labor activities, and improve work safety [27, 28], as well as manage resources for increasing competitiveness and long-term success.

1) Monitoring: The process of visualizing the recording data from real-time monitoring of the status change of production parameters to a specified standard was known as monitoring [29]. It was possible to continuously monitor or audit production data by recording, analyzing, or evaluating traceability data in real-time [30]. When non-standard problems such as critical overloads or insecurity were detected, the collected data could be assessed, and timely corrective action could be ordered [31]. Thus, organizations must regularly monitor and control operations under countermeasures to ensure process reliability, flexibility, and long-term productivity.

2) Tracking: A system that enabled the periodic monitoring of operational procedures and displays the data could be useful for analyzing current and future situations. Controlling and processing operations could also check the status of each step of the process, allowing for real-time tracking of goods flow [20]. Thus, Internet of Things applications that connect all relevant departments to collaborate via the Internet could improve efficiency in operational control decentralization [32, 33]. It provided a real-time tracking system for the manufacturing process, making the operation more convenient and increasing customer confidence.

3) Information: Information was created by processing and analyzing data from the Internet and presenting it in the form of knowledge that was applied to the workplace. Quality information must be accurate and up to date, as well as consistently stored. It is also supported by paperwork reduction, production management access, and inventory management [30]. Complex work systems and real-time decision-making necessitated the use of information as well. Thus, managing the flow of goods in processes enabled supply chain participants to track the flow of goods in real time and make precise decisions based on the information obtained [29]. Consequently, precise data acquisition and management were accomplished, allowing for more efficient operations.

4) Automation: IoT automation connected the entire network and shared data with linked systems based on requirements [31, 34], allowing for cost-effective, convenient, and safe operations. Using automated systems, humans can perform high-accuracy and risky tasks. Furthermore, controlling or receiving commands in real-time [35], between devices or machines based on connectivity requirements resulted in increased production and management capabilities.

B. Cloud Manufacturing

Cloud manufacturing is a data collection and processing system that collects and processes data with servers on computer networks. It also included services and applications that could use a network connection to access and process data or resources on the Internet [7, 25]. In addition, cloud manufacturing was a technology that involved the delivery of computing resources via the Internet, resulting in lower operating costs and increased response time efficiency. It also enabled data access from any location at any time, responsive data sharing, and supply chain support continuity [36, 37]. Therefore, cloud computing technology could make storing and processing large amounts of data more efficient. It could also develop new services, such as quality monitoring and control functions that improve operations and production efficiency [23].

1) Security: Database management security was a critical aspect of security. Particularly concerning was the provision of information to third-party cloud systems [38] because corporate data includes customer data, business data, employee data, intellectual property, and so on [39]. As a result of the various levels of information security requirements that underpin cloud services as well as search and access capabilities, the issue of information security was a huge challenge in services. Data security risks might negatively impact the organization; therefore, trust in cloud service recipients is required [31].

2) Resources as service: The service-oriented manufacturing model was designed to manage resources by providing cloud-based manufacturing solutions, on-demand network technology services, cloud service capabilities, and service platforms [34, 40]. It was a model that combines various technologies to facilitate collaboration, including sharing and managing production resources [41], where cloud manufacturing entailed the transformation of production resources (hardware and software), cloud service capabilities, and the ability to control services, as well as the management of production resources, processes, operations, and Internet transactions [40].

3) Orchestration: Capabilities for customer-focused collaboration and the creation of a new production line from a dedicated manufacturing service [40], were additional cloud manufacturing services for creating capabilities such as intelligent routing, preparation, data integrity checking, and security [41, 42], as well as business process service management that could be modeled and structured to support a smooth business.

4) Resource management: Cloud productivity capabilities enabled organizational planning and resource allocation to achieve goals. In resource management, resource simulation must consider the importance of resource diversity, user demand requirements, and performance requirements [31, 38] by connecting objects and production resource management for supporting manufacturing [43, 44]. Besides, it was a tool for organizing and managing information about production resources, including physical and virtual resources to maximize resource management capabilities [42].

5) Utilization: Using cloud computing technology facilitates and shares data to maximize productivity utilization [45], and the availability of dynamic cloud management would meet real-time needs [34]. By allocating production resources appropriately, cloud manufacturing could reduce costs and increase resource utilization. This allows for a smooth transition from service-oriented manufacturing to remediation [46], resulting in improved utilization or capacity of production resources.

C. Big Data Analytic

Big data analytics is a challenge for corporate performance in driving manufacturing because it can be used for real-time communication, monitoring, and transactions via multi-directional data networks and supported digital supply chain strategies [16]. On the other hand, data collection, analysis, and dissemination from machines and equipment enable efficient data storage, monitoring, and decision-making evaluation. [20, 47]. Meanwhile, predictive analytics, agility, and business insights [21], enable machines and devices to analyze data trends with artificial intelligence, including rapid communication of customer data, which can improve product quality and the business's supply chain [23]. Consequently, big data analytics improves the ability to collect data in real-time to meet objectives on time [48, 49, 50]. Furthermore, it can also learn from data for more accurate predictive production.

1) Predictive: Predictive analytics provided insights into potential trends [51] by utilizing mathematical techniques, simulations, multi-criteria decision-making, planning automation, risk management, and demand forecasting, among other things [52]. Predictive analytics also employed planning and control to accurately forecast production to meet demand and make real-time decisions.

2) Data collection: Data collection enabled real-time decision-making by analyzing data from various sources such as production data, customer feedback, or product ordering systems [33]. The data would be saved and collected in the storage system for further analysis [31]. To derive value from various types of data stored on demand, organizations must collect a large amount of data and require adequate data storage to support big data from resources.

3) Data-driven: Data-driven played an important role in supporting organizations' management, strategic planning, and goal setting, which necessitated large data storages to support big data in multiple formats while maintaining data quality and accuracy. Thus, the ability to process and control operations with big data has been applied to learning, analysis, and event-based decision-making. Consequently, data-driven systems capable of optimizing both production and predictive production are created [33, 52]. Besides, the visual ability of data flow throughout the supply chain, such as service quality, forecast accuracy, production efficiency, and inventory planning, would result in improved supply chain collaboration and effective network response [53].

4) Accuracy: Big data analytics could deliver real-time data quickly and accurately, as well as exchange data in planning, controlling, and forecasting for accurate decision-making [52]. In addition, big data analytics was critical in enabling multidimensional data to be considered and presented with accuracy, resulting in resource utilization that was consistent with demand, reducing losses and costs, controlling quality, and increasing productivity [21, 33] Moreover, big data analytics enabled real-time decision-making based on accurate data analysis results. It could be analyzed using big data platforms, which could generate sophisticated analytical models tailored to the situation and ensure more accurate production.

D. Additive Manufacturing

Additive manufacturing, also known as 3D printing, is a technique for creating parts from computer-aided designs, and it is most used to create prototypes [51, 33]. It is a manufacturing challenge that could replace traditional production, making it more convenient and supporting low-volume production [54, 35, 55]. Additive manufacturing aids product design in meeting needs and complexity, bringing products to market faster, and lowering transportation and storage costs. It also reduces material waste, improves, and develops products, and facilitates prototyping [23, 21, 22]. Thus, additive manufacturing technology has the potential to reduce production time and machine and raw material modification costs, improve reverse engineering ease, rapidly create new products and reproducibility, conduct testing and validation, increase customer satisfaction, and boost competitive advantage.

1) Prototype: Rapid prototyping played an important role in enabling diversified and enhanced production through technology for part design and creation, on-demand prototyping, minimizing design errors, prototyping testing and validation, pre-production customization, and product life cycle in new product development, which could be quickly adapted to change and meet market demands [21]. Furthermore, rapid prototyping increased the versatility of the fabrication process by enabling customization, ease of on-site production, and cost savings in production time and machine tool replacement, including raw materials [33]. Besides, prototyping was cost-effective and could lead to customer satisfaction.

2) Flexibility: The flexibility of additive manufacturing had accelerated the development and production of new products while also enabling convenient improvements or modifications to existing products [33]. Moreover, sophisticated and iterative products could be designed more quickly, saving money and time compared to traditional manufacturing methods while maintaining quality, durability, and functionality. In smart manufacturing, testing and validation were essential components of additive manufacturing [20], allowing manufacturing processes to be managed more efficiently.

3) Optimization: Additive manufacturing was a crucial method for structural design and manufacturing, as well as scalability, product design, and production integration. Furthermore, additive manufacturing could be used to create homogeneous structures with multiple sizes or layers [56]. This was an evolutionary structure optimization in which the optimal structure was determined by appropriate material distribution, resulting in the automatic optimization of specific input data.

4) Reduction: Because of the low scrap rate and easy disassembly of the structure for design and production, using additive manufacturing techniques could help manufacturers save money and support costs while also reducing waste [33]. Additive manufacturing techniques might also reduce the amount of labor, machinery, and equipment used in the manufacturing and prototyping of parts, which is especially labor-intensive and expensive in metal prototyping. Additionally, using additive manufacturing techniques could reduce costs associated with manufacturing, assembly, and inventory, resulting in more efficient prototype production.

E. Cyber-Physical System

A cyber-physical system (CPS) is an integrated engineering system that connects the physical and cyber worlds [42, 44, [57]. The physical world consists of devices, machines, humans, systems, etc. Whereas, the cyber world, also known as the digital world, processes and controls the physical world by connecting it with the Internet of Things (IoT), which oversees the physical world's connectivity, communication, and data transfer to the cyber world. While the cyber world computes, analyzes, or decides to send feedback control data to the physical world automatically. It also encourages the monitoring and control of environmental changes to facilitate manufacturing operations [20, 22] as a technology for management systems that wisely connect physical resources [38, 31, 43] As a result of this, machine performance, quality inspection, and control functions, as well as operation and production, have all improved [56, 40].

1) Collaboration: By directly processing data on the Internet, cyber-physical systems could collaborate indefinitely [33]. Computer collaboration in physical process operations such as machine and equipment control, warehouse management, and production process management could be planned and controlled by analyzing and processing data contained in the system [20, 58] to support collaboration, collaboration and resources, data sharing between service providers, and these capabilities were complementary to integration and collaboration [52].

2) Detects production system: Sensing production systems connected with computers to exchange data (interface), such as the interaction between computer users, machines, and equipment [33]. Whereas cyber-physical systems could monitor, measure, perceive, and control system conditions to meet goals or prevent errors in the manufacturing process, including fault detection (Downtime due to faults, workflow restrictions, and error notification). Besides, identifying the source of the defective process improved systems to prevent faults from recurring.

3) Decision-making: Data obtained from the manufacturing process was used to drive decision-making processes, allowing cyber-physical systems to perform better [59]. Numerous real-time data exchanges occurred between systems to ensure the continuity of decision-making and situation management, resulting in a reduction in data planning errors and complex decision-making processes [20]. Additionally, this enabled analysis and decision-making to control systems and achieve goals. Furthermore, Intelligent decision-making and control, machine control, and troubleshooting all contributed to faster and more accurate decision-making, which increased productivity.

4) Visibility and traceability: Visibility and traceability were components of the cyber-physical system that assisted employees in making simple proactive decisions, inspecting parts, controlling quality, and planning and controlling production. Manufacturing traceability and intelligent forecasting in supply chains are supported to eliminate employee concerns [52] by automatically computing, analyzing data, and transmitting feedback control.

5) Synchronization: the cyber-physical system would be able to synchronize various production activities such as planning and controlling in real-time [52] based on orders, production processes, and responses. By synchronizing the production process with other related processes, it created production process continuity [45].

6) Intelligent: Cyber-physical systems, powered by intelligent data, digitized data that could be used for automated audits, data processing, and monitoring. as well as modeling. Automated learning relied on constantly changing data to make accurate predictions and intelligent decisions. Sensor systems were used in the production and maintenance of machines and equipment [20], to make them more reliable and efficient, and they were a smart manufacturing innovation.

III. MATERIALS AND METHODS

In this study, the survey method was used for data collection, which was followed by quantitative research. The research tool was a questionnaire based on five main factors and a fuzzy analytical network process approach: Internet of Things, cloud manufacturing, big data analytics, additive manufacturing, and cyber-physical system, as shown in Fig. 1.

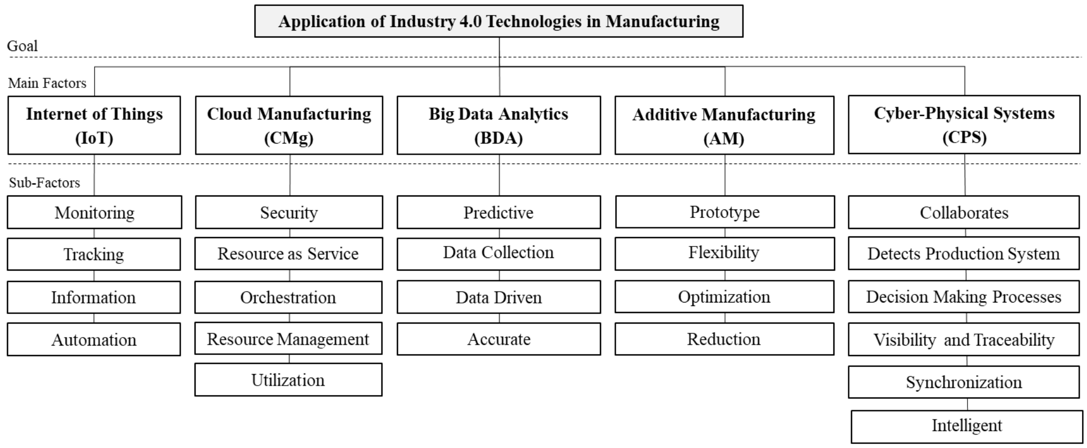


Fig. 1. Conceptual framework of the application of Industry 4.0 technologies in manufacturing.

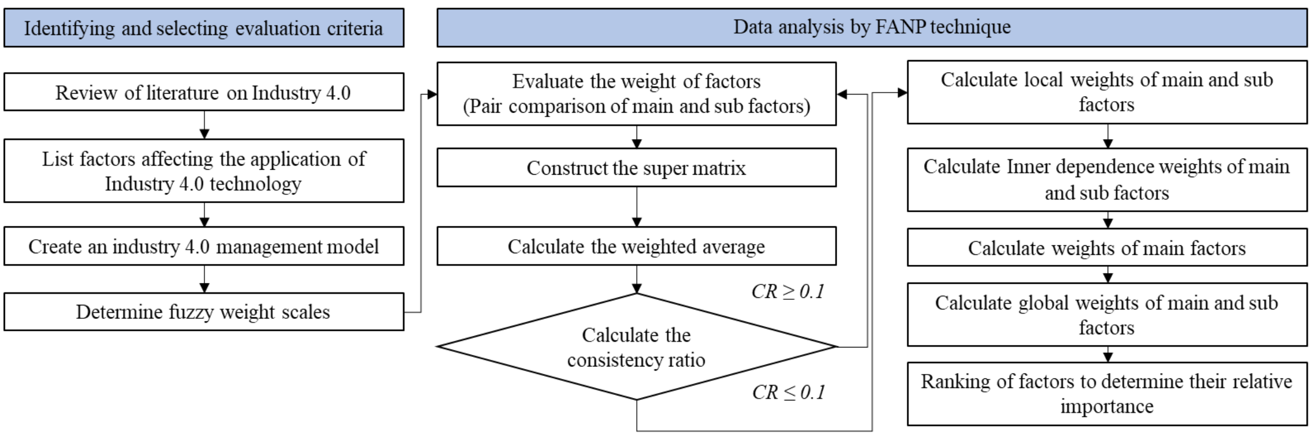


Fig. 2. The research framework of the application of Industry 4.0 technology.

According to the Institute of Electrical and Electronics, the population in this study consisted of 440 electrical appliance manufacturers in Thailand, with the Yamane's formula [68] indicating that the sample size for the study was 82 samples. Following the completion of the questionnaire, the researcher distributed it to 82 respondents sampled using the snowball method who were decision-makers in the implementation of industry 4.0 in electrical appliance manufacturers. At the start of this survey, experts from electrical appliance manufacturers were identified and called in to ask their permission to answer the questionnaire about the factors affecting the application of industry 4.0 technology in electrical appliance manufacturing. Following that, the data from the questionnaire were analyzed and carried out following the steps in the research framework depicted in Fig. 2.

A. *Fuzzy Set*

Zadeh (1965) developed a fuzzy set theory to deal with situations involving uncertainty, imprecision, and vagueness. Currently, fuzzy set theory is frequently used in multiple criteria decision-making techniques for practical applications in uncertain settings, and it is also effectively applied for modeling human decisions and judgments. [69, 70, 71]. The theory also allows for the application of mathematical operations and programming to the fuzzy domain. In general, a membership function defines the fuzzy set, which represents the grade of any element *x* of *X* that has partial membership to *M*. The degree to which an element belongs to a set is determined by a value between zero and one. Thus, the element x belongs to M, $\mu_M(x) = 1$ and $\mu_M(x) = 0$ [69, 72].

The triangular fuzzy number is made up of three parameters: (l, m, u) where $l \leq m \leq u$. The membership function $\mu_M(x)$ of TFN is derived by Eq. (1).

$$\mu_M(x) = \begin{cases} (x-l)/(m-l) & l \leq x \leq m \\ (u-x)/(u-m) & m \leq x \leq u \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Let $X = \{x_1, x_2, \ldots, x_n\}$ represent an object set and $U = \{u_1, u_2, \ldots, u_n\}$ represent a goal set. Chang (1992) extent analysis method takes each object and performs extent analysis for each goal g_i. For each object, m extent analysis values with the following signs can be obtained:

$$M_{g_i}^1, M_{g_i}^2, \cdots, M_{g_i}^m,\ \text{i=1,2,}\cdots\text{,n} \quad (2)$$

Where $M_{g1}^{\,j}$, $(j = 1, 2, \ldots, m)$ represents in triangular fuzzy numbers.

B. *The Fuzzy Analytical Network Process*

The fuzzy analytic network process (Fuzzy ANP) is a multiple criteria decision-making (MCDM) technique that employs both criterion interdependence and criterion inner dependence with the pairwise comparison matrix. In this study, we present Chang's (1992, 1996) extent analysis method because the steps are simpler than those of the other fuzzy ANP techniques. The following are the steps of Chang's (1992, 1996) extent analysis approach:

$$S_i = \sum_{j=1}^{m} M_{gi}^{j} \otimes \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^{j}\right]^{-1} \quad (3)$$

To obtain $\sum_{j=1}^{m} M_{gi}^{j}$, perform the fuzzy addition operation on m extent analysis values for a specific matrix so that:

$$\sum_{j=1}^{m} M_{gi}^{j} = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j\right) \quad (4)$$

And to obtain $\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^{j}\right]^{-1}$, perform the fuzzy addition operation of extent analysis values for a specific matrix so that:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^{j} = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j\right) \quad (5)$$

Then compute the inverse of the vector in the preceding equation so that

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j}\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i}\right) \quad (6)$$

The degree of possibility of $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ is defined as:

$$V(M_2 \geq M_1) = \sup[\min(\mu_{M_1}(x), \mu_{M_2}(y)] \quad (7)$$

and can be equivalently expressed as follows:

$$V(M_2 \geq M_1) = \begin{cases} 1 & \text{if } m_2 \geq m_1, \\ 0 & \text{if } l_1 \geq u_2, \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & \text{otherwise} \end{cases} \quad (8)$$

$$V(M \geq M_1, M_2, \ldots, M_k) = \min V(M \geq M_i), \quad i = 1, 2, \ldots, k \quad (9)$$

Assume that:

$$d(S_i) = \min V(S_i \geq S_k), \quad k = 1, 2, \ldots, n; k \neq i \quad (10)$$

Then the weights vector is computed by:

$$W' = (d(S_1), d(S_2), \ldots, d(S_n))^T \quad (11)$$

And then, weight vectors are normalized as follows:

$$W = (d(A_1), d(A_2), \ldots, d(A_n))^T \quad (12)$$

IV. RESULTS

For this study, the application of Industry 4.0 technology can be divided into five groups: Internet of Things (IoT), Cloud Manufacturing (CMg), Big Data Analytics (BDA), Additive Manufacturing (AM), and Cyber-Physical System (CPS). As illustrated in Fig. 2, each group includes several applications of Industry 4.0 technologies in manufacturing. Consequently, the goal of this research is to propose an industry 4.0 application for electrical appliance manufacturers' management to select appropriate technologies to integrate into their manufacturing processes to increase productivity and gain a competitive advantage. In this section, the research process can be divided into six steps to evaluate the factors affecting the application of 4.0 technology and prioritize those factors for appropriate decision-making by electrical appliance manufacturers in the application of industrial 4.0 technology. The first stage identifies the application of industry 4.0 technology for the framework of factors and sub-factors, as determined by a questionnaire survey using the fuzzy A technique to compute the weights of the factors and sub-factors. The application is built on the steps from the previous section and is explained step by step along with the results.

Step 1: The fuzzy scale is used to determine the important weighting of main factors and sub-factors from a decision-maker in charge of an organization's Industry 4.0 application management. Select appropriate linguistic variables to represent the relative weight of the factors. Using the linguistic scale in Table II, a pair-wise comparison is prepared for each factor based on expert opinion.

Step 2: A questionnaire survey is used to collect data from 82 electrical appliance manufacturers to investigate the application of Industry 4.0 technology for management in practices.

Step 3: Compute the consistency ratio (CR) by Eq. (13).

$$CR = \frac{CI}{RI} \quad (13)$$

When $CI = \frac{(\lambda_{max} - n)}{(n-1)}$ and $\lambda_{max} = \sum_{i=1}^{n}\left[\sum_{j=1}^{n} a_{ij} W_j\right]$

The CR value should be less than 0.1, indicating that the weights determined by decision-makers are consistent. If the CR value is greater than 0.1, it indicates that the determined weights are inconsistent. Thus, the research should begin with a review of the main factors and sub-factors.

Step 4: Compute the average fuzzy evaluation matrix based on the questionnaire survey, as well as the local weights of the main factors and sub-factors, as shown in Table II-VII.

Step 5: As illustrated in Table IX, compute the inner dependence weights and dependencies among the factors considered to be the main factors.

Step 6: To compute the global weights of sub-factors, multiply the local weights of the sub-factors by the interdependent weights of the main factors, as shown in Table X.

When weights of main factors were calculated, it was discovered that the Internet of Things was the most important, accounting for 34.3 percent of the total, the Big Data Analytics accounted for 31.0 percent, Cloud Manufacturing factor 12.5 percent, and Cyber-Physical System 9.4 percent respectively, as shown in Table III.

In terms of Internal of Things, tracking was identified as the most significant sub-factor in this area, accounting for 30.7 percent. Monitoring is the second most significant sub-factor, accounting for 27.2 percent, while Automation and Information had declined in importance, accounting for 25.7 percent and 16.5 percent, respectively, as shown in Table IV.

TABLE II. LINGUISTIC SCALE FOR RELATIVE WEIGHTS OF FACTORS

Risk Level	Triangular Fuzzy Scale	Triangular Fuzzy Reciprocal Scale
Just equal	(1, 1, 1)	(1, 1, 1)
Equally important	(1/2, 1, 3/2)	(2/3, 1, 2)
Weakly more important	(1, 3/2, 2)	(1/2, 2/3, 1)
Strongly more important	(3/2, 2, 5/2)	(2/5, 1/2, 2/3)
Very strongly more important	(2, 5/2, 3)	(1/3, 2/5, 1/2)
More important	(5/2, 3, 7/2)	(2/7, 1/3, 2/5)

TABLE III. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "MAIN FACTORS"

Main Factors	Internet of Things	Cloud Manufacturing	Big Data Analytics	Additive Manufacturing	Cyber-Physical System	Weights
Internet of Things	1.000, 1.000, 1.000	1.282, 1.738, 2.201	1.088, 1.516, 1.967	1.451, 1.951, 2.451	1.555, 2.055, 2.555	0.343
Cloud Manufacturing	0.540, 0.697, 0.933	1.000, 1.000, 1.000	0.459, 0.598, 0.797	0.688, 0.947, 1.301	0.909, 1.354, 1.870	0.125
Big Data Analytics	0.607, 0.813, 1.114	1.402, 1.870, 2.354	1.000, 1.000, 1.000	1.354, 1.854, 2.354	1.305, 1.805, 2.305	0.310
Additive Manufacturing	0.413, 0.523, 0.715	0.929, 1.317, 1.748	0.429, 0.549, 0.764	1.000, 1.000, 1.000	0.769, 1.126, 1.663	0.129
Cyber-Physical System	0.393, 0.489, 0.648	0.604, 0.846, 1.260	0.439, 0.565, 0.797	0.701, 1.039, 1.457	1.000, 1.000, 1.000	0.094

TABLE IV. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "INTERNET OF THINGS"

Sub-factors	Monitoring	Tracking	Information	Automation	Weights
Monitoring	1.000, 1.000, 1.000	0.798, 1.098, 1.431	1.240, 1.701, 2.169	0.766, 1.077, 1.431	0.272
Tracking	0.980, 1.301, 1.685	1.000, 1.000, 1.000	1.201, 1.669, 2.152	0.996, 1.360, 1.778	0.307
Information	0.538, 0.697, 0.947	0.499, 0.665, 0.931	1.000, 1.000, 1.000	0.649, 0.878, 1.230	0.165
Automation	0.929, 1.252, 1.650	0.704, 0.974, 1.305	0.983, 1.407, 1.854	1.000, 1.000, 1.000	0.257

The weight of cloud manufacturing, Resource Management was the most important in this category, accounting for 27.4 percent, The second most important sub-factor was Resource as Service for 22.6 percent, followed by Orchestration for 21.4 percent, Utilization 18.8 percent, and Security 9.7 percent, respectively, as shown in Table V.

The weight of sub-factors in Big Data Analytics revealed that Data-Driven for 36 percent was the most significant sub-factor in this domain. Meanwhile, the second most important sub-factor was Data Collection for 33.4 percent, followed by Accuracy fot 18.2 percent, and Predictive for 12.4 percent, respectively, as shown in Table VI.

The weight of Additive Manufacturing was considered, the findings indicated that the Flexibility was the most important in this area for 26.3 percent, Prototype accounting for 26.0 percent, Optimization for 25.3 percent, and Reduction for 22.4 percent, respectively, as shown in Table VII.

When considering the weight of sub-factors in the Cyber-Physical Production Systems domain, Decision-Making Processes was the most important at 21.4 percent, followed by Visibility and Traceability, which accounted for 20.9 percent. The Detects Production System was the second important sub-factor that accounted for 17.4 percent, followed by Collaboration for 14.7 percent, Synchronization for 14.2 percent, and Intelligent for 11.5 percent, respectively, as shown in Table VIII.

TABLE V. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "CLOUD MANUFACTURING"

Main Factors	Security	Resource as Service	Orchestration	Resource management	Utilization	Weights
Security	1.000, 1.000, 1.000	0.505, 0.681, 0.996	0.488, 0.646, 0.894	1.000, 1.000, 1.000	0.479, 0.632, 0.931	0.097
Resource as Service	1.201, 1.669, 2.152	1.000, 1.000, 1.000	0.808, 1.116, 1.551	0.733, 1.051, 1.421	1.006, 1.441, 1.909	0.226
Orchestration	1.256, 1.724, 2.207	0.821, 1.169, 1.565	1.000, 1.000, 1.000	0.609, 0.850, 1.175	0.887, 1.278, 1.711	0.214
Resource Management	1.354, 1.854, 2.354	0.896, 1.234, 1.695	1.055, 1.457, 1.909	1,000, 1.000, 1.000	1.152, 1.652, 2.152	0.274
Utilization	1.104, 1.604, 2.104	0.577, 0.795, 1.126	0.698, 0.970, 1.386	0.476, 0.632, 0.963	1.000, 1.000, 1.000	0.188

TABLE VI. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "BIG DATA ANALYTICS"

Sub-factors	Predictive	Data Collection	Data Driven	Accuracy	Weights
Predictive	1.000, 1.000, 1.000	0.530, 0.680, 0.900	0.410, 0.516, 0.699	0.776, 1.057, 1.409	0.124
Data Collection	1.330, 1.787, 2.250	1.000, 1.000, 1.000	0.794, 1.120, 1.486	1.110, 1.577, 2.061	0.334
Data Driven	1.451, 1.951, 2.451	0.794, 1.120, 1.518	1.000, 1.000, 1.000	1.087, 1.543, 2.006	0.360
Accuracy	0.924, 1.276, 1.675	0.517, 0.695, 0.992	0.579, 0.762, 1.063	1.000, 1.000, 1.000	0.182

TABLE VII. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "ADDITIVE MANUFACTURING"

Sub-factors	Prototype	Flexibility	Optimization	Reduction	Weights
Prototype	1.000, 1.000, 1.000	0.867, 1.244, 1.640	0.812, 1.189, 1.585	1.000, 1.000, 1.000	0.260
Flexibility	0.776, 1.030, 1.390	1.000, 1.000, 1.000	0.828, 1.211, 1.624	0.974, 1.390, 1.819	0.263
Optimization	0.787, 1.049, 1.427	0.743, 1.014, 1.439	1.000, 1.000, 1.000	0.926, 1.341, 1.770	0.253
Reduction	0.787, 1.081, 1.476	0.675, 0.900, 1.276	0.691, 0.933, 1.374	1.000, 1.000, 1.000	0.224

TABLE VIII. LOCAL WEIGHT AND PAIR-WISE COMPARISON MATRIX OF "CYBER-PHYSICAL PRODUCTION SYSTEMS"

Sub-factors	Collaboration	Detects Production System	Decision-Making Processes	Visibility and Traceability	Synchronization	Intelligent	Weights
Collaboration	1.000, 1.000, 1.000	0.746, 1.010, 1.325	0.541, 0.734, 1.000	0.590, 0.783, 1.049	0.943, 1.333, 1.754	1.061, 1.498, 1.970	0.147
Detects Production System	1.060, 1.423, 1.831	1.000, 1.000, 1.000	0.612, 0.819, 1.122	0.632, 0.884, 1.236	1.112, 1.579, 2.053	1.063, 1.498, 1.955	0.174
Decision-Making Processes	1.226, 1.667, 2.140	1.106, 1.541, 1.998	1.000, 1.000, 1.000	0.970, 1.378, 1.835	1.138, 1.596, 2.134	1.024, 1.433, 1.890	0.214
Visibility and Traceability	1.216, 1.650, 2.108	0.976, 1.394, 1.900	0.638, 0.900, 1.301	1.000, 1.000, 1.000	1.220, 1.693, 2.183	1.116, 1.589, 2.079	0.209
Synchronization	0.741, 0.998, 1.352	0.572, 0.752, 1.053	0.513, 0.720, 1.020	0.504, 0.671, 0.939	1.000, 1.000, 1.000	0.957, 1.463, 1.970	0.142
Intelligent	0.579, 0.797, 1.118	0.621, 0.833, 1.150	0.621, 0.866, 1.199	0.524, 0.705, 1.008	0.524, 0.711, 1.122	1.000, 1.000, 1.000	0.115

TABLE IX. INNER DEPENDENCE WEIGHT OF THE FACTORS

Main Factors	Internet of Things	Cloud Manufacturing	Big Data Analytics	Additive Manufacturing	Cyber-Physical System
Internet of Things	1.000	0.444	0.493	0.402	0.396
Cloud Manufacturing	0.186	1.000	0.187	0.160	0.102
Big Data Analytics	0.461	0.365	1.000	0.353	0.362
Additive Manufacturing	0.201	0.104	0.192	1.000	0.139
Cyber-Physical System	0.152	0.087	0.129	0.085	1.000

TABLE X. TOTAL WEIGHT OF MAIN FACTORS AND SUB-FACTORS

Factors	Interdependent weights	Ranking	Sub-factors	Local weights	Ranking	Global weights	Ranking
Internet of Things (IoT)	0.3198	1	Monitoring	0.2718	2	0.0869	4
			Tracking	0.3067	1	0.0981	3
			Information	0.1647	4	0.0527	7
			Automation	0.2569	3	0.0822	5
Cloud Manufacturing (CMg)	0.1382	4	Security	0.097	5	0.0134	22
			Resource as Service	0.226	2	0.0313	14
			Orchestration	0.214	3	0.0296	15
			Resource management	0.275	1	0.0379	8
			Utilization	0.188	4	0.0260	16
Big Data Analytics (BDA)	0.2963	2	Predictive	0.1243	4	0.0368	10
			Data Collection	0.3342	2	0.0990	2
			Data Driven	0.3600	1	0.1067	1
			Accuracy	0.1815	3	0.0538	6
Additive Manufacturing (AM)	0.1417	3	Predictive	0.2597	2	0.0368	11
			Data Collection	0.2633	1	0.0373	9
			Data Driven	0.2526	3	0.0358	12
			Accuracy	0.2243	4	0.0318	13
Cyber Physical Systems (CPS)	0.1041	5	Collaboration	0.1466	4	0.0153	20
			Detects Production System	0.1735	3	0.0181	19
			Visibility and Traceability	0.2140	1	0.0223	17
			Synchronization	0.2088	2	0.0217	18
			Intelligent	0.1418	5	0.0148	21
				0.1153	6	0.0120	23

The decision-makers performed pairwise comparisons for main factors and sub-factors, which were then analyzed using fuzzy ANP. According to the findings, the "Internet of Things" was the most important factor, with an interdependent weight of 31.98 percent. The second most important factor, with a weight of 29.63 percent, was "Big Data Analytics", while "Additive Manufacturing," "Cloud Manufacturing," and "Cyber-Physical Systems" had interdependent weights of 14.17, 13.82, and 10.41 percent, respectively. In terms of important sub-factors, the results showed that "Data-Driven" was the most important sub-factor that influenced decision-makers, with a global weight of 10.67 percent. Next, "Data Collection" was ranked second with a global weight of 9.90 percent, followed by "Tracking," "Monitoring," and "Automation," with global weights of 9.81, 8.69, and 8.22 percent, respectively. The bottom five sub-factors in terms of importance were "Detects Production," "Collaboration," "Synchronization," "Security," and "Intelligent," with global weights of 1.81, 1.53, 1.48, 1.34, and 1.20 percent, respectively. Therefore, these sub-factors had less influence on the decision-maker when selecting industry 4.0 technologies for their production management.

V. DISCUSSION

According to the findings of this study, the Internet of Things (IoT) is the most important because it can be linked to machines or equipment in the automated manufacturing process, such as the Automated Guided Vehicle (AGV) for loading raw materials into the production line. Robots and packaging equipment must operate in line with production flow to improve work efficiency, productivity, cost reduction, and human error. Furthermore, IoT may also be used to build interactions or data transmission amongst equipment to maintain production running smoothly and in line with standards. Meanwhile, traceability and real-time data exchange can help manufacturing be more efficiently monitored and planned. Because the electronics industry typically uses just-in-time manufacturing methods, raw materials or parts must be arrived on time to meet production plans in both quantity and quality. For this reason, the IoT is used to connect production data with suppliers to control and track the flow of production parts to ensure continuity by predefined production plans against real-time targets. This enables the manufacturer and its suppliers to have access to the same data sources and use data to boost productivity while also storing big data for analysis and intelligent decisions. In terms of inventory management, the IoT can perform quick counts and control the right amount of inventory levels to ensure reliable raw material delivery.

Big data analytics is an important secondary factor in managing big data both inside and outside the organization for analytics to enhance decision-making, machine maintenance predictions, data modeling for summary reporting, data connecting across the manufacturing process, predictive analytics, and data linking between suppliers and customers. Overall, big data analytics can use the information gleaned from data analysis to assist all relevant departments in standardizing and systematically receiving the same information, resulting in more efficient work. By creating prototypes, additive manufacturing is now beginning to play an important role in both product development and manufacturing processes. New product prototypes are created during the stages of concept, prototype development, testing, and production for presenting customers with the product appearance, making decisions, and understanding the product details more easily. Furthermore, additive manufacturing enables manufacturers to more easily and quickly adjust parts to meet customer needs, reducing the time and cost spent on modifying prototype parts and allowing for greater flexibility in product development.

Cloud manufacturing is crucial for connecting accurate and up-to-date data with suppliers and customers. Additionally, cloud computing manages data collection and services both inside and outside the organization by carrying out manufacturing operations over the Internet that responsible is for the supervision of the organization's IT department. Moreover, Electronic Data Interchange (EDI) is used for data exchange between suppliers and customers to reduce errors and duplication of documents. Meanwhile, EDI enables faster and more accurate data transfer, as well as making papers easier to search and available only to authorized users. It also makes use of real-time correlation and efficient data exchange. The cyber-physical system is the least important of the main factors in this study. Robots or automated machines, for example, are pre-programmed for control in a manufacturing process. Subsequently, data obtained from machine or robot operations are not sent back and forth between the physical and cyber worlds for automatic control of machines or robots in the physical world. The design and development of cyber-physical systems in the electrical appliance business demands numerous aspects, as well as the challenges of designing software that analyzes and operates entire systems. Therefore, organizations must invest more effort and finances in developing the software.

In terms of sub-factors, the findings of this study have indicated that data-driven is the most important sub-factor because the electrical appliance industry has a diverse range of models and clearly defined production lines. It is also a type of just-in-time production (JIT) in which raw materials and spare parts from various suppliers are processed into a large number of production lines in a specific time frame. To ensure continuity and accuracy, it is critical to be data-driven, which means that production planning data must be aligned with market demand. While production model sequencing must be consistent with the availability of quality data, raw materials, and spare parts. Moreover, data on the availability of machines and equipment on the production line require predictive maintenance to avoid malfunctions during production. As stated above, data-driven can significantly improve the manufacturer’s production efficiency. Another important sub-factor in collecting data from various production activities is data collection. Large amounts of accurate data must be collected throughout the manufacturing process to analyze it and make decisions quickly and precisely on time. In terms of the significance of data collecting, if a decision is delayed or made incorrectly, the production will not be able to operate as planned. Meanwhile, production costs will rise when manufacturing time is extended. Furthermore, the following manufacturing plan will be interrupted and delayed, resulting in the corporation failing to deliver products on time to customers. Therefore, data collection must be accurate and real-time to meet as several customer requirements as possible.

Traceability is an important sub-factor for tracking the flow of parts and raw materials from suppliers to the production line. Traceability also allows the inspection of parts with serial numbers, lot numbers, and manufacture dates to be rapidly tracked back and standardized, which is very useful in the case of a problem. To keep planning on track and build trust with customers, it is necessary to track and monitor each manufacturing process to assess the availability of all spare parts and raw materials before production begins. Monitoring requires reporting the production situation in real time and comparing it to the target by connecting the data via the internet network. Meanwhile, it displays the operational status of machines and equipment in the production line, as well as performance reports in the manufacturing process as compared to the plan. Consequently, monitoring is an essential component for having a real-time overview of the production situation and managing production efficiently. Automation is crucial for driving the production of electrical appliances to reduce costs and increase manufacturing efficiency. Besides, it can also help to automate the primary and secondary production lines as well. However, the manufacturer must first determine which production processes, such as production line systems or packaging systems, require automation. Finally, automation will aid in the faster and more consistent improvement of production processes, providing customers with confidence.

For less important sub-factors, this study has discovered that the detects production system monitors a variety of parameters due to its connection with machines or equipment and manufacturing processes. Fully automated systems necessitate significant capital investments for ongoing improvement, which compels careful consideration of improvement priorities. For collaboration and synchronization, they collaborate with the physical and cyber worlds to integrate systems in production processes that are time-consuming to improve. By synchronizing communication with machines and equipment, it is necessary to prepare various equipment to support the connection between the cyber-physical system and the production process, making it more convenient and efficient in monitoring and controlling the production process. Furthermore, electrical appliance production activities involve complex multiple steps in production processes that require employee and machine integration. The security of the operating system during the manufacturing process must be stable and free of cyber threats. Only on-premise operating systems or private clouds will be used by electrical appliance manufacturers and business partners for internal storage and services. Therefore, operating system risks require data security countermeasures against accidental data leakage and data theft. For another sub-factor, intelligence is the least important as manufacturers prioritize process automation and robotics to improve manufacturing processes to be fast, low-cost, and data-driven for precision and effective decision-making. As a result, intelligent technology is a self-determining intelligent technology that requires the preparation of a large data structure to drive intelligent manufacturing, which requires the implementation of a continuous development plan.

VI. Conclusion

The fourth industrial revolution is currently posing challenges for many industries, particularly the consumer electronics industry, and IoT is one of the key tools for this change because it can improve and connect machines and equipment on the production line to automate and optimize production efficiency while lowering costs. On the other hand, companies that want to use Industry 4.0 technology, must think carefully due to the high capital investment required. As a result, this research investigates the factors influencing the application of Industry 4.0 technology for electrical appliance manufacturers so they can reap the full benefits of this Industry 4.0 technology, such as increased revenue, profit, productivity, and efficiency. as well as cost savings.

The internet of things, cloud manufacturing, big data analytics, additive manufacturing, and cyber-physical systems were identified as five factors influencing the appliance industry's efforts to adopt Industry 4.0 technologies in this study. The Internet of Things is the most important factor in the application of Industry 4.0 technologies, followed by big data analytics, additive manufacturing, cloud computing, and cyber-physical systems in descending order of importance. Data-driven, data collection, tracking, monitoring, and automation are the top five sub-factors influencing appliance manufacturers in implementing Industry 4.0technologies, while detecting production systems, collaboration, synchronization, security, and intelligence have less influence on decision-makers in applying Industry 4.0technologies to their appliance manufacturing. Based on the study's findings, electrical appliance manufacturers can use factors and sub-factors to select the best industrial technology for improving production management and company performance both financially and profitably.

According to the study's findings, they have identified the factor of the Internet of Things (IoT) as the most important because the IoT can be connected to machines and equipment for production continuity. Furthermore, IoT can track and present real-time data to improve the speed and efficiency of manufacturing. Therefore, IoT capabilities can connect existing machines and devices to new equipment outfitted with IoT systems to improve production and warehouse management performance. It is also critical for big data analytics to manage production resources using data from operating systems. Furthermore, data from sensors, devices, and machines is continuously collected and stored as big data for analysis and process planning, as well as manufacturing to achieve operational precision throughout the procurement of production equipment. It is crucial for just-in-time production planning, which demands high precision in terms of production time and raw material supply cycle in order to deliver products to consumers on time. Additive manufacturing is an important part of new product development because it can reduce the steps and costs associated with prototyping for customers. This makes it easier for customers to understand the product and allows them to quickly modify the prototype to meet their needs, providing flexibility in developing new products. Therefore, it can minimize the cost of new product development and launch products to market faster. Cloud computing is considered as the operating systems used within the enterprise, supplier, and customer networks to ensure the stability and security of real-time connections and data relationships in cloud computing. While cyber-physical systems connect the physical and cyber worlds. Thus, manufacturers must consider numerous dimensions when integrating the entire system, which necessitates adequate personnel and budget planning for continuous development.

In terms of the significance of sub-factors, data-driven is a sub-factor that requires a lot of attention. Because the electrical appliance industry produces a wide range of products and employs just-in-time production systems, manufacturers must manage a large amount of data related to suppliers and customers continuously and in real-time. By connecting to the Internet of Things system, machines and devices will send important data for processing and analyzing big data to monitor the availability of raw materials and spare parts to ensure that manufacturers can produce as planned. Data collection is considered for accurate analysis and decision-making. Because the supply chain in the electronics manufacturing industry is complex, each process generates unique data that is linked both inside and outside the organization, necessitating careful consideration of data importance. Traceability requires tracking the flow of parts, raw materials, and goods to build customer trust and provide real-time production data linked to production processes, resulting in more efficient production management. However, automation is now very important in the production of electrical appliance manufacturers to meet the needs of

Industry 4.0, which requires developing automated machines to work instead of more humans. Improved production lines were required to automatically assemble parts for faster, more accurate, and standardized operations.

In implications for management, this research benefits electrical appliance manufacturers in terms of utilizing the factors in this study to appropriately determine the application of Industry 4.0 technology in their production. Because manufacturers have limited budgets, it is necessary to prioritize and select cost-effective and appropriate Industry 4.0 technologies used in production to increase the enterprise's productivity and competitiveness. In terms of research limitations, the sample size of this study is mostly medium to large-sized companies, tier 1st parts manufacturers, and electrical appliance manufacturers that have the potential to apply Industry 4.0 technology at the moment, whereas the sample size of SMEs is small, and they face obstacles in applying Industry 4.0 technology due to limited funding. However, small, and medium-sized enterprises (SMEs) are numerous and play an important role in driving the country's economy, both in terms of income generation and job creation. Consequently, future research will concentrate on studying problems and obstacles in the application of Industry 4.0 technology by SME businesses for the benefit of SME production development, as well as improving SMEs' competitiveness to compete with other large companies in the market.

REFERENCES

[1] M. Ghobakhloo, “Industry 4.0, digitization, and opportunities for sustainability,” Journal of Cleaner Production. Vol. 252, April 2020.

[2] N. G. P. Carvalho, O. Chaim, E. Cazarini, M. C. Gerolamo, “Manufacturing in the fourth industrial revolution: A positive prospect in sustainable manufacturing,” Procedia Manufacturing. Vol. 21, pp. 671-678, January 2018.

[3] S. Bag, S. Gupta, S Kumar, “Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development,” Internal Journal of Production Economics. Vol. 231, January 2021.

[4] S. Gierej, The Framework of Business Model in the Context of Industrial Internet of Things. Procedia Engineering, Vol. 182, pp. 206-212, April 2017.

[5] Y. A. Fatimah, K. Govindan, R. Murniningsih, A. Setiawan, “Industry 4.0 based sustainable circular economy approach for smart waste management system to achieve sustainable development goals: a case study of Indonesia,” Journal of Cleaner Production. Vol. 269, October 2020.

[6] L. Atzori, A. Iera, G. Morabito, “The internet of things: A survey,” Computer networks. Vol. 54, pp. 2787-2805, May 2010.

[7] S. Mittal, M. A. Khan, D. Romero, T. Wuesta, “A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs),” Journal of Manufacturing Systems, Vol. 49, pp. 194-214, October 2018.

[8] K. N. Qureshi, S. Din, G. Jeon, F. Piccialli, “Link quality and energy utilization based preferable next hop selection routing for wireless body area networks. Computer Communications, Vol. 155, pp. 382-392, January 2020.

[9] J. C. Bendul, H. Blunck, “The design space of production planning and control for industry 4.0,” Computers in Industry. Vol. 105, pp. 260-272, February 2019.

[10] S. Doh, B. Kim, “Government support for SME innovations in the regional industries: The case of government financial support program in South Korea,” Research Policy, Vol. 43, No. 9, pp. 1557-1569, November 2014.

[11] M. Ghobakhloo, “Industry 4.0, digitization, and opportunities for sustainability,” Journal of Cleaner Production. Vol. 252, April 2020.

[12] S. Bag, S. Gupta, S Kumar, “Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development,” Internal Journal of Production Economics. Vol. 231, January 2021.

[13] M. S. Bhatia, S. K. Jakhar, S. K. Mangla, K. K. Gangwani, “Critical factors to environment management in a closed loop supply chain,” Journal of Cleaner Production. Vol. 255, May 2020.

[14] Z. Hassan, H. A. Ali, M. M. Badawy, “Internet of Things (IoT): Definitions, Challenges and Recent Research Directions,” Vol. 128(1), pp. 37-47, October 2015.

[15] S. I. Shafiq, C. Sanin, E. Szczerbicki, C. Toro, “Virtual engineering factory: creating experience base for industry 4.0, Cybernetics and Systems,” Vol. 47, No. 1, pp. 32-47, Febuary 2016.

[16] Schoenherr, T. and Speier-Pero, C. (2015) Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential, Journal of Business Logistics, 36(1), 120-132

[17] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” Future Generation Computer Systems. Vol. 29, No. 7, pp. 1645-1660, September 2013.

[18] H. M. Pereira, S. Ferrier, M. Walters, et al., “Essential Biodiversity Variables,” Vol, 339, No. 6117, pp. 277-278, Jan 2013.

[19] M. S. Bhatia, S. K. Jakhar, S. K. Mangla, K. K. Gangwani, “Critical factors to environment management in a closed loop supply chain,” Journal of Cleaner Production. Vol. 255, May 2020.

[20] M. Javaid, A. Haleem, “Impact of industry 4.0 to create advancements in orthopaedics,” Journal of Clinical Orthopaedics and Trauma. Vol. 11, No. 6, pp. 491-499, March 2020.

[21] J. L. Hopkins, “An investigation into emerging industry 4.0 technologies as drivers of supply chain innovation in Australia,” Computers in Industry Vol. 125, No. 2, pp. 1-20, February 2021.

[22] N. Harikannan, S. Vinodh, “Sustainable industry 4.0 - an exploratory study for uncovering the drivers for integration,” Journal of Modelling in Management, Vol. 16, No. 1, pp. 357-376, April 2021.

[23] G. Büchi, M. Cugno, R. Castagnoli, “Smart factory performance and Industry 4.0,” Technological Forecasting and Social Change. Vol. 150, January 2020.

[24] S. E. Hamdi, M. Oudani, A. Abouabdellah, A. Sebbar, “Fuzzy approach for locating sensors in industrial internet of things,” Procedia Computer Science, Vol.160, pp. 772-777, November 2019.

[25] R. Vasja, M. Maja, K. Alojz, “A Complex View of Industry 4.0”, SAGE Journals, pp. 1-11, June 2016.

[26] K. Zhou, T. Liu, L. Zhou, “Industry 4.0: Towards future industrial opportunities and challenges” 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), January 2016.

[27] L. Ardito, A. M. Petruzzelli, U. Panniello, A. C. Garavelli, “Towards Industry 4.0: mapping digital technologies for supply chain management-marketing integration,” Business Process Management Journal. Vol. 25, No. 2, pp. 323-346, March 2019.

[28] E. Manavalan, K. Jayakrishna, “A review of Internet of Things (IoT) embedded sustainable supply chain for industry 4.0 requirements,” Computers and Industrial Engineering, Vol. 127, pp. 925-953, January 2019.

[29] J. Zhu, H. Zhou, C. Wang, L. Zhou, S. Yuan, W. Zhang, A review of topology optimization for additive manufacturing: status and challenges. Chinese Journal of Aeronautics.

[30] H. Zhong, S. Y. Nof, “The dynamic lines of collaboration model: collaborative disruption response in cyber-physical systems,” Computere and Industrial Engineering, Vol. 87, pp. 370-382. September 2015.

[31] S. Subashini, V. Kavitha, “A survey on security issues in service delivery models of cloud computing,” Journal of Network and Computer Applications. Vol. 34, No.1, pp. 1-11, January 2011.

[32] E. Hofmann, M. Rusch, “Industry 4.0 and the current status as well as future prospects on logistics,” Computers in Industry, Vol. 89, pp. 23-34, August 2017.

[33] J. Schlechtendahl, M. Keinert, K. Kretschmer, A. Lechler, A. W. Verl, “Making existing production systems Industry 4.0-ready,” Production Engineering. Vol. 9, No. 1, pp. 1-6. February 2014.

[34] Q. Zhang, L. Cheng, R. Boutaba, “Cloud Computing: State-Of-The-Art and Research Challenges,” Journal of Internet Services and Applications, No. 1, pp. 7-18, April 2020.

[35] H. P. Breivold, K. Sandstrom, “Internet of Things for Industrial Automation Challenges and Technical Solutions,” IEEE International Conference on Data Science and Data Intensive Systems, pp. 532-539, December 2015.

[36] Ahuett-Garza H. and Kurfess T. “A brief discussion on the trends of habilitating technologies for Industry 4.0 and Smart manufacturing,” Manufacturing Letters, Vol. 15, pp. 60-63. January 2018.

[37] A. Alharthi, V. Krotov, B. Michael, “Addressing barriers to big data,” Business Horizons, Vol. 60, No. 3, pp. 285-292. February 2017.

[38] M. D. Ryan, “Cloud computing privacy concerns on our doorstep,” Communications of the ACM, Vol. 54, No.1, 36-38, Junuary 2011.

[39] K. Popovic´, Z. Hocenski, “Cloud computing security issues and challenges,” The 33rd International Convention MIPRO, May 2010.

[40] R. Henzel, C. Ellwein, G. Herzwurm, O. Riedel “Cloud manufacturing: Theoretical localization within production theory,” Procedia CIRP, Vol. 99, pp. 555-560, 2021.

[41] Y. Ji, P. Liu, Y. Li, B. N. Bekele, “A modified toxicity probability interval method for dose-finding trials,” Chin. Trials, Vol. 7, No. 6, pp. 653-663, October 2020.

[42] L. Ren, L. Zhang, F. Tao, C. Zhao, “Cloud manufacturing: from concept to practice,” Enterprise Information Systems, Vol. 9, No. 2, pp. 186-209, June 2013.

[43] B. H. Li, L. Zhang, S. L. Wang, F. Tao, et al., “Cloud Manufacturing: A New Service-Oriented Networked Manufacturing Model,” Computer Integrated Manufacturing Systems, Vol. 16, No. 1, pp. 1-7, January 2010.

[44] X. V. Wang, X. W. Xu, “An interoperable solution for Cloud manufacturing,” Robotics and Computer-Integrated Manufacturing, Vol. 29, No. 4, pp. 232-247, August 2013.

[45] X. Xun, “From cloud computing to cloud manufacturing,” Robotics and Computer-Integrated Manufacturing, Vol. 28, No. 1, pp. 75-86, February 2012.

[46] F. Tao, L. Zhang, V. C. Venkatesh, Y. Luo, Y. Cheng, “Cloud manufacturing: a computing and service-oriented manufacturing model,” Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, Vol. 225, No. 10, August 2011.

[47] A. Sanders, C. Elangeswaran, J. Wulfsberg, “Industry 4.0 implies lean manufacturing: Research activities in industry 4.0 function as enablers for lean manufacturing,” Journal of Industrial Engineer and Management, Vol. 9, No. 3, pp. 811-33, 2016.

[48] M. Pflanz, M. Schirrmann, H. Nordmeyer, “Drone based weed monitoring with an image feature classifier,” 28th German Conference on Weed Biology and Weed Control, No. 458, pp. 379-384, March 2018.

[49] P. M. Vergara, E. A. de la Cal Marin, J. Villar, V. M. Gonzalez, “An IoT Platform for Epilepsy Monitoring and Supervising,” Journal Sensors, Vol. 17, pp. 1-18, July 2017.

[50] S. R. Prathibha, A. Hongal, M. P. Jyothi, “IoT based monitoring system in smart agriculture,” International Conference on Recent Advances in Electronics and Communication Technology, pp. 81-84, March 2017.

[51] M. Lezoche, J. E. Hernandez, M. M. E. Alemany, H. Panetto, “Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture,” Computers in Industry, Vol. 117, pp. 103-187, March 2020.

[52] A. Adauto, S. Khan, M. A. da Silva, J. A. G. Neto, G. Picasso, M. D. P. T. Sotomayor, “Synthesis, characterization and application of a novel ion hybrid imprinted polymer to adsorb Cd(II) in different samples,” Environmental Research, Vol. 187, No.1, pp.1-10, August 2020.

[53] D. Trabucchi, T. Buganza, “Data-Driven Innovation: switching the perspective on Big Data,” pp. 1-22, January 2018.

[54] N. Santhosh, M. Srinivsan, K. Ragupathy, “Internet of Things (IoT) in smart manufacturing,” IOP Conf. Series: Materials Science and Engineering, Vol. 764, pp. 1-14, September 2019.

[55] H. Kagermann, “Change through digitization-Value creation in the age of Industry 4.0,” Management of Permanent Change, pp. 23-45, December 2015.

[56] J. Zhu, H. Zhou, C. Wang, L. Zhou, S. Yuan, W. Zhang, “A review of topology optimization for additive manufacturing: status and challenges,” Chinese Journal of Aeronautics, Vol. 34, No. 1, pp. 91-110, January 2021.

[57] F. Tao, Y. Cheng, L. D. Xu, et al., “CCIoT-CMfg: Cloud Computing and Internet of Things-Based Cloud Manufacturing Service System,” IEEE Transactions on Industrial Informatics,Vol. 10, No. 2, pp. 1435-1442, May 2014.

[58] C. G. Hidalgo, D. Hidalgo, L. R. Sanchez, et. al., “IoT Heterogeneous Mesh Network Deployment for Human-in-the-Loop Challenges Towards a Social and Sustainable Industry 4.0,” IEEE Access, Vol. 6, pp. 28417-28437, May 2018.

[59] C. Cimini, F. Pirola, R. Pinto, S. Cavalieri, “A human-inthe-loop manufacturing control architecture for the next generation of production systems,” Journal of Manufacturing Systems, Vol. 54, pp. 258-271, January 2020.

[60] F. Rengier, A. Mehndiratta, H. von Tengg-Kobligk, et al., “3D printing based on imaging data: Review of medical applications,” International Journal of Computer Assisted Radiology and Surgery, Vol. 5, No. 4, pp. 335-341, May 2010.

[61] R. Weller1, K. Schmidt1, K. Teinilä, R. Hillamo, “Natural new particle formation at the coastal Antarctic site Neumayer,” Atmospheric Chemistry and Physics, Vol. 15, No. 19, pp. 11399-11410, October 2015.

[62] S. M. Khajavi, J. Holmström, J. Partanen, “Additive manufacturing in the spare parts supply chain: hub configuration and technology maturity,” Rapid Prototyping Journal, Vol. 24, pp. 1178-1192, October 2018.

[63] M. Rehnberg, S. Ponte, “From smiling to smirking? 3D printing, upgrading and the restructuring of global value chains,” Global Networks, Vol. 18, No. 1, pp. 57-80. June 2017.

[64] L. Monostori, Cyber-physical production systems: roots, Expectations and R&D challenges, Procedia CIRP, Vol. 17, pp. 9-13, July 2014.

[65] M. J. Mazzei, D. Noble, “Big Data Dreams: A Framework for Corporate Strategy,” Business Horizons, Vol. 60, No. 3, pp. 405-414, June 2017.

[66] H. F. Cervone, “Organizational considerations initiating a big data and analytics implementation,” Digital Library Perspectives, Vol. 32. No. 3, pp. 137-141, August 2016.

[67] M. Garetti, L. Fumagalli, E. Negri, “Role of Ontologies for CPS Implementation in Manufacturing,” Management and Production Engineering Review, Vol. 6, No. 4, pp. 26-32, December 2015.

[68] T. Yamane, “Statistics: An introductory analysis,” New York, 1967.

[69] L. A. Zadeh, “Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems,” Advances in Fuzzy Systems – Applications and Theory, Vol. 6, pp. 394-432, May 1996.

[70] A. Taweesangrungroj, R. Rattanabanchuen, S. Sinthupinyo. “A Decision-Making Model of Technological-Focused Government Agency Selection of Technological Start-Up Businesses. Trends in Sciences, Vol. 19, No.1, pp. 1749-1749, January 2022.

[71] A. Taweesangrungroj, R. Rattanabanchuen, S. Sinthupinyo, “Decision-making process of Thai government agent for selection of technological startup companies in pitching of concept ideas,” Asian Journal of Information and Communications, Vol.11, No.2, 2019.

[72] S. Sinrat, W. Atthirawong, Integrated Factor Analysis and Fuzzy Analytic Network Process (FANP) Model for Supplier Selection Based on Supply Chain Risk Factors. Research Journal of Business Management, Vol. 9, pp. 106-123, October 2014.

The Prediction of Pediatric Outpatient No-Show Visits by Employing Machine Learning Framework

Abdulwahhab Alshammari[1], Hanoof Alaboodi[2], Riyad Alshammari[3]
College of Public Health and Health Informatics,
King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia[1, 2]
King Abdullah International Medical Research Center, Riyadh, Saudi Arabia[1, 2]
Ministry of the National Guard-Health Affairs, Riyadh, Saudi Arabia[1, 2]
Pediatric Department, Ministry of the National Guard-Health Affairs, Riyadh, Saudi Arabia[2]
National Centre for Artificial Intelligence (NCAI),
Saudi Data and Artificial Intelligence Authority (SDAIA), Riyadh, Saudi Arabia[3]

***Abstract*—Patient no-show for a booked medical appointment is a significant problem that negatively impacts healthcare resource utilization, cost, efficiency, quality, and patient outcomes. This paper developed a machine learning framework to predict pediatric patients' no-shows to medical appointments accurately. Thirty months of outpatient visits data were extracted from data warehouse from January 2017 to July 2019 of the Ministry of National Guard Health Affairs (MNGHA), Saudi Arabia. The researchers retrieved the data from all healthcare facilities in the central region, and more than 100 attributes were generated. The data includes over 100,000 pediatric patients and more than 3.7 million visits. Five machine learning algorithms were deployed, where Gradient Boosting (GB) algorithm outperformed the other four machine learning algorithms: decision tree, random forest, logistic regression, and neural network. The study evaluated and compared the performance of the five models based on five evaluations criteria. GB achieved a Receiver Operating Characteristic (ROC) score of 97.1%. Furthermore, this research paper identified the factors that have massive potential for effecting patients' adherence to scheduled appointments.**

***Keywords*—No-show; machine learning; healthcare medical appointments; predictive analytics**

I. Introduction

Outpatient no-show in a healthcare setting is the non-attendance of confirmed medical appointment by the patient [1-2]. The outpatient medical appointment no-show is a significant issue for healthcare facilities since utilizing resources is ineffective and widens the appointment waiting time. Thus, accurately predicting the patient no-show will reduce the financial cost, increase productivity, enhance the quality of care noticeably, and increase the patient's satisfaction [2-8] considerably.

The demand for outpatient medical services in the Kingdom of Saudi Arabia (KSA) is increasing substantially [5]. There is an average of 70K medical appointments at the Ministry of National Guard Health Affairs (MNGHA) central region medical, of which around 10% are related to pediatric cases. Presently, there is no effectible deployed digital tool in the Electronic Health Records (EHR) that can accurately predict the patient of high risk of no-show [5-6, 9]. As a result, the healthcare in KSA could utilize the advancements in machine learning techniques to construct a digital solution that can identify outpatients with a high possibility of not attending their medical reservation.

Machine learning can be used to build an intelligent digital solution that can provide healthcare with a data-driven system to enhance the patient's adherence to the medical appointment and predict patient behavior toward non-attendance of a scheduled medical appointment. Historical collected data in the EHR system can be utilized to forecast future patient visits. Therefore, an intelligent digital health solution can allow healthcare facilities to strategically create and manage a long-term projection plan for their medical resources [10-16].

This research study aims to develop an intelligent data-driven approach based on a machine learning technique to learn from more than three million extracted pediatric medical records from MNGH database systems to predict outpatient no-show medical appointments smartly. Other objectives are to predict patient behavior and identify the aspects that can be useful to forecast no-shows.

The remaining research paper is organized as follows. Section II covers the related work, while Section III details the methods. The results and discussion are presented in Section IV. Finally, Section V gives the conclusions and future work.

II. Related Work

This section summarized machine learning solutions to predict no-shows in healthcare facilities. Most of the articles in the literature focused on adult cases. Harvey et al. [17] used a logistic regression algorithm to predict the non-appearance of patients from booked radiology examinations. They achieved 75.3% for the Area Under the Receiver (AUC) operator. Chua and Chow [18] extracted a no-show administrative dataset and employed Multiple Logistic Regression (MLR). They scored 72% for AUC. Dantas et al. [19] collected more than 13K records related to bariatric surgery no-show appointments. They achieved an accuracy of 71% using the Logistic Regression algorithm. Likewise, Kurasawa et al. [20] used a Logistic Regression algorithm to forecast the no-show of diabetic patients. Their best model scored 75.7% and 65.9% for precision and recall, respectively.

Furthermore, Mohammadi et al. [21] achieved the best result of 86% of AUC using Naive Bayes on a dataset of 73,811 records, while AlMuhaideb et al. [22] achieved 86.1% of AUC using Hoeffding algorithms on outpatient no-show appointments. Goffman et al. [23] used Logistic Regression on a dataset based on demographic and appointment descriptions and the past patient's activities. They achieved an AUC of 71%. Moreover, Nelson et al. [24] created a predictive model for imaging no-show appointments. They got the best result with the Gradient Boosting algorithm, with 85% and 51.1% for AUC and precision, respectively. Lee *et al.* [25] collected two years of data on no-shows, where they applied three machine learning algorithms: Logistic Regression, Decision Tree, and Random Forest. Random Forest achieved the best results (accuracy of 72.9%).

III. METHODOLOGY

A. Dataset and Attributes

Thirty months of outpatient visits were extracted from the Ministry of National Guard Health Affairs (MNGHA) data warehouse from January 2017 to July 2019. The data were retrieved from all the medical facilities in the central region. The central region has the largest medical facilities of the MNGHA, where more than 70K appointments are booked monthly. The retrieved dataset consists of more than 3,733,580 million pediatric patient visits related to 104,640 pediatric patients. The patient's arrival time at the clinic is used to label the record as show (timestamp) or no-show (no timestamp).

The distribution of genders is almost equal, with 49.3% as female patients and 50.7% as male. In comparison, most patients were Saudi nationals (90%) since the hospital mainly serves employees of national guards and their families, Table I. However, the Saudis have the highest no-show rate of 99% (miss at least one medical appointment), Table I.

TABLE I. STATISTICAL DESCRIPTION OF PATIENTS (N= 104640)

Attributes	*No-Show N%*	*Show N%*	*Total %*
Age: Infant (0-12 Months) Toddler (1-3 Years) Preschool (3-6 Years) School-age (6-12 Years) Adolescent (12-14 Years)	6651 (90.5%) 19187 (64.1%) 7986 (63.9%) 16733 (56.8%) 12422 (48.8%)	701 (9.5) 10729 (35.9%) 4503 (36.1%) 12719 (43.2%) 13009 (51.2%)	7352 (7%) 29916 (28.6%) 12489 (12%) 29452 (28.1%) 25431 (24.3)
Gender: Female Male	22202 (47.5%) 24427 (52.5%)	29373 (50.6%) 28638 (49.4%)	51575 (49.3%) 53065 (50.7)
Nationality: Non-Saudi Saudi	362 (0.7%) 46267 (99.3%)	696 (1.1%) 57315 (98.9%)	1058 (1%) 103582 (99%)

The data warehouse team filtered out unnecessary records to reduce the noise in the extracted data. Categorical attributes such as gender and age were converted into an integer. The age attribute was calculated based on the difference between the birthday and appointment dates. The age attribute was grouped into five categories: i) infant (0-12 months), ii) Toddler (1-3 years,) iii) Preschool (3-6 years), iv) School-age (6-12 years), and v) Adolescent (12-14 years). A new attribute, lead days, was derived from the dataset. The lead days attribute is a derived attribute calculated as the number of days between the booking of the appointment and the scheduling day. The historical behavior of the patients was included since it could contain helpful information to predict future outcomes. Therefore, the number of walk-ins, scheduled appointments, emergency visits, and canceled medical appointments were included for all patients.

TABLE II. DESCRIPTION OF THE ATTRIBUTES

Attribute	*Description*
1) Gender	Male, Female
2) Age Category	Five age categories
3) Nationality	The nationality of the patient
4) Medical Department code	The code of medical/clinic
5) Hospital code	The code of referral hospital /clinic
6) Patient Services Department Type: a) Patient service b) Business center	Type of patient insurance
7) Address code	The primary health care clinic location (used for the patient's address)
8)Appointment type code: New Patient (NP) First visit (FV) Follow-up (FU)	Type of Appointment New Patient First Visit Follow-up Visit
9) Patient Services Department Type Cod: First visit (FV) Follow up (FU)	Type of patient insurance First Visit Follow-up Visit
10) Lead Days	Derived: the time difference in days between the booking of the appointments and the day of the appointments
11) Cancellation Flag: a) Yes= appointment cancelled b) No = not cancelled	Derived: count how many Check-ins Cancellation was Yes as total Derived: count how many Check-ins Cancellation was No as total
12) Medical Treatment Reservation Type Code: a) 1= Schedule b) 2= Walk-in	Derived: count how many Scheduled reservations as the total Derived: count how many Walk-in reservations as total
13) Medical Treatment a) Yes= patient treated b) N= not treated	Derived: count how many patients treated as total Derived: count how many patients not treated as total
14) On Foot visit a) 1= Yes b) 0 = No	Derived: count how many On Foot visit was Yes as total Derived: count how many On Foot visit was No as total
15) Emergency visit a) 1= Yes b) 0 = No	Derived: count how many Emergency visits were Yes as total Derived: count how many Emergency visits were No as total
16) Flu Season	Flu season from Oct-Nov and March-April
17) Distance (km)	The distance between the referring clinic and the Central Medical center of MNGHA at Riyadh City
18) Class No-Show	class show=0 or No-show=1

The historical visits of the patients are grouped together, and nine statistical values were calculated for 11 attributes (attributes 7 to 17 from Table II) which were minimum, maximum, standard deviation, variance, mean, median, skewness, and kurtosis. After the preprocessing phase, the dataset consists of 558,721 records with 106 attributes, including the class; Table II shows the description of the attributes.

B. Machine Learning Algorithms and Evaluation Criteria

This study involves five machine learning algorithms for classification data tasks. That includes Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR), Neural Network (NN), and Decision Tree (C4.5). RF and C5.4 are machine-learning algorithms from the family of decision-tree class algorithms. Gini impurity, information gain, and other techniques are used to build the tree structure; further algorithms description can be found in [26-29].

On the other hand, the Gradient Boosting algorithm is a boosting algorithm. It utilizes a number of weak classifiers through the ensemble method to build strong learners. The algorithm used gradient descent with other techniques to complete building the model; further algorithm descriptions can be found in [26,30. Logistic Regression (LR) can be described as a sample Neural Network (NN). LR can be viewed as a one-layer NN. LR and NN utilized many techniques to build the classification model of a binary class; further algorithm descriptions can be found in [26].

We evaluated and compared the five model's performance based on four evolution criteria: sensitivity (recall), precision, F-score, and accuracy. We considered True Positive (TP) Rate, False Positive (FP) Rate, True Negative (TN), False Negative (FN) Rate, and Receiver Operating Characteristic (ROC). These are defined as:

- True Positive (TP) Rate represents the number of no-show patient events classified as No-show, calculated based on equation 1:

$$TPR=TP / (TP + FN) \tag{1}$$

- False Positive (FP) Rate represents the number of show patient events classified as a no-show, calculated based on equation 2:

$$FPR=FP / (FP + TN) \tag{2}$$

- True Negative (TN) Rate represents the number of show patient events classified as a show, calculated based on equation 3:

$$TPR=TN / (TN + FP) \tag{3}$$

- False Negative (FN) Rate represents the number of no-show patient events classified as a show, calculated based On equation 4:

$$FPR=FN / (FN + TP) \tag{4}$$

- Receiver Operating Characteristic (ROC) score is a classification performance to show a false positive rate versus a true positive rate across a series of cut-off points and selecting the optimal cut-off point.

IV. Results and Discussion

The dataset was divided into 80% training and 20% testing. Table III shows the performance of the machine learning algorithms on an unseen testing dataset. Random Forest (RF) and Gradient Boosting (GB) outperformed the other classifiers on all five evaluation criteria. GB achieved 97.1% for ROC, which was 0.5% higher than RF. On the other hand, the score is higher than GB in terms of false positive rate (5.8% for RF and 5.9% for GB). In general, GB can be considered the champion model with an AUC of 97.1%.

Accurately predicting the intention of patients showing or no-showing to their medical appointment is considered an interesting and challenging goal for healthcare providers. Machine learning has become popular in healthcare research because of machine learning algorithms' ability to discover hidden patterns in the datasets, predict future outcomes and recognize the most relative attributes.

The machine learning algorithms were applied to more than half a million aggregated records with complex relations between attributes (predictors) and class labels with more than 100 attributes. The results of Gradient Boosting in predicting outpatient no-show outperforms traditional algorithms applied by previous research studies in Section II.

Analysis of the most used relevant attributes by machine learning shows that specific attributes significantly impact patient appointment adherence. The attributes are the traveling distance between the patient's residence and the medical center, the increased number of days between booking an appointment and the scheduled day (lead days), clinical medical services, and appointment time and day.

The rate of patient adherence to scheduled appointments can be increased by utilizing other resources. The medical facilities can use Short Message Service (SMS), an automated phone calls reminders 2-3 days before the medical appointment, mainly if the lead time is so long (more than four months). Furthermore, changing the lead days to be no more than 2-3 months [31-32].

Regarding model deployment, the machine learning framework used to build the model is a standard approach that can be integrated effortlessly within the digital healthcare system [33-34]. The research study has two limitations. Firstly, the dataset was retrieved from one region in KSA; therefore, more data from other regions can improve the robustness of the model if the model is deployed on different regions' digital systems. Secondly, the attributes are mainly extracted from an EHR in which other factors that influence the no-show rate are not included. These factors, such as personal reasons, availability of transportation, cultural background, educational level of the parents, and environmental conditions (e.g., weather circumstances), are not available in the digital healthcare system.

TABLE III. THE PERFORMANCE OF CLASSIFIERS ON TEST DATA

Evaluation Criteria	DT	RF	LR	NN	GB
ROC	94.8%	**96.4%**	95.0%	95.2%	**97.1%**
TPR	89.7%	**90.0%**	89.5%	89.7%	**90.0%**
FNR	10.3%	**10.0%**	10.5%	10.3%	**10.0%**
TNR	93.1%	**94.2%**	92.1%	91.9%	**94.1**
FPR	6.9%	**5.8%**	7.9%	8.1%	**5.9%**

V. Conclusion

In summary, machine learning's ability to acceptably predict no-shows provides a new potential digital solution for healthcare facilities to take advantage of intelligent solutions. This paper used five machine learning algorithms, namely, Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and Neural Network, applied to MNGH's extracted dataset from the data warehouse system.

The model was trained and tested using a dataset with different metrics to compare the results. Results show that Gradient Boosting achieves high performance for decision-making by the healthcare appointment management team. Furthermore, the study identified factors that increase the no-show rate, such as location, lead days between the reservation day and appointment day, appointment time of the day, and the type of medical services. Recommendations to increase the adherence of the no-show by patients have also been provided in Section IV.

Future work is to obtain more datasets from healthcare systems (for example, laboratory and medication data) and integrate weather data and patient background data such as educational level, transportation availability, and cultural background.

Disclosures

None of the authors have any competing interests.

References

[1] Huang Y, Hanauer DA. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. Appl Clin Inform. 2014;5(03):836–60.

[2] Denney J, Coyne S, Rafiqi S. Machine learning predictions of no-show appointments in a primary care setting. SMU Data Sci Rev. 2019;2(1):2.

[3] Chariatte V, Berchtold A, Akre C, Pierre-André M (2008) Missed Appointments in an Outpatient Clinic for Adolescents, an Approach to Predict the Risk of Missing. J Adolesc Heal 43:38–45.

[4] Carlsen KH, Carlsen KM, Serup J (2016) Non-attendance, predictors and interventions. Adherence in Dermatology 29–35.

[5] Davies M, Goffman R, May J, Monte R, Rodriguez K, Tjader Y, Vargas D (2016) Large-Scale No-Show Patterns and Distributions for Clinic Operational Research. Healthcare 4:15.

[6] O. AlRowaili M, Ahmed AE, Areabi HA (2016) Factors associated with No-Shows and rescheduling MRI appointments. BMC Health Serv Res 16:1–7.

[7] Norris JB, Kumar C, Chand S, Moskowitz H, Shade SA, Willis DR. An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. Decis Support Syst. 2014;1(57):428–43.

[8] Samorani M, Harris S, Blount LG, Lu H, Santoro MA. Overbooked and Overlooked: Machine Learning and Racial Bias in Medical Appointment Scheduling. Available at SSRN 3467047. 2019 Oct 9.

[9] Srinivas S, Ravindran AR (2018) Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. Expert Syst Appl 102:245–261.

[10] J. Daggy, *et al*., "Using no-show modeling to improve clinic performance," *Health Informatics J.*, vol. 16 pp. 246-259, 2010.

[11] D. Bentayeb, N. Lahrichi, and L. M. Rousseau, "Patient scheduling based on a service-time prediction model: A datadriven study for a radiotherapy center," *Health Care Manag. Sci.*, vol. 22, pp. 768-782, 2019.

[12] Z. Huang, M. Ashraf, H. Gordish-Dressman, and P. Mudd, "The financial impact of clinic no-show rates in an academic pediatric otolaryngology practice," Am. J. Otolaryngol., vol. 38, pp. 127- 129, 2017.

[13] Joseph, J., Senith, S., Kirubaraj, A. A., & Ramson, S. R. J. (2022). Machine Learning for Prediction of Clinical Appointment No-Shows. *International Journal of Mathematical, Engineering and Management Sciences*, *7*(4), 558-574.

[14] Chen J, Goldstein IH, Lin WC, Chiang MF, Hribar MR. Application of Machine Learning to Predict Patient No-Shows in an Academic Pediatric Ophthalmology Clinic. AMIA Annu Symp Proc. 2021 Jan 25;2020:293-302. PMID: 33936401; PMCID: PMC8075453.

[15] Daghistani, T., AlGhamdi, H., Alshammari, R. *et al.* Predictors of outpatients' no-show: big data analytics using apache spark. *J Big Data* **7**, 108 (2020).

[16] Alshammari, Riyad, Tahani Daghistani, and Abdulwahhab Alshammari. "The Prediction of Outpatient No-Show Visits by using Deep Neural Network from Large Data." *International Journal of Advanced Computer Science and Applications* 11.10 (2020).

[17] Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, Pianykh O (2017) Predicting No-Shows in Radiology Using Regression Modeling of Data Available in the Electronic Medical Record. J Am Coll Radiol 14:1303–1309.

[18] Chua SL, Chow WL (2019) Development of predictive scoring model for risk stratification of no-show at a public hospital specialist outpatient clinic. Proc Singapore Healthc 28:96–104.

[19] Dantas LF, Hamacher S, Cyrino Oliveira FL, Barbosa SDJ, Viegas F (2019) Predicting Patient No-show Behavior: a Study in a Bariatric Clinic. Obes Surg 29:40–47.

[20] Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, Waki K, Noguchi T, Ohe K (2016) Machine-Learning-Based Prediction of a Missed Scheduled Clinical Appointment by Patients with Diabetes. J Diabetes Sci Technol 10:730–736.

[21] Mohammadi I, Wu H, Turkcan A, Toscos T, Doebbeling BN (2018) Data Analytics and Modeling for Appointment No-show in Community Health Centers. J Prim Care Community Heal 9:1–11.

[22] AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. Ann Saudi Med. 2019;39(6):373 81.

[23] Goffman RM, Harris SL, May JH, Milicevic AS, Monte RJ, Myaskovsky L, Rodriguez KL, Tjader YC, Vargas DL. Modeling patient no-show history and predicting future outpatient appointment behavior in the Veterans Health Administration. Mil Med. 2017;182(5–6):e1708–14. 2017;14(10):1303–9.

[24] Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. NPJ digital Med. 2019;2(1):1–7.

[25] G. Lee, *et al.*, "Leveraging on predictive analytics to manage clinic no-show and improve accessibility of care," in *Proc. Int.* Conf. Data Sci. Adv. Anal., 2017, pp. 429-438.

[26] Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2020.

[27] Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. Systems Sci Control Engineering Open Access J. 2014;2(1):602–9.

[28] Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogrammetry Remote Sensing. 2012;1(67):93–104.

[29] J R Quinlan, "C4.5: Programs for Machine Learning",Morgan Kaufmann Publishers,isbn=1-55860-238-0, 1993.

[30] Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." *Annals of statistics* (2001): 1189-12.

[31] Agarwal P, Nathan AS, Jaleel Z, Levi JR. Factors Contributing to Missed Appointments in a Pediatric Otolaryngology Clinic. Laryngoscope. 2022 Apr;132(4):895-900. doi: 10.1002/lary.29841. Epub 2021 Aug 24. PMID: 34427327.

[32] Shahab I, Meili R. Examining non-attendance of doctor's appointments at a community clinic in Saskatoon. Can Fam Physician. 2019 Jun;65(6):e264-e268. PMID: 31189640; PMCID: PMC6738377.

[33] Hofer IS, Burns M, Kendale S, Wanderer JP. Realistically Integrating Machine Learning Into Clinical Practice: A Road Map of Opportunities, Challenges, and a Potential Future. Anesth Analg. 2020 May;130(5):1115-1118. doi: 10.1213/ANE.0000000000004575. PMID: 32287118; PMCID: PMC7584400.

[34] Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a Machine Learning Model Applied to Harmonized Electronic Health Record Data for the Prediction of Incident Atrial Fibrillation. JAMA Netw Open. 2020 Jan 3;3(1):e1919396. doi: 10.1001/jamanetworkopen.2019.19396.

Principal Component Analysis Based Hybrid Speckle Noise Reduction Technique for Medical Ultrasound Imaging

Yasser M. Kadah[1]*, Ahmed F. Elnokrashy[2], Ubaid M. Alsaggaf [3], Abou-Bakr M. Youssef [4]
Electrical and Computer Engineering Department, King Abdulaziz University, Jeddah, Saudi Arabia[1, 3]
Electrical Engineering Department, Benha University, Benha, Egypt[2]
Biomedical Engineering Department, Cairo University, Giza, Egypt[1, 4]

***Abstract*—Ultrasound imaging is the safest and most widely used medical imaging technique available today. The main disadvantage of ultrasound imaging is the presence of speckle noise in its images that may obscure pathological changes in the body and makes diagnosis more challenging. Therefore, many techniques were proposed to reduce speckle and improve image quality. Unfortunately, variations of their performance with different scan parameters and due to their methodologies make it hard to choose which one to adopt in clinical practice. In this work, we consider the problem of combining the information from multiple speckle filters and propose the use of principal component analysis to find the optimal set of weights that would retain the most information and hence would better represent the data in the final image. The new technique is implemented to process ultrasound images collected from a research system and the outcomes are compared to the individual techniques and their average using quantitative image quality metrics. The proposed technique has potential for utilization in clinical settings to provide consistently better-quality combined images that may help improve diagnostic accuracy.**

Keywords*—*Image quality metrics; principal component analysis; speckle reduction; ultrasound imaging

I. Introduction

Ultrasound imaging is one of the most widely used medical imaging techniques available today. It offers a versatile tool for scanning soft tissues and blood flow in the body with applications in abdominal, cardiac, vascular, musculoskeletal, obstetrical and gynecological imaging. Since pulse-echo ultrasound imaging uses weak ultrasonic pulses in human tissues, it is presently considered among the safest medical imaging modalities. It is even allowed to be used for obstetrical applications to assess the health of the fetus. Furthermore, ultrasound imaging systems are available in different forms that include smartphone-based systems or portable scanners that target basic applications such as abdominal imaging and are available at very low cost. They also extend to high-end systems that target specialized applications such as surgical or echocardiographic systems at much higher cost. Therefore, the utility of ultrasound imaging technology extends across the different layers of healthcare delivery to include populations in low-income areas all the way to those receiving care at specialized hospitals. The versatility of ultrasound imaging is expected to lead the technology to become a necessary tool for general practitioners just like the stethoscope.

The main barrier to wider adoption of ultrasound imaging in medicine is the noisy appearance of its output images, which makes the task of reading images more challenging compared to several other modalities [1]. This presence of noise causes problems in the interpretation of ultrasound images particularly when slight pathological variations are present within. Consequently, the problem of image denoising in ultrasound images has been addressed by research from academic and industrial researchers. Improving the quality of ultrasound images by reducing noise is likely to expand its utility and introduction into new applications [2].

In addition to the random noise expected from the electronics used in collecting ultrasound imaging echoes, a deterministic yet unknown type of noise-like texture called speckle noise appears as a direct result of the way ultrasound imaging is performed [3]. Ultrasound imaging works by transmitting an ultrasonic wave burst through the tissues from a probe. This wave travels through the different layers of tissues producing reflected and scattered echoes [4]. The reflection of waves happens when the size of interfaces is larger than the wavelength of ultrasound waves, while scattering happens when the interfaces are significantly smaller than this wavelength [1]. The ultrasound wavelength is less than a millimeter in the frequency range of ultrasound imaging and hence, organ boundaries and walls of large blood vessels produce reflected waves, while small vessels and the cellular matrix within organs produce scattered waves [3][4]. Since such interaction modes depend on the ultrasound transducer shape, frequency and orientation as well as the complex three-dimensional shapes, sizes and organizations of the interfaces within the tissues, the interference pattern of echoes from many interactions results in a pattern of constructive and destructive interferences that appears as a pseudo-random noise in the final image called speckle noise [5]. The speckle noise is different from the random noise encountered from thermal noise in the analog front-end electronics for example in that the speckle pattern remains stationary as long as the imaging conditions are maintained. On the other hand, random noise pattern changes with time and therefore temporal averaging is usually done to reduce its effect on the image [6]. Therefore, speckle reduction cannot be done with temporal averaging and requires fundamentally different approaches to perform [7][24].

*Corresponding Author.

In this work, we consider the problem of combining the information from multiple speckle filters and propose the use of principal component analysis to find the optimal set of weights that would retain the most information and hence would better represent the data in the final image. The new technique is implemented to process ultrasound images acquired from an ultrasound imaging research system and the outcomes are compared to the individual techniques and their average using quantitative image quality metrics. The proposed technique has potential for utilization in clinical settings to provide consistently better-quality combined images that may help improve diagnostic accuracy.

II. Literature Review

Many research studies were performed to address the problem of speckle reduction. Based on their fundamental approach, the research in this area can be generally categorized as being either acquisition-based or postprocessing-based. The acquisition-based approach works by collecting and averaging several images with slightly different imaging parameters such that their speckle patterns are different [1][5][6]. Even though this approach offers an apparently simple solution that takes advantage of the speckle formation physics, it is practically difficult to implement and costly on ultrasound imaging systems and adds constraints on the maximum frame rate that may limit some clinical applications.

On the other hand, the postprocessing-based approach has received more attention from most research groups given that it uses the acquired image without any modifications to the data acquisition method on the ultrasound imaging system. The techniques in this category apply image filtration to suppress speckle noise while trying to maintain the other image features such as the edges. Therefore, a general-purpose digital image processing system with access to the image storage of the ultrasound imaging system can be used to perform its job. As such, it can work on an external computer with a frame grabber or other means to collect ultrasound images from an existing ultrasound imaging system without making any changes to it. With the advances made in computing hardware and software including parallel processing and graphics processing units (GPUs), this approach offers versatile solution that meets practical need to work with existing ultrasound imaging systems.

The existing postprocessing methods can be broadly classified based on their filtration strategy into four distinct classes, with many hybrids across them. Based on how the filtration is done, such methods use linear, nonlinear, physics-based methods like anisotropic diffusion, or wavelet shrinkage filtration [6]. The linear filtration works in the image domain using spatial domain filters and includes such techniques as first-order statistics filtering, local statistics filtering with higher moments, and homogeneous mask area filtering [8][9][10]. The nonlinear filtration is similar to linear filters in being image-based method but the spatial domain filters in this class relies on nonlinear functions such as median filtering, linear scaling filter, geometric filtering, and homomorphic filtering [11][12][13][14]. The physics-based class attempts to utilize simulations of physical phenomena such as diffusion to reach a more homogeneous texture and reduce speckle noise. Examples of this class include anisotropic diffusion filtration and its variants [7][15][16][17][18]. The wavelet shrinkage methods decompose the texture in the wavelet domain and truncation of small coefficients related to speckle noise [19][20][21][22]. Several hybrid methods that combine techniques from two or more of these classes between the above methods were also studied in the literature [23][24].

In spite of the good performance reported by the above speckle reduction methods to improve ultrasound images, there remain major issues that hinder their clinical utility in routine practice. The most important of these is that their performance has wide variability across different techniques and even within the same technique for different ultrasound image characteristics. This makes it difficult to choose one technique to adopt in a particular clinical setting given the wide variability of image characteristics with different selected imaging parameters such as ultrasound frequency, depth, and ultrasound probe geometry. Furthermore, the limited success in combining advantages from different categories of speckle reduction techniques with the present hybrid methods indicates a gap in the research within this area. Therefore, the development of a hybrid technique that would allow for a consistent performance while offering a way of combining image quality advantages from different categories would be highly desired to boost performance in clinical use.

Principal component analysis (PCA) was utilized in ultrasound imaging applications using different approaches. In [25], the authors proposed a speckle reduction method based on the segmentation of the ultrasound image into overlapping regions followed by the application of PCA to the set of segmented regions after reshaping each into a vector. This is similar in principle to the approach in [26] in other image denoising applications where local pixel grouping was used before PCA, and a multistage process is applied iteratively to improve denoising performance. Also, a similar approach was used in [27], where the authors also proposed the use of PCA to denoise multi-frame optical coherence tomography data after dividing them into 3D blocks. In [28], the authors presented a survey of blind source separation applications such as PCA in ultrasound imaging and suggested that it can be used to filter out clutter and noise using spectral thresholding. They pointed out the difficulty to make such techniques work adaptively because of the variations in clutter and noise subspaces across different applications. In [29], the authors proposed an approach for speckle noise suppression in medical ultrasound Images by combining PCA and nonlinear diffusion. In this technique, the PCA is used as an orthogonal transformation to a domain where the nonlinear diffusion is applied after bit plane slicing then the results is transformed back using inverse PCA to form the denoised image. In [30], the authors suggested the use of PCA to reduce the speckle noise in echocardiography frame sequences Using PCA. They reshaped each frame into a vector and performed PCA on the set of vectors obtained from the frame sequence. The resultant PCA underwent a process similar to wavelet shrinkage whereby the subspace components corresponding to small eigenvalues below a certain threshold were omitted and the remaining components were used to reconstruct the denoised image. In

[31], the authors proposed a hybrid of several denoising approaches where PCA is used to combine the outputs of those approaches to obtain the denoised image.

Even though the previous studies included interesting applications of PCA in different aspects of ultrasound imaging, their techniques did not address the problem of optimal combination of information from multiple speckle reduction categories. This problem will be addressed in this work.

III. Methodology

The received signal in ultrasound imaging can be modeled as the sum of a true signal coming from reflections and scattering from underling tissue interfaces and a superimposed random noise mainly from the electronics used in the analog front-end of the system. The true signal component can be further subdivided into signal from specular reflectors that delineate the major interfaces within the field of view, and another component coming from the summation of many wavelets coming from scatterers resulting in pseudo-random pattern of signal intensities called speckle. Unlike random noise, the speckle pattern depends on the underlying tissue characteristics such as cell sizes and distributions. Therefore, while the random noise changes with every acquisition, the speckle pattern remains the same provided that the same imaging conditions such as location of probe and imaging frequency and transmission focal points remain unchanged.

This study starts by recognizing the fact that there are many present techniques for speckle reductions that perform very well in some applications while others perform better in other applications. Rather than attempting to propose yet another technique, this study searches for a way to develop a hybrid method that combines several techniques in such a way to perform consistently at the peak performance. The research hypothesis is that using techniques from different speckle reduction categories combined using a technique that minimizes information loss such as principal component analysis would achieve that goal.

From the theory of principal component analysis technique, the first principal component is a smart way of representing the information present in a set of data vectors (or observations) whereby a set of weights are calculated such that it adaptively and optimally retains more information from the original data vectors than any other combination [36]. For example, in theory, the first principal component allows better representation of the original data than simple combination with equal weights as in averaging. This is the basis for our new method to combine data from different speckle reduction filters. In particular, four popular speckle reduction filtering techniques as examples from the four main postprocessing categories (without loss of generality) are used along with the original image as the input to the principal component analysis to find the optimal way of combining their information. These techniques are wavelet shrinkage [19][21], relaxed median (RMedian) filtering [13][14], speckle reducing anisotropic diffusion (SRAD) [15][16][17], and local statistics based filtering (Lee) [8][9]. The implementation parameters were used as in the most recent variant in each technique.

The details of the new methodology are as follows. Ultrasound images are collected as a set of lines (or sticks) that span the scanned region in a linear or a sector manner depending on the imaging probe used. The resulting data are called the stick data and are used to form the properly formatted output image by the image reconstruction technique given the scanning parameters and geometry information. The use of stick data rather than image lines maintains all data points independent by avoiding point correlations from interpolation operations. They also help reduce the computational cost of applying different processing methods on the data. The new technique is implemented by forming a matrix containing the set of vectors with the original data and four filtered versions and use PCA to find the optimal weights to combine them.

Let the original stick data image be $f(m,n)$ of size $M \times N$. The original stick data image is processed using four speckle filtering techniques, namely, wavelet, relaxed median, SRAD and Lee filters. The outcomes from these four filters will be $f_W(m,n)$, $f_R(m,n)$, $f_S(m,n)$ and $f_L(m,n)$ respectively. The observations data matrix D is composed by reformatting the original and four processed stick data images of size $M \times N$ into $MN \times 1$ column vectors, $F, F_W, F_R, F_S,$ and F_L, where each is placed as a column in the D matrix of size $MN \times 5$.

$$\mathrm{D} = [F \mid F_W \mid F_R \mid F_S \mid F_L]. \tag{1}$$

The covariance matrix $\mathrm{COV_D}$ of size 5×5 is estimated as,

$$\mathrm{COV_D} = \left[\, \hat{F} \mid \hat{F}_W \mid \hat{F}_R \mid \hat{F}_S \mid \hat{F}_L \,\right]^T . \left[\, \hat{F} \mid \hat{F}_W \mid \hat{F}_R \mid \hat{F}_S \mid \hat{F}_L \,\right], \tag{2}$$

where $\hat{F},\ \hat{F}_W, \hat{F}_R, \hat{F}_S\ and\ \hat{F}_L$ are the centered versions of their original vectors after subtracting their mean values. Then, the principal component analysis is performed on $\mathrm{COV_D}$ to obtain the vector of coefficients W_{PC} of the principal component (that is, the eigenvector for the largest eigenvalue of the covariance matrix). The reformatted column vector G of size $MN \times 1$ representing the final processed stick data image $g(m,n)$ of size $M \times N$ is then computed as the weighted average of all images as,

$$G = \frac{\mathrm{D} \cdot W_{PC}}{\mathrm{sum}\,(W_{PC})}. \tag{3}$$

Here, $\mathrm{sum}\,(W_{PC})$ is the summation of W_{PC} vector components to ensure that the weighted average weights add to unity. The final processed stick data image $g(m,n)$ of size $M \times N$ is obtained by reformatting G and is sent to image reconstruction to generate the final display image. A block diagram of the new method is presented in Fig. 1.

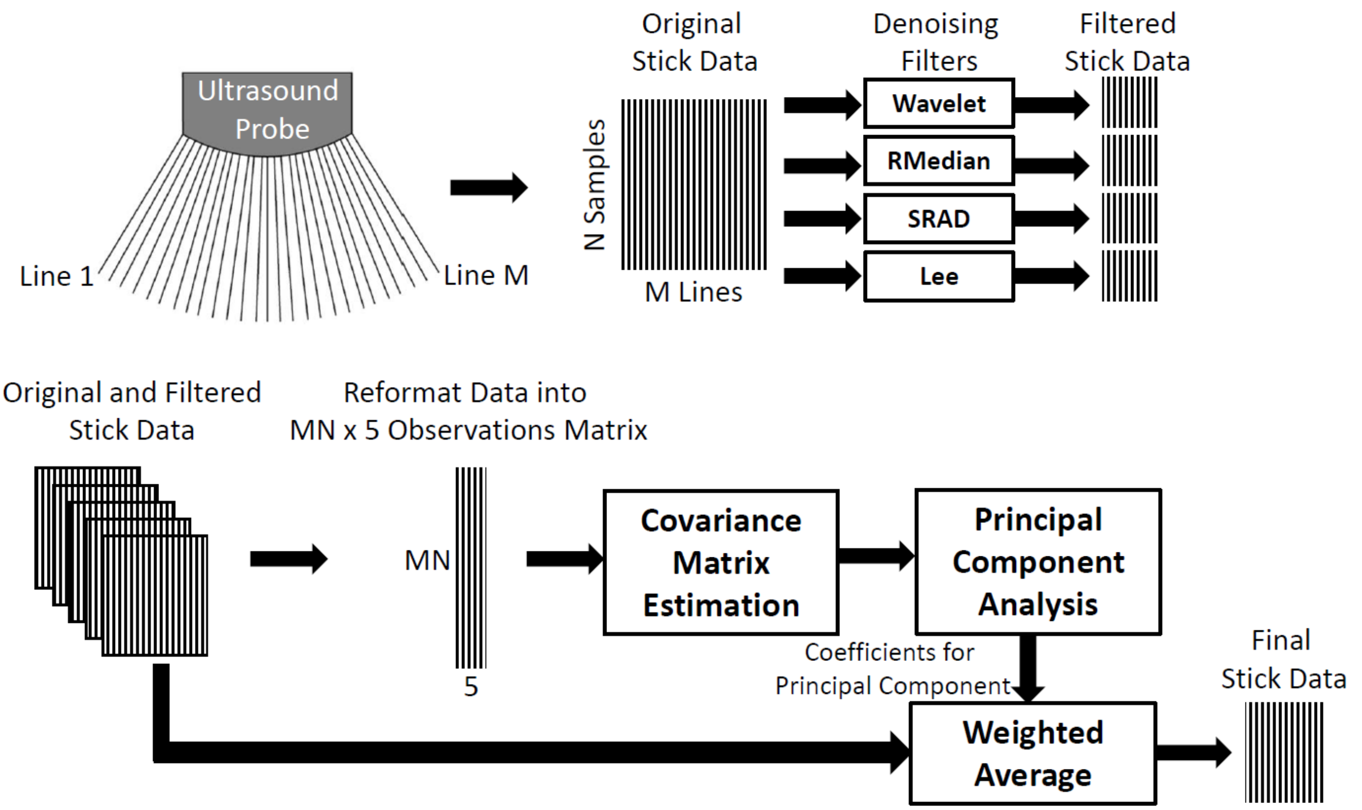


Fig. 1. Diagram showing the steps of the new technique whereby original stick data and its filtered versions multiple speckle reduction techniques undergo PCA to estimate the best weights to combine them for best information retention.

To compare the results between the new technique and averaging in a quantitative manner, several image quality metrics are used to assess the outputs from both techniques as they were applied to several data sets acquired using different imaging conditions (for example, different probes and different applications). The image quality metrics used are root mean squared error [6], Laplacian mean squared error [32], Minkowski error metric [33], structural similarity index (SSIN) [33], , universal quality index (Q) [33][34], signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR) [35]. This allows for objective assessment of the results. The detailed mathematical definitions of such metrics are as follows.

The root mean squared error (RMSE) is the square root of the average of the squared error over the whole image. It is generally considered as an approximation of the standard error.

$$RMSE = \sqrt{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}(g(m,n) - f(m,n))^2}\,. \quad (4)$$

The Laplacian mean squared error (LMSE) is the squared error between the Laplacian of the original and processed images averaged over the entire image. It generally assesses the edge preservation in the processing algorithm. A better-quality processed image will have a lower RMSE value.

$$LMSE = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}(g_L(m,n) - f_L(m,n))^2\,, \quad (5)$$

where $g_L(m,n)$ and $f_L(m,n)$ are the Laplacians of the processed and original images respectively. A better-quality processed image will have a lower LMSE value.

The Minkowski error metric computes the norm of the difference between the original and processed images using different vector norms. The equation for β-norm Minkowski error metric is given as,

$$ErrorM(\beta) = \sqrt[\beta]{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}|g(m,n) - f(m,n)|^{\beta}}\,. \quad (6)$$

Here, we compute the Minkowski error metrics for β values of 3 (ErrorM3) and 4 (ErrorM4). A better-quality processed image will have lower ErrorM3 and ErrorM4 values.

The signal-to-noise ratio (SNR) is an important measure of how the processed signal improved in suppressing noise. A better-quality processed image will have a higher SNR value.

$$SNR = 10\log_{10}\left(\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}(g^2(m,n) - f^2(m,n))}{\sum_{m=1}^{M}\sum_{n=1}^{N}(g(m,n) - f(m,n))^2}\right). \quad (7)$$

The peak signal-to-noise ratio (PSNR) measures the resemblance of the processed image and the original image. A better-quality processed image will have a higher PSNR value.

$$PSNR = -10\log_{10}\left(\frac{RMSE^2}{\max\{f(m,n)\}}\right), \quad (8)$$

where $\max\{f(m,n)\}$ is the maximum intensity value in the original image.

The universal quality index (Q) models any distortion as a combination of three different factors, which are loss of correlation, luminance distortion, and contrast distortion.

$$Q = \frac{\sigma_{gf}}{\sigma_g \sigma_f} \cdot \frac{2\bar{f}\bar{g}}{\bar{f}^2+\bar{g}^2} \cdot \frac{2\sigma_g \sigma_f}{{\sigma_g}^2+{\sigma_f}^2} , \quad (9)$$

where $\bar{f}$ and $\bar{g}$ are mean values of the original and processed images respectively with their standard deviations σ_f and σ_g. Also, σ_{gf} represents the covariance between the original and processed images. A better-quality processed image will have a higher Q value.

The structural similarity index (SSIN) between the original and processed images is a more general form of Q developed by the same research group. A higher value of SSIN indicates a higher quality processed image.

$$SSIN = \frac{(2\bar{f}\bar{g}+c_1)}{(\bar{f}^2+\bar{g}^2+c_1)} \cdot \frac{(2\sigma_{gf}+c_2)}{({\sigma_g}^2+{\sigma_f}^2+c_2)} , \quad (10)$$

where c_1 and c_2 are constants computed as 0.01 and 0.03 of maximum value in image dynamic range respectively (255 in the case of ultrasound images).

IV. Experimental Setup and Data Collection

The experimental ultrasound imaging data were acquired using an ultrasound imaging research system (Digison Digital Ultrasound Research system, Mashreq., Egypt) [38]. A customized research interface was used to set all imaging parameters and allowed the acquisition and storage of raw data samples at high sampling rates of 50 M Samples/s at a quantization of 16 bits. To ensure that the collected data set includes diverse image characteristics, imaging experiments were performed using several multifrequency ultrasound transducers including different linear and convex array probes including an endo-cavity probe. Furthermore, the experiments included different clinical applications on human volunteers and scanning of a tissue-mimicking quality assurance phantom (CIRS Inc., U.S.A.). A total of 10 images were collected for each experiment to do temporal averaging to remove present random noise. The complete data set included the results from 26 different imaging experiments with a total of 260 images. Fig. 2 demonstrates a sample image from each of the 26 imaging experiments to show the wide variation in their nature and characteristics. The conventional ultrasound signal processing stack was used to perform peak detection using Hilbert transformation then resampling to reduce the number of samples in each line to 512. The number of acquired lines for each imaging experiment was 128 and this 128×512 data array of resampled line data was called the stick data. The new technique was applied on the stick data to obtain the final processed image. The final image reconstruction from the processed stick data was subsequently performed to obtain the final processed images. The new methodology was implemented using Matlab 2022b (Mathworks Inc., U.S.A.). The data processing system was a desktop computer with 11th generation Intel® Core™ i7 processor and 32 GB of RAM running a 64-bit Windows 11 Home Edition operating system.

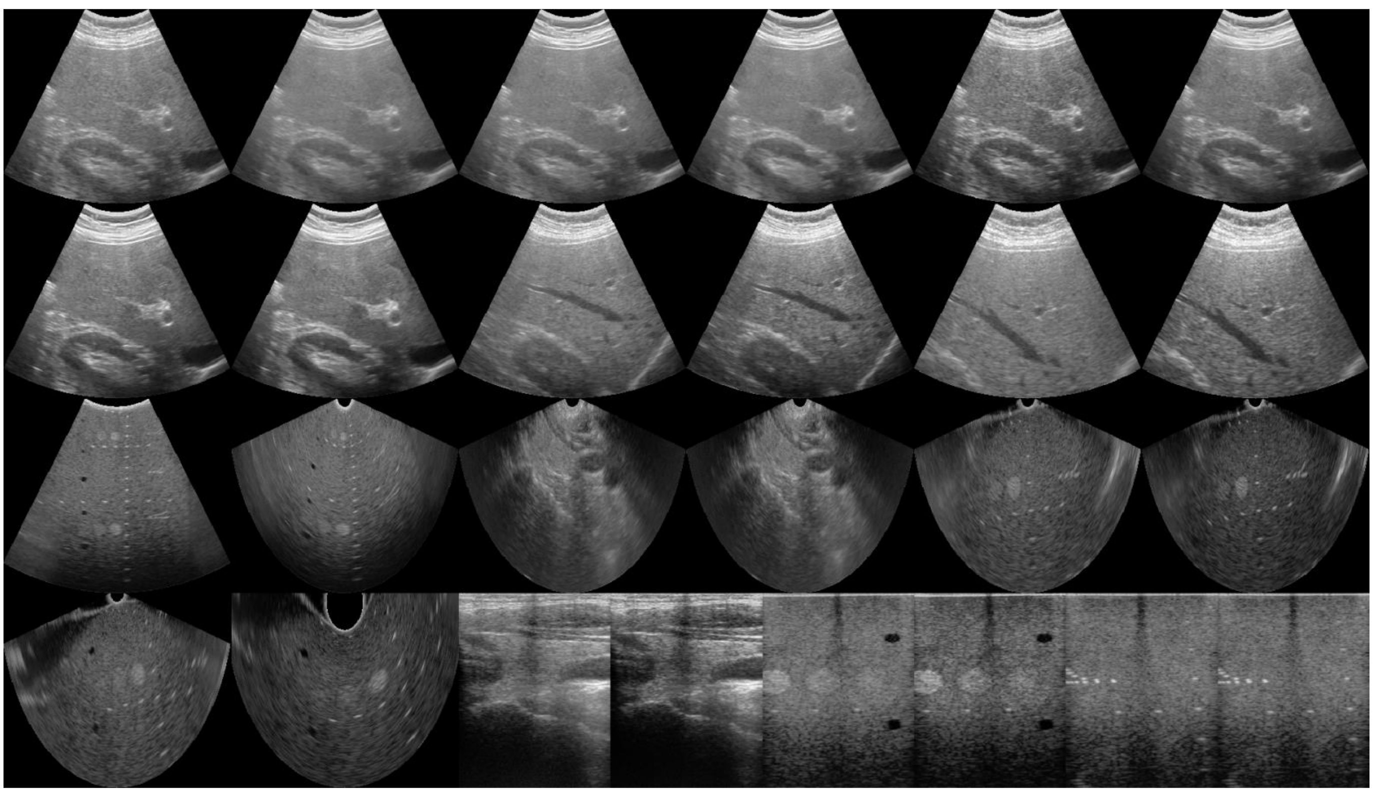


Fig. 2. Illustration of the different imaging experiments performed in this study.

V. Results and Discussion

Example qualitative results of applying the new methodology compared to the original images and the results from individual speckle reduction filters as well as their average are shown in Fig. 3. To better demonstrate the performance, the results from imaging experiments using three different probe geometries were included for both the tissue-mimicking phantom and real human scans. The qualitative evaluation can be done by observing the smoothness of speckle texture as well as the sharpness or definition of edges in each image as compared to the original. As can be observed, variations in both are found across the results from the individual speckle reduction filters where better smoothness comes at the cost of blurry edges such as with the Lee filter results. This can be seen more clearly in the results from phantom experiments given the structure of the phantom. In the results from all image experiments, the proposed PCA-based hybrid technique offers smoothness and edge definition that is comparable to the best of all individual techniques. Furthermore, the results from the simple average of individual techniques appear to partially inherit the quality issue such as blurring more prominently than the new technique.

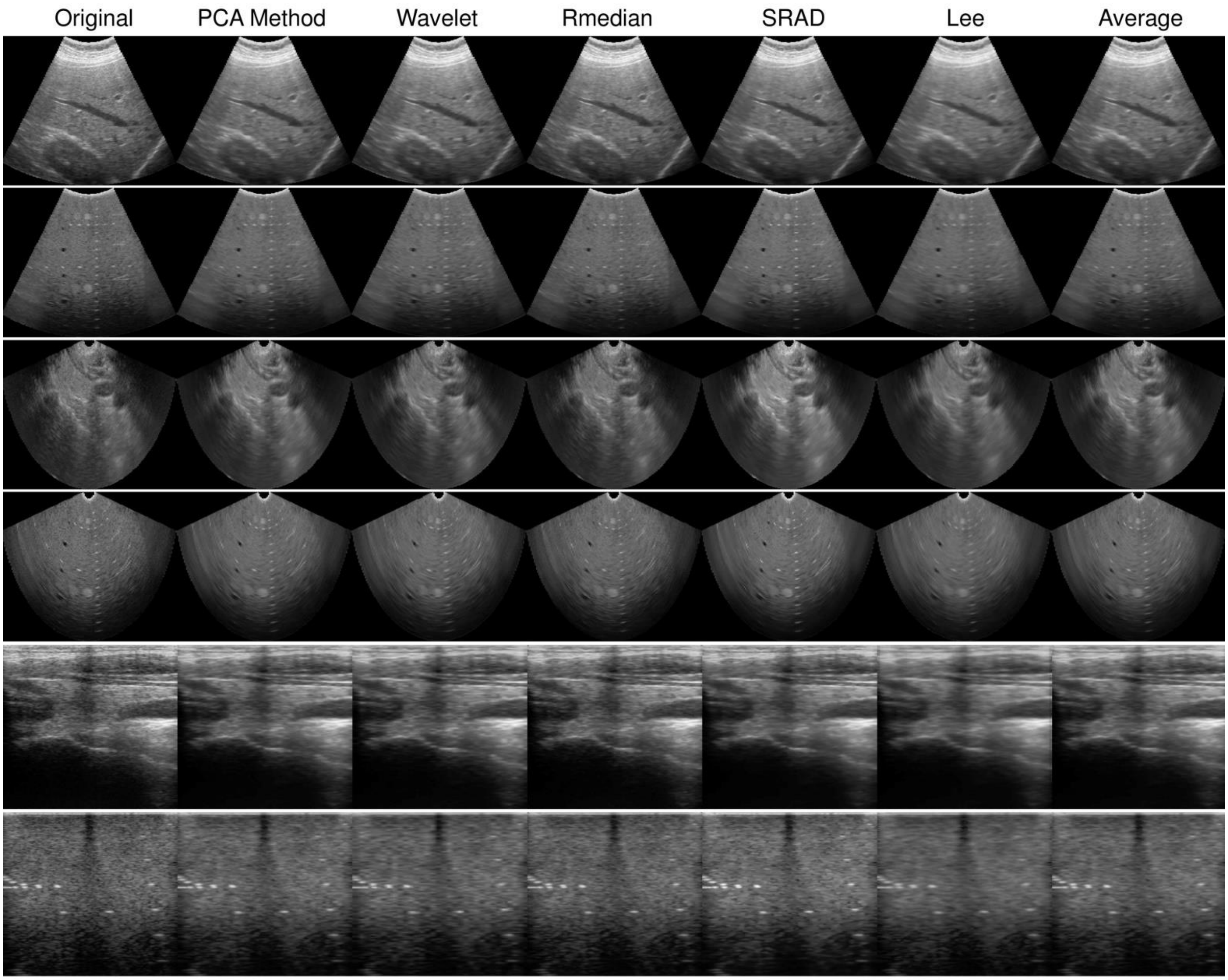


Fig. 3. Experimental results from the new technique are compared to those from the individual techniques used in addition to their average. the results are provided for different probe geometries and scan parameters from both a tissue-mimicking phantom and human volunteers.

The quantitative evaluations of the results are provided in Table I, which presents the percentage mean change in image quality metrics compared to the new PCA-based hybrid technique along with its standard deviation and p-value of statistical significance over the set of 26 experiments. For each metric, the percentage mean change of a particular technique is calculated as the average of the difference between the metric values of that particular technique and the new PCA-based hybrid technique, divided by the average of that metric for the new PCA-based hybrid technique. This formula was chosen to make the percentage change always relative to the same quantity across different techniques to make consistent comparisons. For better inspection of the results, the image quality metrics were grouped into either error or quality metrics as shown in the first column. When a technique has a positive percentage mean change in the error metrics, this means that its error metric is higher than that of the new technique, which would indicate lower performance. ON the other hand, a lower performance in the quality metrics is when a technique has a negative percentage mean change in the error metrics, which means that its quality metric is lower than that of the new technique. As can be observed, the results indicate that the new PCA-based hybrid technique offers better performance across all metrics with varying degrees. For example, in the universal quality index that deterioration varies from -2.77% for relaxed median filter to -30.55% for wavelet filter. It should be noted that the other metrics for these same techniques were very different where for example they were higher by 107.84% and 97.9% respectively for Laplacian mean squared error. This generally demonstrates the issue of variable performance of individual techniques across metrics where some perform better on quality and others on error metrics. One final observation is that the performance of the new PCA-based hybrid technique was consistently superior to that of the average of individual techniques, with significantly higher improvements in error metrics than quality metrics.

In order to ensure that the detected mean changes are statistically significant, a two-sample student t-test was applied and the p-values are listed in the same bin as the percentage mean change in Table I. A significance level of 0.05 was used and higher p-values where statistical significance cannot be confirmed at that level were denoted with "*" in the table. As can be observed, only three results were not statistically significant and they were all for the relaxed median filter. All results for the average of individual techniques were statistically significant indicating that there is a real performance boost. The overall results generally support the research hypothesis that the performance of the new PCA-based hybrid technique is better than or not different from the best individual technique.

TABLE I. PERCENTAGE MEAN CHANGE IN IMAGE QUALITY METRICS COMPARED TO THE NEW PCA-BASED HYBRID TECHNIQUE ALONG WITH ITS STANDARD DEVIATION AND P-VALUE OF STATISTICAL SIGNIFICANCE OVER THE SET OF 26 EXPERIMENTS

Assessment Metric		Wavelet	RMedian	SRAD	Lee	Average
Error Metrics	Root Mean Squared Error (RMSE)	+49.44% ± 14.66 (p<0.001)	+6.69% ± 12.69 (p=0.103*)	+93.25% ± 70.71 (p<0.001)	+46.78% ± 14.33 (p<0.001)	+23.84% ± 3.1 (p<0.001)
	Laplacian Mean Squared Error (LMSE)	+107.84% ± 14.73 (p<0.001)	+97.9% ± 18.9 (p<0.001)	+26.83% ± 17.21 (p<0.001)	+98.31% ± 14.86 (p<0.001)	+58.72% ± 4.51 (p<0.001)
	Minkowski Error Metric (β=3) (ErrorM3)	+49.36% ± 13.32 (p<0.001)	+29.18% ± 12.67 (p<0.001)	+76.49% ± 59.56 (p<0.001)	+49.99% ± 15.05 (p<0.001)	+24.34% ± 2.94 (p<0.001)
	Minkowski Error Metric (β=4) (ErrorM4)	+49.6% ± 12.81 (p<0.001)	+71.18% ± 18.17 (p<0.001)	+64.0% ± 52.31 (p<0.001)	+52.93% ± 15.67 (p<0.001)	+24.82% ± 2.88 (p<0.001)
Quality Metrics	Signal-to-Noise Ratio (SNR)	-12.95% ± 3.1 (p<0.001)	-2.04% ± 3.78 (p=0.471*)	-18.34% ± 10.3 (p<0.001)	-12.56% ± 3.17 (p<0.001)	-6.86% ± 0.67 (p<0.001)
	Peak Signal-to-Noise Ratio (PSNR)	-10.38% ± 3.2 (p<0.001)	-2.31% ± 4.17 (p=0.166*)	-14.44% ± 8.24 (p<0.001)	-11.52% ± 3.0 (p<0.001)	-6.14% ± 0.62 (p<0.001)
	Structural Similarity Index (SSIN)	-23.59% ± 6.34 (p<0.001)	-2.5% ± 4.16 (p=0.013)	-5.84% ± 3.86 (p<0.001)	-23.56% ± 3.75 (p<0.001)	-10.21% ± 1.5 (p<0.001)
	Universal Quality Index (Q)	-30.55% ± 5.95 (p<0.001)	-2.73% ± 5.11 (p=0.027)	-9.88% ± 7.1 (p<0.001)	-31.24% ± 3.4 (p<0.001)	-13.93% ± 1.95 (p<0.001)

* Not statistically significanct.

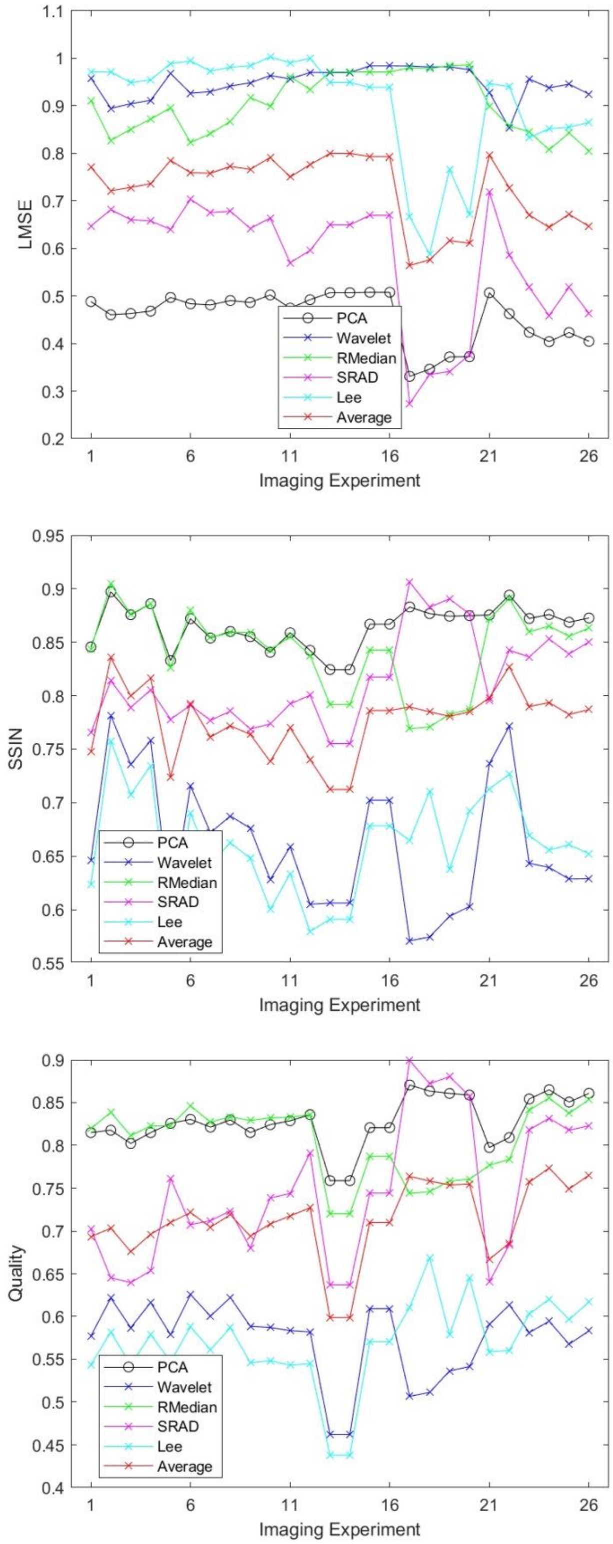


Fig. 4. Experimental results showing example quality metrics Laplacian Mean Squared Error (LMSE), structural similarity index (ssin), and universal quality index (Q) for 26 imaging experiments.

In order to illustrate the variability of metric values across different imaging experiments, Fig. 4 presents the values of Laplacian mean squared error, structural similarity index and universal quality metrics for each of the 26 imaging experiments. It can be observed that the different techniques have significant variability when applied to different experiments with different probe geometry and scan parameters. Furthermore, it is evident that the performance of the new PCA-based hybrid technique is better than or equal to the best individual technique in the majority of experiments, which again confirms the validity of the research hypothesis of developing this method. It can be also observed that the results from the average of individual techniques always lie in the middle of all individual techniques. This means that it may be considered as a way of obtaining consistent results from all techniques, but its performance will not be the best.

In order to verify the importance of including the original image in the observations matrix along with those from the speckle filters, the new technique was implemented with only 4 columns in the observations matrix D by excluding the original image. The quantitative evaluations of the results are presented in Table II. As can be observed, the obtained outcome has mixed performance across each metric as evident by the mixed negative and positive values in each. This indicates the results of excluding the original image are significantly worse than with the original image included. Furthermore, the results of excluding the original image appear to be very close to those of the average of individual techniques. This is evident from the small percentage mean differences and from the fact that there is no statistically significant difference between applying the new method with excluding the original image and that of the average of individual techniques. This indicates the important role of the information in the original image in guiding the estimation of weights used to combine it with the different individual techniques. To demonstrate this further, the computed weights for all experiments are shown in Fig. 5. As can be seen, the weights vary significantly across imaging experiments, which outlines the importance of their adaptive estimation. The weights for the original image were second highest in the majority of experiments indicating its importance in the outcome.

With the performance advantage of the PCA-based hybrid technique established by the results, the main challenge for its widespread application appears to be mainly computational. Rather than using an individual speckle filter, four filters need to be applied in the new method. The computations needed to perform the new technique can be divided into the computations needed for the speckle filters (which may vary with implementation), covariance matrix estimation, principal component analysis of a 5×5 matrix, the weighted average of 5 stick data images. Nevertheless, this computational cost can be considered reasonable for real-time performance especially with current high processing power available in modern digital ultrasound imaging systems. Furthermore, since we require only the first eigenvector corresponding to the maximum eigenvalue, significantly faster solvers can be utilized [37].

TABLE II. PERCENTAGE MEAN CHANGE IN IMAGE QUALITY METRICS COMPARED TO THE NEW PCA-BASED HYBRID TECHNIQUE WITHOUT INCLUDING ORIGINAL IMAGE IN OBSERVATIONS MATRIX ALONG WITH ITS STANDARD DEVIATION AND P-VALUE OF STATISTICAL SIGNIFICANCE OVER THE SET OF 26 EXPERIMENTS

Assessment Metrics		Wavelet	RMedian	SRAD	Lee	Average
Error Metrics	Root Mean Squared Error (RMSE)	+19.43% ± 10.97 (p<0.001)	-14.73% ± 10.44 (p<0.001)	+54.44% ± 55.67 (p=0.002)	+17.31% ± 11.7 (p<0.001)	-1.03% ± 1.91 (p=0.756*)
	Laplacian Mean Squared Error (LMSE)	+32.7% ± 13.07 (p<0.001)	+26.36% ± 15.07 (p<0.001)	-19.02% ± 7.84 (p<0.001)	+26.62% ± 7.11 (p<0.001)	+1.34% ± 1.37 (p=0.66*)
	Minkowski Error Metric (β=3) (ErrorM3)	+19.36% ± 9.73 (p<0.001)	+3.23% ± 10.38 (p=0.337*)	+41.04% ± 47.03 (p<0.001)	+19.86% ± 12.44 (p<0.001)	-0.64% ± 1.51 (p=0.837*)
	Minkowski Error Metric (β=4) (ErrorM4)	+19.54% ± 9.34 (p<0.001)	+36.78% ± 15.58 (p<0.001)	+31.05% ± 41.5 (p=0.001)	+22.2% ± 13.03 (p<0.001)	-0.26% ± 1.19 (p=0.932*)
Quality Metrics	Signal-to-Noise Ratio (SNR)	-6.26% ± 3.26 (p=0.035)	+5.48% ± 4.01 (p=0.075*)	-12.07% ± 11.2 (p=0.002)	-5.85% ± 3.29 (p=0.022)	+0.29% ± 0.57 (p=0.913*)
	Peak Signal-to-Noise Ratio (PSNR)	-4.35% ± 3.26 (p=0.011)	+4.27% ± 4.29 (p=0.018)	-8.69% ± 8.91 (p=0.001)	-5.56% ± 3.17 (p<0.001)	+0.17% ± 0.47 (p=0.899*)
	Structural Similarity Index (SSIN)	-15.31% ± 6.65 (p<0.001)	+8.06% ± 4.95 (p<0.001)	+4.36% ± 4.33 (p=0.002)	-15.28% ± 3.04 (p<0.001)	-0.48% ± 0.53 (p=0.673*)
	Universal Quality Index (Q)	-19.8% ± 8.09 (p<0.001)	+12.33% ± 7.14 (p<0.001)	+4.07% ± 6.54 (p=0.128*)	-20.6% ± 3.22 (p<0.001)	-0.6% ± 0.68 (p=0.737*)

* Not statistically significanct

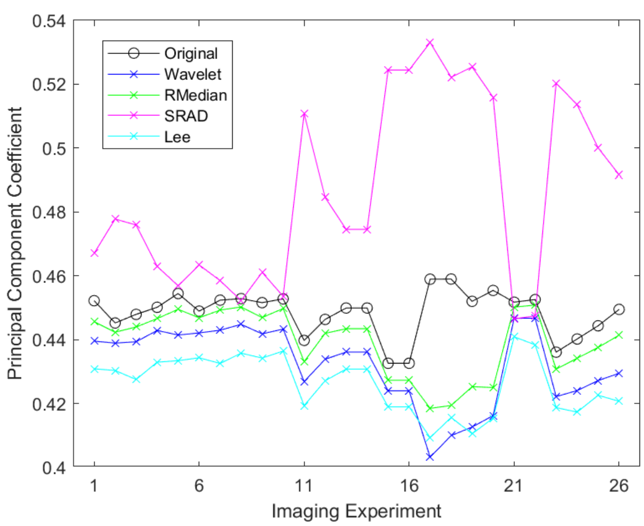


Fig. 5. Illustration of the variations of experimental optimal weights of different images obtained using PCA for the 26 imaging experiments.

VI. CONCLUSIONS

In this study, the combination of information from multiple speckle filters is considered and the use of principal component analysis is proposed to find the optimal set of weights that would retain the most information and hence would better represent the data in the final image. The new technique is implemented to process ultrasound images acquired from a research system and the outcomes are compared to the individual techniques and their average using quantitative image quality metrics. The results confirm that the new PCA-based hybrid technique offers consistent high performance across different experiments. The proposed technique has potential for utilization in clinical settings to provide consistently better-quality combined images that may help improve diagnostic accuracy.

ACKNOWLEDGMENT

This project was funded by the Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah, under Grant No. (CEIES-16-07-01). The authors, therefore, acknowledge the technical and financial support of King Abdulaziz University.

REFERENCES

[1] P. R. Hoskins, K. Martin, A. Thrush, Diagnostic Ultrasound: Physics and Equipment, 2nd ed., Cambridge University Press, 2010.

[2] E. Krupinski, H. Kundel, P. Judy, C. Nodine, “The medical image perception society, key issues for image perception research,” Radiology, vol. 209, pp. 611–612, 1998.

[3] C. B. Burckhardt, “Speckle in ultrasound B-mode scans,” IEEE Trans. Sonics Ultrasonics, vol. SU-25, no. 1, pp. 1–6, 1978.

[4] R. F. Wagner, S. W. Smith, J. M. Sandrik, H. Lopez, “Statistics of speckle in ultrasound B-scans,” IEEE Trans. Sonics Ultrasonics, vol. 30, pp. 156–163, 1983.

[5] A. Perperidis, D. Cusack, A. White, N. McDicken, T. MacGillivray, T. Anderson, “Temporal Compounding: A Novel Implementation and Its Impact on Quality and Diagnostic Value in Echocardiography,” Ultrasound in Medicine & Biology, vol. 41, no. 6, pp. 1749-1765, 2015.

[6] C. P. Loizou, C. S. Pattichis, Despeckle Filtering for Ultrasound Imaging and Video, Volume I: Algorithms and Software, 2nd ed., Morgan & Claypool, 2015.

[7] K. Z. Abdel-Monem, A. M. Youssef, Y. M. Kadah, “Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion,” IEEE Trans. Biomed Eng, vol. 49, no. 9, pp. 997-1014, Sept. 2002.

[8] J. S. Lee, “Digital image enhancement and noise filtering by using local statistics,” IEEE Trans. Pattern Anal. Mach. Intell., PAMI-2, no. 2, pp. 165–168, 1980.

[9] O. Rubel, V. Lukin, A. Rubel, K. Egiazarian, "Selection of lee filter window size based on despeckling efficiency prediction for sentinel SAR images," Remote Sensing, vol. 13, no. 10, p.1887, 2021.

[10] A. F. de Araujo, C. E. Constantinou, J. Tavares, "Smoothing of ultrasound images using a new selective average filter," Expert Systems with Applications, vol. 60, pp. 96-106, 2016.

[11] J. Saniie, T. Wang, N. Bilgutay, "Analysis of homomorphic processing for ultrasonic grain signal characterization," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 3, pp. 365–375, 1989.

[12] M. A. Gungor, I. Karagoz, "The homogeneity map method for speckle reduction in diagnostic ultrasound images," Measurement, vol. 68, pp. 100-110, 2015.

[13] A. B. Hamza, P. L. Luque-Escamilla, J. Martínez-Aroza, R. Román-Roldán, "Removing noise and preserving details with relaxed median filters," Journal of mathematical imaging and vision, vol. 11, no. 2, pp.161-177, 1999.

[14] K. Chauhan, R. K. Chauhan, A. Saini, "Enhancement and Despeckling of Echocardiographic Images," In Soft Computing Based Medical Image Analysis, Academic Press, pp. 61-79, 2018.

[15] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 7, pp. 629–639, July 1990.

[16] Y. Yongjian, S. T. Acton, "Speckle reducing anisotropic diffusion," IEEE Trans. Image Process., vol. 11, no. 11, pp. 1260–1270, November 2002.

[17] H. Choi, J. Jeong, "Speckle noise reduction for ultrasound images by using speckle reducing anisotropic diffusion and Bayes threshold," Journal of X-ray Science and Technology, vol. 27, no. 5, pp.885-898, 2019.

[18] R. G. Dantas, E. T. Costa, "Ultrasound speckle reduction using modified gabor filters," IEEE Trans Ultrason Ferroelec Freq Cont, vol. 54, no. 3, pp. 530-538, 2007.

[19] D. L. Donoho, "Denoising by soft thresholding," IEEE Trans. Inform. Theory, vol. 41, pp. 613–627, 1995.

[20] S. Gupta, R. C. Chauhan, S. C. Sexana, "Wavelet-based statistical approach for speckle reduction in medical ultrasound images," Med Biol Eng Comput, vol. 42, pp. 189–192, 2004

[21] A. K. Bedi, R. K. Sunkaria, "Ultrasound speckle reduction using adaptive wavelet thresholding," Multidimensional Systems and Signal Processing, vol. 33, no. 2, pp.275-300, 2022.

[22] J. Kang, J. Y. Lee, Y. Yoo, "A new feature-enhanced speckle reduction method based on multiscale analysis for ultrasound B-mode imaging," IEEE Trans Biomed Eng, vol. 63, no. 6, pp. 1178 – 1191, 2016.

[23] J. Zhang, G. Lin, L. Wu, C. Wang, Y. Cheng, "Wavelet and fast bilateral filter based de-speckling method for medical ultrasound images," Biomed Sig Proc Cont, vol. 18, pp. 1-10, 2015.

[24] Z. A. Mustafa, B. A. Abrahim, I. A. Yassine, N. Zayed, Y. M. Kadah, "Wavelet Domain Bilateral Filtering with Subband Mixing for Magnetic Resonance Image Enhancement," J Med Imag Health Inform, vol. 2, pp. 230-237, 2012.

[25] J. F. Al-Asad, A. M. Reza, U. Techavipoo, "An ultrasound image despeckling approach based on principle component analysis," International Journal of Image Processing (IJIP), vol. 8, no. 4, pp. 156-77, 2014.

[26] L. Zhang, W. Dong, D. Zhang, G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," Pattern recognition, vol. 43, no. 4, pp. 1531-49, 2010.

[27] H. Lv, S. Fu, C. Zhang, L. Zhai, "Speckle noise reduction of multi-frame optical coherence tomography data using multi-linear principal component analysis," Optics Express, vol. 26, no. 9, pp. 11804-18, 2018.

[28] R. R. Wildeboer, F. Sammali, R. J. Van Sloun, Y. Huang, P. Chen, M. Bruce, C. Rabotti, S. Shulepov, G. Salomon, B. C. Schoot, H. Wijkstra, "Blind source separation for clutter and noise suppression in ultrasound imaging: Review for different applications," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 8, pp. 1497-512, 2020.

[29] M. M. Rahman, P. M. Kumar, M. G. Arefin, M. S. Uddin, "Speckle noise reduction from ultrasound images using principal component analysis with bit plane slicing and nonlinear diffusion method," Proc. 15th International Conference on Computer and Information Technology (ICCIT), pp. 159-163, 2012.

[30] M. Jalali, H. Behnam, P. Gifani, Z. A. Sani, "Simultaneous speckle reduction of echocardiography frame sequence using PCA," Proc. 22nd Iranian Conference on Electrical Engineering (ICEE), pp. 1863-1867, 2014.

[31] I. Kumar, H. S. Bhadauria, J. Virmani, J. Rawat, "Reduction of speckle noise from medical images using principal component analysis image fusion," Proc. 9th International Conference on Industrial and Information Systems (ICIIS), pp. 1-6, 2014.

[32] A. M. Eskicioglu, P.S. Fisher, "Image quality measures and their performance," IEEE Trans. On Communications, vol. 43, no. 12, pp. 2959-2965, 1995.

[33] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error measurement to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, April 2004.

[34] Z. Wang and A. Bovik, "A universal quality index," IEEE Signal Process. Lett., vol. 9, no. 3, pp. 81–84, March 2002.

[35] D. Sakrison, "On the role of observer and a distortion measure in image transmission," IEEE Trans. Comm., vol. 25, pp. 1251–1267, November 1977.

[36] R. Bro, A. K. Smilde, "Principal component analysis," Analytical Methods, vol. 6, no. 9, pp. 2812-31, 2014.

[37] G. H. Golub, C. F. Van Loan, Matrix computations, 3rd ed., Johns Hopkins University Press, 2013.

[38] Y. M. Kadah, A. F. Elnokrashy, U. M. Alsaggaf, A. M. Youssef, "Speckle reduction in medical ultrasound imaging based on visual perception model," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 11, pp. 575–581, 2022.

Application of Artificial Intelligence-Genetic Algorithms to Select Stock Portfolios in the Asian Markets

Luu Thu Quang
Ho Chi Minh University of Banking, Vietnam

***Abstract*—The paper's main goal is to use a genetic algorithm to find the best stock portfolio that meets the criteria of high return and low risk, allowing investors to adjust the appropriate proportion for each share. Using the Python programming language based on the Jupyter Notebook engine, this paper introduces a model of six stock portfolios, each of 30 stocks selected with market capitalization and high liquidity criteria of six markets in the Asian region. The results show that the four portfolios created from the markets of Vietnam, Thailand, Philippines, and Singapore meet both the return and risk objectives. The Malaysian market only meets the risk target, but the portfolio's return is not close to the expected ratio. Meanwhile, the Indonesian market outperformed expectations in terms of profits, but high profits come with high risks, so this market carries a concerning level of risk when compared to the profit and loss of other markets. The suggested stock allocation levels for each portfolio are based on the above results. Finally, the author proposes several policy implications related to the management and operation of the market to limit unnecessary price fluctuations of the stock and affect the business model of companies.**

Keywords—Artificial intelligence; genetic algorithms; optimal portfolio; Sharpe ratio; Asian stock market

I. Introduction

When learning about investment, optimizing the return on a portfolio is one of the most challenging issues in the financial sector. Since it is a process that is not simply a combination of individual stock codes with unexamined investment ratios, investors need to find ways to maximize the profitable investment and limit risk by understanding the market and setting themselves a desired level of return with an acceptable risk level. Dealing with this problem, recently, several artificial intelligence-based methods such as genetic algorithms have been applied to overcome this problem. Genetic algorithms (GAs) are optimal methods. They are widely used in many different fields such as medicine, engineering, electronics, not only the field of financial investment analysis to optimize value and solve problems through the fundamental processes of selection, breeding, and mutation between individuals in the population to create better latter-generation individuals, improving the adaptability of that individual in a more natural environment.

There have been many studies applying genetic algorithms to investments. However, the scope of the above studies is in the overseas market, where the market characteristics are quite different from the market in Southeast Asia. In Southeast Asia, studies of this method of migration algorithms have been applied and studied in many other fields such as health, biology, engineering, construction, information technology to solve different problems. The application of this method to the financial sector is still relatively small. Therefore, implementing this research in the Asian market is necessary and feasible, especially in the growing and constantly changing stock market.

An optimal portfolio is an asset allocation, or in other words, a group of different stock choices that provide the highest rate of return per unit of risk an investor is willing to accept. In addition, the optimal portfolio is considered to be the one with the maximum Sharpe ratio - which measures the amount of return generated per unit of risk when investing in an asset or type of other investments. According to Markowitz's methodology, the linear equations are used to determine the best portfolio, but this traditional method works best when the number of stocks and the size of the search space are both small. When there are a lot of stocks and large searchable space, special techniques, such as GAs, must be used. Compared to other optimization methods, GAs work by simulating evolution and natural selection on a computer starting with a random population. Additionally, we require an adaptive function to choose good individuals and reject bad ones in order to optimize. Genetic algorithms, in contrast to other optimization techniques, have the following advantages:

- GAs work with the variable's code, not directly on the variable.
- Most of the optimization techniques normally search from a single vertex, while GAs always operate on a set of vertices (optimal points), this is an advantage of GAs that increases the chances of reaching the comprehensive optimal, and avoids premature convergence at the local point.
- GAs evaluate the objective function to serve the search process, so it can be applied to any optimization problem (continuous or discrete).
- GA is a subclass of probability algorithms, and their fundamental operations are based on their capacity to incorporate randomness into the processing flow.

The paper will help answer the question: "Should genetic math be applied to the creation of an effective stock portfolio for southeast Asian markets?" To answer that question, the author will study, learn and apply the processes of acquisition to a portfolio of asset classes with a wide range of asset classes with a level of assets risks and ratios that vary to select and create a potential stock portfolio with high profitability criteria with the right level of risk.

II. Theoretical Basis

Yang [1] compares three approaches from genetic algorithms, medium variances, and Bayes methods to improve portfolio efficiency with a sample of MSCI (Morgan Stanley Capital International) datasets of Total Return on Equity of Canada, France, Germany, Japan, the United Kingdom, and the United States include European stocks, bonds, and bonds based on the value of the U.S. dollar between January 1975 and December 2004, taking the risk-free rate of return as the T-bill interest rate of three months. The problem is that the portfolio has a minor standard deviation, and out-of-sample returns also have a significantly higher average value than the standard medium variance portfolio strategy and the Bayes approach. Portfolio weighting using genetic algorithms is less unbalanced and fluctuates much less over time than the standard average method and Bayes method. Furthermore, the author has stated that the algorithm is a model that combines historical information and future uncertainty to significantly improve the accuracy of the average estimate, helping to improve the model's performance by overcoming computational difficulties when approached by the Bayes method

Lin [2] also reports that adopting genetic algorithms would provide optimal value for creating a portfolio. The author asserted that the genetic algorithm could solve the corresponding optimization problems because existing traditional methods cannot be effectively addressed. Moreover, the study based on a team of 40 sample companies collected from the Taiwan market was used to demonstrate the proposed method with the probability of hybridization and random mutations.

Sefiane [3] applied genetic algorithms to optimize a portfolio of five stocks with data including portfolio-wide average returns and portfolio variances for five years from 2007 to 2011. By running the data on Matlab software, the author has produced the result of five proportions of investment in five stock codes and selected a hybrid by one point, two points through the stage of three hybrids. At the end of the research process, the author commented that this problem achieves exceptionally effective results by shortening the calculation time to get an investment solution for the whole portfolio. Dubinskas [4] also evaluate the genetic algorithm method for data of 18 stock codes made from Lithuanian businesses listed on the official list of NASDAQ OMX Baltic companies in 2013 planned portfolio to be built including four businesses representing different sectors. Research indicates that optimizing the portfolio by adopting an algorithmic approach generates four times more returns than portfolios built using a defined or random programming method. Along with high returns, the systematic risk assessment of the portfolio analyzed by genetic algorithms is higher than the portfolio optimized by defined and random programming. However, the authors commented that the portfolio's performance optimized by algorithmic genetics might exceed market performance. Moreover, the author also concludes that a research paper needs to examine several cycles to look for more reliable conclusions, emphasizing the application of genetic algorithms methods during a market downturn.

In addition, Cheong et al. [5] also implement genetic algorithms to support portfolio optimization based on investor information, which is to build portfolios by selecting institutionalized stocks or foreign investors invest more than other stocks. Through data obtained from the Korea Securities Computer Corporation (KOSCOM), specifically the top 90 Kospi 200 companies in terms of market capitalization from September 1, 2007, to May 30, 2014, and through the analysis process, the author commented that long-term information outperforms medium- and short-term information on the Korean stock market. The weight varies when using genetic algorithms how it affects portfolio performance. In particular, the study results add to the behaviors and patterns of investors that can be used to devise investment strategies. Investors build portfolios according to their risk-return level on the average variance framework, and relying on the portfolio strategies of other confident investors can enhance portfolio performance. However, the author also notes that institutional investors anticipate higher returns and lower risks on the Korean stock market, while foreign investors only consider portfolio risk.

Liu et al. [6] shows that genetic algorithms produce results that provide a more reliable portfolio in the same risk and without the need to introduce or mention each country's data. Liu et al [6] argues about the time series length of each stock on the construction of the portfolio and improve the algorithmic efficiency in fundamental stock analysis. With a fairly comprehensive sample of data, including the closing price of 2317 stocks on the Chinese market from February 20, 2016, to February 16, 2017, there are 250 trading days. The author agrees with the researchers above that genetic algorithms have optimized portfolio strategies. Sharpe ratios are relatively high, markedly improved along with an increase in the number of iterations, meaning that the ability to achieve excess returns increases gradually [7]. Furthermore, the author also compared two other methods, the random method and the K-mean algorithm. Their results indicate that with the increase in risk, the rate of profitability when using genetic algorithms increases, indicating the effectiveness of genetic algorithms.

In general, after consulting and researching research articles, the application of genetic algorithms to selectively and creating an optimal portfolio is agreed and encouraged by many authors because the effectiveness of the algorithm is persuasive when going through the analysis process to create an expected profit level. The standard deviation and deviation are in line with the expected level. Therefore, the expansion direction of future studies is to be able to conduct on a broader scale, a long time to test the appropriate level of an optimal category with the criteria of limiting risk, achieving the expected or highest level of profitability.

III. Research Method

A. Data

The data used in this study includes trading histories such as stock codes, trading months and closing prices provided at the DataStream system. The subjects selected in the study data are stocks listed on the stock exchanges of Vietnam, Thailand, Indonesia, Philippines, Malaysia, and Singapore with the following characteristics:

- Shares listed at least two years before the time of inclusion in the portfolio
- The stock must be traded during the study sample selection period

The data includes 180 stock codes, of which in Vietnam there are 30 company codes in VN30 and 30 companies with the largest capitalization in 5 countries, the month of trading started oil from 07/2018 to 04/2021

B. Method

Genetic mathematics is a technique of computer science that predicts the behavior and reactions of humans and organisms. Migration algorithms can be applied in the financial sector by forecasting profits, stock prices, portfolio optimization, and determining trading rules and pricing options. Genetic mathematics includes transformation modalities, including selection, hybridization, and mutation. Identifying individuals in the current generation retained in the next generation is called the selection process. The random combination of two different individuals, using the things available in the previous generation passed on to the next generation, is hybridization. Randomly changing the structure of the individual, and transforming into new individuals is called the process of mutation [8].

The author uses a programming language called Python, which runs on the Jupyter Notebook tool to build a research process, a computer-language method, also known as computer-language code used to replace a string of characters or statements briefly. These codes are technically valuable in terms of information technology. The change in this code does not affect the values and results involved, nor does it explain any economic significance. The code used in the article is inherited from many different authors whose research papers agree with this research paper. The code is only intended to support the function and give the study results. The author has also changed the code to match the author's research direction. The study focuses on variables in profitability, risk, and allocation levels accordingly and delivers the most optimal portfolio outcome. The research process will be as follows:

The prerequisite for implementing the algorithm, the libraries are inserted into python language to help support the calculation of significant data sources such as numpy, pandas, functools. Code:

```
import numpy as np
import pandas as pd
from functools import reduce
```

Step 1, import the data of a country of 30 companies, specifically code as follows. The author takes the example of how Vietnam runs. Code:

```
files=['VIETNAM.csv']
dfs=[]
for file in files:
temp=pd.read_csv(file)
dfs.append(temp)
stocks = reduce(lambda left,right: pd.merge(left,right,on='Date'), dfs)
print(stocks.shape)
stocks.head()
```

In the Step 2, the profit is measured through the historical transaction values. The closing price equal to the average of 1 to 33 months, is takes the timeline of 04/2021 as the standard of calculating profitability. The formula used is the profit margin stated by the author in sector 2, section 2.1.3.2, specifically the total return of the stock plus the dividend paid and divided by the original price of the stock. Here, we assume the percentage of dividends paid is not in companies. Code:

```
def hist_return(months):
 idx=[]
 df=pd. DataFrame()
 for mon in months:35
 temp=(stocks.iloc[0,1:] - stocks.iloc[mon,1:])/(stocks.iloc[mon,1:])
 idx.append(str(mon)+'_mon_return')
 df=pd.concat([df, temp.to_frame(). T], ignore_index=True)
 df.index=idx
 return df
hist_stock_returns=hist_return([3,6,12,24,33])
hist_stock_returns
```

In Step 3, identify the gene or, in other words, encode the problem. In this step, the author will use a binary sequence of n-gene lengths (with n being the number of securities included in the category) as chromosomes to represent a category with n securities to study. Each gene represents the appearance or absence of a stock code, which these genes are regulated in alphabetical order. Code:

```
gene = np.random.rand()
Gene import time
def gen_mc_grid(rows, cols, n, N):
 np.random.seed(seed=int(time.time()))
 layouts = np.zeros((n, rows * cols), dtype=np.int32)
 # layouts_cr = np.zeros((n*, 2), dtype=np.float32)
 positionX = np.random.randint(0, cols, size=(N* n* 2))
 positionY = np.random.randint(0, rows, size=(N*n*2))
 ind_rows = 0
 ind_pos = 0
 while ind_rows < n:
 layouts[ind_rows, positionX[ind_pos] + positionY[ind_pos]*cols] =1
 if np.sum(layouts[ind_rows, :]) == N:
 ind_rows += 1
 ind_pos += 1
 if ind_pos >= N*n*2:
 print("Not enough positions")
 Break return layouts
def gen_mc_grid_with_NA_loc(rows, cols, n, N,NA_loc):
 np.random.seed(seed=int(time.time()))
 layouts = np.zeros((n, rows * cols), dtype=np.int32)
 layouts_NA= np.zeros((n, rows * cols), dtype=np.int32)
 for i in NA_loc:
 layouts_NA[:,i-1]=2
 positionX = np.random.randint(0, cols, size=(N* n* 2))
 positionY = np.random.randint(0, rows, size=(N*n*2))
 ind_rows = 0
 ind_pos = 0
```

```
 N_count=0
while ind_rows < n:
cur_state=layouts_NA[ind_rows,positionX[ind_pos]+positionY[ind_pos]   *
cols]
if cur_state!=1 and cur_state!=2:
layouts[ind_rows, positionX[ind_pos] + positionY[ind_pos] * cols]=1
layouts_NA[ind_rows, positionX[ind_pos] + positionY[ind_pos] * cols] = 1
 N_count+=1
 if np.sum(layouts[ind_rows, :]) == N:
 ind_rows += 1
 N_count=0
 ind_pos += 1
 if ind_pos >= N*n*2:
 print("Not enough positions")
 Break return layouts,layouts_NA
gen_mc_grid(9, 9, 180, 90)
gen_mc_grid_with_NA_loc(9, 9, 180, 90.range(10))
```

Step 4, gather the number of shares in the portfolio and randomly assign the capital allocation ratio, small parts of the total capital specified for each share, to set the investment allocation weighting. A stock portfolio will include fractional or decimal values of all stocks so that the sum of the values is not greater than 100%. Simply put, our portfolio has 30 stocks; the author will randomly create 30 variables corresponding to the proportion of investment in 30 stocks with a total capital of 1 unit. Code:

```
def chromosome(n):
 ch = np.random.rand(n)
 return ch/sum(ch)
child=chromosome(30)
print(child,sum(child))
```

The Step 5 is the first step in the genetic algorithm to initiate populations. The process is quite simple by creating a population of chromosomes, in which each chromosome is a long binary vector n-gene randomly generated. We need to calculate the probability of the securities number that allow research in the portfolio. Having a portfolio with many securities will not be highly effective in terms of objectives. That probability is measured as follows:

$$P = n/(pop\_size)$$

in which n is the number of securities included in the resulting portfolio, pop_size is the number of securities of the whole population. At that time, the initialization process was roughly described as in each gene of chromosomes in the population that generated a random number of realities in the passage [0, 1]. When $r < p$, the stock code appears in the category, i.e., the value of that gene is equal to 1. On the contrary, the gene is zero that stock code will not appear in the portfolio. Code:

```
n=30
pop_size=180
population = np.array([chromosome(n) for _ in range(pop_size)])
print(population.shape)
print(population)
```

Step 6, the author will calculate the target function stated in sector 2 is the Sharpe ratio, taking this ratio as a measure of the performance of the entire portfolio, in the calculation formula, this ratio depends on the variable is a risk-free interest rate, the author will default this ratio to 2%.

$$eval(v_i) = \text{Sharpe ratio} = (E(R_{port}) - r_f)/\delta_p$$

To be able to calculate the Sharpe ratio, calculate the following small components: Calculate the average profit, standard deviation and variance of past stock returns. Code:

```
# First of all, entering and re-universalizing the profit rate calculated in step 2
print(hist_stock_returns.info())
cols=hist_stock_returns.columns
hist_stock_returns[cols]=hist_stock_returns[cols].apply(pd.to_numeric,errors
='coerce')
print(hist_stock_returns.info())
# Calculate the variance of the historical profit, so that it can be easy to
calculate, the author will put the variance of the stock itself by zero
cov_hist_return=hist_stock_returns.cov()
print(cov_hist_return)
for i in range(30):
 cov_hist_return.iloc[i][i]=0
cov_hist_return
# Calculating average historical returns
mean_hist_return=hist_stock_returns.mean()
mean_hist_return
# Calculate the standard deviation of historical profits
sd_hist_return=hist_stock_returns.std()
sd_hist_return
# Calculate the expected return rate of the category
def mean_portfolio_return(child):
 return np.sum(np.multiply(child,mean_hist_return))
mean_portfolio_return(population[0])
# Calculate the variance of the category
def var_portfolio_return(child):
 part_1 = np.sum(np.multiply(child,sd_hist_return)**2)
 temp_lst=[]
 for i in range(30):
 for j in range(30):
 temp=cov_hist_return.iloc[i][j] * child[i] * child[j] * child[j]
 temp_lst.append(temp)
 part_2=np.sum(temp_lst)
 return part_1+part_2
var_portfolio_return(population[0])
# Calculate the Sharpe ratio of the category
def fitness_fuction(child):
 return(mean_portfolio_return(child)-rf)/np.sqrt(var_portfolio_return(child))
fitness_fuction(population[31]
```

The step 7 is a pretty important step in the genetic algorithm. The author will filter the stocks with the highest Sharpe ratio calculated in Step 6. In this step the author defaulted to a selective probability of 0.3, and the author selected the first three stocks in the category to consider the target function of the research paper. Code:

```
def Select_elite_population(population, frac=0.3):
 population=sorted(population,key=lambda x:
fitness_fuction(x),reverse=True)
 percentage_elite_idx = int(np.floor(len(population)* frac))
 return population[:p ercentage_elite_idx]
print(wool(Select_elite_population(population, frac=0.3)))
Select_elite_population(population, frac=0.3)
[fitness_fuction(x) for x in population] [:3]
```

Step 8 is mutating the individuals in the category. It can be said that this is a function that improves the individual more and more healthily than the first life. Especially in this process, the author will randomly select two stocks in the list of 30 stocks to conduct the transformation, and the following computer code will perform the following mutations. Code:

```
def mutation(parent):
 child=parent.copy()
 n=np.random.choice(range(30),2)
 while (n[0]==n[1]):
 n=np.random.choice(range(30),2)
```

```
 child[n[0]],child[n[1]]=child[n[1]],child[n[0]]
 return child
mutation(population[1]),population[1]
```

Step 9, at this step the stock codes will be intersected through real-value investment proportions. In the hybrid task, the author uses the arithmetic hybrid method to bring the most optimal effect. Code:

```
def Arithmetic_crossover(parent1,parent2):
 alpha = np.random.rand()
 child1 = alpha * parent1 + (1-alpha) * parent2
 child2 = (1-alpha) * parent1 + alpha * parent2
 return child1,child2
Arithmetic_crossover(population[2],population[3])
```

Step 10 is the step that helps create a generation of stock codes that come with a more potential weighting perform mutation processes or cross-exchange based on probability. The author defaulted to the probability of mutation as 0.4 and hybridization as 0.6. In the final stages, the probability of mutation will be reduced to 0.1, and hybridization will increase to 0.9. The sum of these two probabilities is always equal to 1, and it should be noted that the probability of hybridization always needs to be greater than the probability of mutation because mutations need to be limited, not always mutations to individuals with good character and adaptability to the habitat or in other words to survive in the category. The input of this process will be the sum of the stocks with the appropriate proportion of investment allocation. The output will be the next generation with a more suitable gene level. Code:

```
def next_generation(pop_size,elite,crossover=Arithmetic_crossover):
 new_population=[]
 elite_range=range(wool(elite))
 while wool(new_population) < pop_size:
 if wool(new_population) > 2*pop_size/3:
 mutate_or_crossover = np.random.choice([0, 1], p=[0.9, 0.1])
 else:
 mutate_or_crossover = np.random.choice([0, 1], p=[0.4, 0.6])
 if mutate_or_crossover:
 indx=np.random.choice(elite_range)
 new_population.append(mutation(elite[indx]))
 else:
 p1_idx,p2_idx=np.random.choice(elite_range,2)
 c1,c2=crossover(elite[p1_idx],elite[p2_idx])
 chk=0
 for gene in range(30):
 if c1[gene]<0:
 chk+=1
 else:
 chk+=0
 if sum(range(chk),0)>0:
 p1_idx,p2_idx=np.random.choice(elite_range,2)
 c1,c2=crossover(elite[p1_idx],elite[p2_idx])
 new_population.extend([c1,c2])
 return new_population
elite=Select_elite_population(population)
next_generation(180,elite)[:3]
elite=Select_elite_population(population)
next_generation(180,elite,Arithmetic_crossover)[:3]
```

Step 11 is a repeat of the entire process until the stocks in the portfolio do not change in terms of maximum profit and minimum risk or the number of fixed iterations selected by the author is 40 iterations. The author ordered that the expected return of the category would be greater than 20% and that the risk through variance calculations must be less than 5%. The loop will stop at the 40th round if that condition is met. If it does not meet, the study will perform through many subsequent loops so that 43 searches for the appropriate proportion for the 43. Stock codes suitable for the portfolio's great purpose achieve maximum profitability with a reasonably limited level of risk. Code:

```
n=30
pop_size=180
population = np.array([chromosome(n) for _ in range(pop_size)])
elite = Select_elite_population(population)
iteration=0
Expected_returns=0
Expected_risk=1
while (Expected_returns < 0.2 and Expected_risk > 0.05) or iteration <= 40:
 print('Iteration:',iteration)
 population = next_generation(180,elite)
 elite = Select_elite_population(population)
 Expected_returns=mean_portfolio_return(elite[0])
 Expected_risk=var_portfolio_return(elite[0])
 Print('Expected returns of {} with risk of {}\n'.format(Expected_returns,
Expected_risk)) )
 iteration+=1
print('Portfolio of stocks after all the iterations:\n')
[print(hist_stock_returns.columns[i],':',elite[0][i]) for i in list(range(30))]
# The ratios will in turn give results combined with the profit corresponding to
each stock code included in the portfolio
print('Portfolio of stocks after all the iterations:\n')
[print(hist_stock_returns.columns[i],':',elite[0][i]) for i in list(range(30))]
print('\nExpected returns of {} with risk of
{}\n'.format(Expected_returns, Expected_risk))
fitness_fuction(elite[29])
```

After this process, result will generate the weight distribution of the stocks included in the portfolio and the expected return and risk level of the entire category most suitable after undergoing the mutation and hybrid loops.

IV. Results

The study used data from six different countries in Southeast Asia to create six portfolios with varying rates of return and risk. Our findings, as seen in Fig. 1, show that Vietnam, with the top VN30 stock portfolio, has produced the highest expected profit result among the six countries with an incredible rate of approximately three times the expected profit margin of 60.23%. Followed by Indonesia reaching 53.13% and the third place is Thailand with a profit rate exceeding the target of 42.76%. However, accompanied by high profitability is a risk that is also a worrying risk when bringing the highest profitability among the six countries considered in Southeast Asia. Vietnam ranked 2nd in terms of portfolio risk at 4.6%, however, significantly the risk of the portfolio of 30 companies in the Indonesian market accounted for 6.04%, reaching the highest risk level of 4.04% reaching the highest risk level. However, these findings failed to meet the author's initial risk expectations. So it can be commented that when comparing the top 30 companies in the stock market with the stock codes reaching the largest capitalization in each country, Vietnam's portfolio gives positive results and is worth investing more than the Indonesian portfolio with higher profits and lower risks. However, does the Vietnamese category produce the best results? The saying "The higher the profit, the greater the risk" is an immovable rule when it comes to investing; everyone has the price of that action, and paying the desired price for a high profit, the higher the risk.

When it comes to investing, any investor wishes to own a portfolio with stock codes that not only bring high returns, but risk must also be an acceptable variable. When it comes to risk, the study results indicate that all five categories meet the standard deviation expectations that the category brings except for one category that does not meet the expected risk requirements of 30 companies in the Indonesian market mentioned above. However, when it comes to profitability, there are only five categories that bring positive results and exceed the expectations of the author: Vietnam, Thailand, Indonesia, Philippines, and Singapore; Only the Malaysian portfolio reached 19.46% of the profit margin, but in return for the risk that the portfolio overcomes only accounts for nearly 1%, an extremely difficult to figure, the fact that the figure of 19.46% is still an adequate number and if we convert all the capital distributed into stock codes less than 1% into the Stocks with higher allocations can still raise the value of the profitability of the whole portfolio but also mean that the risk will increase.

In general, only four over six markets met both in terms of profitability and risk expectations of the author: Vietnam, Thailand, Philippines, and Singapore. On the other hand, when the author sets an expectation of a return of 20%, regardless of the risk of the whole category, five categories meet the requirements, in order of declining profits: Vietnam, Indonesia, Thailand, the Philippines and Singapore. When the author considers only the expected risk of the portfolio of 5%, which means setting the risk level more important than the expected profit level, five categories meet the requirements presented in incremental risk: Malaysia, Singapore, Thailand, Philippines and Vietnam. Therefore, it is difficult to expect a portfolio with high returns with a lower level of risk. Our findings are consistent with the risk and return trade-off.

To be able to provide detailed statistics of six categories in six Southeast Asian markets, the author uses the Sharpe ratio after the optimal portfolio allocation results have produced the same results with the same data on closing prices with the timeline from 07/2018 to 04/2021 of the 30 companies with the largest market capitalization in each country has produced a result. Sharpe's ratio declined, led by Malaysia with 19.81, Singapore with 15.07, Vietnam with 12.67, Thailand with 9.33, Indonesia with 8.46 and Philippines with 6.9 (see Fig. 2). The Sharpe ratio is a ratio that represents the premium value of the portfolio that an investor will receive per unit of risk. The smaller this ratio, the smaller the risk premium, and vice versa. The larger the ratio, the greater the risk premium.

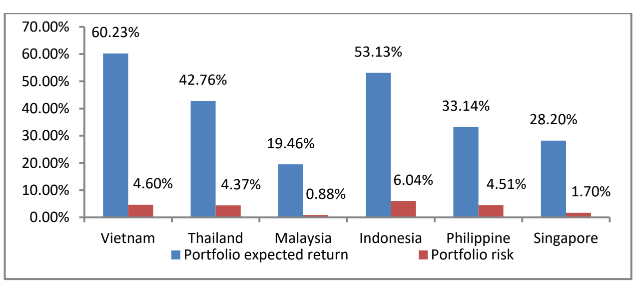


Fig. 1. Portfolio expected return and portfolio risk of Asian market

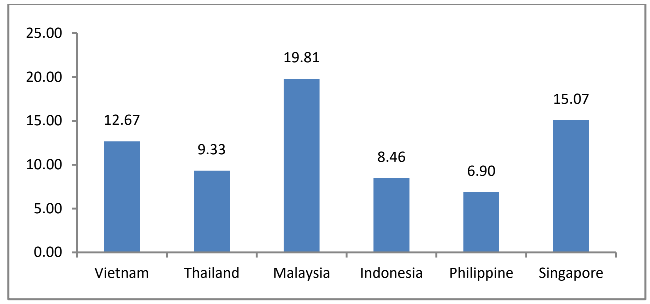


Fig. 2. Sharpe ratio of Asian market

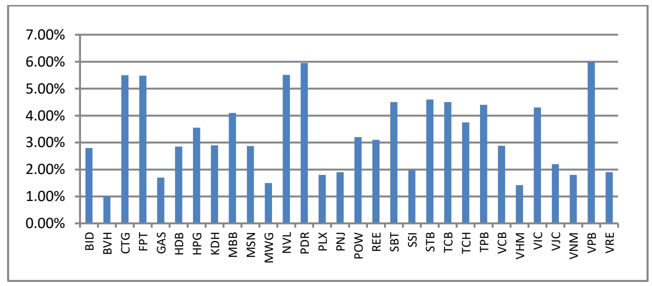


Fig. 3. Optimal portfolio of Vietnam market

The results showed that although Vietnam has the highest profitability and Malaysia offers relatively low profits investing in the Malaysian market still preserves much more profitability than other highly profitable categories. On the other hand, the Philippines market, as analyzed in the above subsection on the level of profitability and risk that the market brings, with the profitability brought although higher than 20% of the expected profit rate of the author. However, the risk of this market is very high, reaching 4.51%, overwhelmed the risk in the Thai market (4.37%). Therefore, when calculating the Sharpe ratio, the Philippines is a market where the adjusted profit above the risk level is much less competitive than the Thai market.

For Vietnamese stocks, After 40 iterations, the model has found an optimal solution for profitability and risk limitation, resulting in the allocation of investment capital into stocks in the VN30 basket. Surprisingly, in the stocks with the historical data levels used as research, there were five stocks with an allocation of over 5% respectively: VPB, PDR, NVL, CTG and FPT with weights of 6.01%, 5.95%, 5.51%, 5.50% and 5.48% respectively. The top five stocks in the allocation table belong to the real estate, technology and banking sectors. However, there are three stocks whose portfolios give relatively low allocation results: BVH, VHM and MWG, 1%, 1.42% and 1.50%, respectively (refer Fig. 3).

For Thai stocks, as can be seen from Fig. 4, the allocation of over 5%, four stocks invested quite a lot in the category: COM7, KTC, CPF and DELTA, respectively. The allocation levels are 7.03%, 6.93%, 6.29% and 6.2%, respectively. In particular, according to the research, we can see that com7 stock code is a company based in Thailand with the business of operating information technology products. This is a company in the form of distributing branded technology products. KTC is a company that serves credit card services and provides personal or business loans. CPF is an agricultural and food corporation specializing in producing the world's largest feed and shrimp. DELTA is an exporting company and a manufacturer of electronics and a flexible power supply. However, in the portfolio, a stock with a fairly low allocation of less than 1% is BBL. This is an operating bank headquartered in Bangkok, Thailand. In general, capital flows are evenly distributed, and no assets are excluded from the stock. This is the leading group of stocks in terms of market capitalization and high liquidity, so the proportion of allocation is evenly arranged is also an inevitable factor.

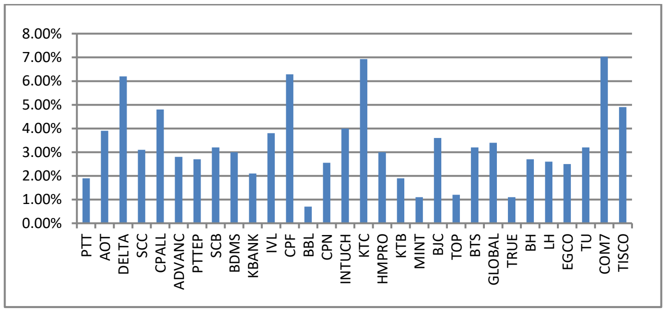


Fig. 4. Optimal portfolio of Thailand market

This is a reasonably particular category for Malaysian stocks because of the profitability and risk that the portfolio generates. Compared to the profitability that the portfolio brings is the lowest of the six velvet research markets. This is also the category with a respectable risk of less than 1%, Five times smaller than the author expected. Moreover, it can be said that although it does not meet the original target set by the author in terms of profitability, with a risk result of only 0.88%, generating a profit of 19.46% of the whole portfolio is a very positive result. The study results have shown the results of the distribution of the proportion of capital investment evenly in stocks. Above the allocation of 5% of the proportion of investment in the portfolio, the result of 4 companies: PMET, TPGC, TLMM and WPHB, with capital allocated of 6.86%, 6.22%, 5.94% and 5.39%, respectively. Most of the above companies are in the form of manufacturing products in aluminum, medical, other services in telecommunications and transportation. However, two stocks are allocated investment levels of less than 1% in the portfolio: PETR (0.88%) and DSOM (0.48%) as evident from Fig. 5.

The results of the research paper in the Indonesian market showed that with the segment of the allocation density of over 5%, there are five stock codes such as EMTK, TBIG, SMMA, TPIA and MEGA with the following proportions as follows 7.39%, 6.8%, 6.78%, 5.62% and 5.28% respectively (see Fig. 6) TBIG and EMTK are two companies specializing in telecommunication technology and the second-largest media company in Indonesia. SMMA specializes in the comprehensive, integrated financial sector, including banking, insurance, asset management and security. TPIA is the leading petrochemical production company using world-class advanced factories. Finally, MEGA is a domestic bank in Indonesia. There are three stock codes: SMGR, BRIS, and ASII, with an allocation of 0.99%, 0.93%, and 0.66%, respectively. SMGR is a state-owned company specializing in domestic cement production, BRIS is a state-owned Islamic bank of Indonesia, ASII is a large corporation specializing in supplying automobile and motorcycle products bearing the brand of Toyota, BMW, Isuzu, Peugeot are the largest in Southeast Asia.

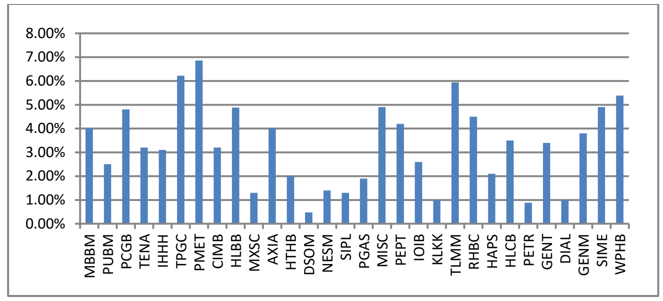


Fig. 5. Optimal portfolio of Malaysia market

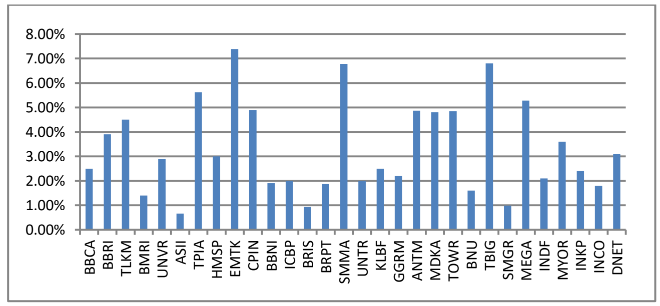


Fig. 6. Optimal portfolio of Indonesia market

For the Philippine market, the author used data from the 30 companies with the highest market capitalization to review and research to deliver fairly consistent attribution results. The results show that there are quite a few companies allocated a proportion of investment above 5%, such as EMP – the largest spirits company in the world with a staggering 7.24%, followed by stock codes such as ACEN, UBP, AGI, HVN, LTG, FGEN, TEL with allocations of 6.85%, 6.31%, 6.26%, 6.19%, 6%, 5.97% and 5.44% respectively as seen in Fig. 7. Among them are companies in electricity production, solar energy, real estate, multi-sectors, serving telecommunication services. However, there are still five companies with an investment capital allocation of less than 1% in the list, including the following stock codes AEV, JGS, SECB, SMC, and ALI. These are primarily multidisciplinary companies in consumption, air transport, and banking.

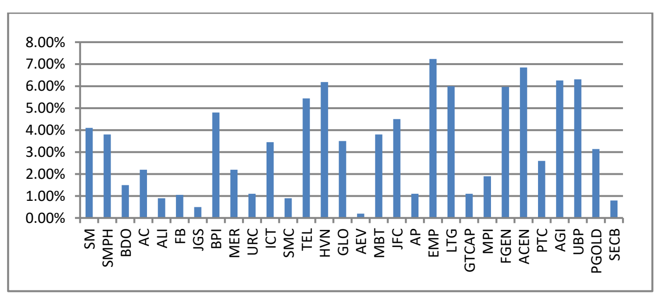


Fig. 7. Optimal portfolio of Philippine market

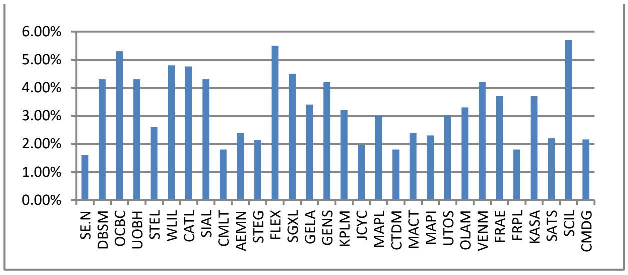


Fig. 8. Optimal portfolio of Singapore market

In the Singapore Stock Exchange, the author used 30 companies listed on the stock exchange with a high market capitalization, and more than half of the portfolio was stock codes in the top straits time index, which is considered a benchmark index for the Singapore market calculated by Singapore Press Holdings, Singapore Exchange and FTSE Group. Moreover, the study results worked when there was no stock code with an investment allocation ratio of less than 1% in this category, which explains that any stock code that appears in the portfolio is highly appreciated and has great capital investment potential. Although the highest allocation in the portfolio has not reached 6% compared to other countries, the even distribution between stocks is very high and homogeneous. The following image shows three stock codes above the 5% allocation: SCIL, FLEX and OCBC, respectively. The proportion is 5.7%, 5.5% and 5.3%, respectively. In particular, SCIL is an urban development and energy company, and FLEX specializes in manufacturing the third-largest multinational electronics in the world. OCBC is a stock code in the banking sector (see Fig. 8).

V. Conclusion

The content of the study is to address and examine the theoretical foundations of portfolio management based on Markowitz's portfolio theory and theories of genetic algorithms such as definitions and cycles contained in the algorithm to establish the portfolios of six different markets in Southeast Asia. The author uses the Python programming language to interpret and assist in producing results and following the steps contained in the genetic algorithm. Based on theory and related studies, the author has obtained very positive results on allocating stocks included in each category with the corresponding profitability and risk level.

The result consists of four categories that meet the author's expectations: the categories in the markets of Vietnam, Thailand, Philippines, and Singapore. The two categories alone are pretty positive in terms of results, namely Malaysia and Indonesia. Although Malaysia has not met the profitability, the risk ratio of less than 1% is a very potential figure. An immovable rule in investing is that the higher the return on their assets, the higher the risk-taking bravery they need. Therefore, for the portfolio of 30 stocks from the Indonesian market with outstanding portfolio results, there is still a level of risk that needs to be accepted, and there must be a trade-off.

Considering the Sharpe ratio of the categories, topped by the 30 companies with the largest market capitalization in Malaysia with 19.81% of the profitability that can be achieved per unit of risk. Although the profitability of this market is not satisfactory to the author, this is still a category worth seriously considering and invests. Ranked second is Singapore, with a Sharpe rate of 15.07%. Thirdly, the Vietnamese market with a not-so-modest figure of 12.67. Ranked fourth in Thailand, the country of golden pagodas reaching 9.33% of the risk compensation level of the whole category. The fourth market among the six countries in Southeast Asia is Indonesia, with a research rate of 8.46% and the Philippine market of 6.9%. After the above statistical step, the significance level is realized when both categories are not satisfactory in terms of profitability, and the expected risk is Malaysia and Indonesia. The Sharpe rate in the Malaysian market is higher than in Indonesia.

References

[1] Yang, X, “Improving portfolio efficiency: A genetic algorithm approach”. Computational Economics, vol 28, 2006.

[2] Lin, C.M., & Gen, M. “An effective decision-based genetic algorithm approach to multiobjective portfolio optimization problem”. Applied Mathematical Sciences, Vol. 1, no. 5, pp. 201-210, 2007.

[3] Sefiane, S., Benbouziane, M. “Portfolio Selection Using Genetic Algorithm”. Journal of Applied Finance & Banking, Vol. 2, No. 4, pp. 143-154, 2012.

[4] Dubinskas, P., & Urbšienė, L. “Investment portfolio optimization by applying a genetic algorithm-based approach”. Ekonomika, Vol. 9, no.6, pp.66-78, 2017.

[5] Cheong, D., Kim, Y.M., Byun, H. W., Oh, K. J., & Kim, T. Y. “Using genetic algorithm to support clustering-based portfolio optimization by investor information”. Applied Soft Computing, Vol. 61, pp. 593-602, 2017.

[6] Liu, C., Gan, W., & Chen, Y. “Research on Portfolio Optimization Based on Affinity Propagation and Genetic Algorithm”. In 2017 14th Web Information Systems and Applications Conference (WISA) (pp. 122-126). IEEE.

[7] Adam, H. Portfolio Management. Book, 2021.

[8] Ahn, W, Hee S. L, Hosun, R, Kyong J. “Asset Allocation Model for a Robo-Advisor Using the Financial Market Instability Index and Genetic Algorithms”. Sustainability, Vol. 12, no. 3, pp. 849, 2020.

Automatic Detection of Alzheimer Disease from 3D MRI Images using Deep CNNs

Nermin Negied[1], Ahmed SeragEldin[2]
School of Engineering and Applied Science, Nile University, Giza, Egypt[1]
School of Communication and Information Engineering, Zewail City, Giza, Egypt[1]
Faculty of Computer Science, MSA, Giza, Egypt[2]

***Abstract*—Alzheimer's disease (AD), also referred to simply as Alzheimer's, is a chronic neurodegenerative disease that usually starts slowly and worsens over time. It is the cause of 60% to 70% of cases of dementia. In 2015, there were approximately 29.8 million people worldwide with AD. It most often begins in people over 65 years of age as it affects about 6% of people 65 years and older, although 4% to 5% of cases are early-onset Alzheimer's which begin before this. In 2015, researchers have figured out that dementia resulted in about 1.9 million deaths. Continuous efforts are made to cure the disease or to delay its progression. Brain imaging is one of the hottest areas in AD research. Techniques like CT, MRI, SPECT, and PET assist in disease detection and help in excluding other probable causes of dementia. Imaging helps to perceive the intended cause of the disease as well as track the disease through its course. This paper applies Image processing and machine learning techniques combined to MRI brain images to help in detection of AD and classify the case either to MDI or Dementia.**

Keywords*—*Alzheimer detection; brain scanning techniques; MRI scanning; image processing; machine learning

I. Introduction

The most common early symptom of AD is the short-term memory loss or in other words the difficulty in remembering recent events. As the disease advances, symptoms can include problems with language, disorientation (easily getting lost), mood swings, loss of motivation, not managing self-care, and behavioral issues [1, 2]. As a person's condition declines, they often withdraw from family and society. Gradually, bodily functions are lost, ultimately leading to death [3]. Although the speed of progression can vary, the average life expectancy following diagnosis is three to nine years [4, 5].

The cause of Alzheimer's disease is poorly understood. About 70% of the risk is believed to be genetic with many genes usually involved [1, 6]. Other risk factors include a history of head injuries, depression, or hypertension. The disease process is associated with plaques and tangles in the brain [6]. A probable diagnosis is based on the history of the illness and cognitive testing with medical imaging and blood tests to rule out other possible causes [7]. Initial symptoms are often mistaken for normal ageing. Examination of brain tissue is needed for a definite diagnosis. Around the world, the current challenge in Alzheimer’s disease is to search certain therapeutic technique that can detain or arrest the disease progression. Neuroimaging can act as a good biomarker used extensively in drug trials. Brain imaging is also helps in diagnosis of AD at its early stage. The clinical studies in AD require heavy cost and thousands of patients are generally incorporated for the accuracy of assessment. Both in academic and industrial studies, the prime focus is on the use of significantly fewer subjects exposed to treatment [8]. The computed tomography (CT), magnetic resonance imaging (MRI), Single-photon emission computed tomography (SPECT) and positron emission tomography (PET) scanning techniques are totally non-invasive and rely on the methods offering molecular diagnostics. These techniques are also extensively used in drug trials. Thus, brain imaging offers a robust and efficient tool and is considered as a biomarker for detecting AD in its prodromal stage.

The remaining of this paper is organized as follows: In Section II is discussed a comparison between different brain scanning techniques. In Section III, an overview of AD detection in literature is presented. Section IV presents the theory. Section V demonstrates the proposed approach, whereas Section VI demonstrates the experimental work and the results achieved. Finally, the paper is concluded in Section VII.

II. Brain Scanning Techniques

The neuroimaging is a powerful biomarker for understanding the pathology of Alzheimer’s disease and its progression. Following are the main different techniques widely used in this area:

MRI scan in AD detection: Magnetic resonance imaging (MRI) scan stand upon the principle of nuclear magnetic resonance. The nucleus of hydrogen atom (proton), residing inside the brain tissue as water or fat, becomes exited with radio wave pulse in a magnetic field. When the pulse is turned off, the proton emits radiofrequency signal and return to its native energy level. Depending upon a combination of different pulses and gradients, a sequence of signals are made and cumulatively form a tissue specific map of the brain [9].

CT scan in AD detection: Defused cerebral atrophy with enlargement of the cortical sulci and ventricular space are the primary criteria for computed tomography (CT) scan diagnosis of AD. Generally, diameters of third and lateral ventricles as well as the bifrontal and bicaudate diameters are linearly measured to evaluate the extent of atrophy. Research reveals that the volume of third and lateral ventricles is increased in AD than control [10]. Scientist have pinpointed that there is a positive correlation between brain atrophy and cognitive loss [11]. Dilated peri-hippocampal fissure forms hippocampal lucency or hypoattenuation in the temporal area medial to the temporal horn and it is a good radiologic marker for AD detection by CT.

PET Scanning in AD detection: Positron emission tomography (PET) scan is a robust, sensitive, powerful, noninvasive technique quantify cerebral blood flow using metabolic property of brain tissue and binging property of amyloidal beta plaque. This technique is essential for accurate diagnosis, evaluation of disease progression and assessment of drug response [12, 13]. In AD, at baseline, PET scan can show relation with decline in mini mental state examination score [14]. A major limitation of PET is it is very expensive.

SPECT Scanning in AD detection: Single-photon emission computed tomography (SPECT) is a noninvasive technique uses photon emitting property of radio nuclei with half-life of 6 to 12 hours. This scan depicts a clear 3D image of brain indicating regional cerebral blood flow. Number of studies with SPECT reveals a positive correlation between severity of dementia and temporoparietal hypofunction [15, 16]. Bilateral temporoparietal hypoperfusion shows 82% positive predictive value for AD [17]. A major limitation of PET is it is very expensive. Fig. 1 shows the four different brain scanning techniques.

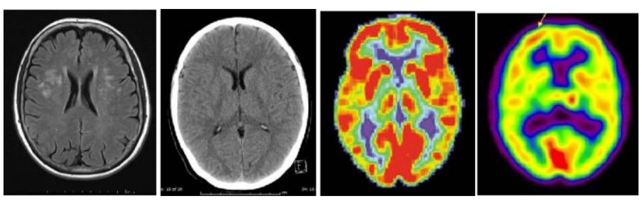


Fig. 1. Different brain scanning techniques, from left to right: MRI scan, CT scan, PET scan, SPECT scan.

III. Literature Review

Many researchers have studied the detection of AD. These studies usually follow two main approaches: the analysis of biomarkers and the examination of patients' decreasing cognitive abilities. The first approach yields reliable results in the detection of AD in its moderate and advanced stages, albeit still performing insufficiently in the early stages of the disease [18]. The second approach has gained more attention in recent years, since, in clinical practice, it has shown promising results in the early detection of AD [19, 20]. Furthermore, when compared to the first approach, the analysis of the decline of cognitive abilities represents an inexpensive and noninvasive alternative. In 2012, Ali El-Zaart and Ali A.Ghosn [21] have proposed a new method for image thresholding (bimodal and multimodal thresholding). Their segmentation method was based on between-class variance and using Gamma distribution. Authors stated that their approach solved the problem of non-symmetric histogram of images by using Gamma rather than Gaussian distribution. They then confirmed that the experimental results obtained by testing their approach showed a good efficiency and satisfying segmentation results, but they mentioned nothing about the accuracy rate or the impact of their segmentation results on the AD automatic detection. Fig. 2 demonstrates the segmented results obtained from that work.

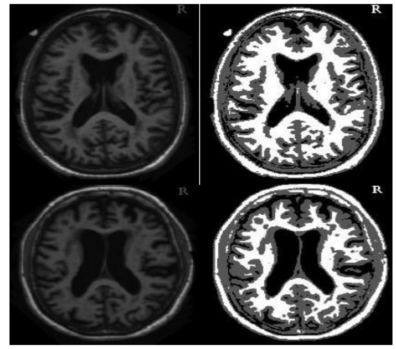


Fig. 2. Original MRI images and segmented Results by El-Zaart and Ghosn approach [21].

Siavash Fashtakeh suggested a research idea in the same year [22]. The idea was based on applying pattern recognition techniques to structural MRI images and then uses automatic classifiers to detect AD in its early stages. ADNI database was used in this study. Brain tissue categorization into white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) was held first followed by customized tissue probability maps created for bias correction. SVM was used then for feature reduction and feature selection. As a future work, the author mentioned some trials about developing better classifier using Neural Network for automate classification. Although the paper contains detailed explanation of the steps and the dataset used, all the results discussed by Fashtakeh research are expected in future! In 2013, Herrera et al [23] have used Discrete Wavelet Transform (DWT) for feature extraction, then Principal Component Analysis (PCA) for feature reduction. Researchers have confirmed that the experimental study showed that the use of dimensionality reduction in this problem led to a worse classification accuracy. The main conclusion of their study accordingly was that all available information is important to achieve a good classification performance. They also confirmed that due to that fact the computational cost of classifiers training will be very large, but the results would be worthy. They then confirmed that their system using SVM without feature reduction showed promising results, but they did not write a clear accuracy rate. S. K. Aruna and S. Chitra [24] two years later used the OASIS MRI brain dataset to compare different machine learning approaches. The researchers firstly used Gabor filter with four different orientations and GLSM to extract features, then they normalized and fused the features. They then used SVM with different kernels to compare between the effectiveness of features obtained by Gabor, GLCM, and the fused features. Authors finally confirmed that their fused features achieve better accuracy than the features obtained by both Gabor and GLCM; however, nothing about the accuracy rates of AD detection was mentioned in their article. In the same year Moradi et al [25] have used SVM and LDS classifiers to detect AD from MRI brain images, with and without feature selection, and compared all the results obtained in all cases. Authors confirmed that LDS classifier with feature selection led to the best results with accuracy rate of 74.74% which was the highest accuracy rate achieved by their work. Bolurchi et al [26] have published the most recent work conducted in this area in 2018, as they used 3D MRI brain images to improve the accuracy rates of automatic AD detection and they assured that they were successful as they used three different feature extractors (PCA, PDF, and contrast homogeneity) with respect to state-of-art, and they fused the features extracted from the three key slices together. Furthermore, they adopted and compared three different classifiers (decision tree, K-NN, and SVM). Authors finally confirmed that SVM achieved the better detection accuracy rate which reached 88.1%. Fig. 3 demonstrates the MRI preprocessing steps.

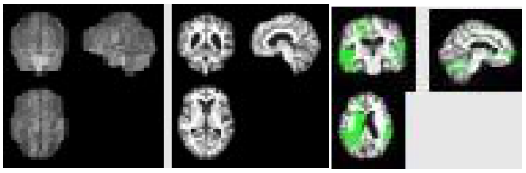


Fig. 3. Preprocessing steps of Bolurchi et al, 3D image (left), key slices (middle), feature detection (right) [26].

IV. Theory

There are mainly three differences between healthy and non-healthy brain images which are: Severely enlarged ventricles often result in changes in the shapes of them, extreme shrinkage of cerebral cortex often results in changes in the outer shape of the brain, and extreme shrinkage of hippocampus [27] (see Fig. 4). Among those three different signs of abnormality this paper limits the automatic detection on the first sign which is the enlargement in ventricles as an obvious abnormality sign that proved to the most promising sign to detect AD automatically from brain MRI images. Two approaches are used here for AD detection: Image thresholding approach and deep learning approach. The following subsections explain the steps of each approach in details.

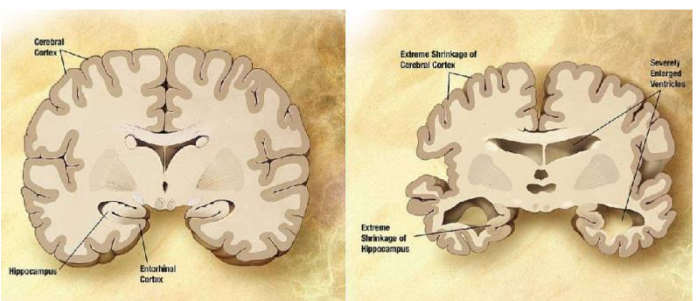


Fig. 4. Comparison of a normal aged brain (left) and the brain of a person with Alzheimer's (right). Characteristics that separate the two are pointed out.

V. Proposed Approach

This work adopts two different approaches to detect Alzheimer form brain MRIs and then combine them together to get a hybrid approach. The following subsections explain the steps of every approach.

A. Image Thresholding

Some image processing steps were conducted to get the MRIs classified to either sick or normal case based on certain threshold. Fig. 6 represents the steps of the image thresholding approach.

The data in ADNI database are stored in dicom files, so firstly the dicom files were converted to normal images. After the dicom file is converted to bng image, this data is an open source which can be found in ADNI website [28]. First, a color space conversion takes place to convert the RGB images to grayscale images. Otsu's threshold [29] is then applied to the grayscale images. Otsu's method is roughly a one-dimensional, discrete analog of Fisher's Discriminant Analysis. It automatically calculates a certain number (Threshold) based on image color analysis. The idea is that it applies color clustering-based image thresholding to convert the image from grayscale image to binary image. The algorithm uses a bi-modal histogram to exhaustively search for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes. Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance. Fig. 5 shows a sample dicom file.

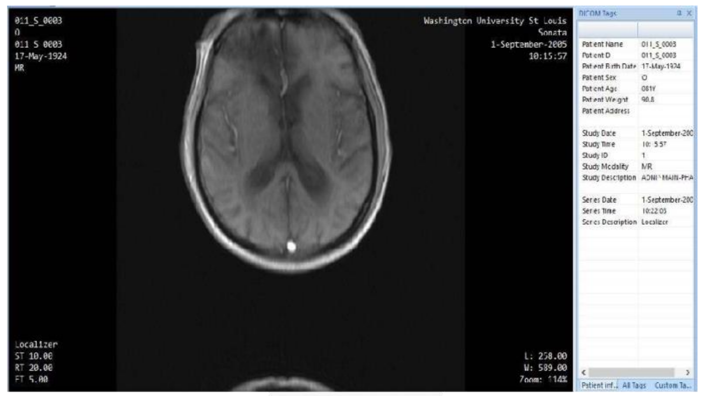


Fig. 5. Sample Dicom file.

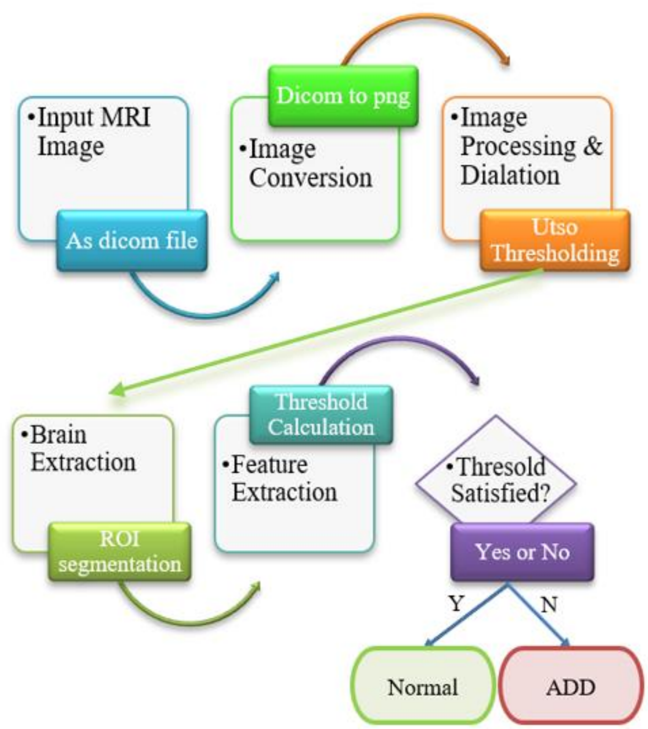


Fig. 6. Image processing phases and steps.

The Algorithm I shows the main steps of the Otsu thresholding method.

Algorithm.1. Otsu Algorithm [29]

1) Compute Histogram and Probabilties of each Intensity Level
2) Set up Initial $\omega_i(0)$ and $\mu_i(0)$
3) Step through all Possible Thresholds t=1, ..., Maximum Intensity
   - Update $\omega_i(0)$ and $\mu_i(0)$
   - Compute $\sigma^2 b(t)$
4) Desired Threshold corresponds to maximum $\sigma^2 b(t)$

Fig. 7 shows image before and after applying Otsu Thresholding.

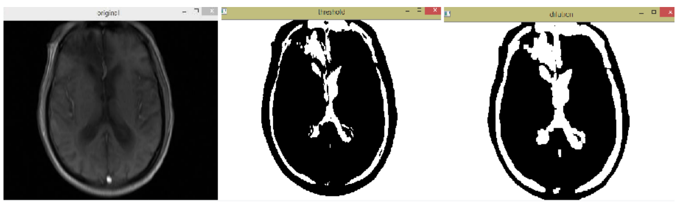


Fig. 7. Axil: The original image (left), the Otsu based threshold image (middle), the dilated image (right).

A morphological image processing is then held to fill the gaps in the ROI (brain ventricles). After wards the brain is then cropped by pixel aggregation. Once the brain image is cropped the ventricles region can be cropped easily as it is in the middle of the image. The ratio of white to black pixels is calculated and the brain is then classified to either normal or sick (with AD).

B. *Deep Learning*

Convolutional Neural Networks (CNN) is then adopted to extract features automatically from the Original MRIs, and then classify the cases. It is mainly composed of five layers: the input layer, the output layer and three layers in between. Where, the first layer contains 16 filters, the second one is composed of 32 filters, then the third layer is composed of 64 filters. Fig. 8 shows the components of the CNN followed by a detailed explanation.

Every layer takes the feature map and enters it to three different components which are Rectified Linear Units (ReLU function), Batch normalization, and Max pooling. ReLU function converts every fraction to 0 and forward the integers as it is to the batch normalization. Batch normalization then normalize the input it received by subtracting the batch mean and dividing it by the batch standard deviation. Max pooling component uses down sampling to reduce the number of features in its input feature map. Following is the Algorithm II of batch normalization transform.

Algorithm.2. Batch normalization transform

Input:
- Values of x over a mini-batch: $\beta = \{x_1,...,m\}$
- Parameters to be learned: γ, β

Output:
- $\{y_i = BN_{\gamma,\beta}(xi)\}$

Mini-batch mean:
- $\sigma_B^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}.(x_i - \mu_B)^2$

Normalize:
- $\ddot{x} \leftarrow \frac{xi - \mu B}{\sqrt{\sigma_B^2 \; + \epsilon}}$

Scale and shift:
- $yi \leftarrow \gamma\,\ddot{x}i + \beta \equiv BN_{\gamma,\beta}(xi)$

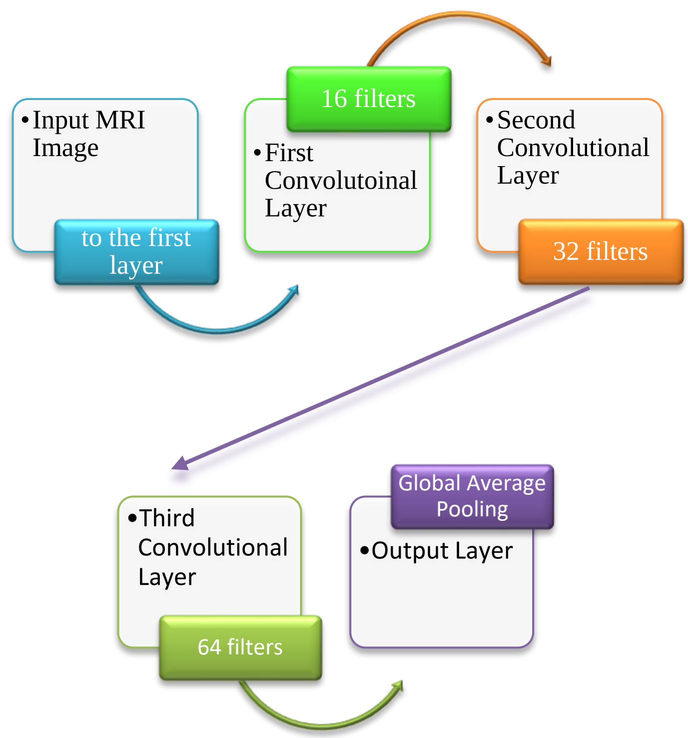


Fig. 8. Layers of CNN.

C. *Hybrid Approach*

In this phase, a combination between Otsu thresholding and CNN approaches was adopted to study the results of a hybrid approach. This time the CNN's input layer takes the output of the Image thresholding approach rather than the original MRI. In other words, the input taken by the CNN is the thresholded binary image of the brain ventricles. This required less computational cost and time as a lot of features are eliminated from the input images entering the input layer of the CNN. Fig. 9 shows the combination flow of the two approaches.

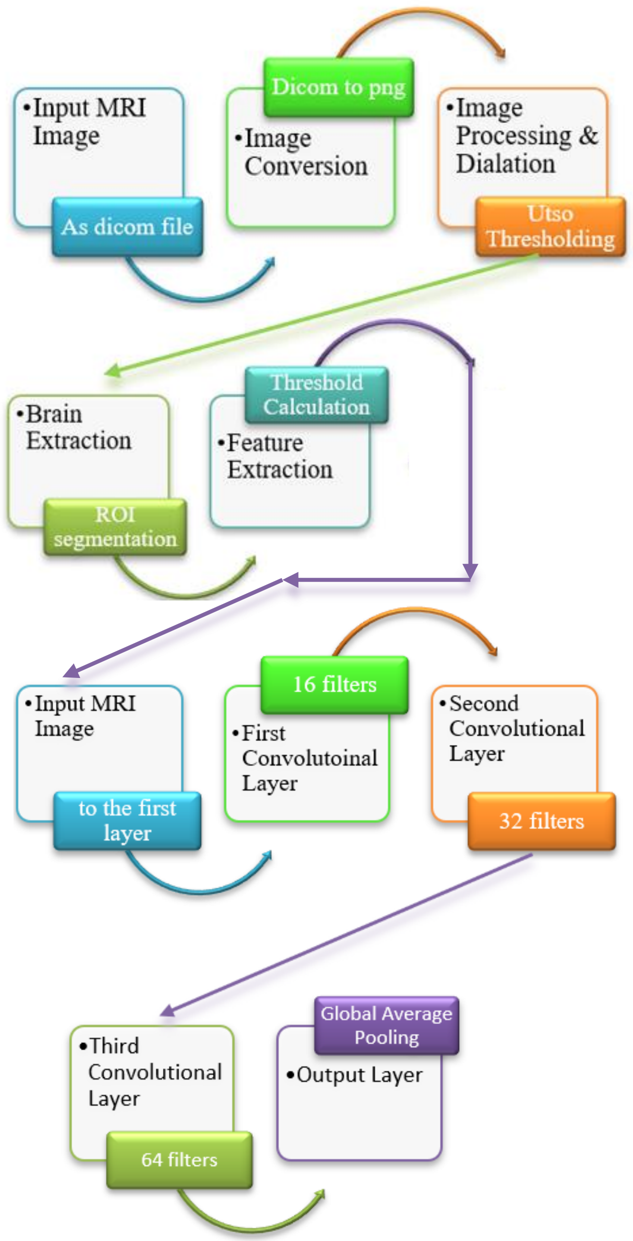


Fig. 9. The hybrid approach (Image Thresholding + CNN).

VI. Experimental Results

ADNI Database was used to test and validate this work. Researchers and collectors of ADNI confirmed that when significant abnormalities are seen on the screening MRI scan, such as a hemispheric infarction, the participant is excluded. Additionally, sites are notified by email if a radiological finding that is not normal for age is identified by the MRI QC team [30].

The following tables show the results obtained by Image thresholding, CNN, and the hybrid approach respectively. The tests were conducted using the 3D MRI images, i.e., Axial, Sagittal, and Coronal.

As it can be observed from Tables I, II, and III, CNN achieved the highest accuracy rate in case of Axial scanning direction. It can also be seen that the Axial direction is the best in AD detection whatever the approach adopted. However, CNN takes more time and computational cost to detect the disease; it outperforms the results of image processing and the hybrid approach concluding that deep learning approaches are much better in such a field. The preceding fact tells that all features are important for automatic AD detection and that feature reduction in not a preferred solution for lowering time and cost. Nevertheless, it must be mentioned that surprisingly the hybrid approach outperforms CNN in case of coronal brain scanning direction.

TABLE I. Results Obtained by Image Thresholding Approach

MRI	Accuracy rate
Axial	76%
Sagittal	68%
Coronal	71%

TABLE II. Results Obtained by CNN Approach

MRI	Accuracy rate
Axial	**90%**
Sagittal	**85%**
Coronal	71%

TABLE III. Results Obtained by the Hybrid Approach

MRI	Accuracy rate
Axial	81%
Sagittal	70%
Coronal	**75%**

VII. Conclusion and Future Work

Three different approaches for automatic AD detection from MRIs were adopted in this work: Otsu thresholding, CNN, and a hybrid approach between them. The three approaches have achieved promising accuracy rates compared to literature. A comparative study was also conducted and proved that CNN outperforms image thresholding and even the hybrid approach in terms of accuracy rates. At the same time, image thresholding methodologies require less time and computational cost. However, by comparing the accuracy rates, CNN worth its computational time and cost. ADNI brain MRI Database was used for the testing process. 3D MRIs were adopted for the validation.

References

[1] Burns A, Iliffe S (5 February 2009). "Alzheimer's disease". The BMJ. 338: b158. doi:10.1136/bmj. b158. PMID 19196745. (Subscription required (help)).

[2] "Dementia Fact sheet N°362". World Health Organization. March 2015. Archived from the original on 18 March 2015. Retrieved 13 January 2016.

[3] "About Alzheimer's Disease: Symptoms". National Institute on Aging. Archived from the original on 15 January 2012. Retrieved 28 December 2011.

[4] Querfurth HW, LaFerla FM (28 January 2010). "Alzheimer's disease". The New England Journal of Medicine. 362 (4): 329–44. doi:10.1056/NEJMra0909142.

[5] Todd S, Barr S, Roberts M, Passmore AP (November 2013). "Survival in dementia and predictors of mortality: a review". International Journal of Geriatric Psychiatry. 28 (11): 1109– 24. doi:10.1002/gps.3946.

[6] Ballard C, Gauthier S, Corbett A, et al. (19 March 2011). "Alzheimer's disease". Lancet. 377 (9770): 1019–31. doi:10.1016/S0140-6736(10)61349-9.

[7] "Dementia diagnosis and assessment". National Institute for Health and Care Excellence (NICE). 30 November 2014.

[8] Pranami Bhaumik, Priyanka Ghosh and Subrata Kumar Dey, " Role of Brain Imaging Techniques in Detection of Alzheimer's Disease", Journal of Ayurveda, Yoga, Naturopathy, Unani, Siddha and Homoeopathy-JAYUSH Volume. 1 Issue. 1. 2017.

[9] K.A. Johnson, N.C. Fox, R.A. Sperling, W.E. Klunk," Brain imaging in Alzheimer disease," Cold Spring Harbor Perspectives in Medicine, vol. 2, no. 4, pp. a006213, Apr 2012, doi: 10.1101/cshperspect.a006213.

[10] J.S. Luxenberg, J.V. Haxby, H. Creasey, M. Sundaram, S.I. Rapoport," Rate of ventricular enlargement in dementia of the Alzheimer type correlates with rate of neuropsychological deterioration," Neurology, vol. 37, no. 7, pp.1135-1140, Jul 1987.

[11] J.M. Bird, R. Levy, R.J. Jacoby, "Computed tomography in the elderly: changes over time in a normal population," British Journal of Psychiatry, vol. 148, pp. 80-85, Jan 1986.

[12] G.A. Marshall, L. Monserratt, D. Harwood, M. Mandelkern, J.L. Cummings, D.L. Sultzer," Positron emission tomography metabolic correlates of apathy in Alzheimer disease," Archives of neurology,vol. 64, no. 7, pp.1015-1020, 2007.

[13] M. Shimojo, M. Higuchi, T. Suhara, N. Sahara, "Imaging Multimodalities for Dissecting Alzheimer's Disease: Advanced Technologies of Positron Emission Tomography and Fluorescence Imaging," Frontiers in Neuroscience, vol. 9, pp. 482, 2015.

[14] W.J. Jagust, M.N. Haan, J.L. Eberling, N. Wolfe, B.R. Reed,"Functional imaging predicts cognitive decline in Alzheimer's disease," Journal of Neuroimaging, vol. 6, pp. 156-160, 1996.

[15] K.A. Johnson, S.T. Mueller, T.M. Walshe, R.J. English, B.L. Holman, "Cerebral perfusion imaging in Alzheimer's disease. Use of single photon emission computed tomography and iofetamine hydrochloride I 123," Archives of neurology, vol. 44, no. 2, pp. 165-168, 1987.

[16] S.T. DeKosky, W.J. Shih, F.A. Schmitt, J. Coupal, C. Kirkpatrick, "Assessing utility of single photon emission computed tomography (SPECT) scan in Alzheimer disease: correlation with cognitive severity," Alzheimer Disease and Associated Disorders, vol. 4, no. 1, pp. 14-23,1990.

[17] B.L. Holman, K.A. Johnson, B. Gerada, P.A. Carvalho, A. Satlin, "The scintigraphic appearance of Alzheimer's disease: a prospective study using technetium-99m-HMPAO SPECT," Journal of Nuclear Medicine, vol. 33, no. 2, pp.181-185, 1992.

[18] Alzheimer's Association. 2015. 2015 Alzheimer's Disease Facts and Figures. Technical report, Alzheimer's Association.

[19] Vanessa Taler and Natalie A. Phillips. 2008. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. Journal of clinical and experimental neuropsychology, 30(5):501–56, July.

[20] Johannes Schroder, Britta Wendelstein, and Ekkehard ¨ Felder. 2010. Language in the Preclinical Stage of Alzheimer's Disease. Content and Complexity in Biographic Interviews of the ILSE Study. In Klinische Neurophysiologie, volume 41, page S360.

[21] Ali El-Zaart and Ali A.Ghosn, "MRI IMAGES THRESHOLDING FOR ALZHEIMER DETECTION", white paper, Department of Mathematics and Computer Science, Beirut Arab University, Beirut, Lebanon. 2012.

[22] Fashtakeh, S. (2012), "Early detection of Alzheimer's disease using structural MRI: A research idea". Life Science Journal-Acta Zhengzhou University Overseas Edition, vol. 9(3), pp. 1072-1079.

[23] Luis Javier Herrera*, Ignacio Rojas, H. Pomares, A. Guillén, O. Valenzuela, O. Baños, (2013), "Classification of MRI images for Alzheimer's diseasedetection", SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013.

[24] S. K. Aruna and S. Chitra, "Machine Learning Approach for Identifying Dementia from MRI Images", World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering, Vol:9, No:3, 2015.

[25] Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen, and Jussi Tohka, "Machine learning framework for earlyMRI-based Alzheimer's conversion prediction in MCI subjects", Journal of NeuroImage, 104 (2015) 398–412, Elseiver. www.elsevier.com/locate/ynimg.

[26] Pouya Bolurchi, Masoud Moradi, and Hassan Demirel, "Alzheimer's disease detection by using key slice selection in 3D MRI images", UKSim-AMSS 20th International Conference on Modelling & Simulation, 2018.

[27] https://en.wikipedia.org/?title=Alzheimer%27s_disease.

[28] https://adni.loni.usc.edu.

[29] Sunil L. Bangare, Amruta Dubal, Pallavi S. Bangare, and Suhas Patil, "Reviewing Otsu's Method For Image Thresholding", International Journal of Applied Engineering Research 10(9):21777-21783, May 2015.

[30] Alzheimer's Disease Neuroimaging Initiative (ADNI), USC Mark and Mary Stevens Neuroimaging and Informatics Institute. 2020.

Multi-Modal Medical Image Fusion Using Transfer Learning Approach

Ms. Shrida Kalamkar, Dr. Geetha Mary A
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India

***Abstract*—Multimodal imaging techniques of the same organ help in getting anatomical as well as functional details of a particular body part. Multimodal imaging of the same organs can help doctors diagnose a disease cost-effectively. In this paper, a hybrid approach using transfer learning and discrete wavelet transform is used to fuse multimodal medical images. As the access to medical data is limited, transfer learning is used for feature extractor and save training time. The features are fused with a pre-trained VGG19 model. Discrete Wavelet Transform is used to decompose the multimodal images in different sub-bands. In the last phase, Inverse Wavelet Transform is used to obtain a fused image from the four bands generated. The proposed model is executed on Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) datasets. The experimental results show that the proposed approach performs better than other approaches and the significance of the obtained fused image is measured using qualitative metrics.**

Keywords*—*Image fusion; discrete wavelet transform; computer vision; inverse wavelet transform

I. Introduction

The process of fusing multiple images of different imaging modalities to obtain an image with a large amount of information for increasing the clinical applicability of medical images is multimodal medical image fusion [1]. Medical Imaging techniques like Magnetic Resonance Imaging (MRI) give the human body structural characteristics. In contrast, Computed Tomography (CT) provides cross-sectional images from within the body, and Positron Emission Tomography (PET) is provided soft tissue. Single-photon Emission Computed Tomography (SPECT) provides 3D images of different human body organs[1], [2]. Different imaging techniques provide different characteristics and information about the same part of the human body. The purpose of the fusion is to obtain better contrast and fusion quality. The result of the fusion should meet the following conditions [3]:

1) The newly obtained fused image should preserve the original information of source images without any information loss [3].

2) The newly obtained fused image should not introduce any inconsistencies or artefacts, and

3) Should not add misregistration and noise in the newly fused image.

Medical Image Fusion finds its applicability in various diagnostic problems. CT and MRI images of the skull can be fused, providing doctors with both structural and cross-sectional information about the skull. This can help in identifying more clearly any skull-based tumors or diseases[4], [5]. Similarly, MRI and ultrasound images can be fused to confirm vascular blood flow. Many reputed medical institutes use very expensive image fusion tools to fuse multiple medical image modalities to diagnose images. Therefore, there is a need to develop a low-cost model for fusing such modalities of images that can benefit small clinics and hospitals [6]–[8]. Recently Deep learning architectures have become more successful for image processing due to the availability of many public datasets and optimization techniques. Deep neural networks have their role in various medical-related applications. Deep networks can be trained for specific clinical applications with the intended dataset, fast computing system, and a huge volume of data. Pre-trained networks can also be used for medical image processing towards specific clinical applications[9], [10]. This paper focuses on the survey and development of a novel transfer learning and wavelet-based technique for multimodal medical image fusion.

The contributions of the paper are stated as follows:

1) This paper reviews the different multimodal image fusion techniques in the spatial domain, transfer domain and deep learning domain.

2) Discrete wavelet transform (DWT) is used initially, which samples the approximation coefficients and detail coefficients at each level.

3) To handle the issue of limited multimodal medical images, a transfer learning approach based on the VGG19 Framework for any modality of a medical image without a training set is proposed.

The contents of the paper are organized as follows. Section II elaborates on the image fusion methods and the need for transfer learning in multimodal medical image fusion. Section III gives information about related research in the field. The problem of image fusion for multimodal medical images using the transform domain approach and VGG19 transfer learning framework and the detailed methodology has been described in Section IV. Section V illustrates experimentation details; datasets used, Fusion metrics and performance analysis used for evaluating the proposed methodology, qualitative metrics, and related comparative analysis in all three domains. Section VI portrays the concluding remarks and pointers to extend the current research work.

II. BACKGROUND

The rapid growth of medical imaging technologies such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and single photon emission computed tomography (SPECT) has provided us much richer information on the physical condition[8], [11].

Table I briefly overviews different medical image modalities used to treat a disease.

TABLE I. DIFFERENT MEDICAL IMAGING MODALITIES

Sr.No	Medical Image Modality	Methodology of taking an image	Type of image obtained
1	CT Scan[7], [8], [12]	Uses X-ray equipment to take images from within the human body from different angles.	Cross-sectional images of all types of body tissue and organs
2	MRI Scan[7], [8], [12]	Detailed images from within the human body using magnetic and radio waves.	Structural Characteristics of the human body are obtained. Gets Anatomic and contrast details between normal and abnormal tissues.
3	SPECT Scan[7], [8], [12]	Uses nuclear medicine for therapeutic and diagnostic procedures.	The body's temperature, blood flow details, etc., are obtained.
4	PET Scan[7], [8], [12]	It uses a radioactive drug to give information on how tissues and organs function.	Sometimes detect disease before it is detected using other imaging tests. It Delivers details about the area of disease.

In literature, image fusion can be carried out in three ways:

A. Pixel Level Image Fusion

In this mode of fusion, the original information in the source images is directly combined without extracting the relevant features [13]. Different Algorithms like the Color space model, Statistical Algorithms, multiresolution decomposition algorithms, and radiometric algorithms can be used to perform pixel-level fusion. Fig. 1 shows an overview of pixel-level image fusion.

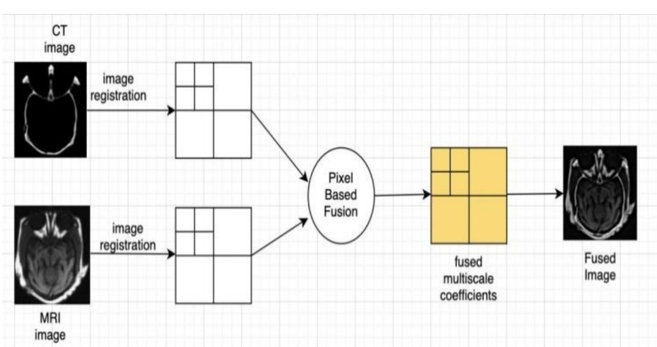


Fig. 1. Pixel-level image fusion.

B. Feature Level Fusion

Feature level fusion, features from the input images using some feature extraction algorithms are extracted to perform the fusion. A region-based fusion approach is used in this type of fusion[14]. Recently, transfer learning algorithms have proven to work well for extracting features from images. Fig. 2 shows an overview of feature-level image fusion.

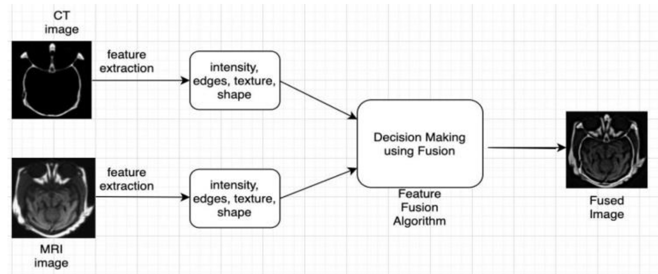


Fig. 2. Feature-level image fusion.

C. Decision-level Fusion

Compared to pixel-level and feature-level fusion, decision-level fusion is accurate and supports real-time data fusion. The main disadvantage of this type of fusion is the high loss of information[10]. Algorithms include voting, Bayesian inference, evidence theory, and fuzzy integral [15].

III. RELATED WORK

There are many different methods used in the medical field to fuse different modalities in medical images. A detailed review of multimodal medical image fusion is extensively covered in [4] Medical Image Fusion methods are divided into spatial domain, transform domain, and deep learning domain shown in fig 3.

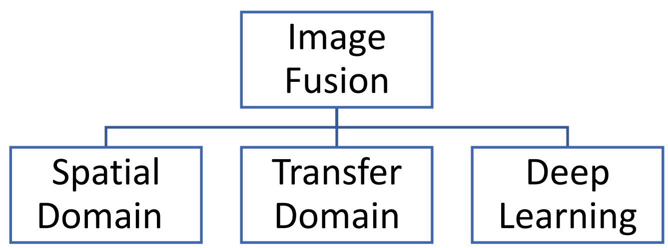


Fig. 3. Image fusion domains.

Table II shows a list of different algorithms in all three domains and systematically reviews different image fusion techniques in all three domains. The algorithms in the spatial domain were popular among the researchers. But spatial domain methods create spatial and spectral distortion of fused images. Hence, researchers used transform domain methods to perform a fusion of images for better fusion results[16]. The fusion methods in the transform domain transform the source image into the frequency domain and then perform reconstruction operations. The disadvantage of the transform domain is that it generates noise in images during fusion processing. With the advancement in deep learning technology, image fusion methods based on deep learning were introduced recently.

TABLE II. MEDICAL IMAGE FUSION METHODS

Domain	Description	Fusion Method	Advantages	Disadvantages
Spatial Domain[1]	Different fusion rules are applied to pixels in the source image.	1. High Pass Filtering 2. Principal Component Analysis 3. Saturation Method of Hue Intensity 4. Minimum and Maximum selection method 5. Brovey Method 6. Average Method	Simple Method. It can be combined as a part of the frequency domain to generate new research methods.	Spectral and Spatial Distortion of Fused Image
Transform Domain[1]	Based on multiscale transform (MST) theory. It transforms the source image from the time domain to the frequency domain. low and high-frequency coefficients are obtained	1. Pulse Coupled Neural Network (PCNN) 2. Discrete Wavelet Transform (DWT) 3. Nonsubsampled Contourlet Transform (NSCT) 4. Nonsubsampled Shearlet Transform(NSST)	Good structure and avoids distortion	Generate noise in image during fusion
Deep Learning[17]	Aim at learning feature hierarchies. Features from higher levels combined with lower-level features. Spatial and Transform Domain-based image fusion methods fail in automatic feature extraction and generation of fusion rules.	1. Convolution Neural Network (CNN) 2. U-Net 3. Stacked Auto-encoders(SAE) 4. Deep Boltzmann Machine(DBM)	Extract the relevant features from the image automatically	Some deep learning models neglect the spatial and temporal topologies of the multimodal data

IV. PROPOSED METHODOLOGY

A. Problem Formulation

We formulate the task of fusion of features from multiple modalities of images as a multimodal medical image fusion problem at the feature level. Several algorithms extract the features of interest from an image, ignore irrelevant features, and eventually apply rules to fuse the important feature during image fusion. These rules play an important role or are the key concept in the fusion process. The selection of proper fusion rules is crucial, resulting in a better fusion process. It is impossible to develop a generic fusion rule for all fusion applications. A more feasible approach in medical image fusion is to develop a new image fusion method that can merge the features extracted from different images into a single fused image.

B. Proposed Methodology for Multimodal Medical Image Fusion

Our proposed work aims to develop a hybrid approach by integrating the structural details provided by CT images with soft tissue details provided by MRI images to get a unique fused image. We have proposed a multimodal medical image fusion based on a blend of Discrete Wavelet Transform and transfer learning VGG19 architecture, which uses the concept of CNN.

The DWT of an image is calculated by passing it through a series of filters, i.e., low pass filter, and then through a high-pass filter to get the approximate and detailed features.

VGG networks are image style transfer methods based on features extracted from the content, style, and generated image. The VGG19 uses the ImageNet dataset for feature extraction and is a pre-trained network. This architecture of VGG19 can be effectively utilized in VGG19-based multimodal image fusion. Fig. 4 graphically represents the proposed CT-MRI medical image fusion architecture.

The following algorithms describe the methodology used for image fusion.

Algorithm 1: Generation of co-efficient from a medical image using DWT.

Input: CT and MRI image
Output: Approximate and detailed co-efficient (LL,LH,HL,HH)

DWT(image, DWT):

1. Decompose CT image using DWT to generate detail and approximate co-efficient.
2. Decompose MRI images using the DWT to generate detail and approximate co-efficient.
3. Pass this co-efficient to VGG19 architecture for fusion.
4. Repeat the function on image 2.

Algorithm 2: Fusion of generated co-efficient using VGG19

Input: Detailed and Approximate co-efficient from CT and MRI images
Output: Fused Image

FUSION(LL,LH,HL,HH):

1. Convert all images to YCbCr format
2. Transfer all images to PyTorch tensors
3. Perform fuse strategy
4. Reconstruct fused image given RGB input images
5. Return fused image

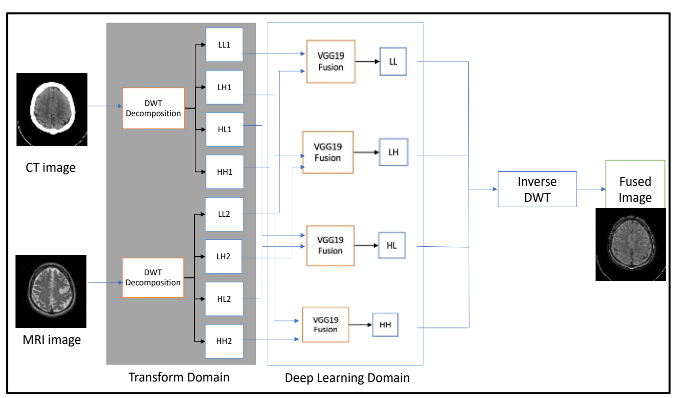


Fig. 4. Proposed CT-MRI Medical Image Fusion Architecture.

In this paper, we have used multimodal medical images, i.e., CT and MRI, the source images given as input to the pre-trained VGG19 model. Initially, we used the Discrete Wavelet decomposition technique of the transform domain to decompose the source images, i.e., CT and MRI images, to generate approximate coefficients of the handcrafted features.

CT images are decomposed to generate LL1 coefficient and horizontal, vertical, and diagonal coefficients, i.e., LH1 (horizontal), HL1 (vertical), and HH1 (diagonal). Similarly, MRI images are decomposed to generate LL2 coefficient and horizontal, vertical, and diagonal coefficients, i.e., LH2 (horizontal), HL2 (vertical), and HH2 (diagonal), as shown in Fig. 5 and Fig. 6.

Later the features from lower-level coefficients, i.e., LL1 and LL2, Horizontal coefficients, i.e., LH1 and LH2, vertical coefficients, i.e., HL1 and HL2, and diagonal coefficients, i.e., HH1 and HH2 of CT and MRI images, respectively are extracted and fused using VGG19 model.

LL, LH, HL, and HH bands are obtained after VGG19 on the input coefficients. The LL band has essential information of an original image. The LH, HL, and HH bands have horizontal, vertical, and diagonal information.

Finally, the frequency sub bands LL, LH, HL, and HH are combined using VGG19 architecture to generate the fused image after applying inverse discrete wavelet transform (IDWT)

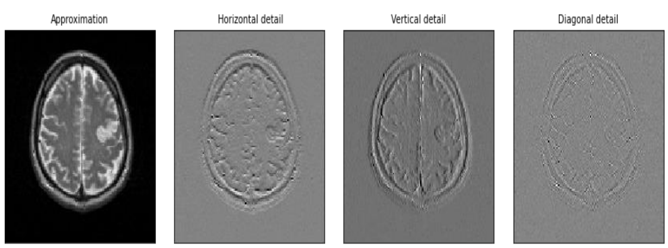


Fig. 5. Wavelet transform of MRI image.

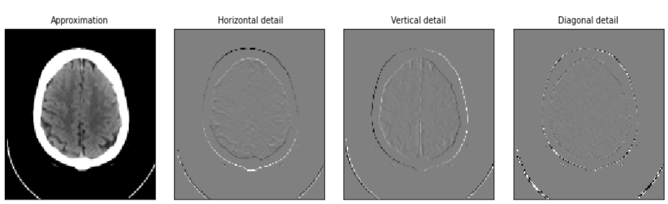


Fig. 6. Wavelet transform of CT Image.

V. Experimental Results on CT and MRI Images

Different fusion algorithms based on spatial, transform, and deep learning domains are used to compare the performance of the VGG19-based medical image fusion.

Spatial Domain methods like Principal Component Analysis (PCA), LPCA, and FCMPCA. Transform domain methods like Standard Wavelet Transform (SWT), Dual-Tree Complex Wavelet Transform (DTCWT), and Non-Subsampled Contourlet Transform (NSCT) are analyzed. Fusion is employed in the wavelet transformation of both CT and MRI images. Wavelet analysis provides a self-adaptive and localized analysis, which is applicable to the time domain and frequency domain. This analysis can focus on any details of the time domain and frequency domain. Multi-scale decomposition of the image based on DWT extracts low-frequency information, as well as horizontal, vertical and diagonal directions of the high-frequency details.

In the VGG19 transfer learning method, fusion is done by applying DWT to CT and MRI images. Both these images are then decomposed into four sub-bands. The sub-bands labelled LH1, HL1 and HH1 represent the detail images co-efficient while the sub-band LL1 represents the approximation image.

A. Evaluation Metrics

1) Peak Signal to Noise Ratio (PSNR): PSNR can be defined as the difference between x and y.

$$\text{PSNR in dB} = 10log_{10}\ \frac{255^2}{\sqrt{\sum_x \sum_y (x-y)^2}} \tag{1}$$

PSNR is calculated using the two source images and the fused image.

$$PSNR_{avg} = \frac{1}{2}\ (PSNR(IM_1, FI) + PSNR(IM_2, FI)) \tag{2}$$

Where,

- IM_1 -> Image 1,
- IM_2 -> Image 2,
- FI -> Fused Image

A higher value of PSNR indicates better quality of the fused image.

2) Structural Similarity Index (SSIM): It contains the structural information present in the images. The loss of structural information leads to distortion of images.

$$\text{SSIM(x,y)}=\left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1}\right)^{\alpha}\left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2}\right)^{\beta}\left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y+C_3}\right)^{\gamma} \ldots..(3)$$

Where,

- μ_x and μ_y are the mean of the images x and y and represents the luminance of the image.
- σ_x and σ_y are the variances. It represents the contrast of the image.
- σ_{xy} is the covariance. It illustrates the correlation between images.
- α, β, and γ - Adjust the relative significance of the above terms. The stability of the metric is C_1, C_2, and C_3- constants that maintain the metric's strength.

Average SSIM is calculated as:

$$SSIM_{avg} = \frac{1}{2}\ (SSIM(IM_1, FI) + SSIM(IM_2, FI)) \tag{4}$$

Where,

- $SSIM_{avg} \rightarrow Average\ SSIM$

- $SSIM(IM_1, FI) \rightarrow$ $SSIM\ between\ source\ image\ 1\ and\ fused\ image$
- $SSIM(IM_2, FI) \rightarrow$ $SSIM\ between\ source\ image\ 2\ and\ fused\ image$

The range of SSIM values is between 0 to 1; 1 indicates a perfect match between the reconstructed image and the original image.

The higher the value of SSIM better is the fused image.

3) Mutual Information (MI): In MI, how much information from two input source images (X, Y) is transferred to the fused image is calculated. The formula is as follows:

$$MI = I(x,f) + I(y,f) \quad (5)$$

$$I(x,y) = \sum_{y \in Y} \sum_{x \in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \quad (6)$$

Where,

- p(x) and p(y) - the edge probability density function of the two images,
- p(x,y) - the joint probability density function of the fused image and the source image X, Y.

Higher the MI value, more information is transferred from the original images to the fused image.

4) Entropy (EN): Information content in the image by taking values between 0 and 8.

$$EN = - \sum_{L=0}^{L-1} p_i\ X\ log_2 p_i \quad (7)$$

Where,

L - number of grey levels and is a probability density function for each gray value i.

B. Experiment and Evaluation on the Group of MRI-CT Images

CT and MRI data were used in this experiment of fusion. The evaluation metrics used in this experiment have been discussed in the above section. The dataset used for this experimentation is available in The Whole Brain Atlas repository hosted by www.harvardmed.edu. All the source images are registered, which is the prerequisite for fusion. The source images are of spatial resolution 256×256×3.

We have conducted detailed experimental studies on different CT and MRI test images. The evaluation metrics used in this paper are PSNR, SSIM, MI and EN. The fusion results have been compared with both wavelet-based and neural network-based fusion techniques. Table III below presents the evaluation index data of different methods of MRI-CT fusion.

The performance of the multimodal medical image fusion methods of both wavelet-based and neural network-based fusion are evaluated using four qualitative metrics such as EN, MI, PSNR, and SSIM.

Deep learning techniques perform automatic and near-accurate feature extraction, so the Convolutional Neural Network algorithm achieves the best Entropy (EN), i.e., the amount of information contained in fused images is the largest. Our proposed DWT+VGG19 algorithm performed best on the PSNR and SSIM evaluation indicators. It can measure a large amount of structural information present in the images. High PSNR means a good quality image and a minor error introduced to the image. Also, the higher the value of SSIM better is the fused image. Hence our proposed algorithm for medical image fusion outperforms the other wavelet-based and neural network-based algorithms in the quality of the image.

The MI and Entropy metrics show that wavelet-based algorithms have poor performance if high noise is present in images. In future, we will use VGG19 architecture with averaging and maximum fusion rules and fuse the detailed contents.

TABLE III. EVALUATION INDEX DATA OF DIFFERENT METHODS OF MRI-CT FUSION

Fusion Methods	**MRI-CT evaluation metrics**			
	PSNR	SSIM	MI	EN
GFF	31.16	0.49	3.43	6.80
MSA	35.07	0.48	3.92	6.38
NSCT+SR	29.56	0.48	3.13	6.89
NSCT+PCNN	31.23	0.50	**5.01**	6.74
NSCT+LE	31.61	0.49	3.29	6.80
NSCT+RPCNN	31.68	0.50	4.42	6.77
NSST+PAPCNN	32.92	0.49	3.23	6.84
DWT	31.97	0.43	1.98	6.59
DWT+WA	30.98	0.49	4.75	6.30
U-Net	26.42	0.32	2.22	5.13
CNN	28.96	0.48	3.08	**7.01**
Proposed model	**40.63**	**0.67**	2.88	4.93

VI. CONCLUSION

Deep learning for multimodal medical image fusion has been a recent research topic. Many researchers are focusing on using deep learning in the fusion of different modalities of images. This paper majorly discusses the following points.

1) Image fusion methods in spatial, transform, and deep learning domains are discussed in this paper.

2) Automatic feature extraction using Deep learning models leads to the generation of the most compelling features, eventually improving model accuracy.

3) The VGG19 transfer learning method proposed in this paper can achieve multimodal medical image fusion. This method can effectively achieve image fusion and help in medical diagnosis. It can help in improving the accuracy of medical diagnoses.

4) Experimental results on CT-MRI data show that the proposed transfer learning method achieves state-of-art performance in terms of qualitative evaluation metrics.

REFERENCES

[1] W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," Neural Comput Appl, vol. 2, 2020, doi: 10.1007/s00521-020-05173-2.

[2] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, “Survey study of multimodality medical image fusion methods,” Multimed Tools Appl, vol. 80, no. 4, pp. 6369–6396, 2021, doi: 10.1007/s11042-020-08834-5.

[3] V. Rajangam, N. Sangeetha, R. Karthik, and K. Mallikarjuna, “Performance analysis of VGG19 deep learning network based brain image fusion,” Handbook of Research on Deep Learning-Based Image Analysis Under Constrained and Unconstrained Environments, pp. 145–166, 2020, doi: 10.4018/978-1-7998-6690-9.ch008.

[4] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, “A Review of Multimodal Medical Image Fusion Techniques,” Comput Math Methods Med, vol. 2020, 2020, doi: 10.1155/2020/8279342.

[5] B. Meher, S. Agrawal, R. Panda, and A. Abraham, “A survey on region based image fusion methods,” Information Fusion, vol. 48, no. December 2017, pp. 119–132, 2019, doi: 10.1016/j.inffus.2018.07.010.

[6] A. Esteva et al., “Deep learning-enabled medical computer vision,” NPJ Digit Med, vol. 4, no. 1, pp. 1–9, 2021, doi: 10.1038/s41746-020-00376-2.

[7] J. M. Dolly and A. K. Nisa, “A Survey on Different Multimodal Medical Image Fusion Techniques and Methods,” Proceedings of 1st International Conference on Innovations in Information and Communication Technology, ICIICT 2019, pp. 1–5, 2019, doi: 10.1109/ICIICT1.2019.8741445.

[8] A. Dogra, B. Goyal, and S. Agrawal, “Medical image fusion: A brief introduction,” Biomedical and Pharmacology Journal, vol. 11, no. 3, pp. 1209–1214, 2018, doi: 10.13005/bpj/1482.

[9] V. Rajangam, N. Sangeetha, R. Karthik, and K. Mallikarjuna, “Performance analysis of VGG19 deep learning network based brain image fusion,” Handbook of Research on Deep Learning-Based Image Analysis Under Constrained and Unconstrained Environments, pp. 145–166, 2020, doi: 10.4018/978-1-7998-6690-9.ch008.

[10] G. Xiao, D. Prasad, B. Gang, and X. Zhang, Image Fusion.

[11] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, “Survey study of multimodality medical image fusion methods,” Multimed Tools Appl, vol. 80, no. 4, pp. 6369–6396, 2021, doi: 10.1007/s11042-020-08834-5.

[12] G. Sreeja and O. Saraniya, Image Fusion Through Deep Convolutional Neural Network. Elsevier Inc., 2019. doi: 10.1016/b978-0-12-816718-2.00010-5.

[13] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, “Pixel-level image fusion: A survey of the state of the art,” Information Fusion, vol. 33, pp. 100–112, 2017, doi: 10.1016/j.inffus.2016.05.004.

[14] M. Ehatisham-Ul-Haq et al., “Robust Human Activity Recognition Using Multimodal Feature-Level Fusion,” IEEE Access, vol. 7, pp. 60736–60751, 2019, doi: 10.1109/ACCESS.2019.2913393.

[15] N. Tawfik, H. A. Elnemr, M. Fakhr, M. I. Dessouky, and F. E. Abd El-Samie, “Survey study of multimodality medical image fusion methods,” Multimed Tools Appl, vol. 80, no. 4, pp. 6369–6396, 2021, doi: 10.1007/s11042-020-08834-5.

[16] M. Kaur and D. Singh, “Fusion of medical images using deep belief networks,” Cluster Comput, vol. 23, no. 2, pp. 1439–1453, 2020, doi: 10.1007/s10586-019-02999-x.

[17] Y. Li, J. Zhao, Z. Lv, and J. Li, “Medical image fusion method by deep learning,” International Journal of Cognitive Computing in Engineering, vol. 2, no. July 2020, pp. 21–29, 2021, doi: 10.1016/j.ijcce.2020.12.004.

The Optimal Route Selection Model of Fresh Agricultural Products Transportation Based on Bee Colony Algorithm

Qingqing Ren
Hainan Vocational University
College of Economics and Management, Haikou, 570100, China

***Abstract*—In order to optimize the distribution route of fresh agricultural products and reduce the distribution cost, the optimal route selection model of fresh agricultural products transportation based on bee colony algorithm is constructed. After establishing the transportation road network model of fresh agricultural products and analyzing the transportation road information, the transportation road network zoning is realized by the dynamic zoning method of transportation road network based on spectral clustering algorithm, taking the unblocked area and congested area as the zoning purpose. Based on the traffic zoning information, the transportation route optimization model of fresh agricultural products with time window is constructed. Solved by bee colony algorithm to obtain the distribution route of fresh agricultural products with the lowest distribution cost and the highest customer satisfaction. The experimental results show that the model can choose the fresh agricultural products distribution route with the lowest distribution cost and the highest customer satisfaction under the two working conditions of smooth traffic and congestion.**

Keywords—Bee colony algorithm; fresh agricultural products; transportation; optimal route selection; road network model; spectral clustering

I. Introduction

Fresh agricultural products refer to primary products such as fresh fruits that can be sold directly on the shelf without further production. At present, fresh agricultural products mainly include fresh vegetables, fruits, flowers, eggs, milk, raw poultry, aquatic products and fresh meat products. These kinds of products occupy a major position in fresh agricultural products, which are traditionally called the three fresh products - fruits and vegetables, meat and aquatic products [1].

Facing the high demand of fresh agricultural products market, the problem to be solved is to reduce the operating cost in the material distribution process as much as possible on the premise of ensuring customer satisfaction. According to the existing research and exploration results, the logistics industry of fresh agricultural products has broad prospects, but the current situation is still relatively lagging behind. The development of the logistics industry related to the distribution of fresh agricultural products is still unbalanced, which needs to be further improved and explored [2]. It is the mission of all kinds of fresh agricultural products logistics parties to do a good job in the transportation and distribution of fresh agricultural products. This requires the promotion and combination of new electronic data tools and logistics industry to further narrow the gap between logistics. Therefore, for fresh agricultural products, we should not only pay attention to the “quality” and “freshness” distribution of fresh agricultural products, but also make full use of information and communication technology and Internet platform, and let the Internet deeply integrate with the sales of fresh food [3]. To solve various problems in the distribution of fresh agricultural products, we need to start with the vehicle route planning between the distribution of fresh agricultural products. At the same time, doing a good job in the route planning between the distribution of fresh agricultural products is also of great significance for solving the obstacles of material sharing and symbiosis between urban and rural areas [4].

The key to material sharing is to do a good job in logistics distribution. More exploration directions in the distribution mode of fresh agricultural products focus on the transportation of fresh agricultural products between urban and rural areas [5]. Joint distribution is a widely adopted logistics distribution method. In brief, it means that the goods distribution orders of multiple companies are distributed and transported by a special logistics company. The logistics company reasonably arranges the distribution time and the transportation path of each order according to the goods demand of different orders and the characteristics of its own means of transportation, so as to achieve the same effect as much as possible [6]. In foreign countries, crowdsourcing platforms are usually established to create relevant information databases to provide basic support for the whole distribution activities. It can exchange information between distribution centers, customers and merchants, so as to maximize distribution revenue and optimize efficiency. In terms of product logistics and distribution, Marcus Bowles studied the application possibility of nanotechnology in the logistics management of the circulation of fresh agricultural products [7]. Omar Ahumada and Rene Villalobos proposed a stochastic tactical planning model for the production and distribution of fresh agricultural products based on the distribution climate and the information of customers changing at any time, which reduced the uncertainty encountered by the fresh agricultural products industry when equation ting growth and distribution plans. As a consumable, the final value of fresh agricultural products depends on the product quality at the time of delivery, including perceived taste, freshness, etc. While the transportation costs, labor and other costs consumed in the whole distribution process affect the final profit and benefit of fresh agricultural products. Therefore, how to reasonably select the distribution mode on the basis of balancing many factors is the top priority, and many factors are interrelated. A good game between various influencing factors is very important for goods suppliers and final consumers to achieve a win-win situation.

Combined with the above background and research foundation, in order to improve the customer satisfaction and transportation cost of fresh agricultural products transportation and distribution, this paper constructs a model of optimal transportation route selection for fresh agricultural products based on bee colony algorithm, and verifies its effectiveness through experiments. The reason for choosing the bee colony algorithm is that it does not need to know the specific special information of the problem to be optimized, and its ability to solve multivariable function optimization problems is excellent. It can choose the fresh agricultural product distribution route with the lowest distribution cost and the highest customer satisfaction, providing a more valuable reference method for the industry to improve service quality and operational efficiency.

II. Related Works

In order to optimize the calculated effect of the path planning scheme in the logistics industry, experts and scholars in the industry have carried out a lot of relevant research. Y. In order to explore the impact of the free trade zone on regional logistics services, Gong et al. selected the adjacent area of a free trade zone in Chongqing as the research object, and used SWOT model to analyze the impact of the free trade zone on international logistics business. Combined with genetic algorithm, an international logistics route planning model is designed. The experimental results show that the model effectively reduces the time consumption of international logistics business [8]. S. O. Ekici et al. found that logistics performance is a key indicator to determine the operating efficiency of the logistics industry. Therefore, they used the enhanced naive Bayesian network and partial least squares method to build a path planning model that considers logistics performance. The experimental results show that the logistics transportation route designed with this model has higher customer satisfaction than that designed with traditional optimization methods [9]. D. T. Xiao's research team believes that the current agricultural product logistics in Heilongjiang Province is relatively inefficient due to the constraints of climate, transportation, supporting facilities and other factors, so it combines convolutional neural network algorithm to build Heilongjiang agricultural product transportation route planning algorithm. After testing, it is found that the transportation cost of the transportation route designed by this model is 15.9% lower than traditional methods [10]. 10. Zhao et al. found that the circulation efficiency of agricultural products in Liaoning Province is insufficient, and some agricultural products have suffered great losses after reaching the transportation destination. Therefore, the author team put forward a series of suggestions on improving transportation efficiency on the basis of understanding the logistics status of local fresh agricultural products [11]. 10. Huang et al. established a real-time and accurate cold chain logistics management system to solve the problems of loose circulation, inaccurate information transmission, untimely processing, and serious loss of fresh agricultural products in the traditional cold chain logistics management mode of agricultural products. In this system, the e-commerce platform is combined with the cold chain logistics network, and the physical real-time emergency management mode of the cold chain of agricultural products is divided into the sub mode of “production warehouse+cold chain”, the sub mode of “e-commerce+cold chain express+smart cabinet” and the sub mode of “cold chain storage+new retail chain”. The test results show that the performance evaluation score of emergency management of fresh agricultural product transport enterprises has been significantly improved after using this system [12].

To sum up, although people have carried out a lot of research to improve the logistics transportation efficiency of all kinds of goods before, most of them only consider one aspect of improving logistics services, such as cost time, customer satisfaction, etc. And few of them consider multiple logistics transportation goals. This research hopes to optimize the cost of agricultural product transportation and customer satisfaction at the same time, which is also the scientific contribution of this research.

III. Optimal Route Selection Model of Fresh Agricultural Products Transportation Based on Bee Colony Algorithm

A. *Transportation Network Model of Fresh Agricultural Products*

According to the characteristics of the road network description of fresh agricultural products transportation, the road network model shall meet the following requirements:

1) Description of geometric characteristics and topological structure of fresh agricultural products transportation network. It includes the location, physical connection, number of lanes, length, width, curvature and intersection of fresh agricultural products transportation roads.

2) Expression of basic modeling unit of fresh agricultural products transportation network. The traffic organization remains unchanged as the principle of road network modeling, that is, the number of lanes in the modeling unit remains unchanged, the lane attributes remain unchanged, and the connectivity relationship between lanes remains unchanged.

3) Description of transportation lanes of fresh agricultural products and topological relationship between lanes. It describes the connectivity of lanes between different basic modeling units, including clarifying the upstream and downstream lanes of each turn, the connectivity of adjacent lanes, etc.

According to the above analysis, the transportation network of fresh agricultural products can be described as a layered network including sub sections, sub section nodes, directed sub sections, lanes and lane connectors. The road network model is divided into two parts: physical layer and logical layer. The graph is used to represent the physical layer Ω_q of the road network model. It is a non-empty set of vertices and edges.

$\Omega_q = (\Gamma, C)$ is the sub segment node set of $\Gamma = \{\Gamma_j\}$. Ω_q representing the endpoint of the sub segment. C is the sub segment set of $C = \{c_j = [\Gamma_p, \Gamma_q]\}$, Γ_p and Γ_q are the two endpoints of the sub segment c_j. Based on the physical layer of the model, the logical layer $\Omega_l = (\Gamma, \partial, Ka, \Omega_{Ka})$ of the model is defined to express the logical connectivity of fresh agricultural products transportation network. Among them, ∂ is the directed sub section, which represents the sub sections with different traffic flows. Based on the directed sub section, the lane elements are described, such as the directed sub section where the lane is located, the number of lanes in the driving direction, lane width, etc.

Based on the above theory, the first layer of the basic road network for the transportation of fresh agricultural products is composed of basic modeling units, including two elements: sub sections and sub section nodes. At the same time, on the basis of sub sections, the sub sections are divided into different directional sub sections according to the traffic flow direction; the second layer refines the lane description based on the first layer, including two elements: carriage lanes and lane connectors for fresh agricultural products. The structure of fresh agricultural products transportation network is shown in Fig. 1.

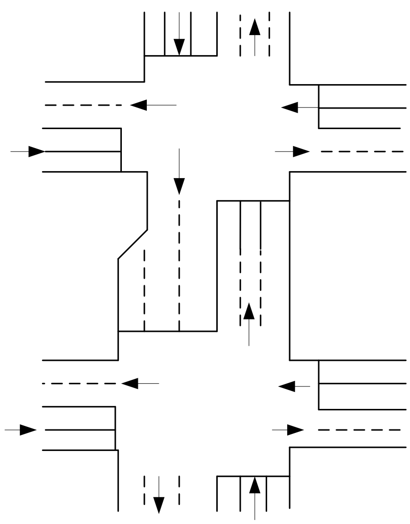


Fig. 1. Schematic diagram of road network.

B. Dynamic Zoning Method of Transportation Network Based on Spectral Clustering Algorithm

In the transportation network model of fresh agricultural products constructed in Section 2.1, the dynamic zoning method of transportation network based on spectral clustering algorithm is used to partition the transportation network of fresh agricultural products. The zoning is mainly to ensure that the transportation of fresh agricultural products can reasonably select the operation path and avoid the traffic congestion area, so as not to increase the transportation cost of fresh agricultural products, but also avoid excessive transportation time of fresh agricultural products and customer dissatisfaction.

Spectral clustering algorithm is a new research hotspot in the field of machine learning in recent years. It is based on the theory of spectral graph. Its essence is to transform the clustering problem into the optimal partition problem of graph, use the similarity relationship between data points to establish a similarity matrix, obtain the multi feature vectors of the matrix, and use them to cluster different data points. Compared with other algorithms, spectral clustering algorithm is simple, easy to implement and not easy to fall into local optimal solution, which is very suitable for many practical problems. The algorithm has been applied to image segmentation, speech recognition, computer vision, text mining and other research fields. Traditional clustering algorithms such as k-means algorithm and fuzzy clustering algorithm do not involve the connection relationship of graph in clustering. Spectral clustering algorithm can realize the segmentation of graph well and cluster a variety of spatial data in this paper. And the zoning research of urban road network needs to consider the connection relationship and distance of traffic flow, signal cycle and actual road network at the same time, so it is appropriate to use spectral clustering method.

Combined with the zoning principle of fresh agricultural products transportation network and the attributes of road intersections, this paper determines the relevant factors for constructing the spectral clustering similarity matrix, and selects the three factors of traffic flow saturation, signal cycle, geographical location of the intersection and the road connection relationship matrix for road network zoning. The first two indicators reflect the actual traffic flow attributes of the intersection, corresponding to the traffic flow saturation principle and the signal cycle principle, and the latter index reflects the topology attribute of the intersection, corresponding to the distance principle.

The traffic flow saturation BH of fresh agricultural products transportation network can be calculated by equation.

$$BH = \frac{CL}{TX} \tag{1}$$

Where, CL and TX are the traffic flow on the fresh agricultural products transportation road respectively. The actual traffic capacity, signal cycle and geographical location of the intersection can be obtained through actual investigation.

In this paper, the NJW algorithm in the k-way division criterion of spectral clustering algorithm is used for dynamic zoning of road network. According to the specific analysis process of NJW multi-channel spectral clustering algorithm and combined with the relevant factors of road network zoning, the specific steps of road network zoning based on spectral clustering algorithm are as follows:

1) Combined with Fig. 1, calibrate the topological structure of fresh agricultural products transportation network and the number of each intersection and road. JX and DL are the intersection and road in the study area.

2) Calculate the similarity matrix Bc. Standardize the original data of the three factors of traffic flow saturation, signal period and intersection geographical location of the road network model shown in Fig. 1, then calculate the similarity between nodes, and calculate the intersection similarity matrix Bc in combination with the road connection relationship matrix.

$$Bc_{ij} = \begin{cases} \exp\left(-\frac{\left\| ys_i - ys_j \right\|}{2\rho^2} \right) & i \neq j \\ 0 & i = j \end{cases} \quad (2)$$

Where, ρ is a parameter selected by people in advance. The larger its value is, the more significant the classification effect is. ys_i and ys_j represent the traffic factors of the i th and j th intersections respectively.

3) Calculate the Laplace matrix lp covering the transportation and traffic information of fresh agricultural products.

$$lp = JH^{-\frac{1}{2}} \left(lp \cdot JH \right)^{-\frac{1}{2}} \quad (3)$$

Where JH is the diagonalization matrix obtained from matrix Bc.

4) Calculate the first maximum eigenvalues tz_n and the corresponding traffic information eigenvector χ_n of the Laplace matrix lp. The transportation information of fresh agricultural products can be obtained, and the corresponding eigenvectors are arranged according to the order of eigenvalues to form a matrix υ.

5) Transform the row vector of the characteristic vector matrix υ of fresh agricultural product traffic information into the unit vector to obtain the matrix Ψ.

6) Regard each row of matrix Ψ as a data point in three-dimensional space, and K clusters are obtained by K-means algorithm.

7) Through step 6, the classification results of the best intersection are obtained, the classification results are visually displayed, and appropriate zoning adjustments are made according to the actual situation.

According to the above steps, the zoning results combined with the road network flow can be obtained. When the zoning results cannot meet the traffic flow performance of each road network area after zoning, the road network zoning needs to be adjusted according to the real-time data of traffic flow.

C. Establishment of the Optimal Route Selection Model for The transportation of Fresh Agricultural Products

Combined with the transportation traffic information of fresh agricultural products, the optimal route selection model of fresh agricultural products transportation is constructed, and the model is solved by bee colony algorithm to obtain the optimal route of fresh agricultural products transportation.

D. Description of Transportation Problems of Fresh Agricultural Products

The vehicle routing optimization problem of fresh agricultural products studied in this paper is to consider a distribution center. The distribution center uses multiple transportation vehicles for collaborative distribution. All vehicles start from the distribution center and distribute in the area composed of multi customer distribution nodes. The location of each distribution node is determined, the demand of each distribution node is known, and the distribution service to all customer distribution nodes is completed within the time window required by each customer. All vehicles eventually return to the distribution center.

A single distribution center is set to deliver goods to the k th customer (or distribution node). The freight volume of the j th distribution traffic route is Υ, the time window is $\left[FH_j, KH_j \right]$. The unloading time is αt_j, the delay cost per hour is s_j, and the average speed and shortest distance between customers and between distribution centers and customers are u_{ji} and e_{ji} respectively; n trucks can be used to transport goods, including m_q trucks in group q, with a loading capacity u_q, a driving cost g_q of per vehicle per kilometer, a waiting cost s of per hour, overtime allowance dr and baggage allowance r of per hour respectively; The vehicle must return to the original distribution center on the same day, and the demand point of p vehicle distribution of group q is expressed as m_{qp}.

IV. Construction of Objective Function

Compared with foreign countries, the domestic cold chain transportation of fresh agricultural products is still in the primary stage, but it has also achieved some results in the development of relevant distribution system and technology (Huang et al. 2020). Through the above analysis, it can be seen that distribution is an important part of the cold chain logistics system. Optimizing the distribution route helps to improve the efficiency of cold chain logistics. Due to the lack of scientific and reasonable planning for the distribution route of fresh agricultural products, it is difficult to meet the daily needs of customers in time. Considering the characteristics of distribution cost, distribution time and fresh agricultural products, an overall optimization model of distribution route with time window is established to minimize the distribution cost.

A. Customer Satisfaction

The delivery target customers of fresh agricultural products are required to receive their products within the specified time period. If they exceed the time period or deliver them in advance, they may be rejected or accept certain punishment measures as compensation. Therefore, only by timely delivering fresh agricultural products to customers according to the time window specified in the order can we meet the requirements of customers, improve the satisfaction with the distribution service of fresh agricultural products, and the online sales of fresh agricultural products can be more accepted by the market [13].

If t_j represents the arrival time of the means of transport and j represents the logistics node number, then $FH_j < t_j < KH_j$. FH_j and KH_j are the earliest and the latest starting time of the distribution service respectively. In the actual distribution process of fresh agricultural products, if the fresh agricultural products are delivered just at the distribution time point required by the customer and look at the time window immediately, the customer satisfaction will reach the maximum, that is, the optimal solution is realized; If the fresh agricultural products are not stuck when they are delivered, and the time is not at the time point required by the customer, that is, the goods may be delivered in advance or the goods may be in a timeout state when they are delivered, namely, the time window can be accepted. At this time, the customer's satisfaction with the service quality of distribution cannot be met [14].

Customer satisfaction can be expressed as a function of fuzzy reservation time. For customers, when the service start time is t_j, the customer satisfaction function can be expressed as:

$$g_j\left(t_j\right)=\begin{cases} 0 & t_j \le FH_j \\ \left[0.7\left(t_j - FH_j \Upsilon\right)/\left(FH_j^e - FH_j \Upsilon\right)\right]+0.3 & FH_j \le t_j \le KH_j \\ 1 & FH_j^e < t_j < KH_j^e \\ \left[0.7\left(KH_j \Upsilon - t_j\right)/\left(KH_j \Upsilon - KH_j^e\right)\right]+0.3 & KH_j^e \le KH_j^e \\ 0 & t_j > KH_j \end{cases} \tag{4}$$

Where, FH_j^e and KH_j^e are the earliest and the latest time window of the reservation in turn.

When driving to position i after completing the transportation task j of fresh agricultural products, the waiting time of the distribution vehicle at position i can be expressed as:

$$\varpi_j\left(t_j\right)=t_i - t_j - t_{ji}\alpha t_j \tag{5}$$

Where, t_{ji} represents the travel time of the distribution vehicle from customer j to customer i; $\varpi_j\left(t_j\right)$ represents the waiting time of the distribution vehicle at the customer i when the starting time is t_i; t_i and t_j are the time taken for the delivery vehicles to arrive at the customer i and j destinations respectively.

B. Loss of Fresh Agricultural Products

Different from other products, fresh agricultural products are perishable, and the fundamental value of fresh agricultural products lies in whether they reach the freshness and quality satisfactory to customers. Even if the current means and methods of transportation can improve the temperature in the transportation process and achieve constant temperature distribution, the final value of fresh agricultural products will continue to decrease with the extension of distribution time in the overall distribution process because the degree of decay of fresh agricultural products is closely related to the length of distribution time; as for the loss degree of fresh agricultural products, generally speaking, it consists of three stages: high value (freshness is the highest after picking) to acceptable value (loss of certain freshness after a period of time, but customers can finally accept) to invalid value (complete decay and damage, which cannot be accepted or eaten by customers). On the whole, fresh agricultural products cannot be delivered when they are obtained, placed an order, and then delivered to customers. More often, fresh agricultural products will be delivered within the stage of acceptable value, resulting in the loss of fresh agricultural products [15].

On the one hand, fresh agricultural products are lost due to the increasing time from the distribution center to the destination, that is, the distribution time has lost fresh agricultural products. The relevant variables are: when the value of vehicle state y_{hk}^v is 1, it means that the vehicle k of distribution center v serves customers h; when the value of y_{hk}^v is 0, which means that the vehicle k does not provide customer h service. The loss cost at this time is:

$$D_1=\sum_{k=1}^{k}\sum_{h=1}^{m+v}\sum_{v=m+1}^{v}\Upsilon f k_h y_{hk}^v \beta_2 t_h^k - \Upsilon f k_h y_{hk}^v \beta_2 t_h^k t_{h-1}^k \tag{6}$$

Where, f represents the unit price of fresh agricultural products; k_h represents the quantity of fresh agricultural products ordered by the customer h; β_2 represents the loss rate of fresh agricultural products during transportation; t_h^k represents the time when the vehicle k arrives at the customer h; t_{h-1}^k represents the time from vehicle k to customer $h-1$.

On the other hand, fresh agricultural products will be unloaded when they arrive at different distribution points in the distribution process. At this time, the opening of the transportation vehicle for unloading will produce internal and external air convection and change the internal air temperature, so as to accelerate the corruption speed of fresh agricultural products and the loss of fresh food in the loading / unloading process. The relevant variables are: when the value of y_{hk}^v is 1, it means that the vehicle k of distribution center v has delivered to point h; if the value of y_{hk}^v is 0, it means that the vehicle k has not delivered to point h. The loss cost at this time is:

$$D_2=\sum_{k=1}^{k}\sum_{h=1}^{m+v}\sum_{v=m+1}^{v} f \Upsilon k_h y_{hk}^v \beta_2\left(O_h^k\right) \tag{7}$$

Where O_h^k is the loading and unloading time of vehicle k at distribution point h.

Then the loss cost D of fresh agricultural products in the process of distribution is the sum of equations (6) and (7).

C. General Model Construction

To sum up, in order to build the model, when the values of the vehicle state judgment variables a_{jiqp}, a_{jqp} and a_{jiqp} are 1 during introducing the transportation of fresh agricultural products. It means that the vehicle k passes through the distribution point (j,i), and when the value of a_{jiqp} is 0, it means that the vehicle k does not pass through the distribution point (j,i); when the value of a_{jqp} is 1, it means that vehicle k serves customer j; when the value of a_{jqp} is 0, it means that vehicle k does not serve customer j.

The following models are established:

$$MinW = \sum_{j=0}^{k}\sum_{i=0}^{k}\sum_{q=1}^{n}\sum_{p=1}^{m_q} e_{ji} f_q a_{jiqp} + \sum_{q=1}^{n}\sum_{p=1}^{m_q}\left(\varpi o_{qp} * r + d\varpi o_{qp} * r\right)$$
$$+ s_j * s\sum_{j=1}^{k}\max\left(t_j - \alpha t_j, 0\right)\left(\alpha t_j - t_j, 0\right) + \Upsilon DBH \tag{8}$$

The constraints are:

$$\frac{\Upsilon}{k}\sum_{j=1}^{k} g_j\left(t_j\right) \geq 80\% \tag{9}$$

$$\sum_{j=1}^{k} f_j a_{jqp}\Upsilon \leq u_q \tag{10}$$

$$\sum_{q=1}^{k}\sum_{p=1}^{m_q} a_{jqp}\Upsilon = 1 \tag{11}$$

$$\sum_{j=1}^{k} a_{jiqp} = a_{jqp}, i = 1, 2, ..., \mathrm{k} \tag{12}$$

Where, equation (8) is the objective function, which means that the distribution cost of fresh agricultural products is the smallest. Equation (9) is the constraint of customer satisfaction, which means that the average value of customer satisfaction should be more than 80%. Equation (10) is the vehicle loading capacity constraint. Equation (11) is used to ensure that customers only complete the distribution by the p th vehicle of group q. Equation (12) represents the only constraint on the vehicle arriving at a customer, that is, each customer has and only one vehicle to serve him. ϖo_{qp} represents the driving time of the p th vehicle in group q, and $d\varpi o_{qp}$ represents the overtime time of the p th vehicle in group q. e_{ji} represents the shortest distance between transportation logistics nodes of fresh agricultural products. s is the hourly waiting cost of the vehicle, and s_j is the delay cost of the j th hour.

D. Solving the Optimal Transportation Path of Fresh Agricultural Products based on Artificial Bee Colony Algorithm

Artificial bee colony algorithm is a new intelligent optimization algorithm proposed by Karaboga in 2005. In the artificial bee colony algorithm, the optimal food source is obtained through the cooperation and role transformation of leading bee, following bee and reconnaissance bee. The location of the food source corresponds to the possible solution of the optimal route for the transportation of fresh agricultural products. The income of the bee colony in the food source represents the fitness of the optimized problem [16].

At the beginning of the algorithm, the initial population M with solutions will be generated randomly, and each solution Y_j is a B -dimensional vector. Then, lead the bees to remember the optimal solution and search in the neighborhood of the food source. After initialization, three kinds of bees search circularly. The search equation is as follows:

$$u_{ji} = y_{ji} + \delta_{ji} y_{ji} - \delta_{ji} y_{\mu i} \tag{13}$$

Where, u_{ji} is the candidate food source and represents the candidate solution of the optimal transportation path of fresh agricultural products; $\mu \in \{1, 2, ..., M\}$, $i \in \{1, 2, ..., B\}$ are generated randomly and unequal. δ_{ji} is a random number uniformly distributed on [-1,1]. The leading bee adopts the greedy criterion to compare the search solution of the optimal route of fresh agricultural products transportation with the previous optimal solution. If the search solution of the optimal path is superior to the previous optimal solution, replace it; on the contrary, it remains unchanged. The probability of following bees to choose food source is:

$$\sigma_j = \xi_j \delta_{ji} / \sum_{j=1}^{M} \xi_j \delta_{ji} \tag{14}$$

Where, ξ_j is the fitness of the j th solution. If a food source remains unchanged after circulation, the food source will be abandoned, and the corresponding leading bee will become a reconnaissance bee, and the solution will be updated through the following equation:

$$\eta_j^i = \eta_{\min}^i + rand\left(0,1\right)\left(\eta_{\max}^i - \eta_{\min}^i\right)\sigma_j \tag{15}$$

Where, η_j^i is the i -th dimensional component of the new food source, which represents the component information of the new path during the transportation of fresh agricultural products; $\eta_{\max}^i$ and $\eta_{\min}^i$ are the maximum and minimum values of the i -th dimensional component respectively.

The classical artificial bee colony algorithm adopts the real number coding method for the transportation path coding of fresh agricultural products, which is obviously not feasible in the distribution route optimization of fresh agricultural products. Distribution center demand is decentralized, so the coding method needs to be reconsidered. In this study, if the natural number coding method is adopted for the demand point, a feasible transportation path of fresh agricultural products can be expressed as $\left(0, \vartheta_{11}, \vartheta_{12}, ..., \vartheta_{1m}, \vartheta_{21}, \vartheta_{22}, ..., \vartheta_{2v}; \vartheta_{n1}, \vartheta_{n2}, ..., \vartheta_{nu}\right)$. The transportation route of this fresh agricultural products means that the first vehicle starts from the distribution center and returns to the distribution center after reaching the demand point $\vartheta_{11}, \vartheta_{12}, ..., \vartheta_{1m}$; The second vehicle starts from the distribution center and returns to the distribution center after reaching the demand point $\vartheta_{21}, \vartheta_{22}, ..., \vartheta_{2v}$. The m-th vehicle starts from the distribution center and returns to the distribution center after reaching the demand point $\vartheta_{n1}, \vartheta_{n2}, ..., \vartheta_{nu}$. If there are 3 vehicles and 9 demand points and the food source is $y = 023601789045$, it means that the first vehicle starts from the distribution center and returns to the distribution center after reaching demand points 2, 3 and 6. The second vehicle starts from the distribution center and returns to the distribution center after reaching demand points 1, 7, 8 and 9, and the third vehicle starts from the distribution center and returns to the distribution center after reaching demand points 4 and 5.

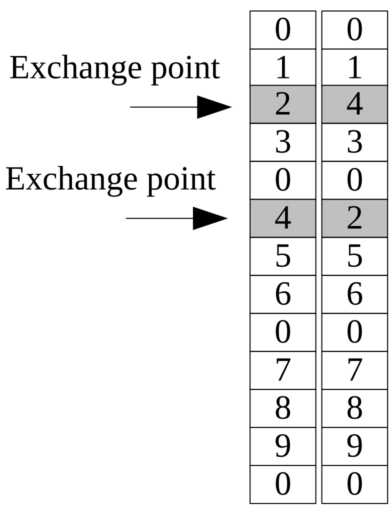


Fig. 2. Food sources before and after exchange.

Since a new food source coding method is adopted in the artificial bee colony, the update of the candidate food source location (transportation path of fresh agricultural products) cannot adopt the mode of equation (15). In this study, by exchanging two neighborhood points in the food source (fresh agricultural product transportation path) are randomly exchanged to obtain the candidate food source. It is illustrated by 9 demand points and 3 vehicles. Fig. 2 shows the food sources before and after the exchange. It can be seen that the candidate food sources can be obtained by exchanging the third and sixth points. The change of food source before and after exchange is small, so many excellent characteristics of food source before change can be maintained; At the same time, the random location exchange increases the diversity of food source selection and avoids falling into local optimization and failing to obtain the global optimal solution of the transportation path of fresh agricultural products [17].

The fitness function is obtained by transforming the objective function of the established transportation route selection model of fresh agricultural products. Since the objective function pursues the minimization of cost and the maximization of customer satisfaction, its reciprocal is taken and transformed into the fitness function of artificial bee colony algorithm:

$$\xi = 1/\min\mathrm{W} = 1/\left[MinW = \sum_{j=0}^{k}\sum_{i=0}^{k}\sum_{q=1}^{n}\sum_{p=1}^{m_q} e_{ji} f_q a_{jiqp} + \sum_{q=1}^{n}\sum_{p=1}^{m_q}\left(\varpi o_{qp} * r + \cdot d\varpi o_{qp} * r\right) + s_j * s \sum_{j=1}^{k} \max\left(t_j - \alpha t_j, 0\right)\left(\alpha t_j - t_j, 0\right) + DBH \right] \quad (16)$$

The specific steps of using artificial bee colony algorithm to solve the optimization of cold chain logistics distribution route of fresh agricultural products are as follows:

1) Generate a certain number of food sources, that is, the initial solution of the transportation path of fresh agricultural products. The initial solution is allocated to each leading bee, and the fitness of the transportation path of each fresh agricultural product is calculated;

2) Set the maximum number of iterations, cycles and the cycles of each solution;

3) Leading bees search the neighborhood of the transportation path of fresh agricultural products, generate new solutions of the same scale, and compare the fitness of the old and new solutions. If the fitness of the optimized new solution for the transportation path of fresh agricultural products is greater than that of the old solution, the new solution will replace the old solution, otherwise the old solution of the transportation path of fresh agricultural products will remain unchanged [18].

4) Following bees choose the transportation path of fresh agricultural products with probability σ_i, and conduct the neighborhood search to generate a new solution and calculate its fitness. Select the maximum fitness value of the new solution and compare it with the fitness value of the old solution. If the fitness of the new solution is greater than that of the old solution, the new solution will replace the old solution, otherwise the old solution will remain unchanged [19].

5) Judge whether the number of iterations reaches the maximum value [20-21]. If it is the maximum value, the optimal solution of the transportation path of fresh agricultural products will be output, that is, the distribution route of fresh agricultural products with the lowest distribution cost and the highest customer satisfaction will be obtained. If not, skip to step (3) and continue the iteration [22]. So far, the transportation route optimization model for fresh agricultural products has been built, and the steps for calculating the transportation route are shown in Fig. 3.

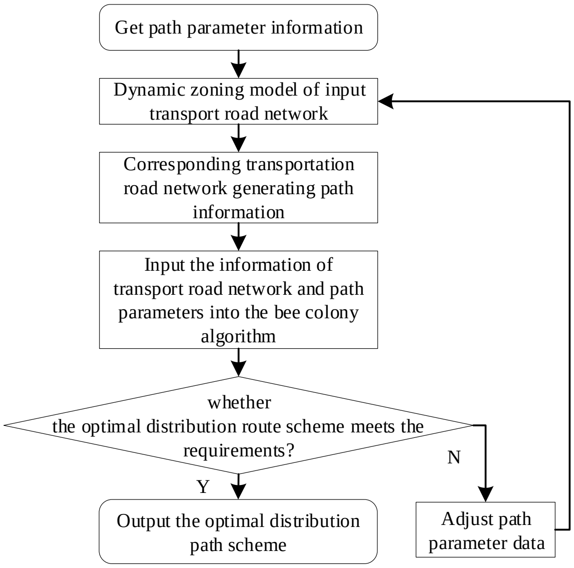


Fig. 3. Flow chart of transportation route optimization model for fresh agricultural products.

V. Results and Analysis

In order to test the effectiveness of the model established above, aiming at the model, this paper will take 18 order information about the distribution process of fresh agricultural products in a city's fresh agricultural products distribution center on a certain day as the research object for analysis and comparison [23-24]. The total transportation cost is 644.1 yuan [25]. Get the optimal transportation route scheme by inputting the data into the program written in Python language.

The basic information of each distribution node is shown in Table I.

TABLE I. Coordinate Nodes of Distribution Points

Distribution node code	Abscissa/m	Ordinate /m
1	19	16
2	17	10
3	21	11
4	20	14
5	26	15
6	23	11
7	26	18
8	16	16
9	17	13
10	15	19
11	18	18
12	24	14
13	20	19
14	23	13
15	12	12

Relevant parameter settings are shown in Table II.

TABLE II. Related Parameter Settings

Parameter name	Parameter value
Vehicle transportation cost (yuan / vehicle / time)	2.2
Vehicle running speed (km / h)	70
Normal salary of driver (yuan / hour)	17
Overtime pays for drivers (yuan / hour)	30
Vehicle waiting cost (yuan / hour)	20
Customer delay fee (yuan / hour)	110

Under the condition of smooth traffic, the path results optimized by the model in this paper are shown in Fig. 3.

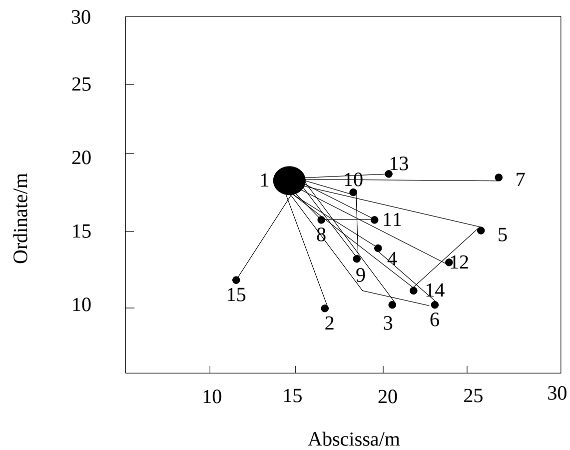


Fig. 4. Path results of 10 vehicles.

The calculation results show that under the use of 10 vehicles, the optimal feasible solution obtained after operation is 1-10-9-1-7-1-8-11-1-12-1-13-1-2-1-14-5-1-15-1-4-6-1-3-1. The transportation time of each vehicle is relatively average, which does not cause excessive use of some vehicles, and the transportation time of each vehicle is not long, which is basically about 12 hours, which can maintain the freshness of fresh agricultural products. From the above results, it can be seen that the model constructed in this paper can effectively reduce the distribution cost and meet the customer satisfaction needs. Table III shows the specific results.

TABLE III. Route Delivery Results of 10 Vehicles

Vehicle code	Route	Journey time/h	Delay waiting/min	Transportation cost / yuan
1	1-10-9-1	12	0	33
2	1-7-1	12	0	23
3	1-8-11-1	11	0	34
4	1-12-1	11	0	35
5	1-13-1	9	0	44
6	1-2-1	10	0	32
7	1-14-5-1	12	0	21
8	1-15-1	11	0	54
9	1-4-6-1	12	0	22
10	1-3-1	12	0	11

As shown in Table III, after the model in this paper is used and fresh agricultural products are transported, the delayed waiting time of the distribution target is 0, and the total transportation cost is 309 yuan, saving 335.1 yuan.

In the case of traffic congestion, after the model in this paper is used, the time-sharing diagram of path optimization is shown in Fig. 4. Among them, the straight line represents the current solution and the wavy line represents the optimal solution. It can be concluded from Fig. 4 that the distribution cost is optimized from 644.1 yuan to about 350 yuan.

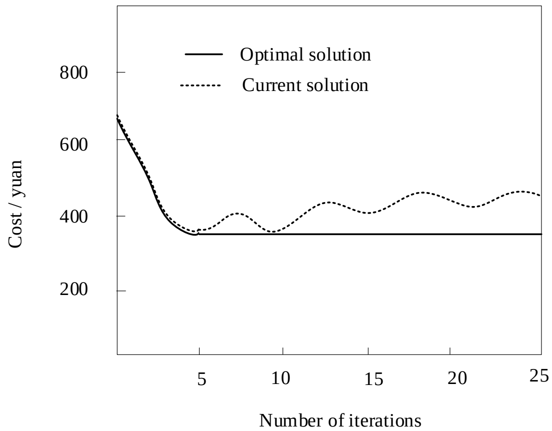


Fig. 5. Time sharing diagram of route optimization in this model under traffic congestion conditions.

In the two working conditions, the test results of customer satisfaction rate are shown in Fig. 5.

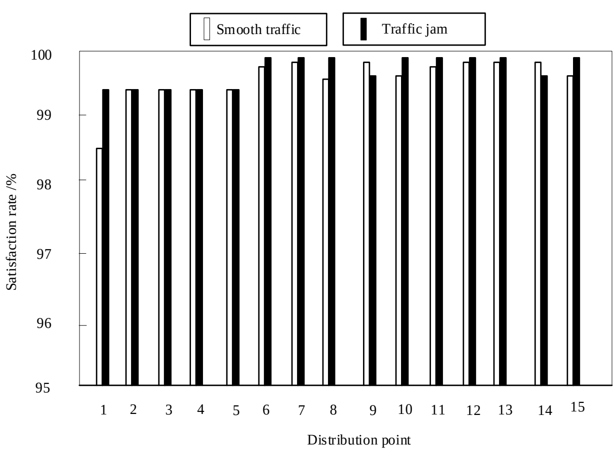


Fig. 6. Test results of customer satisfaction rate under two working conditions.

As shown in Fig. 6, in the two working conditions of smooth traffic and traffic congestion, after the use of the model in this paper, the customer satisfaction is greater than 95%, and the satisfaction is high.

VI. Discussion

Combined with the research content of this paper, the suggestions to improve the matching effect of fresh agricultural products are as follows:

A. *Accelerate the Construction of Agricultural Products Logistics Infrastructure*

The rapid development of agricultural products logistics is inseparable from infrastructure construction. All regions make reasonable planning and purposeful construction according to the actual situation. The following suggestions are mainly put forward:

1) Cold storage construction: The temperature in the cold storage is mainly adjusted according to the different types of agricultural products. The construction of cold storage in cities is closely related to the distribution of agricultural products. The characteristics of fresh agricultural products determine that the cold storage in cities must reach a certain scale in order to ensure that agricultural products are not damaged. Only by investigating the output, types and distribution of urban agricultural products, planning the layout and optimizing the location can we promote the development of agricultural products logistics.

2) Construction of distribution center: The main purpose of distribution center is to distribute to customers at various points, which plays a vital role in the impact of distribution. The distribution cost is directly related to the number of urban distribution centers. There are few distribution centers, resulting in high distribution cost and long distribution distance, and the decay rate of fresh agricultural products is high. Reasonable planning and uniform distribution of distribution centers is an effective measure to effectively reduce the loss in distribution.

3) Transformation and upgrading of cold chain transport vehicle: The cold chain transport vehicle is mainly used for the distribution of agricultural products. Whether the cold chain equipment on the vehicle is advanced or not will also affect the cost. At present, many cold chains transport vehicles in China are refitted through ordinary vehicles. The technical level is relatively backward and the distribution cost is high, which cannot be compared with developed countries. Therefore, it is necessary to introduce advanced cold chain transportation technology from developed countries and upgrade cold chain transportation vehicles.

B. *Expand the Scope of Green Channel*

At present, the transport channel of fresh agricultural products has been expanded to all parts of the country, and the varieties that meet the requirements are given free passage. However, in order to increase taxes, some local governments have not strictly implemented the government documents, which has caused some obstacles to the circulation of fresh agricultural products. In view of the green channel of agricultural products, we should still strive to do a good job in the following aspects.

1) We will continue to implement the policy of exemption and reduction for fresh agricultural products that meet the requirements, and expand this scope to every region and every toll road.

2) Due to the large number of social vehicles, when the agricultural product transport vehicle passes through the toll station, it needs to stop and accept the inspection of the security personnel. It can be released only if there is no problem, resulting in longer transportation time and higher loss rate of agricultural products. Therefore, it is necessary to build a set of security inspection system, which can automatically check whether the products loaded in the vehicle meet the transportation requirements of agricultural products, shorten the vehicle transportation time and effectively ensure the quality of agricultural products.

3) Individuals take advantage of the preferential policy of agricultural product transportation and borrow the green channel to cause some obstacles to the implementation of security inspection personnel. It needs to build a set of credit system and integrate such people into the integrity system, so as to ensure the smooth transportation of agricultural products.

C. Make Full use of Information Technology to Build a Network Information Platform for Cold Chain Distribution

The particularity of fresh agricultural products determines the importance of its distribution link. The optimization of the distribution route of cold chain logistics for fresh agricultural products is to effectively avoid the deterioration and decay of agricultural products and reduce economic losses. Logistics distribution is an important part of the cold chain system, which has high requirements for the timeliness and fresh-keeping function of distribution. In the information age, in order to optimize the cold chain logistics transportation system, we need to make full use of advanced information technology, effectively connect all links of logistics distribution, monitor the status of logistics distribution in real time, and realize the intelligent management of logistics distribution. Through the construction of corresponding network information platform, the collection, sorting and feedback of logistics information are managed in an integrated way, so as to grasp the transport status of fresh agricultural products in real time and improve the safety and freshness of fresh agricultural products. In the implementation of distribution management, big data and other technologies are used to automatically manage logistics transportation and improve the level of distribution management. The one-stop cold chain logistics distribution includes the whole process from production to sales of agricultural products. The intelligent management of the whole service process through the application of information technology can avoid the disadvantages of traditional logistics distribution and meet the optimization requirements of the cold chain logistics distribution route of fresh agricultural products.

VII. CONCLUSION

In recent years, with the continuous development of logistics industry and the increasing variety and output of fresh agricultural products, higher requirements are put forward for cold chain logistics distribution. In the cold chain logistics system, distribution is a very key link, which has high requirements for transportation technology and transportation environment. However, compared with developed countries, China’s cold chain logistics started late, the technology is not mature, the cost of fresh produce distribution is still high, and the corresponding preservation equipment is not perfect. Once the food goes bad or rots, it will seriously affect the surrounding environment. People’s income is increasing, and the demand and quality requirements for fresh agricultural products are increasing. In order to better meet people’s needs, we must take effective strategies to optimize the distribution route of cold chain logistics of fresh agricultural products. When optimizing the distribution route, we can make full use of information technology to build a network information platform for cold chain distribution; strengthen the implementation of relevant policies to provide a good environment for the development of cold chain logistics; optimize the supply chain of fresh agricultural products and build a large distribution center; strengthen alliances of fresh produce-related organizations to promote the long-term stable development of China's cold chain logistics industry. In this context, this paper constructs the optimal route selection model of fresh agricultural products transportation based on bee colony algorithm, and verifies its application value in experiments. However, due to the limitation of personal energy, this study failed to select more data to verify the model performance of the design, which is also the key point to be noted in the follow-up study.

ACKNOWLEDGMENT

The research is supported by: the Education Department of Hainan Province’ Research on the development and application mode of business data analysis course resources in higher Vocational colleges under the background of digital era, project number: Hnjg 2021-124.

REFERENCE

[1] K. Kaewfak, V. Ammarapala, V. N. Huynh, “Multi-objective Optimization of Freight Route Choices in Multimodal Transportation”. *International Journal of Computational Intelligence Systems*, vol.14(1), pp. 794-807, 2021.

[2] A. H. Sodhro, M. S. Obaidat, & Q. H. Abbasi, “Quality of Service Optimization in an IoT-Driven Intelligent Transportation System”. *IEEE Wireless Communications*, vol. 26(6), pp. 10-17, 2019.

[3] L. Hu, W. Naeem, & E. Rajabally, “A Multiobjective Optimization Approach for COLREGs-Compliant Route planning of Autonomous Surface Vehicles Verified on Networked Bridge Simulators”. IEEE Transactions on Intelligent Transportation Systems, vol. 21(3), pp. 1167-1179, 2020.

[4] M. Nowak, C. Archetti & B. Kamiński, “Preface: Special issue on the future of route optimization/vehicle routing”. *Networks*, vol. 73(4), pp. 379-381, 2019.

[5] K. Aliev, E. Traini & M. Asranov, P”rediction and estimation model of energy demand of the AMR with cobot for the designed path in automated logistics systems, *Procedia CIRP*, 2021, vol. 99(13), pp. 116-121, 2021.

[6] S. Yang. “Optimization of Urban Logistics Distribution route under dynamic Traffic Network”, *International Core Journal of Engineering*, vol. 6(1), pp. 243-248, 2020.

[7] Q. Yu, Y. Wang, X. Jiang, B. Zhao, X. Zhang & X. Wang et al. (2021). “Optimization of vehicle transportation route based on iot”. *Mathematical Problems in Engineering*, vol. 2021(41), pp. 1312058.1-1312058.10.

[8] Y. Gong, J. Li. “Research on the Path of International Logistics Development in Chongqing Free Trade Zone,” *Modern Economy*, vol. 11(5), pp. 1053-1063, 2020.

[9] S. O. Ekici, O. Kabak & F. Ulengin. “Improving logistics performance by reforming the pillars of Global Competitiveness Index,” *Transport policy*, vol. 81, pp. 197-207, 2019.

[10] D. T. Xiao. “Research on the development strategy of Heilongjiang agricultural products logistics from the perspective of supply chain,” *Paper Asia*, vol. 35, pp. 40-44, 2019.

[11] X. Zhao. “Research on Supply and Demand of Cold Chain Logistics of Fresh Agricultural Products in Liaoning Province,”. *American Journal of Industrial and Business Management*, vol. 10, pp. 723-733, 2020.

[12] X. Huang, R. Xie & L. Huang. “Real-time emergency management mode of cold chain logistics for agricultural products under the background of “Internet+”, *Journal of Intelligent and Fuzzy Systems*, vol. 38, pp. 1-13, 2020.

[13] Y. Ji. “Decision Optimization for Cold Chain Logistics of Fresh Agricultural Products under the Perspective of Cost-Benefit,” *Open Access Library Journal*, vol. 06, pp. 1-17, 2019.

[14] H. H. Park, S. J. Cho. “An Analysis of the Effect of Logistics Efficiency on the Export of Korean Agricultural Products to New Southern Countries,” *Journal of Korea Trade*, vol 25, pp. 169-183, 2021.

[15] N. Nosan. “The Essence and Role of Logistics Management at Agricultural Enterprises of Ukraine,” *Modern Economics*, vol. 22, pp. 72-76, 2020.

[16] L. M. Taraniuk, J. Wang & H. Qiu. “International aspects of assessing the logistic potential of the main countries of agricultural production,”. *Problems and Perspectives in Management*, vol. 17, pp. 31-44, 2019.

[17] C. Wang. “Research on Optimal Design of Short-Life Cycle Product Logistics Supply Chain Based on Multicriteria Decision Model,” *Security and Communication Networks*, vol. 4, pp. 1-12, 2021.

[18] Z. H. Yang, J. P. Shi. “E-Commerce Fresh Agricultural Products Cold Chain Distribution Time-Consuming Control Algorithm,” *Computer Simulation*, vol. 38, pp. 447-451, 2021.

[19] J. Su. “Coastline climate environment and e-commerce logistics system development: a GIS perspective,” *Arabian Journal of Geosciences*, vol. 14, pp. 1-17, 2021.

[20] S. Hoover, J. D. Porter, & C. Fuentes. “Strategic route planning to manage transit's susceptibility to disease transmission”. *Transportation Research Record*, vol. 2675(8), pp. 369-381, 2021.

[21] T. Y. Hu, Y. C. Hsu, & T. Y. Liao, “Comparisons of a multi-objective compromise weight model and a multi-objective evolutionary algorithm in hazmat transportation route planning” *Transportation Research Record*, vol. 2675(1), pp. 161-171, 2021.

[22] R. W. Calvo, L. Letocart & R. Baldacci, “Preface: emerging challenges in transportation planning” *Networks*, vol. 72(3), pp.309-310. 2018

[23] S. Zhang, M. Chen, & W. Zhang, (2019). “A novel location-routing problem in electric vehicle transportation with stochastic demands” *Journal of Cleaner Production*, vol. 221(JUN.1), pp. 567-581, 2019.

[24] X. Liu, J. Ma, D. Chen, & L. Y. Zhang, “Real-time unmanned aerial vehicle cruise route optimization for road segment surveillance using decomposition algorithm” *Robotica*, vol. 39(6), pp. 1007-1022, 2021.

[25] H. King & M. Wachs, (2020). “Centuries of ballot-box transportation planning in los angeles” *Transportation Research Record*, vol. 2674(12), pp. 155-164, 2020.

An Enhanced Face Detection System using A Novel FIS-CDNN Classifier

Santhosh S.[1], Dr. S. V. Rajashekararadhya[2]

Research Scholar, Department of Electronics and Communication Engineering, Kalpataru Institute of Technology, Tiptur, Karnataka - 572201, India[1]

Professor and Former Principal, Department of Electronics and Communication Engineering, Kalpataru Institute of Technology, Tiptur, Karnataka - 572201, India[2]

***Abstract*—A computer application that can detect, track, identify or verify human faces as of an image or video capture utilizing a digital camera is Face Recognition (FR). Few challenges like low-resolution images, aging, uncontrolled pose, illumination changes, and poor lighting conditions are not tackled even though huge advancement has been created in the Face Detection and Recognition (FDR) domain. Utilizing the Modified Tiny Face Detection (MTFD) and Fuzzy Interference System - Convolutional Deep Neural Network (FIS-CDNN) classifier, a Face Recognition System (FRS) was proposed here to tackle all complications. Primarily, Gamma correction - Based Histogram Equalization (GBHE) technique is utilized to augment the image's input in the pre-processing phase. The MTFD was employed to detect the face. Following that, the features are extracted. The Improved Chehra (IC) landmark extraction method was employed to retrieve the landmark features. And finally, the Geometric Features (GFs) are extracted. Later, the Gaussian - centered Spider Monkey Optimization (GSMO) Algorithm was employed to choose the vital features. To recognizing the face, the chosen features are fed into the FIS-CDNN classifier. When analogized to the prevailing models, it is concluded via the experiential outcomes that higher accuracy was attained by the proposed method.**

***Keywords*—Gamma correction - based histogram equalization (GBHE) technique; modified tiny face detection method (MTFD); Improved Chehra (IC) landmark extraction method; Gaussian-based Spider Monkey Optimization (GSMO) Algorithm; fuzzy interference system-convolutional deep neural network (FIS-CDNN) classifier**

I. INTRODUCTION

Since the face contains numerous important details, it is a vital part of the human body for recognizing each individual [1]. This is what led to Human FR being a significant subject in the artificial intelligence field. Owing to their respective pixel data computation, the query face image is harmonized with the template face image in the facial database for the FR, which is the process of recognizing persons [2]. Without using any further equipment, humans have the natural ability to recognize faces in a scene [3]. Additionally, a complex computer vision problem with a variety of aspects is the automatic FR [4]. In every practical aspect of our lives, it has been recognized as a developing field owing to the wide utilization of this technology. The areas where various applications correlated to this field are currently used widely such as defense, surveillance, banking, media field, social media platforms, fraud identification, e-commerce, trade, diagnostics, space science, research, genomics, bioinformatics, cyber security, internet of things and education [5]. Face Detection, face alignment, face matching and face representation are the four main steps of FRS. Face matching seeks to create a classifier to recognize distinct faces while extracting useful features to differentiate face images as of distinct people is the aim of face representation. The facial landmarks are obtained by the face alignment. Face Detection aims to return the face's position, if they exist.

Some specific conditions still make extremely accurate identification results impossible despite recent advancements in FR. This is typically the case when there is a lack of high-quality data, such as when face images are taken in uncontrolled environments (FR in the wild), where there are numerous variations that can affect facial images, including resolution, background, expression, pose, illumination and occlusions [6]. A critical issue viewed here is the feature's robustness retrieved from face appearance descriptors to tackle these problems. In recent years, a lot of attention was received by deep learning approaches to FR. A turning point was reached when neural networks for facial Feature Extraction (FE) were introduced [7].

Numerous prominent techniques for FE have been developed up to this point, like Scale Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP) etc. [8]. When these features are compared to the tested photographs, the person can be recognized or not. Unfortunately, as there are so many difficult and variable parameters in FD and identification as well as poor evaluation techniques, choosing the optimal facial recognition algorithm is a very difficult procedure [9]. Additionally, redundant features were engendered and struggled to handle high dimensions [10].

A. Problem Definition

Numerous approaches to resolving the FRS issue can be found in prevailing research procedures. Still, there is space for improvement. The following list includes the issues with the prevailing approaches:

- The prevailing approach focuses solely on brightening the images with the issue of locating a face in a hazy image. Thus, it has an impact on the recognition rate.

- Another difficulty for the FR system is the ageing process, which reflects the modifications in the face appearance / texture over a time.
- All the issues like age, expression, illumination, position and occlusion invariants were not addressed by most prevailing works. High recognition performance was attained by the prevailing presented pose invariant FR. However, when age, light, resolution and occlusion were all constant, it was unable to identify faces.

B. Limitations of Existing Works

Mohd Najib Mohd Salleh, Noureen Talpur and Kashif Hussain [27] exhibited Adaptive Neuro-Fuzzy Inference System (ANFIS) is an efficient estimation model not only among neuro-fuzzy systems but also various other machine learning techniques. ANFIS suffers from limitations that halt applications in problems with large inputs; such as, curse of dimensionality and computational expense. Different limitations such as curse of dimensionality, interpretability of rules, and parameter training are the major hurdles that need to be overcome for the implementation in problems with larger number of inputs. This is the reason, ANFIS is often integrated with additional techniques for input selection, rule reduction and parameter tuning, which again increases the complexity of the designed model. Various structural and parameters optimization techniques have been proposed in literature; however, there is enough room of improvement in ANFIS architecture so that applications in larger problems can be achieved easily.

A novel method of FR was proposed in this paper to tackle all the above-mentioned issues. The paper is framed as: the associated work regarding the proposed FR is analyzed in Section II. The concise discussion about FR is displayed in Section III. The proposed system's performance is analyzed in Section IV. The paper is concluded in Section V.

II. Literature Survey

Shokoufeh Mousavi et al.[11] presented a Distinctive Landmark - centered FR (DLFR) system to diminish the issues. Grounded on the number of key points acquired as of a modified SIFT together with the most idiosyncratic face's landmark region, novel features were generated. The weights for the features were optimized by a slightly altered genetic algorithm. A Support Vector Machine (SVM) classifier was utilized to classify the weighted features. The experiential outcomes exhibited higher efficacy than the prevailing papers. However, for SVM, a large number of data was not appropriate.

Xue Lv et al. [12] introduced a deep learning algorithm for the application of an FR model. For the face image's stimulation transformation along with crop pre-processing, an Optimized Multi-task Cascaded Convolutional Network (OMTCNN) algorithm was utilized. To diminish the FR's computational complexity in the embedded system, a lightweight FR algorithm grounded on Convolutional Neural Network (CNN) was developed, which was denoted as Lightweight CNN (LCNN). Finally, to build the multi-core embedded FRS, OMTCNN and LCNN were combined. The outcomes demonstrated that better performance was exhibited by OMTCNN in determining face identity. However, owing to the utilization of diverse datasets for training and testing, the developed system's accuracy was missed slightly.

Fanny Spagnolo et al. [13] developed a hardware architecture grounded on a real-time cascade classifier along with a resource-limited model to perform object detection. To accomplish the FD task along with integrating with an entire heterogeneous embedded system, architecture was tailored grounded on a Xilinx Zynq-7000 FPGA-centered System-on-Chip. The experiential outcomes exhibit that a high frame rate was significantly attained by the accelerator than other benchmark competitors however occupied much fewer resources.

Serign Modou Bah and Fang Ming [14] suggested a model tackling a few issues hampering FR accuracy to elevate the LBP codes by merging the LBP with augmented image processing schemes like Contrast Adjustment, Bilateral Filter, Histogram Equalization (HE), Image Blending and so on. Thereby, the overall FRS accuracy was elevated. The experiential outcomes exhibit that it was an accurate, reliable and robust model for FRS. Long histograms were engendered by this LBP that result in retarding the developed system's recognition speed.

Rameswari et al.[15] developed an automated access control system utilizing FR. The HOG was utilized to retrieve the features. The output was engendered by analogizing the face function that contained a classifier named SVM to classify the face encoding. To make the system more effective, Radio Frequency Identification (RFID) sensor together with an Infra Red (IR) sensor was implemented. The authority to the campus management was offered by engendering a webpage. The prevailing models were outperformed by the experiential outcomes. But, the HOG computation speed was tardy.

Bouchra Nassih et al. [16] demonstrated an efficient 3D FR approach named GD-FM+RF grounded on Riemannian geometry's Geodesic Distance (GD) along with Random Forest (RF). The GD between the points of 3D faces' specified pairs' was computed by employing the Fast Marching (FM) algorithm. To analyze the class separability, the Principal Component Analysis (PCA) algorithm utilized the retrieved features offered by the geodesic facial curves. Later, the input to the RF classifier was these features. Finally, the recognition rate was elevated along with attained most promising outcomes analogized to the benchmark models. However, while utilizing PCA, the information might be lost.

Yassin Kortli et al. [17] exhibited an implementation grounded on the Advanced RISC Machines Cortex-A9 processor utilizing the OpenCV library. Grounded on a hybrid ARM-Field Programmable Gate Array (ARM-FPGA) platform, a co-design was developed. Here, the FPGA processed the traditional hardware Visible Light communication (VLC) architecture. Optimized hardware VLC architecture was engendered to respect the real-time constraints. It was processed subsequently to comprehend the second Hardware / Software (Hw/Sw) co-design. The experiential outcome exhibited that a high speed was provided by the optimized hardware VLC architecture when analogized to the conventional non-optimized framework. The developed system's architecture was more complex.

Masoud Muhammed Hassan et al. [18] exhibited an FR approach grounded on the fusion of Gabor-centered FE, Fast Independent Component Analysis (FastICA) and Linear Discriminant Analysis (LDA). To make the uniform face images, it was converted to grayscale and resized. Later, the Gabor, FastICA and LDA methods were utilized to retrieve the facial features as of the aligned face. To recognize the individual's identity, the nearest distance classifier was utilized. Thus the developed model's efficacy was illustrated by the outcomes, especially under the Cosine distance measure. However, more time consumption along with the high computational cost was the model's major drawback owing to the utilization of '3' sophisticated FE schemes.

Lizzie D'cruz and J. Harirajkumar [19] suggested one-shot learning on the Siamese CNN (SCNN) for FR. Here, the employee attendance was entered in the My Structured Query Language (MySQL) database. For uploading the database, Acute Physiology together with Chronic Health Evaluation (APACHE) Web Server was the application programmable interface employed. Utilizing Send Mail Transfer Protocol (SMTP), get well soon messages were sent to absent employees via Gmail. Better performance was observed by the system with accurate FR along with testing '12' images successfully. However, more memory was requisite by the system to collect the data.

Hao Yang and Xiaofeng Han [20] designed an FR attendance system grounded on real-time video processing. A control group together with an experiential group was comprised in this experiment. Traditional fingerprint check-in was utilized by the control group and real-time video processing was utilized in the experiential group. Two universities were chosen as of a province and a similar number of students were chosen for the experiment, which was then collected, counted and analyzed in the experiment, application space. This may develop FR's prospects along with final issues in the actual check-in. The video FRS' effective performance was exhibited by the experiential data. Owing to the extraction of redundant features, the system was still challenging to offer accurate recognition.

Chaoyou Fu et al. [21] developed integrated large-scale visible data's huge identity information into the joint distribution. Additionally, to ensure their identity consistency, a pairwise identity-conserving loss was imposed on the engendered paired heterogeneous images. As of the noises, immense new distinct paired heterogeneous images with a similar identity could be engendered. To train the HFR network, identity consistency along with identity diversity properties permitted the engendered images via a contrastive learning mechanism. Thereby it yielded both domain-invariant as well as discriminative embedding features. The positive pairs and negative pairs regarded here were the engendered paired heterogeneous images and the images acquired as of diverse samplings. On '7' challenging databases that belong to five Heterogeneous FR (HFR) tasks, encompassing Sketch-Photo, Profile-Frontal Photo, Thermal-Visible and ID-Came, superior performance was attained by the model. Nevertheless, it was more expensive.

Alireza Sepas-Moghaddam et al. [22] presented a CNN grounded a deep Face Expression Recognition (FER) solution, named CapsField incorporated with an extra capsule network that employed dynamic routing to study hierarchical associations between capsules. The spatial features as of the facial images were extracted by CapsField, which learned the angular part-whole relations for a chosen set of 2D sub-aperture images offered from each Light Field (LF) image. Owing to its capacity in learning inter-view along with intra-view relations available in an LF image, the outcomes exhibited that when analogized with benchmark LF-centered FER, a superior performance. However, the system was susceptible to over fitting.

Weipeng Hu et al. [23] introduced an algorithm to explore the cross-modal image's potential domain-invariant Neutral Face (NF) representations by utilizing Dual Face Alignment Learning (DFAL). Feature-level Face Alignment (FFA), Image-level Face Alignment (IFA), as well as Cross-domain compact Representation (CdR) were the '3' effective components included in the model. To encode features for both VIS NF images as well as non- NF images, Teacher-Encoder CNNs (TeEn-CNNs) together with Student-Encoder CNNs (StEn-CNNs) were designed initially. By executing the feature level alignment betweet non- NF and VIS NF the FFA was brought in to learn NF representations. To restore face images, StDe-CNN was engendered to decode features secondarily. By imposing image-level alignment, an NF image was reconstructed by IFA's design. The DFAL approach's effectiveness on '3' challenging Near-Infra Red and VISual (NIR-VIS) databases was illustrated by the experimental outcomes. However, the position and orientation of input were not decoded by CNN.

Jae Young Choi and Bum Shik Lee [24] demonstrated a "Gabor Deep CNN (GDCNN) ensemble" method for FR applications. Distinct along with multiple Gabor face representations were effectively applied as inputs during the GDCNN's training and testing phases. GDCNN ensemble construction as well as GDCNN ensemble combination were the '2' parts comprised in the generated GDCNN ensemble. The experiential outcomes demonstrated that significantly a better FR performance was attained by the approach than the conventional approaches that were grounded only on grayscale or color face images as input illustration. However, more time was acquired by the system owing to the deep learning characteristics of Neural Networks (NN).

Hua Yang et al. [25] designed a Weighted Feature Histogram (WFH) method of Multi-Scale Local Patches (MSLP) utilizing Multi-Bit Local Binary Descriptors (MBLBDs) for FR. Using an MSLP Generation (MSLPG) method, the local patches were extracted to attain multi-scale information. To extract MBLBDs, the MBLBD learning (MBLBDL) method was engendered to diminish the binary descriptors' quantization information loss. To project Pixel Difference Vectors (PDVs) into the MBLBDs in each local patch, a learning mapping matrix along with multi-bit coding rules were engaged in MBLBDL. To discover a set of robust weights for each patch, the Robust Weight Learning (RWL) method was employed to incorporate the MBLBDs into the final face representation. The outcomes as of experiments performed on face datasets exhibited that benchmark recognition performance was attained by WFH and a Coupled WFH (C-WFH).

Harish Sharma, Garima Hazrati, Jagdish Chand Bansal [28], Spider monkey optimization (SMO) algorithm is a recent addition to the list of swarm intelligence based optimization algorithms. The update equations are based on Euclidean distances among potential solutions. The algorithm has extensively been applied to solve complex optimization problems. SMO is applied to solve optimal capacitor placement and sizing problem in IEEE-14, 30 and 33 test bus systems with the proper allocation of 3 and 5-capacitors. SMO is used for the synthesis of sparse linear arrays. The amplitudes of all the elements and the locations of elements in the extended sparse subarray are optimized by the SMO algorithm to reduce the side lobe levels of the whole array under a set of practical constraints. The SMO has also been used to synthesize the array factor of a linear antenna array and to optimally design an E-shaped patch antenna for wireless applications. SMO is a meta-heuristic technique inspired by the intelligent foraging behavior of spider monkeys. The foraging behavior of spider monkeys is based on the fission-fusion social structure. Features of this algorithm depend on social organization of a group where a female leader takes decision whether to split or combine. The leader of the entire group is named here as the global leader while the leaders of the small groups are named as local leaders. With reference to the SMO algorithm, the phenomenon of food scarcity is defined by no improvement in the solution. Since SMO is a swarm intelligence based algorithm, each small group should have a minimum number of monkeys. Therefore, at any time if a further fission creating at least one group with less than the minimum number of monkeys, we define it as the time for fusion. In SMO algorithm, a Spider Monkey (SM) represents a potential solution. SMO consists of six phases: Local Leader phase, Global Leader phase, Local Leader Learning Phase, Global Leader Learning phase, Local Leader Decision phase and Global Leader Decision phase.

III. Proposed Facial Recognition System

For the efficient recognition of the human face, a novel FIS-CDNN Classifier is proposed in this paper. The face is detected initially in this system together with extracting the features and landmarks. Subsequently, the features are chosen. Finally, the face is classified utilizing the FIS-CDNN classifier. Fig. 1 illustrates the proposed model's block diagram.

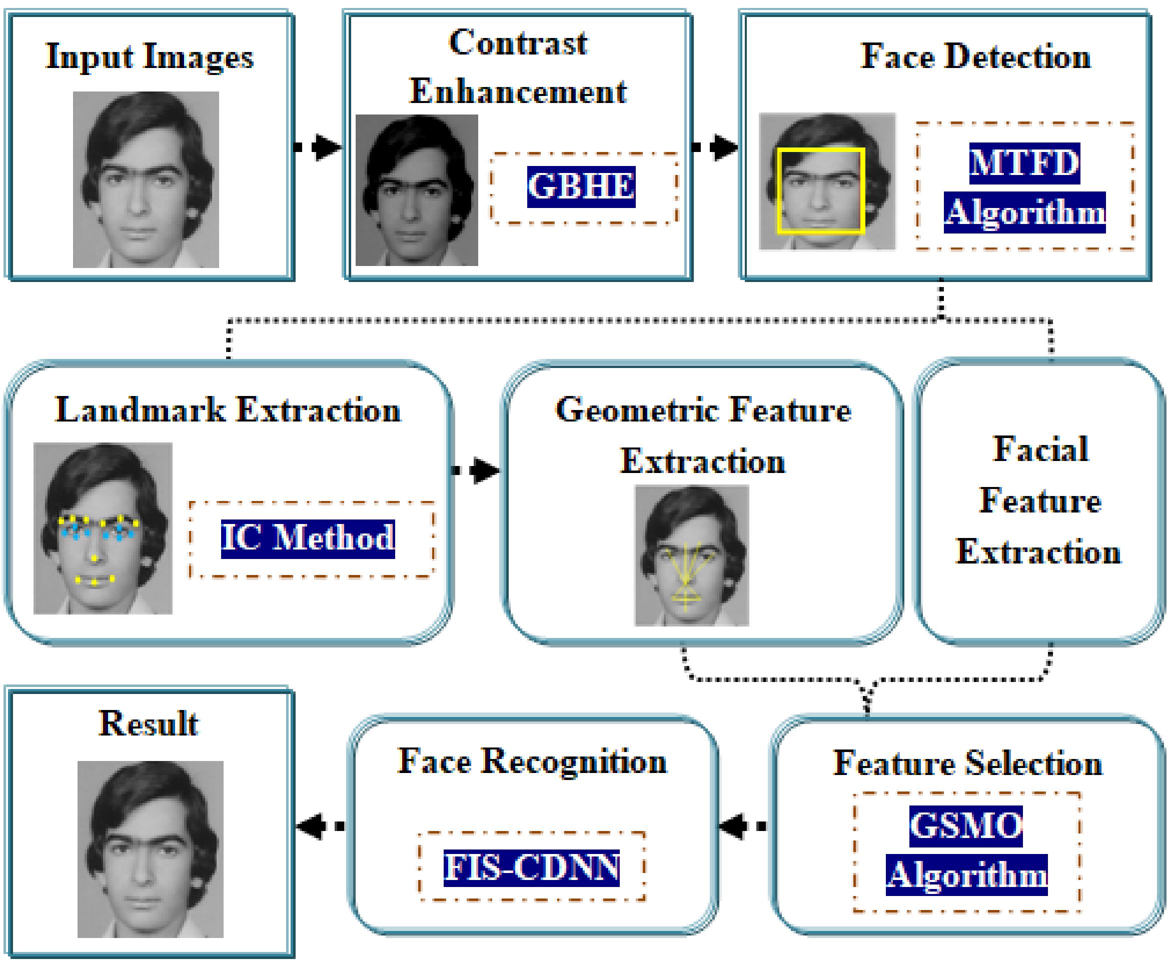


Fig. 1. Block diagram of the proposed methodology

A. Contrast Enhancement

Since the image's quality is low, the face Image (I) is taken as an input and pre-processed in the FRS. By the contrast enhancement technique of Gamma correction-centered Based Histogram Equalization (GBHE), the lower quality is elevated. To augment contrast, HE is a methodology for fine-tuning image intensities. The comparative frequency of various gray levels occurring in the image was represented by the histogram of an image. However, the histogram may bear the intensity saturation issue during the contrast enhancement. So, to control the intensity value, Gamma correction was utilized in this methodology to solve the intensity saturation issue.

A nonlinear operation utilized to encode and decode luminance in the image is Gamma, which is expressed as:

$$G_I = ZI^{\xi} \tag{1}$$

Where, the positive constants are illustrated as Z and I, the gamma correction's output is denoted as G_I, and the gamma value is represented as ξ. Here, the input image's pixel is mapped into the associated pixel of a prepared output image. It is defined as Histogram, which can be expressed as:

$$I_{pre} = \frac{P_{i(G_I)}^{no.oc}}{n(P)} \tag{2}$$

Where, the number of occurrences of a particular pixel with intensity in G_I is denoted as $P_{i(G_I)}^{no.oc}$, the pre-processed image is notated as I_{pre}, the number of pixels is symbolized as $n(P)$ and the image's range of gray level is $[0, L-1]$.

B. Face Detection

The most prominent step in the proposed methodology is face detection. Utilizing the MTFD, the face is extracted by eliminating the unwanted parts from I_{pre}. The Tiny Face Detector is utilized to perform facial extraction on images by implementing the ResNet101 algorithm. In the scale-invariant stage, cosine similarity is utilized rather than Euclidean distance. If '2' similar documents are far distant by the Euclidean distance (owing to the document's size), there may be still a probability of oriented each other as the cosine similarity is having benefits. This is the advantage of cosine similarity.

The input image is shifted into ResNet101 initially. For detecting the face utilizing a pre-trained Artificial Neural Network (ANN), the features are extracted along with amassed in numpy arrays. It expressed the extracted face features as:

$$Q_n = \{Q_1, Q_2, Q_3, ...Q_N\} \tag{3}$$

Where, the number of extracted features is represented as Q_n. By image scale and resolution, the facial extraction process is affected. Between the input image and extracted feature, the similarity score A is computed as:

$$A = \frac{Q_n * I_{pre}}{\|Q_n\|\|I_{pre}\|} \tag{4}$$

The extracted features for FD are valid if the score is identical for extracted features and the input image. The face is not detected otherwise. F denotes the detected face.

C. Feature Extraction

The informative data regarding the face is obtained with the aid of FE. The key features like SIFT, HOGs, Speeded-Up Robust Features (SURF), Binary Robust Independent Elementary Features (BRIEF), LBP, Gabor Feature, Shape, and edge are retrieved as of the face detected region F. The extracted features E_n are expressed as:

$$E_n = \{E_1, E_2, E_3, ...E_N\} \tag{5}$$

1) SURF: It is a fast along with robust algorithm comprising of '4' main parts that are Integral image generation, Fast-Hessian detector, Descriptor orientation assignment and Descriptor generation. By utilizing the integral image for extracting interest points, it is centered on a Hessian-matrix approximation. By applying an approximate Gaussian and derivative scale-space representation, it operates under the entry of integral images. E_1 implies the SURF and the integral image F_Σ can be signified as:

$$F_\Sigma = \sum_{x=0}^{x \le p} \sum_{y=0}^{y \le q} F(x, y) \tag{6}$$

Where, the input image after integral image representation is modeled as $F(x, y)$. For smoothening and space scale representation, a Gaussian filter is also applied. In the pyramid, the scale representation is implemented. For the extraction of the interesting point E_1, the box filter's size doubled the sampling intervals owing to the usage.

2) SIFT: Robustness to illumination, scale and transformation was shown by SIFT. Grounded on specific steps like extremes of scale space detection, determining the feature point's position, feature points direction along with engendering feature descriptors, the features are extracted. To smooth the images, diverse scales of the Gaussian function were utilized by SIFT algorithm. Stable features are chosen along with regarded as a candidate as of the smoothened image. Further, by utilizing the Harris operator, the corner points around the candidate are computed as:

$$E_2 = g(F, S_1, S_2) * F(x, y) \tag{7}$$

Where, the output extracted features are noted as E_2, the Gaussian function is given as $g(F, S_1, S_2)$, the Integral

Length Scale (ILS) is demonstrated as S_1, the scale for the corner point is denoted as S_2, the input image with pixel coordinates are symbolized as $F(x, y)$. If the threshold is lower than the corner points, the corner points are outstanding to represent the object.

3) LBP: A model of retrieving textual features F as of the face detected image F is the LBP, which is a non-parametric operator describing an image's local spatial structure in binary information. Classification, detection and reorganization are the three steps performed by this. The face image's centre pixel is observed as a threshold value along with analogized with the neighbouring pixels. It is denoted as '1' if the neighbouring pixel's value is superior to the threshold value. Otherwise, it is mentioned as 0. An 8-bit value was obtained by merging all these outcomes. Then, the local texture feature's (E_3) LBP value was obtained as:

$$E_3 = \sum_{i=0}^{u-1} \mathrm{T}(P_{(i)\in F} - P_{(c)\in F}) 2^i \tag{8}$$

$$\mathrm{T}\left(P_{(i)\in F} - P_{(c)\in F}\right) = \begin{cases} 1 & P_{(i)\in F} \geq P_{(c)\in F} \\ 0 & P_{(i)\in F} < P_{(c)\in F} \end{cases} \tag{9}$$

Where, the central pixel value is represented as E_3, the neighboring pixel value is proffered as $P_{(i)\in F}$, the threshold function is modeled as T, the number of pixels in the neighbor of $P_{(c)\in F}$ is notated as u and the centre pixel is signified as $P_{(c)\in F}$.

4) HOG: The gradient structure features E_4 is captured by utilizing HOG. Descriptor blocks, block normalization, Orientation binning as well as Gradient computation are the '4' steps involved here. For each pixel, the face image's gradient is computed. The pixel value's magnitude and direction in each block are computed utilizing the gradient. As the image's few parts might show a slight vanish, block normalization is implemented to oversight the image's brightness. Normalization is employed in specific parts to diminish that. The mathematical representation is exhibited as:

$$E_4 = \frac{v}{\|v\|^2 + \sigma} \tag{10}$$

Where, the constant value to avert deviation is σ, the non-normalized vector is signified as v. Conclusively as of the normalization operation, the HOG features are derived along with obtaining the features E_4.

5) BRIEF: A general-purpose feature point descriptor amalgamated with arbitrary detectors is named BRIEF, which can be notated as E_5. Detection, descriptor and descriptor matching were '3' steps incorporated. By smoothing the image utilizing a Gaussian filter, the BRIEF commenced averting the descriptor from being sensitive to high frequency. Then, the random pair of pixels was chosen by the BRIEF around the targeted region. Comparisons may occur between utilizing these pixels. The value of 1 is assigned to the corresponding bit if the initial pixel is found to be brighter compared to the second pixel. Or else it is found 0.

$$E_5 = \begin{Bmatrix} 1 & F(a) \leq F(b) \\ 0 & F(a) \geq F(a) \end{Bmatrix} \tag{11}$$

Where, the two selected points are notated as a and b.

6) Shape: The face image's geometric properties were referred to by the shape feature that is utilized to describe the image content. Grounded on the shape boundary information or boundary plus interior elements where the area, parameter and circularity are deemed as key features, the shape features are extracted. It could be computed as:

$$E_6 = \sum_{x,y} F(x', y') \tag{12}$$

Where, the output of the face's shape feature is represented as E_7, the input face image with the extreme pixels is notated as $F(x', y')$.

7) Edge: The input image's edges E_7 is identified by the Canny algorithm. By utilizing noise removal, differentiation, non-maxima suppression, double thresholding and edge tracking. The edge features are acquired as of the image. These images are smoothened initially and gradients ensure the image's edges are taken. By computing the magnitude together with gradients of directions, the points which are not at the maximal are suppressed in the non-maxima suppression. By employing thresholds namely upper threshold and lower threshold in double thresholding, the strong along with weak edges are identified. The edge tracking is aided in determining the final edges. '2' thresholds namely the upper threshold (h_{high}) and lower threshold (L_{low}) are taken in the tracking. The edges E_8 are classified as strong edge (S_{edge}), weak edge (W_{edge}) and non-edge (U_{edge}) along with deciding the concerning gradients under three conditions that are:

$$E_7 = \begin{cases} S_{edge} & if\ F(t,v) > L_{low} \\ W_{edge} & if\ L_{low} < F(t,v) < h_{high} \\ U_{edge} & if\ L_{low} > F(t,v) \end{cases} \tag{13}$$

Where, the gradient in the gradient direction is proffered as $F(t,v)$.

D. Landmark Extraction

Utilizing the Improved Chehra (IC) landmark Extraction method, the face's landmark is extracted as of F. By training a cascade of regressors, this discriminative facial deformable model was obtained. For extracting the Face appearance, the Chehra method utilized PCA in normal. A drawback here is the original features turn into Principal Components (PC) after implementing PCA on the dataset. The linear combination of original features is the PC. As original features, PCs are not readable and interpretable; also it took more time for executing multiclass datasets. So, LDA is employed in the Face appearance stage rather than this.

For facial region extraction, 49 facial landmark points were computed by the Chehra model on the whole facial image. The face image's 10 regions computed following the landmark detection are the forehead region, right eye region, right face region, lower nose region, left eye region, entire eye region, lip region, left face region and upper nose region. For facial appearance, FE has been executed by utilizing LDA after obtaining the blocks.

LDA is a feature diminution technique grounded on supervised linear transformation. It does so in a way that elevates the ratio between the intra-class scatter $\wp_{\text{int}\,ra}$ and the inter-class scatter $\wp_{\text{int}\,er}$ when projecting a high-dimensional feature vector onto a low-dimensional space. For multiclass l, the scatters are determined as follows:

$$\wp_{\text{int}\,ra} \in F == \sum_{i=1}^{l} \sum_{\Omega \in F} \left(X_\Omega - \mu_i\right)\left(X_\Omega - \mu_i\right)^T \tag{14}$$

$$\wp_{\text{int}\,er} \in F = \sum_{i=1}^{l} \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \tag{15}$$

Where, the average landmark point in i^{th} class is represented as μ_i, the Ω^{th} instance in i^{th} class is indicated as X_Ω. The best projection is then computed along with maximizing the ratio between the projected samples' intra-class scatter matrix and inter-class scatter matrix. It is defined as:

$$X^{best} = \arg\max\left(X\right) \frac{\left|X^T \wp_{\text{int}\,er} X\right|}{\left|X^T \wp_{\text{int}\,ra} X\right|} \tag{16}$$

Where, the best projection point is denoted as X^{best}, the sample point is denoted as X. For further process, the projects are taken as landmark features.

E. Geometric Feature

The face's GF is extracted as of the X^{best} in this phase. The utilization of the distance between the facial landmark points as features is allowed owing to the facial geometry's invariance to the illumination. Hence, the feature utilized here is the distance between landmarks. The GFs like height, width, length of chin, eyebrows, nose, mouth etc. are extracted. The extracted GF is expressed as:

$$G_n \in X^{best} = \left\{G_1, G_2, G_3, \ldots\ldots\ldots G_N\right\} \tag{17}$$

Where, the number of extracted GFs is signified as G_n. The extracted facial together with GFs is amalgamated and the output can be expressed as:

$$\lambda_n = G_n \cup E_n \tag{18}$$

Where, the combined features are represented as λ.

F. Feature Selection

For elevating the proposed facial detection system's learning accuracy, the features are chosen as of λ_n utilizing the GSMO Algorithm after FE. A meta-heuristic method named SMO is motivated by the clever foraging behavior of spider monkeys. The Fission-Fusion Social Structure (FFSS) serves as the base for the foraging habits of spider monkeys. The group's social organization is in such a way that the female leader makes a decision on whether to split or combine the group. In cases of food scarcity, a female creates mutable smaller groups as well as typically serves as the swarm's Global Leader (GL). Availability of food as of a specific territory is the factor upon which the group sizes rely. A direct proportionality is identified between the size and food availability. The GL deemed here is the entire group's leader and the smaller group's heads are named the Local Leaders (LLs). The variation between the current position and positions engendered at random is averaged out to update the leader position. For a specific issue, a random position is engendered within a predetermined range. As a result, the research proposed the alteration to quicken convergence and boost reliability. The updating stage uses the Gaussian Mutation Technique rather than the Convergence rate.

The number of monkeys deemed here were the extracted features. With the uniform distribution, each Spider Monkey (SM) is initialized as:

$$E_i \in \lambda_n = E_{\min(j)} + \Omega(0,1) \times \left(E_{\max(j)} - E_{\min(j)}\right) \tag{19}$$

Where, E_i represents i^{th} SM, the lower and upper limits in j^{th} dimension are indicated as $E_{\min(j)}$, and $E_{\max(j)}$, the random number distributed uniformly in the range of $[0,1]$ is signified as Ω.

The following '6' steps are followed by an algorithm for foraging after initialization. The steps are explained as follows,

a) Local leader phase (LLP): Utilizing the past behaviours of both the LL and the local group members, the SM in LLP shifts its present location. Only when the new location's fitness value is greater than the prior site, the SM's location is updated with the new location. Using the Gaussian mutation technique E_{ij}^{*}, the i^{th} SM of the d^{th} local group's location update equation is:

$$E_{ij}^{*} = E_{ij} + \Omega(G) \tag{20}$$

Where, the j^{th} dimension of i^{th} SM is proffered as E_{ij}, the random Gaussian number is symbolized as $\Omega(G)$.

b) Global leader phase (GLP): Similar to the LLP, but diversely, the swarm updates throughout this phase. Only one randomly chosen solution's dimension is updated here, unlike the LLP. Depending on its probability, which spider monkey will have a probability of being updated is identified.

$$E_{ij}^{*} = E_{ij} + \Omega(0,1) \times \left(\Re_j - E_{ij}\right) + \Omega(-1,1) \times \left(E_{Rj} - E_{ij}\right) \tag{21}$$

Where, the GL position in j^{th} dimension is notated as $\Re_j$. Grounded on the probability that is derived utilizing their fitness, which is computed centered on the classifier's accuracy, the SMs location is updated. The probability of i^{th} SM Φ is calculated as:

$$\Phi = \frac{9}{10} \times \frac{f_i}{f^{\max}} + \frac{1}{10} \tag{22}$$

Where, the fitness of i^{th} SM is illustrated as f_i, the maximum fitness is depicted as $f^{\max}$. The SM's fitness for the freshly engendered position is then analogized to the old position, and a better position is adopted.

c) Global leader learning phase (GLLP): Here, an SM is updated as the swarm's GL whose position has the maximum fitness value of all the members. The global limit count is elevated by 1 if the position of GL is not updated, representing the number of iterations during which the GLs position was not updated.

d) Local leader learning phase (LLLP): At this point, the group's greedy selection is utilized to update the LL's position. The LL's upgraded position is then coordinated with the previous one. The Local Limit Count (LLC) is then elevated by 1 if the LL position has not been updated.

e) Local leader decision phase (LLDP): In this phase, either by random initialization or by means of collective information as of GLs experience utilizing the equation below, the position of every member in that group is updated if any LL does not become restructured to a specific verge known as LL Limit.

$$E_{ij}^{*} = E_{ij} + \Omega(0,1) \times \left(\Re_j - E_{ij}\right) + \Omega(-1,1) \times \left(E_{ij} - \aleph_{dj}\right) \tag{23}$$

f) Global leader decision phase (GLDP): LLD and GLD are equivalent. The GL splits the swarm into a lower maximum number of groups if it does not restructure to a specific boundary known as the GL limit. At that point, the LL was chosen by commencing the LL Learning procedure. Additionally, the GL will unite the entire group if the maximal numbers of groups are engendered but the GLs position is not updated. Until enough food is obtained to feed every SM, the procedure is unremitting. Several features are chosen in this method of opting for leaders, which are expressed as:

$$E_n^{sel} = \left\{E_1^{sel}, E_2^{sel}, E_3^{sel}, \ldots\ldots\ldots E_N^{sel}\right\} \tag{24}$$

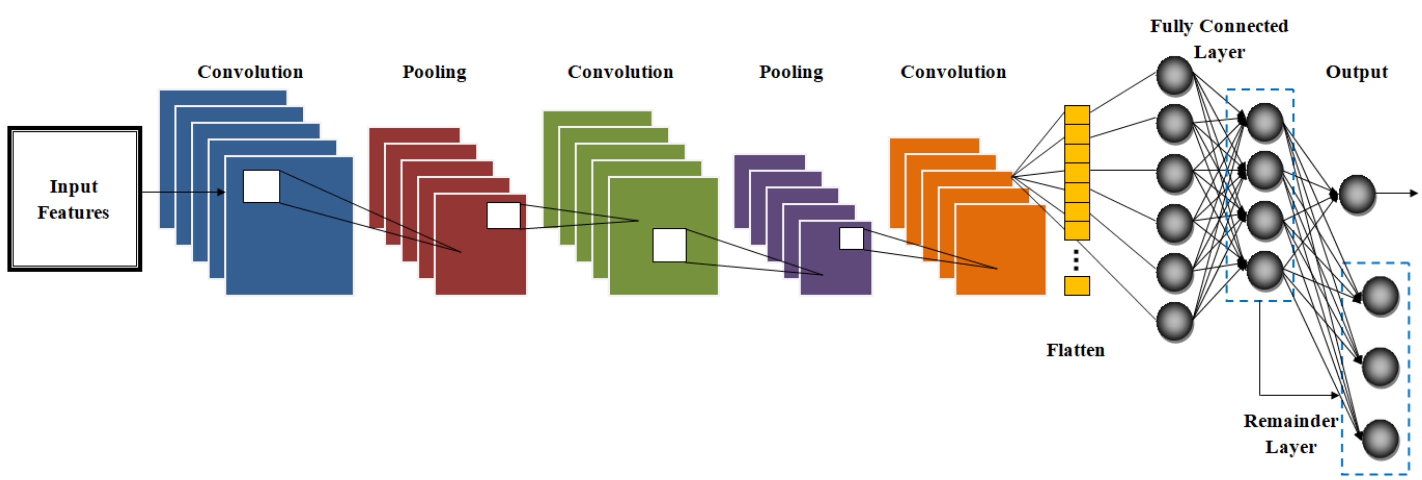


Fig. 2. Architecture of proposed FIS-CDNN

Where, the number of selected features is denoted as E_n^{sel}. Likewise, utilizing a similar algorithm, the features G_n are chosen. Followed by, it is utilized for choosing features E_n along with obtaining the output G_n^{sel}.

G. Face Recognition

Here, for detecting the human's face, the chosen features E_n^{sel} are fed as input to the FIS-CDNN classifier. To process the image's Red, Green and Blue elements at the same time, a 3D NN was utilized by a Convolutional Deep Neural Network (CDNN). Compared to traditional feed-forward NN, the number of artificial neurons requisite to process an image is diminished considerably. Normally, for the reason of weight value selection, the classifier suffers from the accuracy issue. The remainder layer comprising the previous training images behavior is deemed in this proposed methodology to solve this issue. The previous training weight values are amassed in the remaining layer. Therefore, the weight selection is very much aided by the remainder layer along with elevating the accuracy. Using the Fuzzy Interference system (FIS), the weight value is initialized in the proposed classifier. Fig. 2 exhibits the proposed system's architecture.

a) Weight initialization: By utilizing the five layers, the weights are initialized by the FIS system. The fuzzification process is executed by the first layer, which contains the adaptive nodes. The fuzzified output Ψ_i is:

$$\Psi_i = \nabla_1 \left(\Omega_{\tau_k} \right) \quad (25)$$

$$\Psi_i = \nabla_2 \left(\Omega_{\tau_j} \right) \quad (26)$$

Where, the input nodes are notated as ∇_1 and ∇_2, the weight value acquired as of the previous iteration is symbolized as τ_k and τ_j, the membership function is represented as Ω, which indicates how much given nodes satisfy the quantifier.

The output signals acquired as of the preceding layers are multiplied in the second layer along with the second layer's ϖ_i output is:

$$\varpi_i = \nabla_1 \left(\Omega_{\tau_k} \right) * \nabla_2 \left(\Omega_{\tau_j} \right) \quad (27)$$

For normalizing the second layer's output, normalization is executed in the third layer and is expressed as:

$$\left(\varpi_i \right)^* = \frac{\varpi_i}{\varpi_1 + \varpi_2} \quad where, i = 1,2 \quad (28)$$

Where, the normalized features are specified as $\left(\varpi_i \right)^*$. The fuzzy rule's consequent part is executed by the fourth layer that contains the adaptive nodes. Node function has a form:

$$\left(\bar{\varpi}_i \right)^* = \left(\varpi_i \right)^* \left(\varepsilon_i \nabla_1 + \vartheta_i \nabla_2 + \ell_i \right) \quad (29)$$

Where, the linear adaptive parameters are symbolized as ε_i, ϑ_i and ℓ_i, defuzzification is proffered as $\left(\bar{\varpi}_i \right)^*$.

Finally, the proposed system's total weight value W is computed in the final layer as:

$$W = \sum \left(\varpi_i \right)^* \left(\varepsilon_i \nabla_1 + \vartheta_i \nabla_2 + \ell_i \right) \quad (30)$$

b) Convolution layer: A specialized sort of linear operation employed for FE is convolution. Here, a kernel is a small array of numbers applied across the input. At each tensor's location, an element-wise product between kernel's each element and the input array is computed. It is again summed to acquire the output value in the output array's corresponding position. By applying multiple kernels, this procedure is recurring to generate an arbitrary number of feature maps representing the input array's diverse characteristics. The given input data's convolution is mathematically expressed as:

$$c = \sum_d \sum_d \left(E_n^{sel} \right) (\alpha - d, \beta - d) * W(d, d) \quad (31)$$

Where, the convolution operation's output is signified as c, the weight value in the dimension size $d \times d$ is represented as $w(d,d)$, the input matrix's dimension size is α and β. Each convolution layer is specifically pursued by a non-linear Activation Function (AF) implemented element-wise to the preceding layer's output to permit the network to learn non-linear decision boundaries. ReLu is the activation function utilized in this work. The activation function is mathematically expressed as:

$$R = \max(0, c) \quad (32)$$

Where, the ReLu activation function's output is signified as R.

c) Pooling layer: By passing via the AF, the output is obtained, which is again passed to the pooling layer to aggregate the information along with diminishing the representation. The pooling operation's (Δ) result is:

$$\Delta = \frac{c - W}{S} + 1 \quad (33)$$

Where, the kernel's strides are notated as S.

d) Fully connected layer (FCL): The final convolution or pooling layer's output feature maps are transferred into a single-dimensional array of numbers, which means typically flattened. By a learnable weight, each input is allied to every output in this layer. The amount of output nodes contained in the final FCL is typically similar to the number of classes. Then, the flattened output is computed as:

$$\Gamma = \Delta - \left(W \left(d \times d \right) - 1 \right) \quad (34)$$

Where, the FCLs output is indicated as Γ.

e) Softmax layer: To the softmax activation function, the FCL's output is fed directly into this layer. In the output layer, the activation function is utilized along with normalizing the output real values in the range of $[0,1]$ as of the last FCL to target class probabilities. The softmax function Γ_{soft} is given as:

$$\Gamma_{soft} = \frac{e^{\Gamma_\varphi}}{\sum_{\varphi=1}^{h} \Gamma_\varphi} \quad (35)$$

Where, the FCL output at φ^{th} node is notated as Γ_φ , the total number of output nodes is represented as h . With higher accuracy, the faces are recognized by this process. The proposed FIS-CDNN's pseudo code is:

Input: selected features E_n^{sel}

Output: Recognized face I

Begin

Initialize parameters c , R , Γ ,layer b

Compute weight value

$W = \sum (\varpi_i)^* (\varepsilon_i \nabla_1 + \vartheta_i \nabla_2 + \ell_i)$

For $b = 1 to n$

while $b = 1$ (first round of convolution and pooling layer)

Compute convolution operation c

Evaluate activation function

$R = \max(0, c)$

Compute pooling operation Γ

End while

End for

Flattening all the layers

Evaluate softmax activation function Γ_{soft}

End begin

Return I

IV. Results and Discussion

The proposed facial recognition system's performance with FIS-CDNN is analyzed in this section. MATLAB (MATrix LABoratory) is the working platform in which the model is implemented. It is a 4th generation programming language with a multi-paradigm numerical computing environment. For rapid along with simple scientific computations and I/O, it is designed particularly.

A. Database Description

FG-NET, collected from a publically available source, is the dataset used in the proposed framework. The dataset for age assessment along with FR across ages is FG-NET. A total of 1,002 images obtained as of 82 people with an age ranging between 0 to 69 and an age gap of up to 45 years were composed in the dataset.

B. Performance Analysis of Classification Method

Grounded on a few quality measures along with prevailing models like Z- Normalization and Moore Penrose-based Deep CNN (ZMP-DCNN) [26], NN, CNN and Adaptive Neuro-Fuzzy Inference System (ANFIS), the proposed FIS-CDNN method's effectiveness is examined here.

TABLE I. A Comparative Analysis of the Proposed Model and Existing Model with Respect to TP, TN, FP and FN

Classifier	TP	TN	FP	FN
FIS-CDNN	1.853659	161.8537	0.146341	0.146341
ZMP-DCNN	1.52439	161.5244	0.47561	0.47561
CNN	1.487805	161.4878	0.512195	0.512195
NN	1.317073	161.3171	0.682927	0.682927
ANFIS	1.402439	161.4024	0.597561	0.597561

Table I summarises the performance analysis of the proposed FIS-CDNN and the prevailing models Regarding True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The proposed model accurately recognized the face in the outcome named TP. The mismatched face is predicted by the proposed model in the outcome TN. Hence, the proposed model's TP is 1.853659, which is higher than the prevailing techniques. Similarly, 161.8537 is the proposed FIS-CDNN's TN. The face is recognized incorrectly by the model in an outcome named FP. Here, the proposed model's FP and FN are similar at 0.146341. The face is classified more effectively by the proposed model than the prevailing models, which is obtained as of the entire analysis.

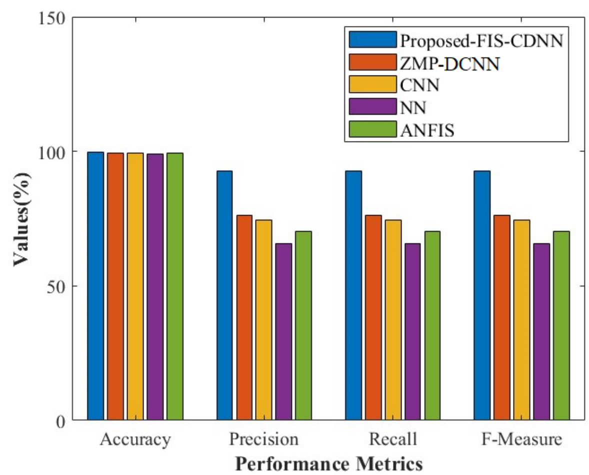


Fig. 3. The performance of FIS-CDNN with the existing methods

The proposed FIS-CDNN and prevailing method's performance analysis are illustrated in Fig. 3, grounded on the Accuracy, Precision, Recall along with F-Measure which is defined as the amalgamation of precision as well as recall. The proposed system's higher performance was exhibited by the proposed method's higher value in every metric. The face was recognized by the proposed method with 99.82% Accuracy. It is higher than the prevailing methods and provided better performance. Likewise, 92.68% is the proposed FIS-CDNN's Precision and Recall value. An F-Measure of 16.47% higher than the existing ZMP-DCNN was exhibited by the proposed one.

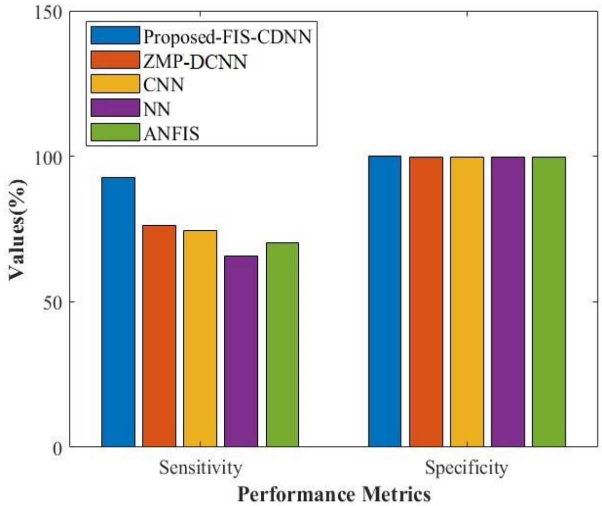


Fig. 4. The performance of FIS-CDNN with the existing methods

The proposed FIS-CDNN and the prevailing method's performance analysis are depicted in Fig. 4, grounded on Sensitivity and Specificity. The degree of similarity between views of an object that evokes an exact recognition response as of a given subject is Specificity. The capability to classify a test between the True and False positive performance correctly is named Sensitivity. While the ZMP-DCNN attains 99.7%, and CNN attains 99.6%, the proposed model's Specificity is 99.9%, which is superior to the prevailing models. Correspondingly, a higher sensitivity of 18.29% was attained by the proposed scheme than the prevailing CNN.

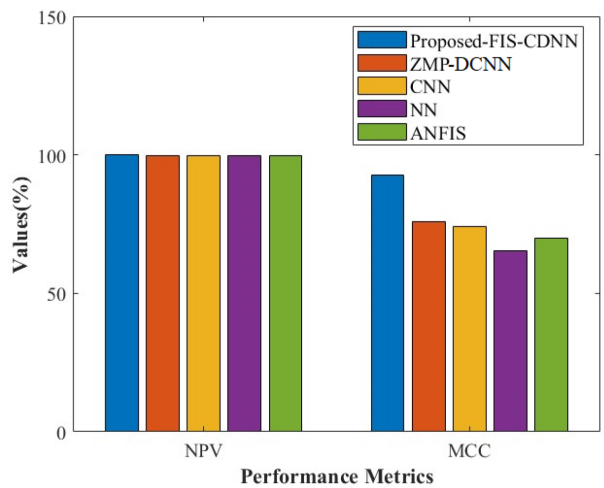


Fig. 5. The performance of FIS-CDNN in terms of NPV and MCC

The proposed model's performance is demonstrated in Fig. 5, on the basis of Negative Predictive Value (NPV) together with Matthew's Correlation Coefficient (MCC). The proposed model's better performance is proved by higher NPV and MCC. 99.9% is the proposed model's NPV, which is superior to the ZMP-DCNN and all other prevailing models. Likewise, 92.59% is the proposed scheme's MCC.

TABLE II. THE PERFORMANCE OF FIS-CDNN IN TERMS OF FPR, FNR, FRR, FDR

Classifier	FPR	FNR	FRR	FDR
FIS-CDNN	0.000903	0.073171	0.073171	0.073171
ZMP-DCNN	0.002936	0.237805	0.237805	0.237805
CNN	0.003162	0.256098	0.256098	0.256098
NN	0.004216	0.341463	0.341463	0.341463
ANFIS	0.003689	0.29878	0.29878	0.29878

Table II summarizes the proposed FIS-CDNN and the prevailing model's performance analysis, Concerning False Positive Rate (FPR), False Negative Rate (FNR), FRR and FDR. The proposed scheme's higher performance was exhibited by lower values of FPR, FNR and FRR. The proposed model's FPR is 0.002033 and 0.002259 lower than ZMP-DCNN and NN respectively. But, it is 0.002786 higher than ANFIS. Similarly, the proposed methodology's FNR, FRR and FDR are identical along with exhibiting an augmentation by 0.073171. This exposed that it is lower than the prevailing CNN along with other prevailing models. So, it is concluded from the metrics that better performance was exhibited by the proposed model for facial recognition.

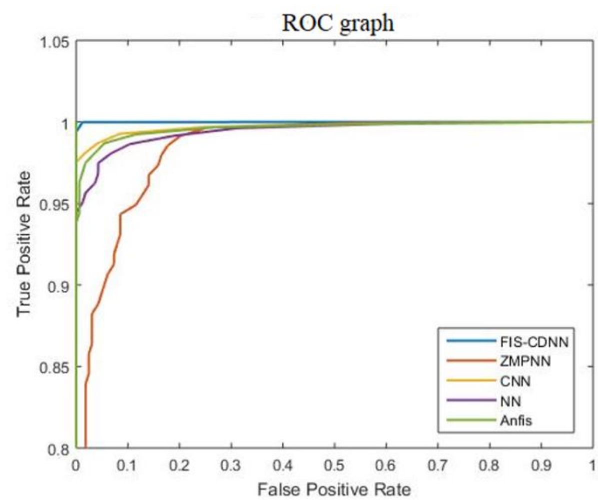


Fig. 6. ROC curve of proposed FIS-CDNN and existing model

Fig. 6 exhibits the FIS-CDNN and the existing model's ROC curve analysis. The facial recognition's performance or quality was evaluated by an effective model named ROC. The True Positive Rate (TPR) is plotted against the False Positive Rate (FPR) to engender a ROC curve. The ROC curve illustrates that higher sensitivity together with specificity than the prevailing models like ZMP-DCNN, CNN, NN and ANFIS was exhibited by the proposed model.

V. CONCLUSION

A novel FIS-CDNN classifier was proposed in this work for facial recognition. Pre-processing, FD, facial FE, face landmark, geometric FE, feature selection and classification are all processes involved in the proposed technique. The experimental analysis is performed after all the steps have been completed, and regarding some quality criteria, the proposed

FIS-CDNN's performance is analogized with that of the prevailing approaches. The final outcomes exposed that a 99.82% accuracy rate was attained by the proposed model. Similar outcomes are obtained for all other metrics, including sensitivity, specificity, recall, precision, recall, recall, f-measure, NPV, MCC, FDR, FPR, FNR and FRR by the proposed scheme. The proposed model is therefore found to be significantly more effective than the prevailing approaches for classifying faces grounded on the findings of all metrics. This work will be extended in the future by extracting features as of various head poses employing utilizing advanced models.

REFERENCES

[1] Kolipaka Preethi and swathy vodithala, "Automated smart attendance system using face recognition", Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021), 06-08 May 2021, Madurai, India, 2021.

[2] Naing Min Tun, Alexander I Gavrilov and Pyae Phyo Paing, "Human face recognition using combination of ZCA feature extraction method and deep neural network", International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 25-29 March 2019, Sochi, Russia, 2019.

[3] Pallavaram Venkateswar Lal, Gnaneswarar Rao Nitta and Ande Prasad, "Ensemble of texture and shape descriptors using support vector machine classification for face recognition", Journal of Ambient Intelligence and Humanized Computing, 2019, https://link.springer.com/article/10.1007/s12652-019-01192-7.

[4] Asif Mohammed Arfi, Debasish Bal, Mohammad Anisul Hasan, Naeemul Islam and Yasir Arafat, "Real time human face detection and recognition based on haar features", IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh, 2020.

[5] Jannat Binta Alam, Momarul Islam Md, Taskeed Jabid and Sabbir Ahmed, "System development using face recognition", International Conference on Automation, Computational and Technology Management (ICACTM), 24-26 April 2019, London, UK, 2019.

[6] Alireza Sepas-Moghaddam, Ali Etemad, Fernando Pereira and Paulo Lobato Correia, "CapsField: Light field-based face and expression recognition in the wild using capsule routing", IEEE Transactions on Image Processing, vol. 30, pp. 2627-2642, 2021.

[7] Belen Rios-Sanchez, David Costa-da-Silva, Natalia Martin-Yuste and Carmen Sanchez-Avila, "Deep learning for facial recognition on single sample per person scenarios with varied capturing conditions", Applied Sciences, vol. 9, no. 24, pp. 1-11, 2019.

[8] Eric-Juwei Cheng, Kuang-Pen Chou, Shantanu Rajora, Bo-Hao Jin, Tanveer M, Chin-Teng Lin, Ku-Young Young, Wen-Chieh Lin and Mukesh Prasad, "Deep sparse representation classifier for facial recognition and detection system", Pattern Recognition Letters, vol. 125, pp. 71-77, 2019.

[9] Farah Deeba, Hira Memon, Fayaz Ali Dharejo, Aftab Ahmed and Abddul Ghaffar, "LBPH-based enhanced real-time face recognition", (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 10, no. 5, pp. 274-280, 2019.

[10] Genevieve M Sapijaszko and Wasfy B Mikhael, "Facial recognition system using mixed transform and multilayer sigmoid neural network classifier", Circuits, Systems and Signal Processing, vol. 39, no. 1, pp. 6142-6161, 2020.

[11] Shokoufeh Mousavi, Mostafa Charmi and Hossein Hassanpoor, "A distinctive landmark-based face recognition system for identical twins by extracting novel weighted features", Computers and Electrical Engineering, vol. 94, pp. 1-17, 2021.

[12] Xue Lv, Mingxia Su and Zekun Wang, "Application of face recognition method under deep learning algorithm in embedded systems", Microprocessors and Microsystems (Pre-proof), 2021, https://doi.org/10.1016/j.micpro.2021.104034.

[13] Fanny Spagnolo, Stefania Perri and Pasquale Corsonello, "Design of a real-time face detection architecture for heterogeneous systems-on-chips", Integration, vol. 74, pp. 1-10, 2020.

[14] Serign Modou Bah and Fang Ming, "An improved face recognition algorithm and its application in attendance management system", Array, vol. 5, pp. 1-7, 2020.

[15] Rameswari R, Naveen Kumar S, Abishek Aananth M and Deepak C, "Automated access control system using face recognition", Materials Today: Proceedings, vol. 45, no. 2, pp. 1251-1256, 2020.

[16] Bouchra Nassih, Aouatif Amine, Mohammed Ngadi, Youssef Azdoud, Driss Naji and Nabil Hmina, "An efficient three-dimensional face recognition system based random forest and geodesic curves", Computational Geometry: Theory and Applications, vol. 97, no. 1, pp. 1-8, 2021.

[17] Yassin Kortli, Maher Jridi, Mehrez Merzougui and Mohamed Atri, "Optical face detection and recognition system on low-end-low-cost Xilinx Zynq SoC", Optik, vol. 217, pp. 1-16, 2020.

[18] Masoud Muhammed Hassan, Haval Ismael Hussein, Adel Sabry Eesa and Ramadhan J. Mstafa, "Face recognition based on gabor feature extraction followed by FastICA and LDA", Computers, Materials & Continua, vol. 68, no. 2, pp. 1637-1659, 2021.

[19] Lizzie D'cruz and Harirajkumar J, "Contactless attendance system using Siamese neural network based face recognition", Materials Today: Proceedings (In Press), 2020, https://doi.org/10.1016/j.matpr.2020.10.462.

[20] Hao Yang and Xiaofeng Han, "Face recognition attendance system based on real-time video processing", IEEE Access, vol. 8, pp. 159143-159150, 2020.

[21] Chaoyou Fu, Xiang Wu, Yibo Hu, Huaibo Huang and Ran He, "DVG-Face: Dual variational generation for heterogeneous face recognition", Journal of Latex Class Files, 2020, https://doi.org/10.48550/arXiv.2009.09399.

[22] Alireza Sepas-Moghaddam, Ali Etemad, Fernando Pereira and Paulo Lobato Correia, "CapsField: Light field-based face and expression recognition in the wild using capsule routing", IEEE Transactions on Image Processing, vol. 30, pp. 2627-2642, 2021.

[23] Weipeng Hu, Wenjun Yan, Haifeng Hu, "Dual face alignment learning network for NIR-VIS face recognition", IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 4, pp. 2411-2424, 2021.

[24] Jae Young Choi and Bum Shik Lee, "Ensemble of deep convolutional neural networks with gabor face representations for face recognition", IEEE Transactions on Image Processing, 2019, http://dx.doi.org/10.1109/TIP.2019.2958404.

[25] Hua Yang, Chenting Gong, Kaiji Huang, Kaiyou Song and Zhouping Yin, "Weighted feature histogram of multi-scale local patch using multi-bit binary descriptor for face recognition", IEEE Transactions on Image Processing, vol. 30, pp. 3858-3871, 2021.

[26] Santhosh S., Dr. S. V. Rajashekararadhya, Efficient Face Recognition System Using Z-Normalization and Moore Penrose-Based Deep Convolutional Neural Network, Advanced Engineering Science, vol. 54, Issue 02, pp. 3177-3202, 2022.

[27] Mohd Najib Mohd Salleh, Noureen Talpur and Kashif Hussain, Adaptive Neuro-Fuzzy Inference System: Overview, Strengths, Limitations, and Solutions, Conference Paper · August 2017, DOI: 10.1007/978-3-319-61845-6_52.

[28] Harish Sharma, Garima Hazrati, Jagdish Chand Bansal, "Spider Monkey Optimization Algorithm", Chapter in Studies in Computational Intelligence · January 2019, DOI: 10.1007/978-3-319-91341-4_4.

Performance Analysis of Machine Learning-based Detection of Sinkhole Network Layer Attack in MANET

Sivanesan N[1], K.S. Archana[2]
Research Scholar, Department of Computer Science and Engineering[1]
Assistant Professor, Department of Computer Science and Engineering[2]
Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India[1, 2]

***Abstract*—This paper proposes an Intrusion Detection System (IDS) against Sinkhole attacks in Mobile Adhoc Networks (MANET) with mobile sinks. A sinkhole attack is where a hacked node advertises a false routing update to draw network traffic. One effect of a sinkhole attack is that it may be used to launch further attacks, such as drops or changed routing information. Sinkhole nodes attempt to forge the source–destination routes to attract the surrounding network traffic. For this purpose, they modify routing control packets to publish fake routing information that makes sinkhole nodes appear as the best path to some destinations. Several machine learning techniques, including Decision Tree (DT), K-Nearest Neighbor (KNN), Convolution neural network (CNN), and Support Vector Machine (SVM), are used to do the categorization. Furthermore, the MANET's node's characteristics, particularly speed, are used for feature extraction. Totally 3997 unique samples, including 256 malicious samples and 3604 normal samples are collected. The categorization results demonstrate the accuracy of DT, KNN, CNN, and SVM at 98.4%, 96.7%, 98.6%, and 97.8%, respectively. The CNN approach is more accurate than other methods, at 98.6%, based on the data. After that, Priority, SVM, KNN, and CNN, in that order, each denotes excellent accuracy.**

Keywords*—Sinkhole; machine learning; MANET; intrusion detection

I. Introduction

A collection of autonomously placed, wirelessly linked nodes makes up a MANET (Mobile Adhoc Network). Each MANET node acts as a router to send the packet from the source node to the destination node. Massive and frequently used networks are remote ad hoc networks. MANET has no centralized management node; each moveable node is autonomous. The moveable system is allowed to go anywhere they are needed.

It allows the nodes to rapidly enter or go away the network [1]. Nodes are not limited in their ability to communicate with one another. Data loss may occur if the association is created. MANET is often utilized in various industries, including scientific, military, search and rescue, etc. Due to increased network connectivity, cyber-attacks are also rising [2]. Ad-hoc WMN (Wireless Mobile Network) is vulnerable to several attacks because of sharing of the channel, an unstable operating location, constrained mobility of resources, frequently changing device topology, and source limitations [3]. The detection of anomalies accepts interference from routine system operations. Due to the intermittent nature of system activity, counting standard system output is difficult [4]. The abnormal process detects recent assaults or unexplained with a high rate of false positives. As an attack detection technique, sign-based IDS is defined by looking for distinctive features in network data, such as a sequence of bytes [5]. It only acknowledges known attacks and misses brand-new assaults for which there is no trend. Safe connectivity in MANET is a difficult problem because of the absence of established infrastructure and complicated topology, among other factors. The idea of intrusion detection keeps the balance by using access control and cryptographic techniques. As an automated detection and source of warning, it is presented to stop an attack that has already occurred or is currently ongoing. IDS only find intrusion that sets off an alert since they are passive and do not take any preventive measures [6].

Marti *et al.*, used Watchdog approaches to reduce routing errors and sinkhole nodes in MANET. The scheme was created by the authors using the DSR protocol. The node delivering the traffic watches promiscuously the transmission of the nearby node and route to detect any malicious conduct on the part of that node. The neighboring node will be regarded as acting inappropriately if it disrupts the data flow. The offending node won't be permitted to participate in the next transaction. The watchdog keeps a copy of recently delivered packets, examines each packet it has received and overhears them for similarities [7]. ML was one aspect of AI that was created in the late 1950s. It has grown and changed through time into algorithms that might be machine based and effective sufficient in engineering, medicine, and computer science to address various issues, including sorting, grouping, regression, and optimization. One of today's most popular technologies is machine learning. ML enables workstations to study without personal involvement and respond consequently dynamically. It automatically, appropriately, and efficiently manipulates complicated data to create a model. ML may profit from a generic framework to have a broad approach to enhancing device performance. It has several scientific uses, including data cleansing, noise reduction, picture identification, automatic spam detection, medical diagnostics, and manual data input [8],[9]. According to the most recent research, ML has been used in WSNs to solve several issues.

By incorporating ML into WSNs, complicated issues like reprogramming, manually navigating through enormous amounts of data, and valuable mining information from the facts are avoided. ML approaches are frequently useful in acquiring enormous amounts of data and providing useable data [10]. This thesis' main goal is to provide a strategy for identifying Sinkhole threats using machine learning techniques.

The lack of a reliable security solution that can shield MANETs from routing assaults is their major problem. The required design solutions are not anticipated to lead to resource limitations like battery life and bandwidth. MANETs' ability to "self-organize" can be both a strength and a weakness in terms of security because it leaves opportunities for both passive and active attacks. MANETs have many weaknesses, and some are brought on by their changeable architecture, constrained power source, bandwidth, and scalability.

Because of the weaknesses mentioned above, existing architectures adapted from wired broadband cannot be used directly for MANETs; as a result, a strong security framework must be designed and deployed exclusively on MANETs. Any future QoS-aware security system must work toward achieving availability, integrity, anonymity, identification, and secrecy as security objectives. Individually created security procedures must be used to produce a safe and QoS-aware network [4]. The issue with MANETs is that they are not widely known for providing combined answers to security or QoS-based issues due to their dynamic nature. This research focuses on sinkhole threats through protocol modification to obtain strong security detail without lowering the quality of service in real time.

II. Literature Review

Wireless networks are extremely susceptible to attacks, and hackers can access communication channels. In MANETs, Programme modules that automatically track harmful network activity might be used to keep an eye on attackers. When creating an intruder identification mechanism for MANETs, we must keep certain things in mind [11]. In [9], a geometric-based black hole and grey hole attack scheme detection is examined. According to [10], a secure data fragment is created to detect and prevent a hole in the ground and Sybil attacks on deployed fixed and dynamic nodes, producing high detection and low false-positive rates. Future directions versus DDoS assaults are provided in [11], along with concerns and taxonomy. [12] provides a summary of Sybil's protection methods utilized in online social webs. The intrusion detection systems for MANETs will operate independently of their wired counterparts. The creation of intruder detection systems for MANETs presents various challenges. Non-collaborative intruder monitoring systems use node-level agents to detect and record any unexpected activity [12]. The biggest obstacle is figuring out where the agents are while the nodes move.

Similarly, the nodes housing the intrusion-detecting agents need more processing power, bandwidth, and battery life. However, those services are constrained in MANETs [13]. Several authors have proposed methods to offer the most suitable answers to an NP-complete problem that involves raising the intruder detection performance with the least amount of resources. There are several intrusion detection architectures available for MANETS [14]. A wide range of assaults is possible, some of which are more devastating in MANETs than in wired networks. The characteristics of these networks prevent the use of conventional methods for identifying attack traffic. Although intrusion detection systems (IDSs) use many different detection methods, anomaly detection is among the most crucial.

Additionally, IDSs based on past attack patterns are less effective if such IDSs are centralized. According to Peterson et al. [15], the detection engine was modified to include a modern Machine Learning approach that recognizes attack traffic live (not to be analyzed and assessed later)[15]. This allowed the IDS rules to be changed instantly. Amouri et al. provide a two-level monitoring approach for spotting rogue nodes in MANETs. The first stage involves the installation of specialized sniffers that operate in promiscuous mode.

Every sniffer uses a decision-tree-based classifier, which generates numbers we apply to every occurrence successfully classified throughout reporting time. The second stage involved sending the categorized instances to the super node that was run on algorithms. Each node being examined establishes the quantities connected to the cumulative fluctuation value of the acquired categorized instances. A workable IDS strategy for wireless sensor networks is the result approach, which has also been expanded [16]. Abd-El-Azim and colleagues proposed MANET's simplified fuzzy-based intrusion detection approach with an automated mechanism using an Adaptive Neuro-Fuzzy Inference System that produces a fuzzy system (ANFIS). The FIS was configured, and this initialization framework was optimized using a genetic algorithm (GA). In the presence of solely blackhole assaults, the network grew by an average of 36% [17]. Soni & Sudhakar recommended the Intrusion Detection Device for the Jamming attack. Depending on time example, the jamming attacker slowly introduced the network packets, quickly increasing their number. The IDS is identified as the attacking node through its unwanted flooding activities, and the attacker's malware is found. The proposed approach continually monitored all network activity, and the harmful node's behaviors were distinct from those of other nodes and did not act normally [18]. Sultana et al. examined the current IDS output when the supposed packet-dropping nodes in a MANET network were present. The reputed intermediate nodes, also known as intermediate bottleneck nodes, lose packets once the number of packets exceeds their handling limits. The effectiveness was calculated using the NS-2 network simulator. The results have demonstrated that the reputational packet-dropping nodes' IDS algorithms' neglect seriously affects network routine [19]. A strategy for cooperative sinkhole identification was put out by Kim et al. in [20].

The sinkhole assault is analyzed, and its characteristics are extracted. The algorithm for sinkhole identification was created to save time and money. When a mobile node receives a route request message with an originator ID matching the receiving mobile node's, it checks the message's sequence number. The current node recognizes the presence of a sinkhole node. It determines that the route request message originated from the sinkhole node if the sequence number in the route request message is higher than the current sequence number of the mobile node. Therefore, it may be said that the sinkhole node is present in the route path of the request message. On routing protocols like, DSR and AODV, Gagandeep. G *et al.*, [21] concentrated on sinkhole attacks and suggested a Security-aware routing (SAR) method to lessen the effects of a sinkhole assault. SAR executes the message routing security and routing update security processes. Presented by Shafiei et al., [22] is a distributed method of sinkhole attack detection. The MANET's nodes are regarded as reliable nodes. These dependable nodes serve as watchdogs nodes. Each monitoring node includes the local knowledge of the network. Furthermore, A base station is necessary for the detecting operation. Depending on the method of assault, MANET attacks can roughly be divided into two categories: passive attacks and active attacks. [28] [29]. An active assault involves disrupting information, modification, or manufacturing, which impairs the MANET's regular operations [34]. A passive attack exchanges data over the network without interfering with communication. The main taxonomy of MANET security exploits is shown in Table I. Passive assaults include eavesdropping, packet analysis, and traffic monitoring. According to the assaults' domain, the assaults can also be divided into external and internal attacks. Nodes are not a member of a network's domain launch external attacks.

III. Proposed Work

In MANETS, both passive and active assaults are available. The attacker aims to disrupt the network by modifying, injecting, forging, inventing, manipulating, and discarding data packets during an active attack. These assaults qualify as serious assaults. Active assaults include packet dropping and denial of service (DOS) attacks [32]. It may be divided into two categories: internal attacks and external attacks. Compared to external attacks, internal attacks are harder to find. Assaults that are focused, such as malicious packet dropping, routing, sleep deprivation, black hole, gray hole, and rushing attacks. Sinkhole attacks are a novel type of assault that require special attention in MANET [33].

A. Sinkhole Attack

Attacks involving sinkholes are challenging to locate for the reasons listed below. When MANET's regular communication occurs, the sinking node makes numerous attempts to draw in the nearby nodes. With the AODV protocol, data packets are either changed or quietly dropped [23][30]. How the rogue node grows and its sequence number is a mystery. As was already mentioned, the AODV sequence number is utilized to indicate how recent a route is. The malicious node listens on the communication channel and keeps track of each node's sequence number. Following that, it assigns itself the highest sequence number among all nodes in the route and suddenly invades the channel, dropping packets [24][31]. For instance, have a look at Fig. 1. The source node is node 1, while the destination is node 5.

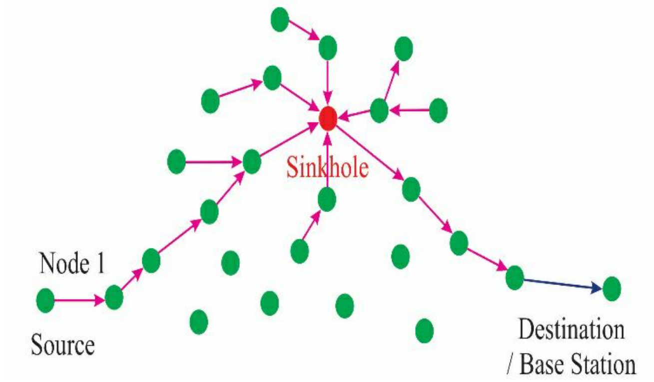


Fig. 1. Sinkhole attack.

B. Support Vector Machine (SVM)

SVM uses a hyperplane, a subset of supervised machine learning, to determine the best classification for each observation in a given data set. SVM is more effective with big datasets and can handle linear and non-linear problems [25]. SVM is incorporated into WSNs to handle various challenges, including congestion control, fault detection, routing, communication, and localization concerns.

Wireless Sensor Networks (WSNs) are susceptible to various software, hardware, and communication-related errors. Given the diversity of deployment scenarios and the limited sensor resources, fault detection in WSNs is difficult [35]. Additionally, the detection must be exact to prevent false alarms and quick to prevent loss. One of the easiest ways to find failure in WSNs appears to be to employ machine learning. Support vector machines (SVMs), a classification technique, are employed in this research to achieve this. SVM is used in our situation to define a decision hu function based on statistical learning theory [27]. This decision function can be used at cluster heads to find sensors because it is a low-resource procedure.

C. K- Nearest Neighbor (K-NN)

K-Nearest Neighbor is the most well-liked example-based method for regressing and classifying issues (k-NN). K-NN is primarily responsible for defining the distance between the sample being measured and the samples being provided. The many distances, including the Chebyshev distance function, Manhattan distance, Euclidean distance, and Hamming distance, are known in k-NN. The measurements are lowered due to the method's ability to identify the absent samples from the highlighted room. K-NN was first developed by using anomaly detection and data aggregation in WSN applications. An effective defensive mechanism for the WSN is provided by an intrusion detection system (IDS), a proactive network proper security solution. In this paper, we propose a carefully planned edge that needs to perform penetration testing when the WSN confronts a DoS attack [26]. To achieve this, we introduce the kNN in computer vision and the arithmetic optimization technique (AOA) in evolutionary computation. We employ a parallel method to improve the interaction between the population and the Lévy flying strategy to modify the optimization to increase the model's accuracy.

D. Deep Learning with Naïve Bayesian Learning

In WSNs, DL addresses various issues, including anomaly and defect finding, energy harvest, calculating data efficiency, and routing. Deep learning models' security applications, including spam filtering, IDS and malware detection have grown in significance in the design of information safety, categorization, and prediction activities. These numerous operations are designed to build a paradigm that typically distinguishes between "regular" and "malicious" samples, such as assaults and typical packets, based on intelligence. The exponential rise in Deep Learning Model usage compounds the complexity of attack plan tools. The mathematical learning method known as Bayesian learning looks for relationships between the datasets by learning separability using various statistical methodologies. Bayesian learning uses a variety of prior probability distributions and fresh information to assess 8 posterior likelihoods. The probability of p() must be increased if Y1, Y2, Y3,..., Yn represent a sequence of input and returns a mark. Numerous issues with WSNs have been handled using Bayesian learning techniques, such as routing, connection issues, fault prediction, data localization, and aggregation.

E. Convolution Neural Network and Decision Trees

The collections of if but then other rules are used by supervised learning machine learning algorithms, such as DT, to increase readability. DT anticipates a class or objective based on the judging criteria and creates a training model using training data. Decision trees provide various benefits, including openness, reduced complexity, and thorough examination of decision-making. Different WSN issues, such as mobile devices, data aggregation, connection, etc., are solved using decision trees.

CNN's have been extensively utilized for deep learning (DL). Convolution layers and completely coupled layers make up this system. Sub-sampling layers may occur between these two levels. With multidimensional, locally correlated input data, they can acquire the greatest results from DNNs with properly scaled complexity. Therefore, the immediate use of CNN occurs in DB, where comparatively many nodes and attributes need to be learned.

The identification of harmful content may be done using our technique. This Sinkhole mitigation is established in a network of normal and malicious output stream monitoring nodes. We first explain the total of healthy nodes and cancerous nodes using their processes. This approach creates a tunnel between the communication or packet and the malicious nodes. These are only sent through the tube. (see Fig. 2).

At that point, receive a message that assists in data collecting and follows data out of each moving node. By defining the crucial function, the system's operation may be increased. At that point, eight key characteristics were chosen to create a dataset tagged with the backing of an exceptional hub address. Consequently, six common machine learning classifiers separate the legitimate and harmful data from research samples into two groups.

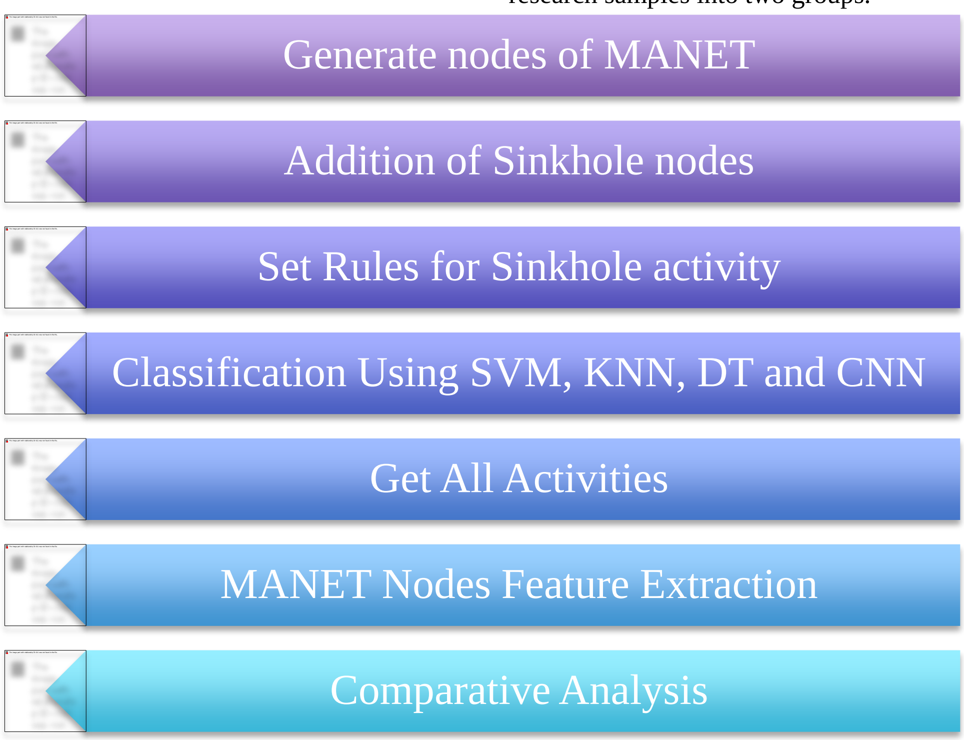


Fig. 2. Conceptual image for detection process.

IV. RESULTS AND DISCUSSION

The device's effectiveness is evaluated using various mathematical standards and contrasted with the new techniques. The developed standalone system for intrusion detection is tailored to work in the network and data link levels of the TCP/IP paradigm, both using the proposed routing mechanism. Because TCP requires ACKs from the destination, it employs two mechanisms to accomplish its task and is designed to work for UDP traffic. The watchdog examines every node that is close enough to a transmission. Before concluding that the node is acting maliciously, it compares the two variables (the threshold value and the counter). The watchdog algorithm calculates the downtime for the node's delay in forwarding the message if a packet is withheld and compares it to the threshold value.

A. *Quality Measures*

Metrics used for evaluation include Sensitivity, Accuracy, and precision mentioned in equations 1 to 3. These measurements are measured using four different parameters: true negative (TN), true positive (TP), false negative (FN), and false positive (FP). The percentage of properly categorized documents among all the data is called accuracy. Precision refers to the performance's pertinent proportion. On the other hand, recall is the proportion of correct classifications the algorithm makes for total functional outcomes. Detection Rate (DR), called True Positive Rate, is the proportion of anomalous records accurately recognized as anomalies to all anomaly records (TPR).

$$\text{Accuracy} = TN + TP / TN + FP + TP + FN \quad (1)$$

$$\text{Sensitivity} = TP / TP + FN \quad (2)$$

$$\text{Precision} = TP / FP + TP \quad (3)$$

B. *Simulation Results*

This work is modeled to detect sinkhole assaults in the Matlab 2019b set with limited nodes. The network protocol, Channel, Computer, and node are combined to create a network topology. In this simulation procedure, many network applications send and receive packets over a network. The simulation execution reaches the principal role and is carried through to the termination stage as packets are created, accepted, and processed. Fig. 3 depicts the nodes' initial positions and the nodes with which they make contact. 48 legitimate nodes and two malicious nodes made up the ad hoc network environment used for this experiment in Fig. 3, 11 sinkhole nodes, are shown as red circles. In contrast, the normal nodes are shown as black circles. Additionally, blue lines connecting the nodes represent the initial connection.

C. *Results of Feature Extraction*

One of the fundamental ideas in machine learning which directly affects performance is the choice of features. Functions unrelated or loosely connected might negatively affect the device's output. Only any data for a full node are included in the output file's full node information. The provided application is educational. When elements that are not important or provide less information

If characteristics that help with classification are left out, it might choose comparable features. There are several advantages of selecting features, including less over-fitting, shorter preparation times, improved precision, etc. It selected eight key components that enhance the functionality of the system. This includes a sum of distances, number of nodes, minimum speed, maximum speed, fastest direction, average speed, and distance to the destination. This proposed work collected 3604 samples which contain standard and wicked samples (normal 3348 and malicious 256). It creates data that is put together and given the eight selected qualities. The high-volume dataset was produced in an ad hoc network setting to identify sinkhole attacks.

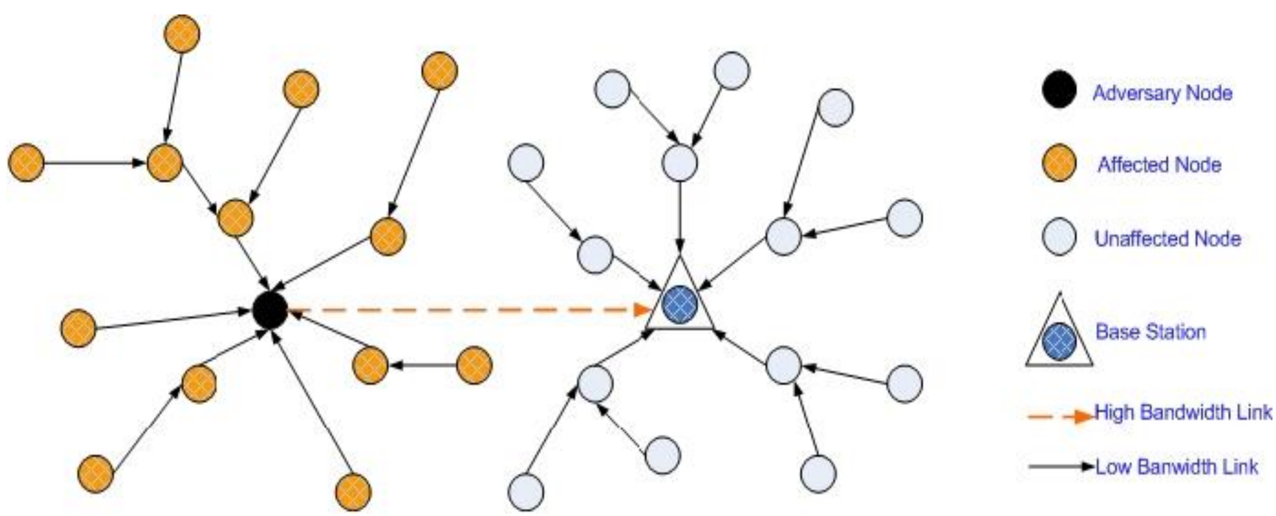


Fig. 3. Position of initial MANET nodes.

D. Classification Results

Fig. 4 shows the classification outcomes using a variety of machine learning techniques, including Convolution Neural Network (CNN), Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbor (KNN). The green arrays in Fig. 4's confusion matrix represent genuine values, whereas the red components represent incorrect ones. The target class is often regarded as a positive class for binary assessment. Our primary goal in this article is to locate sinkhole nodes between conventional nodes. Sinkholes are therefore seen as a good class. The top cell displays the true negative, while the bottom one displays the genuine positive, according to the perplexing matrix of Fig. 4 based on true values. The upper one is a false-negative class from red cells, whereas the bottom is a false-positive class. Malicious and normal nodes are included in the two classes used for classification.

Fig. 4. Example confusion matrix of SVM and KNN.

The vertical grey cells show precision and negative predictive values, while the horizontal grey cells show sensitivity. For instance, 192 (or 75.1 percent) of the 256 sinkhole nodes in the SVM approach are accurately detected. However, 64 (24.9%) are incorrectly identified as normal nodes. In other words, the SVM method's sensitivity is 75.1 percent. The SVM approach has a 99.6% specificity for identifying the normal node. This indicates that just 16 (0.4 percent) of the 3348 normal nodes are incorrectly diagnosed. Additionally, 92.3% (precision) of the discovered sinkhole nodes in the SVM classifier are in a true condition. The SVM classifier, on the other hand, has an accuracy rate of 97.8%. The overall accuracy value that makes up SVM is the value in the confusion matrix's lower-right corner cell. Consequently, the findings demonstrate that the accuracy of the SVM, KNN, DT, and CNN techniques are 97.8%, 96.7%, 98.4%, and 98.6%, respectively. Additionally, the classifier's overall error value is highlighted in red lettering in the lower-right corner.

The input matrix is 8 by 1. Additionally, we employed two convolutional layers with ten 2x2 sizes, stride [1, 1], and padding-free filters. Additionally, we utilized the Tanh and ReLU routines to trigger the layers. Then, correspondingly, 384 and 2 cells are employed in each of the two completely linked layers. Probability is determined, and the final levels are activated using the SoftMax layer

Next, the cross-entropy classification layer is applied while taking into mutually exclusive account classes. Fig. 5 shows the outcomes of the categorization procedure. There are 3000 iterations in the training process. Fig. 5 shows the accuracy and loss value of the training process. The horizontal axis of the ROC curve represents the false positive rate, while the vertical axis represents the true positive rate. In other words, the ROC curve is shown, with the positive class—sinkhole nodes—being considered.

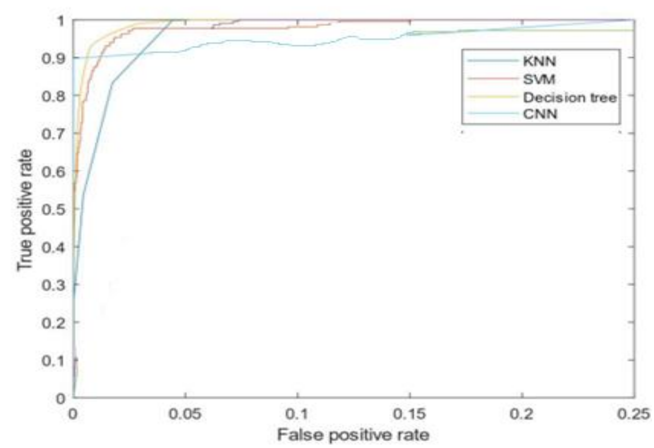


Fig. 5. The ROC curves of different classifiers.

TABLE I. COMPARATIVE ANALYSIS OF VARIOUS METHODS FOR PRECISION IN %

ALGORITHM	EXISTING	PROPOSED
SVM	88.3	92.3
KNN	83.4	89.5
DT	82.4	89.7
CNN	62.4	65.3

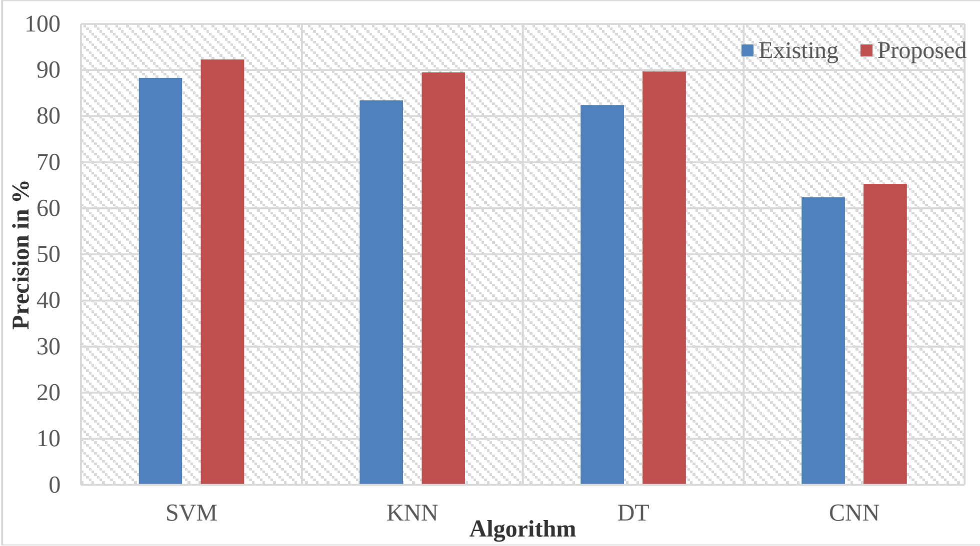


Fig. 6. Comparative analysis of different methods for precision.

Table II compares precision values for different classification algorithms in the existing and proposed systems. The proposed system works better in terms of precision. Fig. 6 represents the comparison between existing and proposed work for various algorithms concerning the precision, and the proposed work outperforms when compared to existing algorithms.

Table II compares precision values for different classification algorithms in existing and proposed systems. The proposed system works better in terms of precision. Fig. 7 represents the comparison between existing and proposed work for various algorithms concerning sensitivity, and the proposed work outperforms when compared to existing algorithms.

The work displays the findings of the assessment of several machine learning techniques. The sensitivity of the DT technique works better than another method, according to the results. The method's sensitivity shows the method's ability to find sinkhole nodes in MANET. As a result, its size signifies the classifiers' potential. The DT classifier has more sensitivity than previous approaches. The accuracy also demonstrates the method's dependability or potential for results.

TABLE II. COMPARATIVE ANALYSIS OF VARIOUS METHODS FOR SENSITIVITY IN %

ALGORITHM	EXISTING	PROPOSED
SVM	72.3	75.1
KNN	60.0	60.2
DT	89.3	91.3
CNN	78.3	80.1

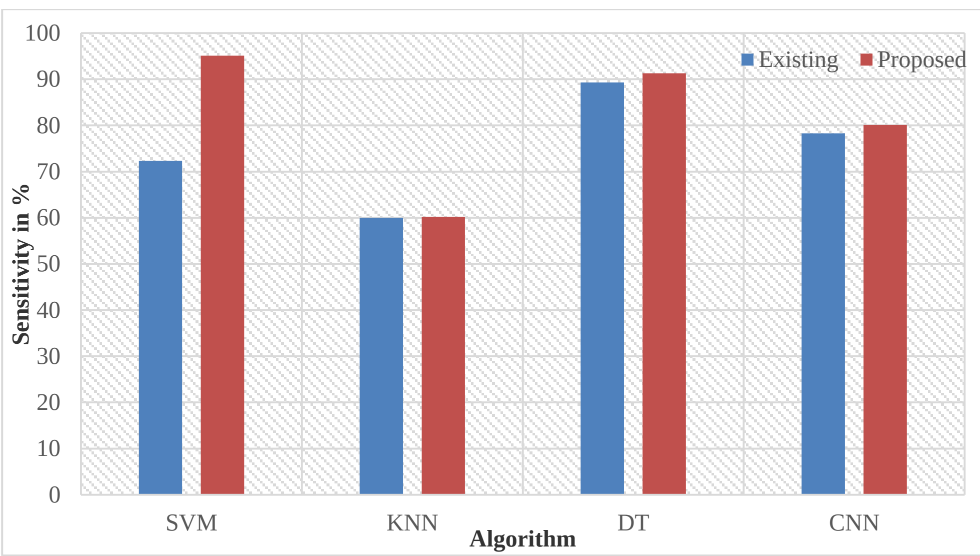


Fig. 7. Comparative analysis of different methods for sensitivity.

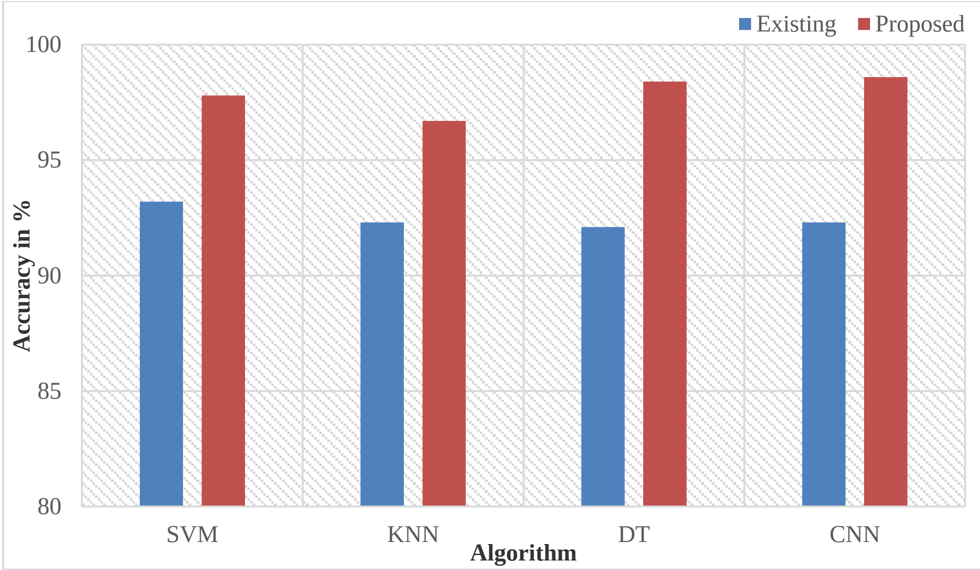


Fig. 8. Comparative analysis of different methods for accuracy.

TABLE III. COMPARATIVE ANALYSIS OF VARIOUS METHODS FOR ACCURACY IN %

ALGORITHM	EXISTING	PROPOSED
SVM	93.2	97.8
KNN	92.3	96.7
DT	92.1	98.4
CNN	92.3	98.6

Table III compares accuracy values for different classification algorithms in existing and proposed systems. The proposed system works better in terms of accuracy. Fig. 8 represents the comparison between existing and proposed work for various algorithms concerning accuracy, and the proposed work outperforms when compared to existing algorithms.

The SVM approach, for instance, has a precision rate of 92.3%. Additionally, the specificity demonstrates how the classifier recognizes a typical node. The accuracy of the CNN approach is 98.6%, which is greater than that of other methods, according to the findings and it has been shown in Fig. 8.

V. CONCLUSION

A network layer assault that mimics routing protocols is called a sinkhole attack. A training dataset is necessary to train models in any training mode to identify sinkhole assaults using machine learning. Real-world situations or exams for categorization can serve as training datasets. The experimental data may be described as a function with a goal value and a descriptive function. 3604 unique samples, both benign and malicious, were gathered for this paper. It creates a dataset that is labeled and composed of eight chosen characteristics. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), and Convolution Neural Network (CNN) are some of the machine learning techniques used in the classification. SVM, DT, and CNN denote excellent accuracy in the following priority. The effectiveness of our technique motivates us to extend this work to tackle the constraints and simulation mentioned in a 3D ad hoc network.

REFERENCES

[1] B. Khalaf, S. Mostafa, A. Mustapha, M. Mohammed and W. Abduallah, “Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods,” IEEE Access., vol. 7, no. 1, pp. 51691–51713, 2019. https://doi.org/10.1109/ACCESS.2019.2908998.

[2] M. Chitkara and M. W. Ahmad, “Review on manet: characteristics, challenges, imperatives, and routing protocols,” Int. J. Comput. Sci. Mob. Comput., vol. 3, no. 3, pp. 432-437, 2014.

[3] M. Sookhak, H. Tang, Y. He and F. R. Yu, “Security and privacy of smart cities: a survey, research issues and challenges,” IEEE Commun. Surv. Tutorials., vol. 21, no. 2, pp. 1718-1743, 2018.

[4] B. Mandal, S. Sarkar, S. Bhattacharya, U. Dasgupta, P. Ghosg et al., “A Review on Cooperative Bait Based Intrusion Detection in MANET,” SSRN., 2020. [Online]. Available: https://bit.ly/3jptCJq.

[5] A. Patcha and J. M. Park, “An overview of anomaly detection techniques: Existing solutions and latest technological trends,” Comput. Networks., vol. 51, no. 12, pp. 3448-3470, 2007.

[6] O. Can, M. O. Unalir, E. Sezer, O. Bursa, B. Erdogdu, “An ontology-based approach for host intrusion detection systems,” In. Proceedings of the research Conference on Metadata and Semantics Research., Tallinn, Estonia, pp. 80-86, 2017. https://doi.org/10.1007/978-3-319-70863-8_8.

[7] F. Al-Dhief, N. Abdul Latiff, N. Noordini Malik, N. Salim, M. Mat Baki et al., “A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms,” IEEE Access., vol. 8, no. 1, pp. 64514–64533, 2020.

[8] J. Bi, H. Yuan and M. Zhou, “Temporal prediction of multiapplication consolidated workloads in distributed clouds,” IEEE Trans. Autom. Sci. Eng., vol. 16, no. 4, pp. 1763-1773, 2019.

[9] J. Bi, H. Yuan, L. Zhang and J. Zhang, “SGW-SCN: An integrated machine learning approach for workload forecasting in geo-distributed cloud data centers,” Inf. Sci., vol. 481, no. 1, pp. 57-68, 2019. https://doi.org/10.1016/j.ins.2018.12.027.

[10] J. Wang, Y. Gao, X. Yin, F. Li and H. J. Kim, “An enhanced PEGASIS algorithm with mobile sink support for wireless sensor networks,” Wireless Commun. Mobile Comput., vol. 2018, no. 9472075, pp. 1-9, 2018. https://doi.org/10.1155/2018/9472075.

[11] G. Somani, M. Gaur, D. Sanghi, M. Conti and R. Buyya, “DDoS attacks in cloud computing: issues, taxonomy, and future directions,” Computer Commun., vol. 107, no.1, pp. 30–48, 2017.

[12] R. Suma, B. G. Premasudha and V. R. Ram, "A novel machine learning-based attacker detection system to secure location aided routing in MANETs," Int. J Networking Virtual Organ., vol. 22, no. 1, pp. 17-41, 2020.

[13] L. Krishnasamy, R. Dhanaraj, D. Ganesh Gopal, T. Reddy Gadekallu, M. Aboudaif et al., "A heuristic angular clustering framework for secured statistical data aggregation in sensor networks," Sensors., vol. 20, no. 17, pp. 4937-4951, 2020. https://doi.org/10.3390/s20174937.

[14] P. Gandotra, R. K. Jha and S. Jain, "A survey on device-to-device (D2D) communication: Architecture and security issues," J. Network Comput. Appl., vol. 78, no. 1, pp. 9-29, 2017.

[15] E. Petersen, M. A. To and S. Maag, "A novel online CEP learning engine for MANET IDS," In. Proceedings of the 2017 IEEE 9th Latin-American Conference on Communications (LATINCOM), Guatemala City, Guatemala, pp. 1-6, 2017. https://doi.org/10.1109/LATINCOM.2017.8240196.

[16] A. Amouri, S. D. Morgera, M. A. Bencherif and R. Manthena, "A cross-layer, anomaly-based IDS for WSN and MANET," Sensors., vol. 18, no. 2, pp. 651, 2018.

[17] M. Abdel-Azim, H. E. Salaha and M. E. Eissa, "IDS Against Blackhole Attack for MANET," Int. J. Network Secur., vol. 20, no. 3, pp. 585-592, 2018. https://doi.org/10.6633/IJNS.201805.20(3).22.

[18] G. Soni and R. Sudhakar, "An IDS Security against Unwanted Flooding of Jamming Attack in MANET," EasyChair Preprint., no. 3789, 2020. [Online] Available: https://bit.ly/3WA1Vw2.

[19] T. Sultana, A. A. Mohammad and N. Gupta, "Importance of the Considering Bottleneck Intermediate Node During the Intrusion Detection in MANET," In. Proceedings of the Research in Intelligent and Computing in Engineering., Thu Dau Mot University, Vietnam, pp. 205-213, 2021. https://doi.org/10.1007/978-981-15-7527-3_20.

[20] S. Mostafa, A. Mustapha, A. Hazeem, S. Khaleefah and M. Mohammed, "An agent-based inference engine for efficient and reliable automated car failure diagnosis assistance," IEEE Access., vol. 6, no.1, pp. 8322–8331, 2018.

[21] G. Gagandeep and A. Aashima "Study on sinkhole attacks in wireless Ad hoc networks," Int. J. Comput. Sci. Eng., vol. 4, no. 6, pp. 1078-1084, 2012.

[22] M. Al-Qurishi, M. Al-Rakhami, A. Alamri, M. Alrubaian, S. Rahman et al., "Sybil defense techniques in online social networks: A survey," IEEE Access., vol. 5, no.1, pp. 1200–1219, 2017.

[23] S. K. Stafrace and N. Antonopoulos. "Military tactics in agent-based sinkhole attack detection for wireless ad hoc networks," Comput. Commun., vol. 33, no. 5, pp. 619-638, 2010. https://doi.org/10.1016/j.comcom.2009.11.006.

[24] A. Razaque and S. Rizvi, "Secure data aggregation using access control and authentication for wireless sensor networks," Comput. Secur., vol. 70, no.1, pp. 532–545, 2017.

[25] R. Srilakshmi and J. Muthukuru, "Intrusion detection in mobile ad-hoc network using Hybrid Reactive Search and Bat algorithm," Int. J. Intell. Unmanned Syst., vol. 10, no. 1, pp. 65-85. https://doi.org/10.1108/IJIUS-09-2020-0049.

[26] A. Umamageswari, S. Deepa and L. S. Beevi, "A novel approach for classification of diabetics from retinal image using deep learning technique," Int. J. Health Sci., vol. 6, no. S1, 2729–2736, 2022. https://doi.org/10.53730/ijhs.v6nS1.5196.

[27] S. Deepa, A. Bhagyalakshmi, V. V. Chamundeeswari and S. G. Winster, "Virtual Image Representation and Adaptive Weighted Score Level Fusion for Genetic Face Recognition," Lect. Notes Electr. Eng., vol 792, 2021. https://doi.org/10.1007/978-981-16-4625-6_77.

[28] Y. Sahu, M. Rizvi and R. Kapoor, "Intruder detection mechanism against DoS attack on OLSR," In. Proceedings of the 2016 Fifth International Conference on Eco-friendly Computing and Communication Systems (ICECCS), Bhopal, India, pp. 99–103, 2016.

[29] M. S. Khan, M. I. Khan, O. Khalid, M. Azim and N. Javaid, "MATF: A multi-attribute trust framework for MANETs," EURASIP J. Wirel. Commun. Netw., vol. 2016, no. 197, pp. 1-17, 2016. https://doi.org/10.1186/s13638-016-0691-4.

[30] R. Jayamma, "Improving the Performances of WSN Using Data Scheduler and Hierarchical Tree," J. Comput. Sci. Intell. Technol., vol. 2, no. 2, pp. 07–16, 2021. https://doi.org/10.53409/mnaa/jcsit/2202.

[31] M. Alqdah, "Intrusion Detection Attacks Classification using Machine Learning Techniques," J. Comput. Sci. Intell. Technol., vol. 2, no. 2, pp. 07–16, 2021. https://doi.org/10.53409/mnaa/jcsit/2201.

[32] Narmatha C. A New Neural Network-Based Intrusion Detection System for Detecting Malicious Nodes in WSNs. J. Comput. Sci. Intell. Technol. 2020; 1(3): 01–08. ©JCSIT, MNAA PUB WORLD, 2020. https://doi.org/10.53409/mnaa/jcsit/2204.

[33] S. Manimurugan, T. Anitha, G. Divya, G. C. P. Latha and S. Mathupriya, "A Survey on Blockchain Technology for Network Security Applications," In. Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT)., Tabuk, Saudi Arabia, pp. 440-445, 2022. https://doi.org/10.1109/ICCIT52419.2022.9711616.

[34] T. Anitha, S. Manimurugan, S. Sridhar, S. Mathupriya and G. C. P. Latha, "A Review on Communication Protocols of Industrial Internet of Things," In. Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT)., Tabuk, Saudi Arabia, pp. 418-423, 2022. https://doi.org/10.1109/ICCIT52419.2022.9711544.

[35] R. Khilar, K. Mariyappan, M. S. Christo, J. Amutharaj, T. Anitha et al., "Artificial intelligence-based security protocols to resist attacks in internet of things," Wireless Commun. Mobile Comput., vol. 2022, no. 1440538, pp. 1–10, 2022.

Friendly Group Architecture for Securely Promoting Selfish Node Cooperation in Wireless Ad-hoc Network

Rajani K C[1], Dr Aishwarya P[2], Manjunath S[3]
Assistant Professor, Department of Artificial Intelligence and Machine Learning, Cambridge Institute of Technology, Bangalore[1]
Prof and HOD, Department of CSE, Atria Institute of Technology, Bangalore[2]
Associate Professor, Department of CSE, Cambridge Institute of Technology, Bangalore[3], India

***Abstract*—Wireless Ad-hoc Network is characterized by a decentralized communication scheme with self-configuring nodes which has witnessed a wide range of practical wireless applications. However, this characteristic also results in various security threats in vulnerable wireless environment irrespective of presence of various routing protocols. Review of existing literatureshows that there is very less emphasis towards securing Dynamic Source Routing (DSR) while majority of solutions uses encryption- based operation. Therefore, this manuscript introduces a novel non-encryption-based scheme called as Friendly Group Architecture which intends to identify the presence of selfish node followed by presenting a method to promote the secure cooperation of it. The complete modelling is analytically designed using probability-based computation and dynamic thresholding. The simulation outcome carried out in MATLAB exhibits that it outperforms existing system with respect to energy, overhead, andsecurity.**

***Keyword*—Wireless Adhoc network; selfish node; DSR; reactive;security**

I. Introduction

A wireless adhoc network is basically a decentralized form of network system where the devices are connected via wireless medium without any dependencies of infrastructure [1]. The decision of routing is undertaken by all the individual wireless nodes carried out dynamically using a routing protocol [2][3]. They can be stated as a self-configuring network with dynamic capabilities. There is no restriction of mobility for such device as the links changes frequency with other connected devices. While doing so, the elementary challenge is to prepare each wireless node to manage information that is demanded for a proper connection [4][5]. This give rises to more challenges as there is a need to forward the data packet to almost all the nodes which comes in the vicinity of its transmission zone. Apart from this, another challenge will be to maintain a required proportion of overhead under control for better routing performance [6]. Although, each nodes is aware of its own data transmission performance but they have no much idea about the demands of others. Apart from this, the dependencies of limited channel capacity are another impediment towards data transmission. Wireless adhoc network is always under a constant threat of attackers owing to its spontaneously changing topology. All the security threats that are present in conventional network are also applicable in wireless ad-hoc network with respect to authentication, confidentiality, integrity, privacy, etc. [7]-[9]. Owing to such security threats, there are an increasing occurrences of intrusion events e.g., inferior monitoring of routing, denial of service attack, injection of counterfeited message, eavesdropping, etc. Essentially, there are two types of routing scheme widely deployed i.e., proactive and reactive protocols. It is found that proactive protocols are (e.g., optimized link state routing, destination sequence distance vector) are more prone to get compromised compared to reactive schemes (e.g., adhoc on-demand distance vector, dynamic source routing). A closer look into existing approaches also showcase that majority of the security scheme is carried out over reactive protocols and currently more research is also dedicated towards securing proactive protocols. However, there are less studies being carried out securing a variant of reactive protocol i.e. Dynamic Source Routing (DSR) to construct on-demand routes while forwarding request of node using source routing. The significant advantage of DSR is its independence from forwarding table update beacon periodically. The overhead is significantly controlled using cache information of route in DSR. However, majority of the security threat in DSR arises from its incapability to repair broken links while lack of updated cache information accelerates to this problem further more. DSR protocol is more prone to flooding attack and it could result in higher delay during connection set up compared to other proactive protocols. Apart from this, DSR protocols is actually meant for static environment and environment with low mobility to some extent [10][11]. They are not suitable for handling communication with higher degree of mobility.

1) Motivation behind the study: With the number of application rising towards the usage of ubiquitous computing, it is necessary that such forms of application should be smart enough to identify threat. However, adoption of sophisticated mechanism may significantly assist in threat detection but at the cost of computational and network resources. Adoption of DSR protocol offers beneficial networking perspective but their mechanism of inherent source routing makes the nodes more vulnerable towards it identity. This is the prime motivational factor to carry out the study towards improving DSR protocol in order to incorporate a capability to perform secure routing operation in presence of routing misbehavior event in wireless adhoc network.

2) Study contribution: Apart from this, it is also seen that irrespective of various studies towards securing wireless adhoc network, there is no benchmarked model which can ascertain this fact. Therefore, the proposed system presents a novel scheme called as Friendly Group Architecture, which uses a simplified analytical model using probability theory exploits the selfish node to promote selfish node cooperation. The study contributes towards a novel modelling of securing DSR protocol by identifying selfish node as well as promotes cooperation of selfish nodes in presence of unknown malicious environment. The core idea of this framework is to balance the security, resource, and overhead demands to enhance DSR protocol in wireless adhoc network. Therefore, the study contribution is as follows:

- The proposed model is developed on the basis of an adversarial model whose information is not predefined with the other normal nodes in network.
- The model presents friendly group architecture which uses both normal and selfish node to participate in data forwarding process.
- The proposed technique introduces a unique incentive allocation scheme which is allocated to all the nodes on the basis of their undertaken action.
- The incentive policy is meant for promoting selfish node to achieve gain if they choose to forward data as a normal node.
- The proposed model is capable of resisting majority of the routing misbehavior in wireless adhoc network with better data transmission performance being noted.

The organization of this paper is as follows: Section II discusses about the existing literatures where different techniques are discussed for detection schemes used in power transmission lines followed by discussion of research problems in Section IIIand proposed solution in IV. Section V discusses about algorithm implementation followed by discussion of result analysis in Section VI and discussion in Section VII. Finally, theconclusive remarks are provided in Section VIII.

II. Related Work

This section briefs of existing approaches towards security in wireless adhoc network especially emphasizing on the work carried out using routing protocols. The work of Almazok et al. [12] have presented an optimized version of DSR protocol using bio-inspired approach as well as time scheduling with main focus on routing performance. Study towards similar direction of improving routing performance is also carried out by Berri et al. [13] where a statistical modelling has been carried out using state of links. Study towards anomaly detection is carried out by Chugh et al. [14] wgere a zone-driven data forwarding scheme has been introduces. Existing study has also discussed about the resistivity techniques against selfish node using reputation based DSR protocol considering the mobility aspect of it as witness in work of Delgado et al. [15]. Hadi et al. [16] have further presented a work which can enhance the detection performance of selfish node under mobility environment using on-demand routing scheme. Study of Liang et al. [17] have presented a DSR scheme by filtering the optimal path for routing to offer more reliable data delivery performance. The work of Mohan Priya et al. [18] have presented a secured DSR scheme in order to resist a specific form of attack i.e. black hole attack. Study towards selfish nature of vehicular network system is investigated by Shan et al. [19] considering static and dynamic nature of the node over multiple environments of communication. The work carried out by Shan et al. [20] has presented a discussion about influence of energy dissipation towards selfish behaviors in mobility environment. Exclusive study towards securing mobility environment is carried out by Srivastava et al. [21] where digital signature is used for authenticating the participating nodes. Azam et al. [22] have studied about various authentication scheme over vehicular adhoc network while work of Faisal et al. [23] have presented a detection of identity attack in wireless adhoc network considering received signal strength. Farahani [24] have used k-nearest method for computing reputation in order to resist black hole attack over mobility environment. Im and Lee [25] have presented a secure covert communication system over two hop adhoc network. Mahmood et al. [26] have studied the issues associated with existing security scheme over vehicular adhoc network and concludes that there are still many issues which require attention in conventional scheme. Naresh et al. [27] have used group key agreement for securing cluster-based communication in adhoc network. Usage of blockchain is reported in work of Ran et al. [28] where on-demand data transmission scheme is used considering QoS constraint. Siddiqui et al. [29] have presented a security approach for resisting wormhole and blackhole attack considering adhoc network integrated with an Internet-of- Security. Wu et al. [30] have presented a secure authentication scheme using key exchange protocol over vehicular adhoc network. Hence, it can be noticed that there are some dedicated investigations towards securing wireless adhoc network. The next section briefs about the issues associated with existing system of secure routing scheme in wireless adhoc network.

1) Limitation and research gap: The prime limiting factors of existing security techniques are mainly associated with the core emphasis on adopting sophisticated technique to identify abnormality in routing. The techniques are either based on data forwarding or it’s based on high-end resource dependent securitytechnique. Another significant limiting factor is that the solution of problem space is more concerned about singular form of attack. Such methodologies make the system non-applicable in different attack scenario. Hence, there is potential research gap between rising of dynamic adversaries and existing problem solution, which is highly symptomatic to specific event. Hence, they cannot be deployed over the scenario which calls for presence of multi-attacker or attackers launching dynamic strategies of attack in wireless adhoc network. Apart from this,there is a still a gap between security solution resiliency and itsdependencies towards resources.

III. Research Problem

The unaddressed problems explored after reviewing the existing system are as follows:

- There are very few potential implementations towards securing DSR protocol over uncertain condition of intrusionin wireless adhoc network.
- Majority of the existing security approaches are based on authentication on priorly known information of the attackerand they are highly specific for attackers.
- The fact that attacker could exhibit dynamic behavior is out of scope of any existing implementation work in wireless adhoc network.
- Existing security approaches doesn't offer much balance between data transmission performance and securityperformance at a same time.

Therefore, the problem statement arrived from the above points is "Developing a novel secure data transmission scheme considering the dynamicity of attacker behaviors and harnessing malicious node to secure communication in wireless adhoc network is challenging task".

IV. Research Methodology

The core target of the proposed system is to introduce friendly group architecture in order to secure DSR protocol in wireless adhoc network. The secondary objective of this architecture is also to ensure better resource management alongwith overhead reduction. The term 'friendly' is stated as this architecture offers identification of selfish node and uses them in order to perform data dissemination in uncertain communication environment in presence of uncertain intruder'sstrategy. Fig. 1 highlights the proposed architecture.

A closer look in Fig.1 shows that proposed DSR implementation scheme is classified into two functionalities i.e., identification of selfish node followed by secure participation of selfish node. In initial operation, all the response-based beacons are analyzed in order to compute probability of cooperation as well as intrusion. These two parameters are further used for computing two empirical forms of trust i.e., degree of conformity which is about data forwarding operation and degree of non- conformity which is about rejecting all possibilities of forwarding data (followed by dropping packet). Further conditions are designed to finalize the trust value from its third type i.e., degree of vagueness in trust, which is about incapability of a node to decide if the other node is regular or selfish node. Finally using dynamic thresholding, selfish node is positively identified. The next round of operation is to further assess the probability of attacker identity for the selfish node followed by another dynamic thresholding in order to allocate incentives. All selfish nodes that comply with proposed secured DSR protocol will be allocated a measurable incentive in order to promote selfish node cooperation. The proposed system ensures that under no circumstances, the selfish nodes once identified as positive attacker will be able to initiate attack next time. The next section discusses algorithms.

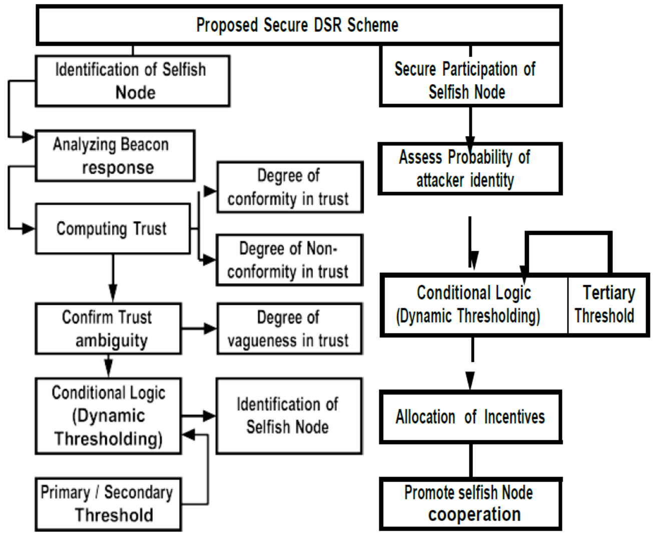


Fig. 1. Proposed friendly group architecture.

V. Algorithm Implementation

This section discusses about the algorithm design and its implication towards developing friendly group architecture with an implicit focus on identification of a selfish node and ensuring participation of a selfish node securely. The novelty of the algorithm implementation is that it exploits the selfish node to participate in data forwarding process even after positively identifying it as a selfish node. This is completely different from any existing system which either isolates the selfish node or takes countermeasure of its non-participation. However, the core idea of this algorithm is to incorporate security in DSR protocolusing retaliation-based concept within the node properties. The discussion of the algorithms is as follows:

A. Algorithm for Identification of Selfish Node

This algorithm is responsible for performing positive identification of a selfish node assuming that it doesn't have anypreviously fed information about its presence. The steps of the algorithm are as follows:

Algorithm for Identification of Selfish Node

Input: n (number of nodes)

Output: n_s (identified selfish node)

Start

1. **For** i=1:n
2. $n_{ad} \rightarrow n_t$(RREP)
3. n_t computes d_c and d_{nc}
4. **If** $d_{nc}>d_c$ & $d_{nc}>T_1$
5. n_s=declare n_{ad} as selfish node
6. **If** (p_c=p_a)
7. compute d_v
8. **If** $d_v>T_2$
9. n_s=declare n_{ad} as selfish node
10. *End*
11. **End**

The algorithm takes an input of n (number of nodes) which after processing yields an outcome of ns (identified selfish node). Considering all the nodes n in simulation area (Line-1), the adjacent node nad forwards the route response beacon RREP to the transmitting node nt after the former receives the route request message from latter (Line-2). Upon receiving this beacon, the transmitting node nt computes degree of conformity in trust dc as well as degree of non-conformity of trust dnc (Line- 3). For a regular form of node, the frequencies of dnc value should reduce or stop which is not the case of malicious form of node. This is because the malicious node could have opted for initially forwarding the regular response in order to increase its individual trust value. In such case, it could become quite impossible for the transmitting node to decide if the adjacent node is regular node or malicious node. Hence, the computation of dc and dnc assist in this regard in the form of mathematical expression as follows:

$$d\text{c}=\text{pc}/\ \text{pc}+\ \text{pa}$$

$$d\text{nc}=\ \text{pa}/\ \text{pc}+\ \text{pa} \qquad (1)$$

In the above expression (1), μ_1 and μ_2 represents probability of cooperation (forwarding data) or probability of intrusion respectively. For any discrete value of p_c and p_a, it's easier to find the case of vulnerability where in such case dnc>>dc. However, if the attacker chooses to increase its trust value by cooperating than in such case, pc= pa which will result in dc=dnc (Line-6). This similar value of degrees will lead to failure of decision of nodes about the regularity and malicious nature. The proposed system considers a primary threshold T_1 which is assigned by user to be compared with d_{nc} (Line-4) to determine its malicious nature (Line-5). However, in case of Line-6, the proposed algorithm further computes degree of vaguess mathematically as follows:

$$d_v=(\lambda.\ P_1)/(P_2.P3) \qquad (2)$$

In the above expression, the variables λ, P1, P2, and P3 represents network coefficient, (pc*pa), $(\text{pc}+\text{pa})^2$, and (pc+pa+1) respectively. This mathematical expression is used only in the condition stated in Line-6. After the dv value is obtained, its is further compared with secondary threshold T2 to ensure that dv should be always within T2 limit (Line-8) otherwise, the monitored adjacent node is termed as malicious node (Line-9). It should be noted that proposed algorithm represents selfish node as malicious node. For regular environment, the value of dv should be reduced and its value is fixed by user based on the application it deploys. Hence, using a simplified probability concept, proposed system can easily identify selfish node ns.

B. Algorithm for Secure Participation of Selfish Node

This algorithm is a continuation of the previous algorithm which detects the selfish node. The prime basis of this algorithm are two folds viz. i) a selfish node will not introduce any form of attack in the preliminary level and will choose to cooperate. As selfish node will not be aware of security protocol running, this is the best way to get them latently introduced within the network, gain trust, and introduce attack, ii) selfish node also participate in data forwarding process; however, they do it with malicious intention. Hence, the above two properties can be harnessed to exploit the data forwarding capability of selfish node as a complimentary to regular node for seamless secure data transmission. However, this operation is strictly monitored on the basis of threat computation for such selfish node and until and unless they are within a permissible limit, the selfish nodes are allowed to propagate data. The success factor of this algorithm completely depends upon how positively the first algorithm works. Apart from this, a second level of controlling is offer, by allocating incentives to the nodes based on their adopted steps of action, the proposed algorithm offers secure participation of selfish nodes. The operational steps of the proposed algorithm are as follows:

Algorithm for Secure Participation of Selfish Node

Input: n_S (selfish nodes)

Output: S_P (secure participation)

Start

1. **For** i=1: *ns*
2. compute p_{aid} of n_S
3. **If** p_{aid}>T3
4. Allocate $I_1 \rightarrow n_S$
5. *Else*
6. S_P=Allocate $I_2 \rightarrow n_S$
7. *End*
8. update S_P, d_V, d_C, and d_{NC}
9. *End*

This above stated algorithm takes the input of ns (selfish nodes) from prior algorithm which after processing ensures an outcome of Sp (secure participation). This algorithm considers all the ns (selfish nodes) (Line-1) followed by conditional assessment to check if the probability of attacker paid is greater than tertiary threshold T3 (Line-3). It should be noted that proposed system performs computation of paid using same expression as that of pa as seen in prior algorithm. If the paid value is found to be more than T3 than it will represent positive presence of many numbers of selfish node. Knowing that fact that a selfish node will need to comply with the proposed secure DSR protocol and hence, its actions will be controlled by allocating an incentive I1, which are computed based on their trust value. The incentive I1 will state allocating of increasing number of profits for selfish node as long as they assist in forwarding data. However, there is also a possibility that the selfish node violates proposed secure DSR protocol and executes its own malicious code. In such case, the malicious node's action is public-ally flagged as a malicious node to all other regular node and in such case; it fails in further data forwarding process to other non-victim regular node. Hence, the selfish node has no other option but to assists in data forwarding or else they will need to isolate themselves from the current network itself (Line- 4). On the other hand, if paid value is found to be lower than tertiary threshold than they are allocated an incentive of I2. It should be noted that I2>I1, which is strategically designed to ensure participation of selfish node (Line-6). All the other variables are further updated and this updated information is shared among all the neighboring nodes (Line-8).

It is to be noted that prime contribution of the proposed secured DSR protocol is basically in the formulation of the incentives which directly control the actions of all nodes on the basis of their actions viz. i) data forwarding, ii) data dropping, iii) raising a notification about intrusion, and iv) introducing intrusion. It could be seen that first two actions could be exhibited by both regular node and selfish node whereas the discrete action of third and fourth could be only exhibited by regular node and selfish node respectively. Hence, on the basis of dv value computation, observing its trend, and comparing with the threshold value. In this process, there is also a likelihood that a regular node could generate false alarm and hence the proposed system introduces an inclusion of penalty factor which is computed for every decision of alarm generated by the regular node. Although, there is a possibility of few instances of false alarm by regular node, but more the regular node updates its variable (Line-8), the occurrences of such false alarm reduces down and network becomes more accurate to capture the event of intrusion. Apart from this, a non-inclusion of conventional cryptography approach is one big advantage of proposed system which not only makes the proposed DSR protocol to offer security but also leverages the data transmission performance in wireless adhoc network.

VI. Result Analysis

This section discusses about the implementation of the proposed system discussed in prior section using MATLAB. The simulation environment consists of randomly distributing 500 nodes in 1000x1000 m2 area with 3000 simulation rounds. The assessment of performance is done using 4 evaluation parameters i.e., overhead, battery usage, identification of selfish node, and consistency in selfish node cooperation. The outcome of proposed system is compared with current work of Delgado et al. [15] termed as DRSR and conventional DSR protocol.

Fig. 2 highlights the traffic overhead reduction which is computed as proportion of control message that lacks application contents. Owing to possession of stale route information, DSR protocols exhibits higher traffic overhead in case of unnecessary cooperation by selfish node, which is avoided to a large scale in DRSR scheme which uses allocation of reputation value for establishing communication. However, DRSR considers static nodes which further doesn't scale up when the iteration is incremented. Proposed system, on the other hand, ensures all its routing decision on the basis of computed probabilities of trust. This ensures higher generation of reliable links causing higher reduction of traffic overhead.

Fig. 3 highlights the battery saving where proposed system excels better energy saving compared to proposed system. Both existing system of DSR and DRSR make use of highly iterative mechanism to find the routes with more emphasis over the destination node proximity and not much into ascertaining the reliability of neighboring node. This is the reason that both the existing schemes doesn't have much significantly different outcomes of battery saving. On the other hand, proposed system has highly structured computation of links, where thresholding, updating operation, and computation of threat probability are spontaneous causing efficient and secure route. This leads to significant energy saving.

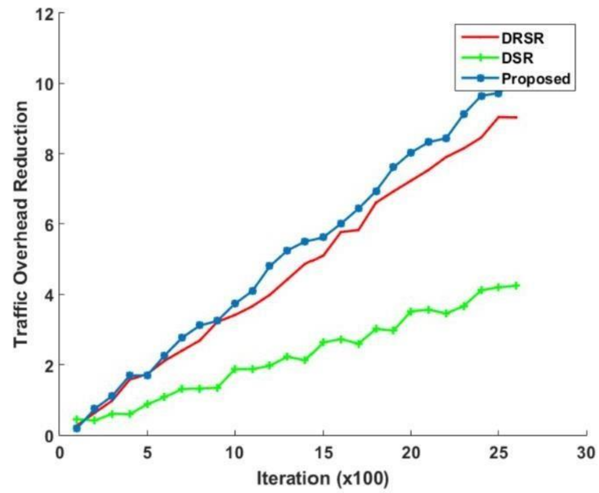


Fig. 2. Analysis of traffic overhead reduction.

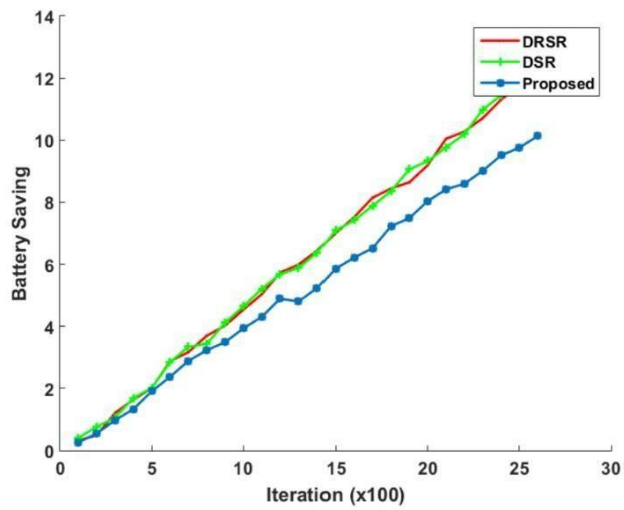


Fig. 3. Analysis of battery saving.

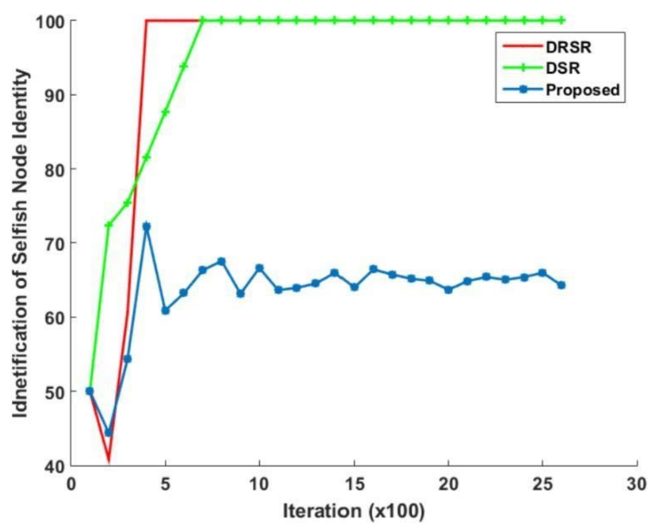


Fig. 4. Analysis of identification of selfish node identity.

Fig. 4 highlights the positive identification of selfish node. It should be noted that proposed system uses three different forms of thresholds in order to confirm the identity of selfish node and therefore, although the trendline of proposed system is lower than existing system, but still the outcome is reliable to consider in presence of uncertain and dynamic environment. This is because, existing DSR and DRSR system showcase higher identification rate as its attackers are well defined in its environment and hence this outcome is not applicable for uncertain intrusion environment.

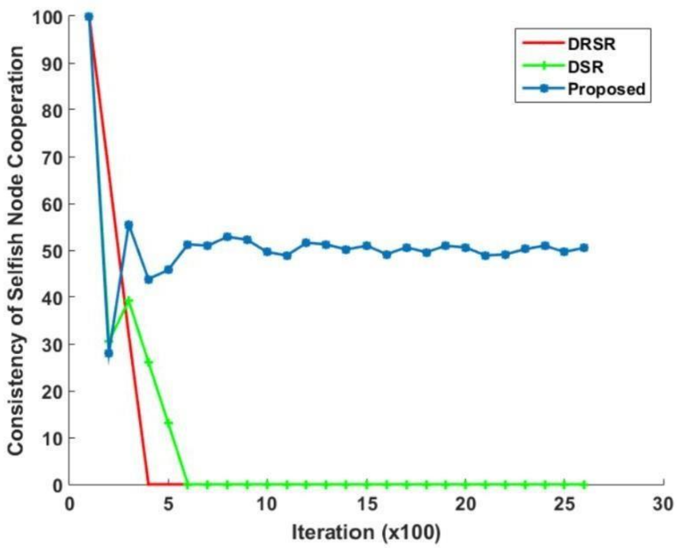


Fig. 5. Analysis of consistency of selfish node cooperation.

Fig. 5 highlights the consistency of the selfish node cooperation. In this case, it is proved that existing system doesn't support selfish node cooperation, where amending the conventional DSR in proposed system logic entails that selfish node has no other choice but to participate in data forwarding process in proposed friendly group architecture. A closer look into the graphical trend shows that proposed system has superior consistency with an increase of iteration. This is because of the fact that proposed algorithm performs spontaneous updates which not only reduces the effort to compute degree of vagueness in trust but it also increases degree of conformity of trust resulting. Allocation of dual form of incentives either forces the selfish nodes to participate till it drains its energy or it isolates the selfish node if the probability of attacker identity is found to be highly significant. Although, this may cause selfish node slightly bypassing the security in preliminary level (0-less than 5 round), however, it soon obtains its consistency and hence proposed system offers more reliable outcomes with respect to selfish node cooperation.

VII. Discussion

From the previous section, it is quite evident that proposed scheme offers better performance in contrast to existing routing schemes. There are multiple points to be highlighted for this perspective of outcomes which are as follows:

- Overhead control is one of the potential contributions of proposed scheme. A closer look into the friendly architecture in joint collaboration with incentive policy states that proposed system carry out its operation on the basis of three different forms of trust scores. This computation is carried for each group in friendly architecture using probability-based conditional logic. Hence, there is much less effort implied by proposed scheme towards trust computation whereas existing DRSR scheme is required to iteratively compute trust score. This fact doesn't suit well in presence of dynamic environment leading to higher traffic overhead in DRSR. At the same time, the higher score of traffic overhead in DSR is accounted for its increasing memory allocation to retain source information.

- From the resource consumption viewpoint, the only operation proposed scheme does are dual fold viz. i) trust computation by neighborhood monitoring and ii) allocation of incentives using conditional logic. The complete operation requires lesser memory, is faster, and is always updated causing better resource retention performance of proposed system with increasing iteration. This is not the case with DSR or DRSR or any other existing scheme briefed in Section II.

- From the security perspective, the proposed scheme emphasized on identification of selfish node on the basis of presented trust computation. The granularity in the trust computation is further ensured by progressive assessment of conditional logic based on dynamic thresholding. This causes uniform performance of selfish node identification over increasing iteration. However, existing scheme e.g. [12]-[25] considers highly sophisticated computation while scheme [26]-[30] offers highly iterative scheme that captures attacker only if attack definition is well defined. Hence, the inconsistency arises in selfish node identification with increasing iteration witnessed with dynamic topology.

- One of the major contributions of the proposed scheme is its prevention technique where the positively identified selfish node is forced to cooperate in network. It is also quite fair enough to ascertain that at certain point of time, such selfish node could violate too. However, there are very less chances for this as if the selfish node chooses to violate, its information have already been updated in hop table of all its neighboring nodes. In such case, selfish node will fail to initiate a new attack in different group. Apart from this, the neighborhood monitoring is a continuous process which is also responsible of capturing any form of anomaly in behaviour of selfish node causing routing misbehaviour.

VIII. Conclusion

The presence of intruder in wireless adhoc network is quite challenging to be explored even by the most potential intrusion detection system. Existing studies carried out considers the predefined information about the attacker doesn't find its applicability over dynamic form of network.

Apart from this, there was so significant security work being carried out considering DSR protocol. These challenges are addressed in current study with following contributions: i) a novel analytical framework of Friendly Group Architecture which promotes the regular node to perform seamless data transmission as well as promotes selfish node for secure cooperation, ii) every actions of selfish nodes are preemptively computed in order to find out its next step of actions and accordingly allocation of incentives are carried out, iii) the proposed model is completely independent of any apriori information of an attacker and hence it can be widely applicable for resisting majority of attacks, iv) the proposed model offers a well-balance between traffic overhead, energy consumption, and security.

REFERENCES

[1] A.H. Wheeb, M.T. Naser, "Simulation based comparison of routing protocols in wireless multihop ad hoc networks", International Journal of Electrical and Computer Engineering, Vol. 11, No. 4,pp. 3186~3192, 2021.

[2] Alamsyah, I K.E. Purnama, E. Setijadi, M. H. Purnomo, "MPR selection to the OLSR quality of service in MANET using minmax algorithm", International Journal of Electrical and Computer Engineering, Vol. 9, No. 1, pp. 417~425, 2019.

[3] Y. Khamayseh, M. B. Yassein, M. Abu-Jazoh, "Intelligent black hole detection in mobile AdHoc networks", International Journal of Electrical and Computer Engineering, Vol. 9, No. 3, pp. 1968~1977, 2019.

[4] J.Mahmood, Z. Duan,Y.Yang, Q. Wang,J. Nebhen, and M.N.M. Bhutta, "Security in Vehicular Ad Hoc Networks: Challenges and Countermeasures", Hindawi Security and Communication Networks, 2021.

[5] Mohammad Al mojamed, "Integrating Mobile Ad Hoc Networks with the Internet Based on OLSR", Hindawi-Wireless Communications and Mobile Computing, 2020.

[6] S.Prabhavat,W.Narongkhachavana,T.Thongthavorn,and C. Phankaew, "Low Overhead Localized Routing in Mobile AdHoc Networks", Hindawi-Wireless Communications and Mobile Computing, 2019.

[7] H. Amraoui, A. Habbani, A. Hajami, and E. Bilal, "Security-Based Mechanism for Proactive Routing Schema Using Game Theory Model", Hindawi Publishing Corporation Mobile Information Systems, 2016.

[8] Y. Mao, C. Zhou, J. Qi, and X. Zhu, "A fair credit-based incentive mechanism for routing in DTN-based sensor network with nodes' selfishness", EURASIP Journal on Wireless COmmunications and Networking, 2020.

[9] Y. Lv,"Security Issues in Multi-hop Device-to-device Communication Networks - Secure Routing Protocols Solution", Journal of Physics: Conference Series, vol.1828, 2021.

[10] H. Yang, "A Study on Improving Secure Routing Performance Using TrustModel in MANET", Hindawi-Mobile Information Systems, 2020.

[11] M. I. Idris, A. Hadi Abd Rahman, P-C Lin, and P. C. K. Hung, "Life Expectancy Analysis of DSR and DSDV Protocol in MANET with Dos Attack", International Journal of Computer Science and Network Security, VOL.20 No.3, March 2020.

[12] S. A. Almazok and B. Bilgehan, "A novel dynamic source routing (DSR) protocol based on minimum execution time scheduling and moth flame optimization (MET-MFO)", EURASIP Journal on Wireless Communications and Networking, 2020.

[13] S. Berri, S. Lasaulce, and M. S. Radjef, "Efficient packet transmission in wireless ad hoc networks with partially informed nodes", EURASIP Journal on Wireless Communications and Networking, vol.148, 2019.

[14] N. Chugh, G. S. Tomar, R. S. Bhadoria, and N. Saxena, "A Novel Anomaly Behavior Detection Scheme for Mobile AdHoc Networks", MDPI-Journalon Electronics, vol.10, 2021.

[15] L. G. Delgado, E. P. Segarra, A. M. Mezher and J. Forné, "A novel dynamic reputation-based source routing protocol for mobile ad hoc networks", EURASIP Journal on Wireless Communications and Networking, vol.77, 2019.

[16] Ahmed. A. Hadi, Zulkarnain Md. Ali, Yazan Aljeroudi, "Improved Selfish Node Detection Algorithm for Mobile Ad Hoc Network", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 4, 2017.

[17] Q. Liang, T. Lin, F. Wu, F. Zhang, W. Xiong, "A dynamic source routing protocol based on path reliability and link monitoring repair", PLOS ONEJournal, vol.16, Iss.5, 2020.

[18] M. Mohanapriya, N. Joshi, M. Soni, "Secure dynamic source routing protocol for defending black hole attacks in mobile Ad hoc networks",Indonesian Journal of Electrical Engineering and Computer Science, Vol. 21, No. 1, pp. 582~590, 2021.

[19] A. Shan, X. Fan, C. Wu and X. Zhang, "Quantitative Study on Impact of Static/Dynamic Selfishness on Network Performance in VANETs," in IEEE Access, vol. 9, pp. 13186-13197, 2021, doi: 10.1109/ACCESS.2021.3051976.

[20] A. Shan, X. Fan, C. Wu, X. Zhang, and S. Fan, "Quantitative Study on the Impact of Energy Consumption Based Dynamic Selfishness in MANETs", MDPI Journal on Sensors, vol.21, 2021.

[21] A. Srivastava, S. K. Gupta, M. Najim, N. Sahu, G. Aggarwal, and B. D. Mazumdar, "DSSAM: digitally signed secure acknowledgement method for mobile ad hoc network", EURASIP Journal on Wireless Communications and Networking, vol.12, 2021.

[22] F. Azam, S. K. Yadav, N. Priyadarshi, S. Padmanaban and R. C. Bansal, "A Comprehensive Review of Authentication Schemes in Vehicular Ad-Hoc Network," in IEEE Access, vol. 9, pp. 31309-31321, 2021, doi: 10.1109/ACCESS.2021.3060046.

[23] M. Faisal, S. Abbas, and H. Ur Rahman, "Identity attack detection system for 802.11-based ad hoc networks", EURASIP Journal on Wireless Communications and Networking, vol.128, 2018.

[24] G. Farahani, "Black Hole Attack Detection Using K-Nearest Neighbor Algorithm and Reputation Calculation in Mobile Ad Hoc Networks", Hindawi-Security and Communication Networks, 2021.

[25] H. -S. Im and S. -H. Lee, "Mobility-Assisted Covert Communication OverWireless Ad Hoc Networks," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1768-1781, 2021, doi: 10.1109/TIFS.2020.3045132.

[26] J.Mahmood, Z.Duan, Y.Yang, Q.Wang, J.Nebhen, and M. N. M. Bhutta, "Security in Vehicular Ad Hoc Networks: Challenges and Countermeasures", Hindawi-Security and Communication Networks, 2021.

[27] V. S. Naresh, S. Reddi, and N.V.E.S. Murthy3, "A provably secure cluster-based hybrid hierarchical group key agreement for large wireless ad hoc networks", Springer Journal on Human-centric Computing and Information Science, vol.9, Iss.26, 2019.

[28] C. Ran, S. Yan, L. Huang, and L. Zhang, "An improved AODV routing security algorithm based on blockchain technology in ad hoc network", EURASIP Journal on Wireless Communication and Networking, 2021.

[29] M. N. Siddiqui, K. R. Malik and T. S. Malik, "Performance Analysis of Blackhole and Wormhole Attack in MANET Based IoT," 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2021, pp. 1-8, doi: 10.1109/ICoDT252288.2021.9441515.

[30] T-Y Wu, Z. Lee, L. Yang, and C-M Chen, "A Provably Secure Authentication and Key Exchange Protocol in Vehicular Ad Hoc Networks", Hindawi-Security and Communication Networks, 2021.

Implementation of Fuzzy Expert System on Skin Diseases

Admi Syarif [1]*, Mayda B Fauzi[2], Aristoteles[3], Agus Wantoro[4]
Department of Computer Science-Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia[1, 2, 3]
Department of Engineering and Computer Sciences, University of Teknokrat Indonesia, Indonesia[4]

***Abstract*—Skin diseases are a group of diseases affecting people of all ages, commonly caused by fungi, bacteria, parasites, viruses, and infections. The disease's main symptoms are usually itching all over the skin. Many patients are often underestimated and embarrassed to consult directly with doctors, which in the end, ignores the symptoms of skin diseases. Since they usually have imprecision symptoms, examining skin diseases is complex and challenging. Recently, many efforts have been made to utilize artificial intelligence approaches for diagnosing various diseases based on the patient's condition. This paper aims to develop a novel fuzzy-based medical expert system based on unprecise existing symptoms. The system uses the specialist Doctor's knowledge (dermatologist) to diagnose and provide the patient's severity level for the disease. We have done numerical experiments using 100 (one hundred) test problems to evaluate the performance of the developed system by comparing the result with the recommendations of doctors (dermatologists). It shows that this system succeeds in all tests with an accuracy value of 95.6%. Thus, this system is very beneficial to support doctors in the assessment of skin diseases.**

Keywords—Artificial intelligence; expert system; fuzzy logic; skin disease

I. Introduction

The skin is the human body's largest organ [1]. Bacteria, viruses, or fungi have usually caused several types of diseases. The statistics indicate that skin diseases are common in half the adult population [2]. In Indonesia, patients with skin diseases are 62.8% men and 37.2% women. With details of the age group 11-20 years 31.4%, in the age group 21-30 years 19.3%, in the age group 31-40 years 14.6% with 76 cases, in the age group 41 -50 years 11.3%, in the age group 51-60 years 7.5%, and the age group > 60 years 3.1%, and the last one in the age group < 10 years 12,8% [3]. Treating skin disease reduces the most negligible effect of disrupting a patient's life quality. Doctors and paramedics often see patients with severe skin diseases without proper treatment. Many patients are often underestimated and embarrassed to consult directly with doctors, which in the end, ignores the symptoms of skin diseases.

Correctly diagnosing a skin disease is essential to patient-centered care. It requires knowledge and thoroughness from expert doctors, but doctors have limitations in consulting with patients because of the tight schedule and many patients. In addition, patients are sometimes embarrassed to consult directly with the Doctor. Many patients underestimate this skin disease, which in the end ignores the symptoms of skin disease.

The treatment therapy given to patients is usually done by giving medicine. Providing the appropriate drugs and the proper dosage is critical for skin disease medication and requires the standard knowledge of pharmacology. Inappropriate or improper dosage would become health problems, such as increased side effects, treatment failure, and resistance [4].

Recently, as computers have become popular, many researchers have developed technology in the form of artificial intelligence that imitates an expert's ability (based on expert knowledge) to make decisions. Expert systems use an expert's knowledge to solve some of our real-life activities, including Interpretation, Prediction, Diagnosis, Design, Planning, Monitoring, and so on [5]-[6]. In medical science, expert systems have been adopted to help doctors and specialists diagnose and get appropriate advice on several medical problems, such as Teeth and Gums, Shortness of Breath in Infants and Children, Eye Diseases, breast cancer, neck pain, and Anemia Diseases. With this expert system, even ordinary people can solve complex issues that usually can only be solved with the help of experts [7].

Each year we see almost ten billion Google Searches related to skin, nail, and hair issues. Two billion people worldwide suffer from dermatologic issues, but there's a global shortage of specialists. While many people's first step involves going to a Google Search bar, it can be challenging to describe what you see on your skin through words alone. Patients consider skin disease a common disease because there is no information about the causes and indications of the disease known to the patient [8]. Many patients experience the adverse effects of skin disease without realizing it, so experts in diagnosing kidney disease are needed. However, because the time specialists have is very limited and the number of patients is large, it is difficult for patients to consult a specialist. It causes patients and specialists to have difficulty conveying information about skin diseases. Therefore, people need an expert system that helps to analyze skin diseases [5].

Another essential issue in diagnosing skin diseases is vague symptoms. Examining skin diseases becomes complex and challenging. Since Zadeh introduced it in 1965, fuzzy logic has been successfully applied for various applications, mainly in control. This 'fuzzy' boom has generated strong interest from researchers in developing new fuzzy-based technology for various applications. Prior to the theory of fuzzy logic, we know as crisp logic, which has a value of true

*Corresponding Author.

and false explicitly [9]. Otherwise, Fuzzy logic is logic that has a value of vagueness or ambiguity (fuzziness) between true and false. In fuzzy logic theory, a value could be true and false simultaneously. People unfamiliar with fuzzy logic would have thought that fuzzy logic is very complicated and unpleasant. However, once people know it, they will be interested and newcomers to studying fuzzy logic [10].

This research aims to develop a novel fuzzy-based medical expert system for diagnosing skin diseases. Based on the imprecise patient's symptoms, the system can produce output depending on the fuzzy inference system's situation. This fuzzy logic-powered system would be an application tool that makes it easier to figure out what might be going on with your skin. By answering your symptoms, the system gives the skin disease diagnosis with a certain confidence level. The AI model analyzes this information and draws from its knowledge to provide you with possible matching diseases. Moreover, it will also show you the severity level and give answers to common questions. The tool is not intended to provide a diagnosis nor be a substitute for medical doctors.

In this study, we used Matlab to develop the system because it is easy to run, performs information retrieval, and assists in calculations. The fuzzy approach is adopted to calculate drug compatibility and types' value. Paramedics and doctors can also use the system to recommend medicine. Numerical experiments using 100 (one hundred) datasets have been conducted to evaluate the effectiveness of the systems. The results are compared with those of the expected results by experts (doctors). It is shown that the system is effective and helps doctors diagnose skin diseases.

We organize this paper into five sections. The literature review of the medical expert system is given in the next section. The third presents a straightforward design of the proposed approach. The numerical experiments and the comparative results are described in the fourth section. Finally, conclusions are drawn in the last session.

II. LITERATURE REVIEW

Several researchers have reported the uses of various Artificial Approaches for medical problems. In the following Table I, several recent AI approaches for various skin diseases are presented.

Based on the above information, it can be seen that most of the previous approaches used a deterministic process. In this research, a novel technique called the fuzzy expert system is developed to diagnose skin disease.

TABLE I. PREVIOUS AI APPROACHES FOR SKIN DISEASES

Id	Authors	Title	Algorithm
1	Hameed et al., 2020 [11]	Multi-Class Multi-Level Classification Algorithm for Skin Lesions Classification using Machine Learning Techniques	Machine Learning
2	Dhiaksa, 2020 [12]	Expert System for Diagnosing Skin Diseases Using the Forward Chaining Method	Forward Chaining
3	Lumini et al., 2020 [13]	A fair comparison of skin detection approaches on publicly available datasets	Convolutional Neural Network (CNN)
4	Yogi & Udjaja, 2018 [14]	Android Application for Detection Of Skin Cancer Using Expert System	Backward Chaining
5	Ma et al., 2017 [15]	Effective Features to Classify Skin Lesions in Dermoscopic images	Support Vector Machine
6	(Waruwu & Simangunsong, 2014 [5]	Application of Certainty Factor Method for Diagnosis Expert System Skin Diseases In Humans	Certainty Factor
7	Sun et al. [16]	A Benchmark for Automatic Visual Classification of Clinical Skin Disease Images	Convo- Lutional Neural Networks (CNN)
8	Alkolifi Alenezi, 2019 [17]	A Method of Skin Disease Detection Using Image Processing and Machine Learning	Machine Learning
9	Rinivasu et al., 2021 [18]	Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM	Deep Learning Neural Networks
10	Srinivasu et al., 2021 [18]	Classification of Skin Disease using Ensemble Data Mining Techniques	Ensemble Data Mining
11	Martiano et al.,2019 [19]	Detection of Potential Skin Cancer Lentigo Maligna Melanoma and Nodular Melanoma with Expert System Using Variable-Centered Intelligent Rule System (VCIRS) Method	Variable-Centered Intelligent Rule System (VCIRS)
12	Pour & Seker, 2019 [20]	Transform domain representation-driven convolutional neural networks for skin lesion segmentation	convolutional neural networks
13	This research	Implementation of Fuzzy Expert System on Skin Diseases	Fuzzy

III. Materials and Methods

This research has several stages of designing an expert system for skin diseases. The process starts with collecting literature studies. Then the second stage is identifying system requirements. The third stage is developing knowledge-based and compiling a membership function to determine the maximum and minimum value of the weight value of each symptom. Next is the development of the system stage. Finally, experiments are done to evaluate the functional system and compare the results with those of experts. The design of the system process in this study is illustrated in the following Fig. 1.

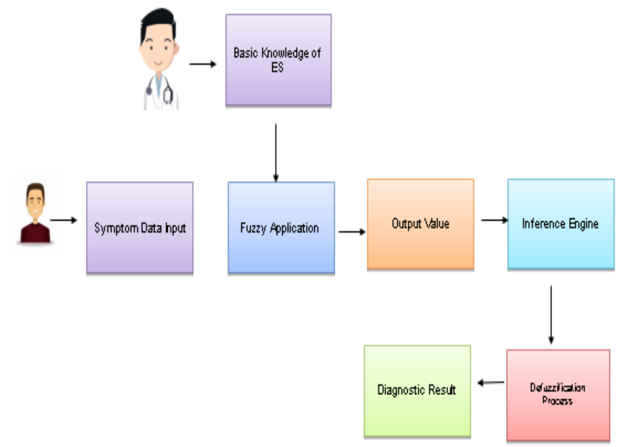


Fig. 1. The design of the system.

A. Symptom and Disease Datasets

This study used 10 (ten) types of skin diseases and 23 (twenty-three) symptoms. The data are given by an expert who is a skin doctor. The types of conditions and the symptoms are presented in Tables II and III. Each sign of the disease will have a weight value. We summarize the weight symptom value of the disease in Table IV.

TABLE II. Skin Disease Data

Code	Skin Disease
D01	Crustose Impetigo
D02	Bullous Impetigo
D03	Neurodermatitis
D04	Cellulitis
D05	Scabies
D06	Candida Cutis
D07	Herpes Zoster
D08	Furuncle
D09	Tinea Versicolor
D10	Tinea Cruris

TABLE III. Skin Disease Symptom Data

S1	S2	S3
Spots appear in the groin area and anus	A rash appears on the facial area	A red rash in the groin area, underarm skin, and breast folds
S4	**S5**	**S6**
Swelling of the skin in the lymph area	Reddish skin blisters on the lower legs, arms, abdomen, or face	Red spots or whitish lesions
S7	**S8**	**S9**
Itchy patches on the wrists, arms, ankles to the genitals or anus	There is a lump	There is a bull that can make wounds
S10	**S11**	**S12**
Have sores that itch	There are red spots	Feeling itchy
S13	**S14**	**S15**
Very itchy, like burning	At night itchy	Feel itchy when sweating
S16	**S17**	**S18**
On one side of the body appears blisters	There is a brownish-yellow scab around the wound as a result of an irritated wound	Skin feels dry, thick, and scaly
S19	**S20**	**S21**
Smooth scaly skin	Fever	Headache or Feeling Dizzy
S22	**S23**	
Pain in the spine	Rash spots on the surface of the skin	

TABLE IV. Symptom Weight Values for the Diseases

Diseases	Symptoms	Weight Value (Percent)
Crustose Impetigo (D01)	S02	30%
	S04	10%
	S17	50%
	S20	10%
Bullous Impetigo (D02)	S02	20%
	S04	10%
	S05	25%
	S09	25%
	S16	10%
	S20	10%
Neurodermatitis (D03)	S10	20%
	S12	20%
	S14	30%
	S15	20%
	S18	10%
Cellulitis (D04)	S01	12,50%
	S02	12,50%
	S03	12,50%
	S04	12,50%
	S05	12,50%
	S09	12,50%
	S16	12,50%
	S20	12,50%
Scabies (D05)	S01	10%
	S03	10%
	S04	5%
	S07	15%

Diseases	Symptoms	Weight Value (Percent)
	S10	10%
	S11	10%
	S12	10%
	S14	15%
	S20	5%
	S24	10%
Candida Cutis (D06)	S01	20%
	S02	10%
	S03	20%
	S10	10%
	S11	20%
	S12	20%
Herpes Zoster (D07)	S01	5%
	S02	7,5%
	S03	5%
	S04	10%
	S05	7,5%
	S09	5%
	S13	15%
	S16	15%
	S20	10%
	S21	10%
	S23	10%
Furuncle (D08)	S04	20%
	S08	70%
	S20	10%
Tinea Versicolor (D09)	S02	5%
	S03	15%
	S06	5%
	S12	35%
	S15	5%
	S19	35%
Tinea Cruris (D10)	S01	25%
	S03	25%
	S10	10%
	S12	20%
	S15	20%

B. Fuzzy Rules

To map the inputs to output, we define several fuzzy rules based on the information given by experts. The fuzzy rules are usually in the form of IF-THEN, made by experts or doctors. The defuzzification process converts the fuzzy output from the system inference engine into a definite value to determine the diagnostic results) [9]. An overview of tree- structure of the expert system on skin diseases is given in the following Fig. 2.

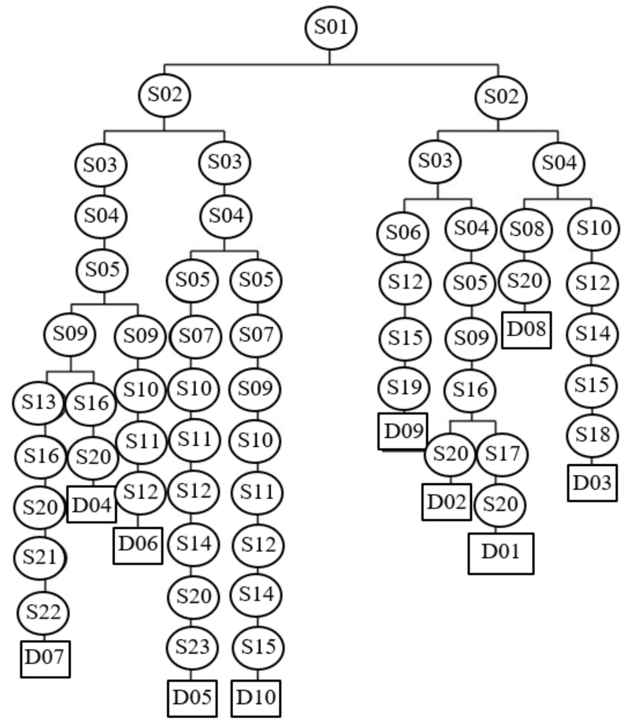


Fig. 2. The tree structure of the system.

C. Fuzzy Logic Controller

The concept of fuzzy logic was then successfully applied to a controller by Zadeh. Recently, it has been used for many applications in our daily life. Thousands of electronic products in the market have adopted the concept of fuzzy logic, ranging from washing machines to high-speed trains. It has been reported that the advantages of fuzzy logic in the sense of performance, simplicity, low cost, and productivity [21]

There are three steps passed before obtaining the crisp value. The processes of fuzzy are as follows:

- Fuzzification is the process of transforming crisp input values into linguistic values, which are usually presented in the form of fuzzy associations with a membership function respectively.
- An interference System is a reference for explaining the relationship between the variables of input and output variables that are processed and the resulting shaped fuzzy. To illustrate the relationship between inputs and outputs, the "IF-THEN" is used.
- Defuzzification is a process of changing the form of fuzzy variables into the data bound (crisp) that can be transmitted to the control equipment.

The step-by-step fuzzy logic process is shown in Fig. 3. The raw input is crisp. The Fuzzification process produces the fuzzy value, which is then calculated by the membership function. Later, the defuzzification process converts it back to the crisp value, as shown in Fig. 3.

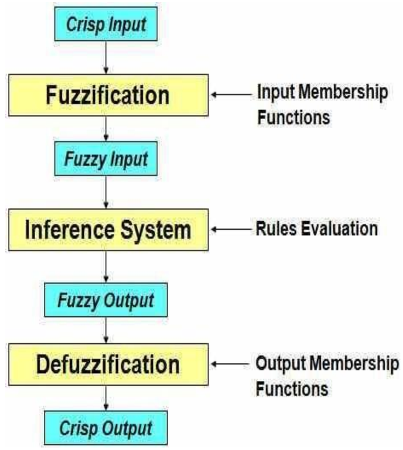
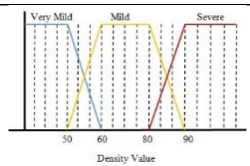
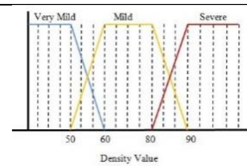
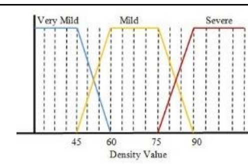
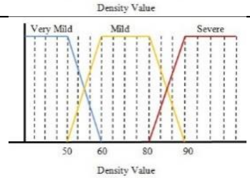
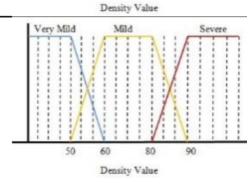
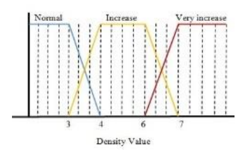
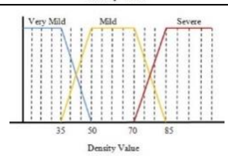
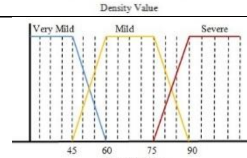
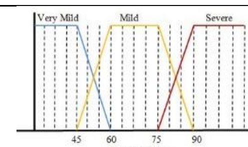
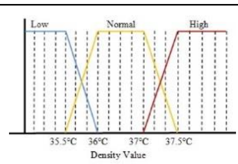
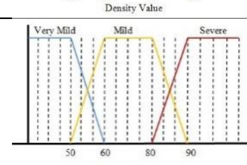
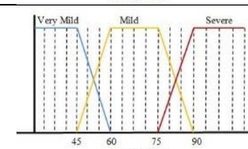
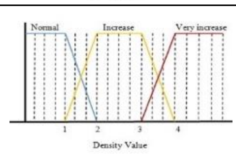
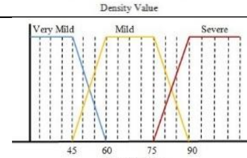
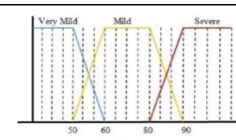
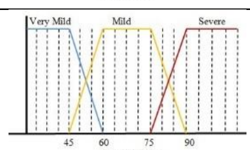
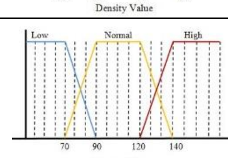
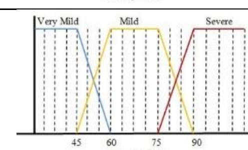
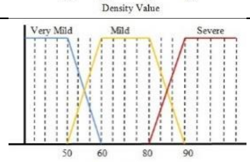
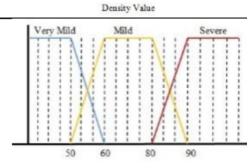
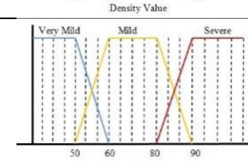


Fig. 3. Step-by-step fuzzy logic approach.

1) Fuzzy membership function: Unlike traditional Crisp Logic, such as Binary Logic, where variables may only take on true and false values represented by 1 or 0, the variables in fuzzy logic may have a truth value in the range of 0 and 1 [22]. The value called membership value explains the degree of truth. Here, the fuzzification process transforms the symptom density value into a linguistic variable using a membership function. The function represents the membership degree of the input variable, denoted by the symbol μ (x), in the interval between 0 on the information given by experts. For this research, we used fuzzy symptoms membership functions, as shown in Table V. We also used criteria and density values for each symptom criterion in this research. The data was obtained from a dermatologist with a practice permit, as shown in Table VI.

TABLE V. MEMBERSHIP FUNCTION OF THE SYMPTOMS

Symptom		**Symptom**		**Symptom**	
S01		S09		S17	
S02		S10		S18	
S03		S11		S19	
S04		S12		S20	
S05		S13		S21	
S06		S14		S22	
S07		S15		S23	

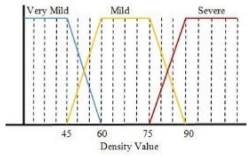
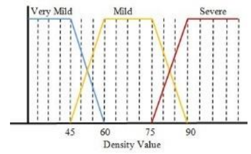
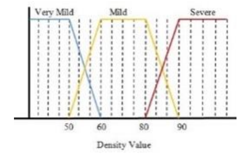
S08		S16		S24	

TABLE VI. THE DENSITY VALUES OF THE SYMPTOMS

Code	Symptom	Criteria	Density Value
S01	Spots appear in the groin and anus area	- Pink spots without freckles	40
		- Pink spots with freckles	55
		- Bright red spots with freckles	70
		- Bright red patches with flaky or scaly spots	82
		- Dark red or black patches with flaky or scaly spots	90
S02	A rash appears on the facial area	- A red rash resembling a wound with nodules filled with clear fluid	78
		- A brownish, crusty rash with nodules filled with clear fluid	85
S03	A red rash in the groin area, underarm skin, and breast folds	- Red rash and red, cracked, dry plaques	75
		- A red rash with pus-filled bumps and red, well-defined plaques	85
S04	Swelling of the skin in the lymph area	- Significant red swelling that feels rough and painful	74
		- Significant red swelling that feels rough and contains pus or blood	88
S05	Reddish skin blisters on the lower legs, arms, abdomen, or face	- The pink skin blisters feel swollen and warm and look shiny	65
		- Red skin blisters that are swollen and inflamed (painful)	78
		- Red skin blisters with bubbles filled with fluid or pus	85
S06	There are red spots or whitish lesions	- Pink lesions	63
		- Red lesions	78
		- Whitish patches	85
S07	There are itchy patches on the wrists, arms, ankles, genitals, or anus	- Itchy patches with red or darker than the surrounding skin	65
		- Itchy patches feel rough and scaly and thick	77
		- Itchy patches with open sores that can lead to infection	85
S08	There is a lump	- The bumps are red and not hard	50
		- Red bumps that are inflamed and hard-filled with pus	74
		- White bumps like a pile of pus, and around the lump is red	85
S09	There is a bull that can make wounds	- There are several bullies in certain body parts	65
		- There are many red bullies like hives that feel itchy	85
S10	Have sores that itch	- The wound is starting to itch	65
		- Burns that feel hot and itchy	78
		- Burns that feel hot, very itchy, and scaly	85
S11	There are red spots	- There are some small red nodules	60
		- There are rashes and red skin blisters filled with clear fluid	77
		- There are nodules and red, inflamed skin blisters filled with clear fluid	85
S12-	Feeling itchy	- itchy	60
		- a bit itchy	78
		- very itchy	85
S13	Very itchy, like burning	- itchy like burning	60
		- a bit itchy, like burning	78
		- very itchy, like burning	85

Code	Symptom	Criteria	Density Value
S14	At night itchy	- itchy at night	60
		- a bit itchy at night	78
		- very itchy at night	85
S15	Feel itchy when sweating	- itchy when sweating	60
		- a bit when sweating	78
		- very itchy when sweating	85
S16	On one side of the body appears blisters	- Red blisters that start to become inflamed and itchy	70
		- Red blisters are painful, itchy, and filled with fluid	83
		- Red blisters that fill with fluid and break into crusts	88
S17	There is a brownish-yellow scab around the wound as a result of an irritated wound	- Yes	85
		- No	0
S18	Skin feels dry, thick, and scaly	- Dry skin	60
		- Skin feels dry and flaky	78
		- Skin feels dry, thickened, and scaly	85
S19	Smooth scaly skin	- Yes	85
		- No	0
S20	Fever	- Low	35 C - 36 C
		- Normal	36,1 C - 37,5 C
		- High	37,6 C - 38 C
S21	Headache or Feeling Dizzy	- Feeling sick at one point	65
		- Like spinning	78
		- Feels floating or head feels heavy	85
S22	Pain in the spine	- Feel pain	65
		- Feel a bit of spinal pain	78
		- Feel very of spinal pain	85
S23	Rash spots on the surface of the skin	- Rash of pink spots	65
		- A rash of bright red spots that start to become inflamed	78
		- A rash of red, inflamed spots that feels rough	88

IV. THE NUMERICAL EXPERIMENTS AND RESULTS

We develop a fuzzy-based expert system for diagnosing kidney diseases using MATLAB R2020a and run on PC Processor Intel® Core™ i3-7020U. For the experiments, we used 100 (one hundred) test problems from the Abdoel Moeloek National Hospital in Lampung. We compared the results of the system with those given by the Doctor.

This system first displays questions from each symptom according to the closest symptom. After selecting the related symptom, the system will show a selection of symptom criteria. Finally, the diagnostic results will be displayed. After choosing the symptoms, the system will present the results as illustrated in Fig. 4.

The system's output will be the diagnostic results and its confidence level. We compared the results with those given by the expert (dermatologist) and the system in Table VII. These results show that the system succeeds in diagnosing the patient's disease for all tests. It can analyze the conditions precisely the same as given by experts.

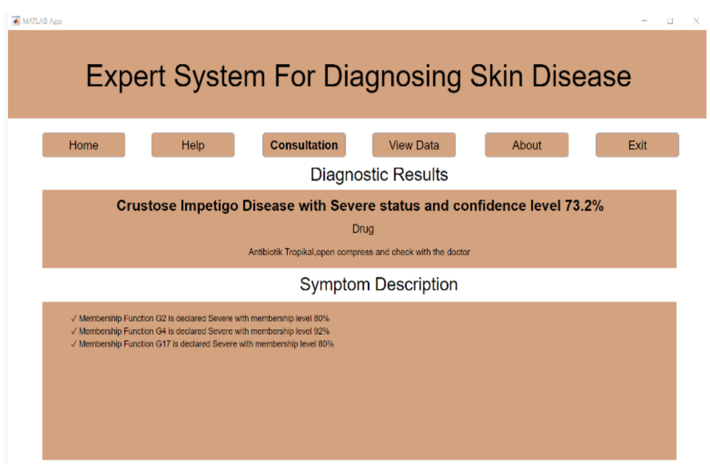


Fig. 4. User Interface of the system.

TABLE VII. COMPARATIVE RESULTS BETWEEN EXPERTS AND SYSTEM

		Predicted by the System									
	Type of Disease	**D01**	**D02**	**D03**	**D04**	**D05**	**D06**	**D07**	**D08**	**D09**	**D10**
Actual by the expert	D01	5	0	0	0	0	0	0	0	0	0
	D02	0	12	0	0	0	0	0	0	0	0
	D03	0	0	8	0	0	0	0	0	0	0
	D04	0	0	0	13	0	0	0	0	0	0
	D05	0	0	0	0	13	0	0	0	0	0
	D06	0	0	0	0	0	13	0	0	0	0
	D07	0	0	0	0	0	0	13	0	0	0
	D08	0	0	0	0	0	0	0	2	0	0
	D09	0	0	0	0	0	0	0	0	13	0
	D10	0	0	0	0	0	0	0	0	0	8

TABLE VIII. NUMERICAL EXPERIMENT RESULTS

				System		
Disease	**Problem**		**Status**	**Very Mild**	**Mild**	**Severe**
D01	5		**Very Mild**	0	0	0
			Mild	0	1	0
			Severe	0	0	4
D02	12		**Very Mild**	0	0	0
			Mild	0	6	0
			Severe	0	0	6
D03	8		**Very Mild**	0	0	0
			Mild	0	2	0
			Severe	0	1	5
D04	13		**Very Mild**	0	0	0
			Mild	0	6	0
			Severe	0	0	7
D05	13		**Very Mild**	0	0	0
			Mild	0	6	0
			Severe	0	0	7
D06	13	**Expert**	**Very Mild**	0	0	0
			Mild	0	5	0
			Severe	0	0	8
D07	13		**Very Mild**	0	0	0
			Mild	0	6	0
			Severe	0	0	7
D08	2		**Very Mild**	0	0	0
			Mild	0	0	0
			Severe	0	0	2
D09	13		**Very Mild**	0	0	0
			Mild	1	4	1
			Severe	0	2	5
D10	8		**Very Mild**	0	0	0
			Mild	0	3	0
			Severe	0	0	5

Next, for each test problem, the system also provides the severity level for each patient. Table VIII shows the comparative results of experts and the developed system on the disease severity for the patients. Based on these results, we calculate the system's accuracy in determining the patients' disease severity, as shown in Fig. 5.

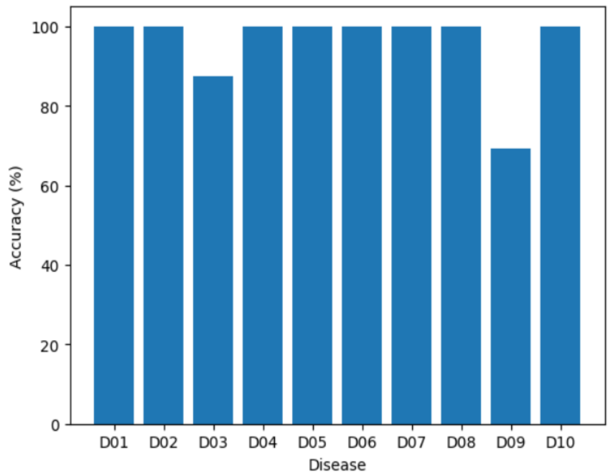


Fig. 5. Expert diagnostic test chart and system.

The above results show that the system's accuracy in determining the severity level of the patient's disease is 95.6%. It shows that this system is effective in diagnosing skin disease. Thus, the system is very beneficial to support doctors in assessing skin diseases.

V. Conclusion

The diagnosis of skin diseases is a difficult task. There is always some possibility of misdiagnosis that leads to miss treatment. This research presents a fuzzy expert system to examine skin diseases based on patient symptoms. The system gives information about the type of disease with confidence based on the existing symptoms and how to handle them. The system's output will be the type of disease and its attributes, namely very mild, mild, and severe. We had several experiments to evaluate the system's performance using 100 test problems and compared the results with those given by the expert (dermatologist). The result shows that the system succeeds in all of the tests. Thus, this system is very beneficial to supporting doctors in the assessment of skin diseases.

Acknowledgments

The authors express their deepest gratitude for using the facilities of the Basic Computing Laboratory, Department of Computer Science, and the University of Lampung to support this research and for the willingness of dr. Hendra Tarigan Sibero, Sp.KK, M. Kes. who has contributed his practical knowledge to this skin disease expert system. A "Fundamental" research Grant supported this research.

References

[1] E. A. Grice and J. A. Segra, "The skin microbiome," Nat. J., vol. 588, no. 4, pp. 244–253, 2013, doi: 10.1038/d41586-020-03523-7.

[2] K. Chiller, B. A. Selkin, and G. J. Murakawa, "Skin microflora and bacterial infections of the skin," J. Investig. Dermatology Symp. Proc., vol. 6, no. 3, pp. 170–174, 2001, doi: 10.1046/j.0022-202x.2001.00043.x.

[3] S. Grover, R. K. Ranyal, and M. K. Bedi, "A Cross Section of Skin Diseases in Rural Allahabad," Indian J. Dermatol., vol. 53, no. 4, pp. 67–70, 2008, doi: 10.4103/0019-5154.44789.

[4] R. N. & colin A. G. P.Wyatt, J., Illingworh, Oxford Handbook of Emergency, Fourth. New York: OXFORD University Press, 2012.

[5] S. K. Waruwu and A. Simangunsong, "Application of Certainty Factor Method For Diagnosis Expert System Skin Diseases In Humans," J. Comput. Networks, Archit. High Perform. Comput., vol. 2, no. 2, pp. 191–194, 2014.

[6] Aristoteles, K. Adhianto, R. Andrian, and Y. N. Sari, "Comparative analysis of cow disease diagnosis expert system using Bayesian network and Dempster-Shafer method," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 4, pp. 227–235, 2019, doi: 10.14569/ijacsa.2019.0100427.

[7] Y. Kumar and Y. Jain, "Research Aspects of Expert System," Int. J. Comput. Bus. Res., no. 6, p. 11, 2012.

[8] J. Kent, "Google Unveils Artificial Intelligence Tool for Dermatology," Health IT Analytics, US, pp. 1–4, 2021.

[9] P. Baranyi, T. D. Gedeon, and L. T. Koczy, "A general interpolation technique in fuzzy rule bases with arbitrary membership functions," IEEE, no. 6, pp. 510–515, 1996, doi: 10.1109/icsmc.1996.569844.

[10] C. Wang, "A Study of Membership Functions on Mamdani-Type Fuzzy Inference S," Theses Diss., vol. January, pp. 20–30, 2015.

[11] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. Hossain, "Multi-Class Multi-Level Classification Algorithm for Skin Lesions Classification using Machine Learning Techniques Nazia Hameed," Expert Syst. Appl., vol. 141, pp. 1–33, 2020, doi: oi.org/10.1016/j.eswa.2019.112961.

[12] A. Dhiaksa, "Expert System for Diagnosing Skin Diseases Using the Forward Chaining Method," J. Intell. Decis. Support Syst. 3, vol. 3, no. 1, pp. 19–26, 2020, [Online]. Available: www.idss.iocspublisher.org.

[13] A. Lumini, A. Lumini, and L. Nanni, "Fair comparison of skin detection approaches on publicly available datasets," Expert Syst. Appl., vol. 160, 2020, doi: https://doi.org/10.1016/j.eswa.2020.113677.

[14] Yogi and Udjaja, "Android Application for Detection of Skin Cancer using Expert," Soc. Econ. Ecol. Int. J., vol. 2, no. 1, pp. 1–8, 2018.

[15] Z. Ma, J. Manuel, and R. S. Tavares, "Effective Features to Classify Skin Lesions in Dermoscopic images," Expert Syst. Appl., vol. 84, pp. 92–101, 2017, doi: doi.org/10.1016/j.eswa.2017.05.003.

[16] X. Sun, J. Yang, M. Sun, and K. Wang, "A Benchmark for Automatic Visual Classification of Clinical Skin Disease Images," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9914 LNCS, p. V, 2016, doi: 10.1007/978-3-319-46466-4.

[17] N. S. Alkolifi Alenezi, “A Method of Skin Disease Detection Using Image Processing and Machine Learning,” Procedia Comput. Sci., vol. 163, pp. 85–92, 2019, doi: 10.1016/j.procs.2019.12.090.

[18] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," Sens. J. MDPI, vol. 21, pp. 1–27, 2021, doi: 10.3390/s21082852.

[19] H. F. Martiano and C. Setianingsih, "Detection of Potential Skin Cancer Lentigo Maligna Melanoma and Nodular Melanoma with Expert System Using Variable-Centered Intelligent Rule System (VCIRS) Method," 2019 6th Int. Conf. Instrumentation, Control. Autom., vol. 01, no. August, pp. 42–46, 2019.

[20] M. P. Pour and H. Seker, "Transform domain representation-driven convolutional neural networks for skin lesion segmentation," Expert Syst. Appl., p. 113129, 2019, doi: 10.1016/j.eswa.2019.113129.

[21] L. A. Zadeh and S. Jose, "The Concept of a Linguistic Variable II," Electr. Eng., vol. 357, pp. 301–357, 1975, doi: 10.1038/s41366-018-0105-2.

[22] T. Sugiura, N. Sugiura, T. Kazui, and Y. Harada, "A self-tuning effect of membership functions in a fuzzy-logic-based cardiac pacing system," J. Med. Eng. Technol., vol. 22, no. 3, pp. 137–143, 1998, doi: 10.3109/03091909809062480.

Dual U-Net with Resnet Encoder for Segmentation of Medical Images

Syed Qamrun Nisa Author[1], Amelia Ritahani Ismail[2]*
Department of Computer Science-Kulliyyah of Information and Communication Technology,
International Islamic University Malaysia, P.O. Box 10, 50728, Kuala Lumpur, Malaysia

***Abstract*—Segmentation of medical images has been the most demanding and growing area currently for analysis of medical images. Segmentation of polyp images is a huge challenge because of the variability of color depth and morphology in polyps throughout colonoscopy imaging. For segmentation, in this work, we have used a dataset of images of the gastrointestinal polyp. The algorithms used in this paper for segmentation of gastrointestinal polyp images depend on profound deep convolutional neural network architectures: FCN, Dual U-net with Resnet Encoder, U-net, and Unet_Resnet. To improve the performance, data augmentation is performed on the dataset. The efficiency of the algorithms is measured by using metrics such as Dice Similarity Coefficient (DSC) and Intersection Over Union (IOU). The algorithm Dual U-net with Resnet Encoder obtains a higher DSC of 0.87 and IOU of 0.80 and beats the other algorithms U-net, FCN, and Unet_Resnet in segmentation of gastrointestinal polyp images.**

Keywords—Segmentation; Medical Images; Deep Convolutional Neural Network; FCN; U-net; Unet_Resnet; Dual U-net with Resnet Encoder

I. Introduction

Image segmentation is one of the most widely and effectively used techniques for image analysis. Image analysis is a method of extracting data from images by analyzing the features within an image. There are a variety of image processing techniques that are used for image analysis, including edge detection, image preprocessing, and image segmentation[1]. The fundamental purpose of image processing is to improve the image or extract relevant information from the images[2]. Segmentation is a crucial yet challenging aspect of image processing for image analysis. Image segmentation techniques are now evolving in a faster and more precise approach[3]. Segmentation is the technique of splitting digital images into several segments to improve image quality.

Segmentation of medical images, as an emerging medical image processing method, has made a significant contribution to long-term medical treatment. For analysis of images in medical, segmentation is performed to locate the area of interest[4]. Before an illness could be diagnosed, medical images must go through numerous processes. Initially, images are gathered, then the pre-processing is performed and after that data must be stored in memory. That demands a significant memory space and a processing time. It is required to process images in medical applications in order to find relevant information[5].

Polyps in the gastrointestinal are abnormal growths of cells in the stomach and colonic mucosa. This abnormal growth occurs gradually and, in most instances, it does not create symptoms till it achieves a significant magnitude[6]. One of the most common causes of gastroenterology is a polyp, which could develop into colorectal cancer[7]. Nonetheless, cancer is preventable and treated, if polyps are diagnosed early[6]. Segmentation of gastrointestinal polyps is a challenging process because of the variation in color intensity and form of gastrointestinal polyps in colonoscopy images. Various techniques have been developed in order to achieve accurate segmentation[8].

In this work, we have applied Deep Convolutional Neural Network algorithms including Dual U-Net with Resnet encoder, U-Net, Unet_Resnet, and FCN to do segmentation on the images of gastrointestinal polyps gathered from the dataset of 'Kvasir-SEG'. The performances are evaluated by using the metrics: Dice similarity coefficient and Intersection over Union. Also, the comparison of these algorithms of segmentation is performed on the bases of metrics.

The rest of the paper is structured as follows: Section II discusses the works related to our study; Section III gives overview of architecture of Dual U-Net with Resnet encoder; Section IV explains the design of methodology for this work; Section V presents the result and discussion; Section VI gives conclusion and presents the future work.

II. Related Work

Medical images are crucial and assume a fundamental part in monitoring and diagnosing the status of a patient's wellbeing[9]. For clinical investigation, diagnosis, and treatment planning, medical imaging including X-Rays, Computed Tomography (CT), Positron Emission Tomography (PET), Ultrasound, and Magnetic Resonance Imaging (MRI) are utilized to provide a clear depiction of the interior of the body[10]. The use of medical imaging techniques has expanded in recent years as a result of technological advancement[11]. Medical images have rich properties, these are images that include a massive amount, high resolution, and complicated features[12]. These medical images are utilized and saved for diagnostic and research purposes on a regular basis[13].

Image analysis is described as a process of extracting quantitative information from images by measuring objects within them. Image analysis has proven particularly valuable for industrial and scientific applications due to its capacity to

*Corresponding Author.

process digital images and objectively analyze parameters such as distance, size, color, number of particles, and so on without affecting the sample[14]. For image analysis, the major image processing approaches used are pre-processing, edge detection, compression, and segmentation. Pre-processing of images often entails eliminating low-frequency background noise, leveling the intensity of individual particle images, and erasing or improving data images before the computational processing. Image compression is a type of data compression in which the original image is encoded with a few number of bits The aim of image compression is to eliminate the image redundancy and store or transmit data in a more efficient manner[1]. Edge detection is a technique for detecting the boundaries of objects or regions that are closely related. The discontinuity of the object is identified using this technique. The edge detection technique is mostly used in image analysis to identify areas of an image where there is a large fluctuation in intensity[15]. Image segmentation is one of the primary phases of image processing where each image is partitioned into several parts, each of which contains some type of information[16].

Image segmentation is an important step in analyzing an image. Image segmentation is the first step in analyzing and extracting information from images[17]. Image segmentation is an image processing technique that partitions an image into contiguous parts[18], that is commonly used to detect objects and borders in images such as lines, and curves[19]. The various segmentation techniques for image segmentation are edge detection, thresholding method, region-based segmentation, clustering-based segmentation[20], and deep neural network-based segmentation[21]. Segmentation of medical images is an essential phase in the examination and evaluation process. For medical images, segmentation is performed to identify and extricate characteristic regions. The purpose of the segmentation of medical images is to identify the tumor location, detect the lesion and other abnormalities, help in treatment planning prior to radiation treatment[22], and help in improving the quality of medical images[4].

Banerjee, et al. [23], proposed an algorithm that clusters pixels into four sections depending on their intensity. This approach employs a thresholding mechanism with one global threshold and two local thresholds. The new aspect of this procedure is the automated establishment of three thresholds based on inherent features of the image. The global threshold is calculated by comparing the neighborhood of the image's localized parts. The global threshold divides pixels into two sets of groups, which are then used to determine the local thresholds. For evaluation of the efficiency of the proposed algorithm, the resulting images are compared to Berkeley's dataset benchmark image. The results are visually examined to determine the region identified by the proposed algorithm. All the regions mentioned in the benchmark images are successfully identified by the proposed algorithm.

Cao, et al. [24], proposed an algorithm for segmentation of the sequence of CT images of the whole heart. They have developed two improved segmentation algorithms based on traditional region-based and edge-based segmentation algorithms and then the improved algorithms are implemented to the whole heart CT sequence images for segmentation. The evaluation of the proposed segmentation algorithm is based on three features in this study: the algorithm's efficiency, subjective evaluation of whole heart image segmentation, and the deviation method, in which the Jaccard and Dice coefficients are used to evaluate the segmentation results of the whole heart CT images. The results of the two improved segmentation algorithms are contrasted with the manual segmentation results, the improved algorithms of segmentation demonstrate superior efficiency in terms of reliability and accuracy of CT image segmentation of the whole heart.

Substantial advancement has been accomplished in recent years in establishing more productive and precise segmentation algorithms for the medical image in machine learning[25]. The machine learning field has been overpowered with a variety of deep learning-based techniques. Deep learning-based techniques such as Recurrent Neural Networks (RNNs), Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNNs) are efficient in image segmentation[26]. Convolutional Neural Network (CNN) works well on medical image challenges, convolutional plays an important role in performance in CNN in image analysis techniques such as segmentation and classification[4]. In a deep neural network, Convolutional Neural Network has several architectures. AlexNet, VGG, FCN, SegNet, Resnet, and U-Net are some of the architectures of Convolutional Neural Networks[27]. AlexNet, VGG, and Resnet are widely regarded as the most prevalent architecture that successfully performs image classification and image recognition tasks[28].

A Fully Convolutional Neural Network (FCN) is a segmentation method that can make a prediction on the bases of pixel-by-pixel using the image's ground truth and directly output the label map. The objective of FCN is to extract significant feature maps and restore them to the labels of the image. The capability to identify features is the fundamental cause of FCN's effectiveness in image segmentation, object detection, and classification. FCN can function effectively only if a sufficient amount of training datasets is accessible. Data gathering is costly in the realm of medical imaging. Furthermore, there are several other aspects that affect data availability, such as privacy, regulation, and ethical concerns[29]. Without enough availability of dataset, FCN provides less accurate results[30] and results in inaccurate segmentation of small organs[31]. As a consequence, FCN has been enhanced to U-Net in order to conduct medical image segmentation, which performs effectively with a relatively small amount of training dataset[29]. SegNet is another architecture of CNN for the segmentation task of images, this architecture has significant features of memory and efficient performance[32]. However, during the training phase, U-Net exhibits a faster processing speed[33], and also U-Net performs better compared to the SegNet[34].

U-Net is among the most widely used architectures of deep Convolutional Neural Networks for segmentation tasks of medical images[35]. The downsampling (contracting path), and the upsampling (expensive path) are the primary components of the U-Net network[36]. The U-Net architecture is shown in Fig. 1.

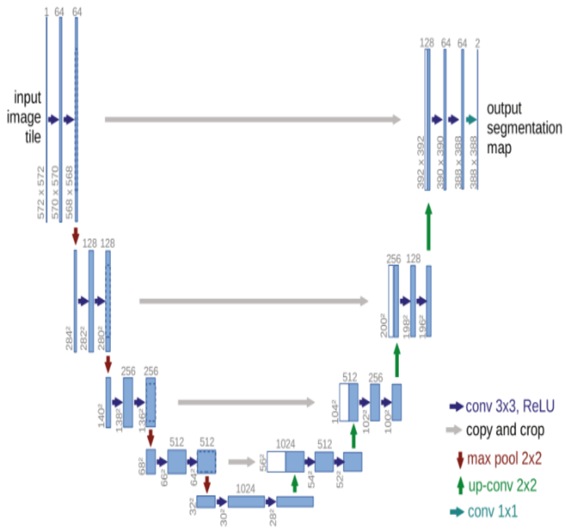


Fig. 1. The architecture of U-Net [37].

Fig. 1 depicts the architecture of U-Net which comprises of a contracting path/downsampling (left side) and an expansive path/upsampling (right side). The convolutional network design is followed throughout the downsampling path. It consists of two 3x3 convolutions, each of which is accompanied by an activation function, a rectified linear unit (ReLU), and a 2x2 pooling with stride 2 for downsampling. The number of channels gets doubled during downsampling. The upsampling or expansive path comprises of 2x2 convolutional that reduces the feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and it also has two 3x3 convolutional layers, each of which is trailed by Rectified Linear Unit activation function. At the final layer, a 1x1 convolutional layer is utilized to map the features to the necessary number of classes[37]. The U-Net architecture has sparked a lot of attention in medical image segmentation, and various variants have been developed based on it[31].

In [38], the authors developed a 3D U-Net, a convolutional neural network technique, for whole heart segmentation. The proposed algorithm comprises two 3D U-Net architectures, the first part of architecture locates the bounding box around the heart and the second part of architecture is to perform the segmentation. This algorithm is evaluated on 20 3D CT images. The dataset of 20 2D CT images is partitioned into two sets: 15 training images and 5 validation images. Because of less dataset, the augmentation procedure is utilized to increase the amount of dataset for training to increase the performance. The performance of the algorithm is calculated by using the Dice coefficient. The mean dice score of 89% is obtained by the proposed model for the whole heart segmentation.

In [39], the authors proposed a U-net segmentation architecture for the CT images of lungs. The dataset presented includes lung cancer thoracic computed tomography (CT) scans with depicted lesions. For the ground truth, manual segmentation for lung parenchyma was performed prior to the experiment. Cropping of images is conducted after manual segmentation to remove details that are not relevant for their analysis. The network's performance is evaluated by the dice coefficient index. They have achieved a 0.9502 dice coefficient for the segmentation.

III. Architecture of Dual U-Net with Resnet Encoder

The architecture of Dual U-Net with Resnet encoder is illustrated in Fig. 2.

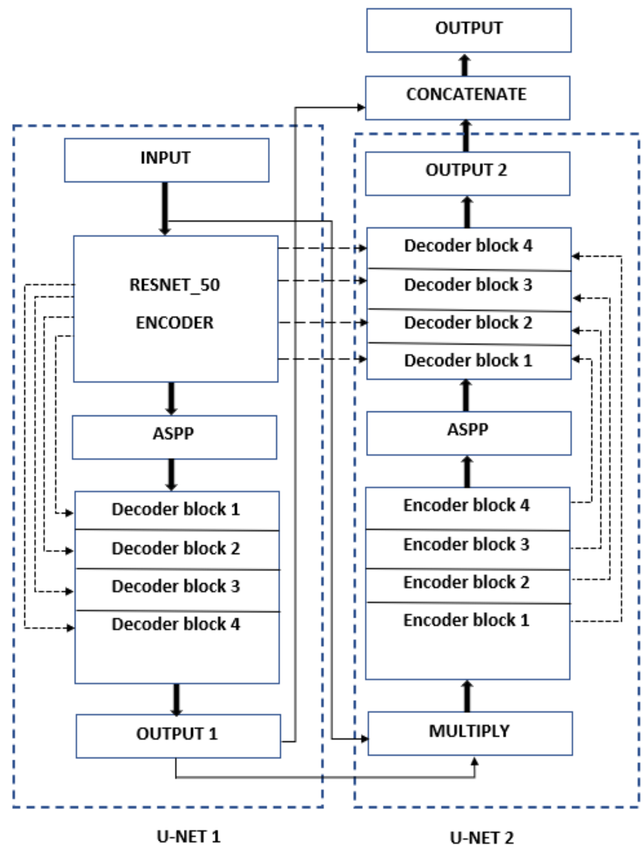


Fig. 2. Architecture of dual U-Net with resnet encoder.

Fig. 2 depicts the architecture of Dual U-Net with Resnet encoder. This architecture consists of two U-net: U-NET 1 and U-NET 2, and an encoder Resnet_50 sub-network. In U-NET 1, the application of Resnet_50, ASPP, and decoder block differentiates the Dual U-Net with Resnet encoder from U-Net. An element by element multiplication is carried out between the results of U-NET 1 with the input of the same network. In U-NET 2, the main distinction between U-Net and Dual U-Net with Resnet encoder is the use of ASPP. The first encoder in Dual U-Net with Resnet encoder purposes pre-trained Resnet_50, while the encoder in U-NET 2 was created from the beginning. The encoder U-NET 2 performs two 3x3 convolution operations, each of which is followed by a batch normalization. An activation function Rectified Linear Unit (ReLU) is applied, which is accompanied by a squeeze and excitation block, significantly improving the integrity of the feature maps. Max-pooling is applied after that with a stride2 and 2x2 window to minimize the spatial dimension of the feature maps. In the architecture, two decoders are used in Dual U-net with Resnet encoder. In the decoder, the input feature is subjected to a 2 × 2 bi-linear up sampling in every block, which increases the dimensionality of the maps of input feature. The appropriate skip connections now concatenate feature maps from the encoder with the result feature maps. Only the skip connection of the encoder in U-NET 1 is utilized in the first decoder, while the skip connection from U-NET 1 and U-NET 2 encoders is utilized in second decoder. After concatenation, two 3 × 3 convolution is performed again, each accompanied by batch normalization and a rectified linear unit activation function. A squeeze and excitation block is then applied. Finally, a convolutional layer is applied with a sigmoid function that produces the mask.

IV. Methodology

Fig. 3 illustrated the flowchart of the method utilized in this work. The methodology design for this work is divided into the following four stages: research design, data collection and data processing, experimental setup, and performance analysis.

A. Research Design

The first step is to measure and evaluate the previous research that has been investigated using the terms 'segmentation', 'medical images', 'polyp image segmentation', 'deep convolutional neural network', and 'U-net'. To identify the problem statement and determine the approaches that have been used, a literature review is required. This section gives an overview of the segmentation of images and identifies the various image segmentation approaches and algorithms, along with their merits and limitations, that have been presented for medical image segmentation using deep convolutional neural networks.

B. Data Collection and Data Processing

This section explains the process for the collection of a dataset. The dataset for this study is gathered from the 'Kvasir-SEG https://datasets.simula.no/kvasir-seg/', which contains Gastrointestinal Polyp images. Once the data is gathered, every image of gastrointestinal polyps is examined for the ground truth. The dataset of gastrointestinal polyps includes 1000 images and corresponding 1000 masks.

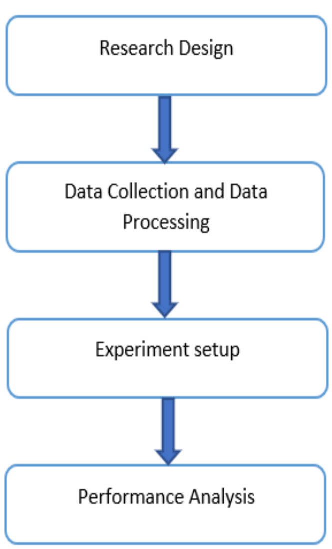


Fig. 3. Illustrates the flowchart of the methodology.

The dataset of gastrointestinal polyp images is partitioned into a training set, a validation set, and a testing set. The training set includes 80% of a dataset which is 800 images, the validation set contains 10% which is 100 images, and the testing set contains 10% which is 100 images. Once the partition of the dataset is done, augmentation of data is performed to the 800 images of the training set, increasing the total number of images in the training dataset up to 13600. For this work, we have used four algorithms of the deep convolutional neural network: Fully Convolutional Network (FCN), U-net, Unet_Resnet, and Dual U-net with Resnet encoder to illustrate the segmentation of an image dataset.

C. Experimental Setup and Peformance Analysis

This experiment is conducted on RTX 2060 GPU with 6 GB of RAM, and python 3.9 software framework with Tensor flow 2.5. All the segmentation algorithms used in this work are trained for 50 epochs.

Image segmentation algorithms such as Unet_Resnet, FCN, U-net, and Dual U-net with Resnet encoder will be compared on the basis of performance. The performance of these algorithms for image segmentation is measured by using the Intersection over Union (IOU), and Dice Similarity Coefficient (DSC).

V. Result and Discussion

The dataset from the 'Kvasir-Seg dataset' of a Gastrointestinal Polyps includes 1000 images of polyps and corresponding 1000 masks. Dataset is partitioned into 80% (800 images) of the training set, 10% (100 images) of the validation set, and 10% (100 images) of the testing set. After the dataset partitioning, data augmentation is implemented to the 800 images of the training set, which increases the dataset of training images to 13,600 images which will help in improving the accuracy rate. After that, the algorithm for segmentation is performed on the gastrointestinal polyp dataset. In this work, we have implemented Unet_resnet, U-net, FCN, and Dual U-net with Resnet encoder to conduct segmentation on images of gastrointestinal polyps.

The performance of the algorithm for segmentation of the images on the gathered dataset is described in this work. Performance of the algorithm is evaluated by, Intersection Over Union (IOU), Dice Similarity Coefficient (DSC) Recall, and Precision. Jaccard index commonly referred to as Intersection Over Union (IOU), is utilized to quantify the percent of overlap between the predicted image and the mask image. To calculate Intersection Over Union, the formula is as follows

$$\mathrm{IOU} = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X|+|Y|-|X \cap Y|} \quad (1)$$

Dice Similarity Coefficient quantifies the percentage of similarity between the predicted image and the mask image. The Dice similarity coefficient (DSC) is calculated by using the formula described below.

$$\mathrm{DSC} = 2 \times \frac{|X \cap Y|}{|X|+|Y|} \quad (2)$$

Precision calculates the percentage between the accurate prediction and the overall prediction. The following is the precision formula:

$$\text{Precision} = \frac{TRUE\ POSITIVE}{TRUE\ POSITIVE + FALSE\ POSITIVE} \quad (3)$$

The quantity of predictions that the algorithm predicted accurately is True Positive, and the total number of predictions is calculated by the formula's denominator.

Recall describes the number of accurate predictions to total observations made by the algorithm. The following is the recall formula.

$$\text{Recall} = \frac{TRUE\ POSITIVE}{TRUE\ POSITIVE + FALSE\ NEGATIVE} \quad (4)$$

True positive is the number of accurate predictions which is recognized by the algorithm, the formula's denominator shows the true positive as well as the number of positives that the algorithm falsely predicted as negative.

A. Results

The deep convolutional neural network algorithms: Dual U-net with Resnet encoder, Unet_Resnet, U-net, and Fully Convolutional Network (FCN) are used in this work to do the segmentation of the image dataset of gastrointestinal polyps. The performance of these algorithms is calculated by the Intersection over Union (IOU), and Dice Similarity Coefficient (DSC) as shown in equation (1) and equation (2) respectively. Both the metrics, Dice Similarity Coefficient and Intersection Over Union, ranges from 0 to 1. Fig. 4 illustrates a few of our results with Unet_resnet, Dual U-net with Resnet encoder, FCN, and U-net.

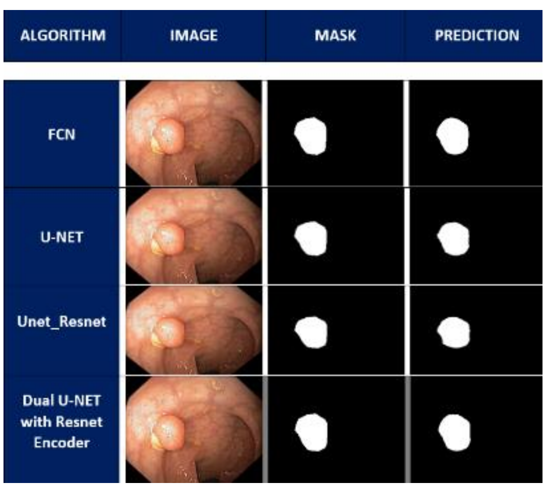


Fig. 4. Results of FCN, U-NET, Unet_Resnet, and Dual U-net with Resnet encoder.

Fig. 4 depicts the results of the segmentation of gastrointestinal polyp images using four algorithms. Table I shows a comparative study of the performances of these four algorithms.

Table I compares the performances of four algorithms based on Dice similarity coefficient (DSC) and Intersection over Union (IOU). As shown in the performance table FCN achieves less result than U-Net, while Unet_Resnet achieves a better result than U-Net. Dual U-net with Resnet encoder has achieved higher DSC and IOU than the other three algorithms.

TABLE I. THE PERFORMANCE OF ALGORITHMS

Algorithm	Dice Similarity Coefficient (DSC)	Intersection Over Union (IOU)
FCN	0.8393	0.7243
U-NET	0.8528	0.7458
Unet_Resnet	0.8603	0.7570
Dual U-net with resnet encoder	**0.8715**	**0.8042**

VI. CONCLUSION

In this work, deep convolutional neural network algorithms; U-Net, FCN, Unet_Resnet, and Dual U-net with Resnet encoder, are proposed to do segmentation of images of the gastrointestinal polyp's dataset from the Kvasir-SEG. The dataset collected from Kvasir-SEG is divided into three sets that are training set, validation set, and test set. Augmentation is applied to the training dataset, which increases the number of images in the dataset, and helps to improve the performance. The performance of the four algorithms of segmentation proposed in this work is evaluated by using the Intersection Over Union (IOU) and Dice Similarity Coefficient (DSC) metrics. Dual U-net with Resnet Encoder achieved DSC and IOU of 0.8715 and 0.8042 respectively. The result of the Dual U-net with Resnet encoder is compared with the other three algorithms, FCN, U-Net, and Unet_Resnet. FCN attained a DSC of 0.8393 and IOU of 0.7243. U-Net obtained DSC of 0.8528 and IOU of 0.7458 and Unet_Resnet acquired DSC of 0.8603 and IOU of 0.7570. The comparison result clearly shows that Dual U-net with Resnet Encoder outperforms the other three segmentation algorithms for gastrointestinal polyp images. In addition, future work will concentrate on enhancing the segmentation algorithm so that more accurate results will be achieved, and segmentation will be performed on various medical image datasets.

ACKNOWLEDGMENT

The publication fees for this paper are funded by Kulliyyah of Information and Communication Technology (KICT), International Islamic University Malaysia.

REFERENCES

[1] P. Prabhu, "Digital Image Processing Techniques – A Survey," no. May 2016, 2017.

[2] B. Idris, "Image Processing Techniques in Computer Vision: An Important Step for Product Visual Inspection," no. November, 2021.

[3] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustain.*, vol. 13, no. 3, pp. 1–29, 2021, doi: 10.3390/su13031224.

[4] R. Merjulah and J. Chandra, "Segmentation technique for medical image processing: A survey," *Proc. Int. Conf. Inven. Comput. Informatics, ICICI 2017*, no. Icici, pp. 1055–1061, 2018, doi: 10.1109/ICICI.2017.8365301.

[5] M. Vardhana, N. Arunkumar, S. Lasrado, E. Abdulhay, and G. Ramirez-Gonzalez, "Convolutional neural network for bio-medical image segmentation with hardware acceleration," *Cogn. Syst. Res.*, vol. 50, pp. 10–14, 2018, doi: 10.1016/j.cogsys.2018.03.005.

[6] M. Billah, S. Waheed, and M. M. Rahman, "An Automatic Gastrointestinal Polyp Detection System in Video Endoscopy Using

Fusion of Color Wavelet and Convolutional Neural Network Features," *Int. J. Biomed. Imaging*, vol. 2017, pp. 1–10, 2017, doi: 10.1155/2017/9545920.

[7] H. M. Afify, K. K. Mohammed, and A. E. Hassanien, "An improved framework for polyp image segmentation based on SegNet architecture," *Int. J. Imaging Syst. Technol.*, vol. 31, no. 3, pp. 1741–1751, 2021, doi: 10.1002/ima.22568.

[8] M. Akbari *et al.*, "Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, pp. 69–72, 2018, doi: 10.1109/EMBC.2018.8512197.

[9] A. Mehmood, I. R. Khan, H. Dawood, and H. Dawood, "Enhancement of CT images for visualization," *ACM SIGGRAPH 2019 Posters, SIGGRAPH 2019*, pp. 3–4, 2019, doi: 10.1145/3306214.3338602.

[10] Nisa, S. Q., & Ismail, A. R. "Comparative Performance Analysis of Deep Convolutional Neural Network for Gastrointestinal Polyp Image Segmentation," vol. 8, no. 4, pp. 149–156, 2021.

[11] Ş. Öztürk, R. Ahmad, and N. Akhtar, "Variants of Artificial Bee Colony algorithm and its applications in medical image processing," *Appl. Soft Comput. J.*, vol. 97, 2020, doi: 10.1016/j.asoc.2020.106799.

[12] W. Bo, W. Ying, and C. Lijie, "Fuzzy clustering recognition algorithm of medical image with multi-resolution feature," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 1, pp. 1–12, 2020, doi: 10.1002/cpe.4886.

[13] S. Sharma and A. Aggarwal, "Content-based retrieval of biomedical images using orthogonal Fourier-Mellin moments," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 7, no. 3, pp. 286–296, 2019, doi: 10.1080/21681163.2018.1493619.

[14] T. Saarinen and G. Petterson, "Image Analysis Techniques," *Track. Environ. Chang. Using Lake Sediments*, vol. 2, pp. 23–39, 2005, doi: 10.1007/0-306-47670-3_3.

[15] Y. Mohamed Y. Abdallah and T. Alqahtani, "Research in Medical Imaging Using Image Processing Techniques," *Med. Imaging - Princ. Appl. [Working Title]*, 2019, doi: 10.5772/intechopen.84360.

[16] W. Khan, "Image Segmentation Techniques: A Survey," *J. Image Graph.*, no. July, pp. 166–170, 2014, doi: 10.12720/joig.1.4.166-170.

[17] R. P. Nikhil and K. P. Sankar, "A Review on Image Segmentation Techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, 1993.

[18] S. Q. Nisa, A. R. Ismail, M. A. B. M. Ali and M. S. Khan, "Medical Image Analysis using Deep Learning: A Review," 2020 IEEE 7th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2020, pp. 1-3, doi: 10.1109/ICETAS51660.2020.9484287.

[19] M. S. Abirami and D. T. Sheela, "Analysis of Image Segmentation Techniques," *Int. J. Recent Adv. Eng. Res.*, vol. 3, no. 3, p. 4, 2014, doi: 10.24128/ijraer.2017.mn89cd.

[20] G. Phonsa and K. Manu, *A survey: Image segmentation techniques*, vol. 741. Springer Singapore, 2019. doi: 10.1007/978-981-13-0761-4_105.

[21] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep Neural Networks for Medical Image Segmentation," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/9580991.

[22] K. K. Gupta, N. Dhanda, and U. Kumar, "A comparative study of medical image segmentation techniques for brain tumor detection," *2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018*, no. December, pp. 1–4, 2018, doi: 10.1109/CCAA.2018.8777561.

[23] A. Banerjee, S. Bhattacharjee, and S. Latib, "Image segmentation using region derived triple thresholding," *2016 3rd Int. Conf. Recent Adv. Inf. Technol. RAIT 2016*, pp. 429–433, 2016, doi: 10.1109/RAIT.2016.7507940.

[24] W. Cao, J. Li, J. Liu, and P. Zhang, "Two improved segmentation algorithms for whole cardiac CT sequence images," *Proc. - 2016 9th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2016*, no. 32, pp. 346–351, 2017, doi: 10.1109/CISP-BMEI.2016.7852734.

[25] H. Seo *et al.*, "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," *Med. Phys.*, vol. 47, no. 5, pp. e148–e167, 2020, doi: 10.1002/mp.13649.

[26] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Comput. Surv.*, vol. 52, no. 4, 2019, doi: 10.1145/3329784.

[27] M. Mittal, M. Arora, T. Pandey, and L. M. Goyal, *Image Segmentation Using Deep Learning Techniques in Medical Images*. 2020. doi: 10.1007/978-981-15-1100-4_3.

[28] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, *A survey of the recent architectures of deep convolutional neural networks*, vol. 53, no. 8. Springer Netherlands, 2020. doi: 10.1007/s10462-020-09825-6.

[29] J. Lai, H. Zhu, and X. Ling, "Segmentation of Brain MR Images by Using Fully Convolutional Network and Gaussian Mixture Model with Spatial Constraints," *Math. Probl. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/4625371.

[30] L. Bi, D. Feng, and J. Kim, "Dual-Path Adversarial Learning for Fully Convolutional Network (FCN)-Based Medical Image Segmentation," *Vis. Comput.*, vol. 34, no. 6–8, pp. 1043–1052, 2018, doi: 10.1007/s00371-018-1519-5.

[31] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *J. Digit. Imaging*, vol. 32, no. 4, pp. 582–596, 2019, doi: 10.1007/s10278-019-00227-x.

[32] M. Z. Alom *et al.*, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," 2018, [Online]. Available: http://arxiv.org/abs/1803.01164.

[33] M. Pashaei, H. Kamangir, M. J. Starek, and P. Tissot, "Review and evaluation of deep learning architectures for efficient land cover mapping with UAS hyper-spatial imagery: A case study over a wetland," *Remote Sens.*, vol. 12, no. 6, 2020, doi: 10.3390/rs12060959.

[34] T. Tsuji and S. Arima, Automatic Multi-class Classification of Tiny and Faint Printing Defects Based on Semantic Segmentation, vol. 189. 2021. doi: 10.1007/978-981-15-5784-2_18.

[35] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-net for medical image segmentation," *Comput. Methods Programs Biomed.*, vol. 192, p. 105395, 2020, doi: 10.1016/j.cmpb.2020.105395.

[36] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, no. Cvd, pp. 12–20, 2015, doi: 10.1007/978-3-319-24574-4.

[38] M. Habijan, H. Leventic, I. Galic, and D. Babin, "Whole Heart Segmentation from CT images Using 3D U-Net architecture," *Int. Conf. Syst. Signals, Image Process.*, vol. 2019-June, pp. 121–126, 2019, doi: 10.1109/IWSSIP.2019.8787253.

[39] B. Ait Skourt, A. El Hassani, and A. Majda, "Lung CT image segmentation using deep neural networks," *Procedia Comput. Sci.*, vol. 127, pp. 109–113, 2018, doi: 10.1016/j.procs.2018.01.104.

Rapid Modelling of Machine Learning in Predicting Office Rental Price

Thuraiya Mohd[1], Muhamad Harussani[2], Suraya Masrom[3]

GreensAFE (GreSFE) Research Group-Faculty of Architecture-Planning and Surveying-Department of Built Environment Studies and Technology, Universiti Teknologi MARA, Perak Branch, Seri Iskandar Campus, 32610 Perak, Malaysıa[1]

Centre of Graduate Studies, Universiti Teknologi MARA, Perak Branch, Seri Iskandar Campus, 32610 Perak, Malaysıa[2]

Malaysia Machine Learning and Interactive Visualization (MaLIV) Research Group Computing Sciences Study-College of Computing-Informatics and Media, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Perak, Malaysia[3]

***Abstract*—This study demonstrates the utilization of rapid machine learning modelling in an essential case of the real estate industry. Predicting office rental price is highly crucial in the real estate industry but the study of machine learning is still in its infancy. Despite the renowned advantages of machine learning, the difficulties have restricted the inexpert machine learning researchers to embark on this prominent artificial intelligence approach. This paper presents the empirical research results based on three machine learning algorithms namely Random Forest, Decision Tree and Support Vector Machine to be compared between two training approaches; split and cross-validation. AutoModel machine learning has accelarated the modelling tasks and is useful for inexperienced machine learning researchers for any domain. Based on real cases of office rental in a big city of Kuala Lumpur, Malaysia, the evaluation results indicated that Random Forest with cross-validation was the best promising algorithm with 0.9 R squared value. This research has significance for real estate domain in near future, by applying a more in-depth analysis, particularly on the relevant variables of building pricing as well as on the machine learning algorithms.**

Keywords—Random forest; decision tree; support vector machine; rapid prediction modelling; office rental price

I. Introduction

Now-a-days, real estate is becoming more digital, automated, and integrated. The fusion of industry 4.0 and digital 4.0 includes connected buildings, wearable technology, data management for buildings and infrastructure, and smart cities. The transformation of the real estate industry was improved due to the advancement in data science technologies such as analytic technologies [1]. The analytic technologies mentioned include Computational Statistics, Artificial Intelligence (AI) and Machine Learning. Machine learning is a sub-field of AI that can learn and re-learn from data exploration and inferences. Nowadays, these analytic technologies have successfully transformed the real estate industry to discover various opportunities, particularly by developing prediction applications that involve fundamental tasks that uncover hidden patterns, unknown correlations, and preferences [2]. Despite opportunities, some challenges appeared, including gaining adequate skills instantly that involves varying knowledge of AI concepts, mathematics, programming, and computer technologies. Thus, rapid software is useful to them and at the same time benefits the expert in accelerating the preliminary analytic tasks.

Considering real estate markets in general, office building markets are more synchronized in terms of exposure to macro-effects and performance of the real estate within the market. The heterogeneity of the office markets makes them more complex to analyse [3], [4]. It can be challenging to understand the market, for which the property's price might be determined on the market, but it may not always equate with the valuation of property in the market [5]. Office markets often relate to good investment opportunities since it draws much capital but with a substantial return [6]. Despite being a well-established investment industry, it has a highly complex market structure due to the lack of a central marketplace and the individuality of each building.

Numerous econometric models have been proposed to predict the office market performance, especially the rental property market. These include office market econometric models [5], and the hedonic regression model [6]. Sadly, limited success was achieved in finding a reliable and consistent model to predict rental property market movements over a five-to-ten-year time frame [7]. It was expected that lacking market data can be the main problem to fault the unreliability of prediction model. Based on the preliminary statistical analysis, the collected data of office rentals has a few problems of variance insufficient, imbalance with very skewed data distribution and most of them are having low dependencies to the target data (dependent variable) to be relied by the prediction model in generating high accurate results.

Acknowledging the advantages of artificial intelligence computing approach that able to learn and redevelop knowledge to self-improve the output target from the given data, the used of machine learning technique in solving issues of real estate industry has started to begin. Even with low-association dataset, machine learning with the intelligent and leaning ability, will use mathematical and heuristics projection to self-improve their performances continuously during the training stage. Despite the wider used of machine learning in various domains of problems, there is still limited work that can be found for the real estate industry. This research attempted to fill the gap by focusing on the flexible and rapid modelling machine learning approach for office rental prediction problem.

This research was funded by NAPREC under grant number 100-TNCPI/GOV 16/6/2 (027/2021).

The contributions of this paper are two-fold. First, it demonstrates the use of AutoModel RapidMiner for the office rental price prediction model as a rapid modelling in the preliminary research activities. Second, it presents a comparison of results between two machine learning training approaches (split, cross-validation) that developed with manual model in RapidMiner and provides discussion on the significant findings in the context of office rental domain.

This paper is organized as follow. The next section II presents the literature review of rapid modelling machine learning and the state-of-the-arts of research for office rentals prediction. The methodology of research is given in section II followed with the results and discussion in section IV. The last section V provides the conclusion remarks.

II. Literature Review

A. Rapid Modelling of Machine Learning

Rapid modelling has been a long-time concern by researchers to solve real-world problems, mainly when complex computing techniques are required. In machine learning, to support rapid modelling, several tools have been introduced, and the most popular is script programming with Python or R programming languages [7]. Although this two-scripting language is considerably easy to utilize with their built-in programming libraries, they still need some help for a non-computing expert who never learns to program.

Graphical User Interface (GUI) based rapid modelling is easier than programming to implement machine learning. To date, some of the popular rapid tools are Weka [8] and RapidMiner [9]. Automated machine learning (AML) [10], [11] is a promising module in Weka and RapidMiner. Auto-WEKA is the AML in Weka, while AutoModel is the one provided in RapidMiner. AML helps to accelerate the modelling tasks by optimizing the machine learning pipelines from the input variables (features selections) step, algorithm suitability selection and hyper-parameters optimal setting. Literature on research that used Auto-WEKA is more available than AutoModel RapidMiner. Thus, focusing on Auto-Model RapidMiner for this study can provide additional benefits to the AML scholarly, mainly to inexpert data scientists in various domains.

B. Office Rent Predictions with Machine Learning

A number of researchers have used machine learning in the real estate or property industry. Researchers in [12] presents the real estate opportunities with machine learning technique modelling. It was reported that one of the advantages of machine learning is for assisting the stakeholders in making important decisions related to commercial or office building. A significant of finding has been presented by researchers in [13] that used machine learning technique to estimate the warehouse rental price. More interesting, by utilising social media web scraping technique, the collected data were analysed with machine learning prediction algorithms and hedonic modelling to monitor the building rental prices in Shenzhen, China [14]. To ensure the development of holistic smart building control effectively, researchers in [15] have utilised deep reinforcement learning, a recent advance method of machine learning. Internet of Things (IoT) is the main elements of smart building hence researchers in [16] introduced an algorithm named as Random Neural Network (RNN) to make used the IoT data to predict the consumption of energy of the smart building. Random Forest machine learning for forecasting shop rents in Guangzhou, China. To conduct mass appraisal in an urban residential area where commercial properties are available, researchers in [17] used multiple regression and random forest as the proposed methods. Similarly, the performances of random forest and multiple regression has been reported from the research findings in [18]. How the neighbourhood environment can influence peer-to-peer accommodation when using random forest is the finding reported in [19]. It seems that random forest is very promising, and it is also one of the suggested algorithms from AutoModel used in this research together with Decision Tree [20] and Support Vector Machine [21]. To the best of our knowledge, rapid modelling on office rental prediction has not been reported yet in the current literature. This research filled the gap by presenting the precise steps and the comparison of results.

III. Methodology

A. The Dataset

The dataset used in this study is a collection of office rentals from the year 2015-2021 in Kuala Lumpur, Malaysia. Table I shows the set of features of developing the machine learning prediction model. This study uses 21 attributes or features as independent variables for office rental prediction.

TABLE I. Features of the Office Rental Prediction Model

Office Rent Determinants	Description
Building Appearance and Design	Physical building appearances and Design
Building Age	Age of Building
Amenities and In-house Services	Amenities and Services Provided by the building
Occupancy	Occupancy rate
Distance to the city centre	Distance to the city centre
Building Frontage	Building allocation towards the main road
Neighbourhood Characteristics	Surrounding areas
Traffic Condition	Traffic Condition (Congested, Free)
Nearest Public Transport	Availability of Public Transport
Transaction Date	Date of Transactions
Floor Level	Rented Floor Level
Rentable Area	Area of premises being rented
Tenancy Duration	Lease duration by the tenant
Service Charge	Service Charge towards the building's occupants
Employment Rate	Employment rate by year
Inflation	Inflation by year
Gross Domestic Product (GDP)	Gross Domestic Product by Year
Finance, Insurance and Real Estate (FIRE)	Business purposes of building
Green Certificate	Green Certification of the building
MSc Status	MSc Certification of the building
Building Grade	Office Building Grading

B. Machine Learning Rapid Modelling Framework

Fig. 1 presents the rapid modeling framework used in this research. On the pre-processing data that was ready to be read in RapidMiner, AutoModel was firstly executed. The purpose of executing AutoModel is to get suitable machine learning algorithms for the dataset and the optimal hyper-parameters. The suggested machine learning algorithms are Generalized Linear Model, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees and Support Vector Machine. The six sets of experiments can be described as listed in Table II. Table III is the optimal hyper-parameters setting suggested by the AutoModel.

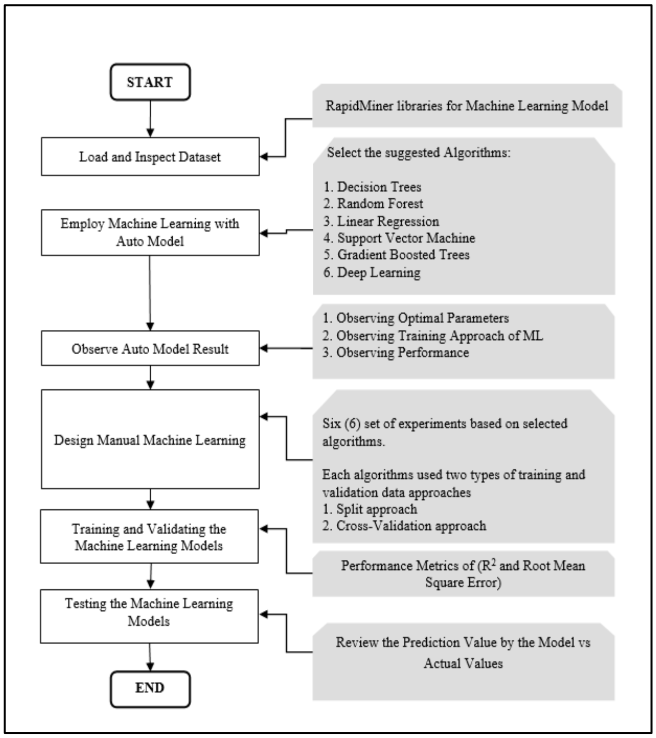


Fig. 1. Machine learning rapid modeling framework.

TABLE II. DESCRIPTION OF EXPERIMENT SET

Experiment set	Algorithm	Training
1	Decision Tree	Split
2	Random Forest	Split
3	Support Vector Machine	Split
4	Decision Tree	CV
5	Random Forest	CV
8	Support Vector Machine	CV

TABLE III. ALGORITHMS SELECTION AND OPTIMAL HYPER-PARAMETERS

List of Algorithms	Maximal Depth	Number of Trees	RBF	C	Error Rate
Decision Tree	15	NA	NA	NA	10.7%
Random Forest	20	7	NA	NA	26%
Support Vector Machine	NA	NA	0.050	1000	43.5%

As listed in Table III, the lowest error rate for the Decision Tree can be achieved for the office rental dataset if the maximal depth of the tree is 15. maximal depth defines the maximum level of the tree minus 1. AutoModel has identified that the worst error rate was achieved if the maximal depth was set to 2. Random Forest has additional hyper-parameters as it is an advancement of Decision Tree. Besides maximal depth, Random Forest has number of trees. Imagine a forest that consists of more than one tree. AutoModel suggested that the optimal values for the number of trees and maximal depth were 20 and 7, respectively. The worst error rate achieved was 42.9% at 140 trees and two maximal depths. Support Vector Machine has Radial Basis Function (RFB) Kernal and C value. The C parameter instructs the SVM optimization to avoid misclassifying each training example. The kernel function is used to transform the data, increasing its dimensional. This enhancement causes the data to be split with a hyperplane with a significantly greater probability and establishes a minimal prediction probability error measure [22]. The configuration of optimal parameters suggested by Auto Model for the Support Vector Machine algorithm is at RBF = 0.050, C = 100, and 43.5% error rate.

C. Training Approaches

AutoModel used a split training approach that allowed the researcher to look at another training approach, namely the cross-validation approach. Therefore, based on the AutoModel findings, manual machine learning modelling in RapidMiner was developed to compare the three machine learning algorithms with the optimal setting but with different training approaches; split and cross-validation (CV). Fig. 2 presents the RapidMiner process for split training, while Fig. 3 presents the CV approach.

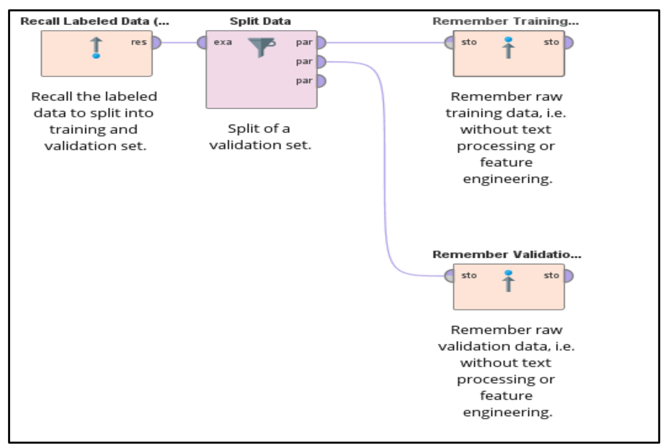


Fig. 2. Split approach.

The “Split Data” operator is a custom operator for splitting a dataset into training and testing datasets [23]. To configure the parameter in the split approach, the researcher will specify the ratio of all partitions. The sum of all partitions’ ratios should be equal to 1. As for this study, the 0.8:02 ratio was used for setting the partitions. Therefore, the training and testing datasets constructed from the original dataset were 80% and 20% of the data, respectively. The next step is to include the algorithms in the office rental price prediction model by connecting all nodes from the chosen parameters.

Fig. 4, Fig. 5 and Fig. 6 present the manual modelling process for the tree algorithms with split training. The Split Data operator can be changed to Cross Validation operator for implementing CV on each model.

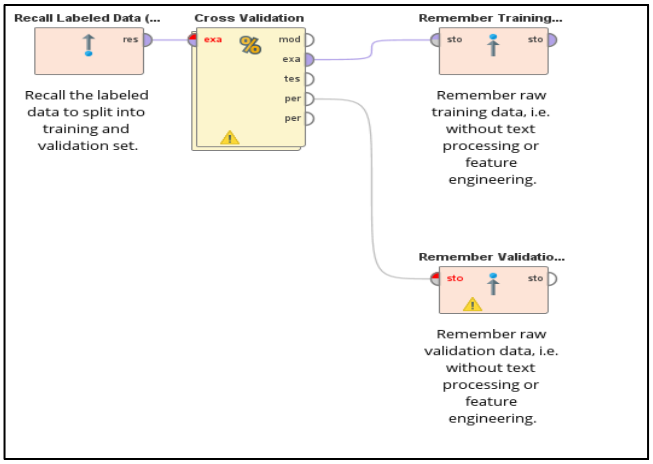


Fig. 3. Cross-validation approach.

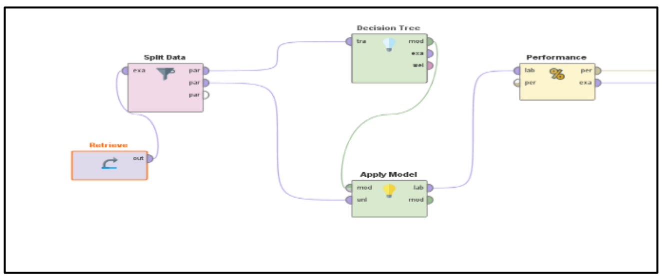


Fig. 4. Decision tree with split training.

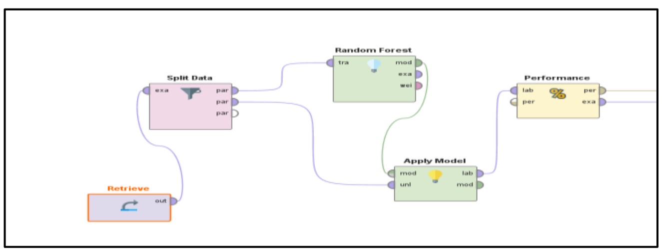


Fig. 5. Random forest with split training.

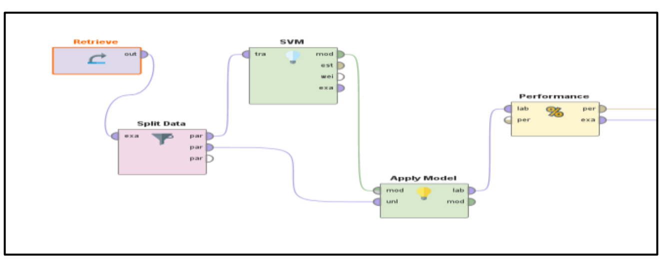


Fig. 6. Support vector machine with split training.

D. *Performances Metric*

In developing the office rent prediction model, this study deployed two (2) performance measurements, namely the Coefficient of Determination "R^2" and "Root Mean Square Error". R-Squared is a statistical metric that indicates the proportion of the variation explained by the independent variables for the dependent variable. The greater the R-squared, the better the model matches the dataset under consideration. R-Squared may be calculated mathematically as Equation (1).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - y)^2} \quad (1)$$

The metric of root means square error (RMSE) is a common way to calculate a model's error in quantitative data prediction. *E* is no absolute good or bad error level but can be identified based on the dependent variable. Generally, the range from 0 to 1000 is classified as small, but if the range is from 0 to 1, it is classified as no longer being small. In evaluating the model, the smaller the value of the root mean square error the better the model has produced [24]. Equation (2) is to calculate root means square error.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}(y_i - \hat{y}_i)^2} \quad (2)$$

IV. Results and Discussion

A. *Machine Learning Prediction Results based on Split Approach*

Table IV compares the prediction results generated by the three different algorithms using the split approach. The "observed value" indicates the actual value/the raw rental data acquired. The prediction value indicates the value generated after considering the optimal parameters, processes, and factors involved.

TABLE IV. Split-Data Prediction Results

Observed Value	**Prediction**		
	Decision Tree	**Random Forest**	**Support Vector Machine**
RM24.53	RM29.20	RM42.67	RM37.50
RM27.55	RM31.89	RM34.60	RM37.57
RM24.58	RM35.47	RM27.37	RM41.53
RM19.85	RM22.42	RM43.09	RM43.80
RM21.37	RM35.47	RM29.06	RM41.58
RM21.98	RM32.45	RM26.52	RM39.48
RM76.84	RM77.47	RM73.72	RM63.35
RM28.00	RM28.72	RM37.16	RM34.77
RM29.85	RM23.97	RM35.37	RM41.99
RM21.37	RM36.28	RM28.44	RM38.78

An extensive result of the prediction with details is provided in the prediction chart, as shown in Fig. 4, 5 and 6. The prediction chart depicted in Fig. 4 to 6 was analysed from the triangle patterns and the diagonal line. The diagonal line indicates the real data/actual office rent values while the triangle patterns indicates the prediction generated by the algorithms used in this study. The accumulated pattern at the diagonal line shows a good prediction. In contrast, the deviated pattern from the line showed less accuracy in predicting the rent and was considered a prediction outlier. The illustrated chart demonstrated that Random Forest (Refer Fig. 8) provides the best predictions can be seen through the high accumulation of the triangle patterns to the diagonal line. In comparison, Decision Tree algorithms also provide a good prediction, as illustrated in Fig. 7. The prediction chart for Support Vector Machine, however, depicts scattered triangle patterns which indicate high frequencies of outliers when predicting office rents, as seen in Fig. 9. To justify Random Forest algorithms as the best predictor, this study employed another approach in Machine Learning called Cross-Validation.

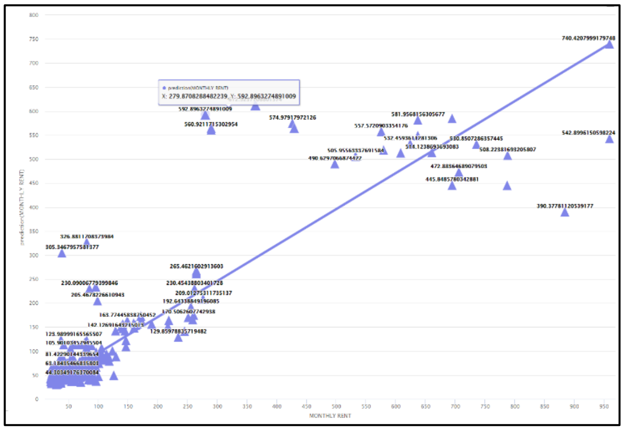


Fig. 7. Prediction chart of decision tree (split).

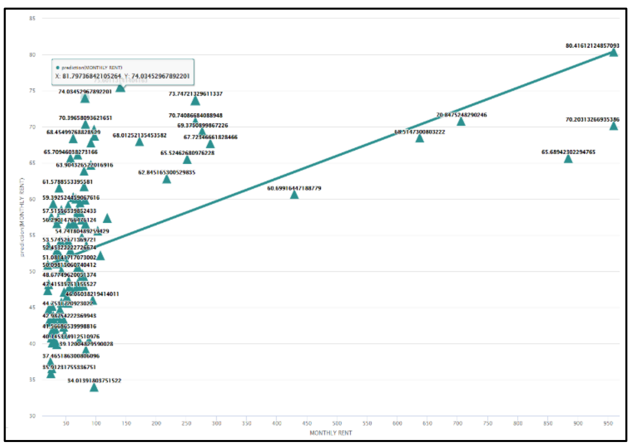


Fig. 8. Prediction chart of random forest (split).

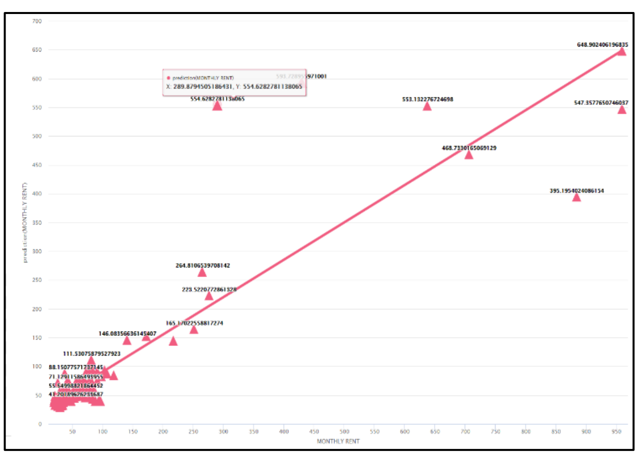


Fig. 9. Prediction chart of support vector machine (split).

B. Machine Learning Prediction Results based on CV Approach

Table V lists some of the prediction results generated by each algorithm using the CV approach. An extensive result of the prediction with details is provided in the prediction charts, as shown in Fig 10 to 12.

TABLE V. CROSS-VALIDATION PREDICTION RESULTS

Observed Value	**Prediction**		
	Decision Tree	**Random Forest**	**Support Vector Machine**
RM24.53	RM29.20	RM46.34	RM22.42
RM27.55	RM31.89	RM34.89	RM39.57
RM24.58	RM35.47	RM31.37	RM41.53
RM19.85	RM22.42	RM43.09	RM35.47
RM21.37	RM35.47	RM29.06	RM41.58
RM21.98	RM32.45	RM26.52	RM39.48
RM76.84	RM77.47	RM73.72	RM63.35
RM28.00	RM28.72	RM37.16	RM34.77
RM29.85	RM23.97	RM35.37	RM28.72
RM21.37	RM36.28	RM23.97	RM38.78

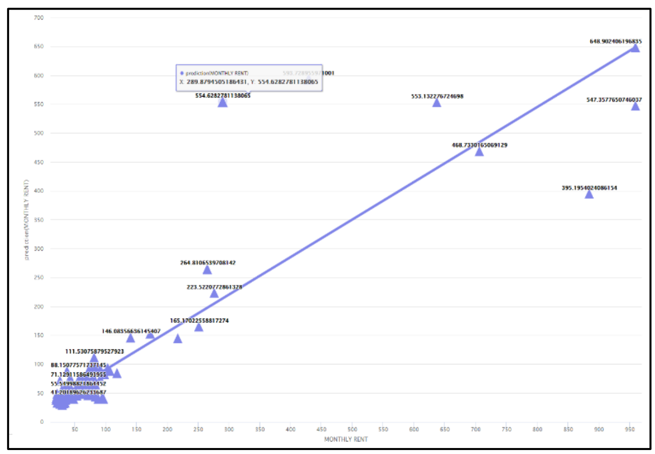


Fig. 10. Prediction chart of decision tree (CV).

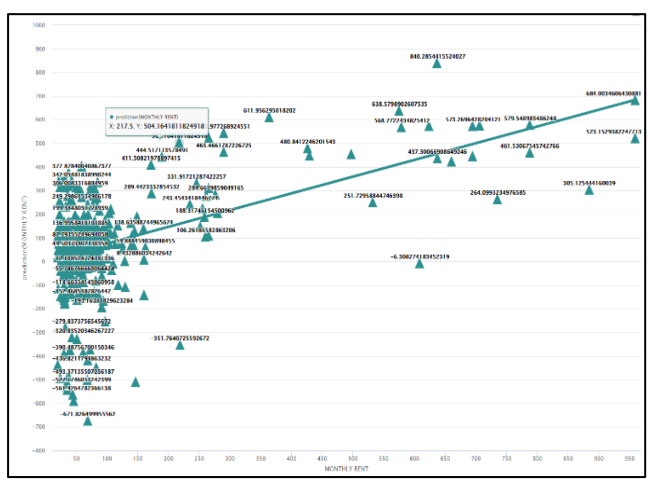


Fig. 11. Prediction chart of random forest (CV).

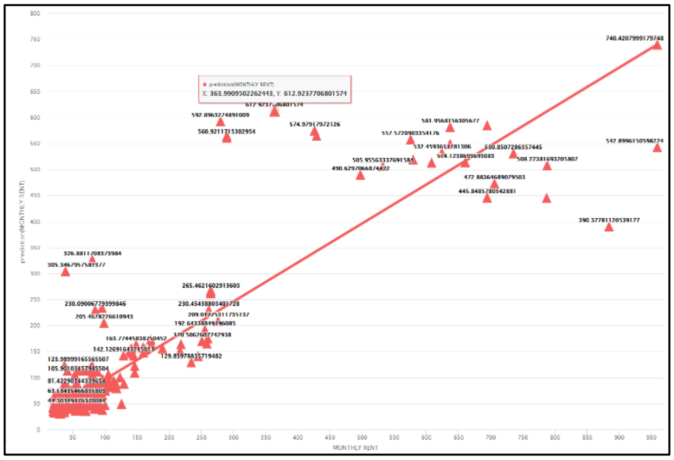


Fig. 12. Prediction chart of support vector machine (CV).

The prediction charts of each algorithm with the CV approach demonstrate that Random Forest provides the best predictions can be seen through the high accumulation of the triangle patterns to the diagonal line. In comparison, Decision Tree algorithms also provide a good prediction, as illustrated in Fig. 4. However, the prediction chart for the Support Vector Machine depicts scattered triangle patterns, which indicate high frequencies of outliers when predicting office rents.

C. Performances Comparisons

Table VI compares the performances of each algorithm concerning the training approaches. Overall, the result shows that the CV approach provides better results with a higher R^2 correlation value and lower RMSE value than the split data approach. The Random Forest algorithm outperformed other algorithms regarding office rental price prediction value. It has generated the most accurate prediction compared to the decision trees and support vector machine. The accuracy of prediction generated by random forest was supported through the observation of the RMSE value. The lowest RMSE values generated by random forest justified a lower error rate when predicting a model evaluation; the smaller the value of root means square error, the better the model produced.

TABLE VI. MACHINE LEARNING ALGORITHMS COMPARISONS

Algorithms	Split		CV	
	R^2	RMSE	R^2	RMSE
Decision Tree	0.852	80.663	0.876	61.294
Random Forest	0.883	75.695	0.906	51.859
Support Vector Machine	0.479	155.465	0.725	120.738

V. CONCLUSIONS

This paper presents the review and findings of using machine learning algorithms for real data of office building rent in Bandar Kuala Lumpur, Malaysia. The rigorous steps for rapid machine learning models initiated with AutoModel preliminary findings were given in this paper. The suggested setting from AutoModel has been used to improve the machine learning by using the suggested split training with cross-validation technique. The performances of all machine learning algorithms can be improved with the cross-validation technique. However, the findings from this study are limited to the tested datasets and therefore require further investigation for different types of problems. Notably, this study provides new knowledge on the application of machine learning in analyzing real estate data, particularly on rental values of the office building. By exploring the machine learning methods, this study will greatly assist future research in solving problems involving prediction and forecasting real estate data.

REFERENCES

[1] J. Barnett, W. Serrano, P. Treleaven, and A. Knight, “Real Estate Data Marketplace,” SSRN Electronic Journal, no. January, 2021, doi: 10.2139/ssrn.3745816.

[2] F. Lorenz, J. Willwersch, M. Cajias, and F. Fuerst, “Interpretable machine learning for real estate market analysis,” Real Estate Economics, 2022, doi: 10.1111/1540-6229.12397.

[3] D. Jaffee, R. Stanton, and N. Wallace, “Energy Factors, Leasing Structure and the Market Price of Office Buildings in the U.S.,” Journal of Real Estate Finance and Economics, vol. 59, no. 3, pp. 329–371, Oct. 2019, doi: 10.1007/s11146-018-9676-x.

[4] H. H. Rong, J. Yang, M. Kang, and A. Chegut, “The value of design in real estate asset pricing,” Buildings, vol. 10, no. 10, pp. 1–26, Oct. 2020, doi: 10.3390/buildings10100178.

[5] M. Ezra, H. Mohd Ali, and Tuti Haryati Jasimin, “Valuers Behavioural Uncertainties in Property Valuation Decision Making,” 2018.

[6] D. Trojanowski, Recent trends and its analysis, no. January. 2019.

[7] Ramachandran Trichur Narayanan, “Novice Programmer to New-Age Application Developer: What Makes Python their First Choice?” 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) , 2019.

[8] Jason Brownlee, Machine learning mastery with Weka, 1st ed., vol. 1(4). 2019.

[9] J. Arunadevi, S. Ramya, and M. R. Raja, “A study of classification algorithms using Rapidminer,” 2018. [Online]. Available: https://www.researchgate.net/publication/325718529.

[10] F. Hutter, L. Kotthoff, and J. Vanschoren, “The Springer Series on Challenges in Machine Learning Automated Machine Learning Methods, Systems, Challenges,” 2019. [Online]. Available: http://www.springer.com/series/15602.

[11] L. Vaccaro, G. Sansonetti, and A. Micarelli, “An Empirical Review of Automated Machine Learning,” Computers, vol. 10, p. 11, 2021, doi: 10.3390/computers.

[12] A. Baldominos, I. Blanco, A. J. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, “Identifying real estate opportunities using machine learning,” Applied Sciences (Switzerland), vol. 8, no. 11, Nov. 2018, doi: 10.3390/app8112321.

[13] Y. Ma et al., “Estimating Warehouse Rental Price using Machine Learning Techniques,” International Journal of Computers Communications & Control, vol. 13, no. 2, pp. 235–250, 2018.

[14] L. Hu, S. He, Z. Han, S. Su, M. Weng, and Z. Cai, “Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies,” 2018.

[15] X. Ding, W. Du, and A. Cerpa, “OCTOPUS: Deep reinforcement learning for holistic smart building control,” in BuildSys 2019 - Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Nov. 2019, pp. 326–335. doi: 10.1145/3360322.3360857.

[16] A. Javed, H. Larijani, and A. Wixted, “Improving Energy Consumption of a Commercial Building with IoT and Machine Learning,” IT Prof, vol. 20, no. 5, pp. 30–38, Sep. 2018, doi: 10.1109/MITP.2018.053891335.

[17] S. Yilmazer and S. Kocaman, “A mass appraisal assessment study using machine learning based on multiple regression and random forest,” Land use policy, vol. 99, Dec. 2020, doi: 10.1016/j.landusepol.2020.104889.

[18] M. Čeh, M. Kilibarda, A. Lisec, and B. Bajat, “Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments,” ISPRS Int J Geoinf, vol. 7, no. 5, May 2018, doi: 10.3390/ijgi7050168.

[19] H. Jiang, L. Mei, Y. Wei, R. Zheng, and Y. Guo, “The influence of the neighbourhood environment on peer-to-peer accommodations: A random forest regression analysis,” Journal of Hospitality and Tourism Management, vol. 51, pp. 105–118, Jun. 2022, doi: 10.1016/j.jhtm.2022.02.028.

[20] J. Kelleher, B. mac Namee, and A. D. ’ Arcy, “Fundamentals of Machine Learning for Predictive Data Analytics.”

[21] D. A. Pisner and D. M. Schnyer, “Support vector machine,” Machine Learning: Methods and Applications to Brain Disorders, pp. 101–121, Jan. 2020, doi: 10.1016/B978-0-12-815739-8.00006-7.

[22] M. Graczyk et al., “Comparative Analysis of Premises Valuation Models Using KEEL, RapidMiner, and WEKA,” LNAI, vol. 5796, pp. 800–812, 2009, doi: 10.1007/978-3-642-04441-0_70.

[23] A. Massaro, V. Maritati, and A. Galiano, “Data Mining Model Performance of Sales Predictive Algorithms Based on Rapidminer Workflows,” International Journal of Computer Science and Information Technology (IJCSIT) International Journal of Computer Science & Information Technology (IJCSIT) International Journal of Computer Science & Information Technology (IJCSIT), vol. 10, no. 3, 2018, doi: 10.5121/ijcsit.2018.10303.

[24] D. Ho, G. Newell, and A. Walker, “The importance of property-specific attributes in assessing CBD office building quality,” Journal of Property Investment and Finance, vol. 23, no. 5, pp. 424–444, 2005, doi: 10.1108/14635780510616025.

Hybrid Modeling to Classify and Detect Outliers on Multilabel Dataset based on Content and Context

Lusiana Efrizoni[1], Sarjon Defit[2], Muhammad Tajuddin[3]
Faculty of Computer Science, STMIK Amik Riau, Pekanbaru, Indonesia[1]
Faculty of Computer Science, UPI YPTK Padang, Padang, Indonesia[2]
Faculty of Computer Science, Bumigora University, Nusa Tenggara Barat, Indonesia[3]

***Abstract*—Due to the linked various matching categories, news article categorization are a rapidly increasing field of interest in text classification. However, the low-reliability indices and ambiguities related to frequently used province classifiers restrict success in this field. Most of the existing research uses traditional machine learning algorithms. It has weaknesses in training large-scale datasets, and data sparseness often occurs from short texts. Therefore, this study proposed a hybrid model consisting of two models, namely the news article classification and the outlier detection model. The news article classification model used a combination of two deep learning algorithms (Long Short-Term Memory dan Convolutional Neural Network) and outlier classifier model, which was intended to predict the outlier news using a decision tree algorithm. The proposed model's performance was compared against two widely used datasets. The experimental results provide useful insights that open the way for a number of future initiatives.**

Keywords—News article classification; machine learning; outlier detection

I. Introduction

Digital or online news is a form of contemporary news where editorial content is distributed over the internet as opposed to published via print or broadcast. News or information contained in online news portals allows errors to occur in grouping/classifying news. For example, news is categorized in the infotainment category, while based on the content of the news or the words contained in it, the news should be categorized in the politics category. Journalists and news monitoring companies (media monitoring companies) often face problems identifying topics in a very large number of news articles around the world [1]. Errors in categorizing or classifying information/news can also occur because the method used is still manual by reading the entire article to find the main topic. This method requires large resources and requires the reader's ability to extract the topic of a news/information document [2]. This fact shows a discrepancy between the news category (or context) and the news content, or the meaning discussed or the news topic (as content) in categorizing or classifying news.

The increase in online news makes it difficult for internet users to access the content they are interested in, so it is necessary to classify news (text) so that it is easily accessible [3]. Coupled with the ever-growing volume of news corpus on the World Wide Web (WWW), it also creates problems in text classification, especially news article classification [4]. Readers can receive much information on the online news portal. Sometimes, they take it for granted without any selection of information. On that basis, much of the current information media classifies the categorization before dissemination to the general public. This classification is useful to make it easier for people to find the desired information. The classification of news articles often suffers from ambiguity due to the various categories that fit and the weak reliability performance of most classification systems used, resulting in low efficiency [5]. Therefore, automatic news (text) classification needs to be developed because manual work is no longer effective. If it is done automatically, people will not be asked to think about which category the news belongs to [2]. The ability to classify texts (news) into certain categories is very helpful in dealing with information overload [6]. Multilabel text classification is an activity to categorize text into one or more categories [7].

Recurrent Neural Network (RNN) is one of the most popular architectures used in Natural Language Processing (NLP) because the recurring structure is suitable for processing text with long variables [7]. Meanwhile, the Long Short-Term Memory (LSTM) was developed to solve the exploding and vanishing gradient problems that can be encountered during traditional RNNs training. Classification of online news texts using LSTM was utilized in this study because the LSTM structure is a complete series or cannot be cut since cutting text document structure changes the meaning of the sentence. Word embedding was used as an input feature in the LSTM before classifying the text. Apart from RNN/LSTM, Convolutional Neural Network (CNN) has also achieved excellent results in the NLP field. CNN can be combined with word vectors in topic classification and semantic analysis to achieve good results [8]. The CNN input matrix only extracts the word vector matrix from the word detail level and ignores the overall semantic feature expression from the text breakdown level, which leads to an incomplete representation of text features and can affect the accuracy of text classification [9].

This study aims to evaluate how extractive summary of the text (news content) and context (similarity of words and relationships between words) can help filter important information from the text either for readers or consumption models in classifying online news in English. The word embedding method is used to represent words in vector space in content-based and context-based representations. The method used in content-based representation is Latent

Semantic Analysis (LSA) and Singular Value Decomposition (SVD), while in context-based representation, the technique used is Word2Vec. In addition to the classification performance of news articles, outlier detection is also the focus of this research since outlier detection is very closely related to the text classification process [10]. Outliers are abnormal patterns or events that do not match the expected events or patterns [11]. Outlier detection is used to detect news that does not fit the category of news articles. The research results are expected to be used by news management agencies on online news portals to classify news articles according to their categories and filter or block if they are found to contain outliers. In addition, automatic news/information categorization is very important for handling multi-label news article classification in online portals [6].

The remainder of this work is structured systematically. The second section provides brief overviews of text classification (or news article categorization) approaches. The research approach for the experiments is presented in Section III. The outcomes of the experiments are reported in Section IV, while the conclusion is presented in Section V.

II. Related Works

Relevant research related to text classification has been carried out in previous studies in accordance with the process of classifying texts or online news articles, including:

- The research of Stein et al. [12] analyzed presentation text (i.e., GloVe, word2vec, and fastTex) with a combination of classification models (i.e., fastText, XGBoost, SVM and CNN) for hierarchical text classification (HTC) using the RCV1 dataset. The analysis results showed that fastText was a classification method that provided very good results as word embedding, although the amount of data provided was relatively small. The precision, recall and F1 measure values were 0.920, 0.922, and 0.920, respectively. In contrast, our study focuses on deep learning to classify news articles based on their categories and machine learning for outlier detection.

- Shao et al. [13] experimented with the word2vec and doc2vec features for a clinical text classification task and compared the results with the traditional bag-of-words (BOW) feature. Learning showed that the word2vec feature performed better than the BOW-1-gram feature. In combination sets that were larger than six modalities (i.e., acupuncture, biofeedback, guided imagery, meditation, tai-chi and yoga), BOW-1,2-gram had better performance compared to the other feature extraction, with an area under curve (AUC) value of 0.91 and an accuracy of 0.85 specifically for the guided imagery. Meanwhile, in the smaller individual modality set, word2vec performed better compared to the other feature extraction, with AUC values between 0.80-0.93 and accuracy values between 0.82-0.86. The dataset used by Veterans Affairs (VA) electronic medical records (EMR) was stored in the Veterans Administration Informatics and Computing Infrastructure (VINCI) database. In contrast, this research only utilized word2vec in feature extraction.

- Research by Yan et al. [14] proposed the LSTM2 model for document classification consisting of repLSTM for the adaptive data representation process and rankLSTM for the integrated learning ranking process. In repLSTM, the supervised LSTM was used to study document representation by inserting label documents. In rankLSTM, the order of document labels was rearranged according to the semantic tree, where the semantics were compatible with and conformed to LSTM sequential learning. The word embedding used was BoW, and the dataset was taken from Bio (10C), email and News. The model achieved F1 Measure results of less than 75% in document classification tasks [14]. Meanwhile, the research in the article builds the news article classification model using LSTM+CNN.

- Gao et al. [9] proposed a text feature that combined the word2vec neural network model and the Latent Dirichlet Allocation (LDA) document topic model. Word2Vec and LDA represented the matrix model. The feature matrix was entered into a CNN for convolution pooling, and text classification experiments were performed. The experimental data come from the Sogou corpus text classification lab with 8,000 documents from sports, military, tourism, finance, IT, real estate, education, entertainment, and eight categories of 1000 experiments. The experimental results suggested that the proposed matrix model had a better classification effect than the traditional text classification method based on word2vec and CNN. At the level of text classification accuracy, the recall rate and F1 of the three evaluation indicators increased by 8.4%, 8.9% and 8.6% [9]. The similarity between the previous and current studies is that both use Word2Vec and LDA in feature extraction.

- Other research conducted by Yuan et al. [15] proposed a weighted word2vec, adding an attention mechanism to the LSTM model for emotion classification. After the text information was encoded into the word vector by word2vec, the weight matrix was combined with TFIDF to form the LSTM input. The dataset used is English data set and Chinese data set. The English dataset is an IMDB film review set consisting of 25,000 movie data, positive and negative values of 12,500 texts each. The Chinese dataset was collected from the hotel review corpora (Chn Senti Corp.) with 6000 data where the positive and negative values were 3,000 texts each. The experimental results showed that this method had a precision, recall and F1 measure of 0.87, respectively [15]. In contrast, the research in the article builds the news article classification model using LSTM+CNN.

- Subsequent research by Wang et al. [14] investigated label embedding for text representation and proposed a label-embedding attentive model. The model embedded words and labels in the same merged space and measured the compatibility of word-label pairs to attend to document representation. The learning framework was tested on five datasets (i.e., AGNews, Yelp Review Full, Yelp Review Polarity, DBPedia and Yahoo! Answers Topic) and clinical text applications. The investigation results indicated that the proposed LEAM (Label-Embedding Attentive Model) algorithm required much lower computational costs, and achieved better performance compared to CNN, LSTM, Simple Word Embeddings-based Models (SWEM) and bi-directional blocks self-attention network (Bi-BloSAN). Following predictive performance [16], F1 and micro mean area was 0.91 and macro average was 0.88 under the ROC (AUC) curve, and precision at n (P@n) was 0.61. In contrast, the research in the article builds the news article classification model using LSTM+CNN. The similarity between the previous and current studies is the use of ROC (AUC) curve to evaluate the model.

- Sun & Chen [17] designed a short text classification method based on word vectors and the proposed LDA topic model by considering the combined weighting factors of grammatical categories and high frequency topic words. In this method, Gibbs sampling was used to train the LDA topic model based on the weight of the part of speech. The model was exercised using the word vector Wor2vec and vectorized with high frequency topics. Then, the ex-tend text feature was tested. After expanding the feature, the SVM algorithm was used to classify the extended short text, and the classification results were evaluated using Precision (83.6), F1-score (84.4), and recall (85.4). The dataset was taken from the news corpus provided by Sogou Lab, which consisted of a total of 6,000 titles after being extracted, divided into six categories: computer, health, sports, tourism, education and military. Each category had 1000 essays, which was less than 200 words and belonged to text data short [17]. The similarity between the previous and current studies is that both used Word2Vec and LDA in feature extraction, but the algorithm used in classification was different.

- Xu et al. [18] proposed a new topic-based skip-gram neural language model to study topic-based word embedding for indexing biomedical literature with CNN. Topic-based skip-grams utilized textual content with topic models, for example, LDA, to capture topic-based precise word relationships and then integrated them into distributed word embedding learning. The combination of topic-based Skip-grams and multimodal CNN architecture outperformed advanced methods in indexing biomedical literature, annotating clinical records, and general textual dataset classification. The performance of the model was measured using the F1 score with a value of 82.7% [18]. In contrast, the research in the article builds the news article classification model using LSTM+CNN.

- In this study, we present a new active learning method for text categorization. The main goal of active learning is to reduce labeling effort without compromising classification accuracy by intelligently choosing which samples to label. The proposed method selects an informative sample set using the posterior probabilities provided by a set of multi-class SVM classifiers, and these samples are then manually labeled by an expert. The datasets used are public datasets in the text category (TC), namely Reuters-21578 document (R8), 20ng dataset and WebKB collection. Word embedding used was TFIDF. The accuracy of each dataset varies where R8, 20ng and WebKB had values of 83.33%, 43.79% and 53.07%, respectively [19].

- Verma [3] compared four very prominent algorithms for news classification, namely Naïve Bayes (NB), SVM, Random Forest and Multi-layer Perceptron (MLP) Classifier. Compared to the other approaches, Naïve Bayes is likely to be a better approach to serve as a text classification model because of its homogeneity. The paper proposed news classification by comparing four classifiers in which several different types of news have been classified, such as business & finance, sports, politics & policy, criminal justice, and health [3]. In contrast, the research in the article builds the news article classification model using LSTM+CNN.

- Next, Azan et al. [20] analyzed the performance of the classification algorithm using the Scopus dataset. In text classification, classification and feature extraction from documents using the extracted features were the main problems in reducing performance of different algorithms. The performance of classification algorithms such as NB and K-Nearest Neighbor (K-NN) showed a better improvement using Bayesian boost and bagging. Data preprocessing and cleaning steps were induced on the selected data set, and class imbalance issues were analyzed to improve the performance of the text classification algorithm. The overall accuracy of NB and KNN was 71.11% and 78.67%, respectively. The experimental results showed that KNN's performance was better than NB's [20]. Our study builds the news article classification model using LSTM+CNN, while the previous research builds the model using a machine learning algorithm.

- Wongso et al. [2] analyzed the suitable algorithm to classify news articles in Indonesian automatically. The dataset was retrieved using a web crawling method from www.cnnindonesia.com. The document first underwent several text preprocessing methods (i.e., lemmatization and stop word removal), followed by applying feature selection (i.e., term frequency-inverse document frequency (TF-IDF) and singular value decomposition (SVD) algorithms) and classification algorithms for Multinomial Nave Bayes, Multivariate Bernoulli Naïve Bayes, and Support Vector Machine. The test results suggested that the combination of TF-IDF and Multinomial Naïve Bayes Classifier gave the highest results compared to the other algorithms, with a precision of 0.9841519 and a recall of 0.9840000. The results outperformed previous similar studies that classified Indonesian-language news articles with an accuracy of 85% [2]. The previous research builds the model using a machine learning algorithm, while the current study builds the news article classification model using LSTM+CNN.

- Fagbola et al. [4] evaluated the accuracy and efficiency of the computational time of Kolmogorov Complexity Distance Measure (KCDM) and Artificial Neural Network (ANN) for large-dimensional news article classification problems. Dataset used by British Broadcasting Corporation (BBC) News. The dataset consisted of 2225 news articles in five categories: politics (417), sports (511), entertainment (386), education and technology (401) and business (510). Porter's algorithm was used for words stemming after tokenization and deletion of stop words, and Normalized Term Frequency–Inverse Document Frequency (NTF-IDF) was adopted for feature extraction. Experimental results showed that ANN performed better in terms of accuracy while KCDM produced better results than ANN in terms of computational time efficiency [4]. The similarity of the previous study with the current study is that both utilized BBC dataset to evaluate the model. Nevertheless, the algorithm used in the feature extraction is different.

- Sunagar & Kanavalli [21] dealt with the complexities involved in the text classification process. The experiment was carried out with the implementation of the RNN+LSTM+Gated Recurrent Unit (GRU) model. This model was compared with RCNN+LSTM and RNN+GRU. The model was tested using the GloVe dataset. The accuracy and recall obtained from the models were assessed. The F1 score was used to compare the performance of the two models. The hybrid RNN model had three LSTM layers and two GRU layers, while the RCNN model contained four convolution layers and four LSTM levels, and the RNN model contained four GRU layers. The weighted average for the hybrid RNN model was found to be 0.74, RCNN+LSTM was 0.69, and RNN+GRU was 0.77. The RNN+LSTM+GRU model showed moderate accuracy in the initial epoch, but the accuracy slowly increased as the epoch increased. In contrast, the research in the article builds the news article classification model using LSTM+CNN.

Several related studies (review papers) on text classification suggest that machine learning algorithms are generally used for text classification combined with word embedding models. Text classification using a deep learning method that combines RNN/LSTM and CNN is still limited. In this study, we proposed a hybrid modeling (RNN/LSTM + CNN) with a word embedding feature (i.e., Word2vec + LSA/SVD) for news text classification and coupled with a model for outlier detection using machine learning algorithms, especially the decision tree algorithm.

III. METHODOLOGY

The hybrid model developed in this study consisted of two models, namely a classification model using a deep learning algorithm (i.e., RNN/LSTM+CNN) and an outlier detection model using a machine learning algorithm (i.e., decision tree). LSTM-CNN was used to classify news articles based on Word2Vec-based context and LSA/SVD-based content. The outlier detection model was intended to predict which news was an outlier. Before the dataset was ready to be used for modeling, the dataset needed to be processed in several stages, such as text processing, data sharing into training and testing data, and feature extraction (word embedding) contained in news texts. After the data was ready to be used, the data was used for model training, which was then employed for data testing to validate the prediction results of the two models. In general, the hybrid model developed is presented in Fig. 1.

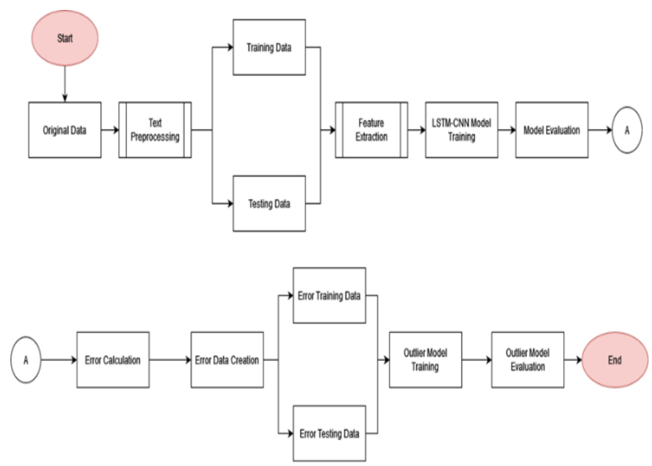


Fig. 1. Hybrid model framework

The classification results from the LSTM+CNN model was used as input in the outlier detection process. There would be errors (outliers) from the classification results generated by the model. For example, if a news item had topic A, the model might predict the probability that the news would be classified as topic B. Specifically, the use of the term “topic” in the research referred to content. Later, each news item would be labeled as an outlier or not an outlier. Furthermore, if topic A is at number 0.4, then there will be an error of 0.6. Once the error calculation result was obtained, the root mean square error (RMSE) for the entire dataset could be calculated. Sample/raw data that had an error of more than the RMSE would be categorized as an outlier. The outlier labeling enabled us to build an outlier classification model.

A. Text Preprocessing

The stages of text preprocessing news articles from raw data to ready-to-use data consisted of four stages (as presented in Fig. 2), namely deletion of symbols and numbers, tokenization, deletion of stop words, and lemmatization. The removal of symbols and numbers was done because symbols and numbers did not have a special meaning that was correlated with the topic of the news. Then, tokenization was done to break sentences or paragraphs in the news into chunks of words so that the model could read them. Furthermore, words with no meaning, such as subject (I, he, she, etc.), were removed to reduce noise in the data. Finally, lemmatization was used to reduce the word form to its simplest form, such as “eating” to “eat”.

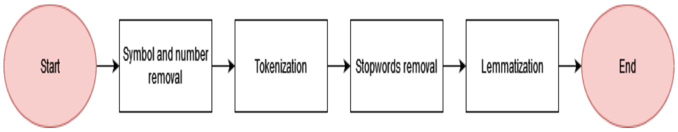


Fig. 2. Stages of text preprocessing

B. Feature Extraction

The news text that had been cleaned was not yet fully usable by the model. Feature extraction was used to extract information from the text. The feature extraction stages are presented in Fig. 3. This research used two extraction methods, namely context and content-based information/news extraction. The Word2Vec model was used to extract context-based information. Each word from the text processing above was converted into a 100-dimensional vector. A window of value 5 was used to obtain the vector, meaning that five surrounding words would be used to understand the context of a word in the paragraph. The results of the Word2Vec model were vectors with dimensions of 100, with the number of vectors as many as the number of words contained in the training data.

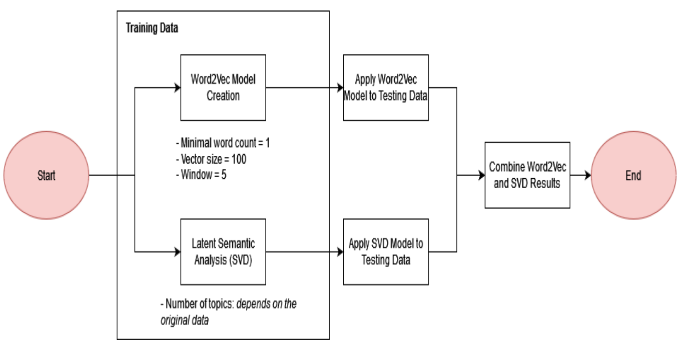


Fig. 3. Stages of feature extraction

The method used to extract content-based news was the SVD model. This model broke down the matrix of words obtained from the text processing results into three matrices, namely a matrix containing the relationship between news and words, a matrix containing the relationship between news and topics, and information containing the relationship between topics and words. The matrix used was the relationship matrix between topics and words so that each word had a vector containing information about its relationship to the topics. After obtaining two types of vectors from the results of Word2Vec and SVD, the two outputs were combined into one vector. This vector was called the embedding vector, which represented one word.

C. Hybrid Model of LSTM+CNN Architecture

Fig. 4 presents the architecture of the classification model of the Hybrid model using two deep learning algorithms (i.e., RNN/LSTM+CNN).

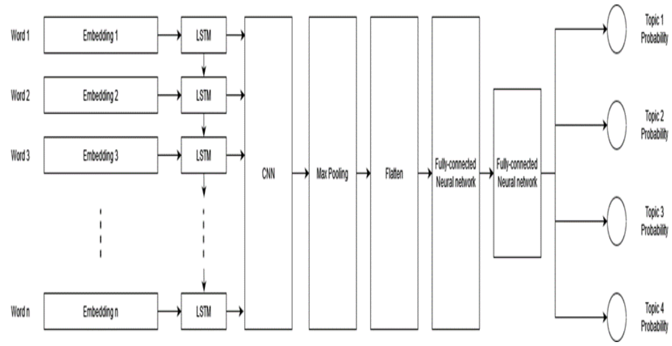


Fig. 4. Hybrid model architecture

Embedding vector from feature extraction was used in LSTM-CNN modeling as an embedding layer. When there was news used as model input, every word in the news would be transformed into the corresponding embedding vector. Then, the embedding vector became the LSTM input layer. The LSTM cells used were 64 cells. The output of each LSTM cell would be used as the CNN input layer, with a filter size of 100, a kernel size of 2, and a ReLu activation function. The CNN output matrix was inserted into the max pooling layer to reduce noise in the output. After the noise in the output was reduced, the output, which had been converted into a 1-dimensional vector in the flattened layer, can be input to a fully connected neural network (NN). Here, we use 2 NN layers, where the first layer had 16 nodes and the ReLu activation function, while the second layer had the number of nodes corresponding to the number of topics in the data. The softmax activation function was used in the second layer to obtain the output in the form of the probability of each topic (content).

D. Model Evaluation

The ROC curve provides a graphical representation of the performance of the classifier. The ROC curve was generated by calculating and plotting the TPR against the false positive rate (FPR) for a single classifier at various thresholds. The TPR and FPR equations are presented in equations 1 and 2.

$$TPR = Sensitivity = TP/(TP + TN) \quad (1)$$

$$FPR = 1 - Specificity = FP/(FP + TN) \quad (2)$$

Where TP is the number of true positives and FN is the number of false negatives. TPR is a measure of the probability that an actual positive event will be classified as positive. FP is the number of false positives and TN is the number of True Negatives. FPR is a measure of how often a "false alarm" will occur or how often an instance of a true negative will be classified as positive.

Visualization of ROC curve uses AUC. The higher the AUC score, the better the classifier performs for a given task. ROC curve with AUC score is presented in Fig. 5.

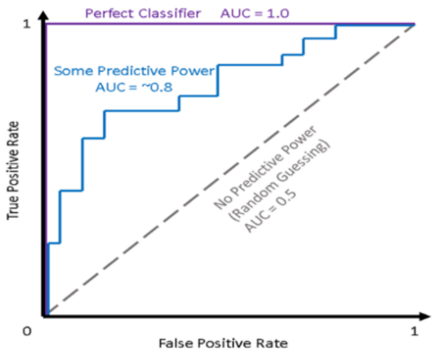


Fig. 5. ROC curve with AUC score

The classification category based on the ROC Curve value is shown in Table I [22]. In general, an AUC score of 0.6 - 0.7 indicates poor classification performance (failure), 0.7 - 0.8 is considered acceptable, 0.8 - 0.9 is considered very good and more than 0.9 is considered extraordinary. Model category based on ROC curve value [23] is presented in Table I.

TABLE I. MODEL CATEGORY BASED ON ROC CURVE

Accuracy	Category
0.90 – 1.00	Excellent Classification
0.80 – 0.90	Good Classification
0.70 – 0.80	Fair Classification
0.60 – 0.70	Poor Classification
0.50 – 0.60	Failure Classification

IV. EXPERIMENTAL RESULT

A. News Classification

The architecture used to build the model with the AGNews dataset consisted of: (1) Embedding matrix (the result of extracting content and context information that was carried out in the previous stage) with input size = 100; (2) LSTM with 64 nodes = 64; (3) CNN with parameters: filter = 100, kernel = 2, number of strides = 1, padding = valid, activation function = Relu; (4) Pooling; (5) Flatten; (6) Dense hidden layer with number of nodes = 16 and activation function = Relu; (7) Dropout rate = 0.5; (8) Dense hidden layer with the number of nodes = 4 (according to the number of targets) and the activation function = softmax. Meanwhile, the hybrid architecture model for the BBC News dataset, which were points 1 to 7, was the same as the AGNews dataset. The difference between the two dataset was in point 8. The dense hidden layer was used in BBC News with five nodes (according to the number of targets). This step was done to achieve the best set of hyperparameters by performing hyperparameter tuning.

The next step was the model fitting. This stage conducted model training, which would be stopped if there was no increase in the accuracy value in data testing for 10 iterations (epochs). The modeling results were then stored (Table II).

TABLE II. HYPER PARAMETERS

Layer (type)	Output Shape	Param #
Embedding_2 (Embedding)	(None, 100, 104)	6533800
Lstm_2 (LSTM)	(None, 100, 96)	77184
Conv1d_2 (Conv1D)	(None, 97, 100)	38500
Global_max_pooling1d_2 (GlobalMaxPooling1D)	(None, 100)	0
Flatten_2 (Flatten)	(None, 100)	0
Dense_4 (Dense)	(None, 10)	1010
Dense_5 (Dense)	(None, 4)	44

Total params: 6,650,538

Trainable params: 6,650,538

Non-trainable params: 0

After the hybrid model was built using training data with hyperparameter tuning, the model was evaluated using the confusion matrix (CM) and ROC curve. The test results after 235 iterations showed a loss value of 0.2818, accuracy of 0.9065, test loss of 0.28183448, test accuracy of 0.90653336. The CM on the model with the AGNews dataset suggested that the data had been predicted correctly for each actual label, as shown in Table III.

TABLE III. CM OF AGNEWS DATASET

	Accuracy 90.65				
Confusion Matrix	***Label***	***World***	***Entertainment***	***Sport***	***Business***
	World	6620	232	373	256
	Entertainment	65	7250	45	73
	Sport	199	83	6703	595
	Business	219	62	602	6623

The evaluation model using the CM on BBC News is presented in Table IV.

TABLE IV. CM OF BBC NEWS DATASET

	Accuracy 0.85					
Confusion Matrix	***Label***	***Entertainment***	***Tech***	***Politics***	***Business***	***Sport***
	Entertainment	64	2	1	14	0
	Tech	2	52	0	6	0
	Politics	0	2	49	12	0
	Business	3	2	5	76	0
	Sport	1	0	1	5	76

The accuracy of the test on the AGNews dataset was 91%, the ROC curve value was 0.9832 and the training validation had a loss value of 0.1146 and an accuracy of 0.9611 (96.11%). The results of accuracy in training, testing and validation as well as the ROC curve value suggested that the model had good performance. The ROC curve for AGNews is presented in Fig. 6.

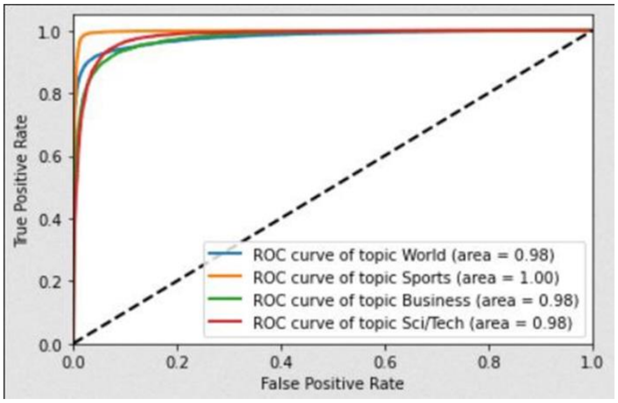


Fig. 6. ROC curve of the AGNews dataset

The ROC Curve of the BBC News dataset is presented in Fig. 7.

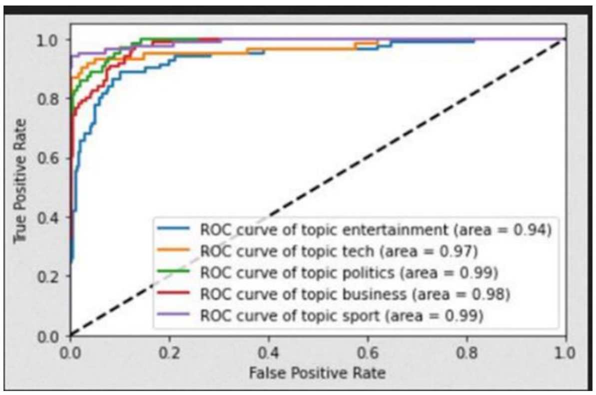


Fig. 7. ROC curve of the BBC News dataset

The ROC curve showed the visualization between TPR and FPR. The classifier that provides the curve closer to the top left corner (perfect classifier) exhibits better performance. The closer the curve is to the 45-degree diagonal of the ROC space, the less accurate the classifier is.

TABLE V. PERFORMANCE EVALUATION OF TWO DATASET

Model	***Class***	***Precision***	***Recall***	***F1-score***	***Support***
AG News	***World***	0.93	0.88	0.91	7481
	Entertainment	0.95	0.98	0.96	7433
	Sport	0.87	0.88	0.88	7580
	Business	0.88	0.88	0.88	7506
			Accuracy	0.91	30000
Model	***Class***	***Precision***	***Recall***	***F1-score***	***Support***
BBC News	***Entertainment***	0.91	0.79	0.85	81
	Tech	0.90	0.87	0.88	60
	Politics	0.80	0.78	0.82	6
	Business	0.67	0.88	0.76	86
	Sport	1.00	0.92	0.96	83
Accuracy				0.85	373

The model was trained and tested against two datasets, namely the AGNews and BBC News datasets. The results of the model evaluation showed that the LSTM+CNN hybrid model had excellent accuracy (as presented in Table V) and ROC AUC scores.

B. Outlier Detection

The first step in outlier detection is processing error data. At this stage, the prediction results of the error model will be collected into 1 (one) separate data frame consisting of test topics and train topics. The magnitude of the error is calculated through the equation: error = 1 – max(xi), i∈{1,2,3,4}error = 1 – max(xi), i∈{1,2,3,4}, where x_i denotes the highest probability score that the model assigns. In this formulation, each prediction error can be calculated by the magnitude of the error, so that the RMSE error data can be calculated. The size of the RMSE dataset error was 0.93. Fig. 8 shows error data train, while Fig. 9 shows error data test on the AGNews dataset.

	Text	Actual Label	Topic 0 score	Topic 1 score	Topic 2 score	Topic 3 score	Topic	Error	outlier_label
2182	[president, hugo, chavez, rid, high, overwhelm...	0	0.03	0.00	0.96	0.00	2	0.04	Normal
75249	[caetano, defender, serginho, take, hospital, ...	0	0.15	0.85	0.00	0.00	1	0.15	Normal
57388	[european, union, agree, lift, long, stand, sa...	0	0.24	0.00	0.75	0.00	2	0.25	Normal
38321	[brussels, european, commission, confirm, want...	3	0.02	0.00	0.72	0.26	2	0.28	Normal
49457	[initial, public, offer, share, newly, establi...	3	0.04	0.00	0.59	0.37	2	0.41	Outlier

Fig. 8. Error data train on AGNews

	Text	Actual Label	Topic 0 score	Topic 1 score	Topic 2 score	Topic 3 score	Topic	Error	outlier_label
82611	[personal, england, wales, another, record, hi...	2	0.12	0.61	0.23	0.05	1	0.39	Outlier
8679	[bangladesh, captain, bashar, rule, next, mont...	0	0.40	0.59	0.00	0.00	1	0.41	Outlier
44571	[internet, phone, service, sipphone, charge, v...	2	0.00	0.00	0.09	0.90	3	0.10	Normal
81887	[people, least, hear, still, think, intend, sp...	2	0.10	0.02	0.31	0.57	3	0.43	Outlier
98948	[cincinnati, still, among, dangerous, cities, ...	2	0.28	0.01	0.32	0.40	3	0.60	Outlier

Fig. 9. Error data test on AGNews

The amount of data with an error greater than the RMSE was around 65% and this data would be considered as an outlier. Fig. 10 shows error data test on the AGNews dataset.

	Text	Actual Label	Topic 0 score	Topic 1 score	Topic 2 score	Topic 3 score	Topic 4 score	Topic	Error
1422	[firm, embrace, ecommerce, firm, embrace, inte...	2	0.00	0.99	0.00	0.01	0.00	1	0.01
1071	[allow, scrutiny, urge, give, watchdogs, freed...	0	0.00	0.98	0.02	0.00	0.00	1	0.02
450	[national, gallery, pink, national, gallery, h...	0	0.00	1.00	0.00	0.00	0.00	1	0.00
264	[franz, seek, government, help, franz, ferdina...	0	0.12	0.88	0.00	0.00	0.00	1	0.12
1072	[game, maker, fight, survival, britain, larges...	1	0.00	0.25	0.00	0.75	0.00	3	0.25
186	[telegraph, newspapers, job, daily, sunday, te...	3	0.00	0.96	0.00	0.04	0.00	1	0.04
1087	[blue, slam, blackburn, savage, birmingham, co...	4	0.00	0.59	0.15	0.00	0.26	1	0.41
89	[campbell, lions, consultant, former, governme...	4	0.00	0.00	0.79	0.00	0.20	2	0.21
664	[orange, colour, clash, court, colour, orange,...	3	0.00	0.98	0.01	0.00	0.00	1	0.02
819	[rule, tackle, weddings, rule, marriage, forei...	2	0.00	0.97	0.01	0.02	0.00	1	0.03
1337	[glastonbury, fan, card, fan, ticket, year, gl...	0	0.00	0.61	0.31	0.08	0.00	1	0.39
135	[portfolio, sale, collection, murder, fashion,...	0	0.04	0.93	0.00	0.00	0.03	1	0.07
369	[lead, interactive, bafta, win, national, thea...	1	1.00	0.00	0.00	0.00	0.00	0	0.00

Fig. 10. Error data test on BBC News

To predict whether a news item is an outlier, two types of systems were designed. In the first system (Fig. 11), the outlier classifier model was installed separately from the LSTM+CNN hybrid model. As for the second system (Fig. 12), the meta-modeling principle was used, namely the outlier classifier model using the prediction results of another model (the LSTM+CNN hybrid model for this case) to determine whether a news item is an outlier news or not. Decision tree models were used in both systems to determine which system had the best performance in determining outliers.

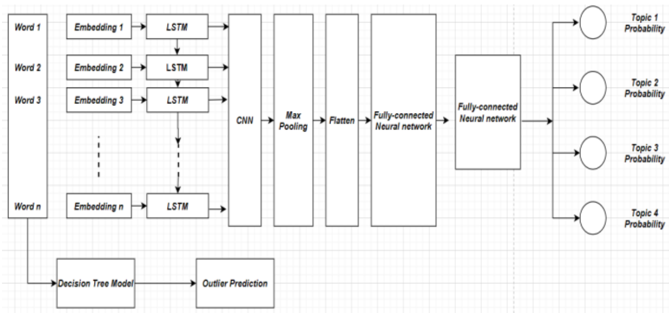


Fig. 11. The First Model of the classification of news outliers

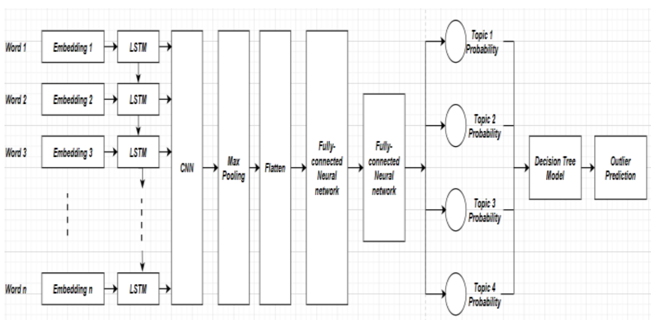


Fig. 12. The second model of the classification of news outliers

C. Validation Model

Validation models of the two classification models are presented in Fig. 13. Fig. 13 shows that system 2 is far superior to system 1. In system 1, the ROC AUC score was only around 0.5-0.6, meaning that the model predictions were not much different from the random prediction results. However, in system 2, the ROC AUC score could reach 0.8 - 0.9, indicating that by obtaining input from the probability prediction results of the LSTM-CNN hybrid model, the outlier classifier model can properly determine whether a news item is an outlier news or not.

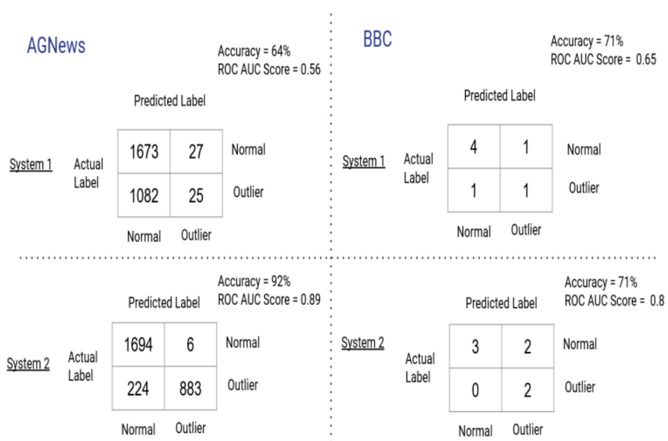


Fig. 13. Comparison of outlier news classifications

D. Comparative Study

Table VI is a comparison of previous research related to news article classification. The results of the performance evaluation of training and testing of the developed hybrid model showed better results. Based on the ROC curve, the hybrid model developed is in the excellent classification category. This can prove that the hybrid model with RNN/LSTM+CNN architecture and feature extraction Wor2Vec+LSA/SVD can be utilized as a good method for classifying sequential text. Furthermore, the hybrid model developed can perform news outlier detection well. The last two approaches in Table VI are the current research results (presented in bold).

TABLE VI. COMPARISON WITH PREVIOUS RESEARCH

Word Embedding	**Text Classification Strategies**	**Dataset**	**Result**
GloVe [21]	• RNN + LSTM + GRU • RCNN+ LSTM • RNN+GRU	GloVe	Accuracy = 0.74 Accuracy = 0.69 Accuracy = 0.77
GloVe, word2vec, and fastText [12]	CNN, SVM, XGBoost	RCV1	Precision value = 0.92, Recall = 0.92 and F1 measure = 0.920
Word2Vec, doc2vec, BoW [13]	SVM	VINCI	AUC value between 0.80 -0.93 and accuracy value between 0.82-0.86
BoW [14]	LSTM2: repLSTM and rankLSTM	Bio (10C), email and News	F1 Measure less than 75%
Word2Vec +LDA [9]	CNN	Sogou Corpus text classificatio n Lab	Accuracy = 0.84, recall = 0.89 and f1-score = 0.86
Word2Vec + TF-IDF [15]	Att-LSTM	IMDB film review and hotel review corpora	Precision value = 0.87, recall = 0.87 and F1 measure = 0.87
[24]	Label-Embedding Attentive Model (LEAM)	AGNews, Yelp Review Full, Yelp Review Polarity, DBPedia and Yahoo! Answers Topic	F1 micro average = 0.91 and macro average = 0.88 under the ROC (AUC) curve, precision = 0.61
LDA + Word2Vec [17]	SVM	Corpus News	Precision value = 83.6, F1-score = 84.4 and recall = 85.4.
LDA [18]	Skip-Gram and CNN	Biomedical literature,	F1 score = 82.7%.

		clinical record annotation	
TF-IDF [19]	SVM	Reuters-21578 document (R8), 20ng dataset and WebKB collection	The accuracy of the dataset varies R8 = 83.33%, 20ng = 43.79% and WebKB = 53.07
[3]	Naïve Bayes, SVM, Random Forest, MLP Classifier	News categories	The Support Vector Classifier has the highest accuracy of 0.6134
TF-IDF [20]	Naïve Bayes (NB) and K-Nearest Neighbor (K-NN)	Corpus Dataset	Naïve Bayes accuracy is 71.11%, and KNN is 78.67%.
TF-IDF dan SVD [2]	Multinomial Nave Bayes, Multivariate Bernoulli Naïve Bayes, and Support Vector Machine	CNN indonesia	TF-IDF and Multinomial Naïve Bayes Classifier provide the highest results compared to other algorithms, with a precision of 0.9841519 and a recall of 0.984
Normalized TF-IDF (NTF-IDF) [4]	Kolmogorov Complexity Distance Measure (KCDM) and ANN	British Broadcasting Corporation (BBC) News	ANN performs better in terms of accuracy while KCDM produces better results than ANN in terms of computational time efficiency.
Word2Vec+LSA/SVD	**RNN/LSTM+ CNN**	**AGNews BBC News**	**AGNews Accuracy = 0.91, BBC News accuracy = 0.85, AGNews ROC curve between 0.98 – 1.00; BBC News ROC curve between 0.94 – 0.99.**
Word2Vec+LSA/SVD	**News Classification (RNN/LSTM+CNN) & Outlier Detection (architecture 2)**	**AGNews BBC News**	**AGNews: Accuracy = 0.92 and ROC curve = 0.89; BBC News: Accuracy = 0.71 and ROC curve = 0.8**

V. Conclusion

In this study, we developed a hybrid model consisting of two models: the news classification model (news categories) and the outlier classification model. The news classification model employed a deep learning algorithm (i.e., LSTM+CNN) based on the context and content of a news story. Meanwhile, the outlier classifier model was intended to predict which news is an outlier. The datasets used for modeling were AGNews and BBC News. The Word2Vec model was used to extract context- and content-based news using the SVD model. Our results on AGNews showed an accuracy of 0.91 with a ROC curve score of 0.97, while BBC News had an accuracy value of 0.86 with a ROC curve score of 0.96. These results suggested that the hybrid model had excellent accuracy and ROC AUC scores.

The process of labeling news included the detection of outliers. The predicted data from the LSTM-CNN model was used and tested with two types of models. First, the outlier classifier model was installed separately from the LSTM-CNN hybrid model. The second model used the meta-modeling principle, namely the outlier classifier model using the prediction results of another model (LSTM-CNN hybrid model). The algorithm used in the outlier classification model was a decision tree. Of the two models tested, the second model was far superior to the first model. In the first model, the ROC AUC score was only around 0.5-0.6, which indicated that the model predictions were not much different from the random prediction results. However, in the second model, the ROC AUC score could reach 0.8 - 0.9, meaning that by obtaining input from the probability prediction results of the LSTM-CNN hybrid model, the outlier classifier model could properly determine which news item was an outlier, with accuracy values of 0.71 and 0.92 for BBC News AGNews, respectively.

The drawback of the results of this study is that the proposed hybrid model did not reach maximum accuracy. The sample used in the dataset to evaluate the performance of the model is still limited, causing the model to have difficulties learning the vocabulary used in classifying news articles based on their categories. This situation may lead to overfitting. Furthermore, the model only tested news articles in English. Further research is needed to develop the proposed hybrid model, combined with other deep learning algorithms and more dataset samples, to obtain optimum model performance. The performance of the developed model also needs to be tested using news articles other than in English, such as Indonesian, Mandarin, Arabic and other languages.

References

[1] D. Liparas, Y. HaCohen-Kerner, A. Moumtzidou, S. Vrochidis, and I. Kompatsiaris, "News articles classification using random forests and weighted multimodal features," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8849, pp. 63–75, October 2014, doi: 10.1007/978-3-319-12979-2_6.

[2] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, and Rudy, "News article text classification in Indonesian language," Procedia Comput. Sci., vol. 116, pp. 137–143, 2017, doi: 10.1016/j.procs.2017.10.039.

[3] P. K. Verma, "Exploration of text classification approach to classify news classification," Int. J. Adv. Sci. Technol., vol. 29, no. 5, pp. 2555–2562, 2020.

[4] T. M. Fagbola, C. S. Thakur, and O. Olugbara, "News article classification using Kolmogorov Complexity Distance Measure and Artificial Neural Network," Int. J. Technol., vol. 10, no. 4, pp. 710–720, 2019.

[5] S. Biradar and M. M. Raikar, "Performance analysis of text classifiers based on news articles-a Survey," Indian J. Sci. Res., vol. 15, no. 2, pp. 156–161, 2017.

[6] I. C. Irsan and M. L. Khodra, "Hierarchical multi-label news article classification with distributed semantic model based features," Int. J. Adv. Intell. Informatics, vol. 5, no. 1, pp. 40–47, 2019, doi: 10.26555/ijain.v5i1.168.

[7] W. K. Sari, D. P. Rini, and R. F. Malik, "Text classification using Long Short-Term Memory with GloVe features," J. Ilm. Tek. Elektro Komput. dan Inform., vol. 5, no. 2, pp. 85, 2020, doi: 10.26555/jiteki.v5i2.15021.

[8] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from TF-IDF to TF-IGM for term weighting in text classification," Expert Syst. Appl., vol. 66, pp. 1339–1351, 2016, doi: 10.1016/j.eswa.2016.09.009.

[9] M. Gao, T. Li, and P. Huang, Text classification research based on improved word2vec and CNN, vol. 11434 LNCS. Springer International Publishing, 2019.

[10] A. Noori, S. S. B. Kamaruddin, and F. B. K. Ahmad, "Towards an outlier detection model in text data stream," Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 6, pp. 2970–2973, 2019, doi: 10.30534/ijatcse/2019/47862019.

[11] P. Krammer, O. Habala, J. Mojžiš, L. Hluchý, and M. Jurkovič, "Anomaly detection method for online discussion," Procedia Comput. Sci., vol. 155, pp. 311–318, 2019, doi: https://doi.org/10.1016/j.procs.2019.08.045.

[12] A. J. Stein, J. Weerasinghe, S. Mancoridis, and R. Greenstadt, "News article text classification and summary for authors and topics," pp. 1–12, 2020, doi: 10.5121/csit.2020.101401.

[13] Y. Shao, S. Taylor, N. Marshall, C. Morioka, and Q. Zeng-Treitler, "Clinical text classification with word embedding features vs. Bag-of-Words features," Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018, pp. 2874–2878, 2019, doi: 10.1109/BigData.2018.8622345.

[14] Y. Yan, Y. Wang, W. C. Gao, B. W. Zhang, C. Yang, and X. C. Yin, "LSTM 2: multi-label ranking for document classification," Neural Process. Lett., vol. 47, no. 1, pp. 117–138, 2018, doi: 10.1007/s11063-017-9636-0.

[15] H. Yuan, Y. Wang, X. Feng, and S. Sun, "Sentiment analysis based on weighted Word2vec and ATT-LSTM," ACM Int. Conf. Proceeding Ser., pp. 420–424, 2018, doi: 10.1145/3297156.3297228.

[16] J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein, "Explainable prediction of medical codes from clinical text," NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 1, pp. 1101–1111, 2018, doi: 10.18653/v1/n18-1100.

[17] F. Sun and H. Chen, "Feature extension for Chinese short text classification based on LDA and Word2vec," Proc. 13th IEEE Conf. Ind. Electron. Appl. ICIEA 2018, no. 1, hal. 1189–1194, 2018, doi: 10.1109/ICIEA.2018.8397890.

[18] H. Xu, A. Kotov, M. Dong, A. I. Carcone, D. Zhu, and S. Naar-King, "Text classification with topic-based word embedding and Convolutional Neural Networks," ACM-BCB 2016 - 7th ACM Conf. Bioinformatics, Comput. Biol. Heal. Informatics, no. April 2019, pp. 88–97, 2016, doi: 10.1145/2975167.2975176.

[19] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A novel active learning method using SVM for text classification," Int. J. Autom. Comput., vol. 15, no. 3, pp. 290–298, 2018, doi: 10.1007/s11633-015-0912-z.

[20] M. Azam, T. Ahmed, F. Sabah, and M. I. Hussain, "Feature extraction based text classification using K-Nearest Neighbor Algorithm," IJCSNS Int. J. Comput. Sci. Netw. Secur., vol. 18, no. 12, pp. 95–101, 2018, [Daring]. Available on: http://paper.ijcsns.org/07_book/201812/20181213.pdf.

[21] P. Sunagar and A. Kanavalli, "A hybrid RNN based deep learning approach for text classification," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 6, pp. 289–295, 2022.

[22] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," J. Thorac. Oncol., vol. 5, no. 9, pp. 1315–1316, 2010, doi: 10.1097/JTO.0b013e3181ec173d.

[23] F. Gorunescu, Data Mining: Concepts, models and techniques. Vol (12). Springer Science & Business Media, 2011.

[24] G. Wang et al., "Joint embedding of words and labels for text classification," ACL 2018 - 56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap., vol. 1, pp. 2321–2331, 2018, doi: 10.18653/v1/p18-1216.

Web based Mitosis Detection on Breast Cancer Whole Slide Images using Faster R-CNN and YOLOv5

Mitosis Detection on Breast Cancer WSI using Faster R-CNN and YOLOv5

Dr. Rajasekaran Subramanian[1], Dr. R.Devika Rubi[2], Dr. Rohit Tapadia[3], Katakam Karthik[4], Mohammad Faseeh Ahmed[5], Allam Manudeep[6]

Associate Professor
Dept of Computer Science and Engineering, Keshav Memorial Institute of Technology, Hyderabad, India[1,2]
Consultant – Pathology, Tapadia Diagnostics Centre, Hyderabad, India[3]
Research Intern, Keshav Memorial Insitute of Technology, Hyderabad, India[4,5,6]

***Abstract*—Histological grading quantifies the tumor architecture and the cytology deviation of breast cancer against normal tissue. Nottingham Grading System grades the breast cancer classification and allots tumor scores. Mitotic detection is one of the major components in the Nottingham Grading System. Using a conventional microscope is time-consuming, semi-quantitative and has limited histological parameters. Digital scanners scan the tissue slice into high-resolution virtual images called whole slide images. Deep learning models on whole slide images provide a fast and accurate quantitative diagnosis. This paper proposes two deep learning models namely Faster R-CNN and YOLOv5 to detect mitosis on WSI. The proposed Deep Learning models uses 56258 annotated tiles for training/testing and provide F1 score as 84%. The proposed model uses a web-based imaging analysis and diagnosis platform called CADD4MBC for image uploading, Annotation and visualization. This paper proposes an end-to-end web based Deep Learning detection for Breast Cancer Mitosis.**

***Keywords*—Nottingham grading system; breast cancer biomarker; whole slide image; mitosis; faster R-CNN; YOLOv5**

I. Introduction

Histological grading system is used to evaluate the behavior and prognosis of breast cancer on Hematoxylin and Eosin (H&E) stained images which quantifies the tumor architecture and the cytology deviation of the breast cancer tissue against the normal tissue. The grading also provides a degree of differentiation in the morphological assessment of breast cancer. As clinical stages are unable to provide early as well as developed lymph-blood meta stages information, Bloom and Richardson defined a histological grading system. This system allotted tumor scores between 1-3 based on individual components such as mitotic nuclei, nuclear pleomorphism, and tubule formation [1]. Their grading system was not accepted as a routine procedure for breast cancer grading classification due to the inconsistent issue in grading [2].

Elston-Ellis modified the Bloom and Richardson system and called their grading system as Nottingham grading system (NGS). NGS is accepted globally as a guide for grading breast cancer classification due to its semi-quantitative assessment of three morphological components such as a number of mitotic figures in the most active area called a high-power field, the size and shape of nuclear variation in tumor tissue against small regular uniform cells (Nuclear Pleomorphism) and the percentage of tumor tubule formation [3]. Hence the detection of mitosis is the major component of NGS. Mitosis is a cell duplication process that divides the cell into two cells which are genetically identical. The rapid and irregular mitosis cell count decides the tumor grading and the selection of tumor treatment options.

Histological grading on glass tissues using conventional microscopes is time-consuming which also has a limited evaluation of histological Parameters and semi-quantitative properties which is subject to high inter-observer variability. These limitations are overcome by the advent of digital pathology and Whole Slide Image (WSI) technology. WSI technology uses digital scanners to create high-resolution virtual images of the glass slide. The digitized images are stored as pyramid structures to view and analyze WSI at various Zoom levels (5x - 40x) [4]. Deep learning models performed over WSI virtual images provide fast and accurate diagnoses [5].

This paper proposes 1) Creation of digitized WSI Breast Cancer dataset; 2) performs mitosis annotations on the created dataset by pathologists; 3) develop two deep learning models, namely, Faster R-CNN and YOLOv5 to detect mitosis on WSI. This proposal reduces the difficulties of pathologists such as time consuming and inter-observer variability. The proposed models produce quantitative results. The paper is organized as follows: Section 2 lists the existing works, Section 3 proposes Faster R-CNN and YOLOv5 models and describes the dataset and preprocessing methods used by the deep learning models, Section 4 discusses the implementation/results and Section 5 conclude the summary of the proposed work.

II. Literature Survey

MP-Mitdet [6] used a multiphase CNN framework to detect mitosis using the public dataset, namely MITOS 12 [7], MITOS 14, AMIDA 13 [8] and TUPAC 16 [9] with the evaluation parameters as F1 Score 0.74 and precision 0.71. The model performs Mean-Standard normalization as the pre-processing procedure on the different public datasets which are captured by different scanners. Model Mask-RCNN is used for automatic labelling and detection of mitosis. Resnet is used for cell-level classification. The model also proved that deep CNN's performance is better than conventional classifiers such as SVM, Logistic Regression, XGBoost, Random Forest and Navies Bayes.

The Deep Mitosis model [10] adopts three components to detect mitosis using a deep detection model on 205 weakly annotated mitosis. The first component DeepDet produces all possible detection of mitosis, the second component DeepVer removes false positives, and the final component DeepSeg uses RPN to segment the mitosis. They conclude that the model DeepVer reduces the performance and other models produce an F1 Score of 0.38.

Meriem Sebai et al [11] use two datasets ICPR12 & ICPR14 for the localization, classification and segmentation of mitosis. The weakly annotated mitosis dataset i.e. ICPR 14 which has labels only the centroid of mitosis is trained by Mask-RCNN Model [12]. The model segments the mitosis by using the pixel-level annotated mitosis dataset called ICPR12. The model Mask-RCNN is used as a two-stage deep learning framework in which the first stage identifies the centroid of the mitosis and in the subsequent stage detects the instance segmentation of mitosis. This model produces an F1 Score of 0.863 on the 2012 ICPR and achieves an F1 Score of 0.475 on the 2014 ICPR datasets.

De Cai et al. [13] segments mitotic cells by using the deep learning model called Faster R-CNN [14]. This two-stage object detector model first identifies the possible mitotic cells and then the second stage detects the target mitotic cells from the result of the first stage which produces higher accuracy than the single-stage detectors and also reduces the computation time. This model achieved a 0.76 recall value and 0.736 F1 Score on Miccai and TUPAC datasets.

Dan C Cireşan et al. [15] proposed a supervised model called Deep Neural network to detect mitosis on H&E images. This model uses Max Pooling layers as a subsampling layer to classify whether a cell is a mitosis or not. The model is trained and tested by two public datasets namely ICPR12 and ICPR14. In the training datasets, the cells in the H&E images are labelled as mitosis or non-mitosis based on the number of pixels which are closer to mitosis. The training dataset also uses bounding boxes to label mitosis. This model achieved an F1 Score of 0.782.

The research done by Gabriel Jiménez et al [16] suggests two deep learning architectures namely CNN and U-net [17] for mitosis classification and detection in histopathological tissue samples. The images used to evaluate the proposed approaches were obtained from two public datasets from the ICPR-2012 competition and the MITOS-ATYPIA-2014 challenge. Convolutional Neural Network is used for binary classification to classify mitotic and non-mitotic cells. The model got 95% testing accuracy with an F1 score of 94.35%. The model U-net is used for the semantic segmentation of mitosis and produces 0.9 F1 score accuracy.

MiNuGAN [18] automatically segments mitoses and nuclei using conditional generative adversarial networks [19] on the public datasets namely TUPAC16, ICPR14, and ICPR12 Datasets. The model uses 618 annotation files for training and 200 file images for testing with an F1 Score of 0.824.

The limitations in the existing literature are most of the researchers applied deep learning models only on the three public datasets namely ICPR 12, ICPR14 and Tupac in which, ICPR14 and Tupac are weakly annotated (only the mitotic centroids are labelled) and ICPR 12 is a strongly annotated dataset (all the pixels of mitotic cells are labelled). The combined data size of the three datasets is around 7000 only. The less data in deep learning training may produce less accuracy. Another main problem in mitosis detection is to differentiate normal cells from mitotic figures. If the pathologist missed this differentiation causes wrong annotations. The existing works are not providing information about the data collection /annotation of the mitosis dataset.

The above-mentioned limitations are overcome by the proposed system by training the deep learning models on both public datasets and private KMIT datasets. KMIT dataset is a private dataset contains breast cancer WSI, tile images with their respective mitotic figure annotations. CADD4MBC is a web-based deep learning platform developed internally by the authors which is used for uploading WSI/tiles, creating annotations, JSON downloading and visualization of WSI. Thus, the proposed system provides a web-based end-to-end AI mitosis detection which connects pathologists, technicians and hospitals.

III. Proposed System

A. Dataset Description

1) Public dataset: The proposed Deep learning models use three public datasets namely ICPR12, ICPR14, and TUPAC. Mitos Dataset is a public dataset of the mitosis competition held by ICPR in 2012. In each slide, the pathologists selected 10 high-power fields at 40X magnification. This dataset consists of 2994 files of size 512 x 512. After applying augmentation methods such as horizontal flipping and vertical flipping increase the files to 4367. The ICPR 2014 dataset was presented in the MITOS-ATYPIA-14 grand challenge and comprises 2400 files of size 512 x 512 at 40x magnification. Pathologists annotated a total of 1502 centroid pixels of mitosis (weakly labelled).

The TUPAC mitosis dataset contains 73 cases collected from the department of pathology at the university medical Centre in Utrecht, The Netherlands. Each case was represented by one WSI region with an area of 2mm x 2mm. The annotated mitotic figures are accepted by two pathologists and the dataset contains 1552 weakly annotated mitotic figures of the size 512 x 512.

2) KMIT dataset: KMIT dataset contains 75 WSI which are collected from Basavatarakam Indo-American Hospital and scanned by Tapadia diagnostics Centre using a Morphile scanner at 40x magnification for the last two years. The dataset contains around 56,000 mitotic figures annotated by pathologists from both hospitals using the CADD4MBC platform and each WSI varies from 8 - 12 GB. The 40X level magnification WSI is divided into several tiles (around 30,000 - 50,000) of size 500 x 500. Each uploaded WSI is divided into several (minimum 100) batches and each batch consist of 150 tiles as shown in Fig. 1.

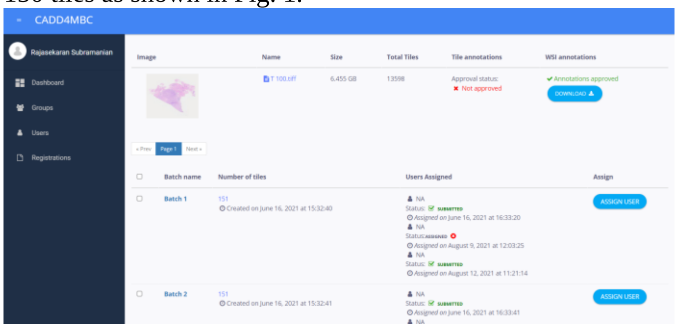


Fig. 1. Uploaded WSI and the batches of tiles

Each tile is annotated by the pathologists using the drawing tools in the CADD4MBC as shown in Fig. 2.

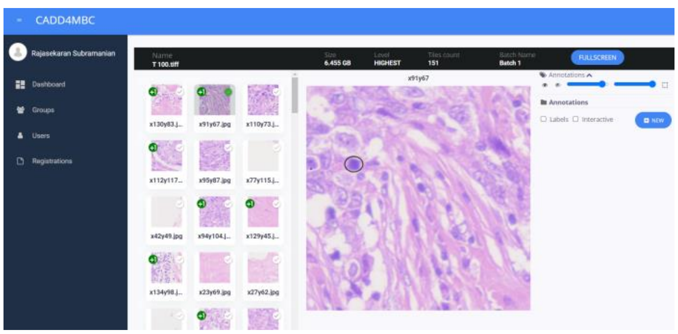


Fig. 2. Mitosis annotations on tiles

Basavatarakam Indo-American Hospital and Tapadia diagnostics centre annotate around 42,000 tiles for mitotic figures using CADD4MBC platform. This Dataset is increased to 56,258 after applying augmentation methods such as horizontal and vertical flipping. The proposed Deep learning (DL) models use both private and public datasets for training. The summary of the dataset used by the proposed deep learning models is shown in Table I.

TABLE I. DATASET SUMMARY

Dataset Name	Size of Dataset	Image Size	Number of Tiles after Augmentation	Labels	Annotation Type
ICPR 12	2994	512 x 512	4367	Strongly Labelled	CSV
ICPR 14	2400	512 x 512	-	Weakly Labelled	CSV
TUPAC	2650	512 x 512	-	Weakly Labelled	CSV
KMIT Dataset	42,000	512 x 512	56,258	Strongly Labelled	JSON

B. KMIT Dataset Preprocessing

1) Color normalization: Hematoxylin and Eosin staining provide a detailed view of the cells in the tissue. Color variations may occur in the image due to the H&E staining affects deep learning predictions and can result in an incorrect diagnosis.

Color normalization provides a uniform standardized staining effect on tissue images. Reinhard color normalization method [20] adjusts the color variations of the input image by comparing the mean color variations between the input image and the standard reference image. After color variation adjustment the input image will be converted back to RGB color space.

2) Mask generation: The input files and the respective annotated JSON files are gathered from the CADD4MBC platform and the mask is created by using the ellipse function in the OpenCV library as shown in Fig. 3.

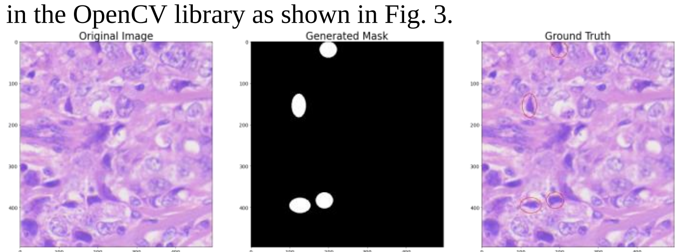


Fig. 3. Generated mask and the respective mitotic figure annotations

The generated mask is used to calculate the area of the ellipse and the semantic segmented points of the mitosis. If the area of the ellipse is very small then the respective annotations are removed as they won't produce good accuracy in deep learning training.

3) Annotation formats: The JSON format of the annotations is downloaded from CADD4MBC and are converted into a COCO dataset format and .txt for Faster R-CNN and YOLOv5 model respectively. The COCO annotation format [21] is used and understood by the most popular advanced neural network libraries such as Facebook Detectron-2. The COCO format can be used for object detection for both binary and multiclass detection. COCO dataset format is a combined file of all the annotations (JSON formats). Hence as shown in the Fig 4, all training image annotations are presented in a single file instead of individual JSON files for each annotation and similarly, testing requires only one COCO dataset JSON file.

The YOLOv5 model requires the annotations in a .txt file format. Each line in the .txt file represents one annotation and has [Class, X-center, Y-center, Width, Height] format, where the class number is used for mitosis and the remaining are the bounding box coordinates of the annotations which is shown as [0, 0.221, 0.579, 0.058, 0.058].

```
{
 "categories": [{ "supercategory": "none", "id": 1, "name": "mitosis" }],
 "images": [
{ "id": 1, "file_name": "T104x100y47.png", "width": 500, "height": 500},
{"id": 2, "file_name": "T99x126y107.jpg", "width": 500, "height": 500}
 ],
 "annotations": [
     {                                   {
     "segmentation": []                  "segmentation": [],
     "id": 1,                            "id": 2,
     "image_id": 1,                      "image_id": 2,
     "area": 375.0,                      "area": 3878.0,
     "category_id": 1,                   "category_id": 1,
     "bbox": [125, 258, 22, 23],         "bbox": [261,281,104,58],
     "iscrowd": 0,                       "iscrowd": 0,
     "ignore": 0                         "ignore": 0
     },                                  },
 ]}
```


Fig. 4. Sample mitotic figure COCO format annotation

C. Proposed DL Models Description

1) Faster R-CNN: Object detection networks primarily depend on algorithms which propose regions. The Proposed work uses Faster R-CNN to detect mitotic figures which is the most popular and advanced CNN-based object detection model. Faster R-CNN consists of three components such as backbone network, region proposal network and region of interest pooling.

a) Backbone network: In this proposed method, the Resnet-FPN [22] model is used as the backbone for automatic mitosis detection which is a combined model of Residual networks (Resnet) [23] and Feature Pyramid Network (FPN). As the input image is given to the first convolution block of the model then the features are extracted till the last convolution block. This layer-by-layer feature extraction sometimes leads to a gradient vanishing problem which makes it to lose the important features of the original image. The gradient vanishing problem is handled by Resnet by using its skip connections property and by FPN. The main objective of FPN is up sampling the low-resolution features with high-level feature maps (feature maps generated from the previous layers).

b) Region proposal network (RPN): Region Proposal network takes feature maps generated by FPN as input. RPN generates anchor boxes by using two parameters such as scales and aspect ratio. After generating anchor boxes there will be a possibility of many boxes which do not contain any object inside them. Region proposal network is mainly used for localizing and classification of the anchor box which are performed by the bounding box regression layer (anchor deltas convolution) and the bounding box classifier layer. These layers produce four regression parameters such as (x, y, w, h) where (x, y) is the centre, and w and h are the width and height of the anchor box.

c) Region of interest (ROI) pooling: The main issue in object detection is each proposal of RPN will be in a different shape. As per the Faster R-CNN architecture after ROI pooling there is a Fully Connected layer which generates a fixed-size feature map from the non-uniform size of input feature maps.

2) YOLOv5: YOLOv5 [24] is a single-stage object detector which requires only a single pass to the neural network and predicts all the bounding boxes. This feature makes YOLOv5 faster than any other traditional detection algorithm.

a) YOLOv5 consists of three important components: Cross Stage Partial Network (CSP) [25] is used as a backbone to extract informative features from an input image. CSPNet is used to achieve a good gradient combination and reduce the amount of computation which is achieved by partitioning the feature map of the base layer into two parts and then merging them through a proposed cross-stage hierarchy. PANnet [26] is used as a model neck which generates feature pyramids and scaled up the features.

The YOLOv5 model uses the YOLO layer as the head layer for the final detection of the objects. This layer generates 3 different sizes (18×18, 36×36, 72×72) of feature maps (bounding boxes) to achieve optimal prediction. This layer enables the model to handle small, medium, and big objects depending on the size of the objects.

IV. Results and Discussions

The proposed system initially uses three public datasets namely ICPR 12, ICPR 14 and TUPAC to detect mitosis. ICPR 12 contains semantic segmentation of mitosis whereas ICPR 14 and TUPAC contain the centroid of the mitosis. The proposed system uses the Faster R-CNN model to detect mitosis by using the ICPR 12 dataset for training and testing around 2994 tiles and produces an F1 score of 82.1. Subsequently, ICPR 14 and TUPAC datasets are tested by the Faster R-CNN model which is tested on around 900 tiles to detect mitosis. From tested results, the false positives are removed by comparing the centroid annotations of mitosis in ICPR 14 and TUPAC. The model produces Fl Score as 75% and 84% for ICPR 14 and TUPAC, respectively.

KMIT dataset is created by scanning 75 breast cancer WSI using a Morphle scanner at 40X magnification. The breast cancer tissue slides are collected from Basavatarakam Indo-American hospital, Hyderabad and Tapadia diagnostic centre Hyderabad. After digitizing WSI, each WSI is divided into tiles and on each tile the preprocessing methods such as normalization and resizing are applied. The tiles are annotated on the CADD4MBC platform by the pathologist of both the hospitals. Then masks are generated for the annotated tiles as well as JSON format of the annotations are downloaded from the platform. The JSON annotations are converted into COCO annotation format which is applicable to Faster R-CNN. The model is trained by 45,006 tiles and tested by 11,251 tiles. The model got an F1 score of 77% and the time taken for training is 10 hours 46 minutes

To improve accuracy the proposed system uses the YOLOv5 model which is the fastest object detection model on the KMIT dataset. YOLOv5 model accepts .txt annotations format. After converting JSON format into .txt format, YOLOv5 is trained by 45,006 tiles and tested by 11,250 files. The model produces an F1 score of 84%, and the time duration for training is around 9 hours 12 minutes.

To improve training and testing time YOLOv5 distributed model is applied to the KMIT dataset. This model used 45,006 files for training and 11,251 for testing and produced an F1 score of 84% within 5 hours and 28 minutes. Table II shows the performance of the proposed models, size of training and testing dataset, Evaluation metrics, and the training time of the proposed deep learning models.

TABLE II. PERFORMANCE OF THE PROPOSED MODEL

Dataset Name	Training	Testing	Model	F1-Score	Recall	Precision	Training Duration
ICPR12	3304	1063	Faster R-CNN	85.48	88.91	82.15	49 mins
ICPR14	-	547	Faster R-CNN	81.46	87.89	76.52	-
TUPAC	-	321	Faster R-CNN	82.15	81.56	81.69	-
KMIT Dataset	45006	11251	Faster R-CNN	75.86	73.83	76.86	10 Hrs 46 mins
KMIT Dataset	45006	11251	YOLOv5 (CPU)	84.23	81.62	86.76	9 Hrs 12 Min
KMIT Dataset	45006	11251	YOLOv5 (Distributed GPU)	84.58	82.31	86.42	5 Hrs 28 mins

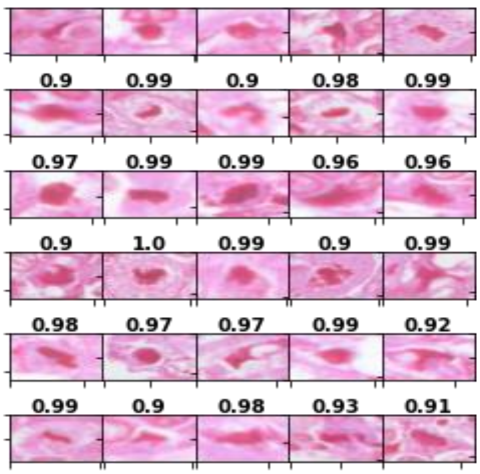


Fig. 5. Single mitosis detection by faster R-CNN

Single Mitosis Detection using Faster R-CNN model is demonstrated in Fig. 5.

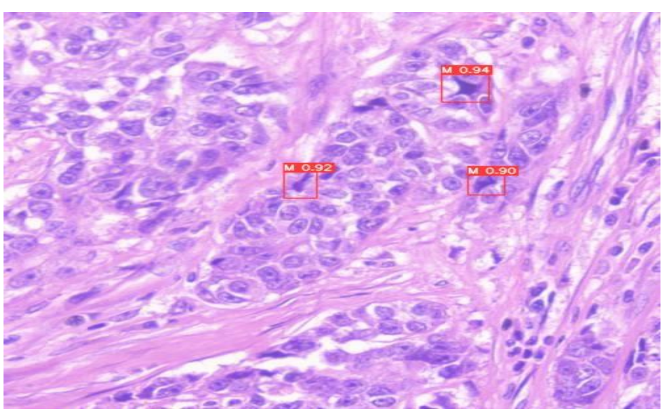


Fig. 6. Mitosis cell detection on tiles by Faster R-CNN

Mitosis cell detection on WSI tiles using Faster R-CNN model is demonstrated in Fig. 6.

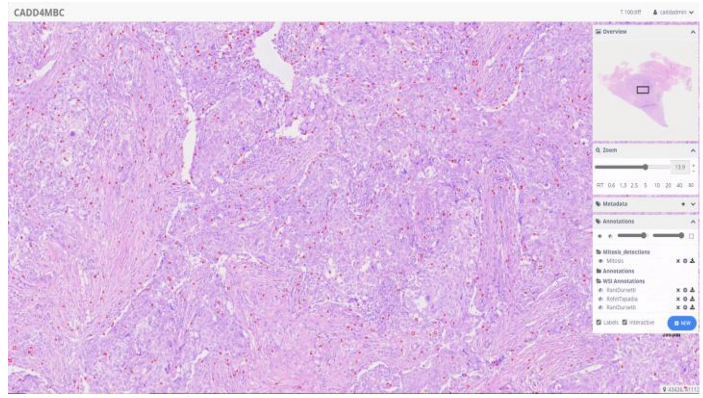


Fig. 7. Mitosis detection by YOLOv5 on WSI image

A number of mitoses (as red dots) detection on WSI image through YOLOv5 model is demonstrated in Fig. 7. The Results shows that the proposed deep learning models Faster RCNN and YOLOv5 produces good accuracy than the other models which are mentioned under the literature survey section. The other models use only the public dataset having 8000 tiles with strong/weak annotation and produce an F1 score between 0.3 and 0.7, whereas the proposed system produces an F1 score of 0.84 by trained and tested around 56,000 mitotic annotated tiles.

V. CONCLUSION

Nottingham grading system is a globally accepted system for breast cancer classification and grading the tumor. Detection of mitosis is the major component of NGS. Conventional microscopic mitosis detection is time-consuming, semi-quantitative and subject to inter observability. Digital pathology digitizes the tumor slides as WSI and applying deep learning models on WSI reduces the workload of the pathologist and assists them for quick accurate report generation. The deep learning models Faster R-CNN and YOLO v5 are learned from both the public datasets and KMIT dataset which has 56,258 tiles with annotated mitosis figures. The proposed web-based deep learning models detect mitosis with an F1 score of 0.84. The accuracy can be improved by increasing the size of the Breast Cancer WSI Mitosis dataset and Mitosis annotations to the models.

ACKNOWLEDGMENT

Prof. Neil Gogte, Director, Keshav Memorial Institute of Technology for the Project Guidance, Finance and Material support.

REFERENCES

[1] Bloom HJ, Richardson WW (1957) Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 11(3):359–377. https://doi.org/10.1038/bjc.1957.43

[2] Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histo-pathology 19(5):403–410. https://doi.org/10.1111/j.1365-2559.1991.tb00229.x

[3] Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991 Nov;19(5):403-10. doi: 10.1111/j.1365-2559.1991.tb00229.x. PMID: 1757079.

[4] Saco A, Ramirez J, Rakislova N, Mira A, Ordi J. Validation of whole-slide imaging for histopathological diagnosis: current state. Pathobiology. (2016) 83:89–98. doi: 10.1159/000442823.

[5] Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J. (2018) 16:34–42. doi:10.1016/j.csbj.2018.01.001

[6] Sohail A, Khan A, Wahab N, Zameer A, Khan S. A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images. Sci Rep. 2021 Mar 18;11(1):6215. Doi: 10.1038/s41598-021 85652-1. PMID: 33737632; PMCID: PMC7973714.

[7] Roux, L. et al. Mitosis detection in breast cancer histological images An ICPR 2012 contest. J. Pathol. Inform. 4, 8 (2013).

[8] Mitos &Amp; Atypia 14 Contest Home Page—MITOS-ATYPIA-14 Grand Challenge. https://mitos-atypia-14.grand-challenge. org/home/. (Accessed: 30th June 2020)

[9] Veta, M. et al. Predicting breast tumor proliferation from whole-slide images: Te TUPAC16 challenge. Med. Image Anal. 54, 111–121 (2019). http://tupac.tue-image.nl/. (Accessed:30th June 2020)

[10] Chao Li, Xinggang Wang, Wenyu Liu, Longin Jan Latecki, DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks, Medical Image Analysis,Volume 45, 2018,Pages 121-133, ISSN 1361-8415, https://doi.org/10.1016/j.media.2017.12.002.

[11] Sebai M, Wang X, Wang T. MaskMitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images. Med Biol Eng Comput. 2020 Jul;58(7):1603-1623. Doi: 10.1007/s11517-020-02175-z. Epub 2020 May 22. PMID: 32445109.

[12] Mask RCNN K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.

[13] D. Cai, X. Sun, N. Zhou, X. Han and J. Yao, "Efficient Mitosis Detection in Breast Cancer Histology Images by RCNN," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 919-922, doi: 10.1109/ISBI.2019.8759461.

[14] Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1–9. https://doi.org/10.1109/TPAMI.2016.2577031 (2015).

[15] Cireșan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Comput Assist Interv. 2013;16(Pt 2):411-8. doi: 10.1007/978-3-642-40763-5_51. PMID: 24579167.

[16] Jiménez G, Racoceanu D. Deep Learning for Semantic Segmentation vs. Classification in Computational Pathology: Application to Mitosis Analysis in Breast Cancer Grading. Front Bioeng Biotechnol. 2019 Jun 21;7: 145. doi: 10.3389/fbioe.2019.00145. PMID: 31281813; PMCID: PMC6597878.

[17] Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-Net: convolutional networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer Assisted Intervention (Munich), 234–241.

[18] Razavi S, Khameneh FD, Nouri H, Androutsos D, Done SJ, Khademi A. MiNuGAN: Dual Segmentation of Mitoses and Nuclei Using Conditional GANs on Multi-center Breast H&E Images. J Pathol Inform. 2022 Jan 20;13:100002. doi: 10.1016/j.jpi.2022.100002. PMID: 35242442; PMCID: PMC8860738.

[19] High-resolution image synthesis and semantic manipulation with conditional gans T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro (Eds.), Proceedings of the IEEE conference on computer vision and pattern recognition (2018), pp. 8798-8807

[20] Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P. (2001). Color transfer between images. IEEE Comput. Graph. Appl. 21, 34–41. doi: 10.1109/38.946629

[21] Lin, T.-Y. et al. Microsoft coco: Common objects in context. In European conference on computer vision 740–755 (Springer, 2014).

[22] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection.".In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125. 2017.

[23] K. He, X. Zhang, S. Ren, J. Sun Deep residual learning for image recognition Proceedings of the IEEE conference on computer vision and pattern recognition (2016), pp. 770-778

[24] Drioua, Wafaa Rajaa, Nacéra Benamrane, and Lakhdar Sais. "Breast Cancer Detection from Histopathology Images Based on YOLOv5." In 2022 7th International Conference on Frontiers of Signal Processing (ICFSP), pp. 30-34. IEEE, 2022.

[25] Lehmberg, A.A. and Olsen, M.N., 2006, September. An Introduction to CSP. NET. In CPA (pp. 13-30).

[26] Liu, Shu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. "Path aggregation network for instance segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8759-8768. 2018.

A Novel Compound Feature based Driver Identification

Md. Abbas Ali Khan[1], Mohammad Hanif Ali[2], AKM Fazlul Haque[3], Md. Iktidar Islam[4], Mohammad Monirul Islam[5]
Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh[1, 3, 4, 5]
Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh[1, 2]

***Abstract*—In today's world, it is time to identify the driver through technology. At present, it is possible to find out the driving style of the drivers from every car through controller area network (CAN-BUS) sensor data which was not possible through the conventional car. Many researchers did their work and their main purpose was to find out the driver driving style from end-to-end analysis of CAN-BUS sensor data. So, it is potential to identify each driver individually based on the driver's driving style. We propose a novel compound feature-based driver identification to reduce the number of input attributes based on some mathematical operation. Now, the role of machine learning in the field of any type of data analysis is incomparable and significant. The state-of-the-art algorithms have been applied in different fields. Occasionally these are tested in a similar domain. As a result, we have used some prominent algorithms of machine learning, which show different results in the field of aspiration of the model. The other goal of this study is to compare the conspicuous classification algorithms in the index of performance metrics in driver behavior identification. Hence, we compare the performance of SVM, Naïve Bayes, Logistic Regression, k-NN, Random Forest, Decision tree, Gradient boosting.**

Keywords*—*Compound feature; driver behavior identification; engine speed; fuel consumption; vehicle

I. Introduction

Every driver has their driving style; therefore, the driver can be identified according to exploration through the driving pattern analysis. It is to be considered as a fingerprint of the driver's manner like acceleration, speed, and braking habits that vary from driver to driver. Driver fingerprinting could lead to important privacy compromises [1]. Today we cannot consider just a vehicle as a modern car, as it is a fully decorated smart device with various functions like multimedia, security system, and different sensors [2]. The sensors were very simple because the driver was informed regarding the features of the engine and the amount of fuel through the magnetoelectric and light display devices [2]. Using state-of-the-art technology in real-time all the microcomputers are communicated with each other through CAN-BUS (Controller Area Network) [3]. To make a car more efficient a good number of technologies are used in the modern engine. To improve engine performance direct injection technology was introduced in the modern car [4]. According to a survey, the researcher predicted that the number of sales of connected cars will reach 76.3 million in the next 2023 [5]. Through state-of-the-art technology, modern engines use less fuel and besides get more power [6]. Most of the cars have partnered with other components which are highly technology-based, such as traffic lights, garage doors, and services [7]. Not only the driving style there is a discount policy on insurance services but also real-time monitoring, maintenance, pathfinding, driving style development, and also consumption of fuel [8]. Vulnerabilities of connected cars will increase the auto-theft which is one of the threats [9]. Top-of-the-range vehicles are targeted by thieves who simply drive off after bypassing security devices by hacking onboard computers [10]. Penny [11] introduced a man-in-the-middle attack or relay attack, to do this radio signals are passed between two devices. Pekaric I et al. (2021) [12] described other attacks such as GPS spoofing and message injection attacks. BMW Connected Drive [13] seamlessly integrates mobile devices, smart home technology, and vehicle's intelligent interfaces into a complete driver's environment. Even though in 2021 they introduced a remote door unlock system through a signal to the driver's door to unlock [13]. CAN-BUS is likely a nervous system used to allow configuration, data logging, and communication among electronic control units (ECU) e.g., ECU is like a part of the body and interconnected through CAN, by which information sensed by one part can be shared with another [14]. Up to 70 ECUs have a modern car e.g., the engine control unit, airbags, audio system, acceleration, fuel unit, etc. [15]. Normally, multi-sensor data is made up of in vehicle's CAN data. The in-vehicle CAN data such as steering wheel, vehicle speed, engine speed, amount of fuel, etc. Several researchers previously proposed a driver identification method based on in-vehicle CAN-BUS data. But direct connectivity is difficult to get data, so onboard diagnostics (ODB-II) is used. (OBD-II, ISO 15765) is a self-diagnostic and reporting capability that e.g., mechanics use to identify car issues, OBD-II specifies diagnostic trouble codes (DTCs) and real-time data (e.g., speed, revolution per minute RPM), which can be recorded via OBD-II loggers from CAN-BUS. Many authors described the problem of CAN-BUS data for identifying the driver [9], [16], [17].

Since this is a big dataset and there are 51 features with 10 labels. Moreover, for analyzing the whole dataset we need more time. To reduce the time complexity we have explained compound feature selection process. In this paper, our objective is to identify driver behavior through telemetric data using machine learning algorithms. We analyze the data in terms of training, testing, and validation to get model accuracy that helps us with driver identification.

There are some significant commercial and personal values of driver identification for insurance, rental companies, personal and tripping regulations as well. The goal of this work is to monitor and maintain unauthorized access.

The paper is organized as follows. We have presented an introduction and literature review Sections I and II as well. The architecture of the proposed system has described in Section III. The discussion of methodology and compound features have existed according to Sections IV and V. Before the conclusion part of Section VII, the results and discussion are mentioned in Section VI.

II. Literature Review

The author [18] uses telemetric data to investigate driver identification and identification accuracy decreases by 15% compared to the method. They use the role of non-public parameters in identifying the driver. Previous work had been done by using car driving simulated [18] data. Investigated the driver's behavior when he follows another car. The features mentioned below are used to observe such as accelerator pedal, car speed, brake pedal, and distance to the next car. Gaussian Mixture Model (GMM) is used to achieve 81% accuracy with 12 drivers and 73% of 30 drivers [18]. Another author [19] analyzes overtaking style for each driver, uses accelerator, and steering data and the accuracy is 85% for about 20 drivers through the Hidden Markov Model (HMM). Some other authors used smartphones to capture driver data. Sensors in the smartphone are- GPS, accelerometer, magnetometer, and gyroscope [20-22]. This data is used for driver profiling and other tasks. An author used an inertial sensor and algorithm was SVM, and k-means methods and got 60% accuracy between two drivers.

In another research, the authors [8, 16], [23-24] acquire data from in-vehicle CAN-BUS via OBD-II. The author [23] uses in-vehicle CAN-BUS sensor data and the accuracy was 99% among 15 drivers. He uses SVM, Random Forest, Naive Bayes, and k-NN methods. Another researcher [9] got 99% accuracy from 51 features of 10 drivers and used Decision Tree, k-NN, Random Forest, and Multilayer Perceptron (MPL). Choi et al. [24] find out the driving detection and driver recognition using both GMM and HMM methods, which are used for analyzing vehicle CAN-BUS data. Kedar-Dongakar et al. [25] recognized the driver classification based on the energy optimization of a vehicle. Based on driving style three types of drivers are classified as aggressive, moderate, and conservative. The author considers the following features for his research work such as vehicle speed, acceleration, torque, acceleration pedal, steering wheel angle, and brake pedal pressure.

Several researches have been going on neural network and deep learning algorithms for a few years back and draw a good impact on driver behavior identification works. Xun et al. [26] introduced Convolutional Neural Network (CNN) and got 99% accuracy for 10 drivers. For driver identification, another paper uses noise-free data and as an algorithm-LSTM-Recurrent Neural Network is used, where it has high accuracy with salient advantages [27].

After reviewing, the difference among this work and the done before.

- For collecting dataset uses specific model's vehicle.
- Specific dataset with different features
- Uses CAN-BUS and public OBD-II.

III. Architecture of the Proposed System

The purpose of the proposed system is to identify the actual driver. After capturing the raw data through OBD-II sent to the cloud is shown in Fig. 1. Once completed the engineering knowledge-based mathematical operation like adding, multiplication, and logging of two features then send for prediction based on model accuracy result.

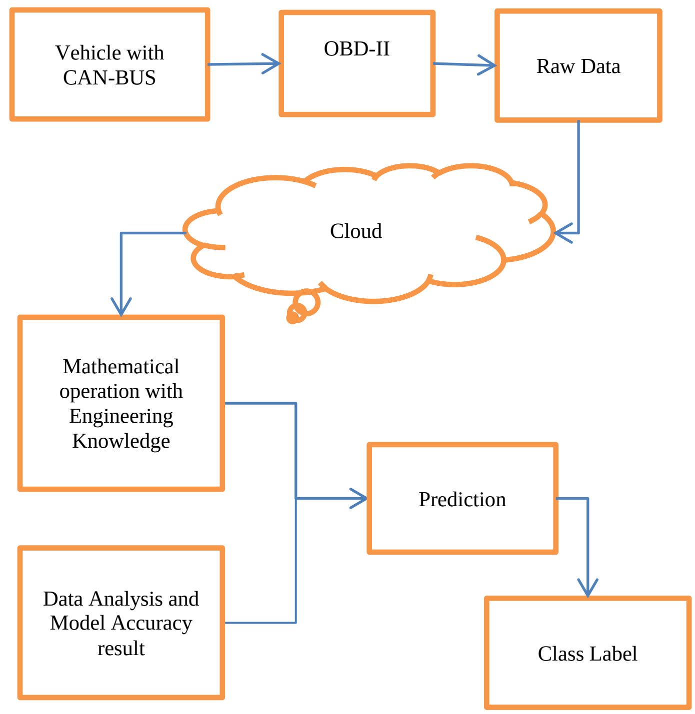


Fig. 1. Architecture of the proposed system

IV. Methodology

A. Approaches of Methodology

The steps of the proposed system are shown in Fig. 2. Before the classifier's dataset preparation, data preprocessing, feature selection, normalization, and other activities have occurred. Several classifiers are used for comparative analysis with state-of-the-art classifiers to find the best model.

In this connection, we need data on the trips for driver identification. Our model is considered an Ocslab driving dataset [28]. This data is used for driver indemnification and personalization based on pattern analysis. KIA motors corporation vehicles in South Korea were performed to collect the data and the experiment has been done since July 28, 2015 [28]. A total of 10 drivers labeled "A" to "J" are included in the trips and cover 23 km length, completing two round trips from 8.00 PM to 11.00 PM [28]. Three types (such as city roads, freeways, and parking lots) of the road are there with their characteristics. There are a total of 94,401 records with 51 dimensions (51 features) [28]. Not all data is possible to get because there are some limitations of OBD-II identifiers and sensors such as it cannot provide body control status or airbag status even wheel angle rotation status. OBD-II has a limited set of identifiers [29] provided by the manufacturer.

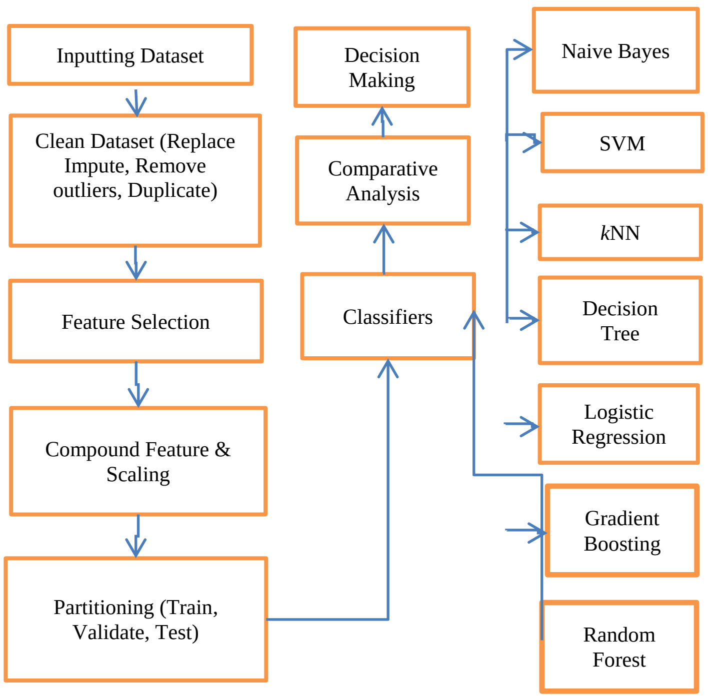


Fig. 2. Step by step process of the work

Among 51 features we have considered two prominent features with around 13,000 samples named Fuel Consumption (FC) and Engine Speed (ES) to make a compound feature. Transform the collected data to our classification model for analysis we follow- feature selection, making a compound feature, scaling, and data processing through state-of-the-art techniques [34]. The sample data are viewed in Eq. (1).

$$X = \begin{bmatrix} X_1^1 & X_2^1 & X_d^1 \\ X_1^2 & X_2^2 & X_d^2 \\ \vdots & \vdots & \vdots \\ X_1^N & X_2^N & X_d^N \end{bmatrix} \quad (1)$$

Where d columns correspond to d variable and N rows correspond to N instances. Fig. 3 shows the histogram of the compound dataset.

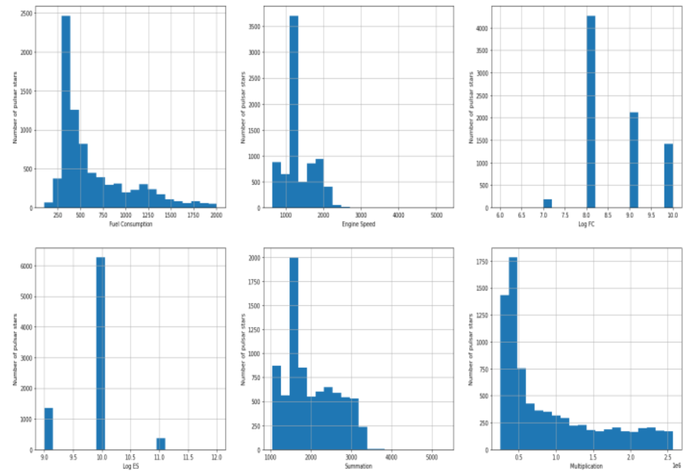


Fig. 3. Histogram of the dataset

To improve the dataset several statistical and logical tools are used to overcome the problem. Significantly different data points are removed by outliers [30].

B. Data Visualization and Relationship in a Dataset

a) Heat-map: A heat map is a data visualization technique that shows the magnitude of a phenomenon as color in two dimensions [31-32]. The color variation may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space [38]. We consider a correlogram a clustered heat map that has the same trait for each axis to display how the traits in the set of traits interact with each other. The correlogram is a triangle instead of a square because the combination of A-B is the same as the combination of B-A and so does not need to be expressed twice. Fig. 4 is shown the feature correlation of the compound dataset. The value of correlation can take any value from -1 to 1, based on that-

- Features such as Fuel Consumption, Log FC, Sum, and Mul are having strong positive correlations.
- Fuel consumption and engine speed, engine speed, and Log FC, Log FC, and Log Es have a weak positive correlation.
- Class has a strong negative and weak positive correlation with every feature.

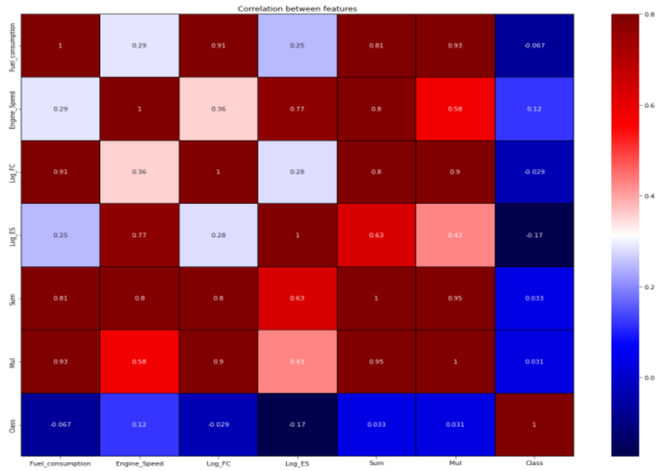


Fig. 4. Correlation among features

C. Pair Plot

A pair plot is a pairwise relationship in a dataset [36]. The pair plot function creates a grid of Axes such that each variable in data will be shared on the y-axis across a single row and the x-axis across a single column [39]. By plotting pair plots, it visualized that most of the classes of each feature overlap with each other. Fig. 5 shows the pair plot of the compound dataset. After analyzing the scatter plot a statistically significant result is shown that non-linear classification models such as k-NN, Decision tree, Random Forest classifier, Gradient boosting classifier, etc. perform better than linear classification models such as logistic regression.

D. Standard Scaler

Standard Scaler comes into play when the characteristics of the input dataset differ greatly between their ranges, or simply when they are measured in different units of measure [37]. Standard Scaler - standardizes a feature by subtracting the mean and then scaling to unit variance. Unit variance means dividing all the values by the standard deviation.

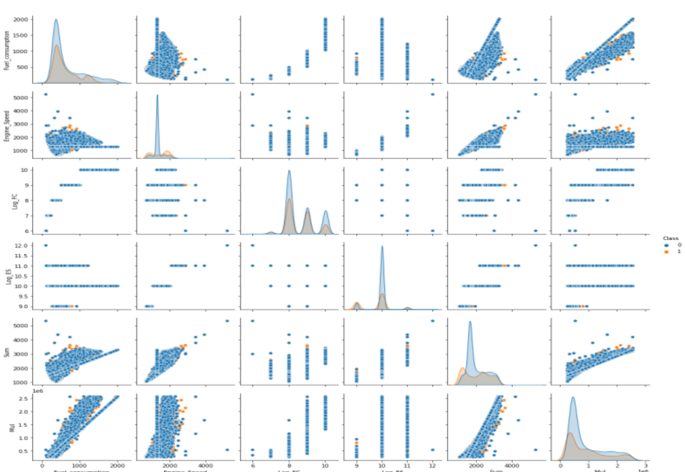


Fig. 5. Pair plot of the dataset

As we see, different scales of data exist in the dataset. Normalization is essential for some machine learning algorithms like k-NN (k-Nearest Neighbor) and SVM [29]. The normalization formulas for integrating data scales are shown in Eq. 2.

$$Z = {}^{(x-\mu)}/_{\sigma} \qquad (2)$$

Here, μ = mean and σ= standard deviation. To evaluate the generalization of the model we have considered K fold cross-validation for low bias and a modest variance [33]. We have used five folds for training and testing the dataset and obtained the mean performance.

In the case of classification application, we have considered supervised machine learning classifiers for performing the metrics named *k*-NN, SVM, Logistic Regression, Decision tree, Random Forest, and Gradient boosting. The *k*-NN (*k*-Nearest Neighbor) is an instance-based traditional machine learning algorithm. Both classification and regression cases *k*-NN can be used and select the number of neighbors through distance calculation of the query points. Eq. 3 is used to calculate the Euclidean distance between two points.

$$D = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (3)$$

SVM stands for Support Vector Machine, used for classification and regression problems. The goal is to find a hyperplane in an N-dimensional space and separately classify the query data point. There is a decision boundary called hyperplane that is used to differentiate the classes. It also creates a margin separator with the nearest observations and it performs better if maximizes the margin. The equation represents the loss function that indicates maximize the margin

$$C(x, y), Where\ Y = f(x) = \begin{cases} 0, if\ y * f(x) \geq 1 \\ 1 - y * f(x), else\ 1 \end{cases} \qquad (4)$$

Logistic Regression predicts whether something is true or false. Instead of fitting a line to the data, it fits an "S" shaped "logistic function" and the curve goes from 0 to 1. The following equation is used to calculate the function, also called the sigmoid function.

$$S(x) = \frac{1}{1+e^{-x}} \qquad (5)$$

Naïve Bayes is a classifier based on Bayes' theorem. It assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. The naïve Bayes model is easy to build a large dataset and outperforms sophisticated classification methods. The way Naïve Bayes is used to calculating the posterior probability shows in the Eq. 6.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \qquad (6)$$

To measure the performance, we have used four indicators named accuracy, precision, F-Measure, and Recall. The computation and the evaluation performance of the classifiers have occurred through the confusion metric. Once the model is generated then the classifier is tested by using a test dataset to check the model's accuracy. Precision indicates how close or dispersed the measurement is to each other. It measures the number of correct positive predictors made. A recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made.

The number of *FP*'s, *FN*'s, *TP*'s, and *TN*'s cannot be calculated directly from this matrix. The values of *FP*'s, *FN*'s, *TP*'s, and *TN*'s for class $\boldsymbol{i}$ ($1 \leq i \leq n$) are determined as per [35].

The final confusion matrix, which has dimension 2 × 2, comprises the average values of the n confusion matrices for all classes. For a binary, i.e. two-class problem, a confusion matrix gives the number of false positives (*FP*s), false negatives (*FN*s), true positives (*TP*s), and true negatives (*TN*s). From this confusion matrix, accuracy, precision, recall, and *F*-1 score are calculated in the following way:

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (8)$$

$$F_1 - score = \frac{2 \times precision \times recall}{precision + recall} \times 100\% \quad (9)$$

$$Accuracy = \frac{TP+TN}{(TP+FN)+(FN+TN)} \times 100\% \quad (10)$$

Fig. 6 shows the time series of a dataset with a window based on RPM. We have considered time t_0 with the window size of w_1 and t_{0+1} time for the next window and so on. Fig. 9 shows the accuracy of the several models based on time and window size.

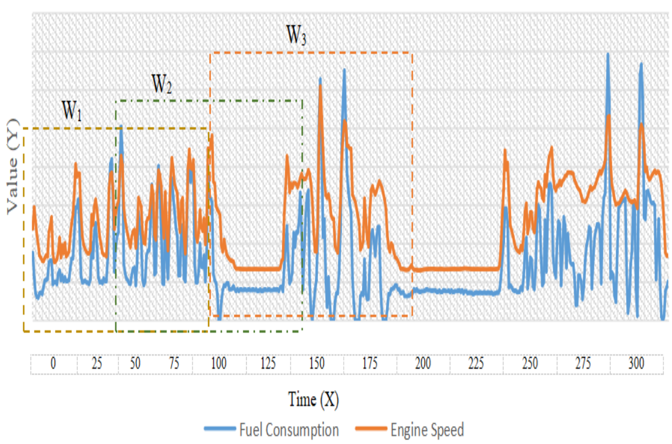


Fig. 6. Time series of dataset with window

V. COMPOUND FEATURE

To make the feature compound we consider a general mathematical operation. We have selected two important features among the 51 features from the Ocslab dataset. Once we have applied mathematical operations to the selected feature to achieve the compound feature. The compound feature processing method is stated in Fig. 7.

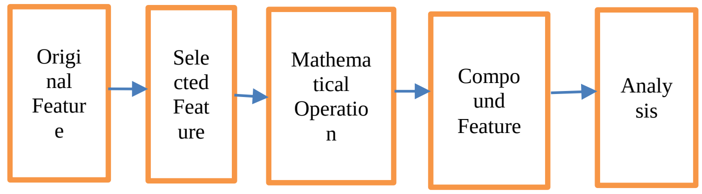


Fig. 7. Preparation of compound feature

From the mathematical operation, we have considered addition, multiplication, and binary logarithm to make the feature compound. Eq. (11), (12), and (13) describe the mathematical operation. Here, k > 0, $k \in \mathbb{N}$ and a_k, b_k are two features with the index of *k*.

Since the value of the features are numerical so we add two features to make a single (Compound) feature is shown in Eq. (11). Eq. (12) is shown for multiplication between two features to build a compound feature. In Eq. (13), another mathematical operation binary log is applied to a single value of two features separately and after that we add them for making compound feature.

$$S = \sum_{k=1}^{n}(a_k + b_k) \quad (11)$$

$$M = \prod_{k=1}^{n}(a_k b_k) \quad (12)$$

$$Y_1 = \log_2 a_k, Y_2 = \log_2 b_k \quad (13)$$

We have considered raw features as the first level and the others as are inner level. Fig. 8 shows the level indexing process.

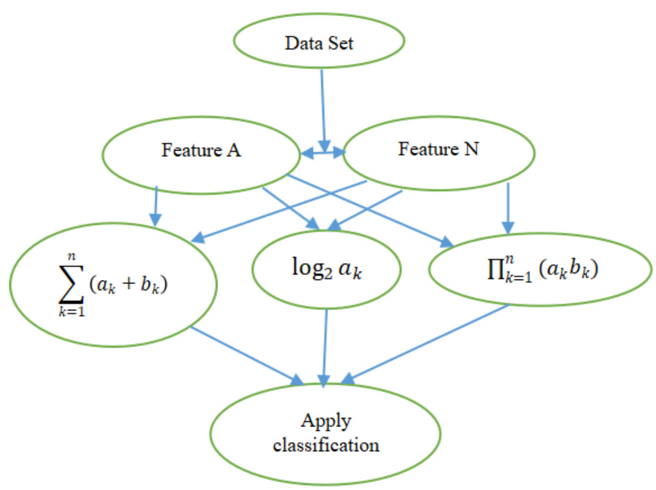


Fig. 8. Inner level processing system

VI. RESULTS AND DISCUSSION

To evaluate the generalization of the model we have considered K fold cross-validation for low bias and a modest variance [33]. We have used five folds for training and testing the dataset and obtained the mean performance.

For the classification of the driver, we have introduced the prominent supervised algorithm named Naive Bayes (NB), Logistic Regression, *k*-NN, Decision Tree (DT), Gradient Boosting (GB), Random Forest (RF), and SVM. Table I shows the result of all classifiers through the confusion metric. Among them, GB shows 83.00 % (highest) and NB performs 65.00% (lowest) accuracy respectively. Mentionable, we have made four compound features from two features among 51 features and considered two drivers among 10 drivers respectively from the Ocslab dataset. The results of non-linear classifiers where GB and RF are given 85.00% and 81.00% accuracy accordingly. Results of confusion metrics are shown in Table II and a comparative analysis of original features vs compound features is visible in Table III. In the Compound feature, GB performs a height accuracy of 85%. Moreover, the ROC curve are shown in true positive in Fig. 10 for all classifiers.

Again we have considered the time series of the drivers' and found out the driving pattern style through the windowing system using the compound dataset. Fig. 9 is representing the accuracy of only two drivers (A, D) and it also represents the W_1, W_2, and W_3 accuracy of 75.00%, 78.00%, and 85.00% respectively. In this research, we have figured out linear and non-linear algorithms to calculate the performance. Moreover, according to pair plot analysis, we have decided that non-linear is better than linear analysis. As a result, the model is statistically significant because of the better performance of the non-linear algorithm.

TABLE I. MODEL CLASSIFICATION REPORT

Classifiers	***Performance parameters***				
		Accuracy	Precision	Recall	F1-Score
Logistic Regression		72.00	76.00	81.00	78.00
	Macro Average	-	71.00	70.00	70.00
	Weighted Average	-	72.00	72.00	72.00
Gradient Boosting		85.00	94.00	74.00	83.00
	Macro Average	-	81.00	83.00	81.00
	Weighted Average	-	84.00	81.00	81.00
Decision Tree		78.00	97.00	66.00	78.00
	Macro Average	-	80.00	81.00	78.00
	Weighted Average	-	84.00	78.00	78.00
Random Forest		81.00	86.00	80.00	83.00
	Macro Average	-	79.00	80.00	79.00
	Weighted Average	-	81.00	80.00	80.00
SVM		76.00	85.87	74.00	79.00
	Macro Average	-	75.00	76.00	75.00
	Weighted Average	-	77.00	76.00	76.00
k-NN		78.00	85.99	78.99	81.99
	Macro Average	-	76.00	77.00	77.00
	Weighted Average	-	78.00	77.00	78.00
Naïve Bayes		65.00	93.00	76.00	84.00
	Macro Average		57.00	64.00	57.00
	Weighted Average		85.00	74.00	78.00

TABLE II. CONFUSION METRICS

Classifiers	***Array metrics***	
Logistic Regression	1189	287
	379	539
Gradient Boosting	1097	379
	72	846
Decision Tree	1076	4000
	33	885
Random Forest	1183	293
	187	731
SVM	1093	383
	200	718
k-NN	1147	329
	210	708

TABLE III. COMPARATIVE ANALYSIS OF ORIGINAL FEATURE AND COMPOUND FEATURE BASED ON MODEL ACCURACY

Classifiers	***Original feature***	***Compound Feature***
Logistic Regression	55	72
Gradient Boosting	80	85
Decision Tree	76	78
Random Forest	81	80
SVM	74	76
k-NN	76	78
Naïve Bayes	57	65

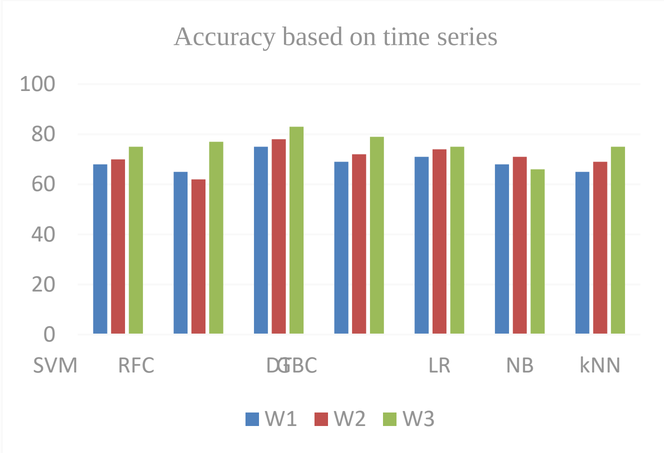


Fig. 9. Accuracy based on time series

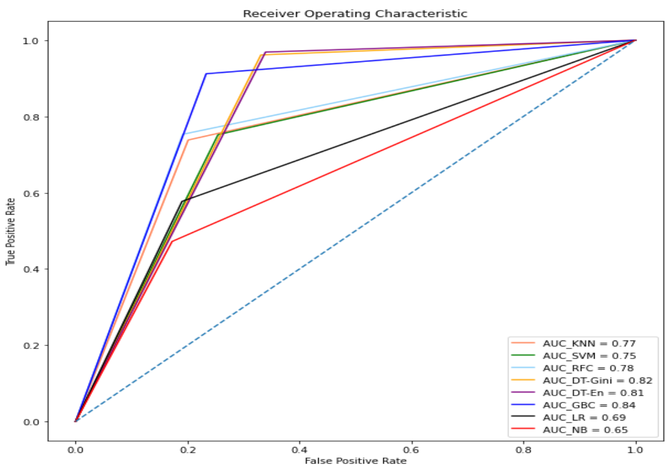


Fig. 10. ROC curve of the several classifiers

VII. Conclusion

In terms of driver identification, the role of machine learning as a cutting-edge technology is immense. Through this work, we found the best model that shows a momentous way to build our proposed anti-theft driving expert system. This research work has presented an in-depth comparison of the performance of seven (7) projecting classifiers in the context of driving behavior identification. As a result, the best and the worst classifiers we have found are GB and NB respectively. This type of result is very effective in developing an online-based anti-theft system. We have executed our experiment with only a few numbers of drivers among ten drivers.

Through online the owner of the vehicle will notify, so there is possibility man-in-the middle attack to another LAN. We will introduce network security to ensure authorized access and overcome the vulnerabilities of the proposed system in the next work.

Acknowledgment

I am grateful to KIA Motor Corporation for the Dataset provided to me.

References

[1] Enev M, Takakuwa A, Koscher K, Kohno T. Automobile Driver Fingerprinting. Proc. Priv. Enhancing Technol.. 2016 Jan; 2016(1):34-50.

[2] K. Uvarov and A. Ponomarev, "Driver Identification with OBD-II Public Data," 2021 28th Conference of Open Innovations Association (FRUCT), 2021, pp. 495-501, doi: 10.23919/FRUCT50888.2021.9347648.

[3] Kaplan S, Guvensan MA, Yavuz AG, Karalurt Y. Driver behavior analysis for safe driving: A survey. IEEE Transactions on Intelligent Transportation Systems. 2015 Aug 26; 16 (6):3017-32.

[4] Oak Ridge National Laboratory for the U.S. Department of Energy and the U.S. Environmental Protection Agency (U.S Department of Energy)

[5] https://www.businesswire.com/news/home/20190523005089/en/Worldwide-Connected-Vehicle-Shipments-Foreca St-to-Reach-76-Million-Units-by-2023-According-to-IDC (Accessed date 9/20/2022)

[6] General Motors. "Chevrolet Malibu Media Archives." (June 1, 2011) http://archives.media.gm.com/division/2003_prodinfo/03_chevrolet/03_malibu/index.html

[7] Jamie Page Deaton "5 Ways Modern Car Engines Differ from Older Car Engines" 16 May 2011. HowStuffWorks.com.<https://auto.howstuffworks.com/5-ways-modern-car-engines-differ-from-older-car-engines.htm> (accessed date 20 August 2022)

[8] Carfora MF, Martinelli F, Mercaldo F, Nardone V, Orlando A, Santone A, Vaglini G. A "pay-how-you-drive" car insurance approach through cluster analysis. Soft Computing. 2019 May; 23(9):2863-2875.

[9] Kwak BI, Woo J, Kim HK. Know your master: Driver profiling-based anti-theft method. In2016 14th Annual Conference on Privacy, Security and Trust (PST) 2016 Dec 12 (pp. 211-218). IEEE.

[10] "dailymail, "http://www.dailymail.co.uk/news/article-2938793/Carhackers-driving-motors-Increasing-numbers-stolen-thieves-simplybypass-security-devices.html, 2015, (accessed: 2022-20-01).

[11] Penny Hoelscher "Information security-what is relay attack" january 31, 2019

[12] Pekaric I, Sauerwein C, Haselwanter S, Felderer M. A taxonomy of attack mechanisms in the automotive domain. Computer Standards & Interfaces. 2021 Apr 23:103539.

[13] BMW center, www.bmwusa.com/ConnectedDrive (accessed date 21-01-2022)

[14] Grant Maloy Smith ,https://dewesoft.com/author/grant-maloy-smith (accessed date 21-01-2022)

[15] Comparison of Event-Triggered and Time-Triggered Concepts with Regard to Distributed Control Systems A. Albert, Robert Bosch GmbH Embedded World, 2004, Nürnberg

[16] Ullah S, Kim DH. Lightweight driver behavior identification model with sparse learning on In-Vehicle CAN-BUS sensor data. Sensors. 2020 Jan;20 (18):5030.

[17] Girma A, Yan X, Homaifar A. Driver identification based on vehicle telematics data using lstm-recurrent neural network. In2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI) 2019 Nov 4 (pp. 894-902). IEEE

[18] Wakita T, Ozawa K, Miyajima C, Igarashi K, Itou K, Takeda K, Itakura F. Driver identification using driving behavior signals. IEICE TRANSACTIONS on Information and Systems. 2006 Mar 1;89 (3):1188-94.

[19] Jun X, Zhao X, Rong J. A study of individual characteristics of driving behavior based on hidden markov model. Sensors & Transducers. 2014 Mar 1;167 (3):194.

[20] Castignani G, Derrmann T, Frank R, Engel T. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent transportation systems magazine. 2015 Jan 19;7 (1):91-102.

[21] Kashevnik A, Lashkov I, Gurtov A. Methodology and mobile application for driver behavior analysis and accident prevention. IEEE transactions on intelligent transportation systems. 2019 Jun 4;21 (6):2427-36.

[22] Van Ly M, Martin S, Trivedi MM. Driver classification and driving style recognition using inertial sensors. In 2013 IEEE Intelligent Vehicles Symposium (IV) 2013 Jun 23 (pp. 1040-1045). IEEE.

[23] Enev M, Takakuwa A, Koscher K, Kohno T. Automobile Driver Fingerprinting. Proc. Priv. Enhancing Technol.. 2016 Jan;2016 (1):34-50.

[24] Choi S, Kim J, Kwak D, Angkititrakul P, Hansen JH. Analysis and classification of driver behavior using in-vehicle can-bus information. InBiennial workshop on DSP for in-vehicle and mobile systems 2007 Jun (pp. 17-19).

[25] Kedar-Dongarkar G, Das M. Driver classification for optimization of energy usage in a vehicle. Procedia Computer Science. 2012 Jan 1; 8:388-93.

[26] Xun Y, Liu J, Kato N, Fang Y, Zhang Y. Automobile driver fingerprinting: A new machine learning based authentication scheme. IEEE Transactions on Industrial Informatics. 2019 Oct 10;16(2):1417-26.

[27] Girma A, Yan X, Homaifar A. Driver identification based on vehicle telematics data using lstm-recurrent neural network. In2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI) 2019 Nov 4 (pp. 894-902). IEEE

[28] Driving Dataset. Available online: https://ocslab.hksecurity.net/Datasets/driving-dataset (accessed on 2-11 2022)

[29] OBD-II PIDs, Available online: https://en.wikipedia.org/wiki/OBD-II_PIDs (accessed on 22-10-2022)

[30] Kwak, Sang Kyu, and Jong Hae Kim. "Statistical data preparation: management of missing values and outliers." Korean journal of anesthesiology 70.4 (2017): 407-411.

[31] Caruso, Pier Francesco, et al. "The effect of COVID-19 epidemic on vital signs in hospitalized patients: a pre-post heat-map study from a large teaching hospital." Journal of clinical monitoring and computing 36.3 (2022): 829-837.

[32] Chen, Tong, Yong-Xin Liu, and Luqi Huang. "ImageGP: An easy-to-use data visualization web server for scientific researchers." iMeta 1.1 (2022): e5.

[33] Jason Brownlee, https://machinelearningmastery.com/k-fold-cross-validation (accessed on 11-02-2022)

[34] Mo, Yujian, et al. "Review the state-of-the-art technologies of semantic segmentation based on deep learning." Neurocomputing 493 (2022): 626-646..

[35] M. T. Habib, A. Majumder, A. Jakaria, M. Akter, M. S. Uddin, & S. Ahmed, (2020). Machine vision based papaya disease recognition. Journal of King Saud University - Computer and Information Sciences, 32(3), 300– 309

[36] Rafatirad, Setareh, et al. "What Is Applied Machine Learning?." Machine Learning for Computer Scientists and Data Analysts. Springer, Cham, 2022. 3-33.

[37] Raju, VN Ganapathi, et al. "Study the influence of normalization/transformation process on the accuracy of supervised classification." 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2020.

[38] Gu Z. Complex heatmap visualization. iMeta. 2022 Sep;1(3):e43.

[39] Pitroda H. A Proposal of an Interactive Web Application Tool QuickViz: To Automate Exploratory Data Analysis. In2022 IEEE 7th International conference for Convergence in Technology (I2CT) 2022 Apr 7 (pp. 1-8). IEEE.

BrainNet-7: A CNN Model for Diagnosing Brain Tumors from MRI Images based on an Ablation Study

Md Harun or Rashid, Salma Akter, Amatul Bushra Akhi
Dept. of Computer Science Engineering
Daffodil International University
Dhaka, Bangladesh

***Abstract*—Tumors in the brain are masses or clusters of abnormal cells that may spread to other tissues nearby and pose a danger to the patient. The main imaging technique used to determine the extent of brain tumors is magnetic resonance imaging, which ensures an accurate diagnosis. A sizable amount of data for model training and advances in model designs that provide better approximations in a supervised environment likely account for most of the growth in Deep Learning techniques for computer vision applications. Deep learning approaches have shown promising results for increasing the precision of brain tumor identification and classification precision using magnetic resonance imaging (MRI). This study's purpose is to describe a robust deep-learning model that categorizes brain tumors using MRI images into four classes based on a convolutional neural network (CNN). By removing artefacts, reducing noise, and enhancing the image, unwanted areas of brain tumors are deleted, quality is improved, and the tumor is highlighted. Several CNN architectures, including VGG16, VGG19, MobileNet, MobileNetV2, and InceptionV3, are investigated to compare or get the best model. After getting the best model, a hyper parameter ablation study was performed on that model. Proposed BrainNet-7 achieved the best results with 99.01% test accuracy and 99.21% test and validation accuracy.**

Keywords*—*MRI image; image pre-processing; transfer-learning; CNN; brainnet-7

I. Introduction

A brain tumor is one of the tenth most common causes of mortality in men and women regarding the brain or central nervous system, often referred to as the CNS [1]. It is estimated that 40% of all cancer types develop brain cancer as a result of metastasis rather than death from brain tumors [2]. In 2000, June 8 was designated World Brain Tumor Day to raise understanding and educate people regarding brain tumors [3]. If abnormal cells begin to grow unnecessarily in the brain or spinal cord, it is known as a brain tumor. The World Health Organization categorized brain tumors into four groups on the basis of molecular characteristics in 2016 -- I, II, III, and IV [3, 4]. Brain tumor patients have a very low life probability when the tumor is in more advanced phase [5]. Therefore, accurate and timely cancer diagnosis and grade estimation enhance illness prognosis and treatment options. A neurological examination, imaging, biopsies, and other methods are used to determine the grade and diagnosis of tumors [3, 6]. Doctors use magnetic resonance imaging (MRI) before and after treatment to determine the tumor's shape. As a result, surgical resections can be planned and monitored as the illness develops [7]. Early classification of brain tumor grade plays an important role in successful prognosis [8]. A good contrast enhancement and noninvasive MRI images make it the preferred imaging technique in glioma diagnosis [9]. Radiologists observe and diagnose tumors using the conventional method, it is laborious and time-consuming. Computer-aided medical diagnosis (CAMD) has made great strides with artificial intelligence (AI) and deep learning, which can assist doctors in interpreting medical images within seconds [10]. A dataset's quality and size significantly impact the performance of deep learning technology. Images with high-quality annotations are required for deep learning techniques. However, labelling large quantities of medical images is quite challenging since annotation is a time- and expertise-intensive process [11]. Two significant barriers to deep learning in medical imaging are insufficient imaging data and a lack of annotations from human experts [11]. The above challenges have been addressed and resolved through numerous efforts. A transfer learning strategy can be helpful when there are only a few domain samples for training. Typically, it refined on the architecture that has already pre-trained on a largest, labelled dataset. The transfer of learning knowledge to the target dataset makes network convergence speed faster while maintaining low computational complexity [12].

In this work, we propose a CNN model BrainNet-7, which is fine-tuned network and classifies brain tumor MRI images most correctly. Firstly, use, five pre-trained models, VGG16, VGG19, MobileNet, MobileNetV2, and InceptionV3, are employed in the dataset and then use proposed a CNN model. After in the CNN model a hyper parameter ablation was performed for getting the robust and fine-tuned model. This model is given the best accuracy among all the previous networks (see Fig. 1).

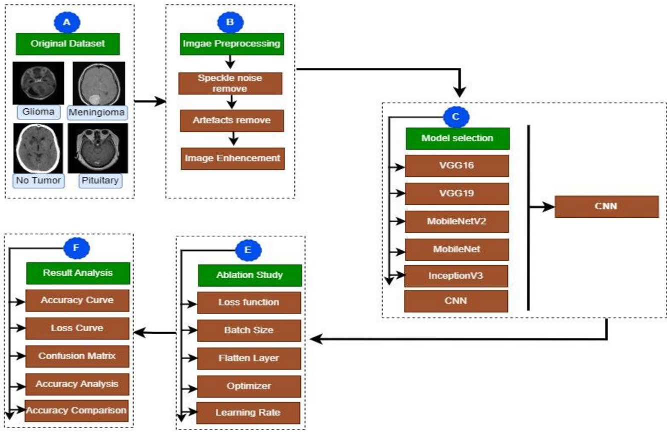


Fig. 1. An overviewof the entire classification process.

II. Literature Review

A classification approach was provided by Santhosh and his colleagues to distinguish between normal and abnormal brain tissues. The segmentation of the system was based on threshold and watershed. SVM provided overall classification accuracy of 85.32 percent [13].

Arunkumar and his research associates created a world-class brain tumor classification architecture on the basis on traditional machine vision techniques like Fourier transform image enhancement, completely automated trainable segmentation, histogram-of-oriented-gradients (HOG) feature extraction, and an ANN-based classification model. Size, circularity, and gray-scale average are used to filter out non-ROI brain components and a k-fold-cross validation technique, the constructed model successfully distinguished between normal and pathological brain slices with the accuracy of 92.14%. [14].

A brain tumor categorization algorithm based on brain MRIs obtained from RD-BVH was proposed by Hafeez Ullah and research fellows. Brain MRI slices' intensity, shape, and texture features were retrieved, and the proposed methodology achieved 97% accuracy [15].

An approach for classifying tumors that uses CNN and a genetic algorithm was proposed by Amin Kabir et al. in 2019 [16]. To reduce validation error, the authors used a genetic algorithm. Prior to the CNN architecture, the images were rescaled using data normalization, and they underwent augmentation to achieve perfect rotation. The accuracy was 94.2% using the recommended procedure. The new method, according to the authors, is sufficiently successful to detect tumors.

Biswas et al. in 2021[17] an efficient training feature. The proposed network construction method, known as "Levenberg-Marquardt," offers 95.4% accuracy, 94.58% sensitivity, and 97.83% specificity. Comparatively speaking, this improved result outperforms other current detection methods. Getting great results comes down to two main things: using the right preprocessing steps and a powerful training function.

Brain cancers can be identified and classified from MRI images using a quicker Region-based CNN (faster R-CNN) technique used by Avşar, E. et al. [18]. Their model's accuracy was 91.66% as measured. In study [19], a method for classifying MRI brain cancer was also suggested that uses SVM with grayscale, symmetry, and texture features to get information about features.

Precious et al. [20] propose three optimizers, including ADAM, SGDM, and RMSprop, from whom detection accuracy of 98.1%, 92.5%, and 83.0% is attained. In order to detect tumors, four supervised machine learning classifiers are used once the features have been retrieved using CNN. Discriminant analysis, Naive Bayes, SVM, and KNN classifiers are among the classifiers that are employed. 96.2%, 94.3%, 75.0%, and 96.2% of the classifiers' accuracy were obtained, respectively.

Papageorgiou et al. [21] created the fuzzy cognitive map (FCM) approach to represent model experts. A computationally sophisticated training technique known as the activation Hebbian algorithm was added to the FCM ranking model to enhance its classification capabilities. Medical resources, which included 100 instances, were used to verify the proposed method. For low-grade and high-grade brain tumors, the FCM model correctly diagnosed patients in 90.26% (37/41) and 99.22% (55/59) of the cases, respectively. Comparing the proposed model's results to those of current algorithms like fuzzy decision trees and decision trees, the proposed model's results show a marginally higher accuracy. The same kind of first-hand data was used to compare them, and while they were able to get high accuracy, they couldn't get high memory.

In order to work with 2D (two-dimensional) images, John Schmeelk [22] used a two-dimensional wavelet transform (2D-WT). The comparison of the two transforms on separated elements was covered in depth by the authors. A similar image was also subjected to a comparison of the global qualities offered by the Fourier transform (FT) approach and the wavelet transform. The Gaussian subfield wavelet was chosen for this study because, for some reason, it made it possible to compare it to the Fourier technique.

In our findings we are improving the model and get the high accuracy from the model. In our approach firstly we decrease noise by using various image processing techniques and then we are developed a model which give us the best accuracy and our approach beat all the existing model.

III. Dataset Description

The Brain Tumor MRI dataset has a total of 7022 MRI images analyzed for this research. There are four classes in the dataset: glioma, meningioma, no tumor, and pituitary. There are 1621 images in the glioma class, 1645 images in the meningioma class, 2000 images in the no tumor class, and 1757 images in the pituitary class. All images of this datasets are 512 × 512 pixels in grayscale presentation. The dataset was taken from the open-source website Kaggle. As shown in Table I, Fig. 2 the dataset is described in detail:

TABLE I. SHOW THE DATASET DESCRIPTION

Name	Description
Total Number of Images	7022
Average Dimension	512 x 512
Color Grading	Grayscale
Data Format	JPG
Glioma	1621
Maningioma	1645
No Tumor	2000
Pituitary	1757

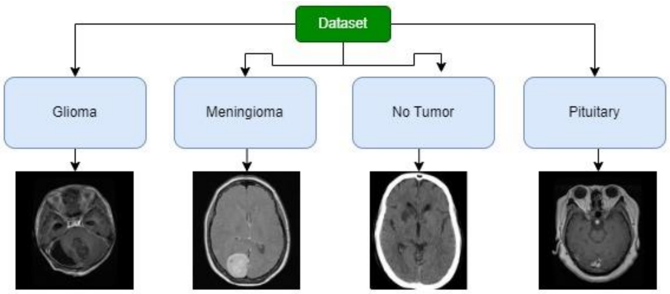


Fig. 2. Brain Tumor MRI dataset containing four classes with various noise and artifacts.

A. Image Processing

There is a lot of noise and artefacts in brain tumor MRI dataset images, so this study focuses on improving the model's accuracy through image processing techniques. Because images are usually filled with noise and artefacts, image processing is the first step in training a deep-learning model. First, a median filter is used to remove noise from this image, then a morphological opening is used to remove artefacts.

B. Remove Spackle Noise

The brain tumour MRI dataset has spackle noise, as previously stated. Median filters are useful for removing spackle noise.

C. Median Filter

Median filter is a well-known order-statistic filter that excels at removing certain types of noise, including Gaussian, random, and salt-and-pepper noise. The image of this step is given in Fig. 3.

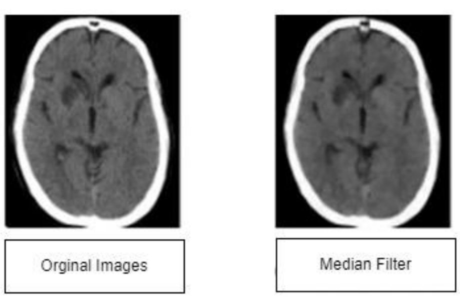


Fig. 3. Median filtered images.

The first photo are the original images and the second photo is the output of the median filter.

D. Artifact Removal

In addition to the brain tumour MRI dataset having artefacts, morphological operations are used to remove these artefacts [23]. Various morphological operations can be used for eliminating artifacts, but this study uses morphological opening techniques.

E. Morphological Opening

The kernel size of the filter depends on the operation to be performed. The output of this step is shown in Fig. 4.

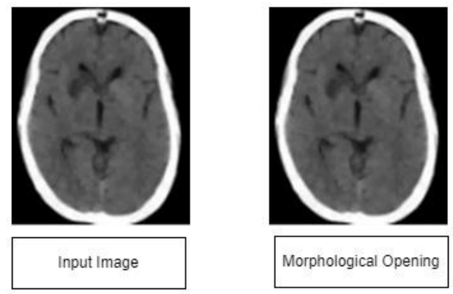


Fig. 4. Output of the morphological opening.

The first image are the input images which we get from the median filter and the second photo is the output of the morphhological opening.

$$g\ (j, s) = f\ (j, s) * u\ (j, s) + (j, s) \quad (1)$$

Where,

g(j,s) – debased image u(j,s) – multiplicative noise

f(j,s) – original image η (j,s) – additive noise

Before diagnosis, removing additive noise from ultrasound images is necessary, but multiplicative noise can be allowed, given in the equation below:

$$g\ (j, s) = f\ (j, s) * u\ (j, s) + \eta\ (j, s) - \eta\ (j, s) \quad (2)$$

$$g\ (j, s) = f\ (j, s) * u\ (j, s) \quad (3)$$

F. Clahe

The Clahe technique is used to balance the overall contrast. A further sophisticated version of adaptive histogram equalization is called CLAHE. Clahe was developed to improve the quality of medical imaging of complex structures [24, 25]. The output of this step is shown in Fig. 5.

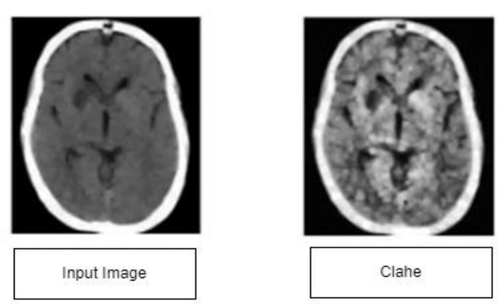


Fig. 5. Output of the clahe.

The first images are the original images and the second images are the output of the CLAHE.

Let an image size be M x M and each tile size for image is m x m then the total number of tiles is calculating as:

$$T = \frac{M \times M}{m \times m} \quad (3)$$

Clip limit $C_L = M_{CL} \times M_{avg}$ is used to construct the histograms for these tiles.

Where,

$M_{CL} =$ normalized contrast limit. $M_{AVG} =$ total pixels average value.

The equation of pixels avarage is (3):

$$M_{AVG} = \frac{Mx \times My}{Mg} \quad (4)$$

Where,

Mg = total gray levels Mx and My = total pixels of x and y dimension

$$M_{CP} = \frac{M \sum cl}{Mg} \quad (5)$$

Where,

$C_L =$ the clipped pixel.

$$M_r = \frac{M_g}{M_r} \quad (6)$$

Where,

M_r is remaining number of clipped pixels

clahe formula:

$$I_c(p,q) = \mathrm{T}(\iota(p,q) = \frac{(L-1)}{PQ} \sum_j^K n_j \quad (7)$$

G. Verification

It is possible to lose a lot of image quality when using many image preprocessing algorithms, so various types of geometric analysis like PSNR, SSIM, MSE, and RMSE are performed to determine if the image quality has been compromised.

H. MSE

MSE describes the pixels of the two pictures under comparison as having a cumulative squared error. A value close to 0 indicates acceptable image quality, while the MSE value ranges from 0 to 1. A value of 0 indicates a picture with no noise. Values higher than 0.5 indicate a decline in quality.

$$MSE = \frac{1}{AB} \sum_{i=0}^{r-1} \sum_{j=0}^{n-1} (O(r,s) - P(r,s))^2 \quad (8)$$

Where,

The ground truth image is O, the image which is processed is P and A, B denote the pixels of O and P, and r, s denote pixel rows of p,q.

I. PSNR

Calculating PSNR begins with calculating MSE. PSNR is then estimated using the following formula:

$$PSNR = 10 \log_{10} \left(\frac{Q^2}{MSE}\right) \quad (9)$$

Q is the highest fluctuation in the input image data type. A maximum of 255 pixels is used as image maximum value. Typically, PSNR should be between 30 and 50 dB for an 8-bit image [26].

J. SSIM

SSIM measures the decline in image condition caused by preprocessing processes. In this estimation, 1 indicates "perfect structural similarity" and 0 means "no structural similarity" [9].

$$SSIM(r,s) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_x^2 + c_1)(\sigma_x^2 + \sigma_x^2 + c_2)} \quad (10)$$

Where,

r, s is two image , $\sigma_r^2, \sigma_s^2$is variance, σ_{rs} is covariance of the images and μ_r, μ_s is the average of two image calculated using the Gaussian window.

K. RMSI

Image quality is measured by RMSE, which compares the original and processed images. RMSE values near 0 indicate good image quality and fewer errors.

$$RMSE = \sqrt{\sum_{J=1}^{N} (d_{fi} - \frac{d_d)^2}{N}} \quad (11)$$

Where, d_{fi} is the different of predict value, d_d is the actual value, N is the Size of the Dataset.

TABLE II. SHOWN PSNR, MSE, SSIM, RMSE

Image	MSE	PSNR	SSIM	RMSE	Image
Image_1	15.17	39.35	0.962	0.13	Image_1
Image_2	13.63	40.66	0.961	0.13	Image_2
Image_3	14.25	42.59	0.964	0.11	Image_3
Image_4	13.13	38.28	0.968	0.12	Image_4
Image_5	12.38	45.47	0.962	0.09	Image_5

In this Table II we are showing some statistical value which proved that out image processing techniques are the best. Image quality is not damaged after image preprocessing.

IV. DATA SPLIT

After analyzing the statistical value of the image, the full dataset is divided into three segments (training set, validation set, and testing set). Three splitting ratios for training-testing data (90:10, 80:20, and 70:30) are often used to evaluate the effect of the overall accuracy of a model is affected by the size of the training-testing data [27]. According to this study, the 70:30 ratio means 70% train sets, 10% validation sets, and 20% test sets.

V. PROPOSED MODEL

As previously said, this research explored with a total of five transfer learning network to identify the ideal network based on accuracy in order to discover the best transfer learning model for the classification issue. Transfer Learning

Model: There are five pre-trained models total—InceptionV3, MobileNetV2, MobileNet VGG16, and VGG19—that are trained on training data and tested on testing data.

A. VGG-16

Simonyan and Zisserman [28] introduced the DCNN model known as VGG16. The model achieving 92.7% top 5 test accuracy in the ImageNet dataset [29]. A VGG16 produced a substantially greater accuracy than a fully trained architecture, according to studies on the efficiency of transfer learning [30]. The kernel may learn more complicated characteristics with the help of the VGG model's enhanced depth. The VGG16 architecture is a convolutional neural network set. It is regarded as one of the best computer vision models to date. VGG16 is unusual in that it only employs 3 x 3 filter convolution layers with a stride 1 and always uses the same padding and maxpool layer with a 2 x 2 filter stride 2. VGG16 has 16 layes. Spatial pooling is achieved in this model by employing five max pooling layers. After a sequence of convolutional layers, three fully connected (FC) layers are added. Finally, there's the softmax layer. In all networks, always set the layers 1 and 2 in the same way.

B. VGG-19

There are 19 layers in the VGG19 model, a variation of the VGG model. The VGG19 model concludes with three additional FC layers, totaling 19 layers with 4096, 4096, and 1000 neurons in each layer. Moreover, five Maxpool layers are included as well as a Softmax layer. It is a characteristic of convolutional layers that ReLU is activated. Non-linearity is included into models using Rectified linear units (ReLu), which increases classification and compute performance. Three layers were built, all of which were entirely interconnected. Finally, there is a softmax function as the model's final layer.

C. MobileNet

The MobileNet model is TensorFlow's first computer vision model designed specifically for mobile applications. MobileNet employs depth-wise separable convolutions. The number of parameters is dramatically reduced compared to a network with regular convolutions of the same depth. As a result, portable deep neural networks have been developed. To generate a depth-separable convolution, two techniques are used.

- In-depth convolution.

- Convolution at the point of interest.

MobileNet, a CNN class that Google freely licenses, is a great starting point for training our ultra-short and ultra-fast classifiers.

D. MobileNetV2

The Google community has suggested MobileNetV2. There are two kinds of blocks in it, and each block has three levels. Each block has 11 convolutional layers with 32 filters in the first, third, and second layers. All layers use the rectified linear activation function (ReLU). In order to prevent non-linearity from corrupting a significant volume of data, longitudinal bottlenecks are essential between layers. There is a difference between the strides of the two blocks, with block 1 having a stride of one and block 2 having a stride of two.

E. InceptionV3

A new InceptionV3 design aims to reduce the needed computational power by modifying earlier Inception designs. It is possible to decrease the computational cost by regularizing, reducing the dimension, factorizing convolutions, and parallelizing computations.

F. BrainNet-7

The BrainNet-7 has three convolutional layers and one max pool layer for each convolutional layer Fig. 6. There are 3 X 3 convolutional kernels in the model. There is a dropout value of 0.5 in the first block of the convolutional kernel, and 32 in the second block of the convolutional kernel. 'Relu' has been selected as the activation function for the final layer, followed by 'SoftMax'. Using batch size 32 and Adam optimizer, categorical crossentropy was used as a loss function. There is an o.oo1 learning rate.

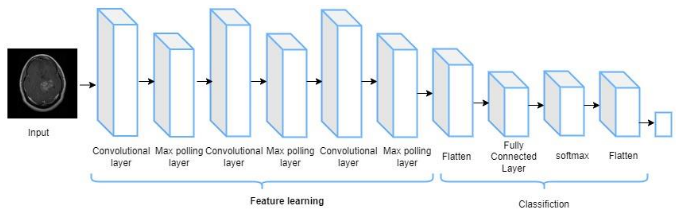


Fig. 6. BrainNet-7 model architecture.

VI. Training Approach

For training the models, the batch size is 16, and the maximum number of epochs is 100[31]. During training, Keras' "callback" function was used to store the weights of the best model based on a minimal loss value [32]. Adam has been used as an optimizer at a learning rate of 0.001. Categorical cross-entropy is the defaulting loss function for multiclass challenges [33]. In order to forecast the probability for each class, 'Softmax' activation is used. Since Softmax normalizes all values between 0 and 1, their aggregate always equals 1.

$$Softmax(y_i) = \frac{\exp(y_i)}{\sum_j \exp(y_i)} \tag{12}$$

VII. Ablation Study

An ablation study is often carried out in CNN-based applications to evaluate the model's stability and performance after deleting or changing various layers and hyper-parameters. Using hyperparameter ablation to develop a robust and fine-tuned network in this study.

VIII. Result and Discussion

In this section, several performance matrix are shown as mathematically.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

$$Recall = \frac{TP}{TP+FN} \tag{14}$$

$$Specificity = \frac{TN}{TN+FP} \quad (15)$$

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

$$ACC = 2\frac{precision*recall}{precision+recall} \quad (17)$$

$$FPR = \frac{FP}{FP+TN} \quad (18)$$

A. Result of Transfer Learning and CNN Model

Table III shows the training accuracy, test accuracy, validation accuracy, and train, test and validation loss for the five transfer learning models. The CNN model has the highest accuracy, as can be seen in the table.

TABLE III. RESULT OF TRANSFER LEARNING AND CNN MODEL

Model	**Train_Accuracy**	**Train_Loss**	**Val_Accuracy**	**Val_loss**	**Test_Accuracy**
VGG19	96.67	0.22	95.63	0.21	95.24
VGG16	97.77	0.18	96.83	0.12	96.13
Mobile Net	95.43	0.18	94.23	0.28	94.24
Mobile Net V2	95.76	0.25	94.63	0.32	94.59
InceptionV3	76.86	0.421	76.21	0.392	76.21
CNN	**97.91**	**0.15**	**97.95**	**0.11**	**97.95**

In this table we are showing the Accuracy of the all model which we employing in this study.

IX. RESULT AND ABLATION STUDY

It is possible to improve classification accuracy by changing several design components to make it more reliable. A total of five studies are run as an ablation study, modifying various BrainNet-7 elements based on the optimized VGG16 architecture.

A. Case Study-1 Changing Flatten Layer

In case study 1, it is shown in Table IV that the flatten layer provides the highest accuracy when it is used. Additionally, global average and global maximum pooling do not provide good accuracy. Flattening the layer gives 96.93% accuracy, while global max and global average pooling give 95.24 and 96.83% accuracy, respectively.

TABLE IV. CHANGING FLATTEN LAYER

		Case Study 01		
Configuration No.	Flatten layer type	Epoch x training time	Test accuracy (%)	Finding
1	**Flatten**	**97 x 5s**	**97.95%**	**Highest accuracy**
2	Global Max pooling	60 x 4s	95.24%	Accuracy dropped
3	Global Average pooling	54 x 5s	96.83%	Accuracy dropped

B. Case Study-2 Changing the Batch Size

Changing the batch size is the subject of case study two. A batch size of 32 is the most accurate, followed by 32, 64, and 16. When the batch size is 32 the test accuracy is 96.39%.

C. Case Study-3 Changing Loss Function

Case study 3 experiments with changing loss functions and finds that categorical cross-entropy gives the best outcomes, (Table V) 96.93%.

TABLE V. CHANGING LOSS FUNCTION

		Case Study 03		
Configuration No.	**Loss Function**	**Epoch x training time**	**Test accuracy (%)**	**Finding**
1	Binary Crossentropy	Error	Error	Error
2	**Categorical Crossentropy**	**43 x 5s**	**98.13%**	**Highest accuracy**
3	Mean Squared Error	96 x 5s	96.82%	Accuracy dropped
4	Mean absolute error	12 x 4s	68.25%	Accuracy dropped
5	Mean squared logarithmic error	45 x 5s	96.83%	Accuracy dropped

D. Case Study-4 Changing Optimizer

Adam optimizer provides the highest accuracy when compared to Nadam, SGD, and ADamax optimizers in case study 4 (Table VI).

TABLE VI. CHANGING LOSS OPTIMIZER

		Case Study 04		
Configuration No.	Optimizer	Epoch x training time	Test accuracy (%)	Finding
1	**Adam**	**97 x 5s**	**98.41%**	**Highest accuracy**
2	Nadam	44 x 5s	98.13%	Previous dropped
3	SGD	90 x 5s	84.13%	Accuracy dropped
4	Adamax	88 x 5s	90.48%	Accuracy dropped

E. Case Study-5 Changing Learning Rate

In comparison to 0.001, 0.0001, and 0.01, when using 0.01 provide the highest accuracy (Table VII).

TABLE VII. CHANGING LEARING RATES

		Case Study 05		
Configuration No.	Learning rate	Epoch x training time	Test accuracy (%)	Finding
1	0.01	92 x 55s	98.41	Accuracy dropped
2	**0.001**	**97 x 5s**	**98.43%**	**Highest accuracy**
3	0.0001	68 x 57s	97.28	Accuracy improved

F. Performance Analysis of Best Model

After executing the ablation study on the suggested BrainNet-7 model, an improvement in classification accuracy is shown on CNN model. A summary of BrainNet-7 final setup is provided in Table VIII.

TABLE VIII. EVAULATED PERFORMANCE OF BEST MODEL

Configuration	Value
Size of images	224 x 224
Epoch	97
Optimizer	Adam
Learning rates	0.001
Batch sizes	16
Activation functions	Softmax
Dropouts	0.5
Momentums	0.9
Accuracy	98.43

G. Performance Analysis and Statistical Analysis

In Table IX show the FPR, FNR, FDR, KC, MCC, MAE and RMSE of the best hyper-tuned CNN (BrainNet-7) model.

TABLE IX. PERFORMANCE ANALYSIS AND STATISTICAL ANALYSIS

Accuracy	FPR (%)	FNR (%)	FDR (%)	KC (%)	MCC (%)	MAE	RMSE
98.43	1.46	2.22	2.59	99.01	88.35	2.09	5.79

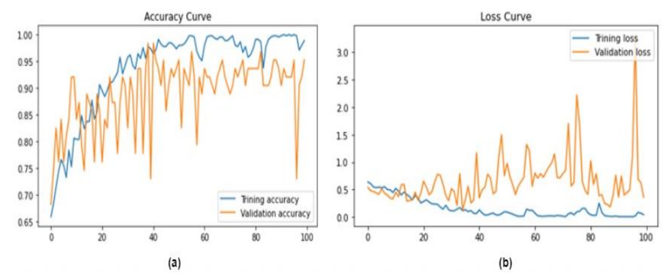


Fig. 7. Losses and accuracy curve.

Fig. 7 shows the accuracy and loss curves for the best-performing model. The training curve converges without bumps from the first to the final epoch. Based on the difference between validation and training accuracy curves, there is no evidence of overfitting in the course of training. As with the training curve, the loss curve in figures 6 unites smoothly to the end epoch. The trainings and losses curves indicate that neither overfitting nor underfitting occurred. The confusion matrix is also shown in Fig. 7.

H. Comparison with Existing Work

This section Table X compares the proposed CNN models (BrainNet-7) to classify. The accuracy, reliability and competency of these earlier investigations and our suggested methods are compared in Table VIII.

TABLE X. COMPARISON BETWEEN PROPOSEDMODEL AND PREVIOUS STUDY

Paper	Dataset	Classifier	Accuracy
[13]	MRI	SVM	85.23%
[14]	MRI	ANN	92.14%
[15]	MRI	Machine Learning Classifer	97.00%
[16]	MRI	CNN	94.2%
[17]	MRI	CNN	95.4%
[18]	MRI	Faster-R-CNN	91.66%
[20]	MRI	Machine learning classifier	98.1%
Proposed Model (BrainNet-7)	**MRI**	**Fine-tuned Transfer Learning**	**98.43**

X. CONCLUSION

For training, deep learning systems for tumor identification in medical imaging need massive, annotated datasets. Subspecialty, radiologists often manually annotate images. AI's progress in medical imaging is hampered by unreasonably high related costs (time and expertise). Transfer learning techniques have been studied to train a competitive classifier with the least amount of annotation expense. Transfer learning is a popular technique that allows models to apply the information they have gained on massive datasets to new recognition and classification tasks. This study proposes a system based on a transfer learning model for classifying brain tumor MRI images more accurately which can reduce the death ratio. This experiment uses various preprocessing techniques to remove speckle noise and artefacts from the image. Five transfer learning models have been experimented with in the Brain Tumor MRI dataset. After that, a hyperparameter ablation study is conducted on the best-performed transfer learning model to get the best outcome from the model. The proposed model has been fine-tuned with proper hyperparameters, that's why it gained the highest accuracy.

XI. LIMITATIONS AND FUTURE SCOPE

The suggested CNN models for multiclass classification outperformed traditional classifiers considerably, according to the whole discourse of this work. The dataset for the suggested model is too small—it only contains 7022 images—and the study's major limitation—the leakage of a significant amount of genuine medical data—can be solved in the future. As a result, it is possible to expand the quantity of unprocessed medical photos and assess the performance of the suggested model using real-time medical data in the future. The suggested model of this research does, however, successfully categorize the four types of brain tumor classes in the majority of test instances. It is feasible to make sure that the recommended fine-tuned CNN model is accurate and improved in all areas of diagnosis, despite a few small disadvantages.

REFERENCES

[1] Cancer statistics [Online]. Available: https://www.cancer.net/cancer-types/bra in-tumor/statistics. Accessed: 17-Feb-2020.

[2] Cancer statistics [Online]. Available: https://www.nhp.gov.in/world-brain -tumour-day2019_pg, 2019. Accessed: 17-Feb-2020.

[3] G.S. Tandel, M. Biswas, O.G. Kakde, A. Tiwari, H.S. Suri, M. Turk, B. K. Madhusudhan, L. Saba, J.S. Suri, A review on a deep learning perspective in brain cancer classification, Cancers 11 (1) (2019) 111, https://doi.org/10.3390/ cancers11010111.

[4] D.N. Louis, A. Perry, G. Reifenberger, D.A. von, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O.D. Wiestler, P. Kleihues, Ellison DW the 2016 World G.S. Tandel et al. Computers in Biology and Medicine 122 (2020) 103804 28 Health Organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (2016) 803–820, https://doi.org/10.1007/ s00401-016-1545-1.

[5] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imag. 35 (5) (2016) 1240–1251, https://doi.org/10.1109/TMI.2016.2538465.

[6] A. Kotrotsou, P.O. Zinn, R.R. Colen, Radiomics in brain tumors: an emerging technique for characterization of tumor environment, Magnetic Resonance Imaging Clinics 24 (4) (2016) 719–729, https://doi.org/10.1016/j.mric.2016.06.006.

[7] S. Bauer, R. Wiest, L.P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies, Phys. Med. Biol. 58 (13) (2013) R97, https://doi. org/10.1088/0031-9155/58/13/R97.

[8] Delattre, J-Y., Bernsen, H. J. J. A., Frenay, M., Tijssen, C. C., and Grisold, W. (2014). Adjuvant Procarbazine , Lomustine , and Vincristine Chemotherapy in Newly Diagnosed Anaplastic Oligodendroglioma : Long-Term Follow-Up of EORTC Brain Tumor Group Study 26951. J. Clin. Oncol. 31 (3), 344–350. doi:10.1200/ JCO.2012.43.2229.

[9] Essig, M., Anzalone, N., Combs, S. E., Dörfler, A., Lee, S-K., Picozzi, P., et al. (2012). MR Imaging of Neoplastic Central Nervous System Lesions: Review and Recommendations for Current Practice. AJNR Am. J. Neuroradiol 33 (5), 803–817. doi:10.3174/ajnr.a2640.

[10] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., and Aerts, H. J. W. L. (2018). Artificial Intelligence in Radiology. Nat. Rev. Cancer 18 (8), 500–510. doi:10.1038/s41568-018-0016-5.

[11] Razzak, M. I., Naz, S., Zaib, A., and Ahmad, Z. (2018). Deep Learning for Medical Image Processing: Overview, Challenges and the Future. Lecture Notes Comput. Vis. Biomech. 26, 323–350. doi:10.1007/978-3-319-65981-7_12.

[12] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., et al. (2016). “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?” IEEE Trans. Med. Imaging 35 (5), 1299–1312. doi:10.1109/tmi.2016.2535302.

[13] Seere, S.K.H. and Karibasappa, K., 2020. Threshold segmentation and watershed segmentation algorithm for brain tumor detection using support vector machine. European Journal of Engineering and Technology Research, 5(4), pp.516-519.

[14] Arunkumar, N., Mohammed, M.A., Mostafa, S.A., Ibrahim, D.A., Rodrigues, J.J. and de Albuquerque, V.H.C., 2020. Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks. Concurrency and Computation: Practice and Experience, 32(1), p.e4962.

[15] Ullah, H., Batool, A. and Gilanie, G., 2018. Classification of Brain Tumor with Statistical Analysis of Texture Parameter Using a Data Mining Technique. International Journal of Industrial Biotechnology and Biomaterials, 4(2), pp.22-36.

[16] AminKabir Anaraki, MoosaAyati, FoadKazemi, “ Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms” , Biocybernetics and Biomedical Engineering, ELSEVIER, vol. 39, Issue. 1, , pp. 63¬ 74, January-March 2019.

[17] Biswas, A. and Islam, M.S., 2021, January. Brain tumor types classification using K-means clustering and ANN approach. In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (pp. 654-658). IEEE.

[18] Salçin, K., 2019. Detection and classification of brain tumours from MRI images using faster R-CNN. Tehnički glasnik, 13(4), pp.337-342.

[19] Nandpuru, H.B., Salankar, S.S. and Bora, V.R., 2014, March. MRI brain cancer classification using support vector machine. In 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science (pp. 1-6). IEEE.

[20] Precious, J., Kirubha, S.P. and Evangeline, I.K., 2022. Automatic Brain Tumor Classification in 2D MRI Images Using Integrated Deep Learning and Supervised Machine Learning Techniques. In Intelligent Vision in Healthcare (pp. 131-144). Springer, Singapore.

[21] Papageorgiou, E.I., Spyridonos, P.P., Glotsos, D.T., Stylios, C.D., Ravazoula, P., Nikiforidis, G.N. and Groumpos, P.P., 2008. Brain tumor characterization using the soft computing technique of fuzzy cognitive maps. Applied soft computing, 8(1), pp.820-828.

[22] Schmeelk, J., 2002. Wavelet transforms on two-dimensional images. Mathematical and computer modelling, 36(7-8), pp.939-948.

[23] Abbas, A.H.; Kareem, A.A.; Kamil, M.Y. Breast Cancer Image Segmentation Using Morphological Operations. Int. J. Electron. Commun. Eng. Technol. 2015, 6, 8–14.

[24] Wang, X.; Liang, G.; Zhang, Y.; Blanton, H.; Bessinger, Z.; Jacobs, N. Inconsistent Performance of Deep Learning Models on Mammogram Classification. J. Am. Coll. Radiol. 2020, 17, 796–803.

[25] Zheng, Y. Breast Cancer Detection with Gabor Features from Digital Mammograms. Algorithms 2010, 3, 44–62.

[26] Van Droogenbroeck, M.; Buckley, M.J. Morphological Erosions and Openings: Fast Algorithms Based on Anchors. J. Math. Imaging Vis. 2005, 22, 121–142.

[27] Beeravolu, A.R.; Azam, S.; Jonkman, M.; Shanmugam, B.; Kannoorpatti, K.; Anwar, A. Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN. IEEE Access 2021, 9, 33438–33463.

[28] Wang, P.; Wang, J.; Li, Y.; Li, P.; Li, L.; Jiang, M. Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing. Biomed. Signal. Process. Control 2021, 65, 102341.

[29] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.

[30] Shuyue, G.; Murray, L. Breast cancer detection using transfer learning in convolutional neural networks. In Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 10–12 October 2017; pp. 1–8.

[31] Shallu; Mehra, R. Breast cancer histology images classification: Training from scratch or transfer learning? ICT Express 2018, 4, 247–254.

[32] Hameed, Z.; Zahia, S.; Garcia-Zapirain, B.; Javier Aguirre, J.; María Vanegas, A. Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models. Sensors 2020, 20, 4373.

[33] Lorencin, I.; Šegota, S.B.; Andeli´c, N.; Mrzljak, V.; ¯ Cabov, T.; Španjol, J.; Car, Z. On Urinary Bladder Cancer Diagnosis: Utilization of Deep Convolutional Generative Adversarial Networks for Data Augmentation. Biology 2021, 10, 175.

Research on Real-time Monitoring of Video Images of Traffic Vehicles and Pedestrian Flow using Intelligent Algorithms

Xiujuan Dong*, Jianping Lan, Wenhuan Wu
School of Electrical & Information Engineering
Hubei University of Automotive Technology
Shiyan, Hubei 442002, China

Abstract—**The development of urbanization has brought many traffic problems, among which the delayed feedback of traffic flow and people flow has led to traffic congestion. In order to effectively analyze the traffic flow and people flow on the traffic road, this research proposes a traffic surveillance video image object detection model based on the improved Vibe algorithm, and uses the moving historical image to track the traffic flow and people flow. Finally, the performance analysis of the algorithm shows that the loss rate of the improved Vibe algorithm proposed in the study is only 0.25%, and its detection accuracy reaches 91.25%. The above results show that the use of Vibe intelligent algorithm can significantly improve the detection effect of traffic flow and pedestrian flow in traffic monitoring video, help to improve urban traffic management ability and promote the development of urban modernization.**

Keywords—Urban development; object detection; traffic video; Vibe algorithm; visitors flowrate; image filtering

I. INTRODUCTION

Rapid economic development promotes urban development. However, in urban development, the increase in the number of motor vehicles and long-term residents increases the difficulty of traffic control. Therefore, a large number of studies have put forward the concept of intelligent transportation on the existing urban traffic problems. In intelligent transportation, it is believed that the processing of traffic data can be realized by using intelligent algorithms, and the traffic pressure can be relieved by understanding the traffic conditions [1, 2]. In the current research, the emphasis is on the systematic and real-time nature of intelligent transportation, which realizes the traffic management of multiple road sections through its system planning, makes timely traffic diversion through real-time nature, and ensures the convenience of people's travel [3]. With the development of computer technology, intelligent algorithms have been continuously introduced in the monitoring and monitoring of traffic flow and people flow. The purpose is to use the moving target tracking algorithm to grasp the changes of vehicle and people flow in the traffic section in real time, and provide a planning scheme for subsequent sections [4]. However, it has to be admitted that my country's traffic video image traffic flow and people flow monitoring technology started later than foreign countries, and the existing detection schemes are difficult to deal with the increasing traffic problems. In this study, a video image target recognition model based on the improved Vibe algorithm is proposed, and the tracking of vehicles and human bodies is realized with the help of motion history images, and the real-time monitoring of traffic flow and people flow is realized, in order to relieve urban traffic pressure and speed up China's Urbanization.

The main structure of the article is divided into four parts. In the second part, the development status of urban traffic monitoring technology and image processing technology is analyzed. In the third part, the intelligent algorithm design for traffic vehicle and pedestrian flow monitoring is proposed. In the fourth part, the algorithm performance analysis and monitoring effect test are conducted.

There are two innovations in this paper. First, image processing is carried out for traffic flow and pedestrian flow monitoring videos; the second is to study the further processing of the surveillance video image and target tracking with the aid of only algorithms, which improves the real-time detection capability of the surveillance video.

II. RELATED WORKS

With the continuous progress of urban traffic, the monitoring technology for the detection of urban traffic vehicles and people flow has also become increasingly prominent. In the same way, aiming at the existing urban traffic flow management problems, the research proposes to introduce intelligent algorithms to realize the monitoring and control of vehicle flow and pedestrian flow. In order to further understand the monitoring status of traffic flow, the research status in this field is analyzed. In order to promote the development of urban modernization, Pan S et al. used network tomography to realize the optimization of urban traffic monitoring, and used FIM and multi-agent reinforcement learning methods to help monitors improve the processing capacity and performance of surveillance video. The simulation experiment proves that the proposed scheme can effectively reduce the error of monitoring video processing compared with the traditional scheme [5]. In order to improve the traffic monitoring ability, Rogowski M gained experience from the traffic monitoring of tourists in national parks. In this study, Rogowski M proposed a comprehensive monitoring method. It has been proved by practice that the proposed method can effectively monitor and evaluate tourists [6]. In

*Corresponding Author.

order to monitor the traffic flow, Zhu L et al. proposed a self-organized flow monitoring measure, and finally proved the feasibility of the measure through safety analysis and experimental verification. However, although this research can achieve traffic flow monitoring, it is difficult to extract effective targets from the monitoring video [7]. In order to obtain traffic conditions in low-light environments such as nighttime, Cai X et al. explored light control strategies in traffic monitoring, aiming to use threshold increments to evaluate disability glare, and showed in the final results, research. The proposed light control strategy can reduce the interference caused by glare in traffic monitoring and improve the effect of traffic monitoring [8]. Almeida TD et al. believe that most of the current traffic monitoring systems require high maintenance costs, so in order to improve the effect of traffic monitoring and reduce costs, a community-based decentralized traffic monitoring system is proposed. The traffic conditions of the road sections are analyzed, and the comparative analysis finally confirms that the accuracy of the system proposed by this study can be higher than 90% in multiple scenarios. However, this research is difficult to monitor the rapid growth of traffic flow and people flow in real time [9].

The processing of traffic monitoring is to process the detected video images. In the research of video image processing, a large number of scholars have proposed effective methods. In order to increase the sensitivity and specificity of medical video image processing, Obukhova N et al. proposed a personalized processing method, which uses artificial intelligence technology to realize the visualization of video image data, which is displayed in the result classification. The proposed video image processing method has high accuracy and improves the diagnostic efficiency of medical images [10]. Long C proposed an edge computing framework to optimally group mobile devices into video processing groups, and at the same time proposed an efficient collaborative video processing scheme that can achieve sub-optimal performance with existing detection accuracy. Each parameter is evaluated, and the effectiveness of the method proposed in the study is confirmed in the simulation experiment, and its superiority compared with the baseline scheme is proved [11]. Ragan-Kelley J et al. believed that there was too little work in different stages in image processing to load the results into the corresponding memory and would add extra cost, so the team wrote a high-performance image processing code that implements The simplification of image processing and the ability to use a dedicated image processor to generate effective results finally confirmed the efficiency of the image processing code through experiments. However, the image processing code proposed in the study has some limitations in application, and is difficult to deal with the environment with huge data [12]. In order to detect acephate, Osman MJ et al. proposed an image processing technology combined with an aptamer sensor, using the sensor to obtain the relevant information of the aptamer, and using the image processing technology to achieve colorimetric detection, and finally proved in the experiment the effectiveness of the current method. Although this research can effectively achieve image detection, it uses sensor technology, which is vulnerable to environmental interference [13]. In order to reduce the mite infection in bee farming, Sevin S et al. proposed an image capture device that can automatically focus detection. The device is compatible with cloud storage and 5G technology, and can distinguish bees from Varroa mites through images, and finally train on samples. The method proposed in the research shows that the method can effectively distinguish between honeybees and Varroa mites [14].

To sum up, with the continuous development of intelligent technology, the research on video image processing in modern technology research is increasing, but in traffic monitoring, there are still few effective processing solutions for monitoring video. For the real-time monitoring of traffic flow and people flow, a monitoring video image processing algorithm based on Vibe intelligent algorithm is proposed to detect the traffic flow and people flow. See Fig. 1 for the structure and flow of research contents.

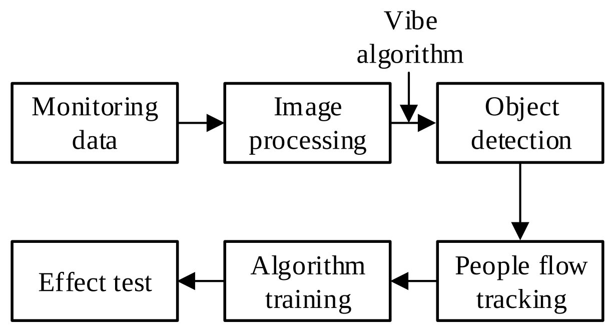


Fig. 1. Workflow diagram.

III. Design of Target Detection Algorithm for Traffic Surveillance Video Image

A. Traffic Flow and People Flow Detection

In urban traffic video detection, affected by the traffic environment and the natural environment, the monitoring video has significant interference noise. Therefore, in order to accurately monitor the real-time traffic flow and people flow, the video image needs to be preprocessed first [15, 16]. In the research, for image denoising in surveillance video, the method of image filtering is used to extract the target features in the image, and the median filtering method is used to suppress the noise in the image. The basic principle of median filtering is to replace the value in the image with the median value of the region where it is located. One-dimensional median filtering is shown in equation (1) [17].

$$y = med(f_0, f_1, ..., f_k) \tag{1}$$

In formula (1), it $(f_0, f_1, ..., f_k)$ represents the sequence value in the one-dimensional sequence; k it represents the sequence length. After filtering, the image needs to be enhanced. In the image enhancement, the research adopts Laplacian sharpening to achieve it. The Laplace calculation method is shown in formula (2).

$$\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \tag{2}$$

In formula (2), it (x, y) represents the coordinates of a certain point in the image.

The detection of traffic flow and people flow in video images is to find out the moving vehicles and people from the video. The target detection algorithm is used to realize the target selection of continuous video. In the research, the background difference method is used to find the appropriate target. The background difference method is often used in the detection of moving objects such as human bodies and vehicles, which can strip the background area in the video image to obtain moving objects [18, 19]. In the application of the background difference method, the Vibe algorithm is widely used, and its calculation is divided into three steps. The first is background initialization. In the background initialization, the video is initialized through the first frame image, that is, to establish the current video. For the background model, see equation (3).

$$BK_m^0 = f^0(x^i, y^i) \mid (x^i, y^i) \in N_G(x, y) \tag{3}$$

In formula (3), it $N_G(x, y)$ represents the neighborhood point. After the background is initialized, the foreground image is detected, and the background model established by the background initialization is used to judge whether the subsequent image sequence belongs to the foreground, and the foreground part is separated. The judgment formula is shown in formula (4).

$$f^k(x, y) = \begin{cases} fore & BK > T \\ back & BK \le T \end{cases} \tag{4}$$

In formula (4), T represents the threshold set according to the experimental environment. When the value of the background model is greater than T that, it is judged as the foreground, and when it is less than, it is judged as the background. In addition, as shown in formula (4), in the selection of the foreground part, the pixel is the selection of points in random. The last step is to update the background model of the image, in order to reduce the error effect of the light intensity on the image [20, 21]. In the process of updating the background model, for any background point, it has a certain probability to update the background model, and randomly selects samples in the updated sample set. After a certain period of time, the probability that the sample value is retained can be expressed as (5).

$$P(t, t+dt) = \left(\frac{N-1}{N}\right)^{(t+dt)} \tag{5}$$

In formula (5), it N represents the number of background models; dt represents the update time; $\frac{N-1}{N}$ represents the probability of not updating at time t. However, the traditional Vibe algorithm is based on the background model. Under the premise of efficient image processing, there is still ghosting phenomenon. The reason is that the algorithm is difficult to accurately detect the slow moving stage. For this reason, in order to solve the ghosting of the Vibe algorithm, the research introduces the frame difference image of adjacent frames to optimize the effect of foreground detection, as shown in Fig. 2.

As shown in Fig. 2, for the calculation of the frame difference image of adjacent frames will judge the selected first frame image after the background initialization. When the selected image is not the first frame image, the pixel points selected multiple times are filtered out by calculating the adjacent frame difference, and they are eliminated to ensure that no background points appear in the foreground part and reduce the appearance of ghosts. In addition, in the actual traffic situation, when there is sufficient light, pedestrians and vehicles on the road will produce obvious shadows during the movement, and the existence of shadows will be judged as the foreground part during detection, resulting in the failure of foreground detection. To achieve better results, a shadow removal algorithm is proposed in this research. First, the brightness in the video image is modeled, see equation (6).

$$s_k(m,n) = E_k(m,n)\rho_k(m,n) \tag{6}$$

In formula (6), $E_k(m,n)$ represents (m,n) the radiance of the pixel point, and $\rho_k(m,n)$ represents the reflection coefficient of the vehicle or pedestrian to the light. On the basis of brightness modeling, HSV is used to eliminate shadows, and the discriminant function of shadow detection is shown in formula (7).

$$HSV(x,y) = \begin{cases} 1 & \partial \le \dfrac{I_k^V(x,y)}{B_k^V(x,y)} \le \beta \cap (I_k^S(x,y) - B_k^S(x,y)) \le T_s \cap \\ & \left| I_k^H(x,y) - B_k^H(x,y) \right| \le T_B \\ 0 & other \end{cases} \tag{7}$$

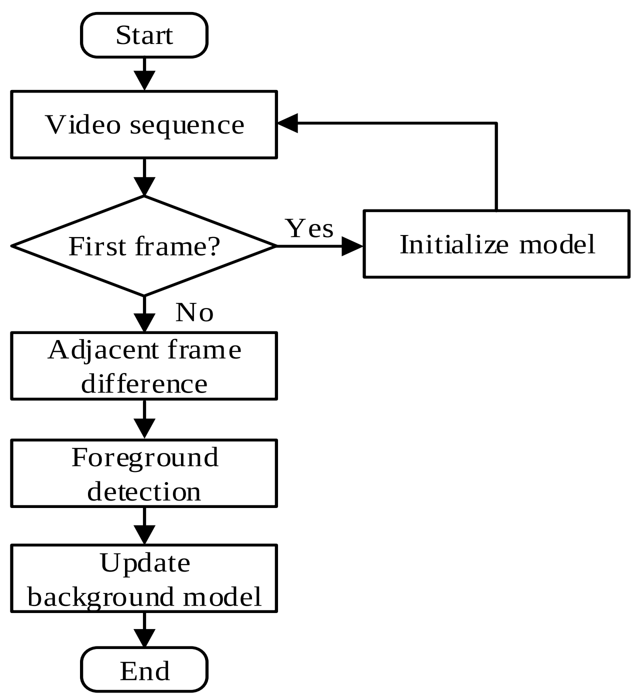


Fig. 2. Image foreground detection.

In formula (7), $I_k^V(x,y)$, $I_k^S(x,y)$, $I_k^H(x,y)$ represent the component of the kth frame image pixel in HSV space; $B_k^V(x,y)$, $B_k^S(x,y)$, $B_k^H(x,y)$ represent the component of the kth frame background pixel in HSV space; ∂, β represent the judgment parameter; T_s, T_B represent the threshold; when the discriminant function When the value is 1, the position is (x,y) defined as the shadow, and the rest are the target. The improved Vibe algorithm and shadow elimination algorithm are combined to realize the detection of traffic flow and human flow. The detection process is shown in Fig. 3.

As shown in Fig. 3, in the moving target detection, the traffic surveillance video is firstly used as the input for detection, and the HSV shadow removal algorithm is used to eliminate the ghost in the target detection. Image denoising is used to remove noise in video images after target detection. Finally, the contours of moving objects in the video images are extracted to analyze the foreground objects of vehicles and human bodies in the images.

B. Traffic Flow and People Flow Tracking and Early Warning

In actual traffic conditions, different traffic conditions have significant differences in the impact of urban traffic flow and people flow, so after solving the problem of vehicle flow and people flow detection, it is necessary to track the moving target. In moving target tracking, it is necessary to perform feature matching on each frame image in the video, that is, mark each moving target, so as to establish the correspondence between adjacent frame images [22, 23]. In the research, in order to track and identify the moving target, the motion history image is used for module tracking, and H is the pixel intensity in the motion history image. The calculation method is shown in formula (8).

$$H(x,y,t)=\begin{cases}\tau & \psi(x,y,t)=1\\ \max(0,H_\tau(x,y,t-1)-\delta) & other\end{cases} \tag{8}$$

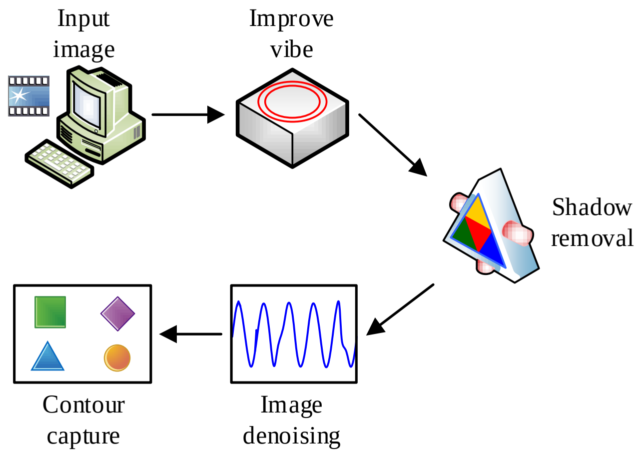


Fig. 3. Detection process of vehicle flow and passenger flow.

In formula (8), τ represents the duration between two frames in the video scene; t represents the time of the target pixel point; δ represents the decay parameter; $\psi(x,y,t)$ represents the inter-frame difference method. In addition, in the tracking process, the size of the weight of the pixel will affect the stability and reliability of the target detection, and the weight calculation is shown in formula (9).

$$k(x)C=\frac{1}{\sum k(\frac{x_i-x_0}{h})}, \quad \sum_{n=1}^{m} q_n \tag{9}$$

In formula (9), C represents the normalization coefficient; q_n represents the color distribution probability; h represents the area radius; m represents the total number of pixel points. The normalization coefficient is expressed as formula (10).

$$C=\frac{1}{\sum(\frac{y-x}{h})} \tag{10}$$

Secondly, similarity is used to measure the matching degree between moving target templates, see equation (11).

$$p=\sum_{n=1}^{m}\sqrt{q_n(y)q_n(x)} \tag{11}$$

In Equation (11), $q_n(x)$, $q_n(y)$ represent the image x axis and y the color distribution probability on the axis, respectively. According to the pixel points in the motion history image proposed by the research, the tracking of traffic flow and people flow is realized, and the method flow is shown in Fig. 4.

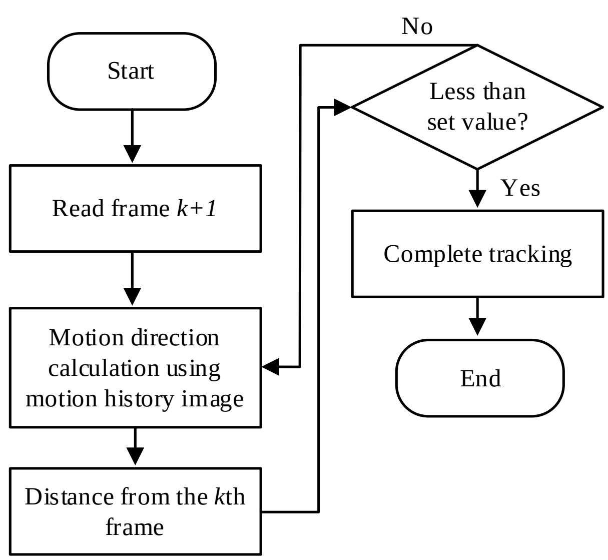


Fig. 4. Tracking process of traffic flow and people flow.

Finally, it is necessary to make a certain early warning in the monitoring of traffic flow and human flow, so as to reduce the incidence of traffic accidents. In the early warning, it is mainly aimed at identifying the distance between the age of the vehicle and the vehicle and between people. A neural network is used for distance monitoring. The distance between cars and cars and between people Z has been expressed since, and its calculation method is shown in formula (12).

$$Z = f * H / y \tag{12}$$

In formula (12), f represents the focal length of the monitoring equipment; H represents the distance from the ground to the particle; y represents the distance from the horizontal contact point of the previous moving target to the vertical distance of the video image. To set parameters for the rear vehicle under monitoring, the distance formula can be updated to formula (13).

$$Z = f * H / y = F * H / Y \tag{13}$$

Formula (13), F represents the equivalent vertical focal length of the rear vehicle in the video image; Y represents the number of pixels at the bottom of the front vehicle and the key position of the image. Through traffic monitoring, it is judged whether there is a possibility of collision between two vehicles, and the time of the vehicles that may collide is estimated. Equation (14) is used to calculate the transformation of vehicles in the image.

$$h = Z_0 / Z_t \tag{14}$$

Equation (14), Z_0 the position of the vehicle detected at the initial time in the image is Z_t expressed, and the position of the vehicle detected at time *t in the image is expressed.* Considering the speed difference between v vehicles, calculating the collision time between vehicles is shown in Equation (15).

$$G = \frac{Z_0 * a}{v_0^2} \tag{15}$$

In formula (15), a represents the braking parameter. Through the motion detection of vehicle information and people flow information in traffic monitoring video images, the monitoring of traffic flow and people flow and accident warning are realized.

IV. Algorithm Performance and Application Test

A. Algorithm Performance Analysis

In the research, the performance of the proposed monitoring video image target detection algorithm for traffic flow and people flow is analyzed. The experiment is set up in Intel Core i5-5300, its main frequency is 2.3GHz, the memory is 4GB, and the operation is performed in Windows7. The data set used in the research is the road surveillance video data of a certain urban area on sunny days, cloudy and rainy days and at night. The data is divided into three data sets, which are 70% training set, 20% test set and 10% application set. , select the training set and test set for calculation training. First, evaluate the loss value of the algorithm through comparative analysis, as shown in Fig. 5.

As can be seen from Fig. 5, the research compares the detection loss rate of the improved Vibe algorithm with the traditional convolutional neural network and A* algorithm. As shown in Fig. 5(a), in the training set, the loss rate of the traditional convolutional neural network during the training process shows a decreasing trend, and after the number of iterations reaches 50, it decreases to near the minimum value. The minimum value of the loss rate is 0.43%. In the training process of the A* algorithm, when the number of iterations reaches 70, the changes of the algorithm begin to gradually stabilize, and finally the loss rate is stabilized at 0.51%. However, the improved Vibe algorithm proposed by the study begins to stabilize after the number of iterations reaches 30, and its loss rate is reduced to 0.25%, which is significantly lower than the other two algorithms. Fig. 5(b) is the test effect in the test set, where the three algorithm changes are consistent with the training set. Secondly, the missed detection rate and the false negative rate index of the three algorithms are analyzed to judge the accuracy of different algorithms in target recognition, as shown in Fig. 6.

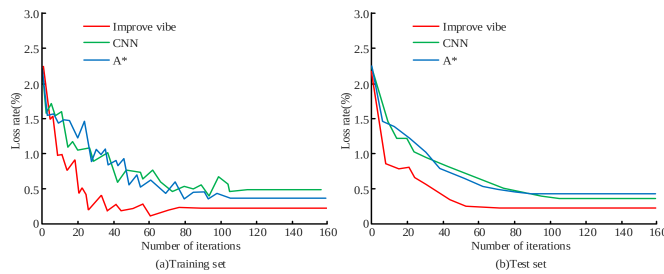


Fig. 5. Training and testing of loss rate.

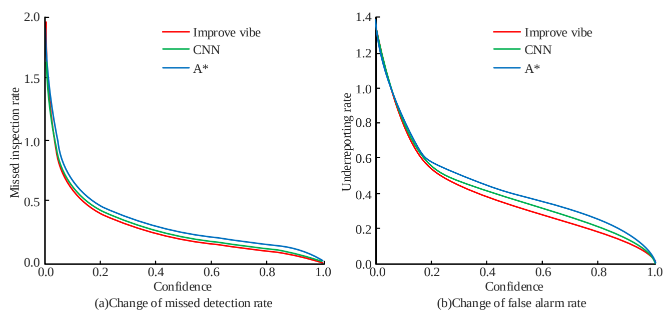


Fig. 6. Test results of missed detection rate and missed report rate.

As can be seen from Fig. 6, with the continuous increase of the confidence, the missed detection rate and the false negative rate of the three algorithms show a decreasing trend. It can be clearly seen that during the training process of the improved Vibe algorithm proposed by the study, the missed detection rate and the false negative rate indicators show a decreasing trend, and show a rapid decline before the confidence level of 0.2, and then gradually show downward trend. It can be seen from the changes of the missed detection rate and the false negative rate indicators of the improved Vibe algorithm that compared with the changes of the FPPI indicators of the traditional convolutional neural network and the A* algorithm, the improved Vibe algorithm can reduce it to a lower level, and its reduction speed is also significantly higher than the rest of the object detection algorithms. After that, the precision and recall rates of all algorithms in target detection training are analyzed, and the PR curve is used to describe them, as shown in Fig. 7.

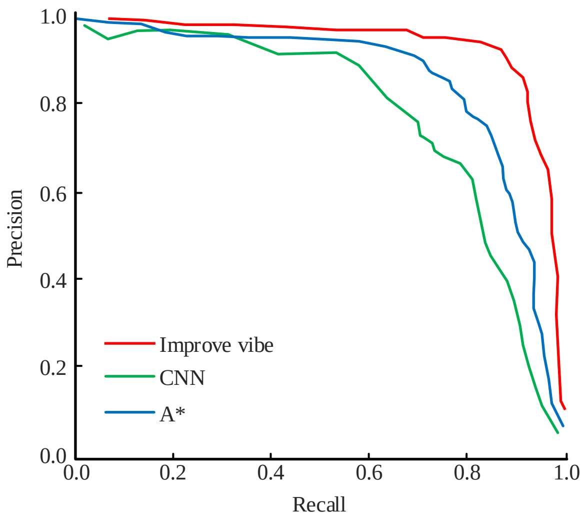


Fig. 7. PR curve comparison.

As can be seen from Fig. 7, in the change of the PR curve, the traditional convolutional neural network is closer to the lower left corner, indicating that the traditional convolutional neural network has a lower detection accuracy than the A* algorithm and the improved Vibe algorithm. . In addition, the precision rate and recall rate of the A* algorithm are in the middle position, and the calculated precision rate and recall rate are 87.62% and 89.17%, respectively. The improved Vibe algorithm proposed in the study is closer to the upper right corner, and its precision and recall rates are 91.25% and 93.33%, respectively. Comparing the PR curves of the three algorithms, it can be seen that the improved Vibe algorithm proposed in the study has higher precision and recall rates, and the target recognition and detection effects reflected in data training are more significant.

B. Application Test of Target Detection Algorithm

The monitoring video traffic flow and people flow detection algorithm proposed by the research is applied and practiced, and 10% of the collected traffic data in an urban area is used as a sample for application testing. Firstly, the detection results of traffic flow and people flow under different algorithm models are analyzed, and the real rate is used as the detection index, as shown in Fig. 8.

From Fig. 8, in the detection of traffic condition data, the true rate of the detection model proposed in the study can finally reach 0.98, and it can be known from the changes of its true rate and false positive rate that its highest false positive rate is only is 0.06. From the change of the true rate of the detection model under the convolutional neural network, it can be seen that its true rate is only 0.96 at the highest, and its false positive rate reaches 0.11. In addition, the application of the A* algorithm detection model shows that the changes of its true rate and false positive rate are significantly different from those of the detection model proposed in the study and the convolutional neural network detection model. The highest value of the true rate is only 0.85, while the false positive rate reached 0.26. The reason for the analysis is that the detection performance of the A* algorithm in the detection of traffic data is interfered by many factors, which leads to the suppression of its true rate. Secondly, the detection time required by different models in the detection of traffic flow and people flow is analyzed, as shown in Fig. 9.

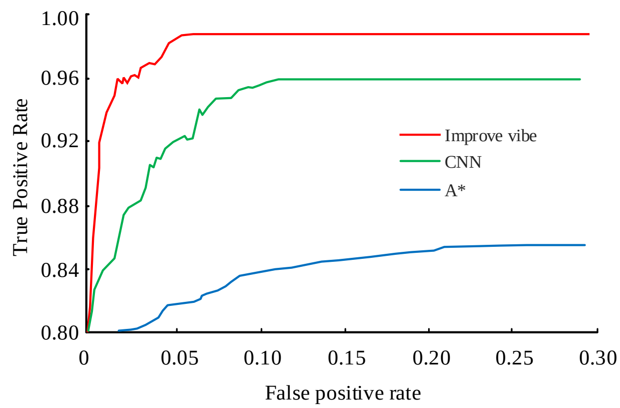


Fig. 8. Real rate test results.

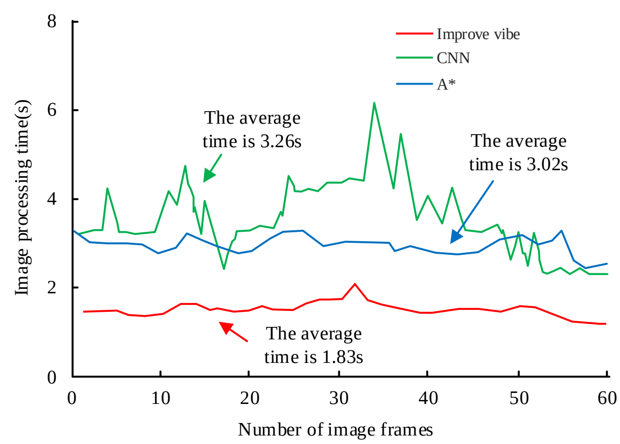
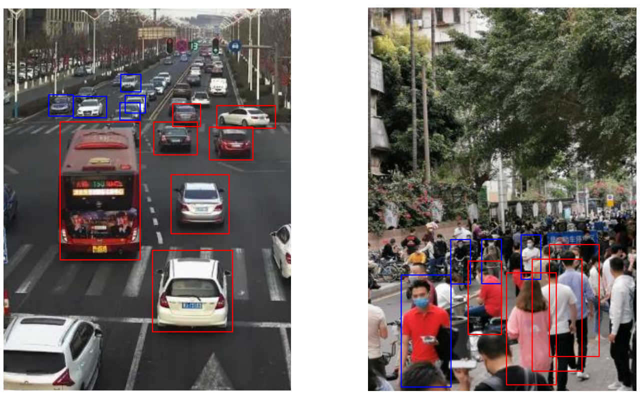


Fig. 9. Comparison of detection time.

It can be seen from Fig. 9 that the average detection time of multiple models in the detection of traffic flow and people flow is within 5s. Among them, the detection time fluctuation of the traditional convolutional neural network detection model is the most significant, and its average time is 3.26s. From the detection time change of this model, it can be seen that with the continuous increase of the number of image frames, its fluctuation frequency continues to increase. The application of the A* algorithm detection model shows that the average detection time of this model is 3.02s, which is a certain reduction compared to the convolutional neural network detection model, and the model has less fluctuation in detection , which is not significantly changed by changes in the number of image frames. Finally, it can be seen from the application results of the detection model proposed in the study that the average detection time of this model is only 1.83s, and with the continuous increase of the number of image frames, the model detection time has no significant change, which is more efficient than the A* detection model for gentle changes. Then, in order to fully understand the application effect of the human flow and traffic flow detection model proposed by the research, we conduct in-depth exploration by analyzing the renderings of its traffic flow and human flow, as shown in Fig. 10.

As can be seen from Fig. 10, the vehicle flow and people flow detection model proposed by the study can count the flow of vehicles and people, and can distinguish between vehicles and people in detail. At the end of the study, a vehicle collision warning model based on convolutional neural network is proposed, which aims to judge the accident rate during the calculation of traffic flow, and use the receiver operating characteristic curve (ROC) curve to evaluate its warning effect, as shown in Fig. 11.



(a)Traffic flow detection (b)Pedestrian flow detection

Fig. 10. Detection effect of traffic flow and pedestrian flow.

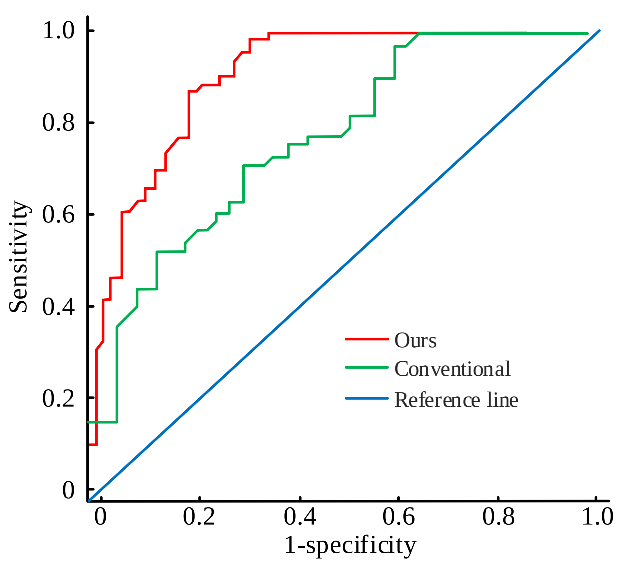


Fig. 11. ROC curve comparison.

As can be seen from Fig. 11, the area under the curve of the early warning model proposed by the study is 0.802, which is significantly higher than the standard value of 0.5, and compared with the traditional early warning model, the area under the curve increases by 0.038. The above results show that the proposed accident early warning model can effectively predict the upcoming traffic accident and provide early warning, and is more efficient than the traditional early warning scheme.

V. Conclusion

Urban development makes urban traffic show a trend of continuous development. Under this trend, how to monitor and monitor the traffic flow and people flow in the city is very important. In the research, a detection model based on Vibe algorithm is proposed for the traffic flow and people flow information in the traffic monitoring video images, and an algorithm for eliminating vehicle shadows is proposed for the traffic conditions in the video images. Finally, the motion history images are used to detect motion. The target is tracked. The algorithm performance simulation shows that the minimum loss rate of the improved Vibe algorithm proposed by the study reaches 0.25%, and the missed detection rate and the false negative rate are also low. Finally, the application test shows that the detection time of the detection model proposed in the study is only 1.83s, and it can count and distinguish the traffic flow and people flow in the traffic situation. In addition, the ROC of the detection model proposed in the study is in traffic accident warning. The offline area reached 0.802. The above results show that the improved Vibe algorithm is effective in detecting traffic flow and people flow in traffic videos, and the lower detection time indicates that it can realize real-time monitoring of traffic vehicles and people flow. However, the improved Vibe algorithm used in the study does not consider the image quality in different weather conditions. Therefore, in the follow-up work, it is necessary to improve the algorithm's ability to deal with complex environments to ensure that the vehicle flow and pedestrian flow can be monitored in various weather conditions

Acknowledgment

This study was supported by Guiding project of Hubei Provincial Department of Education Science Research Plan in 2022 (B2022397).

References

[1] R. Salazar-Cabrera, L. Cruz, and J. Molina, "Design of a Public Vehicle Tracking Service Using Long-Range (LoRa) and Intelligent Transportation System Architecture," J. Inf. Technol. Res., vol. 14, pp. 147-166, January 2021.

[2] D. M. Manias, and A. Shami, "Making a Case for Federated Learning in the Internet of Vehicles and Intelligent Transportation Systems," IEEE Network, vol. 35, pp. 88-94, February 2021.

[3] Y. Q. Xia, C. Yan, X. J. Wang, and X. H. Song, "Intelligent Transportation Cyber-physical Cloud Control Systems," Acta Autom. Sin., vol. 45, pp. 132-142, January 2019.

[4] S. Syedyusuff, R. Subramaniam, and R. Vijay, "Orthogonally Integrated Hybrid Antenna for Intelligent Transportation Systems," Appl. Comput. Electrom., vol. 36, pp. 519-525, May 2021.

[5] S. Pan, P. Li, C. Yi, D. Zeng, Y. C. Liang, and G. Hu, "Edge Intelligence Empowered Urban Traffic Monitoring: A Network Tomography Perspective," IEEE T. Intell. Transp., vol. 22, pp. 2198-2211, April 2021.

[6] M. Rogowski, "Monitoring System of tourist traffic (MSTT) for tourists monitoring in mid-mountain national park, SW Poland," J. Mt. Sci., vol. 17, pp. 2035-2047, August 2020.

[7] L. Zhu, C. Zhang, C. Xu, X. Du, N. Guizani, and K. Sharif, "Traffic Monitoring in Self-Organizing VANETs: A Privacy-Preserving Mechanism for Speed Collection and Analysis," IEEE Wirel. Commun., vol. 26, pp. 18-23, December 2019.

[8] X. Cai, L. Quan, J. Wu, and Y. He, "Night-time disability glare of constant-light LED traffic monitoring fill light:," Lighting Res. Technol., vol. 53, pp. 595-608, January 2021.

[9] T. D. Almeida, J. Nacif, F. P. Bhering, and J. G. R. Júnior, "DOCTraMS: A Decentralized and Offline Community-Based Traffic Monitoring System," IEEE T. Intell. Transp., vol. 20, pp. 1160-1169, May 2019.

[10] N. Obukhova, A. Motyko, and A. Pozdeev, "Personalized approach to developing image processing and analysis methods for medical video systems," Proc. Comput. Sci., vol. 176, pp. 2030-2039, January 2020.

[11] R. Ke, S. Feng, Z. Cui, and Y. Wang, "Advanced framework for microscopic and lane-level macroscopic traffic parameters estimation from UAV video," IET Intell. Transp. Sy., vol. 14, pp. 724-734, July 2020.

[12] Y. Zhang, and Yang B, "Traffic Flow Detection Using Thermopile Array Sensor," IEEE Sens. J., vol. 20, pp. 5155-5164, January 2020.

[13] M. J. Osman, J. Rashid, O. K. Khim, W. M. Z. W. Yunus, A. Noor, N. A. M. Kasim, V. Feizal, and T. C. Chuang, "Optimisation of a gold nanoparticle-based aptasensor integrated with image processing for the colorimetric detection of acephate using response surface methodology," RSC Adv., vol. 11, pp. 25933-25942, July 2021.

[14] S. Sevin, H. Tutun, and S. Mutlu, "Detection of Varroa mites from honey bee hives by smart technology Var-Gor: a hive monitoring and image processing device," Turk. J. Vet. Anim. Sci., vol. 45, pp. 487-491, June 2021.

[15] S. Mittal, "Secure V2V and V2I Technologies for the Next-Generation Intelligent Transportation Systems," Computer, vol. 54, pp. 4-6, February 2021.

[16] O. Gurbuz, and R. L. Cheu, "Survey to Explore Behavior, Intelligent Transportation Systems Needs, and Level of Service Expectations for Student Parking at a University Campus:," Transport. Res. Rec., vol. 2674, pp. 168-177, January 2021.

[17] C. R. Dr, "Context—Aware Smart Reliable Service Model for Intelligent Transportation System Based on Ontology," Lect. Notes Electr. Eng., vol. 597, pp. 23-30, January 2020.

[18] A. B. Khalifa, I. Alouani, M. A. Mahjoub, and A. Menhaj, "A novel multi-view pedestrian detection database for collaborative Intelligent Transportation Systems," Future Gener. Comp. Sy., vol. 113, pp. 506-527, December 2020.

[19] M. Boucher, "Transportation Electrification and Managing Traffic Congestion: The role of intelligent transportation systems," IEEE Electrif. Mag., vol. 7, pp. 16-22, September 2019.

[20] H. Zheng, W. Chang, and J. Wu, "Traffic flow monitoring systems in smart cities: Coverage and distinguishability among vehicles," J. Parallel Distr. Com., vol. 127, pp. 224-237, August 2019.

[21] A. H. Sodhro, M. S. Obaidat, Q. H. Abbasi, P. Pasquale, S. Pirbhulal, A. Yasar et al., "Quality of Service Optimization in an IoT-Driven Intelligent Transportation System," IEEE Wirel. Commun., vol. 26, pp. 10-17, December 2019.

[22] B. Liu, X. Tang, J. Cheng, and P. Shi, "Traffic flow combination forecasting method based on improved LSTM and ARIMA," Int. J. Embed. Syst., vol. 12, pp. 22-30, June 2020.

[23] Y. Liu, J. Q. James, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," IEEE Internet Things, vol. 7, pp. 7751-7763, March 2020.

Towards an Accurate Breast Cancer Classification Model based on Ensemble Learning

Aya Hesham[1], Nora El-Rashidy[2], Amira Rezk[3], Noha A. Hikal[4]

Information Technology Department-Faculty of Computers and Information, Mansoura University, Mansoura, 13518, Egypt[1, 4]
Machine Learning and Information Retrieval Department-Faculty of Artificial Intelligence,
Kafrelsheiksh University, Kafrelsheiksh,13518, Egypt[2]
Information Systems Department-Faculty of Computers and Information, Mansoura University, Mansoura, 13518, Egypt[3]

***Abstract*—Breast cancer (BC) is considered the most common cancer among women and the major reason for the increased death rate. This condition begins in breast cells and may spread to the rest of the body tissues. The early detection and prediction of BC can help in saving a patient's life. In the last decades, machine learning (ML) has played a significant role in the development of models that can be used to detect and predict various diseases at an early stage, which can greatly increase the survival rate of patients. The importance of ML Classification is attributed to its capability to learn from previous datasets, detects patterns that are difficult to comprehend in massive datasets, predicts a categorical variable within a predefined example and provide accurate results within a short amount of time. Feature selection (FS) method was used to reduce the data dimensionality and choose the optimal feature set. In this paper, we proposed a stacking ensemble model that can differentiate between malignant and benign BC cells. A total of 25 different experiments have been conducted using several classifiers, including logistic regression (LR), decision tree (DT), linear discriminant analysis (LDA), K-nearest neighbor (KNN), naive Bayes (NB), and support vector machine (SVM). In addition to several ensembles, the classifiers included random forest (RF), bagging, AdaBoost, voting, and stacking. The results indicate that our ensemble model outperformed other state-of-the-art models in terms of accuracy (98.6%), precision (89.7%), recall, and F1 score (93.33%). The result shows that the ensemble methods with FS have a high improvement of classification accuracy rather than a single method in detecting BC accurately.**

Keywords—Breast cancer; feature selection; classification; machine learning

I. Introduction

Breast cancer (BC) is the major cause of cancer death among women worldwide. Most recent reports showed a significant increase in patients with BC [1]. Early detection of BC is crucial to increasing the survival rates, lowering the risks of treatments, and decreasing mortality. Cancers are uncontrolled cell divisions that can infect other tissues. BC usually starts in ducts that lead milk to the nipple. Symptoms of BC include changes in the size of one or both breasts, dimpling of breast skin, and a lump in one of the armpits. Breast tumors are classified into two types: benign and malignant. Benign tumors are non-cancerous because they are not diffused into the tissue of the breast surrounding the duct and do not cause harm on one's life. Malignant tumors are cancerous, and their cells spread to the surrounding breast tissue through the duct lining, posing a hazard to one's life.

The traditional approach for cancer diagnosis mainly depends on the experience of medical experts and their visual inspections. However, this type of diagnosis consumes long hours and may be susceptible to human errors[2]. Therefore, an automatic diagnosis system must be developed for the early detection of cancer. In this regard, several studies used machine learning (ML) techniques to develop classification and prediction models for medical diagnosis. ML is a type of artificial intelligence that allows a machine to learn by supplying a collection of facts and gaining knowledge through experience rather than through extensive programming [3].

Classification and prediction are the main techniques used to analyze data in ML. Classification refers to the predictive process of class objects whose label is unknown based on the training samples and predicts a categorical variable within a predefined example. Prediction is the process of determining numerical data that are missing or unavailable for a new observation based on previous information.

Different algorithms for classification, such as (KNN), (SVM), artificial neural networks (ANNs), and (RF) have been applied in several studies to classify BC.

Harikumar et al. [4] used DT and KNN algorithms in the classification of BC. Principal component analysis (PCA) technique was used to select the optimal feature set.

Hiba et al. [5]presented a comparison between the performance of four classifiers: SVM, NB, C4.5, and KNN. These studies mainly aimed to determine whether the classification of data is proper in regard to the efficiency and effectiveness of each algorithm.

Marie et al. [6] reviewed ML models that may predict BC in women and compared their performances. A total of 116 participants were measured for various parameters in the Coimbra dataset: insulin, glucose, resistin, adiponectin, homeostasis model assessment (HOMA), leptin, and monocyte chemoattractant protein-1. They implemented classification algorithms that included logistic regression LR, KNN, SVM, DT, RF, gradient boosting method (GBM), and NB.

Tsehay [7] proposed the KNN model for BC prediction using a grid search approach to find the best hyper-parameter. Then, a comparison was made between the default and tuned hyper-parameters. The results demonstrated that the performance significantly increased when the best parameter or K value was utilized to train the KNN using the Wisconsin Breast Cancer dataset.

Salehi et al. [8] proposed an ensemble machine for survivability prediction. The main goal was to lower the feature set's size by deleting the more complex features. This machine depended on three basic learners, namely, multi-layer perceptron (MLP), SVM, and DT, based on 25 simple features. The ensemble multi-stage machine was based on a performance investigation using a 10-fold cross-validation (CV) method.

In this paper, we utilized a stacking ensemble model to detect BC. Ensemble learning refers to the training of a set of models instead of a single model and the combination of results using weighted or unweighted techniques [9]. The aim of combining several models is to achieve an improved predictive performance than any single-component model. The solidity of the ensemble limits the dispersion of single-model prediction. Thus, this type of learning contributes to reduction of variance and the generalization error of these models.

The following problems were considered in this article:

Various studies focused on the use of single ML algorithms in BC classification. These studies did not concentrate on the difference in the algorithms' performance in terms of the various data categories in the dataset (i.e., text and numeric). Several algorithms, such as SVM, perform better with text data than numeric data. Other algorithms, such as DT, perform better with text data. Thus, we suggested the ensemble model to aggregate the benefits of different algorithms.

The main contributions of this dataset are summarized in the following points:

We proposed medically intuitive and accurate ML models for BC prediction based on a medical text dataset, and they can be detailed in the following points:

- Survey of the most recent studies in BC classification and prediction;
- We explore the shortcoming of the previous literature and draw the guidelines for future improvements that could overcome these challenges;
- Proposal of a framework for BC classification based on textural features relating to tumor description;
- Framework model is created to solve a binary classification (0 = benign and 1 = malignant);
- Framework included the study of the effect of using FS and ensemble ML models on improving the quality and efficient for Bc classification;
- Comprehensive analysis of popular ML models (single models), such as NB, KNN, linear discriminant analysis (LDA), LR, DT, and support vector classifier (SVC), and ensemble classifiers, such as RF, bagging, AdaBoost, stacking, and voting classifiers;
- Use of feature selection (FS) to reduce data dimensionality and select the optimal feature set;
- Use of various metrics, including accuracy (Acc), specificity (Sp), precision (P), recall (R), F1 score (F1), and area under the receiver operating characteristic (ROC) (AUC) curve, to evaluate our model performance;
- Investigation of the capability of ensemble learning to perform binary classification tasks after the application of data to pipeline preprocessing steps;
- The results showing that the best Acc was achieved by the ensemble model with FS of 98.6%), which can adapt to several types of classification of medical data.

The rest of the paper is organized as follows. Section II discusses the literature review. Section III elaborates the methods and materials. Section IV explains the proposed framework. Section V presents the results. Section VI presents the discussion. Comparing with other works presented in Section VII. Finally, Section VIII concludes the paper.

II. Related Work

Numerous researchers are interested in BC prediction using ML. The related previous works can be divided into two groups: those that used single ML and those that applied ensemble learning. In the following, the two groups will be discussed.

A. *The First Group Included Studies that Utilized Single ML*

Abdulrahman et al. [10] compared six ML techniques. Their experiment concluded that SVM and RF were the most effective ML classifiers, with an Acc of 97.3% for BC prediction.

R. Karthiga1 et al. [11] proposed a novel strategy of BC diagnosis, which was collected from a visual laboratory, using image analysis and ML techniques. First, they applied image quality enhancement and a region of interest (ROI) selection using the morphological transform process. Then, hot regions, including armpits and neck, were segmented. The image's structural flaws were removed using morphological procedures. The statistical, intensity, and geometric features were extracted from the images of enhanced ROI. Curvelet transform was implemented using the wrapping process. The matrix of gray-level co-occurrence was used to extract textural features from the curvelet coefficients. Various ML techniques were investigated, and the cubic SVM showed the best Acc (93.3%) instead of the limited Acc for SVM.

Shiny et al. [12] proposed a new discriminate ratio for SVM. They compared three of the most common ML algorithms. RF, SVM, and NB have been applied in the original WDBC. The results have shown that SVM exhibited the best performance and an Acc of 97.2%.

Jiande et al. [1] also used the SVM algorithm in addition to KNN, NB, and DT. They utilized gene expression data to classify triple- and non-triple negative BC patients using a ML approach. The implementation of different algorithms indicated that SVM had the highest performance with an Acc of 90%.

Mücahid et al. [13] suggested a model for the detection of BC using ANN and NB based on age, biomarkers, glucose, body mass index, HOMA, resistin, leptin, adiponectin, and insulin. According to the results, the markers can be used to obtain desirable findings from the classification algorithms; BC was classified with high accuracies of 86.95% and 83.54% with the use of ANN and NB algorithms, respectively.

Milon et al. [14] increased the Acc of the ANN classifier and compared five ML techniques, namely, SVM, KNN, RF, ANN, and LR, for BC prediction. The Acc, sensitivity, Sp, P, negative predictive value, false negative (FN) rate, false positive (FP) rate, F1, and Matthews correlation coefficient were all used to evaluate the study's performance. The best Acc achieved was by the ANN (98.57%).

Akizur et al.[15] introduced FS as a promising solution for maximizing Acc. FS was focused on to reduce the dimensionality of data and obtain a high level of BC classification Accs The proposed system was based on two stages. First, FS techniques selected 22 features from the WDBC dataset. For BC classification, the ANN with a 15-neuron (15-neuron ANN) classifier obtained an Acc of 96.4%, 99.9% sensitivity, 98.4% P, 1.6% FP rate, and 0.42 s processing time. The proposed 15-neuron ANN classifier achieved a 98.8% classification Acc in the initial experiment, which was performed without FS. The Acc of classification was improved with a median of 0.6% after the application of FS.

Sapiah et al. [16] also focused on improving the Acc. ML algorithms were used along with particle swarm optimization (PSO) for FS. The results indicated that using PSO achieved better performance than their counterparts in terms of Acc. With the use of PSO, NB, KNN, and reduced error pruning (REP) tree obtained accuracies of 81.3%, 66.3%, and 80%, respectively. Without PSO, the accuracies reached 70%, 75.5%, and 76.3%. Thus, the use of PSO increased the Acc of ML algorithms.

B. *The Second Group Included Studies that Concerned Ensemble ML*

Amal et al. [17] proposed a BC detection method that relies on a bagging ensemble classifier. They used One Rule (OneR) to extract the beneficial features of BC. The results of this method showed that the bagging classifier achieved the best Acc of 92.25% over other single ML classifiers, such as KNN, MLP, and SVM.

Royida et al. [18] proposed a new strategy for enhancing the Acc of BC dataset classification. Hoeffding trees were used for normal classification, whereas NB was applied to reduce data dimensionality. The results of the proposed model have shown an increase in the Acc to 95.9943%.

Meerja et al. [19] proposed an ensemble model for data BC classification using Bayesian network and radial basis (RB) function, which provided a good classification Acc. Various measurements were used to evaluate the model's performance. The results have shown that the proposed ensemble method recorded an Acc of 97.42%.

Tina et al. [20] proposed two methods: stacking -voting model and stacking ensemble models with Bayesian methods as base classifiers. In stacking, the model was used with three Bayesian methods, such as NB, hidden NB, and Bayesian network. Six models were created, and their performances were compared. In voting and stacking models, the classifiers were combined by the voting classifier, and three Bayesian methods were used individually with sequential minimal optimization (SMO) as the base classifier and a voting ensemble of REP Tree and RF as meta classifiers. The results revealed that the method using stacking alone and an ensemble of Bayesian methods exhibited the best performance, and NB combinations with LR and SMO gave the best prediction Acc of 97.8%.

Hajar et al. [21] improved the performance of ML techniques to test their classification performances in large BC databases, such as the Surveillance, Epidemiology, and End Results (SEER) database. The KNN, NB, LR, and MLP techniques exhibited low Acc, contrary to the RF and DT, which proved their high performance. Ensemble approaches, such as voting, stacking, bagging, and boosting, were used to increase the performance of single algorithms. Voting techniques provided the greatest improvement and Acc (99.99%) with NB, DT, and RF. For more information about the use of ML models in BC classification, readers are guided to the following surveys [22],[23].

Previous works presented several shortcomings. (1) Several researchers used an insufficient number of features, (2) most of the researchers disregarded FS as the primary algorithm for BC diagnosis, and (3) others focused on the same algorithm in the classification although most of the different types of data require the use of various types of algorithms to be classified. Most algorithms work best on text data and others on numeric data. Thus, combining two types of data using the same algorithm affects the classification performance. Therefore, we suggested the use of ensemble learning and FS to overcome these limitations and improve the Acc and performance for BC classification**.** Table I shows the characteristic of the reviewed studies.

TABLE I. MAIN CHARACTERISTICS OF THE REVIEWED STUDY

Ref.	Year	Method	Accuracy	Application	Dataset	Data used	Attributes
[10]	2022	SVM	97.3%	Prediction	569 instances from WDBC	Text	10 attributes
		NB	95.7%				
		ANN	95.7%				
		RF	97.3%				

Ref.	Year	Method	Accuracy	Application	Dataset	Data used	Attributes
		KNN	93.8%				
		DT	93%				
[1]	2021	KNN	87%	Classification	1222 samples (110 triple negative 992 non-triple negative; RNA-Seq Data [CrossRef] [PubMed]	Text	30 genes
		NGB	85%				
		DT	87%				
		SVM	90%				
[17]	2021	NLP + ML: J48	88.73%	Prediction	142 case patients from King Abdullah University Hospital in Jordan. Luminal A, 15%; luminal B, 46%; human epidermal growth factor receptor 2 positive, 24%; triple-negative BC, 15%	Text	13 attributes
		NB	89.44%				
		Bagging	92.25%				
		Logistic	85.92%				
		SVM	90.14%				
		KNN	89.44%				
		MLP	90.85%				
		PART	89.44%				
		OneR	90.14%				
[21]	2021	Voting ensemble methods (NB, DT, and RF)	99.99%	Classification	699412 cases (81386 in 2016 and 618026 in 2008–2015) from SEER; the National Cancer Institute from 2008 to 2016	Text	-15 attributes from 2008 to 2015 -11 attributes for 2016
[11]	2021	Curvelet transform - cubic SVM	93.3%	Classification	60 (30 normal, 30 abnormal) from visual laboratory, Brazil, Fluminense Federal University	Image	16 attributes
[19]	2021	Ensemble classifier (Bayesian network and RB function)	97%	Classification	699 instances, WBCD https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic).	Text	10 attributes
[15]	2020	(15-neuron network) with FS	96.4%	Classification	569 samples (357 normal, 212 cancer) from University of California Irvine (UCI) Learning Repository dataset for diagnosis of breast cancer in Wisconsin http:/archive.ics.uci.edu/ml/index.php	Text	32 attributes
[14]	2020	SVM	97.14%	Prediction	699 instances: 458 benign, 241 malignant (UCI) https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cance rwisconsin/breast-cancer-wisconsin.data.	Text	9 attributes
		K NN	97.14%				
		RF	95.71%				
		ANN	98.57%				
		LR	95.71%				
[20]	2020	Stacking ensemble methods (NB with LR and SMO)	97.8%	Prediction	699 instances (16 missing values, 444 benign, and 239 malignant cancers); Wisconsin Breast Cancer Database (WBC) available from UCI between 1989 and 1991	Text	10 attributes
[12]	2019	SVM	97.2%	Prediction	458 benign and 241 malignant (UCI ML Repository)	Text	10 attributes
		RF	96.6%				
		NB	97%				
[13]	2019	ANN	86.95%	Classification	64 patients with breast cancer and 52 healthy controls (UCI ML Repository) http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra	Text	10 attributes
		NB	83.54%				
[18]	2019	Ensemble	95.9943%	Classification	699 instances (444 benign, 239 malignant, and 16	Text	9

Ref.	Year	Method	Accuracy		Application	Dataset	Data used	Attributes
		hoffeding tree and NB				missing) WBC		attributes
[16]	2018		With out (PSO)	With (PSO)	Prediction	198 instances (UCI from the Wisconsin Prognostic Breast Cancer) https://archive.ics.uci.edu/ml/datasets/Breast±Cancer± Wisconsin± (Prognostic	Text	34 attributes
		KNN	66.3 %	75.5 %				
		NB	70%	81.3 %				
		REP	76.3 %	80%				

III. Methods and Materials

A. *Data Set*

In this section, we used the Wisconsin Diagnostic Breast Cancer (WDBC) obtained from UCI ML Repository [24]. This data set was used to detect malignant and benign tumors. All features in the WDBC were extracted from fine-needle aspirate images, which describe the features of the nucleus. The WDBC included the features of 569 patients (357 benign and 212 malignant cases) and 32 attributes in Wisconsin hospitals. Each attribute represents a measurement result. The first and second attributes refer to the identifier number and patient diagnosis status, respectively. The remaining attributes correspond to standard error, mean, and the least of red nucleus features. Table II provides details of the dataset attributes and their description.

TABLE II. Dataset Attributes and their Description

NO	Attribute	Description	Domain value
1	**Id**	Alphanumeric code is used in a health record system to uniquely identify a patient.	8670 to 912 m
2	**Diagnosis**	Process of identifying the condition of the disease	0 or 1
3	**Radius_mean**	Average of distances between the center and points around the perimeter	6.99 to 28.2
4	**Texture_mean**	Gray-scale value standard deviation	9.73 to 39.4
5	**Perimeter_mean**	Average tumor size	44 to 189
6	**area_mean**	Mean area inside the boundary of core tumor	144.1 to 2500
7	**smoothness_mean**	Local variation's mean in the radius lengths for the measurement of the smoothness of breast cells	0.06 to 0.17
8	**compactness_mean**	Calculated using the equation (perimeter^2 / area - 1.0)	0.03 to 0.35
9	**concavity_mean**	Mean of the degree to which the cell contour is concave	0 to 0.44
10	**Concave points _mean**	Mean of the quantity of concave contour parts	0 to 0.4
11	**symmetry _mean**	Mean of similar area of tumor parts that match	0.11 to 0.3
12	**Fractal_ dimension _mean**	Mean for "coastline approximation"-1	0.05 to 0.2
13	**radius_se**	Error of standard for the average length from the center to the perimeter points	0.11 to 2.88
14	**texture_se**	Error of standard for the grayscale values' standard deviation	0.38 to 4.88
15	**perimeter_se**	Error of standard for the tumor perimeter	0.77 to 22
16	**area_se**	Error of standard for the mean area inside the boundary of the core tumor	6.9 to 543
17	**smoothness_se**	Error of standard for variations in radius lengths on a local scale	0 to 0.03
18	**compactness_se**	Error of standard for perimeter^2/area-1.0	0 to 0.15
19	**concavity_se**	Error of standard for the degree to which the cell contour is concave	0 to 0.5
20	**concave points_se**	Error of standard for the quantity of concave contour parts	0 to 0.06
21	**symmetry_se**	Error of standard for the mean of similar area of matching tumor parts	0.01 to 0.08
22	**fractal_ dimension_se**	Error of standard for "coastline approximation"	0 to 0.04
23	**radius_worst**	Biggest or worst mean value for the average distance between the center and points on the perimeter	7.94 to 36
24	**texture_worst**	Biggest or worst mean value for gray-scale value standard deviation	12.1 to 49.6
25	**perimeter_worst**	Biggest or worst mean value size of the core tumor	50.6 to 251
26	**area_worst**	Biggest or worst mean value for the mean area inside the boundary of the core tumor	185 to 4255
27	**smoothness_worst**	Biggest or worst mean value for local variations of radius lengths	0.08 to 0.23
28	**compactness_worst**	Biggest or worst mean value for (perimeter^2 / area - 1.0)	0.03 to 1.07
29	**concavity_worst**	Biggest or worst mean value for the degree to which the cell contour is concave	0 to 1.27
30	**concave points_worst**	Biggest or worst mean value for the quantity of concave contour parts	0 to 0.30
31	**symmetry_worst**	Biggest or worst mean value for a similar area of matching tumor parts	0.17 to 0.77
32	**fractal_dimension_ worst**	Biggest or worst mean value for "coastline approximation"-1	0.07 to 0.22

B. Machine Learning

1) Single machine learning: In this paper, we used popular classification models to test the capability of data features in the classification of BC; these models included SVM [25]–[28], LDA [29], [30], [31] KNN [32]–[35], LR [35]–[38], DT[39], [40], and NB. The selection of classifiers was based on two main reasons: (1) popularity in medical domains (2) and their diversity in terms of classification structure. Table III shows the selected classifiers and their types.

TABLE III. CLASSIFICATION MODELS

Classifier Type	Classifier Name	Label
Tree based	Decision Tree	DT
Linear-based model	Logistic regression	LR
	Support Vector classifier	SVC
Nonparametric-based model	K-Nearest Neighbor	KNN
Statistical based model	Linear discriminant analysis	LDA
	Naïve Bayes	NB

2) Ensemble machine learning: Ensemble classifiers integrate the decision of a set of base classifiers (single classifier) using either weighted or unweighted techniques to obtain a model that outperforms all base classifiers. Table IV describes the ensemble classifiers. Ensemble learning is an initiative technique that simulates the human nature of humans and considers several perspectives before making the final decision [31].

In traditional learning, only a single classifier is used to make the decision and solve a problem, whereas in ensemble learning, several classifiers are used to solve a problem [41]. The two main types of ensemble classifiers include (1) homogeneous classifiers, which use the same classifier such as RF, and (2) heterogeneous classifiers, which use a set of diverse classifiers such as LDA and DT. The sequence of running base classifiers and the decision combination are significant in building ensemble classifiers. Three major kinds of learning and aggregation exist.

First, bagging considers a homogenous classifier in which learning occurs independently (parallel approach). Then, the results are combined using the averaging process as in the following equation $E= (\Sigma\ e_i)/n$, where E refers to the final classifier, and e indicates the base classifier (i.e., RF) [42].

Second, boosting is a homogenous classifier that works sequentially (self-learning technique). Boosting works by assigning weights for every classifier (equal weights in the first stage). Then, the weights are updated every iteration based on the classifier performance. The final model is averaged using a weighted average with the following equation $e= ((\Sigma\ e_i w_i)/{*}\Sigma\ w_i))/n$, where e is the base model, and w is the weight (i.e., AdaBoost, light GBM, GBM, and XGBM). Fig. 1 shows the bagging, boosting, and subspace models [43], [44].

TABLE IV. DESCRIPTION OF ENSEMBLE CLASSIFIERS

Classifier name	Description	Classifiers	Aggregating result technique
RF	Ensemble of DT mode	DT	Majority voting
Bagging	Fits base classifiers on each random subset and aggregates the results (parallel)	LR	Average of five weak classifiers
Boosting	Trains data on a model and then uses the second algorithm to correct errors from the first model (sequentially)	LR	Weighted average of five weak classifier
Voting	Runs several classifiers on all dataset and then outputs class depending on the highest majority voting.	LR, SVC, KNN, and LDA	Soft voting
Stacking	Splits data into several data subsets; each subset runs on different classifiers, and the final results are aggregated.	LR, SVC, KNN, DT, LDA, and NB	Meta model (Level 2 learning)

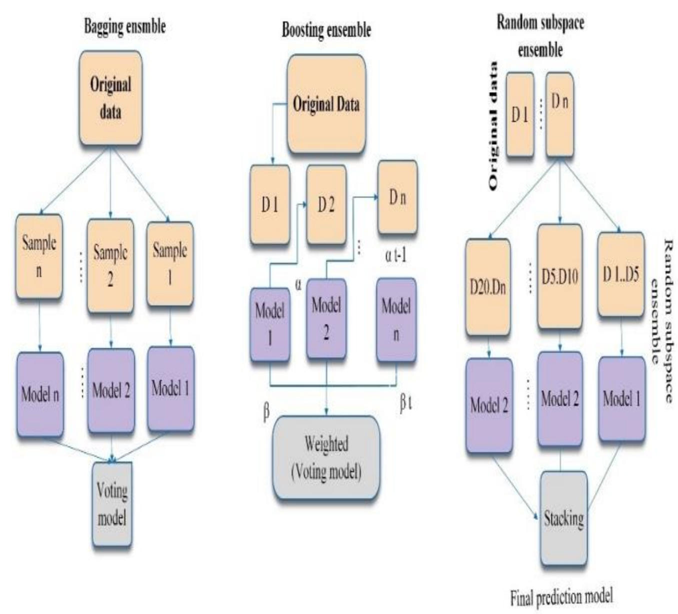


Fig. 1. Bagging, boosting, and subspace models.

Third, stacking considers heterogeneous classifiers that may run in two different ways, including the following. (1) All data are divided into feature subsets, with each subset running on all classifiers (several diverse classifiers) and a replacement for the selection of the most suitable classifier based on performance. Finally, the most accurate classifier is selected for each subset, and the final ensemble model is created. (2) For the second approach, all data sets are run in all classifiers (level 0); then, a meta learner is used to learn the best combination of predictions from all models. Fig. 2 displays the stacking ensemble model.

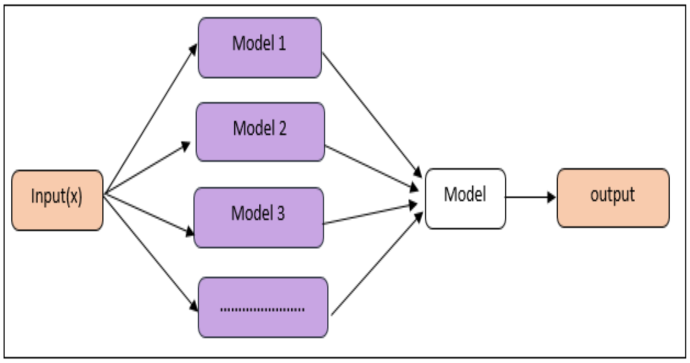


Fig. 2. Stacking ensemble learning.

IV. Proposed Framework

Based on the literature review, it was found that in the majority of the cases disregarded the role of FS algorithm for BC diagnosis. Others focused on the same or single algorithm in the classification although most of the different types of data require the use of various types of algorithms to be classified.

The steps of data pre-processing are as follows:

A. Data Pre-processing

This step aimed to represent data efficiently and improve the quality of the medical data collected. Data processing is

The objective of our proposed work is to construct an ensemble model to aggregate the benefits of different single algorithms until we can come up with an ensemble model which can give the best accuracy. We also use the method of FS that helps for improving the quality and making the process of model more efficient.

This framework is introduced a stacking ensemble classifier for the classification of BC based on tumor features (numeric features). It is a binary classification task in which 0 = benign and 1 = malignant. The model has been evaluated based on the WDBC dataset. As shown in Fig. 3, the proposed framework has several steps. The first step collects the WDBC data and implements the preprocessing steps to improve the quality of data, including (1) data normalization for data scaling to hasten the training process and (2) data balancing to increase the number of cases in the minority class. The second step uses FS to remove the weakest feature. The third step evaluates the performance of six single ML classifiers (KNN, LDA, DT, NB, SVM, and LR) with and without a FS step for BC classification task. The fourth step builds an ensemble model from a set of single classifiers with and without a FS step and compares its performance with other models that do not use ensemble learning. The tested ensemble models included RF, AdaBoost, bagging, stacking, and voting models.

We proposed a customized stacking ensemble model with high performance and Acc rate based on six optimized basic Classifiers required for ML to resolve several types of problems, such as noisy data, redundancy data, and missing data values, making it suitable for building and training ML mode:

1) Data normalization: In this step, we scaled all the data from their original range to be within the range of 0 to 1 through the following equation for normalization. This step speeds up the training processing.

$$y = \frac{(X - \min(x))}{(Max(x) - \min(x))} \tag{1}$$

2) Data balancing: As we mentioned in Section 3.1, WDBC included the features of 569 patients (357 benign and 212 malignant cases). The main problem of imbalanced data is that it affects the performance of ML algorithms. Algorithms, such as DT and KNN, will treat the minor class as not important as the major class and place more concern to the major class. Oversampling and undersampling are the main approaches used to deal with imbalanced data.

Oversampling increases the number of the training set in the minor class, whereas undersampling removes data from the minor class [45]. In this paper, given the shortage of the dataset, we decided to use an oversampling technique known as random oversampling [46]. This technique is used to randomly duplicate training examples in the minor class. The final dataset included 607 patients (357 benign and 250 malignant).

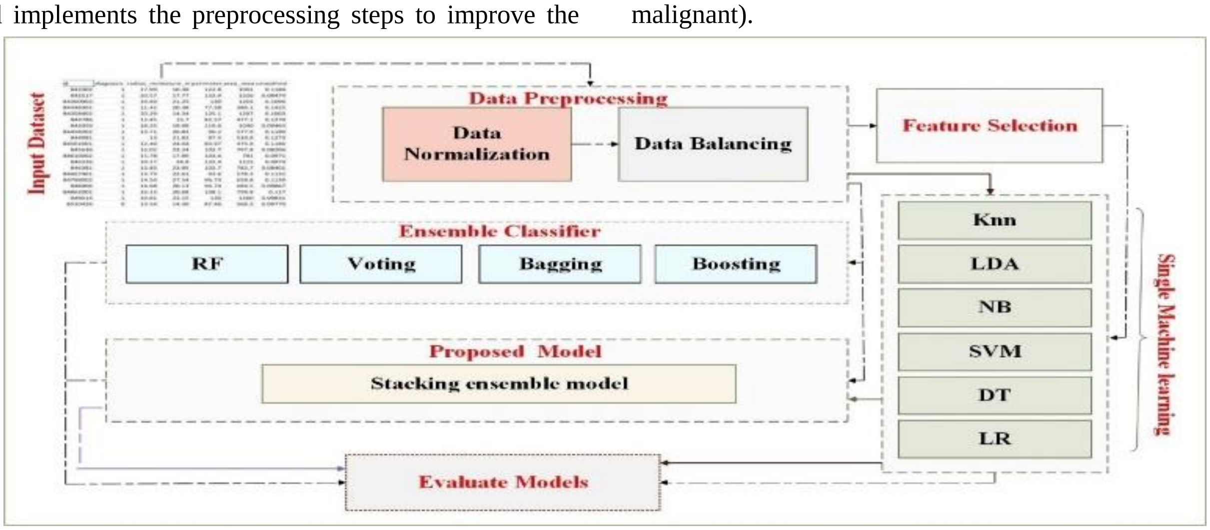


Fig. 3. Proposed framework.

B. Feature Selection

We used recursive feature elimination (REF) for FS. This method is used to reduce the input variable of a model by selecting relevant data only and cutting down the data noise [47].

The REF algorithm is based on filter and removes the weakest feature from our dataset to produce a new subset of features ranked by importance[48]. This process reduced the number of features from 32 to 24.

V. Results

A. Experimental Setup

All experiments were implemented on a PC workstation with an Intel Core i9, two tera hard disks, 32 GB RAM, and Windows 10 64 bit. All ML algorithms were implemented using Python language. Single and ensemble ML models were run using the Scikit-learn library.

All models were trained using 10-fold cross-validation (CV), which is a technique for validating the efficiency of a model in which unobserved testing data are used to evaluate the model performance and ensure the generalization performance of the training models.

We ensured that the training and testing datasets contained no data to prevent the algorithm from memorizing data and giving a better performance in the testing phase without ensuring its generalization capability. Fig. 4 shows the CV process.

B. Evaluation Performance

The effectiveness of ML algorithms was assessed using a set of performance measure parameters. The matrix of confusion, which included true positive (TP), False Positive (FP), true negative (TN), and FN results for actual and predicted data, was created to evaluate the parameters. In our study, we used the following parameters to evaluate terms by their formulas to evaluate our performance study.

The following equations were used to evaluate the comparison study's performance:

Accurcy (Acc): The correct ratio of samples that is classified to the total samples.

$$(\text{Acc}) = \frac{(TP+TN)}{(\text{TP+TN+FP+FN})} \quad (2)$$

Specificity (Sp): The relationship between negative examples observed and all other negative examples; shows the rate of projected existence in the presence of BC in all samples.

$$(\text{Sp}) = \frac{\text{TN}}{(TN+FP)} \quad (3)$$

Precision (P): The division of samples that were positive among all the samples that we anticipated to be positive.

$$(\text{P}) = \frac{\text{TP}}{(TP+FP)} \quad (4)$$

Recall (R): The rate of the case of positive impression divided by the total positive of cases.

$$(\text{R}) = \frac{\text{TP}}{(TP+FN)} \quad (5)$$

F1 score (F1): The mean between P and R

$$(\text{F1}) = 2 * \frac{P*R}{P+R} \quad (6)$$

Area under the ROC curve (AUC): Measures the performance and how the evaluated model distinguishes between models.

$$\text{AUC} = \frac{s_p - n_p + (n_{n+1)/2}}{n_p n_n} \quad (7)$$

C. Experimental Results

Four extensive experiments have been conducted to determine the efficiency of the proposed model using the WDBC dataset. WDBC provided 30 features for 560 patients and summarized the tumor status to determine benign or malignant tumors.

1) Single model

a) Without feature selection: In this section, we evaluated six diverse classifiers, including KNN, LDA, DT, NB, SVC, and LR. We carefully selected the most efficient classifier used frequently on the medical side. Table V shows the performance of all models represented by CV Acc, in addition to details about the testing performance in terms of P, R, F1, AUC, and SP.

From Table V, (1) SVC generated the least performance, with testing results of Acc=78.9%, AUC = 77.88%, and F1 = 69.2%, followed by LR with testing performance results of Acc = 92.1%, AUC = 90.49, and F1=87.32%. DT, KNN and NB performances improved by approximately 1%–2%, whereas LDA provided the best performance, with Acc = 98.2%, AUC = 97.97%, and F1 = 97.22%.

b) With feature selection: As shown in Table VI, in this section [31], [49], we explored the importance of using the FS technique to improve the ML model performance. In this paper, we used REF for FS. This technique works by fitting a model and then removing the weakest features. This process was repeated until the optimum feature set with 24 features was reached. FS techniques improved the Acc of the developed models by about 3%–6%.

SVC enhanced by about 6% and achieved Acc, AUC, and F1 of 89.4%, 82.71%, and 86.6, respectively. NB, KNN, and LR attained 92.9%, 91.22%, and 91.22%, respectively.

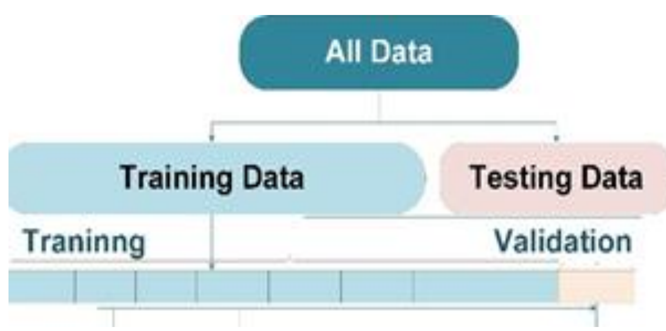


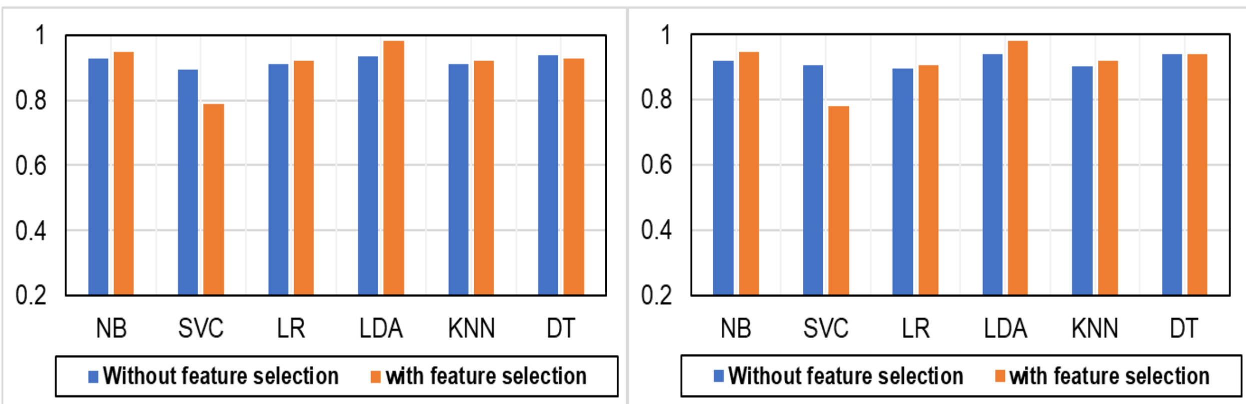
Fig. 4. Process of 10-fold CV.

TABLE V. SINGLE-CLASSIFIER RESULTS WITHOUT FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
NB	94.06	94.73	94.73	94.87	89.47	94.44	91.89	94.65
SVC	81.75	78.94	78.9	80.7	64.2	75.01	69.2	77.88
LR	91.64	92.10	92.1	92.1	88.57	86.11	87.32	90.49
LDA	96.92	98.24	98.2	98.71	97.22	97.22	97.22	97.97
KNN	94.94	92.10	92.1	92.30	84.61	91.66	87.99	91.98
DT	96.9	92.98	92.9	91.02	83.33	97.22	89.7	94.12

TABLE VI. SINGLE-CLASSIFIER RESULTS WITH FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
NB	94.505	92.98	92.9	95.8	92.3	92.3	87.8	91.8
SVC	90.54	89.47	89.4	86.3	71.61	85.11	86.6	82.71
LR	92.08	91.2	91.22	95.8	91.8	82.9	87.1	89.4
LDA	96.9	95.6	93.5	95.6	80.80	87..0	93.5	93.9
KNN	94.72	91.2	91.22	95.5	93.01	85.1	88.8	90.3
DT	98.02	93.8	93.8	94.0	91.6	93.06	92.6	93.8



(a) (b)

Fig. 5. (a) F-score of single classifiers with and without FS. (b) Acc result of single classifiers with and without FS.

The best performance was obtained by the DT, with Acc = 93.8%, F1 = 92.6%, and AUC = 93.8%. From the previous tables (Tables V and VI), we arrived at the following observations: (1) the use of FS improved the total performance with all classifiers; (2) DT worked better with less features. Fig. 5 compares F-score and Acc result of all single algorithms with and without FS.

2) Ensemble models

a) Without Feature Selection: In this section, we investigated the role of popular ensemble models, including voting, bagging, boosting, and RF, and compared them with the proposed stacking algorithm. The proposed stacking algorithm used three diverse classifiers: LDA, DT, and LR. LR was used as a meta classifier at level 2 of learning. These classifiers were selected because their diversity increases their search space and Acc accordingly. Table VII shows the CV Acc and testing performance of the used ensemble models. These results showed that (1) ensemble models improved the results by approximately 2%–4%; RF generated the least performance with the testing results of Acc = 93.8%, AUC = 92.8%, and F1=91.7%.

This finding may be attributed to the structure of RF, which selected a random subset to run in each branch. The other homogenous ensemble classifiers, including bagging and boosting classifiers, achieved better performances, with Acc = 94.3%, AUC = 96.01%, and F1 = 96.5% for the boosting classifier and Acc = 94.73%, AUC = 94%, and F1 = 93% for the bagging classifier. Regarding heterogeneous ensemble classifiers (voting and stacking), stacking provided the best performance, with Acc = 97.2%, AUC = 96.2%, and F1 = 97.72%. The performance improved by about 3%–5% over single classifiers and 1%–2% over homogenous ensemble classifiers.

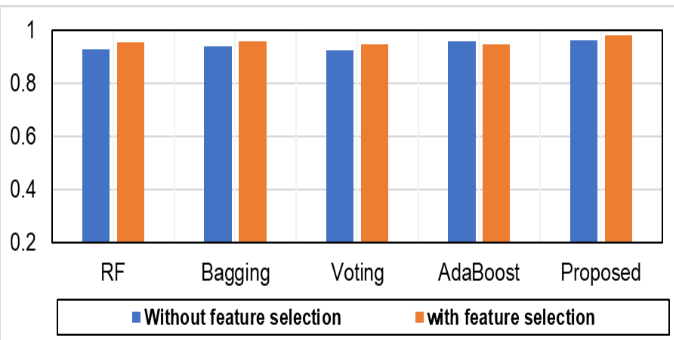
b) With Feature Selection: In this section, we explored the performance after using the FS technique with ensemble classifiers to improve the ML model performance. As shown in Table VIII, the FS techniques improved the Acc of the developed models by about 2%–3%. The RF enhanced by about 3% and achieved Acc, AUC, and F1 of 94.73%, 95.4%, and 92.1%, respectively.

TABLE VII. ENSEMBLE CLASSIFIER RESULTS WITHOUT FEATURE SELECTION

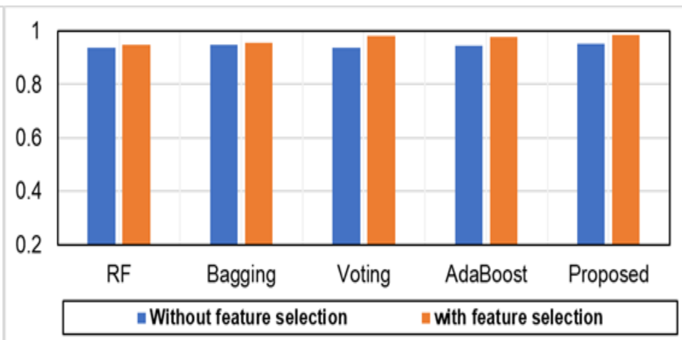
Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
RF	96.4	93.85	93.8	97.1	95.1	88.6	91.7	92.8
Bagging	98.5	94.7	94.73	97.14	95.2	90.9	93.0	94.0
Voting	95.9	93.8	93.85	98.5	97.4	86.3	91.5	92.4
AdaBoost	95.7	97.36	94.3	98.5	97.6	95.45	96.5	96.01
Stacking	99.4	98.2	97.2	98.5	97.72	97.7	97.72	96.2

TABLE VIII. ENSEMBLE CLASSIFIER RESULTS WITH FEATURE SELECTION

Algorithm name	Training score %	Testing score %	Acc %	Sp %	P %	R %	F-score %	AUC %
RF	96.7	94.73	94.73	93.64	87.5	97.2	92.1	95.4
Bagging	99.7	95.61	95.6	94.8	89.7	97.2	93.3	96.04
Voting	96.7	98.2	98.2	98.7	97.22	97.22	97.22	94.9
AdaBoost	94.41	95.40	97.7	94.8	89.4	94.44	91.89	94.6
Stacking	97.81	95.61	98.6	94.8	89.7	97.22	96.33	98.1



(a)



(b)

Fig. 6. (a). F-score of the ensemble classifiers with and without FS. (b) ACC result of the ensemble classifiers with and without FS.

The best performance was obtained from our stacking compared with other studies with F1 score = 9 6.33%, Acc = 98.6%, and AUC = 98.1%. From the previous tables (Tables VII and VIII), we arrived at the following observations; (1) the use of FS improved the total performance with all classifiers; (2) the combination of FS techniques with diverse stacking model improved the overall performance of the model. Fig. 6 compares F-score and Acc result of all ensemble algorithms with and without FS.

3) Model evaluation: We used the Nemenyi critical difference (CD) diagram to calculate the rank of each model [29], [45]; then, we evaluated and compared the performances of all the developed models (stacking, boosting, RF, LDA, and DT) with FS. The diagram in Fig. 7 shows that all algorithms were connected using horizontal lines, and they were ranked from 1 to 5, with rank 1 being alloted for the most accurate model and 5 for the least.

This ranking is equal to the average of ranking results of Friedman's test. We obtained complete information and explored the role of all features, such as radius_mean, area_mean, etc., in BC classification. This diagram proves that the stacking model achieved the best rank with a value 1.6. We calculated the feature importance with two different techniques, including information gain and RF. Table IX shows this importance numerically.

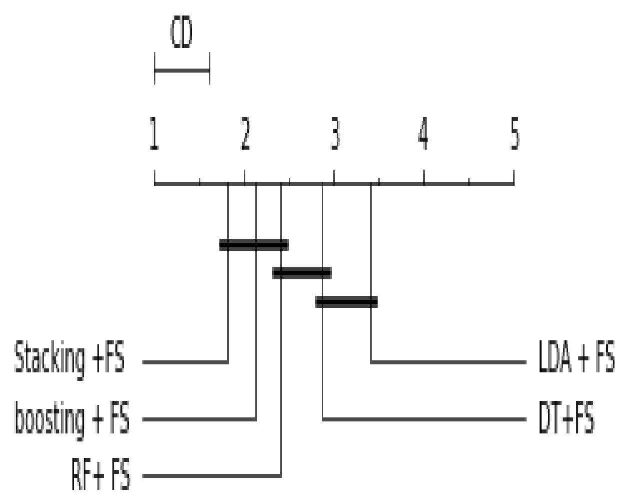


Fig. 7. CD between all classifiers with FS.

TABLE IX. RESULT OF THE FEATURE IMPORTANCE WITH INFORMATION GAIN AND RF

Feature Name	Correlation coefficient	RF
radius_mean	0.02537	0.02848
texture_mean	0.00337	0.00484
perimeter_mean	0.04407	0.05925
area_mean	0.05315	0.03405
smoothness_mean	0.00045	0.00001
compactness_mean	0.00049	0.01874
concavity_mean	0.03083	0.03261
concave points_mean	0.10136	0.14685
symmetry_mean	0.00118	0.00023
fractal_dimension_mean	0.00001	0.00108
radius_se	0.03708	0.02408
texture_se	0.00103	0.00028
perimeter_se	0.02726	0.03035
area_se	0.03464	0.03296
smoothness_se	0.00001	0.00052
compactness_se	0.01012	0.00001
concavity_se	0.00368	0.00435
concave points_se	0.00608	0.00799
symmetry_se	0.00054	0.00031
fractal_dimension_se	0.00001	0.00026
radius_worst	0.16231	0.12677
texture_worst	0.00825	0.00358
perimeter_worst	0.14621	0.15511
area_worst	0.17461	0.16965
smoothness_worst	0.00182	0.00064
compactness_worst	0.01091	0.00912
concavity_worst	0.01233	0.01436
concave points_worst	0.08065	0.09012
symmetry_worst	0.00612	0.00189
fractal_dimension_worst	0.00831	0.00121

VI. DISCUSSION

We introduced an ensemble ML classifier that can distinguish between malignant and benign BC cells. Our model was examined on the WDBC dataset. Our work faces the contribution of accelerating the training process, using data normalization for data scaling and increasing the number of instances in the minority class by data balancing. One of the main challenges in using FS is to remove the weakest feature in our model.

In order to test how well the suggested model works, we performed four experiments with and without the FS method for both single ML and ensemble ML.

In our research, we found that employing ensemble models with FS resulted in better overall performance across all classifiers. Based on six optimized basic classifiers (LR, SVC, KNN, DT, LDA, and NB), we proposed a customized stacking ensemble model that achieved the best performance and accuracy rate.

VII. COMPARISON WITH LITERATURE

In this section, we compared our proposed models with those of other literature. For fairness of comparison, we conducted this comparison only with studies that used the same dataset.

For example, [50] utilized SVM classifier in their studies to achieve an Acc of 97.47%. Similarly, in [51], a single classifier was utilized, and a classification performance of 96.5% was achieved in terms of Acc when using NB and ANN classifiers. Amrane et al. [52] used KNN and attained an Acc of 97.5%. Although these studies also used the same of classifiers, our results were superior with regard to the use of single classifiers. Such a conclusion can be attributed to the use of FS technique, which was used to extract the most important features. Bazila et al. [53] presented an ensemble known as tree augmented NB (TAN) with a boosting technique, and it achieved an Acc of 94.11%. Haifeng et al. [54] focused on an ensemble SVM learning approach named weighted area under the receiver operating characteristic curve ensemble (WAUCE). Their approach obtained an Acc of 97.68%. Table X summarizes this comparison.

Our proposed ensemble model used a heterogeneous ensemble classifier. This type of ensemble provides a high diversity between classifiers, which increases the overall classification performance. Our study achieved a superior Acc (98.6%) compared with other existing studies. On the other hand, our proposed model did not demonstrate a significant variation in the performance between training and testing. The proposed model can be potentially more stable and trustworthy and be used as an alternative to BC prediction.

TABLE X. COMPARISON OF THE ACC OF VARIOUS METHODS FOR BC CLASSIFICATION (EXPERIMENTS CARRIED OUT ON WDBC)

Ref.	YEAR	Method	Accuracy
[50]	2022	SVM	97.47%
[51]	2021	NB-ANN	96.5%
[55]	2021	KNN	97.15
[4]	2019	PCA with K-NN algorithm	95.6 %
[53]	2018	GBM -TAN	94.11%
[54]	2018	SVM -WAUCE model	97.68 %
[52]	2018	KNN	97.5%
[56]	2016	FS+ANN	97.3%
Proposed Model	2022	Stacking Ensemble ML	98.60%

VIII. CONCLUSION AND FUTURE WORK

In this paper, we proposed ML models for BC classification. We focused on the comparison of Acc when using single-classifier models, such as NB, KNN, LDA, LR, DT, and SVC, and ensemble models, such as RF, bagging, AdaBoost, stacking, and voting models, with and without using FS on WDBC. The key results from the study included the following.

In a single model, the SVC generated the least performance with testing results of Acc = 78.9%, AUC = 77.88%, and F1 = 69.2%. When using FS techniques, the total performance was improved, and SVC enhanced by about 6%, with Acc, AUC, and F1 of 89.4%, 82.71%, and 86.6%, respectively.

For the ensemble model, we compared traditional ensemble models combining FS techniques with our stacking proposed model. The results of our experiments revealed that the use of the ensemble model and FS improved the Acc of all classifiers. The performance improved by about 3%–5% over single classifiers and 1%–2% over homogenous ensemble classifiers. The results also showed that the best Acc (98.6%) was achieved by the stacking ensemble ML with FS techniques. The proposed model can help cancer specialists to recognize BC.

In the future, we plan to expand our model to handle different types of data using different feature selection methods. We will investigate the role of machine learning models in dealing with time series data. Finally, we can apply classification with deep learning algorithms to investigate the accuracy enhancement of the classification.

References

[1] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," J. Pers. Med., vol. 11, no. 2, pp. 1–12, 2021, doi: 10.3390/jpm11020061.

[2] A. Ed-daoudy and K. Maalmi, "Breast cancer classification with reduced feature set using association rules and support vector machine," Netw. Model. Anal. Heal. Informatics Bioinforma., vol. 9, no. 1, pp. 1–10, 2020, doi: 10.1007/s13721-020-00237-8.

[3] G. Hindley, O. B. Smeland, O. Frei, and O. A. Andreassen, "Big data and the goal of personalized health interventions," in Mental Health in a Digital World, Elsevier, 2022, pp. 41–61.

[4] H. Rajaguru and S. R. Sannasi Chakravarthy, "Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer," Asian Pacific J. Cancer Prev., vol. 20, no. 12, pp. 3777–3781, 2019, doi: 10.31557/APJCP.2019.20.12.3777.

[5] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," Procedia Comput. Sci., vol. 83, no. Fams, pp. 1064–1069, 2016, doi: 10.1016/j.procs.2016.04.224.

[6] Y. D. Austria, M. L. Goh, L. Sta. Maria Jr., J.-A. Lalata, J. E. Goh, and H. Vicente, "Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset," Int. J. Simul. Syst. Sci. Technol., no. March 2021, 2019, doi: 10.5013/ijssst.a.20.s2.23.

[7] T. A. Assegie, "An optimized K-Nearest Neighbor based breast cancer detection," J. Robot. Control, vol. 2, no. 3, pp. 115–118, 2021, doi: 10.18196/jrc.2363.

[8] J. Razmara, M. Salehi, and S. Lotfi, "Development of an Ensemble Multi-stage Machine for Prediction of Breast Cancer Survivability," J. AI Data Min., vol. 0, no. 3, pp. 371–378, 2020, doi: 10.22044/JADM.2020.8406.1978.

[9] R. Damaševičius, A. Venčkauskas, J. Toldinas, and Š. Grigaliūnas, "Ensemble-based classification using neural networks and machine learning models for windows pe malware detection," Electron., vol. 10, no. 4, pp. 1–26, 2021, doi: 10.3390/electronics10040485.

[10] B. F. Abdulrahman, S. M. N. M. R, S. W. Kareem, and Z. R. Ahmed, "Comparative Evaluation of Machine Learning Algorithms in Breast Cancer," Qalaai Zanist Sci. J., vol. 7, no. 1, pp. 878–902, 2022, doi: 10.25212/lfu.qzj.7.1.34.

[11] R. Karthiga and K. Narasimhan, "Medical imaging technique using curvelet transform and machine learning for the automated diagnosis of breast cancer from thermal image," Pattern Anal. Appl., no. 0123456789, 2021, doi: 10.1007/s10044-021-00963-3.

[12] G. Engineering, "An Outline of Machine Learning Techniques for Breast Cancer Prediction," vol. 3, no. 3, pp. 125–130, 2019.

[13] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," Int. J. Intell. Syst. Appl. Eng., vol. 7, no. 2, pp. 88–91, 2019.

[14] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," SN Comput. Sci., vol. 1, no. 5, pp. 1–14, 2020, doi: 10.1007/s42979-020-00305-w.

[15] M. A. Rahman and R. C. Muniyandi, "An enhancement in cancer classification accuracy using a two-step feature selection method based on artificial neural networks with 15 neurons," Symmetry (Basel)., vol. 12, no. 2, 2020, doi: 10.3390/sym12020271.

[16] S. B. Sakri, N. B. Abdul Rashid, and Z. Muhammad Zain, "Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction," IEEE Access, vol. 6, pp. 29637–29647, 2018, doi: 10.1109/ACCESS.2018.2843443.

[17] A. Alzu'bi, H. Najadat, W. Doulat, O. Al-Shari, and L. Zhou, "Predicting the recurrence of breast cancer using machine learning algorithms," Multimed. Tools Appl., 2021, doi: 10.1007/s11042-020-10448-w.

[18] R. A. Ibrahem Alhayali, M. A. Ahmed, Y. M. Mohialden, and A. H. Ali, "Efficient method for breast cancer classification based on ensemble hoffeding tree and naïve Bayes," Indones. J. Electr. Eng. Comput. Sci., vol. 18, no. 2, pp. 1074–1080, 2020, doi: 10.11591/ijeecs.v18.i2.pp1074-1080.

[19] M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," Eng. Appl. Sci. Res., vol. 48, no. 1, pp. 65–72, 2021, doi: 10.14456/easr.2021.8.

[20] T. E. Mathew, K. S. A. Kumar, and K. S. Kumar, "Breast Cancer Diagnosis using Stacking and Voting Ensemble models with Bayesian Methods as Base Classifiers," vol. IX, no. Ii, pp. 108–121, 2020.

[21] H. Saoud, A. Ghadi, and M. Ghailani, "Breast cancer diagnosis using machine learning and ensemble methods on large seer database," J. Theor. Appl. Inf. Technol., vol. 99, no. 3, pp. 594–604, 2021.

[22] G. Chugh, S. Kumar, and N. Singh, "Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis," Cognit. Comput., no. 0123456789, 2021, doi: 10.1007/s12559-020-09813-6.

[23] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," J. Med. Internet Res., vol. 21, no. 7, pp. 1–22, 2019, doi: 10.2196/14464.

[24] "UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set." https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 (accessed Sep. 14, 2022).

[25] Y. Ding, Y. Wang, and D. Zhou, "Mortality prediction for ICU patients combining just-in-time learning and extreme learning machine," Neurocomputing, vol. 281, pp. 12–19, 2018, doi: 10.1016/j.neucom.2017.10.044.

[26] H. Yao et al., "Severity Detection for the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests," Front. Cell Dev. Biol., vol. 8, no. July, pp. 1–10, 2020, doi: 10.3389/fcell.2020.00683.

[27] A. P. James and B. V. Dasarathy, "A review of feature and data fusion with medical images," Multisens. Data Fusion From Algorithms Archit. Des. to Appl., pp. 491–507, 2017, doi: 10.1201/b18851.

[28] H. Chiroma et al., "Neural networks optimization through genetic algorithm searches: A review," Appl. Math. Inf. Sci., vol. 11, no. 6, pp. 1543–1564, 2017, doi: 10.18576/amis/110602.

[29] M. Karimi Moridani and Y. Haghighi Bardineh, "Presenting an efficient approach based on novel mapping for mortality prediction in intensive care unit cardiovascular patients," MethodsX, vol. 5, no. 29, pp. 1291–1298, 2018, doi: 10.1016/j.mex.2018.10.008.

[30] G. Alfian, M. Syafrudin, M. F. Ijaz, M. A. Syaekhoni, N. L. Fitriyani, and J. Rhee, "A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing," Sensors (Switzerland), vol. 18, no. 7, 2018, doi: 10.3390/s18072183.

[31] N. El-rashidy, S. El-sappagh, and H. M. E. Abuhmed, Tamer, Smair Abdelrazek, "Intensive Care Unit Mortality Prediction : An Improved Patient-Specific Stacking Ensemble Model," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3010556.

[32] V. Srilakshmi, K. Anuradha, and C. S. Bindu, "Optimized deep belief network and entropy-based hybrid bounding model for incremental text

categorization," Int. J. Web Inf. Syst., vol. 16, no. 3, pp. 347–368, 2020, doi: 10.1108/IJWIS-03-2020-0015.

[33] S. Sareen, S. K. Gupta, and S. K. Sood, "An intelligent and secure system for predicting and preventing Zika virus outbreak using Fog computing," Enterp. Inf. Syst., vol. 11, no. 9, pp. 1436–1456, Oct. 2017, doi: 10.1080/17517575.2016.1277558.

[34] A. K. Das, S. Mishra, and S. S. Gopalan, "Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool," PeerJ, vol. 8, pp. 1–12, 2020, doi: 10.7717/peerj.10083.

[35] R. Sadeghi, T. Banerjee, and W. Romine, "Early hospital mortality prediction using vital signals," Smart Heal., vol. 9–10, pp. 265–274, 2018, doi: 10.1016/j.smhl.2018.07.001.

[36] R. Sadeghi, T. Banerjee, and W. Romine, "Early hospital mortality prediction using vital signals," Smart Heal., vol. 9–10, no. March, pp. 265–274, 2018, doi: 10.1016/j.smhl.2018.07.001.

[37] N. Mordvanyuk, F. Torrent-Fontbona, and B. López, "Prediction of glucose level conditions from sequential data," Front. Artif. Intell. Appl., vol. 300, pp. 227–232, 2017, doi: 10.3233/978-1-61499-806-8-227.

[38] I. U. Khan et al., "Computational Intelligence-Based Model for Mortality Rate Prediction in COVID-19 Patients," Int. J. Environ. Res. Public Health, vol. 18, no. 12, 2021, doi: 10.3390/ijerph18126429.

[39] D.-C. Li, C.-W. Liu, and S. C. Hu, "A learning method for the class imbalance problem with medical data sets," Comput. Biol. Med., vol. 40, no. 5, pp. 509–518, 2010, doi: https://doi.org/10.1016/j.compbiomed.2010.03.005.

[40] A. Ferraz, J. H. Brito, V. Carvalho, and J. Machado, "Blood type classification using computer vision and machine learning," Neural Comput. Appl., vol. 28, no. 8, pp. 2029–2040, 2017, doi: 10.1007/s00521-015-2151-1.

[41] K. Thangairulappan and P. Rathinasamy, "Ensemble Neural Network in Classifying Handwritten Arabic Numerals," J. Intell. Learn. Syst. Appl., vol. 08, no. 01, pp. 1–8, 2016, doi: 10.4236/jilsa.2016.81001.

[42] Y. Tounsi, L. Hassouni, and H. Anoun, "An Enhanced Comparative Assessment of Ensemble Learning for Credit Scoring," vol. 8, no. 5, 2018, doi: 10.18178/ijmlc.2018.8.5.721.

[43] B. Peter, Bagging , Boosting and Ensemble Methods Bagging , Boosting and Ensemble Methods, no. January 2012. 2015.

[44] Y. Wang, Y. Chen, and C. F. J. Huang, "Applying Neural Network Ensemble Concepts for Modelling Project Success," no. January, 2015.

[45] S. Ketu and P. K. Mishra, "Empirical Analysis of Machine Learning Algorithms on Imbalance Electrocardiogram Based Arrhythmia Dataset for Heart Disease Detection," Arab. J. Sci. Eng., vol. 47, no. 2, pp. 1447–1469, 2022, doi: 10.1007/s13369-021-05972-2.

[46] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," 2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020, no. May, pp. 243–248, 2020, doi: 10.1109/ICICS49469.2020.239556.

[47] S. Dehdar, K. Salimifard, R. Mohammadi, M. Marzban, S. Saadatmand, and M. Dianatinasab, "Predicting Breast Cancer Diagnosis Delay Using Machine Learning Techniques," SSRN Electron. J., vol. 2022, 2022, doi: 10.2139/ssrn.4088355.

[48] A. Thalor, H. Kumar Joon, G. Singh, S. Roy, and D. Gupta, "Machine learning assisted analysis of breast cancer gene expression profiles reveals novel potential prognostic biomarkers for triple-negative breast cancer," Comput. Struct. Biotechnol. J., vol. 20, pp. 1618–1631, 2022, doi: 10.1016/j.csbj.2022.03.019.

[49] S. El-Sappagh, M. Elmogy, F. Ali, T. ABUHMED, S. M. R. Islam, and K.-S. Kwak, "A Comprehensive Medical Decision–Support Framework Based on a Heterogeneous Ensemble Classifier for Diabetes Prediction," Electronics, vol. 8, no. 6, p. 635, 2019, doi: 10.3390/electronics8060635.

[50] I. Qureshi, B. Mohammad, and M. A. Habeeb, "A MACHINE LEARNING MODEL FOR BREAST CANCER DETECTION USING SUPPORT VECTOR MACHINE," J. Xi'an Shiyou Univ. Nat. Sci. Ed., vol. 65, no. 03, pp. 66–78, 2022, doi: 10.17605/OSF.IO/TSC6J.

[51] İ. ATEŞ and T. T. BİLGİN, "The Investigation of the Success of Different Machine Learning Methods in Breast Cancer Diagnosis," Konuralp Tıp Derg., vol. 13, no. 2, pp. 347–356, 2021, doi: 10.18521/ktd.912462.

[52] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018, pp. 1–4.

[53] A. Bazila Banu and P. Thirumalaikolundusubramanian, "Comparison of bayes classifiers for breast cancer classification," Asian Pacific J. Cancer Prev., vol. 19, no. 10, pp. 2917–2920, 2018, doi: 10.22034/APJCP.2018.19.10.2917.

[54] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," Eur. J. Oper. Res., vol. 267, no. 2, pp. 687–699, 2018, doi: 10.1016/j.ejor.2017.12.001.

[55] S. Varadhan, N. Jeswani, V. Sajnani, and P. Jaiswal, "Analysis of Breast Cancer Diagnosis and Prognosis Using Machine Learning Algorithms," in Lecture Notes in Electrical Engineering, 2021, vol. 700, pp. 3197–3211, doi: 10.1007/978-981-15-8221-9_298.

[56] S. Aalaei, H. Shahraki, A. Rowhanimanesh, and S. Eslami, "Feature selection using genetic algorithm for breast cancer diagnosis: Experiment on three different datasets," Iran. J. Basic Med. Sci., vol. 19, no. 5, pp. 476–482, 2016, doi: 10.22038/ijbms.2016.6931.

A Deep Learning-based Model for Evaluating the Sustainability Performance of Accounting Firms

Cui Hu
Business Big Date College, Chongqing College of Finance and Economics, Chongqing, 402160, China

***Abstract*—The harmonious and stable development of society is strongly related to the sustainable development of enterprises. In order to better face the challenges of environmental resources, sustainable development must be included in the development focus of accounting enterprises. The research proposes a performance evaluation model based on deep learning, improves RBMs model on the basis of deep belief network (DBN), improves the accuracy of the model through reverse fine-tuning technology, and effectively combines multiple restricted Boltzmann machines (RBMs) and Softmax classifiers to build a modular multi classification model to complete the sustainable development performance evaluation of accounting enterprises. The performance of RBM fine tuning classifier is higher than that of RBM expression and PCA (Principal Component Analysis) expression, which mainly shows the effectiveness and stability of feature extraction. The network output results of test samples are converted into prediction performance evaluation. The model is evaluated by average precision (AP), average recall (AR), and prediction accuracy. The AP, AR, and prediction accuracy of the proposed method are 86.95%, 89.74%, and 88.29% respectively, which are higher than Softmax classifiers, Back Propagation (BP) neural networks, and DBN based Softmax methods, It shows that this method is superior to other algorithms in the application of performance evaluation model for sustainable development of accounting enterprises, and it is feasible and effective, which is of great significance to the establishment of performance evaluation model for the accounting industry.**

Keywords*—*Deep learning; RBM; performance evaluation; classification accuracy; sustainability

I. INTRODUCTION

The sustainable development of enterprises is an important guarantee for a harmonious and stable society [1]. The theory of sustainable development includes equity, commonality, sustainability, coordination, demand, limitation, and inherency [2]. Corporate sustainability can be influenced by resource resources such as corporate culture, human capital, intangible assets and ecological nature [3]. From the perspective of corporate sustainability, its essence is clearly different from the meaning of ecological sustainability, which in depth is the optimal allocation of resources. Enterprise performance evaluation is the basis for measuring the production and operation behaviour of enterprises; therefore, it is positive to construct an enterprise sustainability performance evaluation model [4]. Performance evaluation indicators include the evaluation of non-economic aspects and economic aspects, but the previous economic performance indicators of enterprises are no longer applicable to the concept of accounting standards [5]. Domestic and international research on the evaluation of corporate sustainability performance has gone through four evaluation stages: cost performance, financial evaluation, strategic performance, and multidimensional synthesis [6]. With the advantages of efficient storage capacity, parallel processing capability, and associative memory, deep learning has now been more widely used in natural language processing technology, speech recognition, image recognition and other information retrieval. Common deep learning evaluation models include Deep Belief Network (DBN), Stacked Autoencoder (SAE), CNN, etc., and have positive application value in evaluation. In order to realize the evaluation and prediction of the sustainable development of enterprises with the help of the deep learning algorithm, on this basis, the research introduces the requirements of the harmonious financial theory and the new accounting standards, constructs the performance evaluation system of the sustainable development of accounting enterprises, and constructs the corresponding evaluation model through the deep learning algorithm. The contribution of the research is to propose a new feature extraction method with stronger feature learning ability, improve the RBMs model on the basis of DBN, improve the accuracy of the model through reverse fine-tuning technology, and effectively combine multiple RBMs and Softmax classifiers to build a modular multi classification model to complete the sustainable development performance evaluation of accounting enterprises. The goal and significance of the research is to provide a more scientific theoretical basis for the evaluation of the sustainable development of enterprises, and provide advice and guidance for the long-term development of enterprises. The research makes up for the phenomenon that the subjectivity of traditional evaluation methods is too strong, and the evaluation methods constructed have merit.

II. RELATED WORK

Sokil addressed the issue of information support and practical assumptions for sustainable development policy implementation through accounting analysis and mathematical analysis of the quadratic correlation and regression dependence of corporate value added on social and environmental cost scales [7]. Fu et al. used a multidimensional big data matrix model for optimization analysis to systematically construct a corporate performance evaluation system and two dimensions to construct a systematic model for the use of enterprise performance information. They applied logistic regression methods for empirical research, providing a theoretical and empirical basis for the division of the demand dimensions of the enterprise performance evaluation system [8]. Jana et al. proposed a multi-attribute decision method for enterprise financial performance evaluation, applying the new Dombi hybrid operators, which work on parameters have the advantage of good adaptability to complete financial performance assessment through intuitionistic fuzzy methods [9]. Yan Y et al. researchers analysed the performance appraisal management of manufacturing enterprises based on the Analytic Hierarchy Process (AHP) of fuzzy multi-criteria decision making to establish an enterprise performance management evaluation system, and this This exploration can provide a theoretical basis for improving the management innovation ability of manufacturing enterprises, and also has important reference value for expanding the market competitiveness of manufacturing enterprises [10]. Fang H applied the enterprise value management theory to the performance evaluation of university discipline construction members, established a performance value evaluation model, and applied the AHP method and quantitative research methods to costs and the functions achieved were evaluated [11].

Bao through the analysis of many classical methods such as economic value added, key performance indicators, balanced scorecard and current performance evaluation methods, the application analysis shows that the evaluation model is reasonable and the mathematical method is effective [12]. Yang et al. proposed the design of the enterprise environmental performance evaluation system in the new economic situation, according to the basic laws of the new situation theory and big data technology, the evaluation principles and typical index system were applied to the performance evaluation system as a reference, and the experimental results were of reference value for the application of the performance evaluation system in the field of corporate finance, which substantially improved the overall performance of the enterprise[13]. Qi established an enterprise human resource management performance evaluation system, which combined the relevant features of the opposable theory to provide more effective evaluation of enterprise HRM performance, providing a scalable analytical model for enterprise HRM performance evaluation [14]. Xiong et al. researchers evaluated the performance of food cold chain logistics enterprises based on AHP and entropy method, based on the analysis of the current situation, AHP was used to determine the subjective weights and entropy method was used to obtain experts' own weights. Finally, a comprehensive index fusion weight was performed [15]. Cui et al. researchers designed a multi-objective RBM model for training, and experiments showed that the proposed model could effectively improve the data classification accuracy of the heterogeneous network and reduce the loss in the data fusion process [16]. Sun L et al. proposed a gradient-enhanced Softmax classifier based on Convolutional Neural Networks (CNN), which alleviates the gradient disappearance problem in the Softmax classifier and can achieve better results [17]. Pan et al. proposed a novel deep learning network that performs hierarchical feature learning by limiting RBM machines in parallel, and the results demonstrate that the method is able to extract sensitive features for fault detection [18].

As can be seen through the research of domestic and international researchers on corporate innovation and development and performance evaluation models in the context of sustainable development, there are currently more studies that use AHP method for performance evaluation. In terms of building evaluation models based on deep learning, some researchers have proposed using RBM for hierarchical feature learning to improve feature extraction, but there are better methods for feature extraction of evaluation indicators. Therefore, the research will propose a new method of feature extraction with stronger feature learning capability, improve the model of Restricted Boltzmann Machine (RBM)s based on DBN, improve the correctness of the model through reverse fine-tuning techniques, and at the same time effectively combine multiple RBMs and Softmax classifiers to construct a modular multi-classification model to complete the sustainability performance evaluation of accounting firms.

III. RBMs Modular Deep Learning Model for Sustainability Performance Evaluation of Accounting Firms

A. RBMs Modular Deep Learning Models

Because the data type of university performance evaluation is sample data with multiple feature modules, deep learning algorithms have outstanding advantages in performance evaluation. RBM, as a new method of feature extraction, has stronger feature learning capability. The study will improve the model of RBMs based on DBN, improve the correct rate of the model through reverse fine-tuning technique, and at the same time effectively combine multiple RBMs and Softmax classifiers to construct a modular DBN is a concept generation model, which consists of multiple RBMs stacked as in Fig. 1.

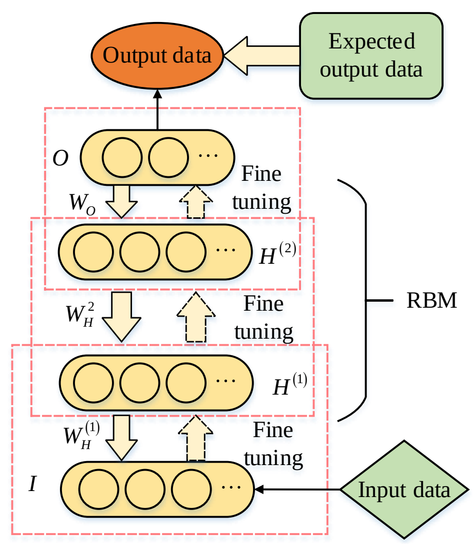


Fig. 1. Classic DBN network structure of two-layer RBM

Fig. 1 shows the structure of a DBN network containing two layers of RBM, which is first trained on the parameters of each layer of RBM by a greedy layer-by-layer unsupervised strategy, and then after confirming that all input data have passed through the full RBM network in order for the me network to obtain useful feature information. Finally, other algorithms such as the Back Propagation (BP) algorithm are used to receive the output data and use it as input data, and the whole network is supervised and adjusted according to the error between the expected and actual output data. At the top is the top layer network for supervised learning, which can become a classifier model such as Softmax depending on the specific practical application.

The RBM is an energy-based model where the visible and hidden layers form a system (v,h) with the energy shown in equation (1).

$$E(v,h) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{j=1}^{n}\sum_{i=1}^{m} h_i w_{ij} v_j \qquad (1)$$

In equation (1), n and m denote the number of neurons in the visual layer and the hidden layer respectively; v denotes the state vector of the visual layer, h denotes the state vector of the hidden layer, a denotes the bias vector of the visual layer, b denotes the bias vector of the hidden layer, and w_{ij} denotes the connection weights of the i neuron in the hidden layer and the j neuron in the visual layer. The writing matrix of equation (1) is shown in equation (2).

$$E(v,h) = -a^T v - b^T h - h^T W v \qquad (2)$$

Equation (2) can be understood as an energy function. This function combined with the RBM corresponding to the state (v,h) yields the expression for the distribution of the expression probability generation model as equation (3).

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \qquad (3)$$

In equation (3), Z is the normalisation factor, i.e. it is the accumulation of the states of all visible and hidden layers, as in equation (4).

$$Z = \sum_{v,h} e^{-E(v,h)} \qquad (4)$$

In contrast, for real life and work problems, the RBM network focuses on the marginal fraction of $p(v,h)$, the probability distribution of the input data v $p(v)$, as shown in equation (5).

$$p(v) = \sum_h p(v,h) = \sum_h \frac{1}{Z} e^{-E(v,h)} \qquad (5)$$

Given that there are no neurons that play a connecting role between the visual or hidden layers that are connectionless, the individual neurons are seamlessly connected to each other, as in equation (6).

$$\begin{cases} p(v|h) = \prod_{i=1}^{n} p(v_i|h) \\ p(h|v) = \prod_{j=1}^{m} p(h_j|v) \end{cases} \qquad (6)$$

When given the state of the visual layer v, the probability of finding the hidden layer neuron $h_j = 1$ is given in equation (7).

$$p(h_j = 1|v) = \sigma\left(b_j + \sum_{i=1}^{n} v_i w_{ji}\right) \qquad (7)$$

When given the state of the hidden layer h, the probability of finding the visual layer neuron $v_i = 1$ is given in equation (8).

$$p(v_i = 1|h) = \sigma\left(a_i + \sum_{j=1}^{m} h_j w_{ij}\right) \qquad (8)$$

In equations (7) and (8), $\sigma(x)$ is the Sigmoid function. The essence of the training process of the RBM is the use of learning algorithms to adjust the parameters of $\Theta = (a,b,W)$ through continuous iterative learning, which eventually makes the energy of the entire RBM move towards a gradient that gradually decreases, and finally reaches a state where the energy of the RBM is minimized, i.e., the model reaches a stable state [19]. When the energy of the RBM network is the lowest, it is known that the marginal distribution range of the RBM network is the largest according to the energy conversion formula, and then the parameter Θ needs to be solved for when the value of the marginal distribution is the largest. Based on the expert product system to fit complex high-latitude data, the system can have a more ideal probability distribution in the processing of high-dimensional model data, it can complete the target transformation training of RBM by maximizing the likelihood function, and the calculation expression is equation (9).

$$L(S,\Theta) = \prod_{i=1}^{N} p\left(v^{(i)}\right) \qquad (9)$$

In equation (9) $S = \left\{v^{(1)}, v^{(2)}, \cdots v^{(N)}\right\}$, $v^{(i)}$ denotes the i th training sample and N is the number of samples. Considering that the function $\ln x$ is strictly monotonically increasing and maximizing $L(S,\Theta)$ and maximizing $\ln L(S,\Theta)$ are equivalent, the objective of training the RBM becomes maximizing the log-likelihood function, as in equation (10).

$$\ln L(S,\Theta) = \sum_{i=1}^{N} \ln p\left(v^i\right) \qquad (10)$$

The gradient ascent method is the most efficient way to maximise the log-likelihood function and it uses an iterative approach to update the parameters as shown in equation (11) [20].

$$\Theta = \Theta + \eta \frac{\partial \ln L(S,\Theta)}{\partial \Theta} \quad (11)$$

In equation (11), $\eta > 0$ is the learning rate. For the training sample S, Eq. (12) is obtained by derivative calculation.

$$\frac{\partial \ln L(S,\Theta)}{\partial \Theta} = \sum_{i=1}^{N}\left(-\left\langle \frac{\partial E(v^{(i)},h)}{\partial \Theta} \right\rangle_{p(h|v^{(i)})} + \left\langle \frac{\partial E(v,h)}{\partial \Theta} \right\rangle_{p(v,h)} \right) \quad (12)$$

In equation (12) $\langle \cdot \rangle_{p(h|v^{(i)})}$ denotes the mathematical expectation based on the training samples and $\langle \cdot \rangle_{p(v,h)}$ denotes the mathematical expectation based on the model reconstruction data.

The study constructs a modular multi-classification model by combining multiple RBMs and Softmax classifiers. The model structure is shown in Fig. 2.

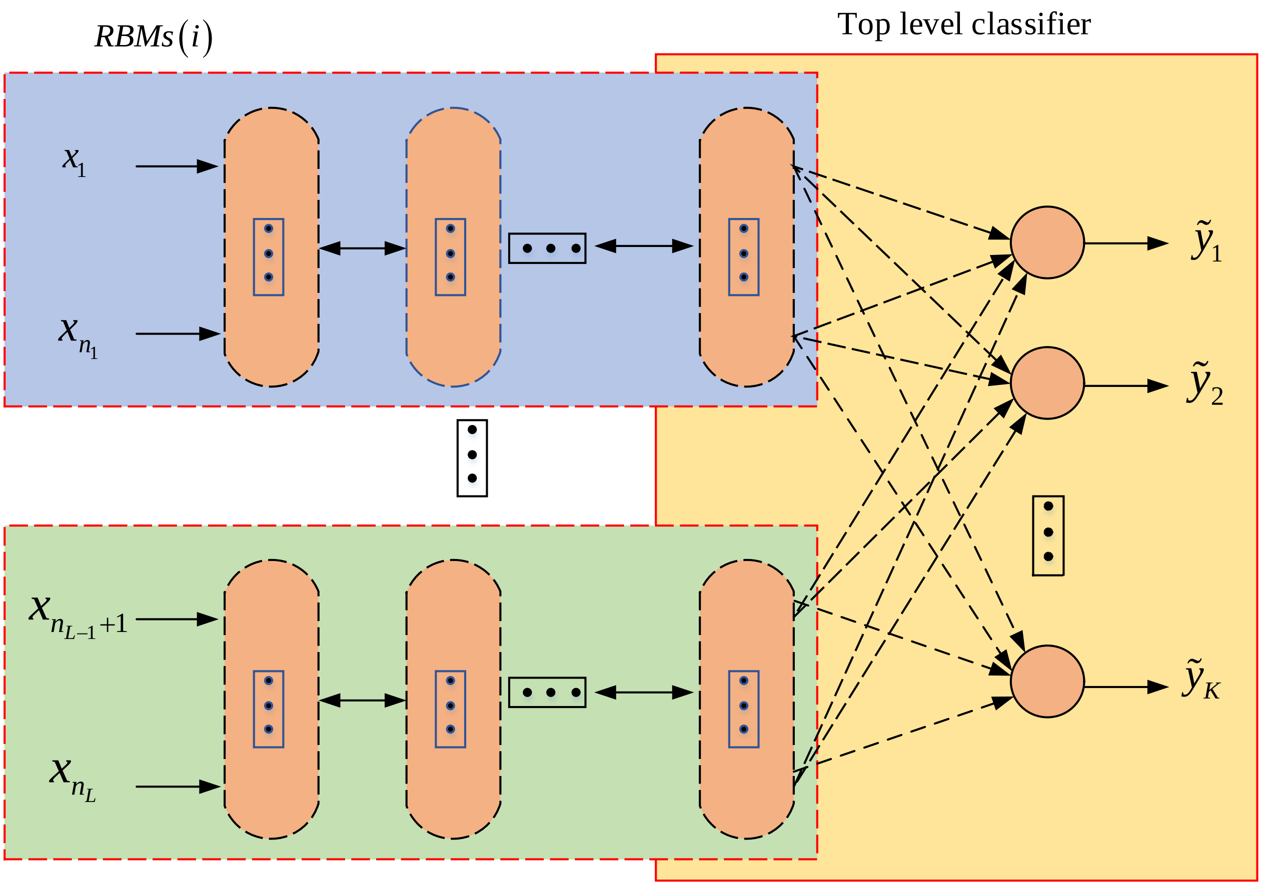


Fig. 2. Structural diagram of RBMs based modular deep learning model

The deep learning model designed in the study is a multiple RBMs feature extractor and a top-level classifier. The deep model first extracts different feature module samples using RBMs features, and then classifies various different types of data by the top-level classifier. The structure of the RBMs is contained in dashed lines, which is a modular feature extractor composed of RBMs, where $RBMs(i)$ is the i th RBM and $i = 1,2,\cdots,L$. Set the input features set to $(x_1, x_2,\cdots,x_n)$, divide them into L feature modules according to the sample characteristics, without considering the sequential nature, the feature modules are noted as $(x_1,\cdots,x_{n_1})$, $(x_{n_1+1},\cdots,x_{n_2})$, $\cdots$, $(x_{n_{L-1}+1},\cdots,x_{n_L})$, then the input layer of $n_L = n$, i feature module is equivalent to the visual layer of the first RBMs of $RBMs(i)$. The hidden layer of the last RBM is the new feature data, then the input of the top-level classifier is the output reconstruction result of each RBMs. The solid line included in the figure shows one of the network structures of the top-level classifier, with the classifier's first j network output value being $\tilde{y}_j$, where $j = 1,2,\cdots,K$.

The top layer of the deep learning model requires the addition of a classifier, and the study uses the Softmax classifier. The Softmax classifier is used as a non-linear model for classification, and it has a high correct classification rate when combined with a deep neural network. The Softmax classifier performs model training through a cost function, and it solves for the optimal parameters of the model through a parametric gradient method [21].

Given the training sample dataset $S=\left\{\left(X^{(1)},y^{(1)}\right),\left(X^{(2)},y^{(2)}\right),\cdots,\left(X^{(N)},y^{(N)}\right)\right\}$, where $X^{(i)}$ is the i sample, the category output values of $X^{(i)}$ are referred to by $y^{(i)}$ and the number of samples is referred to by N. Assuming that the output data contains K categories, the probabilistic classification values are calculated using the hypothesis function approach for $y^{(i)}\in\{1,2,\cdots,K\}$, and $\phi_j=p\left(y=j\middle|X\right)$, which defines the hypothesis function of the form shown in (13).

$$h_\theta\left(X^{(i)}\right)=\begin{bmatrix}p\left(y^{(i)}=1\middle|X^{(i)},\theta\right)\\p\left(y^{(i)}=2\middle|X^{(i)},\theta\right)\\\vdots\\p\left(y^{(i)}=K\middle|X^{(i)},\theta\right)\end{bmatrix}=\frac{1}{\sum_{j=1}^{K}e^{\theta_j^T x^{(i)}}}\begin{bmatrix}e^{\theta_1^T X^{(i)}}\\e^{\theta_2^T X^{(i)}}\\\vdots\\e^{\theta_K^T X^{(i)}}\end{bmatrix}\tag{13}$$

In equation (13), $\theta=\left[\theta_1,\theta_2,\cdots,\theta_K\right]^T$ is the Softmax classifier model parameter; the expression for the probability value of the sample $X^{(i)}$ being classified as the j class is $p\left(y^{(i)}=j\middle|X^{(i)}\right)$, the value normalized to the model output value is $\sum_{i=1}^{K}e^{\theta_j^T X^{(i)}}$, and the sum of the output values of all classes is 1. The cost function of the Softmax classifier is shown in equation (14).

$$C(\theta)=-\frac{1}{N}\left(\sum_{i=1}^{N}\sum_{j=1}^{K}I\left\{y^{(i)}==j\right\}\log\frac{e^{\theta_j^T X^{(i)}}}{\sum_{s=1}^{K}e^{\theta_s^T X^{(i)}}}\right)\tag{14}$$

In equation (14) I is the indicator function, when the expression is false, $I=0$; otherwise, $I=1$. The cost function of the partial derivative processing Softmax classifier yields the gradient expression of the parameters, which is calculated as shown in equation (15).

$$\nabla_{\theta_j}C(\theta)=-\frac{1}{N}\sum_{i=1}^{N}\left[X^{(i)}\left(I\left\{y^{(i)}==j\right\}-p\left(y^{(i)}=j\middle|X^{(i)};\theta\right)\right)\right]\tag{15}$$

B. Sustainability Performance Evaluation Model for Accounting Firms

After obtaining the enterprise performance evaluation data, the modular deep learning model combined with RBMs constructed by the institute is shown in Fig. 3. S I_1 I_2 The RBM training algorithm is used to obtain the output values of the first layer of RBMs u^2 and $\bar{u}^{2}$ connection weights, and the input data of the second RBM are the output values of RBMs u^2 and $\bar{u}^{2}$. The output values and connection weights of each layer are obtained through continuous iteration using the RBM training algorithm and a greedy layer-by-layer unsupervised pre-training strategy, and are used as the initial parameters of the network. After reconstructing the last output values u^{L1} and $\bar{u}^{L2}$, they are fed directly into the top layer Softmax classifier as input data, and the desired output data is used to obtain the initial parameters θ. A deep learning network model with classification function is formed by connecting the pre-trained RBMs(1), RBMs(2) and Softmax classifiers, and the parameters of the network are adjusted using the BP algorithm, and if the number of iterations meets the maximum number of iterations or the classification accuracy does not meet the requirements, the learning is terminated, otherwise the iteration step is repeated.

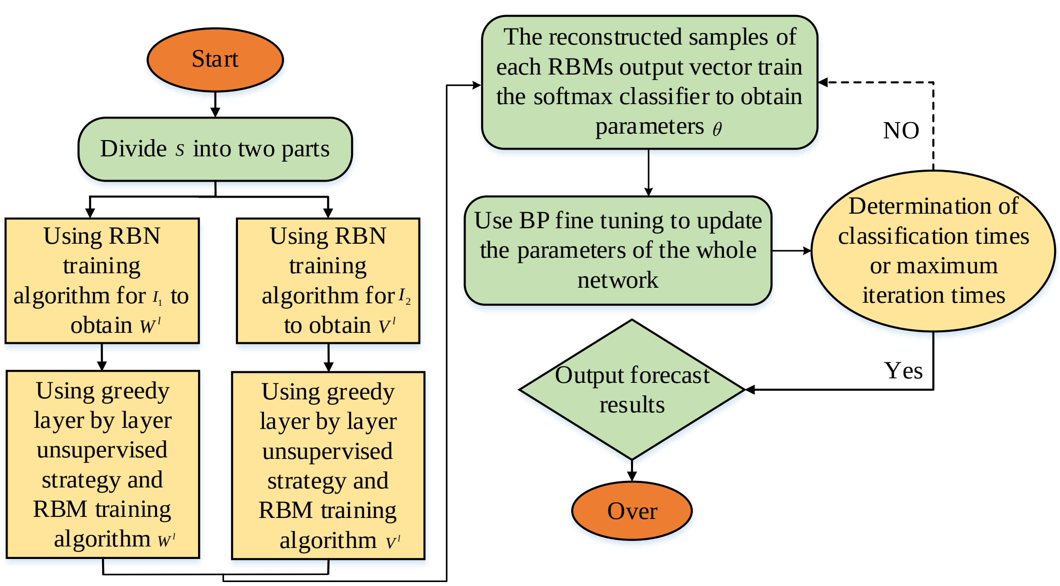


Fig. 3. Flow chart of modular deep learning algorithm based on two RBMs

The sustainable development of accounting enterprises depends on the optimal allocation of resources. Based on the experience of previous studies, the principles to be followed in the evaluation system for sustainable development of accounting enterprises are multiple evaluation subjects, multiple evaluation objectives, multiple evaluation contents and multiple evaluation methods, and the performance evaluation index system includes three aspects: social, resource and economic. According to the basic characteristics, completeness and reliability of the economic performance evaluation system, the study sets five secondary indicators, including cash flow, development capacity, profitability and quality, asset operation capacity and debt servicing capacity. Moreover, there are 15 tertiary indicators, including net asset cash return rate, net asset additional economic value rate, unit asset cash flow from operating activities, capital preservation growth rate, operating income growth rate, net asset additional economic value rate and net operating profit index. The 15 level 3 indicators are cash return on net assets, economic value added to net assets, cash flow from operating activities per unit of asset, capital preservation growth rate, operating income growth rate, economic value added to net assets, operating net profit index, return on total assets, return on net assets, total assets turnover, inventory turnover, accounts receivable turnover, quick ratio, interest earned multiple and gearing ratio.

According to the characteristics of human resources of the resource performance evaluation system, four secondary indicators are constructed including human resources output capacity, human resources utilization capacity, human resources input situation and basic human resources situation. 15 tertiary indicators include the adoption rate of rationalized suggestions, market share of new products, number of patents per capita, new technology development capacity, employees' profit generating capacity, staff turnover rate, employees' satisfaction rate, employees' career development level, staff training expenditure, human resources maintenance capacity, human resources investment level, professional work experience level of technical staff, proportion of technical staff, average age of staff, and education level of staff. Based on the definition of corporate social responsibility, the three secondary indicators of the social performance evaluation system selected for the study are the fulfillment of responsibility to the community, the fulfillment of responsibility to the government and the fulfillment of responsibility to consumers. After-sales service rate, return or repair rate, consumer complaint rate, and customer satisfaction. Table I refers to secondary and tertiary indicators at the social level.

TABLE I. SECONDARY AND TERTIARY INDICATORS OF SOCIETY, RESOURCES AND ECONOMY

Evaluation level	Evaluating indicator
Fulfillment of community responsibilities	Proportion of public service advertising expenditure
	Contribution rate of community activities
	Donation income ratio
Performance of government responsibilities	Employment rate of the disabled
	Employment ratio
	Social contribution rate
	Asset tax rate
Performance of consumer responsibilities	After sales service rate per unit income
	Return or repair rate
	Consumer complaint rate
	Customer satisfaction

The dataset selected for the study was data related to 83 accounting firms for the period 2017-2021, with a total sample data of 415 entries. Given the low relevance of the data sample data to the performance evaluation indicators, missing feature data and most of the data values being zero, the study cleared some of the indicator data. The raw values of the study performance were taken as 56-100, and the performance of the sample data was taken as the output variable, and the performance values were discretized into four category labels with values of 1, 2, 3 and 4 according to K mean clustering, with larger category labels indicating better performance, while the performance single output was discrete into a multiple output form. Given the variation in input feature magnitudes and magnitude units, the study utilises minmax normalisation for sample data processing. Assuming that x_i is the input feature and $x_i^{'}$ is the normalised feature, the normalisation formula can be expressed as equation (16).

$$x_i^{'} = x_i - \min(x_i) / (\max(x_i) - \min(x_i))$$

In equation (16), the maximum and minimum values of the x_i feature are $\min(x_i)$ and $\max(x_i)$, respectively.

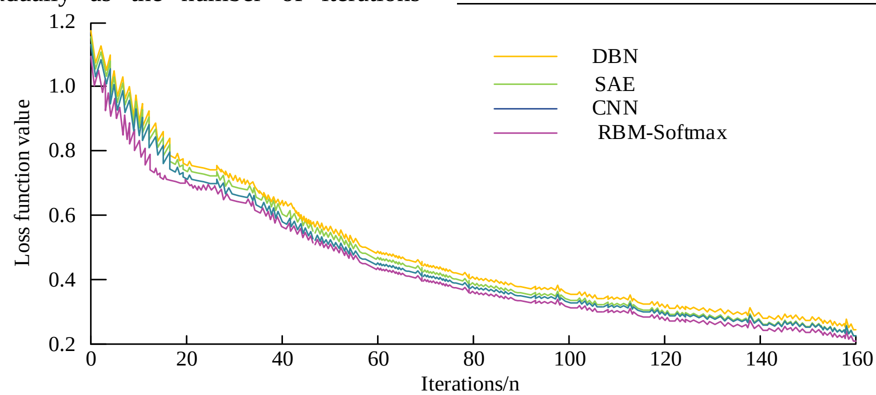
IV. Analysis of the Effectiveness of the Application of the Sustainable Development Performance Evaluation Model for Accounting Firms

The study used the 348 samples close to the centre of clustering as the training sample and the remaining samples as the test sample. Table II shows the results of the mean clustering referred to K. The four centres of performance clustering were 62.7, 68.5, 75.7 and 89.6.

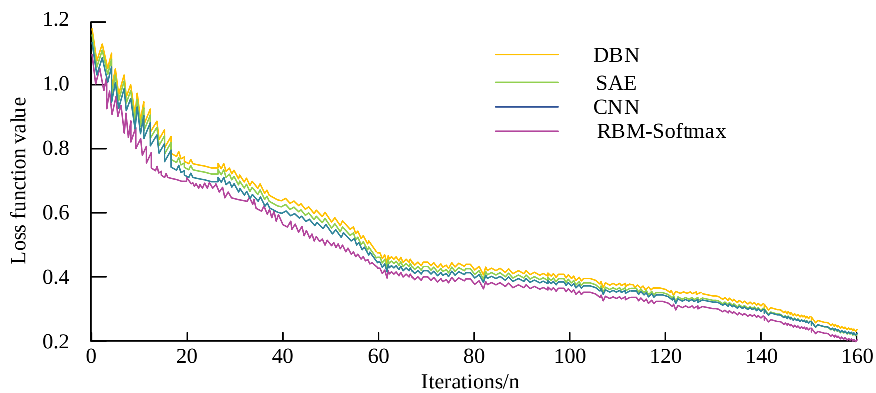
The loss function values of the four different deep learning algorithms for both the training and test sets are shown in Fig. 4(a) and 4(b). From the figures, it can be seen that the RBM-Softmax deep learning algorithm has the lowest value of loss for both the training and test sets, and the reduced value of loss decreases gradually as the number of iterations increases. The results for both the training and test sets show that the minimum loss value is achieved at 160 iterations, with corresponding values of 0.213 and 0.202, respectively.

TABLE II. Results of Mean Clustering K

Performance clustering centre	Category label	Multiple output form	Total samples	Number of training samples	Number of test samples
62.7	1	[1,0,0,0]	168	143	25
68.5	2	[0,1,0,0]	123	103	20
75.7	3	[0,0,1,0]	96	82	14
89.6	4	[0,0,0,1]	28	20	8



(a)Training set results of different depth learning algorithms



(b)Results of different depth learning algorithms

Fig. 4. Loss function values of four different depth learning algorithms in training set and test set

Setting the dimensionality of the input features and the dimensionality of the new features to 4 and 53, respectively, the changes of RBM fine-tuning, PCA (Principal Component Analysis) expression, and RBM expression on the performance of the Softmax classifier were analyzed, and the results are shown in Fig. 5. PCA expression and RBM expression refer to the method of classifying feature data through the Softmax classifier after the extraction of PCA and RBM features, respectively, while RBM fine-tuning refers to parameter tuning of the network based on the RBM expression (Ye F X et al. 2019) [22]. The study set each feature dimension to be run 10 times, and their average result was taken as the final result. Along with the gradual increase in the number of feature dimensions, the RBM expression Softmax classifier performed better compared to the PCA expression, and the RBM expression Softmax classifier extracted features with a correct rate of up to 75%. This is due to the fact that PCA as a linear feature extraction method decreases significantly when the number of feature dimensions increases, while the performance of the RBM fine-tuned Softmax classifier is higher than that of both PCA and RBM expression. There is no significant fluctuation in the performance of the classifier at different dimensions, with accuracy varying within a certain range. This indicates that the data features extracted by RBM fine-tuning are highly effective and stable, which in turn improves the accuracy of the model with a high probability.

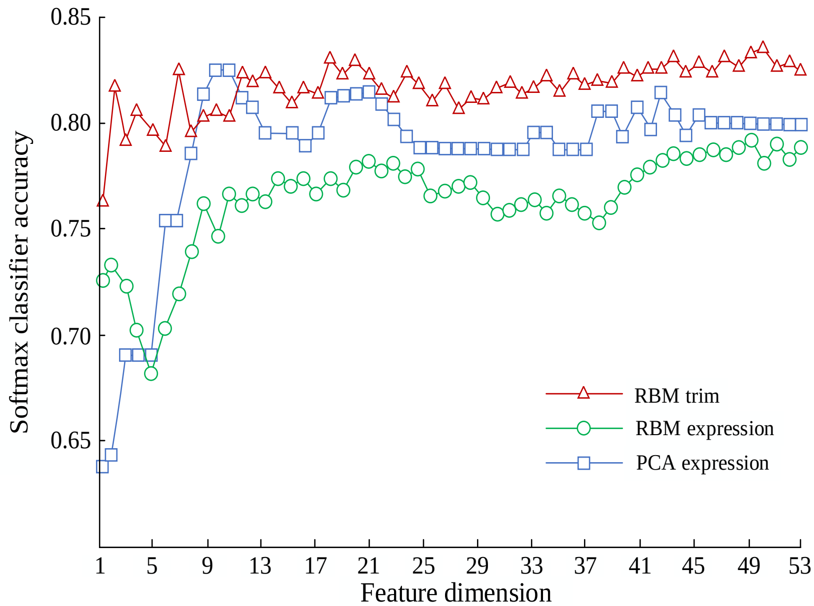


Fig. 5. Curve of accuracy with feature dimension

The study fed test samples into the already trained network and the corresponding classification results for each sample were obtained by classifying them with the top layer Softmax classifier. The input data of five randomly selected samples and the output values of the RBMs at each layer of the RBM were used as data for the performance analysis of the RBMs feature extractor, as shown in Fig. 6. For sample 1, sample 2, sample 3, sample 4 and sample 5 respectively, each dash in Fig. 6 indicates the input data value or RBM output value of a sample. As can be seen from the figure, the samples have high input data complexity and low correlation, which results in poorly differentiated and distinguishable samples. Each dash in Fig. 6 represents a sample's input data value or RBM output value. The output values of the input data after the two-layer RBM action are shown in Fig. 7.

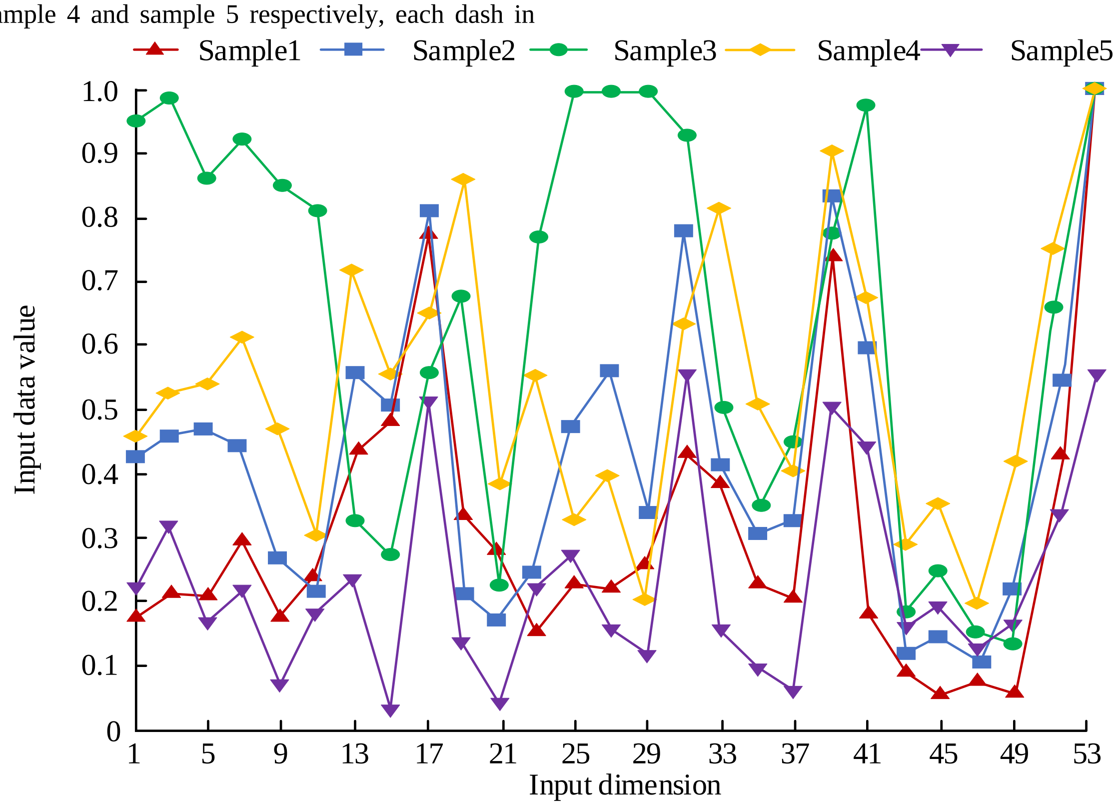


Fig. 6. Input value of sample

As can be seen in Fig. 7, after the extraction of data features by the two-layer RBM network, each sample presents an extremely clear data distribution, and the test samples are highly differentiable between different features. The feature results of the five samples show that the node output values of samples 3 and 4 are all greater than 0.7. The node output values of sample 2 are all relatively intermediate; and the number of nodes of samples 1 and 5 is close to 90% of the Therefore, the study divided the test samples into three categories, namely sample 3 and sample 4, sample 2, sample 1 and sample 5. The longitudinal comparison shows that the relationship between the magnitude of the second layer RBM output values of the samples is, sample 4 > sample 3 > sample 2 > sample 1 > sample 5. The cross-sectional comparison of the samples illustrates the characteristic of differentiability of the sample features.

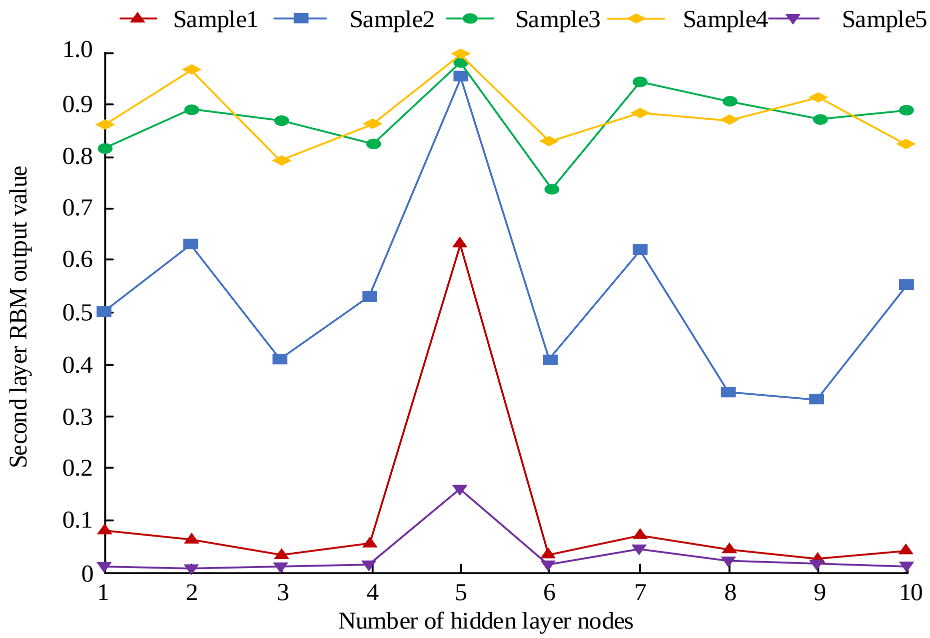


Fig. 7. Output value of the sample in the second layer RBM

The input samples indicate that the data distributions for sample 5 and sample 1 are indeed at a much lower position. The network outputs of the partial test sample data in the Softmax classifier are shown in Table III, where the sum of the probability values of each sample belonging to the four categories is 1. A higher probability value of a sample belonging to a category indicates that the sample belongs to that type. The categories for samples 1 - sample 10 are 2, 2, 3, 1, 2, 1, 2, 4, 4, and 3 in that order. This further validates that the deep learning model, after obtaining abstract feature representations through a multi-layer RBM network structure, the higher the differentiability of the features, which in turn allows for more accurate classification.

TABLE III. PARTIAL TEST SAMPLE DATA OUTPUT VALUE

Sample	Category 1	Category 2	Category 3	Category 4
1	0.0426	0.9402	0.0086	0.0086
2	0.0303	0.9514	0.0151	0.0032
3	0.0016	0.0715	0.8948	0.0320
4	0.9403	0.0423	0.0088	0.0086
5	0.2406	0.7315	0.0188	0.0091
6	0.9388	0.0435	0.0090	0.0087
7	0.0952	0.8816	0.0172	0.0060
8	0.0005	0.0032	0.0460	0.9504
9	0.0030	0.0004	0.0442	0.9524
10	0.0214	0.0006	0.9602	0.0179

The model is evaluated by average precision (AP), average recall (AR) and prediction accuracy, as shown in Fig. 8. It can be seen from the table that the AP of the proposed method is 86.95%, AR is 89.74%, and the prediction accuracy is 88.29%. The three evaluation indicators of the proposed method are greater than Softmax classifier, BP neural network and DBN Softmax. BP neural network algorithm is likely to fall into local extreme value, making training failure. The approximation and generalization ability of networks are closely related to the typicality of learning samples. For the Softmax classifier, the exponential Softmax function can be used to extend the numerical distance with large difference. Therefore, the proposed performance evaluation model for sustainable development of enterprises is more practical than other evaluation models.

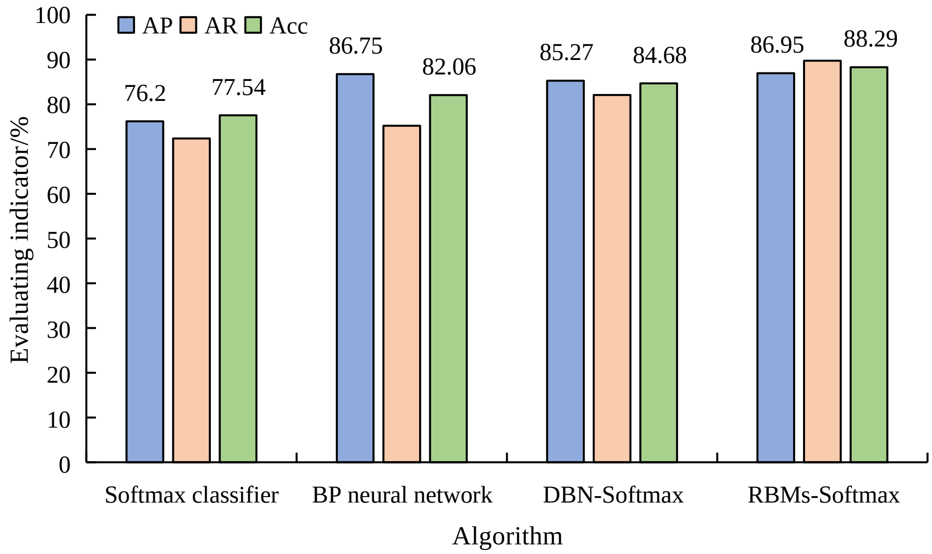


Fig. 8. Comparison of evaluation indexes of each algorithm

V. Conclusion

On the basis of analyzing the sustainable development evaluation system of accounting enterprises, the research constructs a sustainable development performance evaluation model combining multiple RBMs and Softmax classifiers. The results of training set and test set show that the minimum loss value is obtained when the number of iterations is 160 and the corresponding values are 0.213 and 0.202, respectively. The model is evaluated by AP, AR and prediction accuracy. The AP, AR and prediction accuracy of the proposed method are 86.95%, 89.74% and 88.29% respectively, which are higher than the Softmax classifier, BP neural network and DBN based Softmax method. This method is superior to other algorithms in the application of accounting enterprise sustainable development performance evaluation model. Embed sustainable development into the theory of enterprise performance evaluation, realize the comprehensive combination of sustainable development and enterprise performance evaluation theory, theoretically solve the problem of disconnection between enterprise performance evaluation and sustainable development, and enhance the scientificity, integrity and systematicness of performance evaluation theory. In the context of sustainable development, the model has important reference value and significance in the application of accounting enterprise performance evaluation. Although the proposed model has effectively improved the accuracy, the network initialization parameter color setting of this model refers to the empirical value and previous research results, and the training algorithm needs to be further studied in the future. The methods given need to be applied to empirical analysis in the future, and the follow-up research is expected to provide referential ideas and methods for the implementation of the concept of sustainable development at all social levels.

References

[1] G. Raita, T. O.Breaz, "The Era of Digitalization and Future Changes in The Accounting Profession." European Journal of Accounting, Finance & Business, vol.16, no. 26, pp. 63-72, 2021.

[2] Y. Zhang, F. Xiong, Y. Xie, et al. "The Impact of Artificial Intelligence and Blockchain on the Accounting Profession." IEEE Access, no. 8, pp. 461-477, 2020.

[3] E. Nechita, "Analysis of the Relationship between Accounting and Sustainable Development. Sustainable Development." The Audit Financiar journal, vol. 17, no. 155, pp. 520-536, 2019.

[4] J. Bebbington, J. Unerman, "Advancing research into accounting and the UN Sustainable Development Goals." Accounting, Auditing & Accountability Journal, vol. 33, no. 7, pp. 1657-1670, 2020.

[5] N. Iershova, O. Kryvytska, Y. Kharchuk, "Accounting and Analytical Information Formation by Business Entities under Conditions of Sustainable Development Concept Realization." Financial and Credit Activity Problems of Theory and Practice, vol. 2, no. 37pp. 86-94, 2021.

[6] I. Ankova, "Sustainable Development and Challenges to Accounting." Economic Science, education and the real economy: Development and interactions in the digital age, no. 1, pp. 614-625, 2020.

[7] O. Sokil, "Accounting and Analytical Support for Cost and Value-Added Management: The Way to Sustainable Development." Accounting and Finance, vol. 1, no. 87, pp. 59-68, 2020.

[8] H. Fu, "Optimization Study of Multidimensional Big Data Matrix Model in Enterprise Performance Evaluation System." Wireless Communications and Mobile Computing, no. 2, pp. 1-12, 2021.

[9] C. Jana, G. Muhiuddin, M. Pal, et al. "Intuitionistic Fuzzy Dombi

Hybrid Decision-Making Method and Their Applications to Enterprise Financial Performance Evaluation." Hindawi, no. 7, pp. 1-14, 2021.

[10] Y. Yan, D. Chu, "Evaluation of Enterprise Management Innovation in Manufacturing Industry Using Fuzzy Multicriteria Decision-Making under the Background of Big Data." Mathematical Problems in Engineering, no. 11, pp. 1-10, 2021.

[11] H. Fang, "Performance Evaluation Method of Discipline Construction Members in Colleges and Universities from the Perspective of Value Management" Asian Agricultural Research, vol. 12, no. 5, pp. 54-55, 2020.

[12] T. T. Bao, X. L. Xie, P. Y. Long, "Shipping enterprise performance evaluation under uncertainty base on multiple-criteria evidential reasoning approach." Transportation Research Procedia, vol. 25, pp. 2761-2772, 2017.

[13] G. Yang, X. Zhou, "Performance evaluation system design of enterprise environment based on new situation theory and big data technology." Revista de la Facultad de Ingenieria, vol. 32, no. 5, pp. 641-649, 2017.

[14] Y. Qi, "An extension analysis model for performance evaluation of enterprise human resource management based on extension theory." Vol. 42, no. 3, pp. 1243-1246, 2017.

[15] Y. Xiong, J. Zhao, J. Lan, "Performance Evaluation of Food Cold Chain Logistics Enterprise Based on the AHP and Entropy." International Journal of Information Systems and Supply Chain Management vol. 12, no. 2, pp. 57-67, 2019.

[16] Z. Cui, Y. Zhao, Y. Cao, et al. "Malicious Code Detection under 5G HetNets Based on Multi-Objective RBM Model." IEEE Network, vol. 35, no. 2, pp. 82-87, 2021.

[17] L. Sun, L. I. Weijun, X. Ning, et al. "Gradient-Enhanced Softmax for Face Recognition." IEICE Transactions on Information and Systems, vol. 103, no. 5, pp. 1185-1189, 2020.

[18] T. Pan, J. Chen, J. Pan, et al. "A Deep Learning Network via Shunt-Wound Restricted Boltzmann Machines Using Raw Data for Fault Detection." IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 7, pp. 4852-4862, 2019.

[19] S. Feng, C. Chen, "A Fuzzy Restricted Boltzmann Machine: Novel Learning Algorithms Based on the Crisp Possibilistic Mean Value of Fuzzy Numbers." IEEE Transactions on Fuzzy Systems, vol. 26, no. 1, pp. 117-130, 2018.

[20] C. Dewi, R. C. Chen, Hendry, et al. "Experiment Improvement of Restricted Boltzmann Machine Methods for Image Classification." Vietnam Journal of Computer Science, vol. 8, no. 3, pp. 417-432, 2021.

[21] Q. Liu, L. Sun, A. Kornhauser, et al. "Road roughness acquisition and classification using improved restricted Boltzmann machine deep learning algorithm." Sensor Review, vol. 39, no. 6, pp. 733-742, 2019.

[22] F. X. Ye, S. Q. Zhu, "Class-Incremental Learning Based on Feature Extraction of CNN With Optimized Softmax and One-Class Classifiers." IEEE Access, vol. 7, no. 3, pp. 42024-42031, 2019.

A Hierarchical ST-DBSCAN with Three Neighborhood Boundary Clustering Algorithm for Clustering Spatio–temporal Data

Amalia Mabrina Masbar Rus[1], Zulaiha Ali Othman[2], Azuraliza Abu Bakar[3], Suhaila Zainudin[4]
Department of Informatics, Universitas Syiah Kuala, Banda Aceh 23114, Indonesia[1]
Center for Artificial Intelligence Technology (CAIT), Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia[2, 3, 4]

***Abstract*—Clustering Spatio-temporal data is challenging because of the complexity of processing the spatial and temporal aspects. Various enhanced clustering approaches, such as partition-based and hierarchical-based algorithms have been proposed. However, the ST-DBSCAN density-based algorithm is commonly used to process irregularly shaped clusters. Moreover, ST-DBSCAN considers neighborhood parameters as spatial and non-spatial. The preliminary results from our experiments indicate that the ST-DBSCAN algorithm addresses temporal elements less effectively. Therefore, an improvement to the ST-DBSCAN algorithm was proposed by considering three neighborhood boundaries in neighborhood function. This experiment used the El Niño dataset from the UCI repository. The experimental results show that the proposed algorithm increased the performance indices by 27% compared to existing approaches. Further improvement using the hierarchical Ward's method (with thresholds of 0.3 and 0.1) reduced the number of clusters from 240 to 6 and increased performance indices by up to 73%. It can be concluded that ST-HDBSCAN is a suitable clustering algorithm for Spatio-temporal data.**

***Keywords*—Data mining; hierarchical clustering; density-based clustering; spatio–temporal clustering**

I. Introduction

Clustering is a process for grouping data based on similarity distance. It is an effective data mining technique for data segmentation, feature selection, pattern recognition, and anomaly detection. Clustering is an unsupervised method that does not require a prior definition of the input data classes. Clustering techniques help uncover hidden patterns in the examined data [1]. Unlike conventional clustering algorithms that process nonspatial or non-temporal data, clustering spatio–temporal data is challenging. Spatio-temporal data is different from relational data, in which computational approaches are developed for both spatial and temporal attributes [2]. Because of spatio–temporal data structure features, which record the general variables and the corresponding location and time [3], it became challenging to cluster the spatial and temporal data together.

Spatio–temporal clustering is an emerging research area. It gains actual location and time information from the enormous amount of geographical data provided by GPS, satellite, wireless technology, sensor networks, and other devices that could transmit location and time-stamped data. For instance, an organization has invested more resources in obtaining hidden knowledge and information from spatio–temporal data making research in this field more critical. The work by Bogorny and Shekar [4], Mazimpaka and Timpf [5], and Atluri *et al.* [6] have reviewed spatio–temporal data mining and its applications in surface ozone (O3) variations [7], forest fire [8], groundwater potential zone [9], citizens security [10], traffic congestion [11], healthcare, and social media.

Kisilevich *et al.* [12] categorized spatio–temporal data based on temporal extension, spatial extension, and spatial location. Spatial extension expands the spatial shape from points into lines and then areas. The expansion starts from a single snapshot to an updated snapshot and the completed time series in temporal extension. The spatial location is divided into two categories depending on the data collection location (fixed or dynamic). Based on these extensions, Kisilevich *et al.* [12] defined the following five types of spatio–temporal data: ST-events, geo-referenced variable, geo-referenced time series, moving objects, and trajectories.

M. Y. Ansari *et al.* [13] presents six categories of spatiotemporal clustering algorithms: event clustering, geo-referenced data item clustering, geo-referenced time series clustering, moving clusters, trajectory clustering, and semantic based trajectory data mining. ST-events have a fixed location and only store one snapshot of variable values. The geo-referenced variable also has a fixed location but stores only the current updated values. Similarly, geo-referenced time series also have a fixed location, but it stores the entire history of variable values as time-series data for each location. Moving objects and trajectories have dynamic locations. The object changes its location over time; however, only current values are recorded. In contrast, trajectories record each object's movement as a completed time series. Similar to Madraky *et al.* [14], this research focuses on clustering geo-referenced time-series data and utilizes a nature-inspired approach.

Various clustering approaches have been proposed to enhance the traditional algorithm, such as density-based, partition-based, and hierarchical-based [15]. The density-based algorithm is widely used because it can process irregular-shaped clusters, e.g. ST-DBSCAN [16], ST-OPTICS [17], ST-Shared Nearest Neighbors (SNN) and ST-SEP-SNN [18], P-DBSCAN [19], ST-DCONTOUR [20], RT-DBSCAN [21], CorClust [22], and MDST-DBSCAN [23]. One of the commonly used spatio–temporal density-based clustering approaches is Spatio–temporal (ST)-Density-based Spatial

Clustering of Applications with Noise (DBSCAN) [16]. However, the algorithm does not address the temporal elements of data. The neighborhood parameters are defined as spatial and non-spatial objects only. Therefore, the maximum temporal distance was later introduced to improve the ST-DBSCAN algorithm.

Other techniques extended to other traditional density-based clustering algorithms. For example, Spatio–temporal Shared Nearest Neighbor (ST-SNN) and Spatio–temporal Separated Shared Nearest Neighbor (ST-SEP-SNN) extend SNN by Ertöz *et al.* [24] but introduce the polygon distance algorithm to search core polygon [18]. Their results have been compared with that of PDBSCAN [19]. 4D+SNN also added weighting for spatial and temporal attributes [25]. The ST-DCONTOUR algorithm by Zhang and Eick [20] emphasizes the batching process of SNN density-based clustering algorithms.

The ST-DBSCAN algorithm by Birant and Kut [16] is one of the standard benchmark algorithms for spatio–temporal clustering. In ST-DBSCAN, introducing a new distance limit for spatial data called *Eps 2* enhanced the DBSCAN algorithm. The sea surface temperature, sea surface height, and wave height in the Black, Marmara, Aegean, and Mediterranean seas were clustered. The result was presented as a map with labels of the cluster numbers for each group. Besides, utilizing a cluster map emphasizes the spatial aspect of the data rather than the temporal aspect. Therefore, this research aims to improve ST-DBSCAN by considering spatial and non-spatial.

The following Sections II discuss how the ST-DBSCAN algorithm is limited and how to improve it by proposing Three Neighborhood Boundary in neighborhood function and improving it again using the hierarchical ward's method. Section III discusses the dataset and parameter settings; Section IV discusses the results obtained from experiments, including comparison results with the benchmark algorithm, and the conclusion and future work.

II. Related Work

Density-Based clustering is an unsupervised learning technique that locate distinct groups or clusters in the data. These techniques are based on the notion that a cluster in data space is a contiguous region of high point density, separated from other such clusters by contiguous regions of low point density. The fundamental density-based clustering algorithm is called Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Outliers and noise-filled massive amounts of data can be used to find clusters of various sizes and shapes.

Three minor modifications to DBSCAN, known as ST-DBSCAN , are proposed by D. Birant *et al*. [16]. ST-DBSCAN deal with the detection of (i) core items, (ii) noise objects, and (iii) neighboring clusters. The ability to identify clusters on spatial-temporal data is the primary motivation behind this modification. When clusters of various densities occur, the second adjustment is required to locate noisy objects. D. Birant et al. [16] introduce the idea of density factor. Each cluster is given a density factor, which indicates how dense the cluster is. The third adjustment compares the cluster's average value with a new value that will soon be available.

In ST-DBSCAN, the DBSCAN algorithm was enhanced by introducing a new distance limit for spatial data called Eps 2. In terms of its research, sea surface temperature, sea surface height, and wave height in the Black Sea, the Marmara Sea, the Aegean Sea, and the Mediterranean Sea are clustered. The clustering result was presented on a map labeled with a cluster number for each area. Using a cluster map, the result focused more on the spatial part of data and lacked in showing the temporal part.

The problem with ST-DBSCAN is the lack of an algorithm on the temporal aspect of spatio-temporal data. This is because the temporal data were separated manually by filtering the data that occurred on a consecutive day or the same day in a different year [16]. Thus, the algorithm did not identify the cluster with a pattern that exists continuously in two different years. The cluster generated took into account the spatial and non-spatio-temporal part of the data; however, it lacks the usage of the temporal aspect of the data.

The clustering algorithm "Spatio-Temporal-Ordering Points to Identify Clustering Structure (ST-OPTICS)" was proposed by K.P. Agrawal *et al.* [17] and is a modified version of the previous density based technique "Ordering Points to Identify Clustering Structure (OPTICS)." The proposed algorithm can produce spatio-temporal clusters and address issues such as (i) handling spatio-temporal data, non-spatial values to take into account temporal dimensions, (ii) algorithm does not depend on dimensions of data, so it is ready to handle n-dimensions, and (iii) independence of ordering of observations in database can be observed by the working principle of the proposed technique like it first performs Orde, (iv) locating nested and nearby clusters, and (v) ultimately, it is also scalable.

The ST-OPTICS algorithm employs ST-DBSCAN to extract the clusters and then performs hierarchical clustering to aggregate the clusters. The result generated has better performance indices compared to ST-DBSCAN. However, comparing to a similar version of ST-DBSCAN, ST-OPTICS still performs clustering using spatial and non-spatio-temporal aspects of the data but lacks the temporal aspect of the data. Similarly, the results are also shown in the form of maps which limit the description of the temporal part of data. Therefore, there is still a need for a clustering algorithm that could cluster spatio-temporal data that will account not only the spatial and non-spatio-temporal parts of data but also the temporal part of data.

To cluster overlapping polygons that can change their positions, sizes, and shapes over time, Sujing Wang et al. [18] introduce two new spatiotemporal clustering algorithms, called Spatiotemporal Shared Nearest Neighbor clustering algorithm (ST-SNN) and Spatiotemporal Separated Shared Nearest Neighbor clustering algorithm (ST-SEP-SNN). The core polygon notion and the spatial closest neighborhood of a polygon are both redefined by Sujing Wang *et al.* [18]. Even with high-dimensional data with outliers, ST-SNN and ST-SEP-SNN may locate clusters of various sizes, shapes, and densities.

P-DBSCAN, a novel DBSCAN for analysis of locations and events utilizing a collection of geo-tagged photographs, was proposed by Slava Kisilevich *et al.* [19]. Two new ideas were introduced by P-DBSCAN: (i) the density threshold, which is based on the population of the neighborhood, and (ii) adaptive density, which is utilized to quickly converge to high-density areas.

Yongli Zhang *et al.* [20] proposed the ST-DCONTOUR algorithm, a unique serial density-contour-based spatio-temporal clustering technique that uses location streams as input and a model-based clustering approach to produce spatio-temporal clusters. In our method, the incoming data is divided into batches, and we use a serial technique to construct spatial clusters for each batch individually first. The formation of spatio-temporal clusters is then accomplished by establishing ongoing connections between spatial clusters in successive batches. Our method makes use of contouring algorithms to recognize spatial clusters as closed contours of a region whose density exceeds a predetermined threshold and contour analysis methods to recognize persistent, transient, and freshly formed spatial clusters in batches.

Yikai Gong *et al.* [21] proposed a technique called RT-DBSCAN to enable continuous cluster checkpointing-based real-time data clustering. In order to provide scalability, the platform is developed using container-based technologies and Apache Spark running on massive Cloud resources. The real-time data streams were the exclusive focus of the DBSCAN algorithm. With the use of density-based clustering, the DBSCAN algorithm handles rapidly expanding high velocity data streams.

Based on empirical geographical correlations across time, M. Hüsch *et al.* [22] proposed a method called CorClustST. CorClustST compares and interprets clustering outcomes for various scenarios, such as those involving several underlying variables or various time scales. The clustering technique is extended in a way that enables massively parallel execution and works around memory constraints.

C. Choi *et al.* [23] proposed a clustering method, called MDST-DBSCAN for large-scale, multidimensional spatiotemporal data in a reliable and efficient manner. The MDST-DBSCAN is applied to idealized patterns and a real data set, and the results from both examples demonstrate that it can identify clusters accurately within a reasonable amount of time. The MDST-DBSCAN has a limitation in that the method of defining the neighbors in the clustering process operates conservatively.

III. Proposed Method

The spatio-temporal dataset is different from the traditional dataset in many ways. Firstly, it contains implicit spatial or temporal data that initially need to be calculated. Secondly, spatio-temporal data has granularity, and making its selection differently will impact mining results. Thirdly, spatio-temporal data has an auto-correlation.

The ST-DBSCAN algorithm was proposed by Birant and Kut [16] and has been widely utilized in various studies, achieving impressive results on spatio–temporal data. The distance between two objects *o* and *p* denoted as dist *(o, p)* is defined by measures such as the Manhattan, Euclidean and Haversine distance functions. The neighborhood of object o is defined as $\{p \in D \vee dist(o,p) \leq Eps\}$, where p is another object in database D and *Eps* is the predefined minimum distance between object *o* and *p*. Eq. 1 formally defines the neighborhood utilized in ST-DBSCAN as follows: *Eps 1* for spatial data and *Eps 2* for non-spatiotemporal data, whereas Eq. 2 defines the core objects. As per Eq. 2, objects that satisfy the minimum number of neighborhoods within the radius of *Eps* and are greater than or equal to the *MinPts* are defined as core objects. The ST-DBSCAN is also defined by its density-reachable, density-reachable, density-connected, density-based cluster and border objects [16].

The temporal data aspect was separated manually by filtering daily observations in different years. Thus, this method cannot identify clusters that contain patterns of two different years. The neighborhood in Eq. 1 only considers the spatial and non-spatial distances. It does not include the temporal features, as the algorithm uses two boundary neighborhood-clustering methods. Fig. 1 illustrates the limitation of ST-DBSCAN. Only the spatial dimension is considered in the clustering process; thus, only one cluster will be generated. However, the distance becomes apparent if the data is viewed from the right or topside. Hence, the blue and grey boxes will not be grouped into the same cluster.

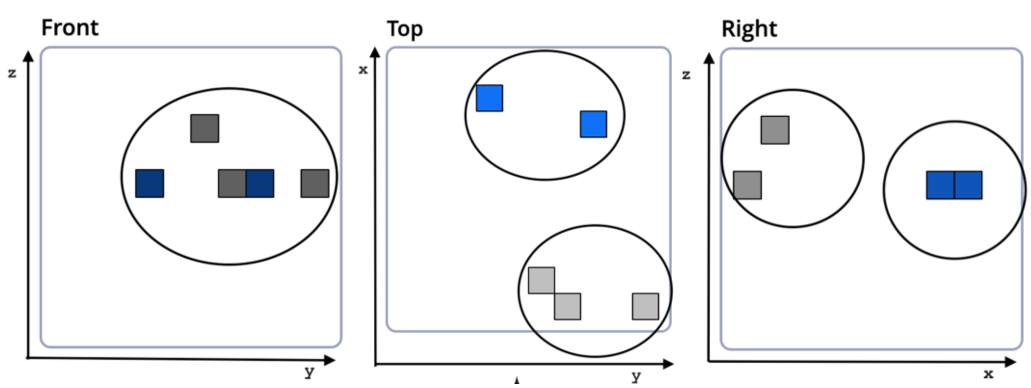


Fig. 1. Two-dimensional view of spatio–temporal data with views from the front, right and top.

The proposed Hierarchical ST-DBSCAN (ST-HDBSCAN) algorithm was developed in two parts. The first part improves the neighborhood of the ST-DBSCAN by introducing Three Neighborhood Boundary. The second part introduces the hierarchical ward's method to improve the performance of DBSCAN.

$$spatial\_dist(o,p) \leq Eps1) \wedge (var\_dist(o,p) \leq Eps2) \rightarrow Neighbor(o,p) \quad (1)$$

$$NumNeighbor(o) \geq MinPts \rightarrow CoreObject(o) \quad (2)$$

A. ST-DBSCAN with Three Neighborhood Boundary

Fig. 2 illustrates the spatio–temporal data in three dimensions to incorporate the temporal aspect. The spatial dimension is represented along the z-axis and y-axis, whereas the temporal dimension is represented along the x-axis. In the three-dimensional view, it becomes clear that when the temporal aspect is taken into consideration, proper separation is achieved: the grey boxes are on the positive side of the x-axis, and the blue boxes are on the negative side of the x-axis. Therefore, ignoring the temporal aspect could lead to inconsistent clustering.

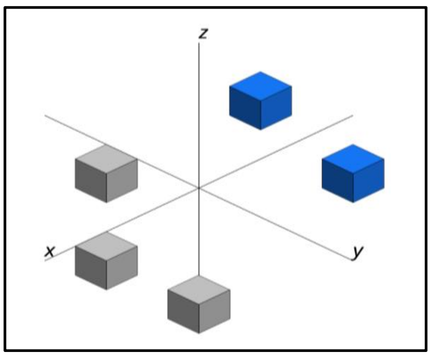


Fig. 2. A three-dimensional view of spatio–temporal data (the spatial dimension is represented along the z-axis and y-axis while the temporal dimension is represented along the x-axis).

Eq. 3 defines the spatio–temporal neighborhood utilized in the improved ST-DBSCAN algorithm. It is composed of Three Neighborhood Boundary: *Eps 1* for spatial data, *Eps 2* for non-spatio-temporal data, and *Eps 3* for temporal data.

In the improved ST-DBSCAN algorithm, temporal neighbors are defined by *Eps 3* to limit temporal distance. As a result, two objects are temporally close if the distance between them in terms of time is in the range of *Eps 3*. For example, the temporal distance could be defined over 1 year or 1 month. If values are collected daily, a temporal distance defined over 1 year will limit the data to the range of a year, that is, 182 days before and after the date for each value. Therefore, to spatially and temporally cluster the data, there is a need to develop a new algorithm that could incorporate the spatial and temporal aspects of data. Based on this information, in this research, the density-based algorithm ST-DBSCAN is enhanced by incorporating a temporal distance limit called maximum temporal distance (*Eps 3*). The maximum temporal distance is defined in Eq. 4.

The proposed approach can be illustrated as follows. Given objects O (*s1*, *s2*, *n1*, *n2*, *t1*) and P (*s3*, *s4*, *n3*, *n4*, *t2*), each comprising of five variables (latitude, longitude, wind speed, sea surface temperature, and date), Eq. 5 (haversine function) is utilized to compute the spatial distance (*spatial_dist)* between O and P and Eq. 6 (Euclidean distance) computes the non-spatiotemporal distance (*non_st_dist*). Also, Eq. 7 computes the absolute date difference between the two objects measured in days. The Euclidean distance was an unsuitable measure for the earth's curved surface due to the utilized spatial data (latitude and longitude coordinates). Therefore, the haversine distance was the preferred option for spatial distance computation.

$$(spatial_{dist}(o,p) \leq Eps1) \wedge (non_{st_{dist}}(o,p) \leq Eps2) \wedge (temp_{dist}(o,p) \leq Eps3) \rightarrow Neighbor(rs,p) \quad (3)$$

$$spatial_{dist} = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{s_3-s_1}{2}\right) + \cos(s_1)\cos(s_3)\sin^2\left(\frac{s_4-s_2}{2}\right)}\right) \quad (4)$$

where:

h: is the haversine distance

r: is the radius of Earth with value 6371 km

φ_1: is the latitude of point 1

φ_2: is the latitude of point 2

λ_1: is the longitude of point 1

λ_2: is the longitude of point 2

$$spatial_{dist} = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{s_3-s_1}{2}\right) + \cos(s_1)\cos(s_3)\sin^2\left(\frac{s_4-s_2}{2}\right)}\right) \quad (5)$$

$$non_{st_{dist}} = \sqrt{(n1-n3)^2 + (n2-n4)^2} \quad (6)$$

$$temp\_dist = |t2 - t1| \quad (7)$$

Pseudocode 1 describes the pseudocode of ST-DBSCAN with maximum temporal distance (ST-DBSCAN *Eps 3*). The algorithm starts by initializing the cluster number as described in line 13. Afterward, it searches for the neighbors of unlabelled data utilizing the retrieve-neighbor function (lines 15–16). The expectation is that if the retrieve-neighbor function meets all the conditions, the objects are considered neighbors. Conversely, if the number of neighbors does not satisfy the *MinPts*, these objects will be labeled as noise (lines 19–20). Otherwise, its neighbors are labeled as the same cluster (lines 23–24). Therefore, the algorithm checks the neighbors of each discovered neighbor (lines 26–29). Similarly, the neighbor would be labeled the same cluster (lines 30–33). These processes are repeated until all objects have been observed. Lastly, the Clustered_Data set, which contains all data labeled with a Cluster ID, is returned as the result of the algorithm (line 42).

Line	Pseudocode 1: Pseudocode of ST-DBSCAN with Eps3 Algorithm

```
1.  Inputs:
2.    D = (d1, d2, … , dn) Set of objects
3.    N = (d1, d2, … , dn) Set of Retrieved Neighbors
4.    R = (r1, r2, … , rn) Set of Retrieved Neighbors of N
5.    Eps1: Maximum spatial distance value
6.    Eps2: Maximum non-spatial distance value
7.    Eps3: Maximum temporal distance value
8.    Mint: Minimum number of points within Eps1, Eps2, and Eps3
9.
10. Output
11.   Clustered_Data : Data that has been labeled with ClusterID
12.
13. Initialize cluster label as 0
14.
15. For i in range of number of data rows in D
16.   If label of the current data with index i set as "unmarked"
17.     Retrieve neighbors of data with index i based on Eps1, Eps2, and Eps3 values
18.     Store the neighbors in N
19.     If number of neighbors are less than MinPts
20.       Set label of data with index i as "noise"
21.     Else
22.       Create new cluster label
23.       For all neighbors in N
24.         Set cluster label for each neighbor as new cluster label
25.       End For
26.       For all neighbors in N
27.         Get one neighbor data and store it as current object
28.         Retrieve neighbors of current object based on Eps1, Eps2, and Eps3
29.         Store the neighbors of current object in R
30.         If number of neighbors in R less than or equal to MinPts
31.           For all neighbors in R
32.             If the neighbor label's is not "noise" or "unmarked"
33.               Set cluster label for the current object as new cluster label
34.             End If
35.           End For
36.         End If
37.       End For
38.     End Else
39.   End If
40. End For
41.
42. Return data D with cluster label as Clustered_Data
```

Pseudocode 2 shows the pseudocode of the retrieve-neighbor function. This function returns the neighbors of an object by computing the spatial, temporal, and non-spatiotemporal distance between the object and other objects in the dataset. The algorithm starts by excluding the current object from the data (lines 12–13). It then computes the spatial distance between the two objects using the Haversine function (Eq. 5) and stores the distance in the spatial_dist variable (lines 15–16). Afterward, it computes the non-spatiotemporal distance utilizing the Euclidean distance function (Eq. 6) and stores the result in the non_st_dist variable (lines 17–18). Finally, it calculates the temporal distance by calculating the date differences between the current object and other objects in the dataset (Eq. 7) and stores it in temporal_dist (lines 19–21).

The next step determines the neighbors of each object in the dataset in turn by comparing the spatial_dist, non_st_dist and temporal_dist with the values of *Eps 1*, *Eps 2* and *Eps 3*. For each evaluation, if each distance value is less than the corresponding *Eps* value, the current object (*p*) is considered a neighbor of the examined object (*o*) and will be added to the set of the object's neighbors (Data_Neighbors) (lines 22–26). The Data_Neighbors is returned to the ST-DBSCAN *Eps 3* Algorithm (line 30).

Pseudocode 2: Pseudocode of the Retrieve Neighbors for ST-DBSCAN with maximum temporal distance (Eps 3)

```
Inputs:
  D     : Data of all objects
  n     : Index of current object
  Eps1: Maximum spatial distance value
  Eps2: Maximum non-spatial distance value
  Eps3: Maximum temporal distance value

Output
  Data_neighbors : Data of current object's Neighbors

For i in range of data row in D
  If i is equal to n
    Continue the program
  Else
    Calculate spatial distance of object index i with object index n using Harversine function
    Store the spatial distance as spatial_dist
    Calculate non-spatio-temporal distance of object i with object index n using Euclidean function
    Store the non-spatio-temporal distance as non_st_dist
    Calculate temporal distance of the object index i with object index n by
    subtracting time of object index i with object index n
    Store the temporal distance as temporal_dist
    If spatial_dist is less than Eps1 and
      non_st_dist is less than Eps2 and
      temporal_dist less than Eps3
        Set object index i as neighbor of object index n
        Add object index i to Data_neighbors
    End If
  End If
End For
Return Data_neighbors
```

B. *An Improved ST-DBSCAN with Hierarchical Ward's Method*

ST-HDBSCAN improves the ST-DBSCAN algorithm by utilizing the hierarchical ward's method. Pseudo code 3 shows the ST-HDBSCAN algorithm, which consists of the clustering and hierarchical phases. Lines 15–41 outline the clustering procedure with *Eps 3*. After clustering, the hierarchical ward's method is employed to aggregate the results (line 50). It combines close and similar clusters into new larger ones based on the temporal distance value (computed by fast dynamic time warping) (lines 43–48). For the hierarchical phase, the Fast Dynamic Time Warping (FastDTW) algorithm warps the timeline of data points into compressed values (line 47) to simplify hierarchical cluster grouping. After computing the similarity pair of each cluster, the distance is condensed to minimize duplication (line 49). The minimum distance is selected as the cut point of the hierarchy to obtain the proper clusters (line 51).

Pseudocode 3: Pseudocode of the ST-HDBSCAN Algorithm

```
1.  Inputs:
2.    D = (d1, d2, … , dn) Set of objects
3.    N = (d1, d2, … , dn) Set of Retrieved Neighbors
4.    R = (r1, r2, … , rn) Set of Retrieved Neighbors of N
5.    Eps1: Maximum spatial distance value
6.    Eps2: Maximum non-spatial distance value
7.    Eps3: Maximum temporal distance value
8.    Mint: Minimum number of points within Eps1, Eps2, and Eps3
9.
10. Output
11.   Clustered_Data : Data that has been labeled with ClusterID
12.
13. Initialise cluster label as 0
14.
15. For i in range of number of data rows in D
16.   If label of the current data with index i set as "unmarked"
17.     Retrieve neighors of data with index i based on Eps1, Eps2, and Eps3 values
18.     Store the neighbors in N
19.     If number of neighbors are less than MinPts
20.       Set label of data with index i as "noise"
21.     Else
```

```
                Create new cluster label
                For all neighbors in N
                  Set cluster label for each neighbor as new cluster label
                End For
                For all neighbors in N
                  Get one neighbor data and store it as current object
                  Retrieve neighbor of current object based on Eps, Eps2, and Eps3
                  Store the neighbor of current object in R
                  If number neighbor in R less than or equal to MinPts
                    For all neighbor in R
                      If the neighbor label's is not "noise" or "unmarked"
                        Set cluster label for the current object as new cluster label
                      End if
                    End for
                  End if
                End for
              End Else
            End If
          End For
          Return data D with cluster label as Clustered_Data

          For i in range of number of cluster in Clustered_Data
            For j in range of number of cluster in Clustered_Data
              Get non spatial data with ClusterID i
              Get non Spatial data with ClusterID j
              Calculate FastDTW time series distance between cluster1 and cluster2
              Store the calculated value in variable 'dist'
          Convert pair distance 'dist' into condensed distance
          Generate hierarchy with ward method
          Get new cluster label from hierarchy with cutting level of threshold value
          For i in range of Clustered_Data
            Replace cluster label with the new cluster label from hierarchy
          Return Cluster_Data with new label
```

IV. Experiment Results and Discussion

A. Equations

The El-Niño dataset from the UCI Repository was utilized for this research. This dataset was provided to study the El-Niño/Southern Oscillation cycle phenomena, which is well-known in climatology as the cause of the climate anomalies worldwide [26]. The data was collected from 1980 to 1998 and comprised different observations for each buoy. In total, there are 75 buoys and 178,080 observations in the dataset. The dataset was pre-processed using several techniques such as data cleaning, data transformation, and data reduction. The data cleaning technique employed the K-Nearest Neighbor Algorithm as the imputation technique. A new variable was added for data transformation, and Min-Max normalization was implemented. Since the wind speed value is vital in determining El-Niño cycles, a new variable, 'Wind Speed,' is also estimated using the Zonal and Meridional wind values. On the other hand, air temperature and humidity were removed from the dataset because air temperature values are similar to sea surface temperature. It does not contribute much to determine El-Niño because 36% were missing the humidity values.

B. Parameter Setting

Table I shows the parameter settings used in the two experiments. The settings are as follows. Experiment 1 improves the ST-DBSCAN algorithm with the maximum temporal distance, whereas Experiment 2 improves it with the hierarchical ward's method.

TABLE I. PARAMETERS SETTINGS OF EXPERIMENT 1 AND 2

Parameters	**Experiment 1**		**Experiment 2**	
	ST-DBSCAN	***ST-DBSCAN Eps 3***	***ST-HDBSCAN 0.3***	***ST-HDBSCAN 0.1***
Eps 1 (spatial)	5000	5000	5000	5000
Eps 2 (non-spatio–temporal)	0.3	0.3	0.3	0.3
Minimum points	Log(178080) = 12.0899	Log(178080) = 12.0899	Log(178080) = 12.0899	Log(178080) = 12.0899
Eps 3 (temporal)	N/A	182	182	182
Threshold value	-	-	0.3	0.1

C. ST-DBSCAN with Three Neighbourhood Boundary Result

Table II and Table III show the clustering results for ST-DBSCAN compared with ST-DBSCAN Eps 3. The results indicate that the ST-DBSCAN algorithm does not consider the dataset's temporal aspect. Obs 87567–174633 was recognized as one cluster: Cluster 6. Its similarity term was based on the spatial distance (latitude, longitude) and non-spatio–temporal distance (wind speed and sea surface temperature) only. Similarly, in Table III, Obs 87799–151340 was grouped into Cluster 4 only. On the other hand, ST-DBSCAN *Eps 3* grouped Obs 87567–174633 into two clusters; 233 (Obs 174633–177614) and 166 (Obs 87230–87567), as shown in Table II. In Table III, Obs 87799–151340 was grouped into four clusters, which are 217, 220, 228, and 166. The results show that ST-DBSCAN *Eps 3* clusters the data according to the temporal distance. It separated observation 177614, collected on November 5, 1998, from observation 87230, collected on April 5, 1994, which have about a four-year difference (Table II).

TABLE II. SAMPLE DATA 1 CLUSTER RESULT FOR ST-DBSCAN COMPARED WITH OTHER ALGORITHMS

Obs	Date	Lati-tude	Longi-tude	Wind Speed	Sea Surface Temp	Buoy	ST-DBSCAN Eps 3	STHDB SCAN (0.1)	STHDB SCAN (0.3)	ST-DBSCAN
174633	21/05/1998	1	−95.02	0.8591	0.4721	28	233	9	2	6
174692	22/05/1998	1	−139.96	0.7599	0.3475	20	233	9	2	6
174750	23/05/1998	1	−109.96	0.8296	0.4640	38	233	9	2	6
174930	26/05/1998	1	−139.88	0.8052	0.4815	41	233	9	2	6
174932	26/05/1998	1	−94.95	0.8922	0.1858	48	233	9	2	6
175050	28/05/1998	−1	−95.07	0.7038	0.3980	57	233	9	2	6
175161	30/05/1998	1	−109.96	0.7656	0.4383	38	233	9	2	6
175273	01/06/1998	1	−125.05	0.8303	0.5245	18	233	9	2	6
175584	06/06/1998	−1	−110	0.6650	0.3932	49	233	9	2	6
175736	15/06/1998	1	−95.04	0.7965	0.5333	28	233	9	2	6
175890	20/06/1998	1	−94.95	0.8520	0.5123	67	233	9	2	6
177012	01/10/1998	−1	−95.1	0.7074	0.5374	74	233	9	2	6
177614	05/11/1998	−1	−95.1	0.7664	0.6282	74	233	9	2	6
87230	29/04/1994	1	−179.9	0.7779	0.5883	47	166	17	3	6
87232	29/04/1994	1	−124.98	0.8210	0.3705	39	166	17	3	6
87234	29/04/1994	−1	−154.99	0.8699	0.4464	52	166	17	3	6
87288	30/04/1994	1	−110.1	0.6636	0.3485	17	166	17	3	6
87459	03/05/1994	1	−125.02	0.6858	0.3758	18	166	17	3	6
87460	03/05/1994	1	−179.85	0.7786	0.3967	27	166	17	3	6
87461	03/05/1994	1	−179.89	0.7994	0.3079	47	166	17	3	6
87462	03/05/1994	0.04	−124.35	0.6190	0.2969	2	166	17	3	6
87513	04/05/1994	0.04	−139.99	0.7297	0.4579	3	166	17	3	6
87567	05/05/1994	1	−140.26	0.7311	0.5450	0	166	17	3	6

TABLE III. SAMPLE DATA 2 CLUSTER RESULT FOR ST-DBSCAN COMPARED WITH OTHER ALGORITHMS

Obs	Date	Lati-tude	Longi-tude	Wind Speed	Sea Surface Temp	Buoy	ST-DBSCAN Eps 3	STHDB SCAN (0.1)	STHDBS CAN (0.3)	ST-DBSCAN
151340	01/04/1997	1	−125.01	0.6679	0.3474	61	217	15	2	4
153131	01/05/1997	1	156.01	0.9001	0.3355	23	220	11	2	4
153133	01/05/1997	−1	164.86	0.8605	0.1847	71	220	11	2	4
153134	01/05/1997	−1	164.33	0.8807	0.3707	34	220	11	2	4
154959	01/06/1997	1	147.05	0.8648	0.2051	42	220	11	2	4
156725	01/07/1997	1	147.05	0.8562	0.2225	42	220	11	2	4
158575	01/08/1997	−1	156.01	0.8706	0.3821	53	220	11	2	4
160492	01/09/1997	1	164.98	0.8850	0.1768	25	220	11	2	4
162354	01/10/1997	1	−179.86	0.8713	0.2448	47	228	8	2	4
164284	01/11/1997	−1	−170.03	0.7692	0.3121	35	228	8	2	4
87679	07/05/1994	−0.5	166.92	0.8534	0.1437	16	166	17	3	4
87681	07/05/1994	1	−154.93	0.8296	0.4567	43	166	17	3	4
87682	07/05/1994	−1	−170.03	0.8454	0.3735	35	166	17	3	4
87742	08/05/1994	1	−125.07	0.6147	0.5030	18	166	17	3	4
87743	08/05/1994	−0.03	−170.02	0.8541	0.3364	12	166	17	3	4
87744	08/05/1994	1	−170.04	0.8864	0.3627	26	166	17	3	4
87745	08/05/1994	1	−154.92	0.8224	0.3427	43	166	17	3	4
87746	08/05/1994	0.04	−139.97	0.6276	0.1990	3	166	17	3	4
87799	09/05/1994	1	−179.89	0.8929	0.1243	47	166	17	3	4

A heatmap can be utilized to visualize the performance of the algorithms. Fig. 3 shows the heatmap of the 14 clusters by ST-DBSCAN. Each color represents a different cluster, and each line represents the timeline of a buoy. Each buoy changes its cluster membership accordingly in its timeline, represented by the changing color in the buoy timeline. The heatmap shows that the clusters generated by ST-DBSCAN do not separate the data based on time. The small clusters indicate this problem with various colors spread across the timeline. This problem shows that clusters contain data with similar values but could occur at any time.

Fig. 4 shows the heatmap of the ST-DBSCAN *Eps 3* algorithm. The Fig. 4 clearly shows the separation of data based on time, although 240 clusters were generated in the process. Though the cluster areas are enormous, the members in each cluster are small. However, the clusters are well-ordered based on time. However, since the number of clusters is enormous, it is difficult to interpret the data pattern, and each cluster has a tiny number of members. Therefore, to reduce the number of clusters and increase the understanding of the patterns generated, an improvement is conducted by aggregating the similar clusters based on the temporal similarity of the clusters. This technique was previously employed in ST-OPTICS, where the algorithm combined the density-based and hierarchical-based methods to cluster spatio–temporal data [17].

D. ST-HDBSCAN Result

The ST-HDBSCAN algorithm was utilized to minimize the number of clusters generated by ST-DBSCAN with *Eps 3*. The clusters generated from the previous experiment were aggregated according to their similarity to form a hierarchy of clusters. FastDTW was employed as the cluster aggregation function. If the FastDTW value is small, the cluster is considered similar and is grouped into a larger cluster.

The cut point of the dendrogram decides the number of clusters generated by ST-HDBSCAN. The higher the cut point affect to the fewer clusters generated. The number of vertical lines at the different cut-point levels indicates the number of clusters. Fig. 5 shows the hierarchical clustering dendrogram for the ST-HDBSCAN algorithm using the hierarchy threshold (0.1). The cut point of the dendrogram is calculated by multiplying the threshold with the maximum hierarchy distance or maximum FastDTW distance of all clusters. For the data utilized in our experiment, the maximum distance was 249.95.

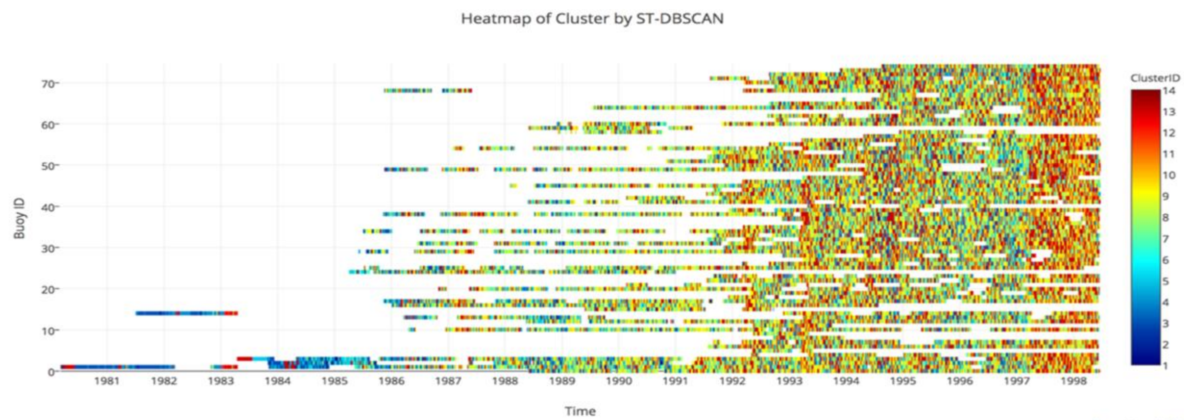


Fig. 3. The cluster heatmap of ST-DBSCAN.

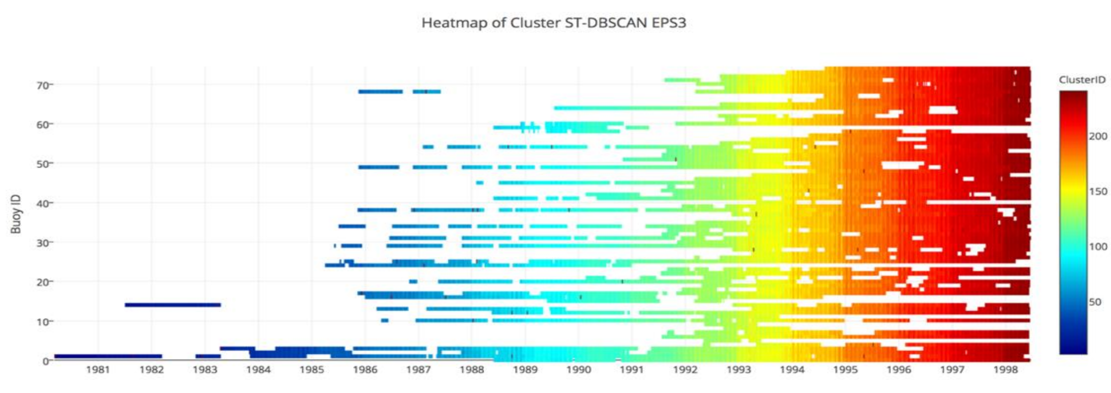


Fig. 4. The cluster heatmap of ST-DBSCAN with Eps 3.

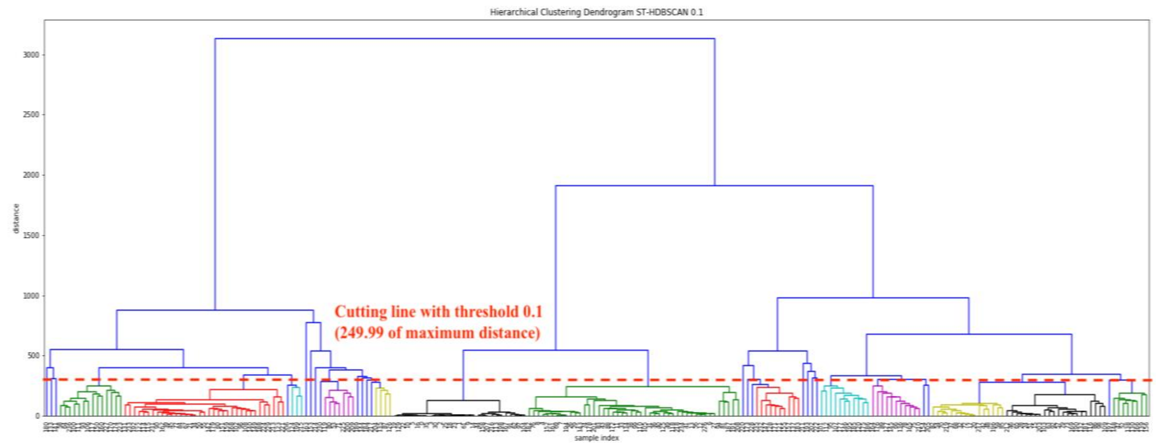


Fig. 5. Hierarchical Clustering Dendrogram ST-HDBSCAN 0.1.

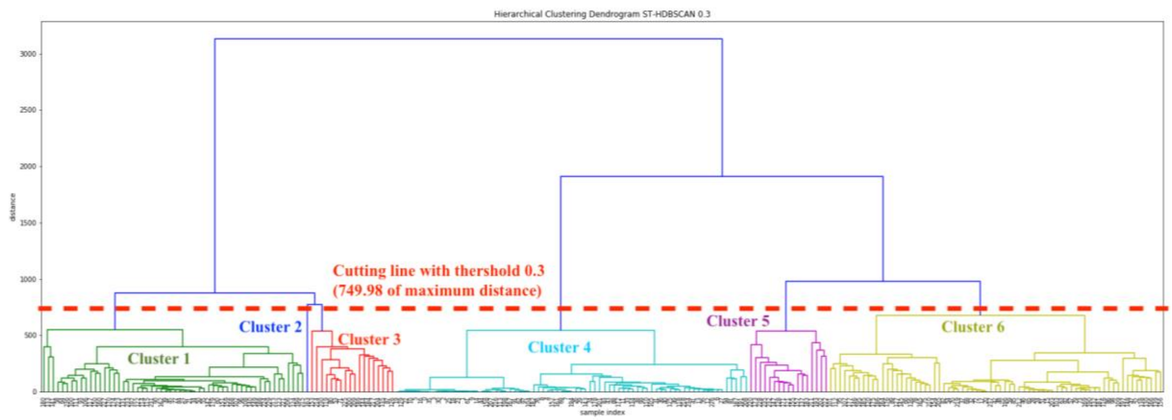


Fig. 6. Hierarchical Clustering Dendrogram ST-HDBSCAN 0.3.

On the other hand, Fig. 6 shows the cut point hierarchy for the threshold (0.3). It achieved a maximum distance of 749.98 and generated six clusters.

It is evident from Fig. 5 and 6 that the cluster colors are similar because the groups are the same; only the level of detail is different. Also, the result of the ST-HDSBCAN algorithm with the threshold (0.1) is more detailed. It is observed when the smaller cluster size and more significant cluster number in Fig. 5 compared with Fig. 6. However, the result of ST-HDBSCAN with a threshold (0.3) has interesting cluster patterns.

Furthermore, Fig. 7 shows the heatmap of ST-HDBSCAN with a threshold (0.3). It indicates how each cluster is distinguished. This pattern appears at a particular timeline, identified by 'ClusterID 3' in cyan color. This distinctive cluster occurred between 1994 and 1995. According to the Earth System Research laboratory, one of the National Oceanic and Atmospheric Administration research centers, the El-Niño that occurred within the above specified years is considered one of the top 24 strongest between 1895 and 2015. It is also regarded as one of the new El-Niño events that warmed the central–equatorial area of the Pacific Ocean rather than the eastern part of the Pacific Ocean.

It is evident from the heatmaps of ST-HDBSCAN with a threshold (0.3) that the clusters are spread across the timeline, except for Cluster 3. Cluster 1 (dark blue) mostly appears between 1990 to 1995 and starts to disappear after a few years. On the other hand, cluster 2 (light blue) appears close to Cluster 1 and spreads through the entire timeline. Cluster 4 (yellow color) appears mostly early in the timeline but starts to disappear between 1990 to 1995. It is replaced by Cluster 6 (brown color), although Cluster 4 reappears in the following years. Cluster 5 only appeared in 1990, and its effect on the environment was significant in later years. Cluster 6 appears in the early years of the timeline but mostly has a strong presence between 1990 to 1995. It gradually diminishes, and it completely vanishes in early 1997.

Fig. 8 shows the heatmap of the ST-HDBSCAN of threshold (0.1). The heatmap indicates a similar pattern to that of Fig. 7. Furthermore, the clusters are not as clear compared with the ST-HDBSCAN threshold (0.3). A cluster 1 to 6 mostly appears from 1990 to 1995, whereas Cluster 7 to 9 appears after. Clusters 1 to 6 shows a resurgence in 1998. This pattern is similar to Cluster 1 and 2 in ST-HDBSCAN of threshold (0.3).

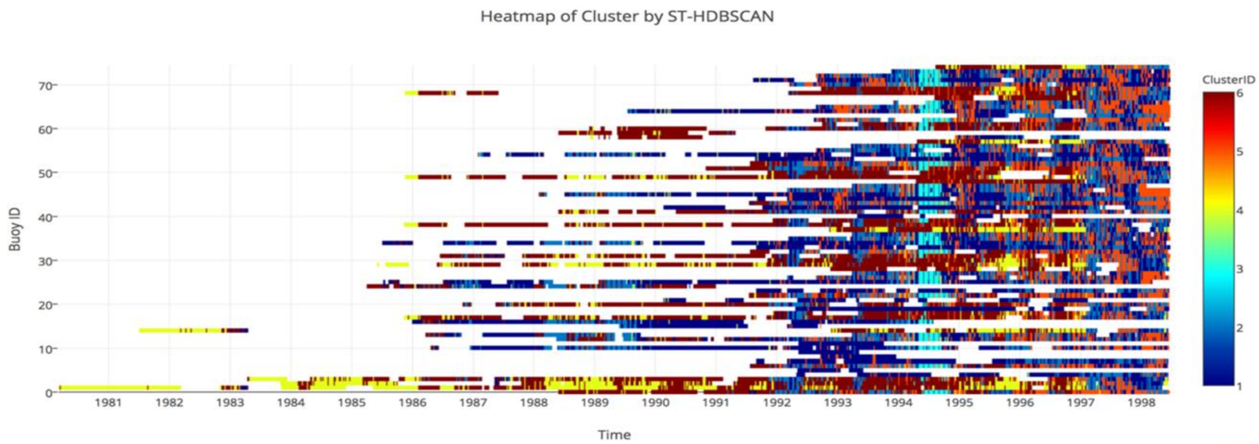


Fig. 7. Seven cluster 3 heatmaps of Cluster ST-HDBSCAN with Threshold (0.3).

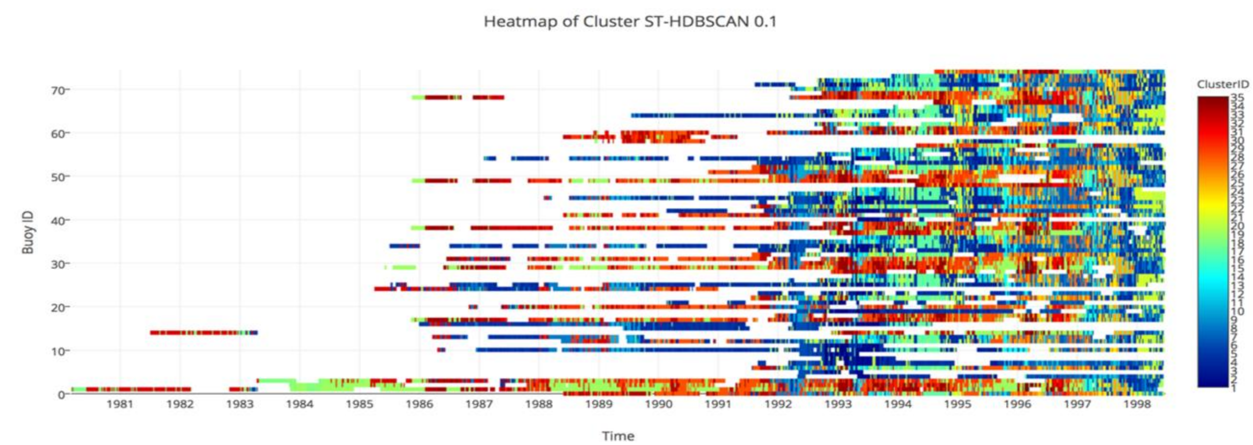


Fig. 8. Cluster heatmap of ST-HDBSCAN with Threshold (0.1).

Similar to Clusters 4 and 6 in ST-HDBSCAN of threshold (0.3), Clusters 16 to 19 start appearing early (1981) and are dominant from 1994 to 1995. Likewise, Clusters 28 to 30 appear early (1988), but they appear primarily between 1990 and 1994. They are replaced by Clusters 24 to 26 but reappear in early 1997 and are replaced again by Clusters 24 to 26. However, other clusters generally spread across the timeline and are not noticeable in the heatmap.

E. Performance Indices of the Proposed Method

Table IV shows six performance indices for ST-HDBSCAN, ST-DBSCAN, and ST-DBSCAN with Eps 3. The high and low values differ according to the type of performance indices method, as indicated in column Best If Value. For example, in the performance index Ball-Hall, the higher value indicates the best performance, whereas Det Ratio is in direct contrast to it. The result specifies that ST-HDBSCAN with a threshold (0.3) has the best performance indices values for all performance indices except for KsqDetW. Table V specifies the comparison of the performance indices of ST-DBSCAN with ST-HDBSCAN 0.3. From Table V, it is evident that ST-HDBSCAN with 0.3 improves the performance indices of ST-DBSCAN by more than 40%. Similarly, the GDI51 and Trace W performance indices of ST-HDBSCAN 0.3 improve ST-DBSCAN by more than 100%.

The negative percentage of Det Ratio and Log Det Ratio were converted to positive values to calculate the average percentage improvement. This conversion is because a negative percentage indicates that the index is getting lower. Note that lower indices imply better performance indices. The average percentage improvement in the Dat Ratio index by ST-HDBSCAN 0.3 is 75%. Thus, adding the maximum temporal distance (*Eps 3*) and hierarchical aggregation using the ward's method improves all performance indices. However, this is not the case for the KsqDetW index.

The KsqDetW index value is lower because the number of clusters generated by ST-HDBSCAN (6 clusters) is smaller than that by ST-DBSCAN (14 clusters). Given that the function to evaluate the KsqDetW is highly dependent on the number of clusters, a smaller number of clusters results in lower KsqDetW values. Eq. 8 evaluates the value of KsqDetW, where K is the number of clusters and WG is the sum of the within-group scatter matrix.

$$KsqDetW = K^2 det(WG) \quad (8)$$

TABLE IV. PERFORMANCE INDICES COMPARISON OF ST-DBSCAN, ST-DBSCAN WITH EPS 3, ST-HDBSCAN 0.1 AND ST-HDBSCAN 0.3

Indices	Best If Value	ST-DBSCAN	ST-DBSCAN Eps 3	ST-HDBSCAN 0.1	ST-HDBSCAN 0.3
Ball-Hall	high	0.01615323	0.0157	0.02011439	**0.0230969**
Det Ratio	low	14.4842	6.41	4.69	**3.55**
GDI51	high	0.0331	0.0315	0.0779143	**0.0944**
KsqDetW	high	1.69E+08	**1.11E+11**	3.27E+09	1.27E+08
Log Det Ratio	low	476018.2	330972	275336.7	**225691**
Trace W	high	1904.39	2951	3386.383	**3912**

TABLE V. PERFORMANCE INDICES COMPARISON OF ST-DBSCAN AND ST-HDBSCAN 0.3

Indices	Best If Value	ST-DBSCAN	ST-HDBSCAN 0.3	Difference	Percentage
Ball-Hall	high	0.01615323	**0.0230969**	0.00694367	43%
GDI51	high	0.0331	**0.0944**	0.0613	185%
KsqDetW	high	1.69E+08	**1.27E+08**	-42000000	25%
Trace W	high	1904.39	**3912**	2007.61	105%
Det Ratio	low	14.4842	**3.55**	-10.9342	75%
Log Det Ratio	low	476018.2	**225691**	-250327.2	53%

V. CONCLUSION AND FUTURE WORK

The proposed ST-HDBSCAN effectively incorporates the spatial and non-temporal data aspects in clustering. Likewise, it utilizes the temporal aspect of data in the clustering process, thus significantly improving the clustering performance indices. The algorithm introduces an additional temporal distance boundary (*Eps 3*), known as the Three Neighborhood boundary. ST-HDBSCAN further improves the clustering performance by employing the hierarchical ward's method and fast dynamic time warping. We made comparisons with the ST-DBSCAN algorithm utilizing heatmaps and some performance indices. ST-HDBSCAN outperformed by 185% and indicated interesting patterns with spatio–temporal data. Several directions in future work are possible such as adapting the algorithm to process streaming data, indexing spatial and temporal aspects of data, and implementing parallel computing to improve the efficiency of the ST-HDBSCAN algorithm. Also, implementing the experiments for other spatio–temporal datasets in other fields such as criminology, medical and social media will also be beneficial to discovering hidden knowledge in the data.

ACKNOWLEDGMENT

This work was supported by the Universiti Kebangsaan Malaysia under Grant FRGS/1/2019/ICT02/UKM/02/7.

REFERENCES

[1] U. Khalil, O. Ahmed Malik, D. T. Ching Lai, and O. Sok King, "Cluster Analysis for Identifying Obesity Subrouops in Health and Nutritional Status Survey Data," Asia-Pacific J. Inf. Technol. Multimed., vol. 10, no. 02, pp. 146–169, Dec. 2021, doi: https://doi.org/10.17576/apjitm-2021-1002-11.

[2] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-Temporal Data Mining," ACM Comput. Surv., vol. 51, no. 4, pp. 1–41, Jul. 2019, doi: https://doi.org/10.1145/3161602.

[3] A. Madraky, Z. A. Othman, and A. R. Hamdan, "Hair-oriented data model for spatio-temporal data representation," Expert Syst. Appl., vol. 59, pp. 119–144, Oct. 2016, doi: https://doi.org/10.1016/j.eswa.2016.04.028.

[4] V. Bogorny and S. Shekhar, "Spatial and Spatio-temporal Data Mining," in 2010 IEEE International Conference on Data Mining, Dec. 2010, pp. 1217–1217. doi: https://doi.org/10.1109/ICDM.2010.166.

[5] J. D. Mazimpaka and S. Timpf, "Trajectory data mining: A review of methods and applications," J. Spat. Inf. Sci., no. 13, Dec. 2016, doi: https://doi.org/10.5311/JOSIS.2016.13.263.

[6] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-Temporal Data Mining: A Survey of Problems and Methods," ACM Comput. Surv., vol. 51, no. 4, pp. 1–41, Jul. 2019, doi: https://doi.org/10.1145/3161602.

[7] H. Mahidin et al., "Spatio-temporal of surface ozone (O3) variations at urban and suburban sites in Sarawak region of Malaysia," in IOP Conference Series: Earth and Environmental Science, Oct. 2021, vol. 880, no. 1, p. 012004. doi: https://doi.org/10.1088/1755-1315/880/1/012004.

[8] M. Tonini, M. G. Pereira, J. Parente, and C. Vega Orozco, "Evolution of forest fires in Portugal: from spatio-temporal point events to smoothed density maps," Nat. Hazards, vol. 85, no. 3, pp. 1489–1510, Feb. 2017, doi: https://doi.org/10.1007/s11069-016-2637-x.

[9] N. Ahmed, M. A.-A. Hoque, B. Pradhan, and A. Arabameri, "Spatio-Temporal Assessment of Groundwater Potential Zone in the Drought-Prone Area of Bangladesh Using GIS-Based Bivariate Models," Nat. Resour. Res., vol. 30, no. 5, pp. 3315–3337, Oct. 2021, doi: https://doi.org/10.1007/s11053-021-09870-0.

[10] U. M. Butt et al., "Spatio-Temporal Crime Predictions by Leveraging Artificial Intelligence for Citizens Security in Smart Cities," IEEE Access, vol. 9, pp. 47516–47529, 2021, doi: https://doi.org/10.1109/ACCESS.2021.3068306.

[11] B. Priambodo, A. Ahmad, and R. A. Kadir, "Spatio-temporal K-NN prediction of traffic state based on statistical features in neighbouring roads," J. Intell. Fuzzy Syst., vol. 40, no. 5, pp. 9059–9072, Apr. 2021, doi: https://doi.org/10.3233/JIFS-201493.

[12] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal clustering," in Data Mining and Knowledge Discovery Handbook, Boston, MA: Springer US, 2009, pp. 855–874. doi: https://doi.org/10.1007/978-0-387-09823-4_44.

[13] M. Y. Ansari, A. Ahmad, S. S. Khan, G. Bhushan, and Mainuddin, "Spatiotemporal clustering: a review," Artif. Intell. Rev., vol. 53, no. 4, pp. 2381–2423, Apr. 2020, doi: https://doi.org/10.1007/s10462-019-09736-1.

[14] A. Madraky, Z. A. Othman, and A. R. Hamdan, "Analytic Methods for Spatio-Temporal Data in a Nature-Inspired Data Mode," Int. Rev. Comput. Softw., vol. 9, no. 3, pp. 547–556, 2014.

[15] Z. A. Othman, A. A. Bakar, A. M. Adabashi, and Z. Muda, “A Similarity Normal Clustering Labelling Algorithm for Clustering Network Intrusion Detection,” J. Appl. Sci., vol. 14, no. 10, pp. 969–980, 2014, doi: https://doi.org/10.3923/jas.2014.969.980.

[16] D. Birant and A. Kut, “ST-DBSCAN: An algorithm for clustering spatial–temporal data,” Data Knowl. Eng., vol. 60, no. 1, pp. 208–221, Jan. 2007, doi: https://doi.org/10.1016/j.datak.2006.01.013.

[17] K. P. Agrawal, S. Garg, S. Sharma, and P. Patel, “Development and validation of OPTICS based spatio-temporal clustering technique,” Inf. Sci. (Ny)., vol. 369, pp. 388–401, Nov. 2016, doi: https://doi.org/10.1016/j.ins.2016.06.048.

[18] S. Wang, T. Cai, and C. F. Eick, “New Spatiotemporal Clustering Algorithms and their Applications to Ozone Pollution,” in 2013 IEEE 13th International Conference on Data Mining Workshops, Dec. 2013, pp. 1061–1068. doi: https://doi.org/10.1109/ICDMW.2013.14.

[19] S. Kisilevich, F. Mansmann, and D. Keim, “P-DBSCAN: a density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos,” in Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application - COM.Geo ’10, 2010, p. 1. doi: https://doi.org/10.1145/1823854.1823897.

[20] Y. Zhang and C. F. Eick, “ST-DCONTOUR: a serial, density-contour based spatio-temporal clustering approach to cluster location streams,” in Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming, Oct. 2016, pp. 1–4. doi: https://doi.org/10.1145/3003421.3003429.

[21] Y. Gong, R. O. Sinnott, and P. Rimba, “RT-DBSCAN: Real-Time Parallel Clustering of Spatio-Temporal Data Using Spark-Streaming,” in Computational Science -- ICCS 2018, Cham: Springer International Publishing, 2018, pp. 524–539. doi: https://doi.org/10.1007/978-3-319-93698-7_40.

[22] M. Hüsch, B. U. Schyska, and L. von Bremen, “CorClustST—Correlation-based clustering of big spatio-temporal datasets,” Futur. Gener. Comput. Syst., vol. 110, pp. 610–619, Sep. 2020, doi: https://doi.org/10.1016/j.future.2018.04.002.

[23] C. Choi and S.-Y. Hong, “MDST-DBSCAN: A Density-Based Clustering Method for Multidimensional Spatiotemporal Data,” ISPRS Int. J. Geo-Information, vol. 10, no. 6, p. 391, Jun. 2021, doi: https://doi.org/10.3390/ijgi10060391.

[24] L. Ertöz, M. Steinbach, and V. Kumar, “Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data,” in Proceedings of the 2003 SIAM International Conference on Data Mining, May 2003, pp. 47–58. doi: https://doi.org/10.1137/1.9781611972733.5.

[25] R. Oliveira, M. Y. Santos, and J. M. Pires, “4D+SNN: A Spatio-Temporal Density-Based Clustering Approach with 4D Similarity,” in 2013 IEEE 13th International Conference on Data Mining Workshops, Dec. 2013, pp. 1045–1052. doi: https://doi.org/10.1109/ICDMW.2013.119.

[26] Earth System Research Laboratory, “Top 24 Strongest El Nino and La Nina Event Years by Season,” 2015. https://www.esrl.noaa.gov/psd/enso/climaterisks/years/top24enso.html (accessed Apr. 07, 2022).

Research on Asymmetry of Two-Queue Cycle Query Threshold Service Polling System

Man Cheng[1], Dedu Yin[2], Xinchun Wang[3*]
Department of Physical and Electronics, Chuxiong Normal University
Chuxiong Normal University 546 South Lucheng Road, Chuxiong, Yunnan, China

***Abstract*—Based on establishing the mathematical model of the system provided system parameters, using the discrete-time Markov chain and a function set by a non-negative integer random variable as Probabilistic methods, discrete time variable, two-queue different gated polling system has been fully analyzed, the low- and higher-order properties and cycle period of the system are deduced, and average queue pair length and average waiting delay for message packets are calculated accurately. The simulation experiments agree well with the theoretical calculations. The analysis further deepens the cognition of the asymmetric threshold polling system, lays a foundation for researches on the asymmetric threshold polling system, and has positive significance for a better and more flexible control periodic query polling work system.**

***Keywords*—Discrete time query; asymmetric two-queue; threshold service; second-order characteristic quantity; information packets average delay**

I. Introduction

Due to the difficulty of researches on asymmetric threshold service polling system [1], it generally begins by studying an asymmetric threshold service polling system with two queues. Therefore, the performance [2] of the two-queue asymmetric threshold service system is deeply analyzed and studied in this paper, whose purpose is to further explore the inherent laws [3] of the multi-queue threshold service polling system. As is known to all, the periodic [4] query queuing service system has been widely used in industrial process control [5], communication network technology [6], and computer network technology. However, for the periodic query queuing service system, parameters [7] such as the number of packets of information entering the network system at each moment, the time of packet service and the transfer time are all random variables, so the analysis of its related performance has considerable complexity [8]. Especially in the study of asymmetric systems [9], because there are still some problems in the analysis method [10], the results are local results obtained under certain limited conditions, and some precise parsing is given in the literature [11]. Based on the literature [12], This article uses the discrete-time Markov chain and a function [13] set by a non-negative integer random variable to parse the asymmetric threshold service system of discrete-time type and the double queue of periodic query type, deduces the low- and higher-order characteristic quantities [14] of the system, calculates the average length and cycle time of message grouping, and the mathematical reasoning of the average waiting delay [15]. The simulation experiments are in good agreement with the theoretical numerical calculations [16]. This topic discusses a deeper understanding of the multi-queue asymmetric threshold service polling system, which has positive significance for a more flexible control polling system. The content and scope of this paper further expand the research space of polling system.

II. Theoretical Model of the Asymmetric Threshold Service

A. Physical Model of the Asymmetric Threshold Service

The dual-queue asymmetric [17] threshold service receives service on a first-in, first-out basis in the same queue. Both Queue one and Queue two are running in threshold service policy mode [18]. In the first queue, until the current threshold information packet is serviced, the newly added information packets in the service process should be accumulated until the next polling cycle before service is performed, transition to queue two service after a transition time, and the service of queue two only serves the current threshold information packet until the service [19] is complete. And then, after a transition time, it is switched to another queue for the next round of service. The physical structure [20] of the asymmetric threshold service is shown in Fig. 1.

Buffer queue for any input, the numbers of information packets arriving in the buffer in each unit slot are independent of each other and have the same probability distribution [21]. The probabilities generating function, mean and variance of the distribution of the number of information groups in the queue are $A_i(z_i)$ 、 $\lambda_i = A_i^{'}(1)$ 、 $\sigma_{\lambda_i}^2 = A_i^{''}(1) + \lambda_i - \lambda_i^2$. The probabilities generating function, mean and variance of the transformation time distribution are $R_j(z_j)$ 、 $r_j = R_j^{'}(1)$ 、 $\sigma_{\beta_k}^2 = R_k^{''}(1) + r_k - r_k^2$. The probabilities generating function, mean and variance of service time distribution are $B_k(z_k)$ 、 $\beta_k = B_k^{'}(1)$ 、 $\sigma_{\beta_k}^2 = B_k^{''}(1) + \beta_k - \beta_k^2$. The above $i, j, k = 1,2$. Large buffer storage capacity for each group of queues is enough not to cause packet loss. Services in each queue are done on a first-come, first-served basis.

*Corresponding Author.

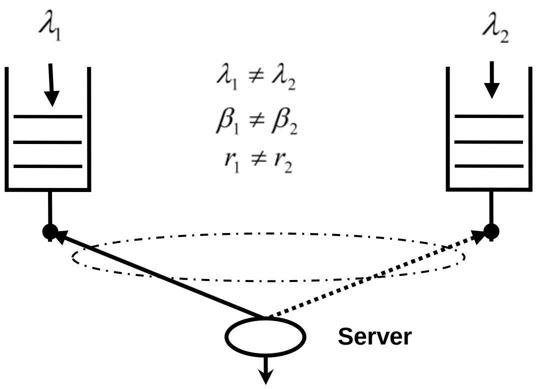


Fig. 1. Physical model of the asymmetric dual-queue threshold service

In order to facilitate the analysis of the queuing system [22], it is necessary to define the following random variables. $u_i(n)$ is defined as the query conversion while for waiter from service site i to service site $i+1$ at t_n. $v_i(n)$ is defined as the waiter starts serving while for the service site i at t_n. The number of information packets[23] entered by the waiter in the 3rd column of the queue is within 3 times。 $\mu_j(u_i)$ is defined as the number of information packets entered by the waiter in the j Queue at u_i. $\eta_j(v_i)$ is defined as the number of information packets entered by the waiter in the service site j within v_i. $\xi_i(n)$ is the number of information packets queued in the service object i at the t_n. For asymmetric two-queues $i, j = 1,2$.

B. State Variable Expression [24]

For a two-queue asymmetric threshold service polling system, when querying the threshold service queue at the t_{n+1}, the relation [25] is as follows:

$$\xi_1(n+1) = 0 + \eta_1(v_1) + \mu_1(u_1) \quad (1)$$

$$\xi_2(n+1) = \xi_2(n) + \eta_2(v_1) + \mu_2(u_1) \quad (2)$$

C. A Function set by a Non-Negative Integer Random Variable of State Variable Distribution [26]

For a two-queue asymmetric threshold service polling system, systematic probability generating function expression is as follows:

$$G_{i+1}(z_1, z_2) = \lim_{n\to\infty} E\left[\prod_{j=1}^{2} z_j^{\xi_j^{(n+1)}}\right] \quad (3)$$

Bring (1), (2) into (3), a function set by a non-negative integer random variable [27] is as follows

$$G_1(z_1, z_2) = R_2(A_1(z_1)A_2(z_2))G_2(z_1, B_2(A_1(z_1)A_2(z_2))) \quad (4)$$

$$G_2(z_1, z_2) = R_1(A_1(z_1)A_2(z_2))G_1(B_1(A_1(z_1)A_2(z_2)), z_2) \quad (5)$$

This $G_1(z_1, z_2)$ in (4) is a function set by a non-negative integer random variable of state variable probability distribution of information packet in threshold 1st Queue. This $G_2(z_1, z_2)$ in (5) is a function set by a non-negative integer random variable of state distribution of information packet in threshold 2nd Queue.

III. Theoretical Analysis of System Performance Indicators

A. Parse Low-Order Characteristic Quantities Expression [28]:

$$g_i(j) = \lim_{z_1, z_2 \to 1} \frac{\partial G_i(z_1, z_2)}{\partial z_j} \quad i = 1,2;\ j = 1,2 \quad (6)$$

The following expression [29] can be obtained by parsing (4), (5) and (6).

$$g_1(1) = r_2\lambda_1 + g_2(1) + g_2(2)\beta_2\lambda_1 \quad (7)$$

$$g_1(2) = r_2\lambda_2 + g_2(2)\beta_2\lambda_2 \quad (8)$$

$$g_2(1) = r_1\lambda_1 + g_1(1)\beta_1\lambda_1 \quad (9)$$

$$g_2(2) = r_1\lambda_2 + g_1(2) + g_1(1)\beta_1\lambda_2 \quad (10)$$

The following expression can be obtained by parsing (8) and (9).

$$g_1(2) = \frac{(r_2 - \rho_1 r_2 + \rho_2 r_1)\lambda_2}{1 - \rho_1 - \rho_2} \quad (11)$$

$$g_2(1) = \frac{(r_1 - \rho_2 r_1 + \rho_1 r_2)\lambda_1}{1 - \rho_1 - \rho_2} \quad (12)$$

The following expression is average length of queue by bringing (11) and (12) into (7) and (10).

$$g_1(1) = \frac{(r_1 + r_2)\lambda_1}{1 - \rho_1 - \rho_2} \quad (13)$$

$$g_2(2) = \frac{(r_1 + r_2)\lambda_2}{1 - \rho_1 - \rho_2} \quad (14)$$

B. Resolve Two-Order Characteristics [30] Define:

$$g_i(j,k) = \lim_{z_1, z_2 \to 1} \frac{\partial^2 G_i(z_1, z_2)}{\partial z_j z_k} \quad i = 1,2;\ j = 1,2; k = 1,2 \quad (15)$$

The equation can be obtained by reduction and resolving (4) and (5) according to (15).

$$g_1(1,1) = \lambda_1^2 R_2''(1) + \left(r_2 + \beta_2 g_2(2)\right) A_1''(1) + \lambda_1^2 g_2(2) B_2''(1) + 2 r_2 \lambda_1 g_2(1) + 2\beta_2 r_2 \lambda_1^2 g_2(2) + g_2(1,1) + 2\beta_2 \lambda_1 g_2(1,2) + \beta_2^2 \lambda_1^2 g_2(2,2) \tag{16}$$

$$g_2(2,2) = \lambda_2^2 R_1''(1) + \left(r_1 + \beta_1 g_1(1)\right) A_2''(1) + \lambda_2^2 g_1(1) B_1''(1) + 2 r_1 \lambda_2 g_1(2) + 2\beta_1 r_1 \lambda_2^2 g_1(1) + g_1(2,2) + 2\beta_1 \lambda_2 g_1(1,2) + \beta_1^2 \lambda_2^2 g_1(1,1) \tag{17}$$

$$g_1(1,2) = g_1(2,1) = \lambda_1 \lambda_2 R_2''(1) + \lambda_1 \lambda_2 g_2(2) B_2''(1) + r_2 \lambda_1 \lambda_2 + r_2 \lambda_2 g_2(1) + (2r_2 + 1)\rho_2 \lambda_1 g_2(2) + \rho_2 g_2(1,2) + \rho_2 \beta_2 \lambda_1 g_2(2,2) \tag{18}$$

$$g_2(2,1) = g_2(1,2) = \lambda_1 \lambda_2 R_1''(1) + \lambda_1 \lambda_2 g_1(1) B_1''(1) + r_1 \lambda_1 \lambda_2 + r_1 \lambda_1 g_1(2) + (2r_1 + 1)\rho_1 \lambda_2 g_1(1) + \rho_1 g_1(2,1) + \rho_1 \beta_1 \lambda_2 g_1(1,1) \tag{19}$$

$$g_1(2,2) = \lambda_2^2 R_2''(1) + \left(r_2 + \beta_2 g_2(2)\right) A_2''(1) + \lambda_2^2 B_2''(1) + 2\rho_2 r_2 \lambda_2 g_2(2) + \rho_2^2 g_2(2,2) \tag{20}$$

$$g_2(1,1) = \lambda_1^2 R_1''(1) + \left(r_1 + \beta_1 g_1(1)\right) A_1''(1) + \lambda_1^2 B_1''(1) + 2\rho_1 r_1 \lambda_1 g_1(1) + \rho_1^2 g_1(1,1) \tag{21}$$

The higher-order feature [31] can be obtained by solving (16), (17), (18), (19), (20) and (21).

$$\begin{aligned} g_1(1,1) = \{ & \lambda_1^2 [(1 + 2\rho_2 - 2\rho_1 \rho_2 - 2\rho_2^3 - 4\rho_1 \rho_2^2 + \rho_1^2 \rho_2^2 + 2\rho_1 \rho_2^4 + 2\rho_1^2 \rho_2^3) R_1''(1) + (1 - \\ & \rho_1^2 \rho_2^2) R_2''(1)] + \frac{(1-\rho_1\rho_2)(r_1+r_2)}{1-\rho_1-\rho_2} [(1 - \rho_2^2 - \rho_1\rho_2 - 2\rho_1\rho_2^2 + \rho_1\rho_2^3) A_1''(1) + (1 + \\ & \rho_1\rho_2) \beta_2^2 \lambda_1^2 A_2''(1)] + \frac{(r_1+r_2)\lambda_1^2}{1-\rho_1-\rho_2} [(1 + 2\rho_2 - 2\rho_1\rho_2 - 2\rho_2^3 - 4\rho_1\rho_2^2 + \rho_1^2\rho_2^2 + \\ & 2\rho_1\rho_2^4 + 2\rho_1^2\rho_2^3) \lambda_1 B_1''(1) + (1 - \rho_1^2\rho_2^2) \lambda_2 B_2''(1)] + \frac{2\lambda_1^2}{1-\rho_1-\rho_2} \{ [(1 - \rho_2^2 - \rho_1\rho_2 - \\ & 2\rho_1\rho_2^2 + \rho_1\rho_2^3)] \times [\rho_1 (1 + 2\rho_2 - \rho_1\rho_2) r_1 (r_1 + r_2) + \rho_2 r_1 (r_2 - \rho_1 r_2 + \rho_2 r_1) + \\ & r_2 (r_1 - \rho_2 r_1 + \rho_1 r_2) + \rho_2 (1 + \rho_1\rho_2) r_2 (r_1 + r_2)] + \rho_2^2 (1 + \rho_1\rho_2) [\rho_1 (1 + \rho_1\rho_2) \times \\ & r_1 (r_1 + r_2) + r_1 (r_2 - \rho_1 r_2 + \rho_2 r_1) + \rho_1 r_2 (r_1 - \rho_2 r_1 + \rho_1 r_2) + \rho_2 (1 + 2\rho_1 - \rho_1\rho_2) \times \\ & r_2 (r_1 + r_2)] \} + \frac{2\rho_2\lambda_1^2}{1-\rho_1-\rho_2} \{ [(1 - \rho_2^2 - \rho_1\rho_2 - 2\rho_1\rho_2^2 + \rho_1\rho_2^3)] \times [(1 - \rho_1 - \rho_2) \times \\ & (r_1 + \rho_1 r_2) + \rho_1 (1 + \rho_2)(r_1 + r_2)] + \rho_1\rho_2 (1 + \rho_1\rho_2) [(1 - \rho_1 - \rho_2)(r_2 + \rho_2 r_1) + \\ & \rho_2 (1 + \rho_1)(r_1 + r_2)] \} \} / [(1 - \rho_1^2 - \rho_1\rho_2 - 2\rho_1^2\rho_2 + \rho_1^3\rho_2)(1 - \rho_2^2 - \rho_1\rho_2 - \\ & 2\rho_1\rho_2^2 + \rho_1\rho_2^3) - \rho_1^2\rho_2^2 (1 + \rho_1\rho_2)^2] \end{aligned} \tag{22}$$

$$
\begin{aligned}
g_2(2,2) = \{ & \lambda_2^2 [(1-\rho_1^2\rho_2^2)R_1''(1) + (1+2\rho_1 - 2\rho_1\rho_2 - 2\rho_1^3 - 4\rho_1^2\rho_2 + \rho_1^2\rho_2^2 + \\
& 2\rho_1^4\rho_2 + 2\rho_1^3\rho_2^2)R_2''(1)] + \frac{(1-\rho_1\rho_2)(r_1+r_2)}{1-\rho_1-\rho_2}[(1+\rho_1\rho_2)\beta_1^2\lambda_2^2 A_1''(1) + \\
& (1-\rho_1^2 - \rho_1\rho_2 - 2\rho_1^2\rho_2 + \rho_1^3\rho_2)A_2''(1)] + \frac{(r_1+r_2)\lambda_2^2}{1-\rho_1-\rho_2}[(1-\rho_1^2\rho_2^2)\lambda_1 B_1''(1) + \\
& (1+2\rho_1 - 2\rho_1\rho_2 - 2\rho_1^3 - 4\rho_1^2\rho_2 + \rho_1^2\rho_2^2 + 2\rho_1^4\rho_2 + 2\rho_1^3\rho_2^2)\lambda_2 B_2''(1)] + \\
& \frac{2\lambda_2^2}{1-\rho_1-\rho_2}\{[(1-\rho_1^2 - \rho_1\rho_2 - 2\rho_1^2\rho_2 + \rho_1^3\rho_2)] \times [\rho_2(1+2\rho_1-\rho_1\rho_2)r_2(r_1 + \\
& r_2) + \rho_1 r_2(r_1 - \rho_2 r_1 + \rho_1 r_2) + r_1(r_2 - \rho_1 r_2 + \rho_2 r_1) + \rho_1(1+\rho_1\rho_2)r_1(r_1+r_2)] + \\
& \rho_1^2(1+\rho_1\rho_2)[\rho_2(1+\rho_1\rho_2)r_2(r_1+r_2) + r_2(r_1 - \rho_2 r_1 + \rho_1 r_2) + \rho_2 r_1(r_2 - \rho_1 r_2 + \\
& \rho_2 r_1) + \rho_1(1+2\rho_2 - \rho_1\rho_2)r_1(r_1+r_2)]\} + \frac{2\rho_1\lambda_2^2}{1-\rho_1-\rho_2}\{\ [(1-\rho_1^2 - \rho_1\rho_2 - \\
& 2\rho_1^2\rho_2 + \rho_1^3\rho_2)] \times [(1-\rho_1-\rho_2)(r_2 + \rho_2 r_1) + \rho_2(1+\rho_1)(r_1+r_2)] + \rho_1\rho_2(1 + \\
& \rho_1\rho_2)[(1-\rho_1-\rho_2)(r_1+\rho_1 r_2) + \rho_1(1+\rho_2)(r_1+r_2)]\}\} / [(1-\rho_1^2 - \rho_1\rho_2 - \\
& -2\rho_1^2\rho_2 + \rho_1^3\rho_2)(1-\rho_2^2 - \rho_1\rho_2 - 2\rho_1\rho_2^2 + \rho_1\rho_2^3) - \rho_1^2\rho_2^2(1+\rho_1\rho_2)^2] \quad (23)
\end{aligned}
$$

C. The Average Waiting Time of the System

In the periodic query threshold service, the average waiting time of the service object is the average waiting time [32], of the information packet from entering the queue to the start of the service. From [33], the average waiting delay of the two queues of the gate service [34] can be obtained.

$$\overline{W}_{G1} = \frac{(1+\rho_1)g_1(1,1)}{2\lambda_1 g_1(1)} - \frac{A_1''(1)}{2\lambda_1^2} \quad (24)$$

$$\overline{W}_{G2} = \frac{(1+\rho_2)g_2(2,2)}{2\lambda_1 g_2(2)} - \frac{A_2''(1)}{2\lambda_2^2} \quad (25)$$

IV. NUMERICAL CALCULATION AND EXPERIMENTAL SIMULATION

Numerical calculations from theoretical analysis, it is time to carry out programming and simulation experiments according to the operating mechanism of the system. Let the arrival process of each queue in any time slot obey the Poisson distribution. The channel rate is 54Mbit/s; the message packet length is 2700 bit; the query conversion time is 10μs; the slot width is 10μs. And the numerical calculation is consistent with the simulation experimental parameters. Set polling number $M = 10^6$. When $r_1 = r_2 = 1$, $\lambda_1 = 0.01$, $\lambda_2 = 0.04$ the average-length and average waiting time of the first and second queue changes with service rate is in Tables I and II. If you set the number of cycles $M = 3\times 10^6$. Under $r_1 = r_2 = 1, \beta_1 = 1, \beta_2 = 3$ conditions, the average-length and average waiting time of the first and second queue will change with arrival rate in Tables III and IV. If you set the number of cycles $M = 5\times 10^7$ under $\beta_1 = \beta_2 = 10$, $\lambda_1 = 0.01, \lambda_2 = 0.04$ conditions, the average waiting time will change with shifting-time, which is shown in Tables V and VI. If the first and second queue is exactly the same, running the number of polling cycles $M = 5\times 10^7$ the average-length and average waiting time of the first and second queue changes will change with the variation-load, which is shown in Fig. 2 and Fig. 3.

TABLE I. THE AVERAGE QUEUE-LENGTH AND AVERAGE WAITING-TIME OF QUEUE ONE CHANGES WITH SERVICE RATE

β_{1i}	$g_1(1)$ Numeral Calculations	$g_1(1)$ Simulation results	$\overline{W}_{G1}$ Numeral Calculations / slot	$\overline{W}_{G1}$ Simulation results / slot
1	0.0209	0.0208	0.5888	0.5586
2	0.0223	0.0225	0.7372	0.7402
3	0.0236	0.0235	0.9598	0.9530
4	0.0251	0.0249	1.2597	1.2578
5	0.0266	0.0263	1.6536	1.6550
6	0.0285	0.0285	2.1565	2.1966
7	0.0309	0.0310	2.7938	2.7039
8	0.0334	0.0336	3.5954	3.7181
9	0.0362	0.0355	4.6048	4.8738
10	0.0399	0.0390	5.8625	5.4977

TABLE II. THE AVERAGE QUEUE-LENGTH AND AVERAGE WAITING-TIME OF QUEUE TWO CHANGES WITH SERVICE RATE

β_{2i}	$g_2(2)$ Numeral Calculations	$g_2(2)$ Simulation results	$\overline{W}_{G2}$ Numeral Calculations / slot	$\overline{W}_{G2}$ Simulation results / slot
1	0.0843	0.0843	0.6209	0.6194
2	0.0888	0.0882	0.8125	0.8144
3	0.0946	0.0950	1.0863	1.0769
4	0.1001	0.1002	1.4598	1.4344
5	0.1066	0.1066	1.9529	1.9254
6	0.1139	0.1129	2.5930	2.5883
7	0.1238	0.1220	3.4172	3.4498
8	0.1331	0.1329	4.4598	4.4635
9	0.1452	0.1436	5.7832	6.1967
10	0.1601	0.1560	7.4961	7.0018

TABLE III. THE AVERAGE QUEUE-LENGTH AND AVERAGE WAITING-TIME OF QUEUE ONE CHANGES WITH ARRIVAL RATE

λ_{1i}	$g_1(1)$ theory	$g_1(1)$ experiment	$\overline{W}_{G1}$ Theory / slot	$\overline{W}_{G1}$ Experiment / slot
0.010	0.0207	0.0207	0.6308	0.6041
0.025	0.0555	0.0552	0.7711	0.7802
0.040	0.0953	0.0954	0.9561	0.9488
0.055	0.1409	0.1410	1.1672	1.1567
0.070	0.1942	0.1935	1.4101	1.4006
0.085	0.2574	0.2571	1.6928	1.7012
0.100	0.3338	0.3351	2.0323	2.0523
0.115	0.4256	0.4249	2.4409	2.4518
0.130	0.5412	0.5409	2.9506	2.9511
0.145	0.6912	0.6941	3.6053	3.6452

TABLE IV. THE AVERAGE QUEUE-LENGTH AND AVERAGE WAITING-TIME OF QUEUE TWO CHANGES WITH ARRIVAL RATE

λ_{2i}	$g_2(2)$ theory	$g_2(2)$ experiment	$\overline{W}_{G2}$ Theory / slot	$\overline{W}_{G2}$ Experiment / slot
0.010	0.0207	0.0207	0.6254	0.6234
0.025	0.0555	0.0559	0.8354	0.8451
0.040	0.0951	0.0946	1.0771	1.0822
0.055	0.1409	0.1411	1.3584	1.3311
0.070	0.1942	0.1941	1.6871	1.6642
0.085	0.2572	0.2563	2.0768	2.0662
0.100	0.3336	0.3361	2.5452	2.5758
0.115	0.4261	0.4259	3.1197	3.1301
0.130	0.5415	0.5396	3.8358	3.8472
0.145	0.6915	0.6927	4.7583	4.7846

TABLE V. THE AVERAGE QUEUE-LENGTH AND AVERAGE WAITING-TIME OF QUEUE ONE CHANGES WITH TRANSFER TIME

γ_{1i}	$g_1(1)$ theory	$g_1(1)$ experiment	$\overline{W}_{G1}$ Theory / slot	$\overline{W}_{G1}$ Experiment / slot
1	0.0400	0.0398	5.8771	5.8765
2	0.0800	0.0802	8.0773	8.0752
3	0.1200	0.1203	10.271	10.208
4	0.1600	0.1607	12.477	12.479
5	0.2000	0.2003	14.679	14.682
6	0.2400	0.2401	16.872	16.821
7	0.2800	0.2803	19.079	18.987
8	0.3200	0.3203	21.275	21.231
9	0.3600	0.3581	23.479	23.487
10	0.4000	0.4004	25.679	25.789

TABLE VI. THE AVERAGE QUEUE-LENGTH AND AVERAGE WAITING-TIME OF QUEUE TWO CHANGES WITH TRANSFER TIME

γ_{2i}	$g_2(2)$ theory	$g_2(2)$ experiment	$\overline{W}_{G2}$ Theory / slot	$\overline{W}_{G2}$ Experiment / slot
1	0.1600	0.1597	7.5051	7.5047
2	0.3200	0.3232	10.3010	10.4120
3	0.4800	0.4801	13.1009	13.0360
4	0.6400	0.6402	15.9050	15.9380
5	0.8000	0.7999	18.7040	18.7035
6	0.9600	0.9568	21.5058	21.4448
7	1.1200	1.1201	24.3069	24.2318
8	1.2800	1.2797	27.1050	27.1030
9	1.4400	1.4378	29.9050	29.9020
10	1.6000	1.6011	32.7068	32.7691

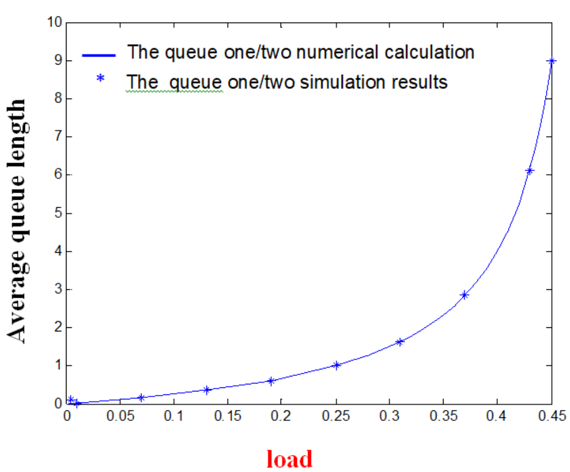


Fig. 2. The curve of the average queuing length following the load changes

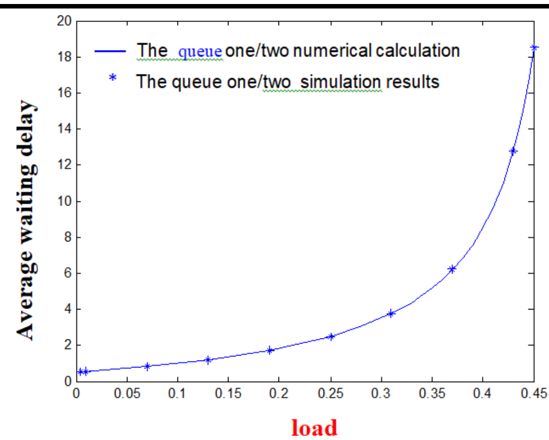


Fig. 3. The curve of the average waiting delay following the load changes

V. Discussion on Simulation Experiment And Numerical Calculation Results

Referring to Tables Ⅰ and Ⅱ, it can be seen that the average queuing length and average waiting delay changes with the variety of β_i when λ_i and γ_i remains constant. Looking at the two types of plots in Fig. 2 and 3, it can be seen that the system delay changes faster with the changing of λ_i . System delay is bigger when λ_i is bigger and when β_i remains constant. Experiment results and theoretical analysis expression (24) and (25) are consistent. In practice, the experimental results are approximately equal to the theoretical numerical calculations, there is a certain deviation. The main factor is that the number of polling statistics loops is too small, which is only $M = 10^6$. If the number of polling cycle statistics is further increased, the error between the experimental value and the theoretical value will be further reduced.

By viewing Tables Ⅲ and Ⅳ, it can be seen that the average queuing length and system delay changes with the variety of λ_i when β_i and γ_i remains constant. It can be seen that the average waiting delay changes faster as β_i is variety. System delay is bigger as β_i is bigger and λ_i remains constant. Experiment results and theoretical analysis expression (24) and (25) are consistent. Experimental data are subject to minor deviations, which is because the polling statistics loop count reaches $M = 3\times10^6$. If the number of polling cycle statistics is further increased, the error between the experimental value and the theoretical value will be smaller.

Referring to Tables Ⅴ and Ⅵ, It can be seen that the average queuing length and average waiting delay changes with the variety of γ_i when λ_i and β_i remains constant. It can be seen that the system delay changes faster as ρ_i is changing. System delay of the ones if lighter loads are smaller and shifting time remains constant. Experiment simulations and theoretical calculation expression (24), (25) are consistent. Experimental data are subject to minor deviations, which is because the polling statistics loop count reaches $M = 3\times10^6$. If the number of polling cycle statistics is further increased, the error between the experimental value and the theoretical value will be smaller.

By viewing Fig. 2 and 3, it can be seen that the system queuing length and system waiting delay changes when the variety of ρ_i if the same γ_i under the two queues are set to be symmetrical. Experiment results and theoretical analysis expression (24) and (25) are consistent. There may be minimal deviation of experimental data because the polling statistics loop count reaches $M = 5\times10^7$. If the number of polling cycle statistics is further increased, it can also improve the quality of experimental data analysis and can further reduce the uncertainty of the system.

Comparing Tables Ⅰ and Ⅱ, the load on the 10th service terminal is $\rho_{\max1,10} = 0.1$, $\rho_{\max2,10} = 0.4$, Comparison Table Ⅲ and Ⅳ, the maximum load on the 10th service terminal is $\rho_{\max1,10} = 0.145$, $\rho_{\max2,10} = 0.435$, Comparison Table Ⅴ and Ⅵ, the maximum load on the 10th service terminal is $\rho_{\max1,10} = 0.1$, $\rho_{\max2,10} = 0.4$. In a word, the value of the maximum load on the service terminal is $\rho_{\max} < 0.5$, System service terminals are running under light load. In the following research work, it is also necessary to analyze and discuss the system under heavy load, so as to strive for the integrity of the analysis.

By comparing Tables I, Ⅱ, Ⅲ, Ⅳ, Ⅴ and Ⅵ as well as Fig. 2 and Fig. 3, it can be seen that the deviation between the simulated experimental value and the numerical calculation value is further reduced with the increase of the threshold polling times. Proven by experiment, if you take the statistical polling times $M \ge 10^8$, it can prove that the error can be controlled within 1% under heavy load.

When the network system operates with symmetric parameters, Formula (24) and (25) of theoretical analysis is simplified to formula (20) in [26] under the $N = 2$ condition is exactly the same.

VI. Summary

The article, using the discrete-time Markov chain and a function set by a non-negative integer random variable, Accurate Analysis of Two-Queue Asymmetric Threshold Service System, is based on the establishment of a theoretical model and the definition of system-related parameters, in which the low-order characteristic quantity, higher-order characteristic quantity and query period of the system are deduced, and the system queuing length and system waiting delay are accurately calculated. The simulation experiment comes from the operating mechanism of the system and the theoretical analysis and sample calculation are in good agreement. The article lays a foundation for the researches on the related problems of multi-queue asymmetric threshold service polling system, further deepens the cognition of the asymmetric threshold polling system, and further expands the research space of polling system.

Acknowledgment

The authors extend their appreciation to the Funded Project that Modeling and Performance Analysis of Asymmetric Polling System of Scientific Research Fund Project of Department Education of Yunnan Province (2019J0401), Modeling Performance Analysis and Application Research of Asymmetric Polling System of Joint Special Projects of Local Colleges and Universities of Department of Science and Technology of Yunnan Province (202001BA070001-200).

References

[1] YANG Z J, KOU Q L, DING H W. "Research on polling system with retransmission mechanism in 5G slicing architecture". Computer Engineering, Vol.48, No.10, 2022.

[2] YANG Z J, KOU Q L, DING H W. "Research on performance of polling service with failure retries adapted to WSN". Modern Electronics Technique. Vol.45, No. 9, pp. 13-20, 2022.

[3] LI B, LIANG Z G, YAO Y H, GUAN Z, DING H W. "Optimal design of intelligent traffic light system based on FPGA and asymmetric polling control". Modern Electronics Technique. Vol.45, No.5, pp. 139-145, 2022.

[4] LIANG Z G, LI S N, LIANG Z G, DING H W. "Asynchronous polling system with priority data forwarding based on FPGA". Application Research of Computers. Vol.39, No.2, pp. 500-520, 2022.

[5] CHENG M, WANG X C, HE J H, DING C Y. Performance Analysis of Periodic Query Polling System Based onAsymmetric Dual-Queue Threshold Service. Journal of Chuxiong Normal University. Vol.37, No.3, pp. 7-14, 2022.

[6] YANG J H, LI Q, ZUO Y J, XIONG H G. "Delay analysis of WAIC network based on priority-deficit round robin scheduling". Journal of Beijing University of Aeronautics and Astronautics. Vol.47, No.9, pp. 1849-1856, 2021.

[7] XU J H. "WSN node capture detection method based on multi - polling". Video Engineering. Vol.45, No.4, pp. 121-128, 2021.

[8] SONG X H, LI W L. "Serial Communication Packet Polling Multiple Access Method Based on Nutaq Platform". Computer simulation. Vol.38, No.8, pp. 143-137, 2021.

[9] SHI W G, ZHANG Y C, WANG S W. "Non-fragile Extended State Estimation Based on Round-robin Protocol". Control Engineering of China. Vol.28, No.11, pp. 2230-2238, 2021.

[10] Z. J. Yang, H. Y. Zheng, H. W. Ding, "Research of polling MAC multi-robot system in MLAN," Application Research of Computers. Vol. 39, No. 4, pp. 1178-1182, 2022.

[11] Z. J. Yang, Mao, H. W. Ding, Q. L. Kou, "Analysis of Continuous Time Two-Level Exhaustive Polling Access MAC Protocol," Computer Engineering and Applications. Vol. 58, No. 9, pp. 136-143, 2022.

[12] Z. J. Yang, L. Mao, H. W. Ding, Z. Liu, "Analysis of characteristics of asymmetric two-level polling system for distinguishing priority business," Modern Electronics Technique. Vol. 44, No. 7, pp. 11-16, 2021.

[13] Z. J. Yang, Q. L. Kou, H. W. Ding, "Research on MAC protocol of polling system based on blockchain," 2021, Vol. 43, No. 2, pp. 228-236.

[14] X. H. Song, W. L. Li, "Serial Communication Packet Polling Multiple Access Method Based on Nutaq Platformr," Computer Simulation. Vol. 38, No. 3, pp. 143-147, 2021.

[15] Z. J. Yang, L. Mao, H. W. Ding, "Performance prediction of polling system based on Long Short-Term Memory network," Journal of Yunnan University: Natural Sciences Edition. Vol. 42, No. 6, pp. 1046-1052, 2020.

[16] Z. J. Yang, Z. Liu, H. W. Ding, "Research of continuous time two-level polling system performance of exhaustive service and gated service," Journal of Computer Applications, 2019, Vol. 39, No. 7, pp. 2019 - 2023.

[17] Y. Q. Chu, Z. M. Liu, "The impact of priority policy in a two-queue Markovian polling system with multi-class priorities," Proceedings of the 12th International Conference on Queueing Theory and Network Applications. Berlin: Springer, Vol. 2017, pp. 282 -296.

[18] Y. Su, Z. J. Yang, Y. Y. Ding, et al. "Implementation of polling system control in wireless sensor networks," Electronic Measurement Technology, Vol. 41, No. 4, pp. 66-70, 2018.

[19] Z. Guan, Z. J. Yang, M. He, et al. "Study on the delay performance of station dependent two-level polling systems," Acta Automatica Sinica, Vol. 42, NO. 8, pp. 1207 -1214, 2016.

[20] W. H. Mu, L. Y. Bao, H. W. Ding, Y. F. Zhao, "An exact analysis of discrete time two-level priority polling system based on multi - times gated service policy," Acta Electronica Sinca, Vol. 26, No. 2, pp. 276-280, 2018.

[21] M. A1-Maqri, M. Othman, B. Ali, Z. Hanapi. "Adaptive multi--polling scheduler for QOS support of video transmission in IEEE 802.11e WLANs," Telecommunication Systems, Vol. 61, No. 4, pp. 773-791, 2016.

[22] M. Y. Wang, H. W. Ding, Y. F. Zhao, Z. G. Liang, L. Z. Zhu, "Theapplication of the exhaustive polling theory in intelligent traffic system [A]. Matec web of Conferences," Les Ulis, France: EDP Sciences, Vol. 2016, pp. 2013-2017.

[23] D. F. Zhao, H. W. Ding, Y. F. Zhao, "An analytical model of a discrete-time polling MAC protocol for wireless LANs using m-gated services," ActaElectronica Sinica, Vol. 38, No. 7, pp. 1495-1499, 2010.

[24] D. Everitt, "Simple approximations for token rings. IEEE Trans Commun," 1986, Vol. 34, pp. 719-721.

[25] D. F. Zhao, "Study on production flowos control of work stations with priority classes," Journal of China Institute of Communications, Vol. 27, No. 5, pp. 265-368,1998.

[26] D. F. Zhao, S. M. Zheng, "Message waiting time analysis for apolling system with gatedservice," Journal of China Institute of Communications, Vol. 15, No. 2, pp. 18-23, 1994.

[27] M. J. Ferguson, Y. J. Aminetzah, "Exact results for nonsymmetric token ring systems," IEEE Trans Commun, Vol. 33, No. 3, pp. 223-231, 1985.

[28] O. C. Ibe, X. Cheng, "Approximate analysis of asymmetric single-server token-passing systems," IEEE Trans Commun, Vol. 37, No. 6, pp. 572-577, 1989.

[29] B. Mukherjee, C. K. Kwok, A. C. Lantz, et al. "Comments on exact analysis of asymmetric polling systems with single buffers," IEEE Trans Commun, 1990, Vol. 38, No. 7, pp. 944-946.

[30] F. Hou, D. F. Zhao, "An Analysis of a Multimedia Translationg System Based Asymmetric Cyclic Polling," Journal of Yunnan University: Natural Science Edition, 2000, Vol. 22, No. 3, pp. 177-180.

[31] J. H. Shi, D. F. Zhao, G. H. Cai, "Model Analysis for Tracking Senso Rmult-service com m unication system," Journal of Yunnan University: Natural Science Edition, Vol. 25, No. 2, pp.105-109, 2003.

[32] M. G. Wang, "Modeling and Characteristic Analysis of Two Kinds of Polling Systems," Kunming pp. 19-37, 2009.

[33] D. F. Zhao, "Research on Asymmetry in Token Network," Journal of Communications, Vol. 19, No. 1, pp. 75-80, 1998.

[34] J. S. Zhao, D. F. Zhao, H. W. Ding, "Research on Asymmetric Problems in Threshold Service Polling System," Journal of Yunnan University: Natural Science Edition, Vol. 32, pp. 1, pp.30-35, 2010.

Transformer-based Cross-Lingual Summarization using Multilingual Word Embeddings for English - Bahasa Indonesia

Achmad F. Abka[1], Kurniawati Azizah[2], Wisnu Jatmiko[3]
National Research and Innovation Agency, Jakarta, Indonesia[1]
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia[1, 2, 3]

Abstract*—Cross-lingual summarization (CLS) is a process of generating a summary in the target language from a source document in another language. CLS is a challenging task because it involves two different languages. Traditionally, CLS is carried out in a pipeline scheme that involves two steps: summarization and translation. This approach has a problem, it introduces error propagation. To address this problem, we present a novel end-to-end abstractive CLS without the explicit use of machine translation. The CLS architecture is based on Transformer which is proven to be able to perform text generation well. The CLS model is a jointly trained CLS task and monolingual summarization (MS) task. This is accomplished by adding a second decoder to handle the MS task, while the first decoder handles the CLS task. We also incorporated multilingual word embeddings (MWE) components into the architecture to further improve the performance of the CLS models. Both English and Bahasa Indonesia are represented by MWE whose embeddings have already been mapped into the same vector space. MWE helps to better map the relation between input and output that use different languages. Experiments show that the proposed model achieves improvement up to +0.2981 ROUGE-1, +0.2084 ROUGE-2, and +0.2771 ROUGE-L when compared to the pipeline baselines and up to +0.1288 ROUGE-1, +0.1185 ROUGE-2, and +0.1413 ROUGE-L when compared to the end-to-end baselines.

Keywords—Cross-lingual summarization; multilingual word embeddings; transformer; automatic summarization

I. Introduction

Automatic summarization is a process of automatically producing a shorter version of an original while retaining contents and meanings that are considered essential. This shorter version is called a summary. The purpose of summarizing is to produce a summary that contains the main content of a document in less space [1]. Automatic summarization helps users get the main idea of a document without having to read the whole document thus saving time and effort compared to doing it manually [2]. In general, automatic summarization can be categorized into two based on the approach in summarizing: extractive summarization and abstractive summarization. The extractive approach produces a summary by copying words or sentences from source documents that are considered important. The abstractive approach produces a summary using its own words or sentences. In addition to text, summaries of images and videos can also be produced [3] [4] [5]. In the context of this work, automatic summarization is an activity that uses a machine (computer) to automatically summarize a document using a certain algorithm or method.

The study on automatic summarization was first reported by Luhn [6]. The system is based on bag-of-words. The frequency and relative position of a word in a sentence are the main features in determining how important the sentence is. Gradually, linguistic information such as word type and structure are utilized using natural language processing (NLP). The extractive approach is then widely used because of its simple approach without the need for extensive NLP [7] [8]. The success of the sequence-to-sequence (Seq2Seq) recurrent neural network (RNN) model has made the rapid development of abstractive approaches [9] [10]. Inspired by the attention-based neural machine translation (MT) model [11] [12], the RNN is used by adding the attention mechanism so that the model can focus on some parts of text while diminishing other parts at a certain time/context.

Cross-lingual summarization (CLS) is a task to produce a summary in the target language, from source documents in another language [13] [14]. Traditionally, CLS can be done by involving two processes: translation and monolingual summarization (MS). However, there is a problem when doing it with two processes in a pipeline scheme. The pipeline process introduces error propagation which adversely affects the final quality of the summary. This problem can be solved by conducting end-to-end model training [15] [16]. Another challenge in CLS research is the limited data or corpus that can be used to train the model. Research on end-to-end CLS is still relatively new, so resources related to it are still quite scarce. There is no dataset that is considered a standard that can be used as a benchmark. The researchers still tend to create their own dataset for the language pair domain that is of interest to them.

In this paper, we present a novel end-to-end abstractive CLS that solves the error propagation problem found in the pipeline scheme. We adapt transformer-based architecture and extended it by including multilingual word embeddings (MWE) [17] components. These components allow the model to be able to represent words in two different languages. This paper is an extended study of our previous work on the end-to-end abstractive CLS model [18]. The model is trained in the English domain for source documents and the Bahasa Indonesia domain for its summary using the dataset that we have constructed. Inspired by [19] and [20], we use a round-trip translation technique to generate CLS dataset from MS dataset. This technique is similar to that used by [15], but we apply it to the document and its summary, not just the summary. The resulting dataset is then evaluated using bilingual evaluation understudy (BLEU) [21]. BLEU is a method for evaluating the quality of text that has been machine translated from one language to another automatically. The CLS model can produce a Bahasa Indonesia summary from an English source document end-to-end without explicitly using a machine translator.

The main contributions of this work are:

- A novel framework for generating end-to-end abstractive cross-lingual summary by incorporating MWE components in the architecture. These components are used to represent words in both languages used in the source document and its summary.
- This works produces a new cross-lingual dataset. This dataset can be used for CLS research with the English domain as the source document and the Bahasa Indonesia domain as the summary.

The rest of this paper is organized as follows. In Section II, we review the related work in CLS. In Section III, we explain the proposed method. In Section IV, the details of the experimental setup and the evaluation metrics are presented. The experimental results are given in Section V. Finally, the conclusions of this research are found in Section VI.

II. Related Works

CLS is the process of generating a summary in the target language from source documents in another language. Unlike MS, CLS involves at least two different languages. Traditional approaches treat CLS as a pipeline scheme. The newer approach does it end-to-end to avoid error propagation that occurs in the pipeline process.

A. Traditional Cross-Lingual Summarization

CLS research has been conducted [22] [23] [24]. The approaches are generally divided into extraction-based and compression-based. Summarization is done in a pipeline manner which is divided into two steps: summarizing and translating. There are two patterns of utilization or use of machine translation. The first pattern is to first translate the source document using a machine translator and then the results of the translation are used to summarize. The second pattern summarizes the source document first and then translates the results using a machine translator into the target language. Another approach utilizes existing source documents in two languages to generate summaries in the target language [25] [26]. This target language is one of the two languages. In addition to the extraction process, a compression process is also carried out in the compression-based approach. Summarization begins with a selection process to obtain relevant content in the form of sentences or phrases (bilingual). The compression is carried out by removing parts of sentences or phrases that do not meet some criteria, such as information content, legibility, and grammar/structure [27] [28] [29].

In contrast to the CLS previously described which uses an extractive approach, Zhang, Zhou, and Zong [30] developed an abstractive CLS system. The system works by first translating the source document from English to Chinese using Google Translate. The next step is to extract bilingual concepts and fact pairs. Then the score of translation and salience is calculated from this set of pairs. Based on the scores, a set of pairs is selected to be used in the summary. This selection considers several criteria, including compatibility between concepts and facts, number of sentences, summary length, etc. Post-processing begins with determining which pairs of concepts and facts are included in a particular sentence. Then the last step is to put them in order. Abstractive summarization has the ability to paraphrase. In the system developed by [30], the abstractive process is carried out by combining pairs of concepts and facts into one sentence. This process is indeed a form of paraphrasing, but these pairs of concepts and facts are taken extractively and used as they are.

Similar to [30], our work also falls under the abstractive CLS category. The difference is that our approach is based on neural networks and summarization is done directly, rather than in two steps in a pipeline scheme. This straightforward CLS can avoid error propagation that occurs in the pipeline process.

B. Neural Network-based Cross-Lingual Summarization

In 2019, neural network-based CLS began to emerge. Ouyang, Song, and McKeown [31] use neural networks for CLS in pipeline schemes, on the translation side using Marian [32] and on the summarization side using pointer-generators [33]. The system summarizes Somali, Swahili, or Tagalog source documents into an English summary. Zhu et al. [15] are the first to report on end-to-end CLS. A CLS from English to Chinese was built using transformer architecture that has been proven to have good performance in text generation. The network is trained using a multi-task learning approach. Specifically, they combine CLS loss with MS loss. The CLS dataset used was built by modifying the CNN/Dailymail dataset. The dataset is translated into Chinese using a machine translation service. Duan et al. [16] see CLS as a zero-shot problem. They do not have a CLS dataset to train the model directly. The network is trained by adapting a paradigm on neural machine translation (NMT) called triangular NMT systems. A network called a student network is a CLS network that is trained to imitate the behavior of a teacher network which is an MS network. Ladhak et al. [34] proposed a benchmark dataset for abstractive CLS named WikiLingua. The data is taken from WikiHow in the form of document and summary pairs in 18 languages. Not all data is available in 18 languages, availability in each language varies. The dataset is tested for CLS with pipeline and end-to-end approach. The end-to-end approach uses mBART [35] which is fine-tuned using the source language document and the target language summary.

Our approach is inspired by [15]. However, we propose to add MWE components which we believe will improve the performance of the model. MWE represents the two languages in the CLS task in the same vector space. This facilitates the mapping of the relation between inputs and outputs during model training.

C. Multilingual Word Embeddings

Multilingual word embeddings (MWE) are word embeddings that represent words in various languages in one vector space [17]. Conneau et al. [36] developed MWE using an unsupervised approach. Supervised approaches generally require bilingual resources such as dictionaries or parallel corpus. They proposed an unsupervised way to map monolingual word embeddings without the need for bilingual data. Word embeddings for each language are initially trained independently. Then, each word in the two languages is mapped into the same vector space using deep adversarial networks [37]. Heinzerling & Strube [38] developed MWE for word segmentation (subword embeddings). Word segmentation is done by using the byte pair encoding (BPE) approach. BPE performs segmentation to words that rarely appear into several tokens. Embeddings are trained in a simple way. All articles from Wikipedia in various languages are combined, then used to train subword embeddings. Artetxe & Schwenk [39] also developed multilingual embeddings but specifically developed embeddings for sentences. These sentence embeddings were trained using a single bidirectional LSTM (BiLSTM) encoder.

In this work, we use MUSE from Conneau et al. [36]. MUSE is used because the embedding is word-based so it is expected to be able to cover most of the vocab. MWE based on word segmentation such as BPEmb from Heinzerling & Strube [38] may be better than MUSE because it can recognize tokens that rarely or never appear in the training data. However, it is not suitable for the case in this work because the embedding is trained using various language versions of the articles available on Wikipedia. The amount of data which is an aggregate of all articles in various language versions can also result in many words being segmented due to their small frequency, thus causing the size of vocab to be very large. In general, a vocab size that is too large can adversely affect the performance of NLP system.

III. Methods

CLS can be expressed as an input sentence in the source language $x = (x_1, x_2, x_3, \dots, x_m)$ and $x_i \in V_s$, where m is the input sentence length and V_s is the source vocab in the source language, with an output summary in the target language $y = (y_1, y_2, y_3, \dots, y_n)$ and $y_i \in V_t$, where n is the output summary length, $n \leq m$, and V_t is the target vocab in the target language.

A. Transformer-based Cross-Lingual Summarization

NLP systems that are built specifically for a particular language are generally trained using datasets in that language. This approach has a problem when applied to other languages. The system needs to be retrained with separate datasets in the other specific language. This is equivalent to building the system from scratch. Another way is to use machine translation. The system is combined with machine translation in a pipeline so that it can produce output in the desired language. MWE can be used to tackle this problem. Both languages are represented by embeddings that are mapped into the same vector space. This mapping makes the two languages seem to be the same language, at least at the word level. The CLS architecture proposed in this work is based on transformer architecture [40] by adding MWE components to handle cross-lingual problem. Transformer was chosen because it is state-of-the-art in many NLP topics including CLS and has proven to be able to perform text generation well [41] [42].

Inspired by Zhu et al. [15], the architecture also added a second decoder with a single shared encoder. During training, the input of CLS decoder (decoder1) is a summary in Bahasa Indonesia, while the input of MS decoder (decoder2) is a summary in English. MS decoder is only used during training. So, in the training process, encoder-decoder1 is trained for CLS problems, while encoder-decoder2 is trained for MS problems. The losses from the two are combined and can be calculated as follows:

$$J(w) = -\frac{1}{N}\sum_{i=1}^{N}[y_i^a \log(\hat{y}_i^a) + (1 - y_i^a)\log(1 - \hat{y}_i^a)] + \\ -\frac{1}{N}\sum_{i=1}^{N}\left[y_i^b \log\left(\hat{y}_i^b\right) + \left(1 - y_i^b\right)\log\left(1 - \hat{y}_i^b\right)\right] \quad (1)$$

where w are parameters in the model, y_i^a and y_i^b are the correct labels for both tasks, and $\hat{y}_i^a$ and $\hat{y}_i^b$ are predictive labels for both tasks. The combined loss was used to update the weight. During test, the MS decoder was ignored. The architectural diagram can be seen in Fig. 1.

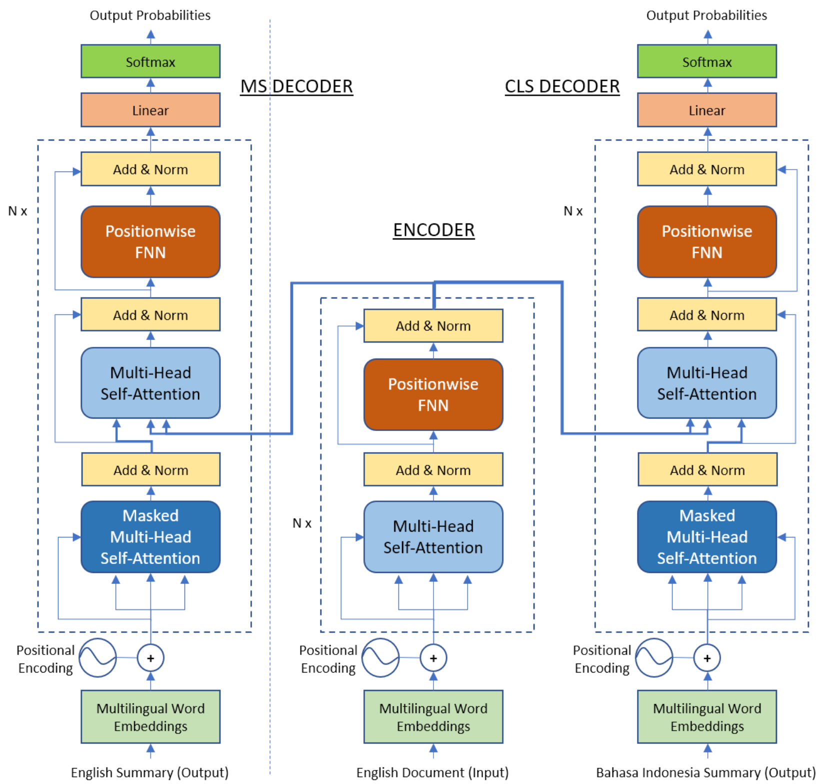


Fig. 1. CLS architecture

B. Cross-Lingual Summarization Architecture Components

1) Input and output: The input of the encoder is an English document $\mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots, \mathrm{x}_\mathrm{m})$. This document is tokenized using a subword tokenizer. Words that occur infrequently will be segmented into several tokens, while words that occur frequently (high frequency) are left as is. This sequence of tokens is then converted into a sequence of indexes according to the vocab. The output of CLS decoder is a Bahasa Indonesia summary and the output of MS decoder is an English summary. At the time of training, the tokens here are also an input that will be converted into a sequence of indexes according to the vocab. Vocab for Bahasa Indonesia is different from vocab for English. At the time of test, the CLS decoder generates a Bahasa Indonesia summary $\mathrm{y} = (\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3, \ldots, \mathrm{y}_\mathrm{n})$ while the MS decoder is ignored.

2) Multilingual word embeddings: The proposed CLS architecture incorporated pre-trained MWE from MUSE [36]. We use MUSE because it can cover most of the words in vocab while maintaining vocab cohesiveness. MWE Component maps the input English document $x = (x_1, x_2, x_3, \ldots, x_m)$ into a sequence of embeddings $Z = (Z_1, Z_2, Z_3, \ldots, Z_m)$ whose size varies with respect to the source sequence length.

3) Positional Encoding: This component provides position information to the tokens in the sequence by adding the "positional encodings" to the embeddings. The positional encodings are obtained using the following formula:

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d})$$

$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d}) \quad (2)$$

where pos is the position $(0 \leq pos < \frac{L}{2}$, where L is the length of the sequence), d is the embedding dimension, and i is used for mapping to indices of elements in positional encoding vector $(0 \leq i < \frac{d}{2})$. Positional encoding is necessary because the proposed transformer-based architecture does not consider positional information or word order. Without positional encoding, this architecture is essentially a bag-of-words model.

4) Multi-head self-attention: Multi-head self-attention consists of several heads which correspond to scaled dot-product attention (self-attention). Called self-attention because it can generate its own value of query (Q), key (K), and value (V). These three values are abstractions that represent the input needed to calculate the attention weight. It is obtained using the following formula:

$$Q = ZW_Q$$
$$K = ZW_K$$
$$V = ZW_V \quad (3)$$

where Z is the input embeddings and W_Q, W_K, and W_V are learnable matrices. The scaled dot-product attention is then calculated using the scaling factor of $\frac{1}{\sqrt{d_k}}$.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

where $d_k$is dimension of K. The output of each head is then concatenated to get the final values.

$$MultiHead(Q,K,V) = Concat(head_1, head_2, \ldots, head_h)W^O$$
$$where\ head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (5)$$

where h is the number of heads and W^O, W_i^Q, W_i^K, and W_i^V are learnable matrices. Multi-head self-attention on the decoder is also called encoder-decoder attention because it receives input from the encoder and decoder. There is a mapping of the relation between input and output here. This component can be calculated in parallel because the attention weight here is independent of one another.

5) Masked multi-head self-attention: This component has a similar function to the multi-head self-attention component. However, unlike multi-head self-attention, masked multi-head self-attention has a function to mask output that has not been seen/predicted. The model must predict the output based on the results of the previous output and must not look at the output that appears later. When the matrix operation is performed, the output that has not been seen/predicted will be masked/changed to zero so that it cannot be seen by the model.

6) Positionwise FNN: Positionwise FNN is a simple feedforward neural network that runs for all attention weights. Its main task is to convert the attention weight into an acceptable form for the next step, such as encoder layer, decoder layer, or linear layer. This component consists of two linear layers. The first layer uses ReLU activation.

$$FFN(x) = \max(0, xW_1 + b_1)\,W_2 + b_2 \quad (6)$$

7) Linear Layer and Softmax: Linear layer is a component in the form of a feedforward layer that acts as a classifier. This component is used to adjust the dimensions as needed. For example, to accommodate the number of words (classes) in the vocab. Finally, there is the softmax layer which converts the vector into a probability distribution.

IV. EXPERIMENTS

This section describes the experiments. First, we describe the dataset which includes English documents and their summary, Bahasa Indonesia documents and their summary, and English-Bahasa Indonesia parallel corpus. Then, we explain the implementation and model variations including the baseline model. Finally, we discuss the evaluation metrics used to measure the performance of CLS models.

A. Dataset

Developing a CLS system is a challenging task because it involves two different languages, especially if the language is a low-resource language. The experiments conducted in this work require a cross-lingual dataset from English documents to Bahasa Indonesia summaries which is not readily available. Therefore, the dataset needs to be created first. Creating a dataset from scratch for deep learning-based models can take a lot of time and effort. The alternative is to create a new dataset by utilizing existing data. The closest problem to CLS is the MS problem. The best dataset as a basis to work with is a dataset of MS for Bahasa Indonesia because the output target of our CLS model is a summary in Bahasa Indonesia. We use IndoSum [43] as the basis for creating a new CLS dataset. IndoSum is an attempt to create a benchmark dataset for Bahasa Indonesia summarization. This dataset contains nearly 20000 news articles taken from online websites. This number is still relatively small when compared to the available English summarization dataset. Each article has an abstract summary that was created manually by 2 native Bahasa Indonesia speakers. The dataset has 6 categories: entertainment, inspiration, sports, show world, headlines, and technology. The dataset has been divided into 5-fold cross-validation and has been divided into training sets, development sets, and test sets. In this work, only the first fold was used. This dataset is written in JSON format.

Based on the need to train the CLS models, IndoSum's articles and summaries need to be translated into English. English articles are used as input for the CLS model. In addition, these English articles and their English summary are used to train the English MS model used in the pipeline approach. The strategy for creating the CLS dataset can be seen in Fig. 2. At first the original dataset of Bahasa Indonesia documents and summaries are translated into English using Google Translate (forward translation). The result of this English translation is then translated back into Bahasa Indonesia (back translation). This is done to ensure the quality of the CLS dataset. The back translation results were evaluated against the original dataset using BLEU [21]. The evaluation of the back translation can be seen in Table I. The evaluation is the cumulative BLEU score calculated using the NLTK. The results show that the quality of the cross-lingual dataset is quite good. As a comparison to get the intuition of the BLEU score, specific machine translators from English to Bahasa Indonesia have been reported to have BLEU scores of 25.3 [44] and 24.5 [45]. In the end, we obtained the Sum (Ina) dataset to train Bahasa Indonesia summarization, the Sum (Eng) dataset to train English summarization, and the CLS (Eng-Ina) dataset to train CLS as can be seen in Fig. 3. The total of data is 18774 document-summary pairs. The statistics of the CLS dataset are presented in Table II. To train machine translators from English to Bahasa Indonesia, the Pan Asia Networking Localization (PANL) parallel corpus dataset was used [45]. This data is divided into 15373 train data, 3845 validation data, and 4806 test data.

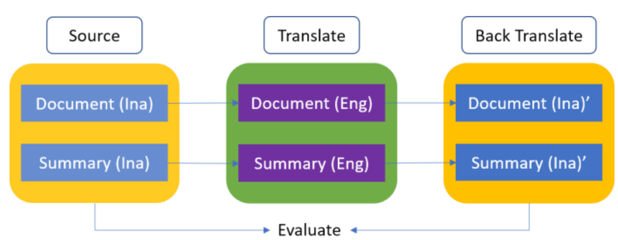


Fig. 2. CLS dataset creation strategy

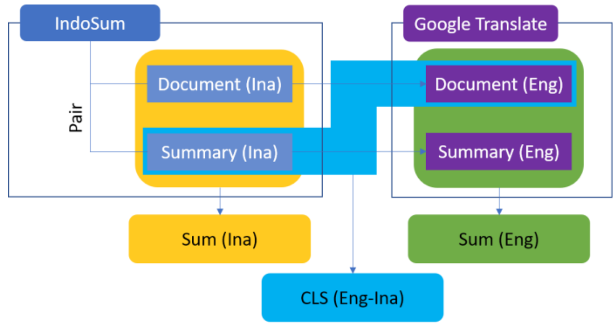


Fig. 3. CLS dataset diagram

TABLE I. CLS DATASET EVALUATION

Back Translation	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Document	0.8056	0.7046	0.6251	0.5527
Summary	0.7991	0.7076	0.6327	0.5626

TABLE II. DATASET STATISTICS

CLS Dataset	Train	Val	Test
# Documents/Summaries	14262	750	3762
# Average Sentence in Document	20.08	19.89	19.92
# Average Sentence in Summary	4.69	4.65	4.69
# Average English Words in Document	377.50	377.86	374.41
# Average Bahasa Indonesia Words in Document	326.58	327.38	323.84
# Average English Words in Summary	75.35	75.18	75.40
# Average Bahasa Indonesia Words in Summary	64.39	64.22	64.37
# Average English Words in Sentence (Document)	16.80	16.96	16.79
# Average Bahasa Indonesia Words in Sentence (Document)	14.72	14.99	14.71
# Average English Words in Sentence (Summary)	14.54	14.63	14.55
# Average Bahasa Indonesia Words in Sentence (Summary)	12.45	12.57	12.44

B. Implementation

In this work, we train 10 CLS models. Two models are pipeline and eight models are end-to-end. All models use transformer as their basis. The model is implemented using Python programming language and TensorFlow library. TensorFlow is an open-source library for machine learning and artificial intelligence. TensorFlow is developed by Google Brain Team. TensorFlow can run on CPUs and GPUs and is available for Linux, macOS, Windows, Android, and iOS operating systems. TensorFlow is widely used for NLP and computer vision applications. Each model is accompanied by a subword tokenizer which is used to prepare the input to be submitted to the model. The vocab size for English is 26331 and the vocab size for Bahasa Indonesia is 27373. All models have the same hyperparameters as can be seen in Table III. Each model uses Adam's optimization [46] with a custom learning rate scheduler. Regarding randomness, all models are trained on the same seed value of 777, whether it is a seed for Python, NumPy, or TensorFlow. The same seed is applied to global conditions as well as to any operations involving randomness. The pre-processing carried out are: 1) converting all text to lowercase; 2) removing symbols, special characters, HTML tags, and emoticons; 3) performing Unicode normalization; and 4) brackets: (), [], {} and all characters in between are discarded. The 26 characters of the alphabet are preserved, digits/numbers are preserved, and some punctuation marks are preserved, such as period, comma, exclamation mark, and question mark.

TABLE III. CLS MODEL HYPERPARAMETERS

BUFFER_SIZE	15000
BATCH_SIZE	64
NUMBER_LAYERS	4
DIMENSION_MODEL	300
DIMENSION_FNN	512
NUMBER_HEADS	5
DROPOUT_RATE	0.1
POSITIONAL_INPUT	202
POSITIONAL_OUTPUT	122

C. Model Variations

In this work we build ten variations of the model which are grouped into three main groups: 1) pipeline CLS; 2) end-to-end CLS; and 3) end-to-end CLS with MWE components. We also conducted experiments by implementing two strategies of truncating the input, namely using only the head of the document and using the head and tail of the document. This input truncation strategy is inspired by Sun et al. [47] and Mutasodirin & Prasojo [48] who use this strategy for text classification problems. Both use the BERT model [42] but get different conclusions about this input truncation strategy, so this strategy cannot be generalized yet. In this work, the head strategy means using the first 200 tokens from the document, while the head-tail strategy means using the first 100 tokens from the document and concatenating them with the last 100 tokens from the document. So, the models in groups 2 and 3 are further divided into two groups: models that use the head of the document as input and models that use the head and tail of the document as input. Four out of 10 are baseline models: two pipeline models, PipeTS and PipeST, an end-to-end model that uses a vanilla transformer, VCLS, and an end-to-end model based on Zhu et al. [15], MCLS. The following is an explanation of each model variation:

1) PipeTS: Translation → Summarization. The model is built in a pipeline scheme, starting with the English translation process first and then continuing with the Bahasa Indonesia summarization process. The machine translation model is trained using the PANL dataset. The Bahasa Indonesia summarization model is trained using the Sum (Ina) dataset. Both the translation and the summarization models use a vanilla transformer.

2) PipeST: Summarization → Translation. The model is built in a pipeline scheme, starting with the English summarization process first and then continuing with the English translation process. The English summarization model is trained using the Sum (Eng) dataset. The machine translation model is the same as that used in PipeTS. Both the summarization and the translation models use a vanilla transformer.

3) VCLS: Head + CLS. The model is built end-to-end with head truncation strategy for the input. The model uses a vanilla transformer architecture but is trained using the CLS (Eng-Ina) dataset.

4) MCLS: Head + CLS-MS. The model is built end-to-end with head truncation strategy for the input. The model is based on Zhu et al. [15] CLS architecture. The model is trained with 2-task learning, which is jointly training CLS and MS using CLS (Eng-Ina) dataset and Sum (Eng) dataset.

5) VCLS_T: Head-Tail + CLS. The model is built end-to-end with head-tail truncation strategy for the input. The model uses a vanilla transformer architecture but is trained using the CLS (Eng-Ina) dataset.

6) MCLS_T: Head-Tail + CLS-MS. The model is built end-to-end with head-tail truncation strategy for the input. The model is based on Zhu et al. [15] CLS architecture. The model is trained with 2-task learning, which is jointly training CLS and MS using CLS (Eng-Ina) dataset and Sum (Eng) dataset.

7) MWE_VCLS: Head + CLS + MWE. The model is built end-to-end with head truncation strategy for the input. The model uses a vanilla transformer architecture but is trained using the CLS (Eng-Ina) dataset. This model is equipped with MWE components.

8) MWE_MCLS: Head + CLS-MS + MWE. The model is built end-to-end with head truncation strategy for the input. The model uses the proposed CLS architecture explained in section 3. The model is trained with 2-task learning, which is jointly training CLS and MS using CLS (Eng-Ina) dataset and Sum (Eng) dataset.

9) MWE_VCLS_T: Head-Tail + CLS + MWE. The model is built end-to-end with head-tail truncation strategy for the input. The model uses a vanilla transformer architecture but is trained using the CLS (Eng-Ina) dataset. This model is equipped with MWE components.

10)MWE_MCLS_T: Head-Tail + CLS-MS + MWE. The model is built end-to-end with head-tail truncation strategy for the input. The model uses the proposed CLS architecture explained in Section III. The model is trained with 2-task learning, which is jointly training CLS and MS using CLS (Eng-Ina) dataset and Sum (Eng) dataset.

D. Evaluation Metrics

This work uses a quantitative research method with the independent variable being the transformer-based CLS model and the dependent variable being the evaluation of the model's performance measured quantitatively. Recall-oriented understudy for gisting evaluation (ROUGE) is a suite of measurement metrics and a software package used to evaluate automatic summarization and machine translation [49] [50] [51]. ROUGE is a popular metric widely used to evaluate automatic summarization. The advantages of using ROUGE:

- ROUGE is easy and fast compared to evaluation by humans,
- ROUGE is used by "everyone" so it is easy to compare one research result with another.

ROUGE compares the summary generated by the system with a reference summary or a set of reference summaries. Here are some of the available comparison methods:

- ROUGE-N: Compares n-grams between system summary and reference summary.
- ROUGE-L: Compares the longest matching sequence (LCS) between system summary and reference summary.
- ROUGE-S: Compare skip-bigrams, whether there are any word pairs in a consecutive sentence, gaps are possible. For example, skip-bigram measures the overlap of word pairs that can have a maximum of two gaps between words. The sentence "I eat fried rice" then the skip-bigram is "I eat, I fried, I rice, eat fried, eat rice, fried rice."

There are still other measurement metrics, for more information refer to Lin [49].

Bilingual evaluation understudy (BLEU) is a method for automatic evaluation of machine translation. The BLEU score is a measurement metric that assesses the degree of similarity between sentences produced by machine translation and reference sentences made by human translators. The more similar, the higher is the value. This method works by counting the number of n-grams in the produced sentence that matches the n-grams in the reference sentence. BLEU was proposed by papineni et al. [21].

V. Results and Discussion

We evaluated the models trained using the ROUGE metric described in the previous section. Specifically, we used ROUGE-1, ROUGE-2, and ROUGE-L. The discussion regarding the experimental results is divided into three parts. First, we discuss the comparison of the pipeline model and the end-to-end model. Second, we discuss the comparison of the end-to-end model with the MWE components and the end-to-end model without the MWE components. In the end, we discuss the strategy of input truncation.

A. Comparison between Pipeline CLS Model and End-to-End CLS Model

Traditionally, CLS is done in a pipeline which involves two steps: summarization and translation. There are two patterns of using a machine translator in a pipeline scheme. The first pattern does the translation on the source document first and then the results of the translation are used to generate the summary. The second pattern performs summarization on the source document first and then the summary is translated into the target language. To construct CLS for these two patterns, three constituent models are needed: English summarization model, Bahasa Indonesia summarization model, and English translation model. We trained a transformer-based English summarization model on Sum (Eng) dataset and a transformer-based Bahasa Indonesia summarization model on Sum (Ina) dataset. Then, we trained a transformer-based English translation model on PANL parallel corpus. The performance of these three models individually can be seen in Table IV and Table V.

We use the two patterns of the pipeline scheme as baseline models (PipeTS and PipeST). These two pipeline models were constructed using MS model and MT model which had previously been trained independently. Both are then tested using the CLS (Eng-Ina) dataset. The results are shown in Table VI. The pipeline approach introduces error propagation. To solve this, we trained several end-to-end CLS models (4 variations of VCLS models and 4 variations of MCLS models) on CLS (Eng-Ina) dataset. The performance of these end-to-end models can also be seen in Table VI. All the end-to-end models outperform the model that uses the pipeline scheme. The end-to-end model can achieve improvement from +0.1298 ROUGE-1, +0.0763 ROUGE-2, +0.1083 ROUGE-L up to +0.2981 ROUGE-1, +0.2084 ROUGE-2, +0.2771 ROUGE-L compared to the pipeline baselines. This verifies the purpose of performing end-to-end summarization. Unlike the pipeline method, the end-to-end model performs CLS in one direct step. This approach avoids the error propagation that occurs when doing it in two steps, namely in the pipeline scheme.

TABLE IV. PERFORMANCE OF OUR MONOLINGUAL SUMMARIZATION

Model	ROUGE-1	ROUGE-2	ROUGE-L
SumINA	0.6456	0.5580	0.6355
SumENG	0.6434	0.5123	0.6088

TABLE V. PERFORMANCE OF OUR MACHINE TRANSLATION

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Tran	0.3683	0.2078	0.1191	0.0628

TABLE VI. ROUGE F1 SCORE

Type	Model	Variations	ROUGE-1	ROUGE-2	ROUGE-L
Pipeline	PipeTS	Translation → Summarization	0.1266	0.0142	0.1038
	PipeST	Summarization → Translation	0.1661	0.0278	0.1313
End-to-End	VCLS	Head + CLS	0.2959	0.1041	0.2396
	MCLS	Head + CLS-MS	0.4087	0.2112	0.3652
	VCLS_T	Head-Tail + CLS	0.2981	0.1093	0.2448
	MCLS_T	Head-Tail + CLS-MS	0.3809	0.1893	0.3385
End-to-End MWE	MWE_VCLS	Head + CLS + MWE	0.3617	0.1519	0.3038
	MWE_MCLS	**Head + CLS-MS + MWE**	**0.4247**	**0.2226**	**0.3809**
	MWE_VCLS_T	Head-Tail + CLS + MWE	0.3347	0.1311	0.2767
	MWE_MCLS_T	Head-Tail + CLS-MS + MWE	0.3886	0.1919	0.3438

B. *Comparison between End-to-End CLS Model with MWE and End-to-End CLS Model without MWE*

We also use two end-to-end models as baseline besides the two pipeline baselines which were discussed in the previous section. The first end-to-end baseline is an end-to-end CLS model built using a vanilla transformer (VCLS). The second end-to-end baseline is an end-to-end CLS model based on Zhu et al. [15] (MCLS). The model is trained with 2-task learning, which is jointly training CLS and MS using CLS (Eng-Ina) dataset and Sum (Eng) dataset.

CLS is a Seq2Seq problem, when given a text sequence in the source language it produces a shorter version of an original text sequence in the target language. In Seq2Seq problem, it is important to learn the relation (alignment) between the input sequence and the output sequence. The CLS model that cannot do a proper alignment mapping is unable to produce a correct and meaningful summary. We propose using MWE to help with this alignment. MWE can represent words in various languages in one vector space. Words that have similar meanings are in proximity. This pre-mapping of words is expected to facilitate the model in aligning word sequences between input and output. The VCLS and MCLS are then modified by adding MWE components (MWE_VCLS and MWE_MCLS).

The experimental results can be seen in Table VI. We can find that the model that utilized MWE components beats the underlying baseline model that does not use it. This shows that the MWE helps to better map the relation (alignment) between source document input in English and its summary output in Bahasa Indonesia. This can happen because every word in both languages is already in the same vector space. The MWE_MCLS model can achieve maximum improvement up to +0.1288 ROUGE-1, +0.1185 ROUGE-2, and +0.1413 ROUGE-L when compared to the end-to-end baselines. This is the best-performing model in our experiments.

C. *Input Truncation Strategy*

Sun et al. [47] and Mutasodirin & Prasojo [48] use an input truncation strategy on text classification problems. The strategy is not only to take the head of the data as input but also to combine it with the tail of the data. Generally, only the head of the data is used because it is assumed that the core information is here. However, their experimental results did not reach the same conclusion, so this strategy cannot be generalized yet. We adapt this strategy to our CLS models which previously only took the head part of the document into a combination of the head and the tail. This adaptation produces models VCLS_T, MCLS_T, MWE_VCLS_T, and MWE_MCLS_T. The experimental results of these four models can be seen in Table VI. However, in general, the results have not been able to exceed the score obtained by the model that only uses the head of the document. The best score is obtained by MWE_MCLS_T. A decrease in performance of -0.0361 ROUGE-1, -0.0307 ROUGE-2, and -0.0371 ROUGE-L when compared to the best model uses the head truncation strategy (MWE_MCLS).

The IndoSum dataset that we use is constructed from online news articles. Upon further examination of this dataset, we found that most of the important information is at the beginning of the document. This information appears in the summary, while the end of the news article generally contains additional information or explanatory information that does not appear in the summary. Strategies that use a combination of head and tail from the document are proven to be unable to improve the performance of the models that originally uses head truncation strategy, it even can decrease the model performance due to including insignificant information in the summary.

VI. Conclusion

In this work, we present the end-to-end abstractive CLS for English documents to Bahasa Indonesia summary. The CLS architecture is based on transformers, modified by adding MWE components to address cross-lingual problems. The architecture also has a second decoder with a shared encoder. This second decoder is used only during training to carry out joint learning between CLS and MS. During the test, the second decoder is ignored. The model is also trained using two input truncation strategies: head and head-tail. The head truncation strategy cuts off and takes the head part of the document as input while the head-tail truncation strategy combines the head and the tail of the document as input. To train the model, we create a new CLS dataset from MS dataset by adapting the round-trip translation technique. The resulting CLS dataset is evaluated using BLEU to ensure its quality.

Based on the experimental results, it can be concluded that the use of MWE improves the performance of the CLS model, specifically for summarizing an English source document into a Bahasa Indonesia summary. The proposed model successfully outperformed the baseline model and improved its performance. The strategy of utilizing information at the end of the data failed in improving the performance of the model. Using only the head part of the data is still better. Furthermore, it can be concluded that the end-to-end model is better than the pipeline model. The use of machine translation is a weak point of the pipeline model because it introduces error propagation.

Acknowledgment

This work was funded by PUTI Q3 Universitas Indonesia 2020-2021 with number: NKB-4375/UN2.RST/HKP.05.00/2020 and supported as part of Visiting Professorship Project on Gaining International Accreditation and Enhancing Academic Reputation Program Universitas Indonesia 2020-2021.

References

[1] D. Radev, E. Hovy and K. McKeown, "Introduction to the special issue on summarization," Computational linguistics, vol. 28, p. 399–408, 2002.

[2] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A survey of automatic text summarization: Progress, process and challenges," IEEE Access, vol. 9, p. 156043–156070, 2021.

[3] Y.-F. Ma, L. Lu, H.-J. Zhang and M. Li, "A user attention model for video summarization," in Proceedings of the tenth ACM international conference on Multimedia, 2002.

[4] J. Xu and T.-C. Lu, "Seeing the big picture from microblogs: Harnessing social signals for visual event summarization," in Proceedings of the 20th International Conference on Intelligent User Interfaces, 2015.

[5] K. Zhang, W.-L. Chao, F. Sha and K. Grauman, "Video summarization with long short-term memory," in European conference on computer vision, 2016.

[6] H. P. Luhn, "The automatic creation of literature abstracts," IBM Journal of research and development, vol. 2, p. 159–165, 1958.

[7] C. D. Paice, "Constructing literature abstracts by computer: Techniques and prospects.," Inf. Process. Manage., vol. 26, p. 171–186, 1990.

[8] J. Kupiec, J. Pedersen and F. Chen, "A trainable document summarizer," in Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, 1995.

[9] S. Chopra, M. Auli and A. M. Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016.

[10] R. Nallapati, B. Zhou, C. Gulcehre, B. Xiang and others, "Abstractive text summarization using sequence-to-sequence rnns and beyond," arXiv preprint arXiv:1602.06023, 2016.

[11] D. Bahdanau, K. Cho and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[12] B. Sankaran, H. Mi, Y. Al-Onaizan and A. Ittycheriah, "Temporal attention model for neural machine translation," arXiv preprint arXiv:1608.02927, 2016.

[13] X. Wan, H. Li and J. Xiao, "Cross-language document summarization based on machine translation quality prediction," in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[14] F. Boudin, S. Huet and J.-M. Torres-Moreno, "A graph-based approach to cross-language multi-document summarization," Polibits, p. 113–118, 2011.

[15] J. Zhu, Q. Wang, Y. Wang, Y. Zhou, J. Zhang, S. Wang and C. Zong, "NCLS: Neural cross-lingual summarization," arXiv preprint arXiv:1909.00156, 2019.

[16] X. Duan, M. Yin, M. Zhang, B. Chen and W. Luo, "Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019.

[17] X. Chen and C. Cardie, "Unsupervised Multilingual Word Embeddings," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, 2018.

[18] A. F. Abka, M. Pratama and W. Jatmiko, "Cross-Lingual Summarization: English - Bahasa Indonesia," in 2021 6th International Workshop on Big Data and Information Security (IWBIS), 2021.

[19] R. Sennrich, B. Haddow and A. Birch, "Improving Neural Machine Translation Models with Monolingual Data," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016.

[20] G. Lample, A. Conneau, L. Denoyer and M. Ranzato, "Unsupervised Machine Translation Using Monolingual Corpora Only," in International Conference on Learning Representations, 2018.

[21] K. Papineni, S. Roukos, T. Ward and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002.

[22] W. Ogden, J. Cowie, M. Davis, E. Ludovik, H. Molina-Salgado and H. Shin, "Getting information from documents you cannot read: An interactive cross-language text retrieval and summarization system," in Joint ACM DL/SIGIR workshop on multilingual information discovery and access, 1999.

[23] H. Saggion, D. R. Radev, S. Teufel, W. Lam and S. M. Strassel, "Developing Infrastructure for the Evaluation of Single and Multi-document Summarization Systems in a Cross-lingual Environment.," in LREC, 2002.

[24] L. Yu and F. Ren, "A study on cross-language text summarization using supervised methods," in 2009 international conference on natural language processing and knowledge engineering, 2009.

[25] X. Wan, "Using bilingual information for cross-language document summarization," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011.

[26] X. Wan, F. Luo, X. Sun, S. Huang and J.-g. Yao, "Cross-language document summarization via extraction and ranking of multiple summaries," Knowledge and Information Systems, vol. 58, p. 481–499, 2019.

[27] J.-g. Yao, X. Wan and J. Xiao, "Phrase-based compressive cross-language summarization," in Proceedings of the 2015 conference on empirical methods in natural language processing, 2015.

[28] E. L. Pontes, S. Huet and J.-M. Torres-Moreno, "A Multilingual Study of Compressive Cross-Language Text Summarization," in Mexican International Conference on Artificial Intelligence, 2018.

[29] E. L. Pontes, S. Huet, J.-M. Torres-Moreno and A. C. Linhares, "Cross-language text summarization using sentence and multi-sentence compression," in International Conference on Applications of Natural Language to Information Systems, 2018.

[30] J. Zhang, Y. Zhou and C. Zong, "Abstractive cross-language summarization via translation model enhanced predicate argument structure fusing," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, p. 1842–1853, 2016.

[31] J. Ouyang, B. Song and K. McKeown, "A robust abstractive system for cross-lingual summarization," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019.

[32] M. Junczys-Dowmunt, R. Grundkiewicz, T. Dwojak, H. Hoang, K. Heafield, T. Neckermann, F. Seide, U. Germann, A. Fikri Aji, N. Bogoychev, A. F. T. Martins and A. Birch, "Marian: Fast Neural Machine Translation in C++," in Proceedings of ACL 2018, System Demonstrations, Melbourne, 2018.

[33] A. See, P. J. Liu and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," arXiv preprint arXiv:1704.04368, 2017.

[34] F. Ladhak, E. Durmus, C. Cardie and K. McKeown, "WikiLingua: A new benchmark dataset for cross-lingual abstractive summarization," arXiv preprint arXiv:2010.03093, 2020.

[35] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis and L. Zettlemoyer, "Multilingual denoising pre-training for neural machine translation," Transactions of the Association for Computational Linguistics, vol. 8, p. 726–742, 2020.

[36] A. Conneau, G. Lample, M. Ranzato, L. Denoyer and H. Jégou, "Word Translation Without Parallel Data," arXiv preprint arXiv:1710.04087, 2017.

[37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," Advances in neural information processing systems, vol. 27, 2014.

[38] B. Heinzerling and M. Strube, "BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages," in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, 2018.

[39] M. Artetxe and H. Schwenk, "Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond," Transactions of the Association for Computational Linguistics, vol. 7, p. 597–610, 2019.

[40] 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.

[41] S. Edunov, M. Ott, M. Auli and D. Grangier, "Understanding back-translation at scale," arXiv preprint arXiv:1808.09381, 2018.

[42] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, 2019.

[43] K. Kurniawan and S. Louvan, "Indosum: A new benchmark dataset for indonesian text summarization," in 2018 International Conference on Asian Language Processing (IALP), 2018.

[44] T. W. Guntara, A. F. Aji and R. E. Prasojo, "Benchmarking multidomain english-indonesian machine translation," in Proceedings of the 13th Workshop on Building and Using Comparable Corpora, 2020.

[45] A. Hermanto, T. B. Adji and N. A. Setiawan, "Recurrent neural network language model for English-Indonesian Machine Translation: Experimental study," in 2015 International conference on science in information technology (ICSITech), 2015.

[46] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proceedingsof the International Conference on Learning Representations (ICLR), 2015.

[47] C. Sun, X. Qiu, Y. Xu and X. Huang, "How to fine-tune bert for text classification?," in China national conference on Chinese computational linguistics, 2019.

[48] M. A. Mutasodirin and R. E. Prasojo, "Investigating Text Shortening Strategy in BERT: Truncation vs Summarization," in 2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2021.

[49] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," in Text Summarization Branches Out, Barcelona, 2004.

[50] C.-Y. Lin, G. Cao, J. Gao and J.-Y. Nie, "An information-theoretic approach to automatic evaluation of summaries," in Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, 2006.

[51] K. Ganesan, ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks, 2018.

Analysis of Content Based Image Retrieval using Deep Feature Extraction and Similarity Matching

Anu Mathews[1], Sejal N[2], Venugopal K R[3]
Dept. of AI&ML, K S Institute of Technology, Bengaluru, India[1]
Dept. of AI&ML, BNM Institute of Technology, Bengaluru, India[2]
Bangalore University, Bengaluru, India[3]

***Abstract*—Image retrieval using a textual query becomes a major challenge mainly due to human perception subjectivity and the impreciseness of image annotations. These drawbacks can be overcome by focusing on the content of images rather than on the textual descriptions of images. Traditional feature extraction techniques demand for expert knowledge to select the limited feature types and are also sensitive to changing imaging conditions. Deep feature extraction using Convolutional Neural Network (CNN) are a solution to these drawbacks as they can learn the feature representations automatically. This work carries out a detailed performance comparison of various pre-trained models of CNN in feature extraction. Features are extracted from men footwear and women clothing datasets using the VGG16, VGG19, InceptionV3, Xception and ResNet50 models. Further, these extracted features are used for classification using SVM, Random Forest and K-Nearest Neighbors classifiers. Results of feature extraction and image retrieval show that VGG19, Inception and Xception features perform well with feature extraction, achieving a good image classification accuracy of 97.5%. These results are further justified by performing a comparison of image retrieval efficiency, with the extracted features and similarity metrics. This work also compares the accuracy obtained by features extracted by the selected pre-trained CNN models with the results obtained using conventional classification techniques on CIFAR 10 dataset. The features extracted using CNN can be used in image-based systems like recommender systems, where images have to be analyzed to generate item profiles.**

Keywords*—Convolutional neural network; deep learning; feature extraction; accuracy; similarity

I. Introduction

Image retrieval systems browse, search and retrieve visually similar images from large image databases. Traditional image retrieval methods utilize the image annotations for obtaining the metadata that help in finding the similar images, but this is a laborious process and is subjective to human perceptions. Efficient image retrieval is the backbone of most of the search engines and recommender systems. Search engines may not be able to retrieve relevant information according to user's preferences due to imperfect textual query, less knowledge about the search query or due to wrong tagging of the database images. This gap between real intention of a user's search and his understanding of the object is called as semantic gap [1]. With the large amount of image data encountered in social networks, online e-commerce website, medical imaging, etc., it is a challenging problem to search the humongous databases for similar images, especially when it comes to real time image retrieval. Feature extraction techniques help in getting a representation of the attributes of an image, which gives information about the image contents and hence helps in efficient retrieval of visually similar images. Feature extraction transforms data into more informative forms with efficient representation for analysis and classification [2]. Advantages of feature extraction include:

- Reduction of redundancy.
- Reduction in the number of processing resources.
- Helps in avoiding overfitting problem in machine learning models.
- Increase in accuracy.

Motivation: Any image recognition or image retrieval task requires a good feature representation in order to achieve high performance. However, since it is not feasible to define a good feature set manually, feature extraction plays a major role in image retrieval tasks. Features obtained using traditional methods of feature extraction are not capable of expressing the semantic information of the image. Deep Learning techniques make the task of feature extraction automatic and more efficient. Since deep learning techniques require a large amount of labelled training data, transfer learning can be used to reduce this overhead [3]. Fine tuning a CNN allows pretrained models to be used in a new task with a new dataset.

CNN models have two stages, of which, the first is feature extraction and the second is classification. This paper focusses on feature extraction using pretrained CNN and comparing their performance.

The open research questions which are addressed in this work include:

i) How effective are deep learning methods for learning good feature representations from images for Content Based Image Retrieval?

ii) How do the feature extraction and image retrieval time vary with the different models?

A. *Contributions:*

- This work contributes to the existing knowledge of feature learning by exploring feature extraction and transfer learning techniques using the five selected pre-trained CNN architectures.

- Comparison of the goodness of the features extracted by the selected CNN models by comparing the image retrieval performance using similar image search technique and also by using the extracted features for image classification.

The following sections of this paper are organized as follows. Related work in the domain of feature extraction and image retrieval is discussed in Section II. The fundamental concepts in image retrieval and the methodology adopted are given in Section III. Section IV gives the framework and the algorithm used in Reverse Image Search. Section V discusses the implementation details and results in Feature Visualization and Feature Extraction. Section VI concludes the paper by discussing the open research issues in the field of image retrieval and deriving an outlook for further research.

II. RELATED WORK

The goodness of an effective image retrieval system is attributed to the effectiveness of image feature extraction techniques used by the system. Various techniques have been proposed for feature extraction, starting from extraction of handcrafted features using traditional feature extraction techniques like SIFT, SURF, HOG to recent techniques utilizing Deep Learning. Work proposed in [4] investigated the use of various color, texture and shape features for image retrieval in CBIR (Content Based Image Retrieval) and biometrics systems, by mapping the image content onto low-level features.

An integrated image representation model has been proposed by complementary fusion of SIFT and CNN, to describe contents at multiple levels in images, followed by a series of operations of L2 normalization, PCA and whitening, on the integrated representation for a more compact representation [5].

The various techniques for extraction of texture and color features include color correlogram, color histogram, color co-occurrence matrix, wavelet transform and Tamura feature [6]. Several recent works have focused on the use of machine learning and Deep Learning for feature extraction [7-9]. The convolution layers in CNN can change the shape of the output, thereby enabling the learning of basic object shapes in the primary layers and more complex objects in the deeper layers, with a drastic reduction in the error rate [10].

Studies reveal that high level features are of different behavior from low and middle level features under certain conditions, and hence, CNN features in each of the layers can be utilized for representation of knowledge. The level of noise included in these features have been studied and a thresholding approach is proposed to remove as much noise as possible, thereby generating efficient CNN embedding spaces [11].

Some of the models with combination of CNN features and handcrafted images have also shown good results. The work in [12] automates the extraction of electroencephalogram (EEG) signals the feature extraction methods of Wavelet Packet decomposition and Genetic Algorithm-Based Frequency-Domain Feature Search. A combination of handcrafted and CNN features has been used for liver MR image adequacy assessment yielding a model performance of AUC = 0.94, wherein, HC features of intensity values, topological structure, texture information (GMM, ECC, and GLCM features) are extracted from the images and used for classification using Random Forest classifier.

Many algorithms use Hashing based methods for image retrieval. While some of the algorithms identified approximate regions of the objects in an image using attention sub-network, focusing mainly on the foreground objects [13], image features extracted using deep hashing methods with deep neural networks (DNN) gave better image retrieval performance [14].

Histogram features have also given good results in CBIR. Work proposed in [15] used color spectral histogram for computing the similarity between images, whereas the method proposed in [16] used correlated primary visual texton histogram features.

The use of color and texture features have been proposed in several works. The work proposed in [17], calculates the image similarity by giving equal priority to dominant color Hu moments for CBIR, wherein, the texton template detects and extracts the consistent zone of an image and uses the seed point's selection approach for initiating the clustering process. Study [18] uses both Hu moments and dominant color descriptor on the pixels. Research [19] uses LBP-DBN training algorithm with adaptive momentum learning rate to overcome the issues of longer training time and reduction in classification accuracy. The author in [20] uses Directional Magnitude Local Hexadecimal Patterns, for image retrieval, by reducing the semantic gap problem using a learning- based approach. Quadruplet loss function and feature fusion are also used for clothing retrieval [21]. The features extracted using Resnet-50 are merged with middle-level features to get a combined feature representation. Work proposed in [22] uses a visual re-ranking approach, by using a correlation matrix of an image retrieval list and a CNN model learns the relevance of each of the image pairs simultaneously. The model is further optimized using a weighted MSE loss, which also considers the sparsity of labels. Feature reduction using an improved CNN and PCA quadratic dimensionality reduction has been proposed [23]. Deep Learning techniques for feature extraction are more promising due to their capability to extract and learn features of an image automatically.

Recent advances in image retrieval have started focusing on graph neural networks for CBIR. The work done in [24] proposes graph neural networks to reframe the re-ranking process for CBIR by using a 2-layer GNN to aggregate neighbor information of the entire data. Several new approaches have been proposed for deep feature embedding. In [25] a GNN characterizes and predicts the local correlation structure of images in the feature space, using which, neighboring images collaborate and refine their feature embeddings based on local linear combination. The representation capability of graphs has led to various graph-based image recommendation systems also [26]. Table I gives a performance comparison of some of the results obtained in CBIR.

TABLE I. PPERFORMANCE COMPARISON OF REVIEWED CBIR SYSTEMS

Author	*Method*	*Dataset*	*Performance*
Salamh et al. [7]	Multi-descriptor-local binary patterns	Extended Yale B & Grimace database	Accuracy=99.4%
Alsuwaiket et al.[9]	Wavelet Packet decomposition (WPD), Genetic Algorithm-Based Frequency-Domain Feature Search (GAFDS)	Children's Hospital Boston: EEG recordings of 23 cases	CNN: 97.93% accuracy SVM: 94.49% accuracy RF : 88.03% accuracy
Manjunath et al. [10]	AlexNet inspired CNN	CIFAR-10	KNN:28.2% accuracy SVM:37.4% accuracy CNN: 85.97% accuracy
Y. Zhang et al. [13]	Supervised hashing method which fusions simplex feature similarity & location similarity among multiple objects	VOC 2007 VOC 2012 NUS-WIDE	6% increase on Weighted MAP -3.0% increase on NDCG@1000 -2.9% on ACG@1000
X. Zhang [19]	Classification: Image local feature multi-level clustering and image-class nearest	RPC: Defiance Technology Nanjing Research Institute	Average Classification accuracy=84% as compared to ISDH(Instance Similarity Deep Hashing)

III. FUNDAMENTAL CONCEPTS AND METHODOLOGY

The basic steps involved in image retrieval include extraction of features and computation of image similarity. The features of the collection of database images are extracted and stored. The query image undergoes the feature extraction process and subsequently the similarity matching algorithm computes the similarity between the features of query image and the features of the database images. Distance metrics like Cosine distance and Euclidean distance are used for similarity computation. Finally, top-n similar images are retrieved.

A. CNN Models

Convolutional Neural Networks use supervised learning and train the network by backpropagation. CNNs have the advantage of reduced amount of preprocessing when compared to other deep learning networks. A CNN consists of two basic parts of feature extraction and classification.

The feature extraction part has several convolution layers followed by max pooling layers and an activation function. The classifier part consists of fully connected layers. Normalization layers in CNN help in keeping signals from each layer at suitable levels.

The output size at each of the convolution layer is given by the following formula:

$$\text{Output size} = [(W-F+2P)/S] + 1$$

where, W is the size of the input image, F is the size of the receptive field, S is the stride, and P indicates the zero-padding used on the border.

1) ResNet50: The ResNet model was developed by He et al. in the year 2015 with the concept of residual connections [27]. ResNet50 consisting of 50 weighted layers, has five times lesser memory requirement as compared to VGG model because of the usage of Global Average Pooling layer, converting the 2D feature maps of the last layer to an n-classes vector calculating the probability of belonging to each class.

Fig. 1 shows the Residual block which is the basic building block of ResNet model. In a plain CNN network, the input X gets transformed into H(X) by passing through the different layers. In the residual model, the identity connection transforms the input X into F(X)+X, i.e., the original H(x) gets modified into F(X)+X. The ResNet model can learn from much deeper network, as the identity mapping from the input acts as a shortcut for the gradient to pass through, thereby avoiding the vanishing gradient problem.

2) VGG16: VGG 16 architecture proposed by Simonyan and Zisserman in 2014 [28], consists of blocks with increasing number of convolutional layers with 3x3 filters as shown in Fig. 2.

The size of the activation maps reduces by half due to the max-pooling blocks between the convolutional blocks. The classification blocks consist of two layers, each with 4096 neurons, followed by the final output layer of 1000 neurons.

3) VGG19: VGG19 model consists of 16 convolution layers, three Fully connected layers, five Max Pool layers and one SoftMax layer. RGB image of dimension (224 * 224) is given as input and the kernels used are of (3 * 3) size with a stride of 1. Max pooling over a (2 * 2) window with stride 2, followed by Rectified linear unit introduces non-linearity for better classification and improved computation time as compared to the models using sigmoid or tanh functions. Two fully connected layers of size 4096 is followed by a layer with 1000 channels and final layer is softmax function.

4) InceptionV3: Szegedy et al. proposed the Inception V3 architecture [29]. This model uses varying sizes filters and a concatenation of these filters is used to extract features at different scales as shown in Fig. 3. In Inception, the input is compressed using 1x1 convolutions after which, filters of different sizes are used on each of these spaces.

5) Xception: The Xception model is an "extreme" version of the Inception module that involves "Depthwise Separable Convolutions" [30]. The number of computations is high in the case of a basic convolution operation because the operation of applying filters on every input channel and the combining of these values is done in one step.

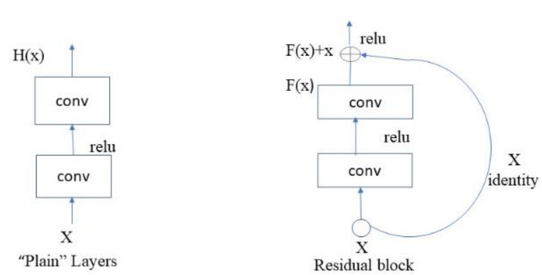


Fig. 1. Residual block in ResNet model.

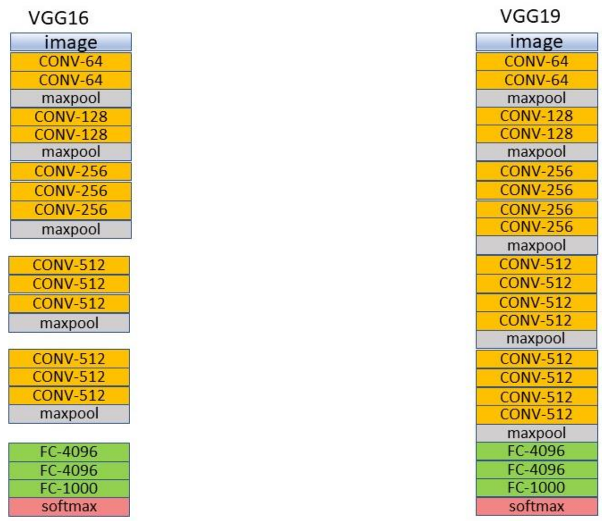


Fig. 2. VGG16 and VGG19 models.

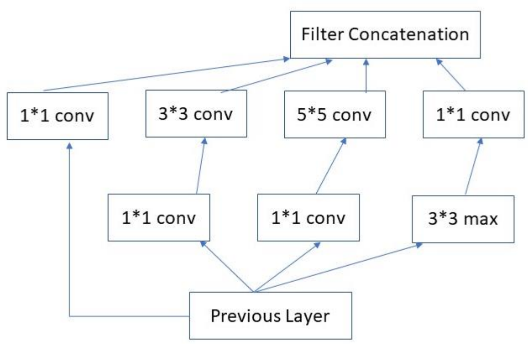


Fig. 3. The inception module.

Depthwise separable convolutions separates these steps into two, the first being the depthwise separation which is the filtering stage, followed by combining stage of pointwise convolution. Depthwise convolution applies convolution to the input image ($D_f * D_f * M$) with filters of depth 1 ($D_k * D_k * 1$), to a single input channel at a time as shown in equation (1) and then pointwise convolution compresses the input by applying N filters of dimension 1*1*M across the depth as shown in equation (2).

$$D_f * D_f * M \rightarrow D_g * D_g \quad (1)$$

$$D_g * D_g * M \rightarrow D_g * D_g * N \quad (2)$$

where D_f is dimension of input image, D_g is dimension of feature map and M is number of channels in input image

Table II gives a summary of the five pre-trained CNN models used in feature extraction.

B. Fine Tuning CNN

Fine tuning is the technique for speeding up the training process by overcoming the problems posed by small size dataset, by taking the weights of a pre-trained neural network and using it to initialize a new model which is being trained on similar domain data. This can be broadly categorized into Transfer Learning and Feature Extraction.

TABLE II. SUMMARY OF CNN MODELS USED IN FEATURE EXTRACTION

Model	*Input Size*	*Length of Feature Vector*	*Parameters*
Inception V3 (2015)	299*299*3	2048	Total parameters: 23,851,784 Trainable parameters: 23,817,352 Non-trainable parameters: 34,432
ResNet50 (2015)	224*224*3	2048	Total parameters: 25,636,712 Trainable parameters 25,583,592 Non-trainable parameters: 53,120 SGD - mini-batch size = 256 Learning rate starts from 0.1 and is divided by 10 when the error plateaus Weight decay = 0.0001 Momentum = 0.9 No Dropout
VGG16 (2014)	224*224*3	4096	Total parameters: 138,357,544 Trainable parameters: 138,357,544 Non-trainable parameters: 0 SGD(lr=0.1,decay=1e-6, momentum=0.9)
VGG19 (2015)	224*224*3	4096	Total parameters: 143,667,240 Trainable parameters 143,667,240 Non-trainable parameters 0
Xception (2015)	299*299*3	2048	Total parameters: 22,910,480 Trainable parameters: 22,855,952 Non-trainable parameters: 54,528

Most of the popular CNN architectures including VGG16, VGG19, ResNet50, AlexNet, InceptionV3, Xception have been pre-trained on ImageNet dataset. ImageNet consists of around 1.2 million training images, 50,000 validation images and 100,000 testing images, belonging to 1000 categories.

1) Transfer learning: Transfer learning works by loading pre-trained weights into a base CNN model, followed by freezing all the base model layers, making them non-trainable. This is followed by creating a new model on top of the output of any of the base model layers and training this newly created model on the new dataset.

2) Feature extraction: Feature Extraction is a lighter approach, wherein, after loading pre-trained weights into a base CNN model, the new dataset is run through the base model to extract the output of the layers of this model, and use this output as the input data for a new model.

C. Classification Algorithms

Classification is a “Supervised Learning” technique for identification of the category of new observations based on the training data. In classification, a program learns from the given observations and then classifies new observations into categories. In this work, classifiers are used for classification of images by utilizing the features extracted by the CNN models on the fashion dataset having men footwear and women clothing.

1) Random forest: Random Forest classifiers use a combination of many classifiers to solve complex problems. The advantages of Random Forest include its capability to maintain high accuracy through cross validation with higher dimensionality dataset, prevention of overfitting and the capability to handle missing data.

2) Support Vector Machine: Support Vector Machine (SVM) is a "supervised" machine learning algorithm that finds a hyperplane in an N-dimensional space that distinctly classifies the data points. SVM is very effective in high dimensional cases with the number of features deciding the dimension of the hyperplane. SVM maximizes the margin between the data points and the hyperplane.

3) K Nearest Neighbors: KNN is a "supervised learning" algorithm for regression and classification. KNN takes into consideration K nearest data points to predict the category or continuous value of the newly observed data. KNN uses all of the training instances for output prediction for new data. Model learning process is performed only at the time when prediction is requested on the new instance.

D. *Similarity Metrics*

The similarity metrics used for similarity matching are:

- Minkowski Distance: This distance metric takes two vectors and computes the distance between these vectors. The parameter "p" is called the "order", which allows calculation of different distance measures.

$$d(x,y) = (\sum_{i=1}^{n}|x_i - y_i|^p)^{1/p} \quad (3)$$

For Manhattan distance, p = 1.

For Euclidean distance p = 2.

- Manhattan Distance: This distance is computed as the sum of the absolute differences between two vectors.

$$d(x,y) = \sum_{i=1}^{n}|x_i - y_i| \quad (4)$$

- Euclidean Distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (5)$$

- Cosine Similarity:

$$sim(a,b) = \frac{a.b}{||a||.||b||} \quad (6)$$

- Jaccard Similarity:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (7)$$

IV. Reverse Image Search Framework and Algorithm

Problem Definition: For a given image query, retrieve the top-n similar images.

The features extracted using the different CNN models are utilized for reverse image search, using the framework shown in Fig. 4. For dimensionality reduction, Principal Component Analysis is used. Manhattan, Euclidean, Cosine distances and Jaccard similarity are used as the similarity measures.

A. *Framework*

The Reverse Image Search framework for retrieving visually similar images, shown in Fig. 4, consists of the following modules:

1) Data collection and preprocessing module: The Crawler utility fetches the retailer-specific webpages and extracts images. The preprocessing module prepares the images for the feature extraction process according to the input specifications of the selected pre-trained model. The image preprocessing utilities transform the raw image data to dataset objects, that can be further used as inputs to the model.

2) Feature extraction module: Text-based image retrieval methods may fail to retrieve visually similar relevant images because of the absence of query terms in the description of image. The feature extraction module overcomes this disadvantage by incorporating CNN layers for extraction of features. This module consists of the specific CNN model with the last fully connected layers removed, to extract the various levels of features. The extracted feature vectors are of high dimensionality and are therefore computationally expensive, with high memory requirement. Principal component analysis transforms the data into fewer number of dimensions, thereby giving summarized feature vectors. This helps to reduce the complexity in data and at the same time retains the patterns in the images.

3) Similarity Calculation module: This module computes the similarity between the feature vector of the input query image and the feature vectors of the database images extracted using the CNN models and the ranking of images is done according to their similarity. Finally, the top-n ranked images will be retrieved using the similarity metrics as in Equation (3) – Equation (7).

B. *Algorithm: Top-n similar Images Retrieval*

The algorithm for Top-n similar images retrieval is as follows:

Input : User query Image q, Image database I_{db} with image I.

Output: Top-n similar images

begin

Offline:

Pre-process the images in the image database I_{db} based on the input size of the CNN models

Extract Visual features F using the pre-trained CNN models for each image I in I_{db}

Reduce dimensionality using Principal Component Analysis

Online:

Extract Visual features F_q of the input query image q using the CNN models

for each feature vector Fj in F **do**

Compute the feature similarity *Sim(Fq, Fj)* using Equation (6)

Retrieve the Top-n similar images based on similarity calculation.

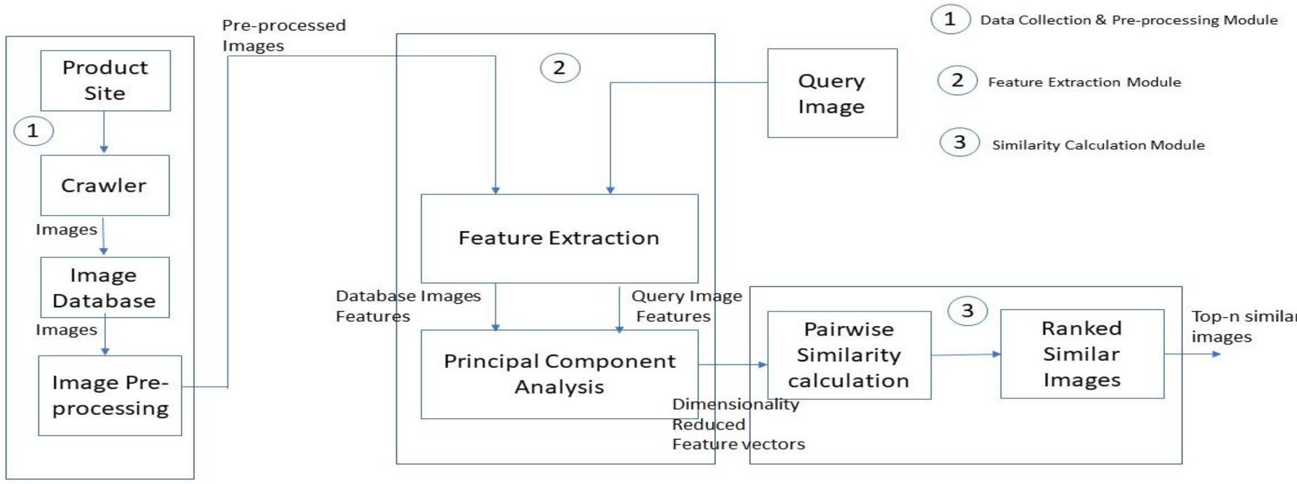


Fig. 4. Reverse image search framework.

V. Implementation and Results

A. Dataset and Experiment Setup

This work is carried out on 11th Gen Intel Core i7 processor with NVIDIA GeForce RTX 3050 Ti GPU. The men footwear dataset consists of a total of 2167 images belonging to three different classes of casual sandals, formal shoes and flip flops. The women clothes dataset consists of 4339 images belonging to five different classes of jeans, salwars, shirts, shorts and tops. The data is split into train and test sets with 80:20 ratio. The work concentrates on feature extraction using the pre-trained models VGG16, VGG19, InceptionV3, Xception and ResNet50 models, followed by evaluating the feature extraction efficiency by giving the extracted features as input to classifiers. Feature Visualization is done to know the depth and granularity of different features extracted at different layers.

B. Feature Extraction and Classification

This work utilizes the Keras Implementation of the pre trained CNN models. After initializing the input image size and applying the corresponding preprocessing functions, the pretrained weights are loaded into the model. The preprocessing function of the ImageDataGenerator module is used for data augmentation to improve generalization. Next, for feature extraction, the training images are loaded after expanding their dimensions according to the input size of the respective CNN models. Random Forest Classifier from Scikit learn library with n estimators' value of 50 is used in this work. SVC from the Scikit learn library is used here with a Linear kernel for SVM classification. KNN classifier with K=20 is used for the nearest neighbour classification.

For the men footwear dataset, as shown in Fig. 5, Random Forest gave an image classification accuracy of 97.5% with VGG19 features, SVM gave an accuracy of 97.5% with Xception features and KNN gave an accuracy of 95.8% with Xception features.

Fig. 5. Classification accuracy with features extracted by the different CNN models on the men footwear dataset.

For the women clothes dataset, as shown in Fig. 6, RF and SVM gave image classification accuracy of 89.1% and 96.8% respectively with InceptionV3 features and KNN gave an accuracy of 85.9% with Xception features.

The accuracy results show the efficiency of our selected pre-trained CNN models in feature extraction as compared to the work proposed in [10], where the authors achieved classification accuracy of 28.2% and 37.4% with KNN and SVM respectively on CIFAR 10 dataset, which consists of objects belonging to 10 classes. The results obtained with pretrained CNN model feature extraction and classification, outperform these with a classification accuracy of 72.4% and 86% with KNN and SVM classifiers respectively, using the features extracted by the Xception model on CIFAR 10 dataset.

Fig. 7 gives the classification accuracy obtained with the features extracted by the pretrained models on the CIFAR 10 dataset, along with the results obtained in [10] given as M1.

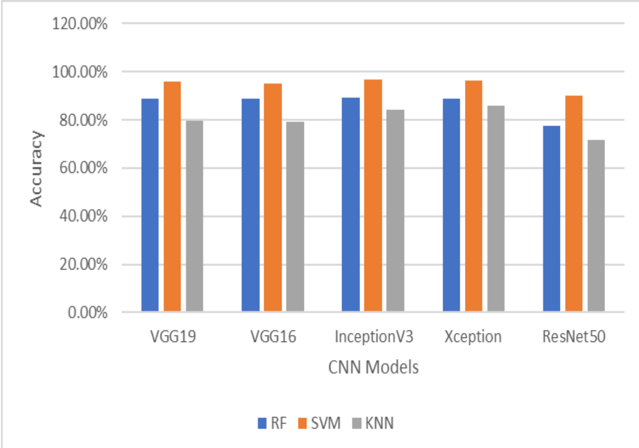


Fig. 6. Classification accuracy with features extracted by the different CNN models on the women clothes dataset.

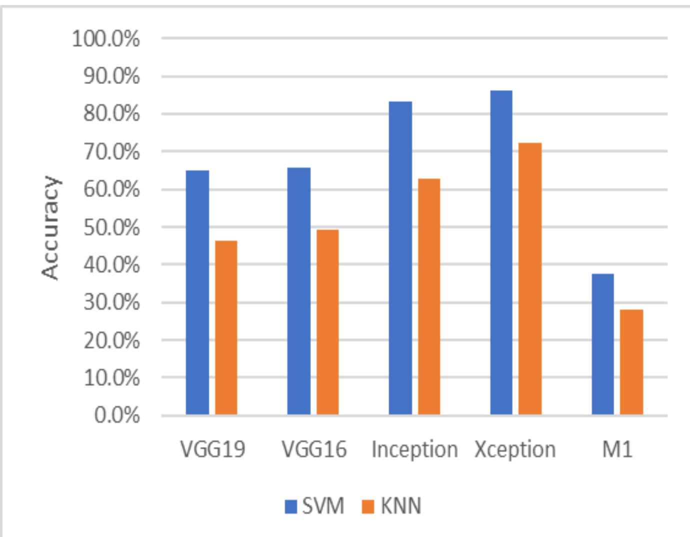


Fig. 7. Classification accuracy on the CIFAR 10 dataset.

C. *Feature Visualization*

The application of different convolutional filters to the input image, results in different feature maps, wherein, each pixel of each feature map is an output of the convolutional layer. Visualization of feature map for an image helps in understanding the detected features. The feature maps closer to the input layer of the neural network detect details of features like edges. The feature maps closer to the output layer of the model capture much more abstract features like texture.

Fig. 8 shows the feature map visualization of the different blocks of convolution layers of InceptionV3 model on a men footwear input image.

D. *Reverse Image Search*

The image retrieval results with features extracted by the five CNN models on Men footwear and women clothes dataset using Manhattan, Euclidean, Cosine and Jaccard similarity as the similarity measures, is compared in this section, in terms of the time taken for feature extraction and also the visual similarity of the retrieved images. The utilities for image preprocessing is imported from tf.keras.preprocessing. This transforms the raw image data to a tf.data.Dataset object, that can be further used to train the model. The extracted feature vectors are of high dimensionality and have high memory requirement. For dimensionality reduction, Principal Component Analysis with 300 components is used.

1) Time taken for feature extraction and image retrieval: The Men Footwear train dataset consists of 1732 training images belonging to three classes and the Women Clothes dataset consists of 3469 training images belonging to five classes. On an average, VGG19 takes more time for extracting the features of the training set images followed by VGG16, ResNet50, Inception V3, and Xception.

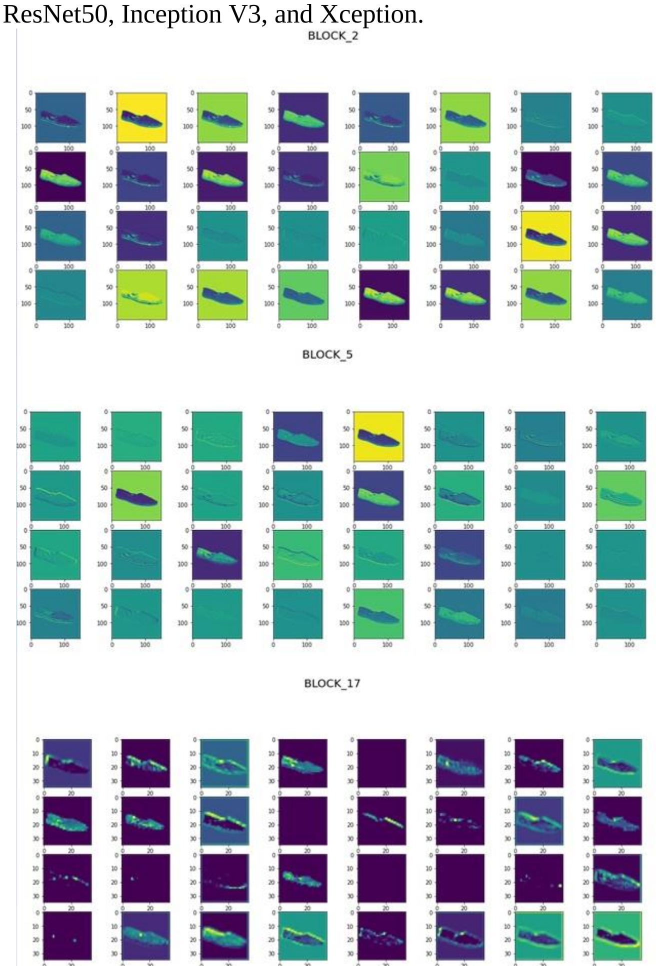


Fig. 8. Feature visualization using Inceptionv3 on men footwear image.

Fig. 9 shows the time taken for extraction of Men Footwear database image features by the different CNN models. Fig. 10 shows the time taken for query image feature extraction by the different CNN models and retrieval of top five similar images.

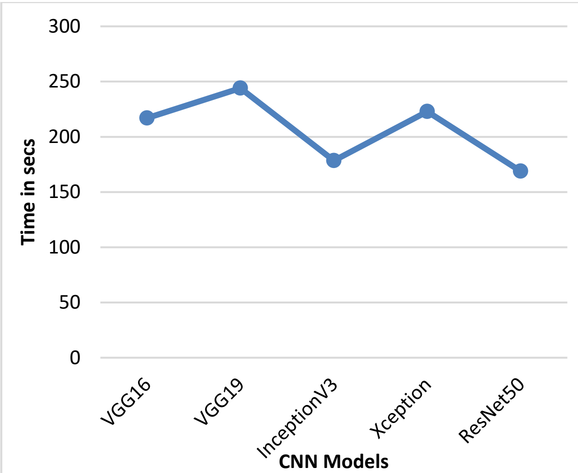


Fig. 9. Time taken in seconds for extraction of men footwear database image features by the different CNN models.

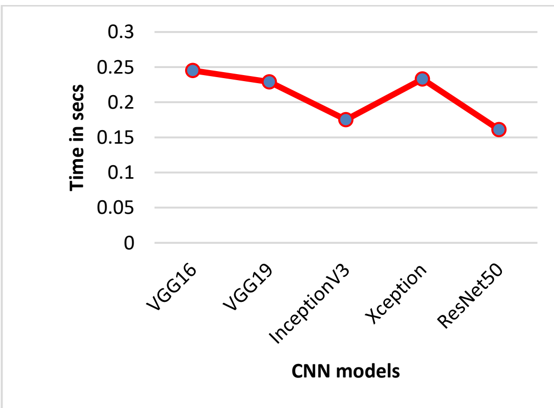


Fig. 10. Time taken in seconds for query image feature extraction by the different CNN models and retrieval of top 5 similar images.

Fig. 11 and Fig. 13 show the image retrieval results with features extracted by the five CNN models on Men footwear and women clothes dataset using Manhattan, Euclidean and Cosine similarity as the similarity measures.

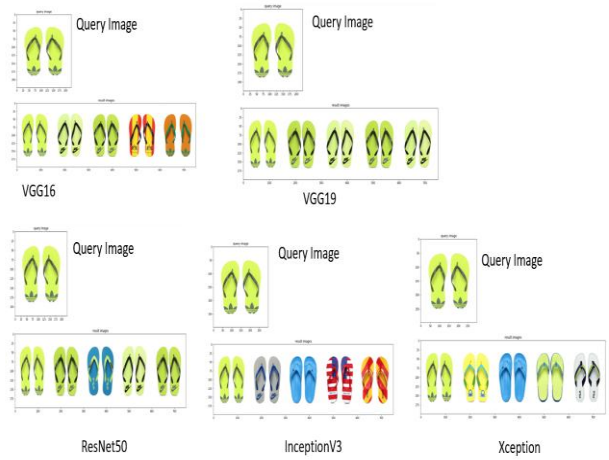


Fig. 11. Men footwear reverse image search with Cosine, Euclidean and Manhattan similarity metrics using the different CNN models for feature extraction (query image on top and top 5 result images below for each CNN model).

Image retrieval results with Jaccard similarity show poor results as compared to Manhattan, Euclidean and Cosine as can be seen from Fig. 12.

Fig. 12. Men footwear reverse image search with Jaccard similarity metrics using VGG19 and InceptionV3 CNN models for feature extraction (query image on top and top 5 result images below).

Fig. 13. Women clothes reverse image search with Cosine, Euclidean and Manhattan similarity metrics using the different CNN models for feature extraction (query image on top and top 5 result images below for each CNN model).

VI. Conclusion

The work carried out focuses on feature extraction using VGG19, VGG16, ResNet50, InceptionV3 and Xception models, followed by evaluating the feature extraction efficiency by feeding the extracted features as input to classifiers. Results of feature extraction and image retrieval show that VGG19 features show best results with both men footwear and women clothes dataset. The classification accuracy results show that VGG19 features give more accuracy with men footwear, whereas InceptionV3 and Xception features are better with women clothes. This work also compares the time taken for feature extraction and image retrieval with each of the five CNN models. The features extracted using the deep learning models can be used to analyze images in various image-based systems. One practical application of this would be in image-based recommender systems. The features extracted from images can be used to build item profiles in content-based image recommendations. Recent advances in the field of recommender systems use graph-based methods like Graph Convolutional Network, as they are capable of representing the complex embeddings.

Future research directions include:

- Utilization of Deep Learning models for transfer learning with optimization techniques.
- Using a fusion of different handcrafted local features and CNN features is also one of the challenging future directions for research.

References

[1] Enser, Peter, and Christine Sandom. "Towards a comprehensive survey of the semantic gap in visual image retrieval." In International Conference on Image and Video Retrieval, pp. 291-299. Springer, Berlin, Heidelberg, 2003.

[2] Salau, Ayodeji Olalekan, and Shruti Jain. "Feature extraction: a survey of the types, techniques, applications." In 2019 International Conference on Signal Processing and Communication (ICSC), pp. 158-164. IEEE, 2019.

[3] Zhuang, Fuzhen, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. "A comprehensive survey on transfer learning." Proceedings of the IEEE 109, no. 1 (2020): 43-76.

[4] Ryszard S Choras. Image feature extraction techniques and their applications for cbir and biometrics systems. *International journal of biology and biomedical engineering*, 1(1):6–16, 2007.

[5] Ke Yan, Yaowei Wang, Dawei Liang, Tiejun Huang, and Yonghong Tian. Cnn vs. sift for image retrieval: Alternative or complementary? In Proceedings of the 24th *ACM international conference on Multimedia*, pages 407–411, 2016.

[6] Jigisha M Patel and Nikunj C Gamit. A review on feature extraction techniques in content based image retrieval. In 2016 *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 2259–2263. IEEE, 2016.

[7] Salamh, Ahmed B. Salem, and Halil Ibrahim Akyüz. "A Novel Feature Extraction Descriptor for Face Recognition." *Engineering, Technology & Applied Science Research* 12, no. 1: 8033-8038, 2022.

[8] Alasadi, Abdulmalik A., T. H. Aldhayni, Ratnadeep R. Deshmukh, Ahmed H. Alahmadi, and Ali Saleh Alshebami. "Efficient feature extraction algorithms to develop an arabic speech recognition system." *Engineering, Technology & Applied Science Research* 10, no. 2: 5547-5553, 2020.

[9] Alsuwaiket, Mohammed A. "Feature Extraction of EEG Signals for Seizure Detection Using Machine Learning Algorthims." *Engineering, Technology & Applied Science Research* 12, no. 5: 9247-9251, 2022.

[10] Manjunath Jogin, MS Madhulika, GD Divya, RK Meghana, S Apoorva, et al. Feature extraction using convolution neural networks (cnn) and deep learning. In 2018 3rd *IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*, pages 2319– 2323. IEEE, 2018.

[11] Dario Garcia-Gasulla, Ferran Par´es, Armand Vilalta, Jonatan Moreno, Eduard Ayguad´e, Jesu´s Labarta, Ulises Cort´es, and Toyotaro Suzumura. On the behavior of convolutional nets for feature extraction. *Journal of Artificial Intelligence Research*, 61:563–592, 2018.

[12] Wenyi Lin, Kyle Hasenstab, Guilherme Moura Cunha, and Armin Schwartzman. Comparison of handcrafted features and convolutional neural networks for liver mr image adequacy assessment. *Scientific Reports*, 10(1):1–11, 2020.

[13] Yingqi Zhang, Yong Feng, Jiaxing Shang, Mingliang Zhou, and Baohua Qiang. Attention-aware joint location constraint hashing for multi-label image retrieval. *IEEE Access*, 8:3294–3307, 2019.

[14] Huimin Lu, Ming Zhang, Xing Xu, Yujie Li, and Heng Tao Shen. Deep fuzzy hashing network for efficient image retrieval. *IEEE transactions on fuzzy systems*, 29(1):166–176, 2020.

[15] Maruthamuthu Ramasamy and John Sanjeev Kumar Athisayam. 2d matrix based indexing with color spectral histogram for efficient image retrieval. *Journal of Systems Engineering and Electronics*, 27(5):1122–1134, 2016.

[16] Ahmad Raza, Hassan Dawood, Hussain Dawood, Sidra Shabbir, Rubab Mehboob, and Ameen Banjar. Correlated primary visual texton histogram features for content base image retrieval. *IEEE Access*, 6:46595–46616, 2018.

[17] LK Pavithra and T Sree Sharmila. An improved seed point selection-based unsupervised color clustering for content-based image retrieval application. *The Computer Journal*, 63(3):337–350, 2020.

[18] Guangyi Xie, Baolong Guo, Zhe Huang, Yan Zheng, and Yunyi Yan. Combination of dominant color descriptor and hu moments in consistent zone for content based image retrieval. *IEEE Access*, 8:146284–146299, 2020.

[19] Xiaoli Zhang. Content-based e-commerce image classification research. *IEEE Access*, 8:160213– 160220, 2020.

[20] Ayesha Khan, Ali Javed, Muhammad Tariq Mahmood, Muhammad Hamza Arif Khan, and Ik Hyun Lee. Directional magnitude local hexadecimal patterns: A novel texture feature descriptor for content-based image retrieval. *IEEE Access*, 9:135608–135629, 2021.

[21] Yongwei Miao, Gaoyi Li, Chen Bao, Jiajing Zhang, and Jinrong Wang. Clothingnet: Cross- domain clothing retrieval with feature fusion and quadruplet loss. *IEEE Access*, 8:142669–142679, 2020.

[22] Jianbo Ouyang, Wengang Zhou, Min Wang, Qi Tian, and Houqiang Li. Collaborative image relevance learning for visual re-ranking. *IEEE Transactions on Multimedia*, 23:3646–3656, 2020.

[23] Chen, Rongyu, Lili Pan, Yan Zhou, and Qianhui Lei. "Image retrieval based on deep feature extraction and reduction with improved CNN and PCA." Journal of Information Hiding and Privacy Protection 2, no. 2 (2020): 67.

[24] Zhang, Xuanmeng, Minyue Jiang, Zhedong Zheng, Xiao Tan, Errui Ding, and Yi Yang. "Understanding image retrieval re-ranking: a graph neural network perspective." arXiv preprint arXiv:2012.07620 (2020).

[25] Kan, Shichao, Yigang Cen, Yang Li, Mladenovic Vladimir, and Zhihai He. "Local Semantic Correlation Modeling Over Graph Neural Networks for Deep Feature Embedding and Image Retrieval." IEEE Transactions on Image Processing 31 (2022): 2988-3003.

[26] Wang, Shoujin, Liang Hu, Yan Wang, Xiangnan He, Quan Z. Sheng, Mehmet A. Orgun, Longbing Cao, Francesco Ricci, and Philip S. Yu. "Graph learning based recommender systems: A review." arXiv preprint arXiv:2105.06339 (2021).

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[28] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[29] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818-2826, 2016.

[30] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251-1258. 2017.

Emotion Recognition on Multimodal with Deep Learning and Ensemble

David Adi Dharma, Amalia Zahra
Computer Science Department-BINUS Graduate Program-Master of Computer Science
Bina Nusantara University, Jakarta, Indonesia 11480

***Abstract*—Emotion Recognition on multimodal dataset is a difficult task, which is one of the most important tasks in topics like Human Computer Interaction (HCI). This paper presents a multimodal approach for emotion recognition on dataset MELD. The dataset contains three modalities, audio, text, and facial features. In this research, only audio and text features will be experimented on. For audio data, the raw audio is converted into MFCC as an input to a bidirectional LSTM, which will be built to perform emotion classification. On the other hand, BERT will be used to tokenize the text data as an input to the text model. To classify the emotion in text data, a Bidirectional LSTM will be built. And finally, the voting ensemble method will be implemented to combine the result from two modalities. The model will be evaluated using F1-score and confusion matrix. The unimodal audio model achieved 41.69% of F1-score, while the unimodal text model achieved 47.29% of F1-score, and the voting ensemble model achieved 47.47% of F1-score. To conclude this research, this paper also discussed future works, which involved how to build and improve deep learning models and combine them with ensemble model for better performance in emotion recognition tasks in multimodal dataset.**

Keywords—Emotion recognition; deep learning; ensemble method; transformer; natural language processing

I. Introduction

Humans, along with technological developments pour their emotions or feelings either through some media such as text, photos, audio, or video recordings. Human emotions are complex, which make it difficult to be studied or predicted, and it takes a high level of intelligence to be able to recognize the emotions expressed by people in the current media [1]. Due to the complexity of human emotions, the variety of human's feelings, and the media where they convey their emotions, the AI model's learning has evolved to multimodal datasets, where existing media, audio, video, text, biological information, can improve the model's ability to classify emotions more accurately [2]. This emotion recognition task is involved in some study subjects, which are Natural Language Processing (NLP) and Machine Learning (ML).

The algorithms used to classify emotions are also being actively researched and developed. The same algorithm used for classification can have different results, depending on the dataset used. The datasets used for classification are media such as text, or images, or EEG (Electroencephalogram) signals, as well as sound. Examples of Machine Learning (ML) method that was used for classifying emotions are SVM (Support Vector Machine), KNN (K-Nearest Neighbor), and Bayesian Network in the research [3]. Then the emotion classification method developed to neural networks-based model which are Deep Learning (DL) such as Recurrent Neural Network (RNN) in [4] research, Deep Neural Network (DNN) in [5], DialogueRNN which is RNN-based model in the research by [6], LSTM (Long Short-Term Memory) and CNN (Convolutional Neural Network) conducted in the [7] experiment. Lastly, a Bidirectional LSTM-CNN model was used to learn context and classify emotions in the [8] and [9] works.

Transformer which is proposed in 2017 [10] is the neural network architecture. This model outperformed any RNN and LSTM model that were popular in NLP. Thus, Transformer develops into BERT in 2019 [11]. BERT stands for Bidirectional Encoder Representation of Transformer, and until today BERT has been widely used to extract contextual information from texts. BERT also used in emotion recognition task on the text dataset, in [11] research published in 2022.

An ensemble learning [12] is a technique that combines base predictors model and improves it to become more outstanding predictor. There are few kinds of ensemble learnings which are called bagging, boosting, and stacking [13]. In the [14], there is ensemble learning called voting, which choose the highest probability value as the final classification result. In this experiment, the voting ensemble learning will be performed.

There are some datasets that are widely used in multimodal emotion classification research, such as IEMOCAP (The Interactive Emotional Dyadic Motion Capture) and MELD (Multimodal EmotionLines Dataset). MELD was made and published in 2018 by [15]. These works have used IEMOCAP dan MELD for building and developing models that are used to classify emotion in multimodal. This paper built a model using MELD dataset, which contains 7 labels of emotions, such as anger, disgust, sadness, joy, neutral, surprise and fear. All data in the MELD dataset are in English. There are three modalities that are provided the dataset, video which are audio and facial data, and textual data.

The purposes of the experiment are to extract emotion features in multimodal dataset and build a model that could recognize the emotion that is learned from the features. This experiment also intends to evaluate the model built and compare the evaluation result with the existing models.

There are two tasks that will be conducted in this research, Speech Emotion Recognition (SER) and Text Emotion Recognition (TER). Based on the state-of-the-art mentioned above, BERT will be built to tokenize the textual dataset, then a Bidirectional LSTM model is built to classify the emotion for the text dataset. On the other hand, a Bidirectional LSTM model will be used for SER task. Finally, a voting ensemble learning model will be applied as the ensemble learning. This paper will utilize the advantages of models that were mentioned above for improving emotion recognition model's performance.

The remainder of the paper is structured as follows: Section II will review related research for emotion recognition for both unimodal and multimodal dataset, Section III will provide proposed method in this research, Section IV on experiment scenario, discussion on Section V, and finally, conclusion and future works are provided on Section VI.

II. RELATED WORKS

Research related to emotion classification has been developed using various combinations of algorithms with various datasets. Datasets can be in the form of text, sound, images, to multimodal such as a combination of text with sound and video recordings (image and sound). Research conducted in [4] uses the IEMOCAP dataset as the multimodal dataset, where the researcher takes audio and text data. In the audio dataset, the features are extracted using the tools openSMILE for audio and NLTK for text. In the classification of emotions, the results of the extraction of these features are processed by an RNN model. Another study by [16] used the OMG (One-Minute Gradual-Emotion Recognition) dataset and divided the dataset into 2 classes, namely arousal and valence. To extract video data features, researchers used OpenFace and VGG Face, to extract audio data features using OpenSMILE, and for text using Lexicon.

Then another study using the AFEW (Acted Facial Expression in the Wild) dataset conducted in [2] and compared several feature extraction methods, such as TF-IDF (Term Frequency-Inverse Document Frequency) and WV (Word Vectors) for text data, C3D and VGG for visual datasets, then MFCC, SoundNet, and VGG for audio data. Then each data type, namely audio, text, and visual is classified using 3 methods, namely RF (Random Forest), SVM (Support Vector Machine), and LSTM (Long Short-Term Memory).

Subsequent research conducted in [6] proposed a model called DialogueRNN, which was used to classify emotions from the IEMOCAP and AVEC datasets. In this study, audio data features were extracted using OpenSMILE tools, 3D-CNN to extract visual features, and CNN to extract text features. Further research by [17] used the IEMOCAP dataset by taking audio and text data. Text data is extracted using Word2Vec, while audio dataset is extracted to new low-level features and then processed into high-level features or contextual features using CNN-LSTM. The method used for emotion classification is Deep Neural Network (DNN) which consists of four layers, namely input layer, hidden layer 1, hidden layer 2, and softmax layer.

Research by [18] uses the IEMOCAP dataset and retrieves audio and text data. The method used to classify is divided into two, namely, several Machine Learning (ML) models and several Deep Learning (DL) models. The ML methods used are Random Forest (RF), Gradient Boosting (XGB), Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), and Logistic Regression (LR). Meanwhile, the DL method used is Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM).

Another study conducted by [5] used the IEMOCAP dataset, taking audio and text data from IEMOCAP. Text data feature extraction using GloVe, while audio data is handcrafted. Researchers combined CNN (Convolutional Neural Network) with LSTM (Long Short-Term Memory) to classify emotions from audio data, and Bi-LSTM (Bidirectional Long Short-Term Memory) to classify text data. Then in the classification of multimodal data, the method used is DNN (Deep Neural Network) by combining high-level features from 2 data, namely audio and text, then using a softmax layer to perform the classification.

Furthermore, research conducted by [19] used the MELD dataset where the data taken were audio and text data. Audio data is converted to MFCC first, while text data is processed using BERT. Then the audio and text data are entered into the GRU and then forwarded to the proposed model, namely MMFA (Multi-Level Multi-Head Fusion Attention).

The next study conducted in [17] used the MELD dataset and took the audio and text data as multimodal data, where the proposed model is GNN (Graph Neural Network). Research by [20] uses the MELD dataset and uses all data modalities, namely audio, text, and visual. The proposed model is HFGCN (Hierarchical Fusion Graph Convolutional Network). HFGCN provides output in the form of utterance-level features which are then entered into a fully connected layer for emotional classification.

In the [21] research in 2018, Bi-LSTM and LSTM are combined with Attention Layer to classify emotions, while the datasets are crawled from the internet. After bi-LSTM and LSTM layers are being utilized, a soft voting model is deployed as the ensemble strategy to get the final emotion. In [11] used ResNet for emotion recognition in audio modality and used BERT based model to emotion recognition in text modality. Then final score is being calculated using weighted score formula mention in [11].

In the [15] along with MELD dataset, a Bidirectional LSTM model was built to classify the emotions. This bidirectional LSTM will be included as a baseline model for the unimodal emotion model in this experiment. Also, in the [22], there are some bidirectional architectures that were experimented on. The best architecture will be included as the baseline model in this experiment. The unimodal model that will be built is the modified version from these two baseline models. Then, for the multimodal classification, voting will be performed from the predictions which are generated by these baseline models as the benchmarks.

III. Research Framework

The topic of this research is to recognize emotion through classification from the model that will be built with multimodal dataset. Many Machine Learning and Deep Learning models have been researched to do emotion recognition tasks in the multimodal dataset. Common multimodal datasets such as IEMOCAP, AFEW, OMG, and MELD are being widely used to train the emotion recognition model. Most research uses bimodal such as audio and text dataset, there are also few research that use audio, facial, and text features. There are many combinations of methods in multimodal dataset, but to find the most effective one for emotion classification, it will need a lot of experiments.

This research proposed deep learning models to recognize emotions in the two dataset modalities in the MELD dataset. This research will be done in supervised learning as MELD has already been labeled to train deep learning models. For audio features, bidirectional LSTM (Long-Short Term Memory) will be used to classify emotions. Then, as for text modality, BERT will be deployed to extract contextual features and classify the emotions. Lastly, voting ensemble learning is also being experimented in research.

As shown in Fig. 1, this research starts from choosing the topic of the research, and then studies all the literature related to the topic. After that the method and deep learning model is chosen. The next part will be preparing the dataset, preprocessing the data, and extracting the features from the dataset. Then, in the next part of the research, the model will be tuned, then train the model for each modality, which are audio and text. After that, voting ensemble learning is applied. At last, after all models are trained, the models will be evaluated using evaluation metrics such as F1-score and confusion matrix, then analyze the experiment result, and conclude the experiment result in the paper.

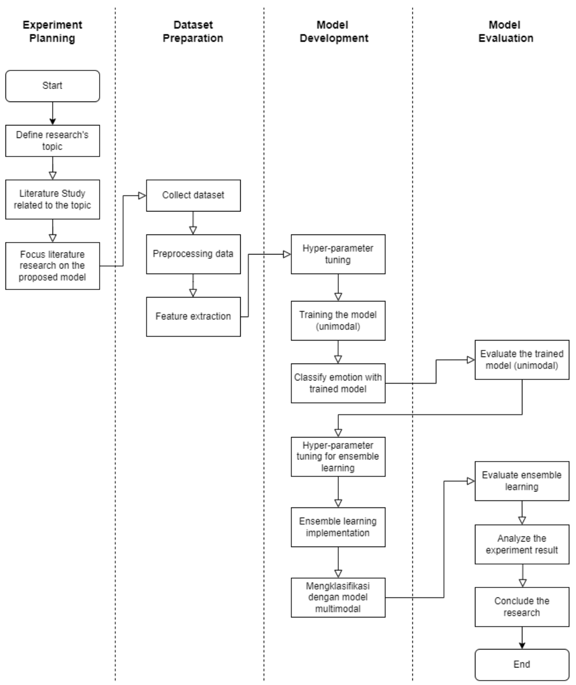


Fig. 1. Experiment scenario.

A. Preparing Dataset

In this research, MELD dataset will be used to train and build the model. MELD [15] contains 13,000 utterances from 1,433 dialogs in the TV series called “Friends”. The dataset is divided into three parts, train, test, and dev. Every utterance from 1,433 dialogs already has emotion label and sentiment label. The dataset contains three modalities, which are audio, facial, and text modality. There are seven emotion labels annotated in the dataset, Joy, Sadness, Fear, Anger, Surprise, Disgust, and lastly Neutral. There are also three annotations for sentiment analysis, such as neutral, positive, and negative. But in this research, only emotion labels will be used as the label to train the model.

The distribution of the emotion in the dataset is shown the Fig. 2. The train part contains the most data, while test and dev parts contain less data respectively. Fig. 3 below will show how the emotion labels are divided in percentage using pie chart. Neutral label is dominant with 47% label from all emotion label that exist in the MELD dataset. Then followed by Joy with 17%, Anger and Surprise with 12% each, Sadness is 7%, Disgust is 3% and Fear with only 2% from the dataset.

As the dataset is greatly imbalance, which will affect the model’s performance, the emotion labels in this experiment will be converted to only five emotion labels, which are neutral, anger, joy, sadness, and surprise. Fear and disgust are excluded because of the lack of data. As for the train data, the labels distribution will be divided into 683 data for neutral, anger, joy, and surprise, and sadness label. The total training data for the model will be 3415. The data distribution bar chart is represented in Fig. 4.

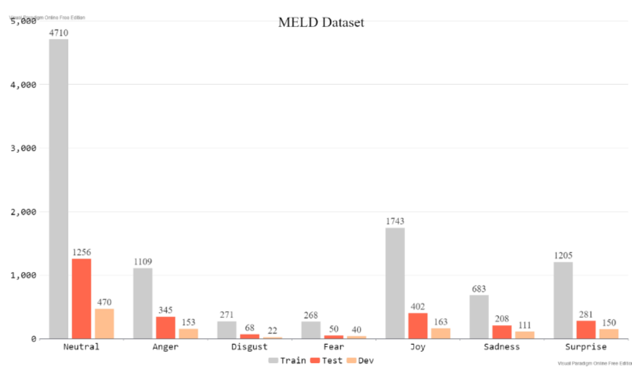


Fig. 2. MELD label distribution.

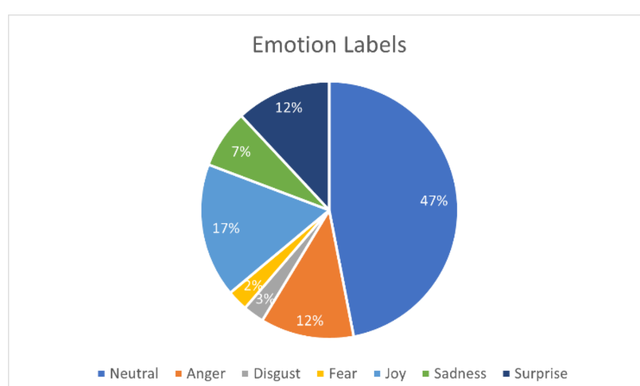


Fig. 3. Labels percentage.

Fig. 4. Training data distribution.

B. Feature Extraction

There are two feature extraction methods in this research. For text data, raw texts are tokenized with Distil-BERT model. As for the audio modality, the audio file will be extracted into MFCC [23]. Feature extractor methods are being used to convert the raw data so that the data can be learned by the model. After features are extracted from the dataset, there will be unimodal emotion recognition to classify emotion from given features. As for text data, an embedding layer will be applied before Bidirectional LSTM to understand the context of the text.

IV. Theory and Methods

A. Emotion Recognition

Emotion recognition is one of the topics that has been widely researched along with the development of AI (Artificial Intelligence), the main reason being that the application of emotion classification is carried out on many difficult AI tasks, such as creating dialogues, understanding user behavior, and multimodal interactions [15]. The classification of emotions in a conversation can be used to make a suitable response by analyzing the emotions of the user.

Human emotions become the most important aspect to perform and develop natural human-machine interaction (HMI). While AI is one of the most developed topics in recent years, it'll need data from several modalities. Also, there are many methods to be explored to make AI learn effectively.

B. MFCC

The mel-frequency cepstral coefficient (MFCC) is an interpretation of an audio file in the form of a sequence of numbers [23]. The sequence of numbers is obtained from dividing raw audio files into frames, then performed some steps that are shown in Fig. 5. Lastly, the MFCC array is transposed using arithmetic formula. These MFCC numbers represented human voices amplitude.

C. LSTM

LSTM (Long Short-Term Memory) is one of the Deep Learning models derived from the RNN (Recurrent Neural Network) model which has special properties [24]. LSTM is the proposed solution model after finding the shortcomings of the traditional RNN model, where the RNN has a vanishing gradient and an exploding gradient, causing the RNN to fail to capture long-term dependencies, and as a result, the prediction accuracy of the RNN model decreases.

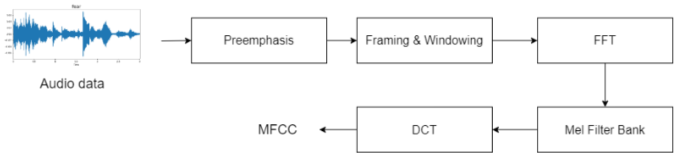


Fig. 5. MFCC process.

The LSTM network architecture treats the hidden layers contained in the neural network as memory units, where the memory units collected in one recurrent hidden layer are called memory blocks. These memory blocks can be used to memorize temporal state from the neural network so that LSTM can remember the correlation between features in a sequence of time.

D. BERT

BERT (Bidirectional Encoder Representations from Transformers) is a new Transformer model for language representation designed to understand the contextual relationship between words from unlabeled text [25]. BERT also penetrated the start-of-the-art because it can see the context of a sentence from two directions, from left to right and from right to left. The BERT model architecture consists of a multi-layer bidirectional Transformers encoder.

BERT can be used to process not only embedding in text data, but also perform classification. The total Transformers blocks contained in the BERT baseline model are 12 blocks, with 768 hidden units, and 12 self-attention heads. The corpus used for pre-training in BERT is Books Corpus (800 million words) and English Wikipedia (2,500 million words).

V. Proposed Method

The proposed method for this research is shown in Fig. 6, Fig. 7, and Fig. 8. There are two models that are used to classify emotion from each modality. For the audio data, Bidirectional LSTM is used to recognize emotion in audio dataset, where the model learns audio features that were already extracted into MFCC.

This SER model's architecture is based on Bidirectional LSTM that was experimented on [15] and [22]. The baseline models have two layers of Bidirectional LSTM. These twp layers can learn the MFCC features very well, better than 1 layer or 3 or more layers, based on the first experiment by the authors. Bidirectional LSTM that are too complex couldn't perform well with the MFCC features, so the Bidirectional LSTM adjusted into two layers. The LSTM unit in the layers are 256 and 128 units, respectively. Then, Dense Layer and Dropout is applied to reduce the complexity of the output from Bidirectional LSTM layers. The SER model architecture is shown in Fig. 6.

On the other hand, Distil-BERT pretrained model is utilized to tokenize the raw words from the text dataset. Then, the tokenized words are fed into an embedding layer. This embedding layer is trained with training dataset to understand the context from the text data. The output from the embedding layer is inputted to 3-layered Bidirectional LSTM and classifies the emotion.

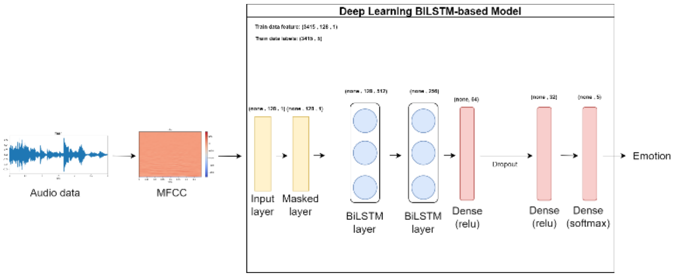


Fig. 6. Proposed SER deep learning model.

This TER model's architecture is based on Bidirectional LSTM that was built in [15] and [22]. There are two layered Bidirectional LSTM that were applied on both experiments. Also, in the [22] there are six layers of BiLSTM and two Dense Layers. But the performance of the 6-layered BiLSTM is worse than the 2-layered BiLSTM. So, the 2-layered BiLSTM can be the baseline model of this experiment. Then, 2-layered BiLSTM is adjusted into 3-layered BiLSTM which can perform better than the 2-layered. Then, a flatten layer and three layers of Dense layers are added to reduce the complexity of the output from the Bidirectional LSTM layers. The TER model architecture is shown in the following Fig. 7.

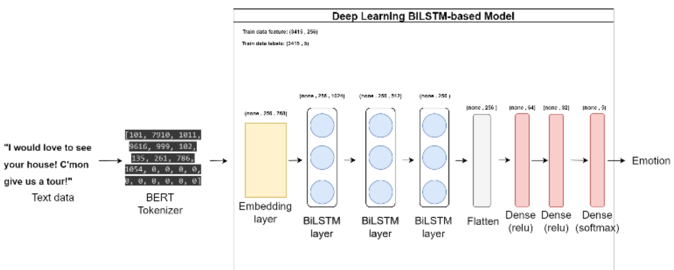


Fig. 7. Proposed TER deep learning model.

Lastly, a voting ensemble method is deployed to classify the final emotion. The voting ensemble model is shown in the following Fig. 8. First, the unimodal models, SER model and TER model are trained. After that, each model, both SER and TER model predict using the features of the test data. The output would be on one-hot encoded format, because of the softmax activation function in the last layer in both models. This softmax function produces numbers that could be interpreted as probability. For example, the prediction generated from the first model is ([0.654 0.346]). The second model generates ([0.781 0.219]). Because the probability from the second model which is 0.781 is higher than 0.654, the voting will choose prediction from the second model.

So, every prediction that is generated from the unimodal models will be compared, which model produces higher probability. If the SER model generates output with higher probability, then the final emotion label will be selected from the SER model's prediction. On the contrary, if TER model generates output with higher probability, then the emotion label will be selected from TER model.

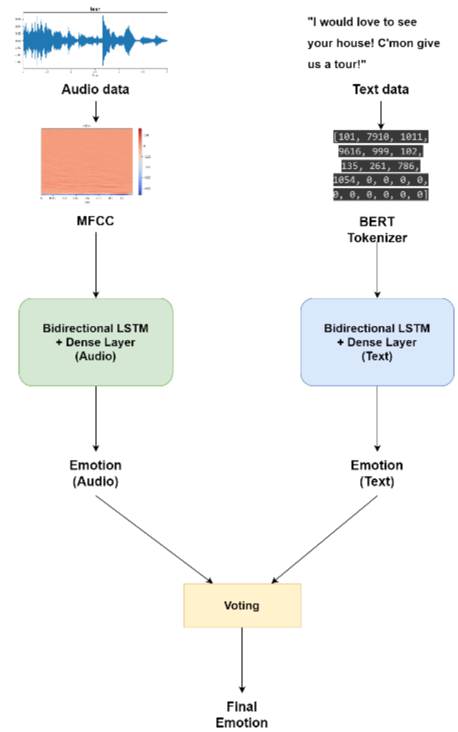


Fig. 8. Proposed voting ensemble learning.

The experiment is implemented using Keras library with the help of librosa library as the audio feature extractor and BERT tokenizer from transformers library built by the Hugging Face. The librosa library is utilized to extract raw audio files into MFCC while the BERT model was used as the word tokenizer in the experiment.

The details of the bidirectional LSTM and deep learning layers which is the proposed method for classifying emotions from audio data, text data, and the voting result from the two modalities are written in the following tables. The details of the SER model are shown in Table I.

On the other side, the construction bidirectional LSTM and deep learning layers for the textual emotion recognition (TER) are written in the following Table II.

TABLE I. BIDIRECTIONAL LSTM LAYERS FOR SER

Bidirectional LSTM				
Layers	**Units**	**Dropout**	**Output Size**	**Activation**
Input layer	1	-	(None, 128, 1)	-
Masking	1	-	(None, 128, 1)	-
Bidirectional LSTM	256	0.4	(None, 128, 512)	tanh
Bidirectional LSTM	128	0.3	(None, 256)	tanh
Dense	64	-	(None, 64)	ReLU
Dropout	0	0.2	(None, 64)	-
Dense	32	-	(None, 32)	ReLU
Dense	5	-	(None, 5)	Softmax

TABLE II. BIDIRECTIONAL LSTM LAYERS FOR TER

Bidirectional LSTM				
Layers	**Units**	**Dropout**	**Output Size**	**Activation**
Embedding		-	(None, 256, 768)	-
Bidirectional LSTM	512	0.1	(None, 256, 1024)	tanh
Bidirectional LSTM	256	0.1	(None, 256, 512)	tanh
Bidirectional LSTM	128	-	(None, 256)	tanh
Flatten		-	(None, 256)	-
Dense	64	-	(None, 64)	ReLU
Dense	32	-	(None, 32)	ReLU
Dense	5	-	(None, 5)	Softmax

VI. RESULT AND DISCUSSION

Results from the experiment provided in Table III presented model's performance on the single modality and multimodality. Since the MELD has already been divided into train and test datasets, the model is being built using train data. The validation dataset takes 15% randomly from the train dataset to show the learning progress of each epoch.

The Bidirectional LSTM for the SER task trained with 30 epochs with batch size 32. While the TER task trained with only 20 epochs using the same batch size, which is 32. Both proposed SER and TER model used the l2 kernel regularizer where the kernel regularizer is set to $1x10^{-4}$. This regularizer is added to reduce or prevent the models that were built in the experiment from overfitting. The optimizer used to train both SER and TER model is Adam. For the TER model, the learning rate was set to $1x10^{-5}$. On the SER model, the learning rate was set to $1x10^{-4}$ because the number of epochs to train SER model is higher compared to TER model. These hyperparameters are manually tuned until receive the best result.

The experiment results are compared to some research papers shown in the following Table III. The benchmark models are trained using the same amount data as the proposed model with the same hyperparameters. The proposed method could surpass the baseline models with small gaps in decimals. If the f1-score is converted from decimal into percent, the proposed model does not surpass more than 1%. The gap between voting performed on model built by [15] and the proposed model approximately 1.5%.

From the evaluation result that is shown in Table III, proposed SER model, TER model, and the vote result based on SER and TER models could surpass the F1-score benchmarks. The emotion labels are strongly related with the context spoken by the speaker. This shows that the embedding layer that has been trained could understand contextual features better compared to MFCC, so that the TER model could perform better in recognizing the emotions given in the dataset.

As for the baseline TER model that was built in [15], an embedding layer is added, so the baseline model could learn the contextual features from tokenized word like model that was built in [22]. Both embedding layers are trained with the same parameter from the proposed model, so that the architecture of Bidirectional LSTM can be evaluated and compared fairly.

The details of the evaluation metrics such as precision, recall, and f1-score for each emotion label are provided in the following Tables IV, V, and VI. Table IV contains the details from SER model, while Table V contains the details from TER model and voting result, respectively.

The metrics that are shown in Table IV above interpret the details of how the proposed SER model predicted each label from the test dataset. As the test data mainly contain neutral emotion, the precision, recall, and f1-score for neutral emotion are good. But, for other emotions such as joy and surprised, the model performed poorly on predicting both emotions.

TABLE III. EXPERIMENT RESULT ON MELD DATASET BASED ON F1-SCORE

Modality	Approach	F1-score
Audio	BiLSTM [15]	0.408
	BiLSTM [22]	0.402
	BiLSTM (Proposed)	0.417
Text	BiLSTM [15]	0.443
	BiLSTM [22]	0.464
	BiLSTM (Proposed)	0.473
Multimodal	Voting from 2 modalities [15]	0.459
	Voting from 2 modalities [22]	0.468
	Voting from 2 modalities (proposed)	0.475

TABLE IV. EVALUATION METRICS SCORE FOR PROPOSED SER MODEL

Emotion	Precision	Recall	F1-score	Support
Anger	0.297	0.183	0.226	345
Joy	0.196	0.067	0.1	402
Neutral	0.521	0.705	0.599	1256
Sadness	0.145	0.255	0.185	208
Surprised	0.145	0.039	0.062	280

TABLE V. EVALUATION METRICS SCORE FOR PROPOSED TER MODEL

Emotion	Precision	Recall	F1-score	Support
Anger	0.310	0.446	0.366	345
Joy	0.344	0.236	0.280	402
Neutral	0.746	0.584	0.655	1256
Sadness	0.132	0.284	0.179	208
Surprised	0.476	0.489	0.482	280

The metrics that are shown in Table V interpret the details of how the proposed TER model predicted each label from the test dataset. Just like SER model, the TER model could predict neutral emotion well enough. The difference between the SER and the TER model is how well the TER model predicted surprised emotion, compared to the SER model. Overall, the TER model performed better except in predicting sadness emotion label, but only by inches.

TABLE VI. EVALUATION METRICS SCORE FOR PROPOSED VOTING ENSEMBLE

Emotion	Precision	Recall	F1-score	Support
Anger	0.313	0.452	0.370	345
Joy	0.363	0.246	0.293	402
Neutral	0.745	0.582	0.654	1256
Sadness	0.132	0.288	0.181	208
Surprised	0.477	0.488	0.482	280

The metrics that are shown in Table VI above interpret the details of the performance from two modalities TER and SER combined. Overall, the F1-score in the models is not as good as F1-score on the other dataset, for example research by [14] on WASSA2018 dataset. The main reason of the overall scores only in range 40% of F1-score is because the dataset is greatly imbalanced. To cope the imbalanced label from the train dataset, undersampling was performed so that the training data is much fewer compared to all training data, which is 3415 compared to 9989. So, the models only trained with one third of the full dataset.

Secondly, some of the data also does not contain keywords that are related to the emotion, for example, in the dataset, there is 1 word that is used in different emotions, which is “Hey.” This word is used in joy, anger, and neutral expression in the dataset. It creates confusion so the model can’t recognize the emotion in each sentence or utterance.

VII. CONCLUSION

This study experimented with a deep learning architecture which is bidirectional LSTM for doing unimodal emotion classification. This study also includes voting ensemble learning to combine two different modalities of the input information.

Experimental results on the MELD dataset demonstrated the good result of the proposed method. The performance of the proposed method could improve the performance from the previous state-of-the-art strategies both [15] and [22] around 1% performance improvement in F1-score on both SER and TER model. However, both unimodal models and the multimodal model performed poor in predicting some emotions such as sadness and joy.

For future studies, other input modalities such as different physiological measurements, or sentiment, or maybe context of the dialogue should also be added and included to the dataset, so that the input for the deep learning model become more complex and richer. Also, the facial features from the video dataset could be utilized as a feature to predict the human’s emotions.

The imbalanced dataset also could be tackled by upsampling the dataset, especially on some emotion labels that only have a little data, which leaves some spaces to work in the future. Therefore, more multimodal datasets should be evaluated in future work.

ACKNOWLEDGMENT

The present paper was supported by the University of Bina Nusantara as thesis research to complete the graduate program. The writers are genuinely thankful and honored to those who help with the work on this paper and hopefully can produce many other great research or projects in the future.

REFERENCES

[1] S. H. Park, B. C. Bae, and Y. G. Cheong, “Emotion recognition from text stories using an emotion embedding model,” in Proceedings - 2020 IEEE International Conference on Big Data and Smart Computing, BigComp 2020, Feb. 2020, pp. 579–583. doi: 10.1109/BigComp48618.2020.00014.

[2] Z. Lian, Y. Li, J. Tao, and J. Huang, “Investigation of Multimodal Features, Classifiers and Fusion Methods for Emotion Recognition,” 2018. [Online]. Available: https://arxiv.org/abs/1809.06225.

[3] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. J. Jerči´c, “Evaluating Classifiers for Emotion Recognition Using EEG,” 2013.

[4] S. Yoon, S. Byun, and K. Jung, “Multimodal Speech Emotion Recognition using Audio and Text,” SLT, pp. 112–118, 2018.

[5] L. Cai, Y. Hu, J. Dong, and S. Zhou, “Audio-Textual Emotion Recognition Based on Improved Neural Networks,” Math Probl Eng, vol. 2019, 2019, doi: 10.1155/2019/2593036.

[6] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, and E. Cambria, “DialogueRNN: An Attentive RNN for Emotion Detection in Conversations,” Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.00405.

[7] M.-H. Su, C.-H. Wu, K.-Y. Huang, and Q.-B. Hong, LSTM-based Text Emotion Recognition Using Semantic and Emotional Word Vectors, ACII Asia. 2018.

[8] J. L. Wu, Y. He, L. C. Yu, and K. Robert Lai, “Identifying Emotion Labels from Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model,” IEEE Access, vol. 8, pp. 66638–66646, 2020, doi: 10.1109/ACCESS.2020.2985228.

[9] C. Li, Z. Bao, L. Li, and Z. Zhao, “Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition,” Inf Process Manag, vol. 57, no. 3, May 2020, doi: 10.1016/j.ipm.2019.102185.

[10] A. Vaswani et al., “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762.

[11] S. Padi, S. O. Sadjadi, D. Manocha, and R. D. Sriram, “Multimodal Emotion Recognition using Transfer Learning from Speaker Recognition and BERT-based models,” Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.08974.

[12] K. Pham, D. Kim, S. Park, and H. Choi, “Ensemble learning-based classification models for slope stability analysis,” Catena (Amst), vol. 196, Jan. 2021, doi: 10.1016/j.catena.2020.104886.

[13] A. Verma and S. Shikha Mehta, A Comparative Study of Ensemble Learning Methods for Classification in Bioinformatics. 2017.

[14] Q. Zhou, Z. Zhang, and H. Wu, “BiLSTM-Attention and LSTM-Attention via Soft Voting in Emotion Classification,” in WASSA 2018 - 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Proceedings of the Workshop, 2018, pp. 149–155. doi: 10.18653/v1/P17.

[15] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, “MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations,” Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.02508.

[16] D. Deng, Y. Zhou, J. Pi, and B. E. Shi, “Multimodal Utterance-level Affect Analysis using Visual, Audio and Text Features,” May 2018, [Online]. Available: http://arxiv.org/abs/1805.00625.

[17] Y. Gu, S. Chen, and I. Marsic, "Deep Multimodal Learning for Emotion Recognition in Spoken Language," 2018.

[18] G. Sahu, "Multimodal Speech Emotion Recognition and Ambiguity Resolution," Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.06022.

[19] N. H. Ho, H. J. Yang, S. H. Kim, and G. Lee, "Multimodal Approach of Speech Emotion Recognition Using Multi-Level Multi-Head Fusion Attention-Based Recurrent Neural Network," IEEE Access, vol. 8, pp. 61672–61686, 2020, doi: 10.1109/ACCESS.2020.2984368.

[20] Z. Lian, J. Tao, B. Liu, J. Huang, Z. Yang, and R. Li, "Conversational emotion recognition using self-attention mechanisms and graph neural networks," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020, vol. 2020-October, pp. 2347–2351. doi: 10.21437/Interspeech.2020-1703.

[21] S. Tang, Z. Luo, G. Nan, Y. Yoshikawa, and I. Hiroshi, "Fusion with Hierarchical Graphs for Mulitmodal Emotion Recognition," Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.07149.

[22] B. Tris Atmaja, K. Shirai, and M. Akagi, "Speech Emotion Recognition Using Speech Feature and Word Embedding," 2019.

[23] M. G. de Pinto, M. Polignano, P. Lops, and G. Semeraro, Emotions Understanding Model from Spoken Language using Deep Neural Networks and Mel-Frequency Cepstral Coefficients. 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), 2020.

[24] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for Short-term traffic forecast," IET Intelligent Transport Systems, vol. 11, no. 2, pp. 68–75, Mar. 2017, doi: 10.1049/iet-its.2016.0208.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805.

Fault Diagnosis Technology of Railway Signal Equipment based on Improved FP-Growth Algorithm

Yueqin Yang
Urban Rail Transit College, Jilin Railway Technology College, Jilin, 132299, China

***Abstract*—The rapid development of computer information technology has made various fault diagnosis and detection technologies emerge in an endless stream. As one of the main transportation vehicles, the detection efficiency of fault diagnosis of railway signal equipment has important practical significance for maintaining the overall safe operation of railways. On the basis of the traditional FP-Growth algorithm, improve the TF-IDF algorithm to realize the weight discretization of text features, and realize the improvement of the FP-Growth algorithm by adjusting the adaptive confidence and support. The FP-Growth algorithm will be improved. FP-Growth algorithm is used for performance tests and applications. The results show that the minimum running time-saving of the proposed algorithm is 1500ms, and the average accuracy of P@N exceeds 85%, which is higher than that of the FP-Growth algorithm (81.4%) and VSM algorithm (82.1%). The PR curve of the improved algorithm is closer to the upper right, which effectively ensures the processing of correlated data, and the overall average precision performance under the influence of positive and negative signal-to-noise ratio values exceeds 95%. And the signal curve generated by the algorithm. The error range of the data under the four fault types of track circuit, turnout, signal, and connecting line floats between 1% and 5%. The improved FP-Growth algorithm can effectively analyze railway fault types and data. Perform analysis and data processing to minimize diagnostic errors.**

Keywords—FP-growth algorithm; gailway; signal equipment; fault diagnosis; relevance; frequent itemsets; vector space model

I. Introduction

The railway has always been in the backbone position in the modern comprehensive transportation system. It is an important infrastructure and the main means of transportation for the public in China, and also an important industrial sector to promote the national economy and social development. With the rapid development of China's railways in recent years, the safety requirements for railway transportation are becoming increasingly strict. With the in-depth development and application of information technology, the era of big data has come [1]. In the aspect of railway, the important premise to ensure traffic safety is to ensure the normal operation of railway signal equipment. Under the railway big data application platform, using data mining technology to process fault text data is one of the current research hotspots. The fault text data is extracted and mined, and the final results can guide maintenance personnel to process the fault equipment. As important technical equipment, railway signal equipment plays a key role in organizing and directing train operations, ensuring running safety and transportation efficiency. However, the types and causes of railway signal faults are complex, and traditional detection methods are difficult to detect railway signals. Once the signal equipment fails, it is easy to cause a major safety accident, which will seriously affect the maintenance efficiency and endanger the life safety of the staff. The rapid development of information technology provides a variety of diagnostic methods for the detection of railway signals. Different methods are used to diagnose railway faults, such as the application of neural networks, fuzzy logic fault diagnosis, expert diagnosis systems, etc. Data analysis of fault signals is one of the key contents of detection. Common text data association analysis algorithms include Apriori Algorithms and Frequent Pattern Growth (FP-Growth) algorithms. The correlation algorithm can provide a practical basis for fault diagnosis based on ensuring the correlation analysis of signal detection data. The FP Growth algorithm is often used in data mining processing because of its fewer traversal times and the ability to compress the data, but its application effect will be affected by the data type [2]. Therefore, it is studied to improve the traditional FP Growth algorithm and discretize the weight value of the text feature based on considering the type and characteristics of the railway fault causes, to ensure the objectivity and standardization of data processing, and adaptively set the confidence and support to reduce the non-objectivity of manually set parameters, to detect and analyze the railway signal fault problems with the improved FP Growth algorithm. The research ensures that the improved FP Growth algorithm can better analyze and process the fault information by discussing the status quo of relevant research methods and designing experimental methods, and verifies the effectiveness of the proposed algorithm with the help of simulation experiments and algorithm comparison, in order to provide new data diagnosis and treatment means for future railway fault maintenance and operation maintenance, so as to ensure the normal operation of railway construction.

II. Related Works

The construction and operation of the railway system and the information transmission of the command and dispatch of the trains are all centered on the signal equipment. With the increase of traffic volume, it is very important to strengthen the fault detection of the signal equipment to ensure the normal development of railway safety operation and maintenance. Different scholars have put forward different views on the detection and processing of railway signals. For example, Fang et al. used FP growth algorithm to carry out secondary equipment defect monitoring, and used frame algorithm to establish an exception model [3]. Jiang et al. used FP growth algorithm for association analysis and data mining, but found that this algorithm still has problems of low query efficiency and frequent item set traversal [4]. Yang and other scholars used the text mining intelligent fault classification model of signal equipment to realize the feature extraction and transformation of text information, and combined the characteristics of different classifiers. The ensemble study of learning classification was carried out under the experimental results. The experimental results showed that the model has high accuracy and recall [5]. Yuan and his team used the biased classification algorithm to divide the data flow of railway fault signals, and divided the data into line segments. The data block matrix is constructed to realize the integrated calculation of data consumption, and the experiments have verified that the algorithm can effectively reduce the memory consumption and greatly improve the data operation efficiency [6]. Han combines the hash table technology with the original FP growth algorithm, and conducts database experiments on it. It is found that the improved algorithm effectively shortens the running time, and its advantages gradually become prominent with the decrease of support [7]. Scholars such as Jiyang QI fully consider other external and objective factors when carrying out fault diagnosis, not only using a single indicator such as fault symptoms as the judgment standard, but also using the Weibull distribution model to calculate the failure rate based on, and carry out example analysis, the experimental results prove that the proposed algorithm can greatly improve the accuracy of fault diagnosis, and effectively reduce the interference of multi-information fusion on fault diagnosis [8]. Due to the diversity and complexity of information storage and application, word segmentation algorithms and topic models are used to extract features from information data, and then use support vector machines to diagnose equipment faults. It is found that the algorithm can provide algorithm guidance value for railway operation and maintenance [9]. Zhang scholars use association rules to achieve fault diagnosis of railway signals, and use TF-IDF algorithm to achieve fault feature extraction. The average diagnostic rate of this algorithm is 11.44% higher than that of Bayesian network [10]. Scholars such as Cheng integrate particle swarm optimization and minimum entropy to improve back - rolling the product method is used to detect the faults of rolling bearings, and the auxiliary coordinate transformation is used to optimize the algorithm. In the experimental simulation, it is found that the algorithm has good deconvolution performance, which effectively improves the diagnosis accuracy [11]. Zheng et al. designed constraints and screened data features to diagnose UAV flight faults. The experimental results show that the rule base of the improved algorithm has good scale growth and good application effect [12]. Ding scholars added the tail attribute to the traditional FP Growth algorithm to achieve tire data analysis. The experimental results show that the improved algorithm can effectively find the quality influencing factors and data feature mining [13]. Li et al. The fault feature vector acquisition and neural network improvement are carried out with the particle swarm optimization algorithm. The improved algorithm can show a short training time under the fault signal data of different states and its diagnosis accuracy rate is above 95%, which has a good application potential [14]. Jin and his team of scholars the power grid fault diagnosis based on layer data fusion is used to construct the model to process the data of different information sources, and the restoration path algorithm based on the fault rules is used to make the optimal judgment of the power supply restoration path. The results show that the method has strong practicability and effective the accuracy of power grid fault diagnosis has been improved [15]. At the same time, Cho and other scholars have achieved a safe and efficient train by reviewing and analyzing the control unit and current relay interface of the automatic train protection system, and detecting the fault type after analyzing the cause of the fault [16].

Fuzzy logic method, artificial neural network method and other methods will be interfered by artificial thinking to a certain extent when detecting railway signals and the application of FP-growth algorithm to railway fault detection has attracted the attention of many scholars. Fang et al. used FP growth algorithm to conduct secondary equipment defect monitoring, and used Hadoop framework and Mapreduce algorithm to establish an exception model. The results show that the algorithm can effectively find equipment abnormal defects and provide a data basis for diagnosis [17]. In order to reduce the disadvantage that feature extraction of fault data mostly depends on manual work, Jiangquan scholars use convolutional neural network to perform two-dimensional transformation and feature extraction of signals, and the algorithm proposed has good application timeliness under different cost and load conditions [18]. Scholars such as Syakur MA use the FP-Growth algorithm to understand the clustering results and classification of sales data and customer data, and use It can better explore the correlation between the subject and object of commodity consumption, and combine it with the frequent pattern tree to determine the item set. The test results show that the algorithm can effectively mine the data relationship, improve the average processing time of the data, and achieve better smooth operation efficiency. Guarantee [19]. Scholars such as Ariestya WW use frequent patterns and a priori algorithms to process selection, transformation and evaluation of data, and then conduct marketing strategy analysis to analyze the degree of consumer demand. The experimental results show that this method is used in data processing. It has good application performance and effectively provides a supporting basis for the formulation of consumer marketing strategies [20].

The railway fault signal detection can analyze its fault signal according to the classification idea. Some scholars use text data mining, biased classification algorithms and frequent patterns to identify and analyze the fault signal. Considering the complexity of the fault signal and the multi-source of the data, some scholars use wavelet decomposition, particle swarm optimization algorithm and convolution neural network to transform the features, and they have achieved certain results. Some research results of FP growth algorithm application, defect detection and data classification also show that it can be used in better classification and recognition. Based on this research idea, the text feature data with discrete weight is proposed as the input value of FP Growth algorithm, and the algorithm is improved through adaptive setting, effectively taking into account the characteristics and changes of railway signal fault data.

III. Fault Diagnosis Analysis of Railway Signal Equipment based on Improved FP-Growth Algorithm

A. Application of Improved FP-Growth Algorithm in Signal Fault Diagnosis

The continuous increase of railway lines and the continuous updating of signal equipment make the fault diagnosis records in the operation and maintenance of the current railway system mostly exist in the form of text, and the traditional railway text data are mostly analyzed from the aspects of safety management, and are not based on the type of fault diagnosis. It is difficult to improve the efficiency and accuracy of fault diagnosis by analyzing the characteristics and causes of faults. It is difficult to improve the efficiency and accuracy of fault diagnosis. The phenomenon of multiple meanings, the difference of signal fault text records and the existence of redundant information make the diagnosis of railway signal faults difficult. There are problems such as untimely and inaccurate problems, which further increase the difficulty of railway signal equipment diagnosis. Therefore, on this basis, the research is based on text data to carry out feature extraction and data mining to improve the accuracy of fault diagnosis, that is, to establish an association rule based the fault detection of the fault signal can realize the diagnosis and real-time analysis of the fault type and cause from different aspects. First, the fault signal text is segmented and feature extracted to ensure the vectorization of its data. The Vector Space Model (VSM) will the feature item is used as the smallest unit of the text, and the corresponding feature weight is given to it, and the mathematical expression of the model is shown in the formula [21].

$$V(d_i) = (t_1, w_{i1}; t_2, w_{i2}; t_3, w_{i3}; ....; t_n, w_{in}) \tag{1}$$

In formula (1), is t_n the first n feature item w_{in} of the text, and is the feature weight corresponding to the d_i feature item t_n. The weighted algorithm TF-IDF (Term Frequency - Inverse Document Frequency) algorithm is often used in data mining, and its core idea is to evaluate the importance of a word to a certain document in the corpus, its calculation formula is shown in the formula.

$$W_{t.d} = tf_{t.d} * idf_{t.d} \tag{2}$$

formula (2), is $W_{t.d}$ the weight value of the $tf_{t.d}$ feature value in the text, t and $idf_{t.d}$ is the word frequency and the inverse document frequency, respectively. The calculation formulas of the word frequency and the inverse document frequency are shown in formula (3) and formula (4).

$$tf_{t.d} = n_{t.d} / \sum k_{nt.d} \tag{3}$$

$$idf_{t.d} = \log\left(N / N_{t.d} + 1\right) \tag{4}$$

equations (3) and (4), $n_{t.d}$ is the number of times the feature value appears in the $\sum k_{nt.d}$ text, is the sum of the number of feature values in the text, is the total number N of texts, and is the number $N_{t.d}$ of texts with feature values in the text. The individual differences of text data and the appearance of synonyms will increase the difficulty and accuracy of fault diagnosis. The research adjusts the traditional Term Frequency-Inverse Document Frequency (TF-IDF) algorithm to obtain the formula (5)).

$$W_{t.d}' = tf_{t.d} * idf_{t.d} = tf_{t.d} * \log\left(N / N_{t'.d} + 1\right) \tag{5}$$

In the formula, is $W_{t.d}'$ the final weight of the feature item, $tf_{t.d}$ and t_n is $idf_{t.d}$ the word frequency and inverse document frequency of the feature item, $N_{t'.d} = N_{tn.d} + N_{tm.d}$, respectively, and is the sum of the number of texts t_n containing the feature item t_n, t_m. At the same time, the fault feature weight is discretized and divided into three categories according to its importance., and the value of 0-2 indicates the low, general and high importance. Fig. 1 is a schematic diagram of the level division under the weight discretization.

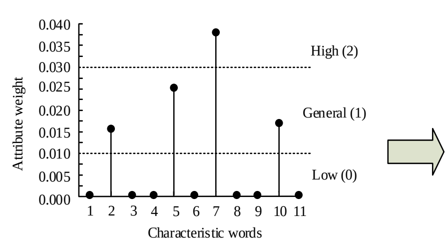
When marking the keywords extracted by the algorithm, it is evaluated by means of the hit rate (Recall), the accuracy rate (Precision) and the comprehensive evaluation (F measure).

$$\mathrm{Re}\,call = \frac{\left|Y_{real} \cap X_{re}\right|}{Y_{real}} * 100\% \tag{6}$$

$$\mathrm{Pr}\,ecision = \frac{\left|Y_{real} \cap X_{re}\right|}{Y_{real}} * 100\% \tag{7}$$

$$F - measure = \frac{2 * \mathrm{Re}\,call * \mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} * 100\% \tag{8}$$

In formulas (6)-(8), it Y_{real} represents the keyword set obtained by the measurement algorithm, which is Y_{real} the keyword set under manual marking. To diagnose and detect railway signal faults, it is possible to determine the fault feature layer, fault type layer and the fault cause layer realizes the subdivision and classification of fault information, and then provides a reference for subsequent countermeasures and improvements. For a variety of railway fault signal information, the text information data is often analyzed by means of association analysis. The value correlation or association pattern between itemsets is found in a large number of databases, emphasizing the interdependence between data, and then helping decision making and judging the feasibility of data. The key lies in the decomposition of association tasks and the formulation of association rules. The running framework of association rules is shown in Fig. 2.



Characteristic words	Attribute weight	Weight dispersion result
1	0	0
2	0.015	1
3	0	0
4	0	0
5	0.025	1
6	0	0
7	0.038	2
8	0	0
9	0	0
10	0.017	1
11	0	0

Fig. 1. Schematic diagram of grade division under weight discretization.

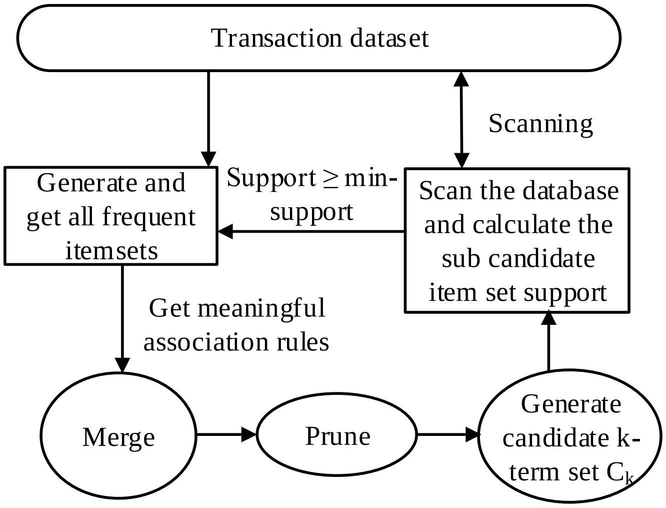


Fig. 2. Schematic diagram of association rule operation framework.

Association rules mainly realize the processing of itemsets by scanning the data set and determining the support degree, which is $A \to B$ an implication of the form, which A, B is an itemset containing one or more items, $A \cap B = \Phi$, which is the antecedent A, B of the association rule. and consequent. The constraints of association rules are the minimum values of support and confidence, and their mathematical expressions are shown in equations (9) and (10).

$$\sup(A) = count(A \subseteq T)/|D| \tag{9}$$

Equation (9) is the calculation formula of the support degree, which is expressed as the ratio of the frequency of a certain item set appearing in the transaction to the corresponding transaction number, where D, A is the corresponding transaction database and item set, and $\sup_{-} count$ is the corresponding support degree count. Confidence is expressed as the ratio of the probability that two related rules appear at the same time to the probability that a single rule appears $confidence(A \Rightarrow B)$, and its mathematical expression is Equation (10).

$$confidence(A \Rightarrow B) = \sup_{-} count(A \cup B)/\sup_{-} count(A) \tag{10}$$

Frequent Pattern-Growth (FP-Growth), as a typical algorithm for solving association problems, is different from the Apriori algorithm that scans each potential frequent itemset, which only scans the database twice and does not generate candidate sets. In the case of compressing the database containing frequent itemsets into a frequent pattern tree (Frequent Pattern-Tree, FP-tree), the frequent itemsets are obtained. The FP-Growth algorithm reduces the processing of itemsets to a certain extent. In the transaction database processing, the sorted frequent itemsets are used to reconstruct the database, then the root node of the tree is created, and the transactions are compressed into the FP-Tree branch in the form of merge prefix to realize the frequent itemsets. Fig. 3 shows the construction process of the algorithm pattern number.

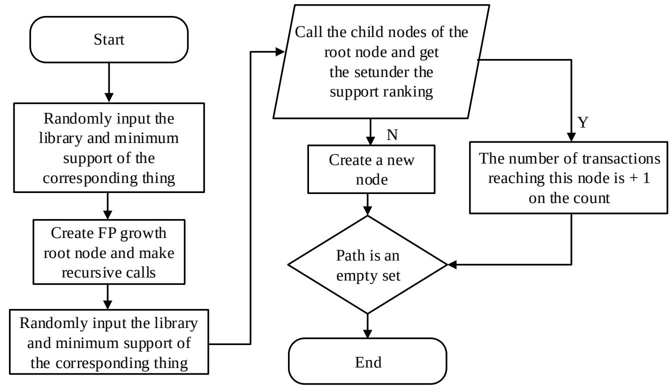


Fig. 3. Construction process of algorithm pattern number.

However, the traditional FP-Growth algorithm will artificially set the minimum support degree and the minimum confidence degree when dealing with related transactions, which makes the selection of algorithm parameters to a large extent subjective and arbitrary, and it is difficult to ensure the authenticity of the algorithm and the operation. Feasibility. Also, the ambiguity of determining the size of frequent itemsets will also cause the pattern tree to consume more memory during the traversal process, thereby reducing the efficiency of data mining. Therefore, the research is based on this to eliminate the defects of artificially setting parameters., based on the characteristics of the data itself, the minimum support and confidence are set. The improved FP-Growth algorithm first sorts the support number of the transaction set in descending order, and builds a sequence table corresponding to "order-value", in which There is a monotonically decreasing relationship trend between the sequence value and the ordinal value. The polynomial curve fitting function is set by the numbers corresponding to the ordinal-value table, and its mathematical expression is shown in formula (11).

$$y = f(x) = \sum_{i=0}^{t} m_i * x^i \tag{11}$$

In formula (11), y, x is the serial number value and the sequence value, t is the support number, i is the item set, and m is the upper limit of the interval satisfied by the function. Perform the quadratic derivative of the obtained function to obtain the second-order derivative function, as shown in formula (12). Show:

$$y' = f''(x) = \sum_{i=2}^{t} i(i-1) * m_i * x^{i-2} \tag{12}$$

When the sequence value of the second derivative function of the fitting curve appears for the first time in the interval, $y' = f''(x) = 0$ the sequence number value under the point is rounded down, and the rounded result is used as the algorithm parameter. Then set according to the adaptive support degree Threshold, get deleted items less than support degree, merge them to get sub-database, build its TP-Tree with sub-database, at this time, the data sub-database realizes the association of frequent itemsets under the recursive call tree structure Mining, but also to ensure that the obtained results are strong association rules.

B. Design of Railway Fault Signal Diagnosis System

The detection and diagnosis of railway signals based on the improved FP-Growth algorithm is to mine the rules in the fault storage knowledge base, and the reasoning flow chart is shown in Fig. 4.

By judging the fault characteristics of railway signals and saving them in a comprehensive database, and judging whether there are countermeasures for the type of problem, if so, the fault diagnosis results will be output directly, if not, the information characteristics will be screened in the historical fault database. Extraction, reasoning and correlation judgment, and consider a variety of factual basis in the process of fault type and cause analysis, and directly seek manual diagnosis when it is difficult to obtain a clear diagnosis result. This fault diagnosis process can effectively realize the processing of information data, detection, judgment, etc. effectively ensure the efficiency and accuracy of detection. The diagnosis of railway fault signals is the basis for staff to carry out equipment maintenance. The design of railway signal equipment fault maintenance system can make the diagnosis results intuitive, so as to carry out corresponding diagnosis countermeasures Research. On the basis of improving the FP-Growth algorithm, the research proposes a multi-level signal diagnosis structure including fault diagnosis, fault data management, post-maintenance management and user management. The picture shows the architecture diagram of the railway signal equipment fault repair prototype system.

As shown in Fig. 5, the architecture diagram of the fault repair prototype system includes the presentation layer, the business logic layer and the data access layer to realize the input and storage of the fault data, the logical interaction of the data business information and the human-computer interaction of the personnel related to the fault information., and then realize the cause and effect of fault data and interactive processing. The most important one is the business logic layer. The fault word database is established through the fault signal record, and then the knowledge base and data reasoning are used to diagnose the strong correlation. The fault causes of the rules are identified, and this is used as a reference for the maintenance plan to judge whether it needs manual modification. Continuously improve the fault case database of railway signal equipment and improve the overall fault diagnosis efficiency.

At the same time, when combining association rule mining and manual identification and diagnosis results, in order to avoid the omission of important fault causes, $P@N$ is used as the evaluation index, and its meaning is expressed as the accuracy rate of the retrieved fault cause as the real cause, and its calculation formula as shown in formula (13).

$$P@N = \sum_{j=1}^{n} I(j)/n \tag{13}$$

In formula (13), n is the set number of output fault causes, $I(j)$ and is the ratio between the input value of the test set conforming to the fault cause under the diagnosis algorithm and the actual cause of the fault, if the diagnosed cause is consistent with the actual cause, then $P@N$ The value of 1.

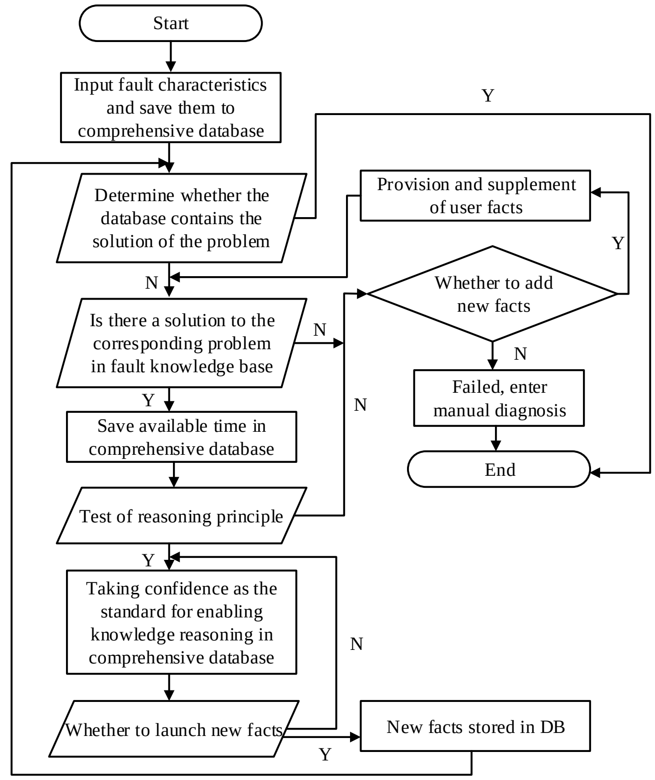


Fig. 4. Schematic diagram of fault diagnosis process.

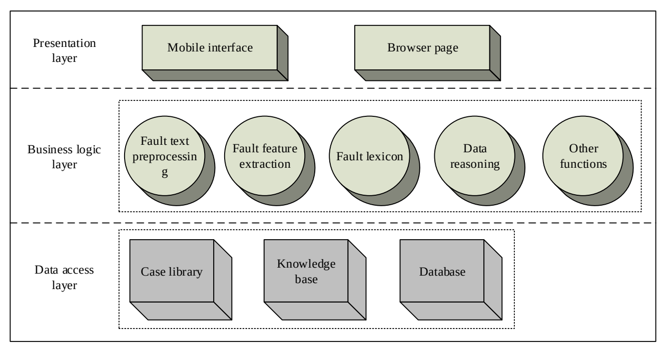


Fig. 5. Architecture diagram of railway signal equipment fault maintenance prototype system.

IV. Application Analysis of Railway Signal Equipment Fault Diagnosis under the Improved FP-Growth Algorithm

The development momentum of railway construction projects has shown a good trend with the enhancement of China's economic strength. As one of the main means of transportation in China, the safety monitoring awareness of railways increases with the development of safety accidents. Research is improving FP- With the support of the Growth algorithm, the fault signal system is designed and analyzed experimentally. First, the experimental development environment is set up. The fault detection application environment includes two parts: browser and mobile terminal. The main parameter setting conditions are shown in Table I.

TABLE I. DEVELOPMENT ENVIRONMENT OF THE SYSTEM

Software type	**Development environment**		**Version**
Browser	***Operating environment***		***Windows7 64bit***
	Hardware environment	Hardware environment	i7-4790 CPU @ 3.60GHz 8G R
		Development tool	MyEclipse8.5
	Software environment	Database	MySQL (5.5.2)
		JDK	JDK1.8
		The server	Tomcat7.0
Mobile end	Operating system		Android4.0.4 or above
	Hardware environment		RAM 512M ROM 1G or above
	Development tool		Android Studio

The key to the detection of railway faults is to accurately judge the causes of the fault types. However, the levels and contents involved in railway operation and maintenance are relatively complex, and the original fault storage data has a certain amount of meaningless content. Research Using the VSM algorithm as the comparison algorithm, randomly select transaction data in the fault database for analysis, and ensure the consistency of the corresponding algorithm parameter settings when comparing the algorithms. The VSM algorithm uses the space vector of the feature items to represent the text information. The text similarity problem is transformed into a dimensional space vector problem. The data is counted and sorted on the algorithm running time and correlation accuracy, and the results are shown in Fig. 6.

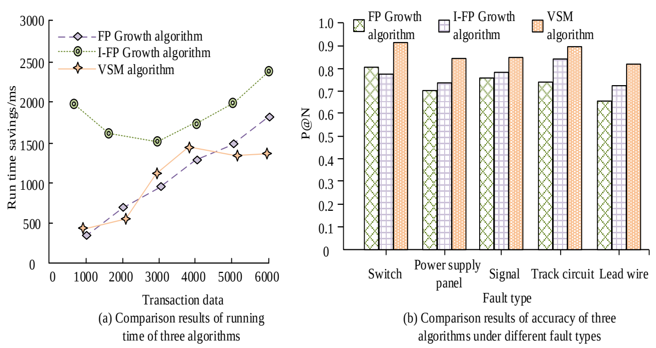


Fig. 6. Comparison results of running time and fault correlation accuracy of three algorithms.

In Fig. 6, "I - FP Growth algorithm" represents the improved FP-Growth algorithm proposed by the study. It can be seen from Fig. 6(a) that with the increase of transaction data volume, the performance of the three algorithms. The running time is quite different. The specific performance is that the overall running time of the traditional FP-Growth algorithm shows an upward trend. The minimum and maximum values are 350ms and 1700ms respectively, while the improved FP-Growth algorithm is It shows a downward trend first and then upward trend, and takes 3000 data sets as the dividing point, the minimum and maximum values are 1500ms and 2400ms, respectively. The VSM algorithm shows a relatively obvious numerical fluctuation. The reason is that the improvement of FP-Growth The algorithm can reduce the frequent itemsets occupying memory by dividing the database into subsets, and avoid the redundant database under the condition tree generation of the original algorithm. The results in Fig. 6(b) show that the P@ N is higher than the other two algorithms, and its average accuracy exceeds 85%, which is about 5 times and 9 times that of the other two algorithms. The average accuracy of the traditional FP-Growth algorithm and the VSM algorithm are 81.4% and 82.1, respectively. %. The reason is that the improved FP-Growth algorithm can mine the correlation between the fault information by discretizing the weight feature words, which greatly improves the diagnostic accuracy of the algorithm. However, the traditional FP Growth algorithm and VSM algorithm lack of consideration of the potential information in the fault text data, and when encountering texts with similar contents but with synonyms, the calculated similarity will be far from the actual. Because the number of topics is artificially determined, there is a certain degree of experience, which ultimately leads to less fault features, and it is difficult to deeply mine the associated meaning of fault information, so it is difficult to maintain the accuracy of inferred information. The accuracy of the three algorithms is calculated, the result is shown in Fig. 7.

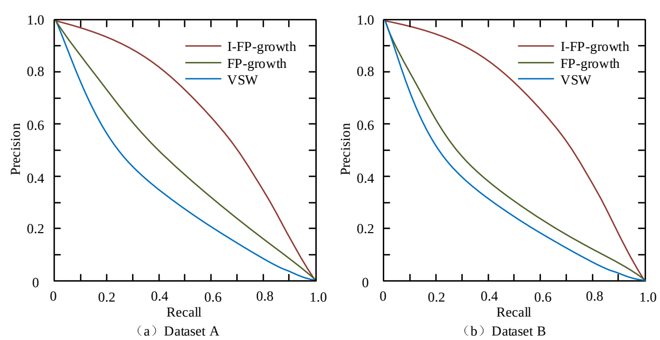
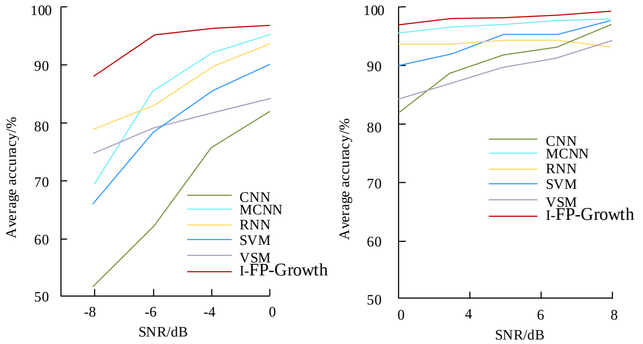


Fig. 7. PR curves of three algorithms.

The research uses the Precision-Recall (PR) curve as a judging tool. The more the PR curve is convex to the upper right, the better the data classification effect is. The PR curves on A and dataset B are closer to the upper right, and their average accuracy reaches 87.15%, which is much higher than the 83.12% and 81.06% of the FP-Growth algorithm and the VSM algorithm. The reason is that the VSM algorithm and the FP-Growth algorithm. It is difficult to reduce the negative impact of objective factors and algorithm parameters on the results of data processing only with vector features and correlations. The improved FP-Growth algorithm can effectively take into account the characteristics of railway signal faults for correlation analysis, reducing redundant and redundant data. Then, the statistical results of the algorithm are organized, and the results are shown in Fig. 8.



(a) Average precision results of different algorithms under positive signal-to-noise ratio

(b) Average precision results of different algorithms under negative signal-to-noise ratio

Fig. 8. Average precision results of different algorithms under signal-to-noise ratio.

Fig. 8 shows the average precision results of SNR under different algorithms. Fig. 8(a) shows that when the SNR is negative, the average precision results shown by various algorithms are quite different. When the ratio is -8 to -6, the average precision is from large to small as I- FP-Growth > RNN > VSM > MCNN > SVM > CNN, and when the signal-to-noise ratio is -6 to -0, the average precision is from large to the smaller is I- FP-Growth > MCNN > RNN > SVM > CNN. The average precision of the I- FP-Growth algorithm and the CNN algorithm shows a good upward trend, and the I-FP-Growth algorithm is better than the CNN algorithm. In Fig. 8(b), when the signal-to-noise ratio is positive, the overall average accuracy performance of the algorithms from good to poor is I -FP-Growth > MCNN > SVM > CNN > RNN > VSM, where I -FP-Growth algorithm The average precision shown is more than 95%, and the maximum value is 98.2%. The accuracy performance is second only to the I -FP-Growth algorithm for the MCNN algorithm, and the maximum value of the MCNN algorithm is 96.3%. As the ratio of signal power and noise power, the signal-to-noise ratio is one of the important indicators to characterize the signal quality. It may be difficult to express a good communication effect if signal power or noise power is used to represent the quality of signal processing. However, the reduction of signal-to-noise ratio will increase the bit error rate, which will greatly reduce the visual presentation effect of signal quality. When the SNR value is very low, it indicates that the working environment suffers from great noise interference, and the method in this paper still obtains satisfactory diagnostic accuracy in the task of adding very low SNR value noise signal. Although the noise causes a great difference in the data distribution of the training and test data sets in the target task, the method in this paper can still achieve effective fault diagnosis, which shows that the method in this paper has excellent robustness, can complete high-precision and stable fault diagnosis when the noise is seriously disturbed, and has good feature stability. The FP-Growth algorithm is less affected by the value of the signal-to-noise ratio, and is in an upward trend as a whole, and its average precision performance is better than other algorithms. The proposed algorithm is tested for signal generation and selected under different fault types. The signal data of the real curve is compared with it, and the results are shown in Fig. 9.

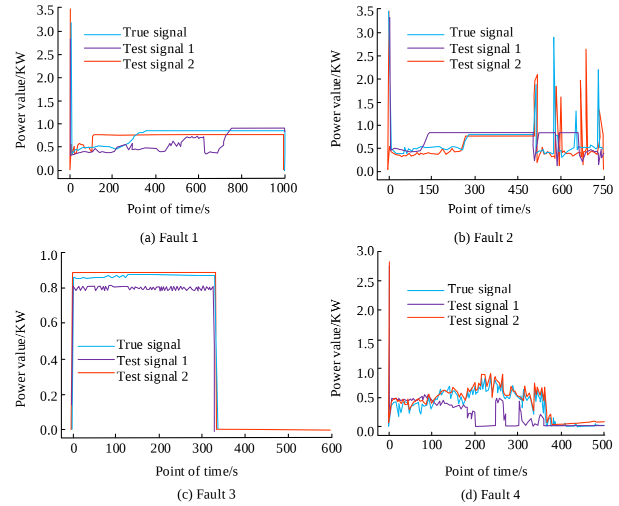


Fig. 9. Signal generation results under four fault types.

Fig. 9 represents the traditional FP-Growth algorithm, "Test signal 2" represents the improved FP-Growth algorithm, Fig. 9(a)-(d) are the track circuit fault, switch fault, signal fault and the connecting line is faulty. The results in Fig. 9(a) show that the test data 2 has a large difference between the real signal and the data before the time point of 300s, but with the increase of time, the magnitude of the power difference between the two is not More than 5%. However, the error between the signal curve graph of test data 1 and the real signal curve is large, and the fluctuation between the time point range (0-750) s is large. In Fig. 9(b), the test data the generated signal curve of 2 is continuous, and the time point at which the peak appears is basically the same as that of the real signal, and the trend is roughly the same. The signal curve of the improved algorithm in Fig. 9(c) is in the range of time points (0-380)s. The error range amplitude The value is less than 3%. The traditional FP-Growth algorithm curve has short and small fluctuations in a large range, and the error value compared with the real signal curve exceeds 6.8%. In Fig. 9(d), the curve of the improved FP-Growth algorithm is compared with the real signal. The fluctuation of the curve is basically the same, and the error margin value is less than 1%. However, the signal trend curve of the traditional FP-Growth algorithm has begun to deviate from the real signal curve after the time point is 100s, and the signal missing phenomenon occurs at the time point of 400s. The above results show that the improved FP-Growth algorithm can effectively detect and analyze signals of different fault types, and its accuracy and algorithm performance are better, which can effectively provide data support for fault diagnosis and maintenance countermeasures.

V. Conclusion

At present, the detection and diagnosis of railway faults mostly rely on manual operation. The complexity and redundancy of data types increase the difficulty and accuracy of fault analysis, and it is difficult to provide a guarantee for railway safety operation and maintenance. Based on the analysis and extraction, the improved FP-Growth algorithm is used to detect the fault signal data, and its performance test and application analysis are carried out. The results show that the improved FP-Growth algorithm takes 3000 data sets as the dividing point. The maximum and minimum time are 2400ms and 1500ms, respectively, which are significantly higher than the 350ms and 1700ms of the traditional FP-Growth algorithm, and its P@N accuracy is 5 times and 9 times that of the other two algorithms. The classification performance of data mining is better, and its average accuracy reaches 87.15%, which is much higher than 83.12% and 81.06% of the FP-Growth algorithm and VSM algorithm. The average accuracy of the Growth algorithm is better than other algorithms, with a maximum value of 98.2%. The error values between the signal generation results and the real signals under the four fault types are small, all lower than 5%. The above results show that the improved FP-Growth algorithm proposed by the study is less affected by the value of the signal-to-noise ratio, and its performance is better than other algorithms in terms of running time saving and accuracy, and can effectively diagnose fault data. The research only selects text data as the data source for analysis, which makes the fault problems it solves more limited and less comprehensive. For example, the data generated by microcomputer monitoring may also be an important source of fault problems. Strengthening the diagnosis and analysis of multiple data is one of the important contents to be studied in the future. At the same time, the research only uses the feature weight number as the source of fault analysis, without taking into account the post fault risk, Therefore, it is necessary to carry out combined weighting in the later stage to ensure the accuracy of fault information features.

VI. Findings

The research is supported by Provincial Project: Research and Practice of "223" Talent Training Model for Railway locomotive operation and maintenance professional (No. JGJX2022D721).

References

[1] E. Wang, D. Yang, C. Wang, et al., "Optimized Fault Detection Algorithm Aided by BDS Baseband Signal for Train Positioning," Chinese Journal of Electronics, vol. 29(1), pp.34-40, 2020.

[2] N. Xu, X. Xiu, J. Liu, et al., "A novel scheme for multivariate statistical fault detection with application to the Tennessee Eastman process," Mathematical Foundations of Computing, vol. 4(3), pp.167-184, 2021.

[3] X. Fang, W. Huang, D. Ye, et al., "Application of distributed parallel FP-growth algorithm in secondary equipment defect monitoring," Power System Protection and Control, vol. 49(8), pp.160-167, 2021.

[4] X. Jiang, X. Xiao, W. Wen, et al., "Operation optimization of thermal power unit based on Parallel FP-Growth algorithm," Guangxi Electric Power, vol. 42(1), pp.11-14,19, 2019.

[5] L. Yang, P. Li, R. Xue, et al., "Intelligent classification of faults of railway signal equipment based on imbalanced text data mining," Tiedao Xuebao/Journal of the China Railway Society, vol. 40(2), pp.59-66, 2018.

[6] C. Yuan, P. Li, and Y. Zhang, "Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing," Future Generation Computer Systems, vol. 88, pp.279-283, 2018.

[7] C. doves Han, "Research and improvement of FP-Growth algorithm," Journal of Longyan University, vol. 38(5), pp.8-13, 2020.

[8] Q. I. Jiyang, and L. Wang, "Fault diagnosis method of equipment based on multi-information fusion," International Equipment Engineering and management, vol. 25(2), pp.77-97, 2020.

[9] Z. Zhong, T. Tang, and F. Wang, "Research on fault feature extraction and diagnosis of railway switches based on PLSA and SVM," Tiedao Xuebao/Journal of the China Railway Society, vol. 40(7), pp.80-87, 2018.

[10] X. Zhang, "Research on fault diagnosis method of railway signal equipment based on association rules," Railway Standard Design, vol. 66(4), pp.175-181, 2022.

[11] Y. Cheng, N. Zhou, W. Zhang, et al., "Application of an improved minimum entropy deconvolution method for railway rolling element bearing fault diagnosis," Journal of Sound & Vibration, vol. 425:53-69, 2018.

[12] W. Zheng, Q. Li, and Y. Zhang, "Uav flight fault diagnosis based on improved FP-Growth algorithm," Science & Technology Innovation and Application, vol. 2019(13), pp.16-19, 2019.

[13] D. Ding, Y. Yi, and M. Li, "Tire quality data analysis based on FP-growth algorithm," China Mechanical Engineering, vol. 30(2), pp.244-251, 2019.

[14] Y. Li, and Z. Pan, "Analog circuit fault diagnosis methods based on rbf neural network," Telecommunications and Radio Engineering, vol. 78(13), pp.1193-1201, 2019.

[15] M. Jin, H. Li, J. Xie, et al., "Rule based intelligent fault diagnosis and recovery power delivery algorithm," Dianli Xitong Baohu yu Kongzhi/Power System Protection and Control, vol. 47(4), pp.176-181, 2019.

[16] B. H. Cho, and H. Lee, "Study on operation characteristics and failure causes of LED type railway traffic lights," Journal of the Korean society for railway, vol. 21(8), pp.767-773, 2018.

[17] X. Fang, W. Huang, D. Ye, et al., "Application of distributed parallel fp-growth algorithm in secondary equipment defect monitoring," Power System Protection and Control, vol. 49(8), pp.160-167, 2021.

[18] J. Zhang, S. Yi, G.U.O. Liang. et al., "A new bearing fault diagnosis method based on modified convolutional neural networks," Chinese Journal of Aeronautics, vol. 33; 167(02), pp.54-62, 2020.

[19] M. A. Syakur, B. K. Khotimah, E. Rochman, et al., "Using K-means algorithm and FP-growth base on FP-tree structure for recommendation customer SME," Journal of Theoretical and Applied Information Technology, vol. 96(4), pp.1102-1113, 2018.

[20] W. W. Ariestya, W, Supriyatin, and I. Astuti. "Marketing strategy for the determination of staple consumer products using fp-growth and apriori algorithm," Jurnal Ilmiah Ekonomi Bisnis, vol. 24(3), pp.225-235, 2019.

[21] C. Tan, G. Ji, R. Liu, et al., "LTSA-LE: A local tangent space alignment label enhancement algorithm," Tsinghua Science and Technology, 26(2), pp.135-145, 2020.

A Systematic Literature Review: Internet of Things on Smart Greenhouse

Dodi Yudo Setyawan[1], Warsito[2]*, Roniyus Marjunus[3], Nurfiana[4], Rahmalia Syahputri[5]
Doctoral Program of Mathematics and Natural Sciences, Lampung University[1]
Department of Computer System, Faculty of Computer Science, Institute Informatics and Business Darmajaya[1, 4]
Department of Physics, Faculty of Mathematics and Natural Sciences [2, 3]
Department of Informatics Engineering, Faculty of Computer Science, Institute Informatics and Business Darmajaya[5]
Lampung University, Bandar Lampung, Indonesia[1, 2, 3]

***Abstract*—The Increasing food needs and climate instability require researchers to innovate agriculture using smart greenhouses that are integrated with the Internet of Things (IoT). The Systematic Literature Review (SLR) begins with determining the topic keywords followed by searching the publisher link. It obtains 301 publications to be reviewed, 58 of which address the research questions posed. This study aims to collect and analyze in depth various knowledge about the Internet of Things Smart Greenhouse regarding sensors, methods used and publishers who publish the most related topics and the possibility of a research gap. The findings are as many as 12 publications that use temperature and humidity sensors and use the research and development methods integrated with Artificial intelligence methods, of which 62.1% do not use the datasets and 37.9% use the datasets. It obtains two possibilities of a research gap, namely improvising the algorithm and the dataset used and placing full control on the microcontroller development board and making IoT a supporting tool.**

Keywords—IoT; SLR; Smart greenhouse; agriculture; research gap

I. INTRODUCTION

As the world's population grows and the need for food continues to increase and the decrease in agricultural land that is used as land for residence requires us to continue to innovate to meet these needs. Not only that climate change, which is increasingly erratic, make the planting period of food crops unstable. One of the innovations that can be done is to use a greenhouse in the agricultural process, by using a greenhouse the planting process can be carried out continuously regardless of the current climatic conditions. The placement of a greenhouse can also be done in places with a limited area such as the rooftop of a building. However, the greenhouse is just a place to grow crops, there needs to be innovation which will later be known as the Internet of Things Smart Greenhouse (IoTSG). These technological innovations include automation, artificial intelligence and the internet of things. The integration of these innovations is proven to have a good impact on agricultural processes and results in greenhouses as described below. This study dissected and dag deeper into various IoTSG literature regarding the sensors used, the methods applied, publisher statistics and the possibility of research gaps that were obtained so that improvisation of the IoTSG could be done.

More specifically about smart green houses, by using smart greenhouses, pest control of food crops can be done easily. As has been done by [1], the robot sprays pesticides around the plants for pest control. Preventive measures by predicting plant diseases in greenhouses can also be carried out as done by [2], [3] on cucumber and vegetable crops.

Preventive measures not only in controlling plant pest diseases, the harvesting process of mushroom plants can also be carried out by robots in a greenhouse environment [4], [5]. Irrigation management in greenhouses and large-scale planting areas can be done easily with the integration of IoT technology and remote sensing [6], [7]. Using the water more effectively and efficiently is an important aspect and needs to be taken seriously. Adequacy of water sources can be monitored and controlled properly by farmers, maintainers and administrators using mobile apps.

Various mobile apps based on the IoT platform can be used as part of an agriculture information system, such as in the FIWARE case [8], [9], this system allows every mobile app holder to convey information models sent to the cloud such as planting, harvesting, product packing, transportation and distribution of agricultural products. All the information is stored properly in the cloud database.

The development of smart greenhouses equipped with sensors, actuators and cloud connections as well as artificial intelligence algorithms has also been carried out, but the IoT platform system is still suspended [10]. The system that is built is called an automated hybrid, some systems work automatically and some are controlled manually.

In sensor data acquisition, the measurement accuracy of the measuring instrument is very important so that the resulting decision will be very accurate. What needs to be done is the calibration of measuring instruments made by comparison with standard measuring instruments. Not only on measuring instruments or IoT sensor systems, calibration is also carried out on measuring instruments or geomagnetic sensor systems and soil temperature sensors that produce the MAG3119 standard deviation sensor for x coordinates of 8.5, y coordinates of 2.66 and z coordinates of 1.9 and standard deviation of the DHT11 sensor of 0.1161 [11].

*Corresponding Author.

II. Research Methodology

This SLR was carried out with the ultimate goal of knowing the main information about the IoTSG that was documented. The SLR that was carried out was different from Traditional Literature Review (TLR), the SLR was carried out methodically and systematically. A good review is a review that can be imitated by other researchers and produces a better scientific value. In addition, it can also evaluate all existing evidence on a validated research topic. The SLR process that has been carried out can be seen in Fig. 1.

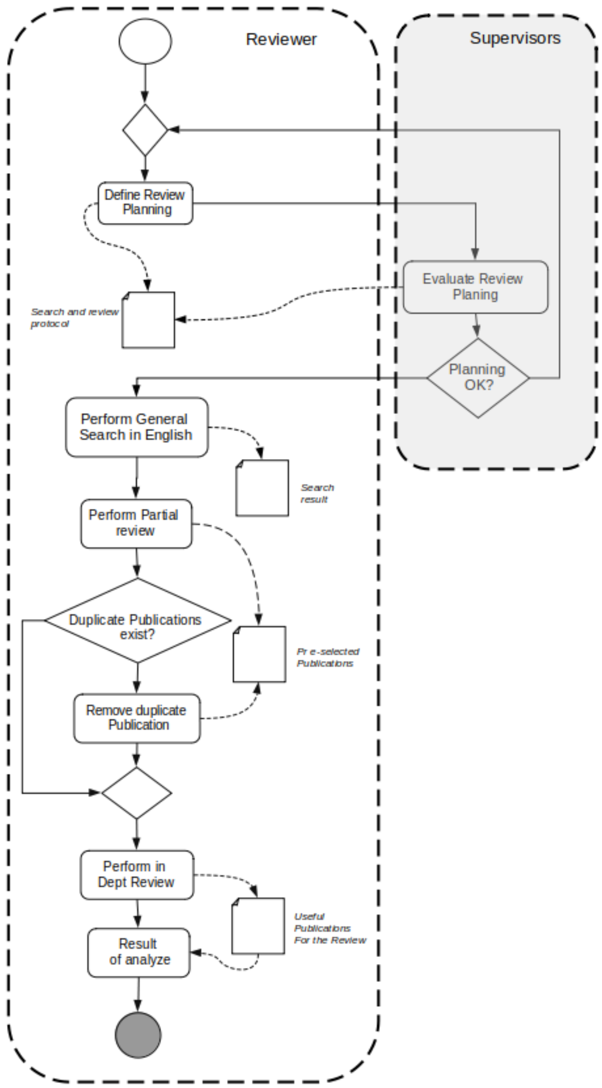


Fig. 1. SLR process [12].

The first thing to do in this SLR is planning a literature review by considering the needs of future research and defining a literature search protocol. Next, the supervisor analyzes and assesses its adequacy. Literature searches were carried out on various sources such as Sciencedirect and IEEE. This search was conducted using only English keywords. Next, conduct an initial review of the search results that have been obtained to optimize publications that are very useful. To delete duplicate publications and the rest of the results of this initial review, an in-depth analysis is carried out. The planning and realization stages of the SLR are described next, while the analysis of the results is presented in Section 4.

A. SLR Planning

To achieve the objectives as described above, Research Questions (RQ) are prepared to define research needs in more detail.

RQ1: What are the sensors used

RQ2: What are the methods used

RQ3: What publishers most related topics

RQ4: What are the research gap possibility

Literature search links include: Science Direct (https://www.sciencedirect.com/), IEEE (https://ieeexplore.ieee.org/Xplore/home.jsp) and MDPI publisher (https://www.mdpi.com/).

B. Search String

The search string uses keywords from the pre-planned research topic. There are four keywords used in this search, namely: IoT, farming, agriculture and greenhouse. The combination of four keywords above for literature searches on each search link is as follows:

C. IoT and Farming and Agriculture and Greenhouse

The search year is limited to 2017 to 2022. Search results are entered into the Zotero reference manager (https://www.zotero.org/) by forming a folder with the name of each folder, including: Sciencedirect and IEEE.

III. Result and Discussion

A. Search String Review

At this stage, a partial review of literature obtained was carried out to obtain the most potential literature in this study. This partial review was carried out on the title and abstract, but in certain specifications a review was carried out on the introduction and conclusion. Definitions related to RQ were included and potentially useful literature was obtained. The realization of this SLR was only done with one search string. The total articles obtained were 301 articles as shown in Fig. 2.

The results were generally displayed at the beginning and the answer for each RQ was placed after it. From each reviewed paper, on average, each paper uses more than one sensor. The general number of sensors used were 26 sensors as shown in Fig. 3. The most widely used sensors in IoTSG were temperature and humidity sensors as many as 12 and the second most sensors are soil moisture sensors with 10.

The method used in each paper was a research and development method integrated with various other methods such as the addition of artificial intelligence as decision support such as recurrent neural networks, artificial neural networks, fuzzy logic, decision trees and so on.

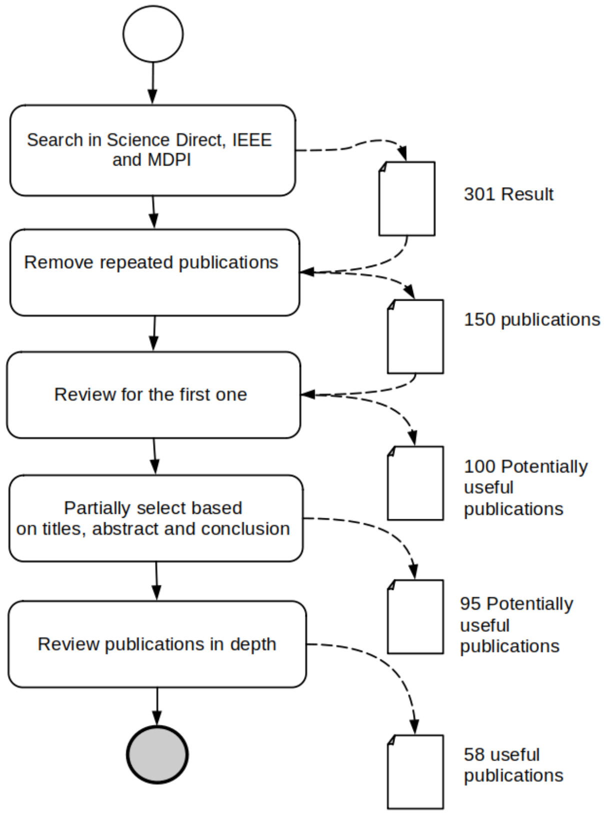


Fig. 2. SLR Step.

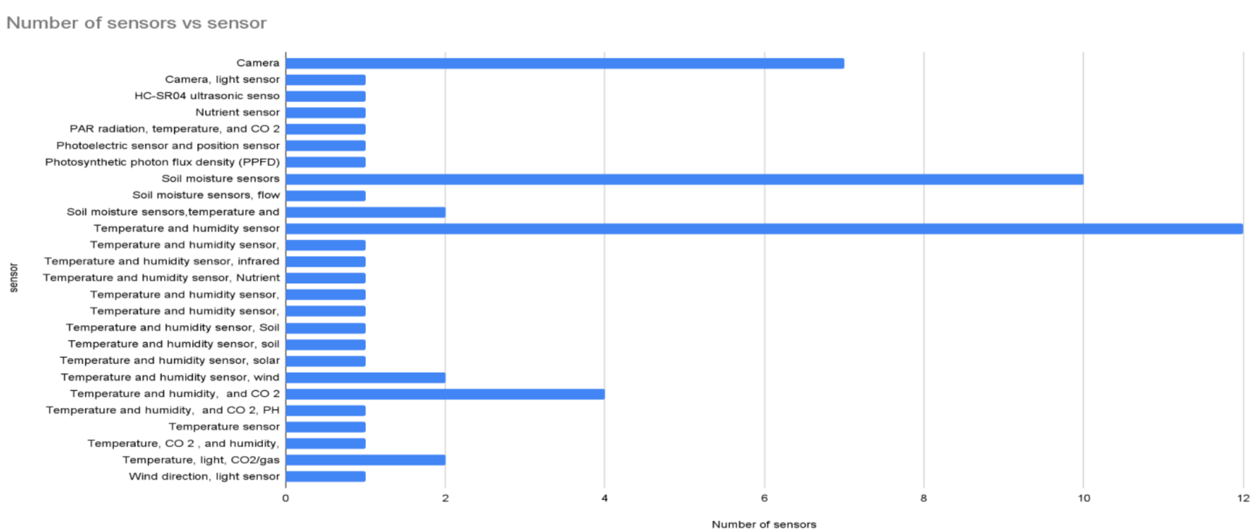


Fig. 3. Sensors used

Development of a meteorological information system using the KOSEN Weather Station (KWS) [13], the Google web toolkit implementing the inter-browser communication method of remote method calls [14], sharing and linking data using partnerships for data innovation that provide the information needs of farmers [15]. The system can help farmers with weather forecasts that are stored in the cloud and information can be viewed via a web that is easy for farmers to understand. To compare the level of accuracy of KWS, the same data was also built using the WXT520 sensor and it was proven that KWS was more accurate.

In greenhouses, there is a way to maximize plant growth by adding lighting using LEDs [16]. The method used was research and development to build an automatic control system for adding lighting using three LEDs among others: red LED, blue LED and a combination of both. Plant growth was observed using IoT, the results obtained from the addition of lighting were better seen from the leaf area index, leaf area rate index and leaf weight rate, besides that the photosynthetic activity of plants was the highest if green light was added in terms of redistribution of quanta absorbed by the plant the entire leaf thickness [17].

By utilizing the application of IoT technology in agriculture, it is proven to increase productivity and production, harvest quality and resource efficiency. This research and development has been carried out using case studies and models as a service approach [18]–[21] real time and historical data and predictive models can be accessed via the web [22]. Firebase servers [23], one of the protocols used is LoRa which is considered sufficient to transmit small data from sensors to the cloud [24]–[26]. The system helps farmers because it produces a Decision Support System (DSS).

To improve the sustainability of agricultural cultivation, research and development methods are used integrated with deep learning, maps of planting sustainability forecasts can be obtained in real time using IoT networks [27].

Energy use in agricultural IoT systems is very important, it is necessary to pay attention and take into account the peak times and prediction models of energy use in IoT-based greenhouses have been made. Various algorithms have been used with research and development integrated with machine learning algorithms such as artificial neural networks, support vector machines, extreme gradient boost and random forest, in research [28] it was found that random forest produces predictions with the highest accuracy compared to other algorithms of 92 %. Integrate the Multi agent System to build greenhouse energy management to optimize energy use for two seasons in one year [29]. The original genetic algorithm was also used for predictive lighting for lighting control in the greenhouse, from the result of measurements that was carried out on growth there was no difference between greenhouses with lighting control and those without lighting control, but there was an efficient use of electrical energy from greenhouses that used lighting control [30].

Various algorithms have been developed for smart agriculture in greenhouses for various estimates that produce contributions, one of which is the algorithm that has been developed is a Repetitive Neural Network (RNN) which can estimate plant growth rates in greenhouses from 0.75 to 0.81 in all growth periods [31].

To obtain data from physical dynamics, a greenhouse requires sensors that are integrated in a wide range. A sensor network was built in a wireless sensor network (WSN) using a wireless Zigbee module with a mesh topology to sense air temperature, CO_2 levels [32]. In addition to the Zigbee module, you can also use the SL900A module in the UHF band [33]. The measurement of wave attenuation for the 2.4 GHz frequency band that penetrated the greenhouse was also measured to determine the effect of the waves on plants [34]. The data loss rate between data acquisition units in WSN was also measured, where data loss at the gateway was 1.52%, between gateways and server was 0.4% [35].

Not only in terms of sensor networks in greenhouses, the development of robot agriculture with a three-device system (3DS) for growth plant monitoring and pest management used the main features of spectroscopy in Automated Guided Vehicles (AGV) [36]. In addition, by utilizing image analysis, image processing technology based on IoT cameras on tomato cultivation in greenhouses was able to estimate the number of blooming flowers, ripe fruit and harvest date with an error of approximately 2.03 days which can support plant growth management [37]. Sending the water level measurement value from the HC-SR04 sensor using a short message service shows good results and efficient [38] , in maximum water level is 3.182m have resolution about 19.2 mm [39] and using MW22B multi turn Potentiometer have obtained vertical resolution about 0.03 m and error is 1.11% [40]. Detection of whole tomatoes used research and development methods by integrating Convolutional Neural Network (R-CNN) with an accuracy of 87.83% [41]. While using deep learning based classification and detection, the accuracy rate reached 99% [42], this method showed that the method used had great potential for predicting tomato ripeness and yield. A genetic algorithm was used to predict tomato growth using data on air temperature, PAR radiation and CO2 concentration [43].

The algorithms used to estimate plant growth in greenhouses can also be used to estimate temperatures in cold greenhouses which cause the greenhouse to freeze, the method used is research and development by integrating Artificial Neural Network (ANN) algorithms and Fuzzy Associative Memory (FAM) with the effectiveness of the results reaching 90% accurate [44] The root mean square error for temperature prediction with a value of 0.2, CO_2 1.29 and humidity 0.14 [45], using a network of long short term memory and Recurrent NN with a root value mean square error of 0.289 [46]. The estimation of soil moisture content or soil moisture as a planting medium in a greenhouse can also use the ARX, ARMX, BJ and State algorithms with values of 91.31%, 91.09%, 91.08% and 90.75% [47]. The development model of Soil Moisture Forecasting (SMF) to activate Constant Moisture Automatic Irrigation (CAIS) using soil moisture data using sensors on planting media soil with different depths with the estimated error values of 0.011 [48], 0.962 [49], 0.014 [50]. The use of capacitive soil moisture sensors in measuring soil moisture is more resistant to rust than resistive soil moisture [51]. While using a model based on long short term memory, the error value is 0.72% [52] and if using a synthesis algorithm and combining it with a triangular affiliation function 0.99% [53]. By controlling soil moisture in the planting medium in the greenhouse automatically the estimation results of the algorithm used can save both, water irrigation and plant nutrients [54]. To overcome the problems of irrigation control due to the weakness of the internet network, an edge of the network was created. All sensor data is stored in the cloud database if there is a network problem it will affect the resulting decision [55].

The application of fuzzy inference engine determines for control and monitoring of hydroponic plants in greenhouses, the defuzzification obtained from sensor network data was used to control irrigation. Plant growth was monitored as input fuzzy parameters [56]. The application of fuzzy logic [57] and Neural Network [58] in this irrigation control can improve water use efficiency. Sensor networks have also been created in several zoning to facilitate obtaining fuzzy parameter input data based on crop area [59]. The water needs of the plants are properly fulfilled and the efficiency of water use is obtained.

Generally, the roofing material for greenhouses is Ultra violet (UV) plastic, but the presence of a transparent photovoltaic solar module that is used as a roof as well as a source of energy for the greenhouse is a distinct advantage [60]. Sunlight as a light source for the photosynthesis process of plants in the greenhouse can enter through transparent photovoltaic which is also a source of electrical energy for the greenhouse.

In addition to the sensor network in the greenhouse which is used as data input in controlling the application of intelligent robots based on the Intelligent Mechatronic System, research has also been carried out. This intelligent robot was tasked with monitoring plant growth conditions in the greenhouse, this robot was used as a substitute for officers in the greenhouse and had an operating efficiency of 95.86% [61]. The use of drones in greenhouses can also help manage greenhouses. Simultaneous visual localization and mapping algorithms and ORB-SLAM utilize the camera as a sensor [62].

The publishers who published the most papers with related topics are MDPI as much as 56.9%, IFAC-Paper Online as much and Information Processing Agriculture as much as 10.3% and the remaining less than 5% as shown in Fig. 4, from this information we can obtain information highest probability of publishers publishing related topics. So that the technical paper that will be produced from this SLR is most likely to be published by the publishers who publish the most related topics.

To find out the quality of publishers at the Quartile level, each paper was analyzed through https://www.scimagojr.com/ by using the ISSN number in each paper as a search keyword. The results of the journal rank analysis can be seen in Fig. 5.

Paper in quartile 1 is 72.4%, quartile 2 is 17.2% and quartile 3 is 10.3%. This shows that the papers on this SLR, which will later become references to technical papers, occupy journal quality in quartiles 1 to 3.

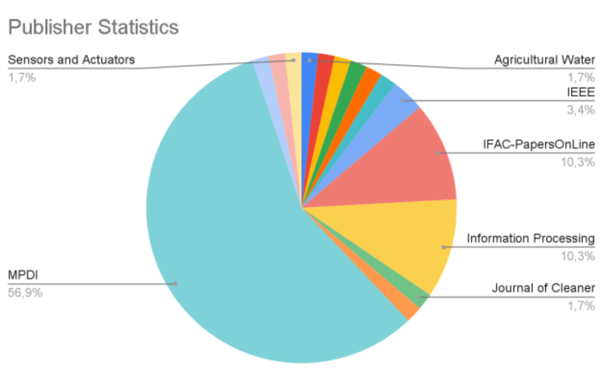


Fig. 4. Publisher statistics.

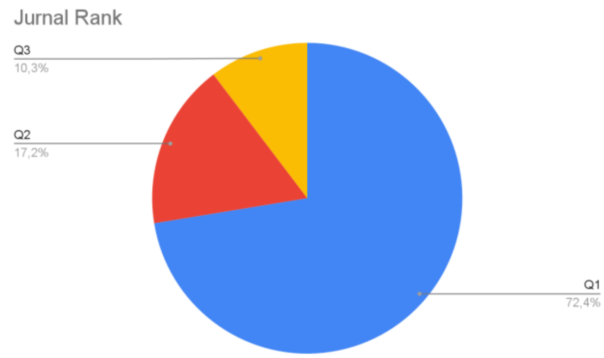


Fig. 5. Rank journal.

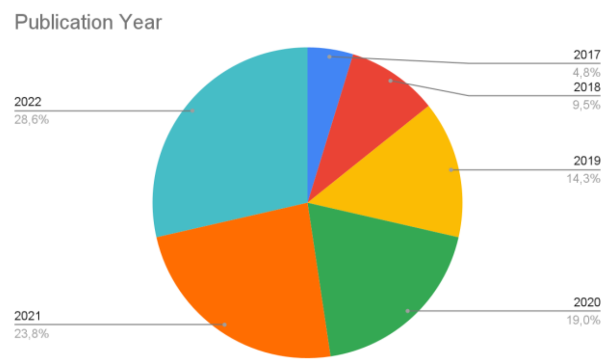


Fig. 6. Publication year.

To find out the research position of related topics, an analysis of the year of publication of each paper, the results are as shown in Fig. 6 above with the number of papers published in 2022 of 28.6% (until August 2022), 23.8% in 2021, 19.0% in 2020, 14.3% in 2019, 9.5% in 2018 and 4.8% in 2017. Judging from the increase in the number of related topics from year to year, it shows that related topics are still good topics for research.

Answering RQ4 to obtain a research gap, one of the analysis carried out was to group paper papers into two groups, namely groups of papers that use datasets and those that do not. The result obtained was 62.1% not using a dataset and 37.9% using a dataset as shown in Fig. 7 using a dataset.

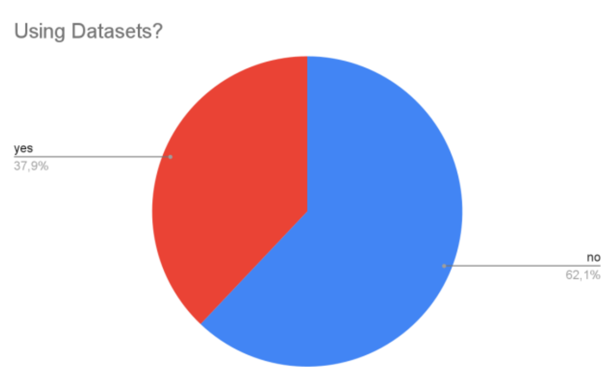


Fig. 7. Dataset.

B. The Possibility of a Research Gap 1

To obtain the possibility of a research gap, a more in-depth analysis was carried out, both on papers that used datasets or not. Papers that used datasets were papers whose research used artificial intelligence such as machine learning using artificial neural network algorithms, decision trees, and so on. The dataset used was a private dataset built by collecting data for some time, then the dataset obtained was used to train the algorithm created. Prediction accuracy and speed as well as memory usage was recorded as a result of the study. To improvise the algorithms and datasets that were created, we corresponded to researchers using the correspondence email contained in each paper. We asked for the source code and dataset, then compiled and tested it on another microprocessor, after obtaining the accuracy and predictive speed and using the same memory, we tried to improvise and produced a research gap or novelty, and the results were presented in another technical paper. For example, an algorithm used to create a greenhouse temperature model using the Light Gradient Boosting Machine algorithm (LGBM) [63] median error /^{o}C -0.027, mean error/^{o}C -0.058 and Root Mean Square Error /^{o}C 0.645, it is hoped that after improvement the algorithm (call the x algorithm) a smaller predictive error value is obtained as shown in Fig. 8.

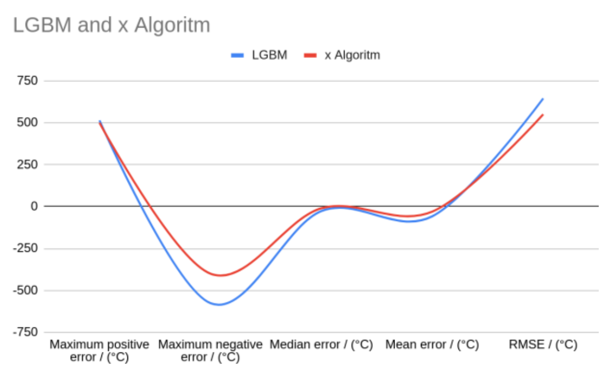


Fig. 8. x algorithm.

Not only did the predictive error value improve, the time training x algorithm also improved as shown in Fig. 9.

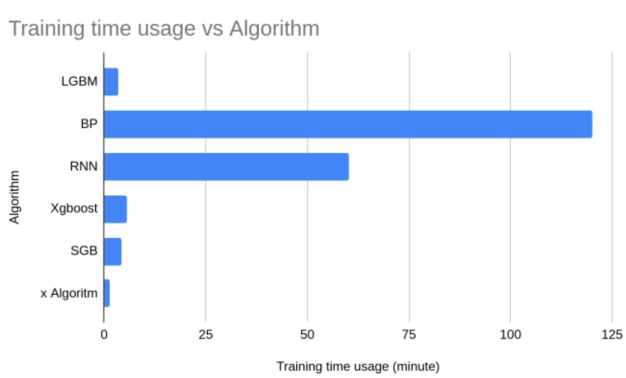


Fig. 9. Time training algorithm.

For papers that did not use datasets in their research, they only built systems for control and monitoring in greenhouses. In general, the system built in each paper can be seen in Fig. 10 as follows.

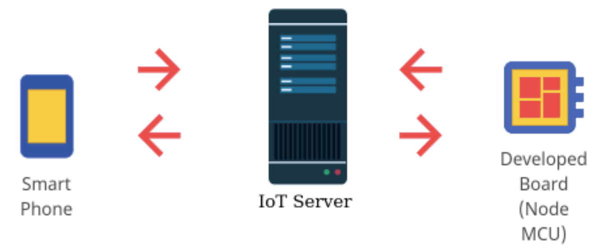


Fig. 10. General IoT system.

The IoTSG system that was created for both control and monitoring was completely on the server or on smart phone mobile apps. The developed board or nodeMCU is only a transmitter and receiver of the sensors and actuators in the greenhouse. The system is very dependent on internet connection, if the internet connection is good then monitoring greenhouse conditions such as temperature, humidity, soil moisture and soil nutrient levels were monitored properly as well as controlling actuators for the process of watering plants and fertilizing can be done properly too. If the internet connection is not good or disconnected, the system will die, monitoring and controlling cannot be done, which can be fatal to plants.

C. The Possibility of a Research Gap 2

To overcome the weaknesses of the IoT system, as stated above, it is necessary to add a developed board that is connected to sensors and actuators in the greenhouse to carry out real time control and monitoring and to compose the IoT system as supporting tools instead of full monitoring and control. In addition, artificial intelligence algorithms can be written on this developed board, for the results of this second novelty will be presented in the next technical paper.

IV. CONCLUSION AND FUTURE WORK

The results of the SLR confirmed that The Internet of Things research on Smart greenhouses was still a trend, as seen from the increasing number of related topics from year to year in quartile 1 to 3 journals. The most widely used sensors were temperature and humidity sensors with research and development methods that were integrated with other algorithms as a decision support system. The decision obtained was used to control the actuators in controlling conditions in the greenhouse such as temperature conditions, air humidity, soil moisture, robot control, nutrient levels of both soil and water planting media.

From the publisher statistics, information was obtained that published the most related topics so that we can choose where to publish the next technical paper; this increased the probability of an accepted paper.

For future work, we will conduct research with topics obtained from two possibilities of a research gap, the first future work: the improving algorithm to obtain better prediction and error values than before. The second future work: the addition of a developed board like developed board which is connected to sensors and actuators in the greenhouse to carry out real time control and monitoring. The developed board communicates serially with the cloud-connected Nodemcu.

V. Declaration and Competing Interest

The researchers declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] U. Zangina, S. Buyamin, M. S. Z. Abidin, and M. S. A. Mahmud, "Agricultural rout planning with variable rate pesticide application in a greenhouse environment," *Alex. Eng. J.*, vol. 60, no. 3, pp. 3007–3020, Jun. 2021, doi: 10.1016/j.aej.2021.01.010.

[2] K. Liu, C. Zhang, X. Yang, M. Diao, H. Liu, and M. Li, "Development of an Occurrence Prediction Model for Cucumber Downy Mildew in Solar Greenhouses Based on Long Short-Term Memory Neural Network," *Agronomy*, vol. 12, no. 2, p. 442, Feb. 2022, doi: 10.3390/agronomy12020442.

[3] F. Lin, Y. Weng, H. Chen, and P. Zhuang, "Intelligent greenhouse system based on remote sensing images and machine learning promotes the efficiency of agricultural economic growth," *Environ. Technol. Innov.*, vol. 24, p. 101758, Nov. 2021, doi: 10.1016/j.eti.2021.101758.

[4] J. Rong, P. Wang, Q. Yang, and F. Huang, "A Field-Tested Harvesting Robot for Oyster Mushroom in Greenhouse," *Agronomy*, vol. 11, no. 6, p. 1210, Jun. 2021, doi: 10.3390/agronomy11061210.

[5] H. H. Lund, M. Exner, NE. Jensen, M. Leggieri, M. Outzen, G. Ravn-Haren, M. von Sechested, A. Vaering, R. Andersen., "GrowBot: An Educational Robotic System for Growing Food," *Appl. Sci.*, vol. 12, no. 11, p. 5539, May 2022, doi: 10.3390/app12115539.

[6] A. Selmani,H. Oubehar, M. Outanoute, A. Ed-Dahhak, M. Guerbaoui, A. Lachhab, B. Bouchikhi., "Agricultural cyber-physical system enabled for remote management of solar-powered precision irrigation," *Biosyst. Eng.*, vol. 177, pp. 18–30, Jan. 2019, doi: 10.1016/j.biosystemseng.2018.06.007.

[7] N. Bono Rossello, R. Fabrizio Carpio, A. Gasparri, and E. Garone, "A novel Observer-based Architecture for Water Management in Large-Scale (Hazelnut) Orchards," *IFAC-Pap.*, vol. 52, no. 30, pp. 62–69, 2019, doi: 10.1016/j.ifacol.2019.12.498.

[8] R. Guirado-Clavijo, J. A. Sanchez-Molina, H. Wang, and F. Bienvenido, "Conceptual Data Model for IoT in a Chain-Integrated Greenhouse Production: Case of the Tomato Production in Almeria (Spain)," *IFAC-Pap.*, vol. 51, no. 17, pp. 102–107, 2018, doi: 10.1016/j.ifacol.2018.08.069.

[9] Y. Wu, L. Li, M. Li, M. Zhang, H. Sun, N. Sygrimis, W. Lai., "Remote-Control System for Greenhouse Based on Open Source Hardware," *IFAC-Pap.*, vol. 52, no. 30, pp. 178–183, 2019, doi: 10.1016/j.ifacol.2019.12.518.

[10] D. Cafuta, I. Dodig, I. Cesar, and T. Kramberger, "Developing a Modern Greenhouse Scientific Research Facility—A Case Study," *Sensors*, vol. 21, no. 8, p. 2575, Apr. 2021, doi: 10.3390/s21082575.

[11] D. Y. Setyawan, D. Yuliawati, W. , and W. Warsono, "Calibration of Geomagnetic and Soil Temperatur Sensor for Earthquake Early Warning System," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 16, no. 5, p. 2239, Oct. 2018, doi: 10.12928/telkomnika.v16i5.7592.

[12] I. Cisternas, I. Velásquez, A. Caro, and A. Rodríguez, "Systematic literature review of implementations of precision agriculture," *Comput. Electron. Agric.*, vol. 176, p. 105626, Sep. 2020, doi: 10.1016/j.compag.2020.105626.

[13] J. Kim, D. Minagawa, D. Saito, S. Hoshina, and K. Kanda, "Development of KOSEN Weather Station and Provision of Weather Information to Farmers," *Sensors*, vol. 22, no. 6, p. 2108, Mar. 2022, doi: 10.3390/s22062108.

[14] J. Wang, J. Zhou, R. Gu, M. Chen, and P. Li, "Manage system for internet of things of greenhouse based on GWT," *Inf. Process. Agric.*, vol. 5, no. 2, pp. 269–278, Jun. 2018, doi: 10.1016/j.inpa.2018.01.002.

[15] T. P. Kharel, A. J. Ashworth, and P. R. Owens, "Linking and Sharing Technology: Partnerships for Data Innovations for Management of Agricultural Big Data," *Data*, vol. 7, no. 2, p. 12, Jan. 2022, doi: 10.3390/data7020012.

[16] J. Jamal, S.Azizi, A. Abdollahpouri, N. Ghaderi, B. Sarabi, A. Silva-Ordaz, Victor M. C.Meneses., "Monitoring rocket (Eruca sativa) growth parameters using the Internet of Things under supplemental LEDs lighting," *Sens. Bio-Sens. Res.*, vol. 34, p. 100450, Dec. 2021, doi: 10.1016/j.sbsr.2021.100450.

[17] V. Lysenko, Y. Guo, A. Kosolapov, E. Usova, T. Varduny, and V. Krasnov, "Polychromatic Fourier-PAM fluorometry and hyperspectral analysis of chlorophyll fluorescence from Phaseolus vulgaris leaves: Effects of green light," *Inf. Process. Agric.*, vol. 7, no. 2, pp. 204–211, Jun. 2020, doi: 10.1016/j.inpa.2019.09.009.

[18] S. Santiteerakul, A. Sopadang, K. Yaibuathet Tippayawong, and K. Tamvimol, "The Role of Smart Technology in Sustainable Agriculture: A Case Study of Wangree Plant Factory," *Sustainability*, vol. 12, no. 11, p. 4640, Jun. 2020, doi: 10.3390/su12114640.

[19] M. Muñoz, J. L. Guzmán, J. A. Sánchez, F. Rodríguez, and M. Torres, "Greenhouse Models as a Service (GMaaS) for Simulation and Control," *IFAC-Pap.*, vol. 52, no. 30, pp. 190–195, 2019, doi: 10.1016/j.ifacol.2019.12.520.

[20] G.-A. Musat, M. Colezea, F. Pop, C. Negru, M. Mocanu, C. Esposito, A. Castiglione., "Advanced services for efficient management of smart farms," *J. Parallel Distrib. Comput.*, vol. 116, pp. 3–17, Jun. 2018, doi: 10.1016/j.jpdc.2017.10.017.

[21] W. Yong, L. Shuaishuai, L. Li, L. Minzan, L. Ming, K.G. Arvanitis, Cs. Georgieva, N. Sigrimis., "Smart Sensors from Ground to Cloud and Web Intelligence," *IFAC-Pap.*, vol. 51, no. 17, pp. 31–38, 2018, doi: 10.1016/j.ifacol.2018.08.057.

[22] M. Munoz, J. L. Guzman, J. A. Sanchez-Molina, F. Rodriguez, M. Torres, and M. Berenguel, "A New IoT-Based Platform for Greenhouse Crop Production," *IEEE Internet Things J.*, vol. 9, no. 9, pp. 6325–6334, May 2022, doi: 10.1109/JIOT.2020.2996081.

[23] H. A. Méndez-Guzmán,J. A. Padilla-Medina, C. Martínez-Nolasco, J. J. Martinez-Nolasco, A. I. Barranco-Gutiérrez, L. M. Contreras-Medina,M. Leon-Rodriguez., "IoT-Based Monitoring System Applied to Aeroponics Greenhouse," *Sensors*, vol. 22, no. 15, p. 5646, Jul. 2022, doi: 10.3390/s22155646.

[24] S. Zhang, Y. Guo, S. Li, Z. Ke, H. Zhao, J. Yang, Y. Wang, D. Li, L. Wang, W. Yang, Z. Zhang., "Investigation on environment monitoring system for a combination of hydroponics and aquaculture in greenhouse," *Inf. Process. Agric.*, vol. 9, no. 1, pp. 123–134, Mar. 2022, doi: 10.1016/j.inpa.2021.06.006.

[25] P. Placidi, R. Morbidelli, D. Fortunati, N. Papini, F. Gobbi, and A. Scorzoni, "Monitoring Soil and Ambient Parameters in the IoT Precision Agriculture Scenario: An Original Modeling Approach Dedicated to Low-Cost Soil Water Content Sensors," *Sensors*, vol. 21, no. 15, p. 5110, Jul. 2021, doi: 10.3390/s21155110.

[26] R. Singh, M. Aernouts, M. De Meyer, M. Weyn, and R. Berkvens, "Leveraging LoRaWAN Technology for Precision Agriculture in Greenhouses," *Sensors*, vol. 20, no. 7, p. 1827, Mar. 2020, doi: 10.3390/s20071827.

[27] A. Pena, J. C. Tejada, J. D. Gonzalez-Ruiz, and M. Gongora, "Deep Learning to Improve the Sustainability of Agricultural Crops Affected by Phytosanitary Events: A Financial-Risk Approach," *Sustainability*, vol. 14, no. 11, p. 6668, May 2022, doi: 10.3390/su14116668.

[28] S. Venkatesan, J. Lim, H. Ko, and Y. Cho, "A Machine Learning Based Model for Energy Usage Peak Prediction in Smart Farms," *Electronics*, vol. 11, no. 2, p. 218, Jan. 2022, doi: 10.3390/electronics11020218.

[29] I.-S. Choi, A. Hussain, V.-H. Bui, and H.-M. Kim, "A Multi-Agent System-Based Approach for Optimal Operation of Building Microgrids with Rooftop Greenhouse," *Energies*, vol. 11, no. 7, p. 1876, Jul. 2018, doi: 10.3390/en11071876.

[30] S. Afzali, S. Mosharafian, M. W. van Iersel, and J. Mohammadpour Velni, "Development and Implementation of an IoT-Enabled Optimal and Predictive Lighting Control Strategy in Greenhouses," *Plants*, vol. 10, no. 12, p. 2652, Dec. 2021, doi: 10.3390/plants10122652.

[31] J.-W. Lee, T. Moon, and J.-E. Son, “Development of Growth Estimation Algorithms for Hydroponic Bell Peppers Using Recurrent Neural Networks,” *Horticulturae*, vol. 7, no. 9, p. 284, Sep. 2021, doi: 10.3390/horticulturae7090284.

[32] M. Erazo-Rodas, M. Sandoval-Moreno, S. Muñoz-Romero, M. Huerta, D. Rivas-Lalaleo, and J. Rojo-Álvarez, “Multiparametric Monitoring in Equatorian Tomato Greenhouses (III): Environmental Measurement Dynamics,” *Sensors*, vol. 18, no. 8, p. 2557, Aug. 2018, doi: 10.3390/s18082557.

[33] A. Rennane, F. Benmahmoud, A. Tayeb Cherif, R. Touhami, and S. Tedjini, “Design of autonomous multi-sensing passive UHF RFID tag for greenhouse monitoring,” *Sens. Actuators Phys.*, vol. 331, p. 112922, Nov. 2021, doi: 10.1016/j.sna.2021.112922.

[34] D. Cama-Pinto, J. A. Holgado-Terriza, M. Damas-Hermoso, F. Gómez-Mula, and A. Cama-Pinto, “Radio Wave Attenuation Measurement System Based on RSSI for Precision Agriculture: Application to Tomato Greenhouses,” *Inventions*, vol. 6, no. 4, p. 66, Oct. 2021, doi: 10.3390/inventions6040066.

[35] J. Wang, M. Chen, J. Zhou, and P. Li, “Data communication mechanism for greenhouse environment monitoring and control: An agent-based IoT system,” *Inf. Process. Agric.*, vol. 7, no. 3, pp. 444–455, Sep. 2020, doi: 10.1016/j.inpa.2019.11.002.

[36] V. Thomopoulos, D. Bitas, K.-N. Papastavros, D. Tsipianitis, and A. Kavga, “Development of an Integrated IoT-Based Greenhouse Control Three-Device Robotic System,” *Agronomy*, vol. 11, no. 2, p. 405, Feb. 2021, doi: 10.3390/agronomy11020405.

[37] U. Lee, M. P. Islam, N. Kochi, K. Tokuda, Y. Nakano, H. Naito, Y. Kawasaki, T. Ota, T. Sugiyama, D.H. Ahn., “An Automated, Clip-Type, Small Internet of Things Camera-Based Tomato Flower and Fruit Monitoring and Harvest Prediction System,” *Sensors*, vol. 22, no. 7, p. 2456, Mar. 2022, doi: 10.3390/s22072456.

[38] G. A. Pauzi, A. Supriyanto, B. L. Putra, and S. Suciyati, “ANALYSIS AND REDUCTION OF TIME SHIFT BETWEEN SENDING AND RECEIVING ON DATA ACQUISITION USING SMS GATEWAY SYSTEM,” . *Vol.*, no. 20, p. 8, 2005

[39] Warsito, G. A. Pauzi, S. W. Suciyati, and Turyani, “Design and characterization of water level detector using MW22B Multi-Turn potentiometer,” Bandung, Indonesia, 2012, pp. 174–177. doi: 10.1063/1.4730714.

[40] G. A. Pauzi, S. W. Suciyati, and K. Imaniar, “STUDY OF INCLINATION ANGLE OF REFLECTOR OBJECT IN SIMPLE WATER LEVEL INSTRUMENT USING 40 KHz ULTRASONIC TRANSDUCER,” . *Vol.*, vol. 42, p. 5, 2005.

[41] Y. Mu, T.-S. Chen, S. Ninomiya, and W. Guo, “Intact Detection of Highly Occluded Immature Tomatoes on Plants Using Deep Learning Techniques,” *Sensors*, vol. 20, no. 10, p. 2984, May 2020, doi: 10.3390/s20102984.

[42] Y.-Y. Zheng, J.-L. Kong, X.-B. Jin, X.-Y. Wang, and M. Zuo, “CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture,” *Sensors*, vol. 19, no. 5, p. 1058, Mar. 2019, doi: 10.3390/s19051058.

[43] F. García-Mañas, F. Rodríguez, and M. Berenguel, “Leaf area index soft sensor for tomato crops in greenhouses,” *IFAC-Pap.*, vol. 53, no. 2, pp. 15796–15803, 2020, doi: 10.1016/j.ifacol.2020.12.230.

[44] A. Castañeda-Miranda and V. M. Castaño-Meneses, “Smart frost measurement for anti-disaster intelligent control in greenhouses via embedding IoT and hybrid AI methods,” *Measurement*, vol. 164, p. 108043, Nov. 2020, doi: 10.1016/j.measurement.2020.108043.

[45] I. Ullah, M. Fayaz, N. Naveed, and D. Kim, “ANN Based Learning to Kalman Filter Algorithm for Indoor Environment Prediction in Smart Greenhouse,” *IEEE Access*, vol. 8, pp. 159371–159388, 2020, doi: 10.1109/ACCESS.2020.3016277.

[46] G. Codeluppi, L. Davoli, and G. Ferrari, “Forecasting Air Temperature on Edge Devices with Embedded AI,” *Sensors*, vol. 21, no. 12, p. 3973, Jun. 2021, doi: 10.3390/s21123973.

[47] E. A. Abioye, M. S. Z. Abidin, M. S. A. Mahmud, S. Buyamin, M. K. I. AbdRahman, A. O. Otuoze, M. S. A. Ramli, O. D. Ijike., “IoT-based monitoring and data-driven modelling of drip irrigation system for mustard leaf cultivation experiment,” *Inf. Process. Agric.*, vol. 8, no. 2, pp. 270–283, Jun. 2021, doi: 10.1016/j.inpa.2020.05.004.

[48] S.-C. Huang and Y.-Z. Lin, “A Low-Cost Constant-Moisture Automatic Irrigation System Using Dynamic Irrigation Interval Adjustment,” *Appl. Sci.*, vol. 10, no. 18, p. 6352, Sep. 2020, doi: 10.3390/app10186352.

[49] Y. Y. Hilal, M. K. Khessro, J. van Dam, and K. Mahdi, “Automatic Water Control System and Environment Sensors in a Greenhouse,” *Water*, vol. 14, no. 7, p. 1166, Apr. 2022, doi: 10.3390/w14071166.

[50] A. M. Okasha, H. G. Ibrahim, A. H. Elmetwalli, K. M. Khedher, Z. M. Yaseen, and S. Elsayed, “Designing Low-Cost Capacitive-Based Soil Moisture Sensor and Smart Monitoring Unit Operated by Solar Cells for Greenhouse Irrigation Management,” *Sensors*, vol. 21, no. 16, p. 5387, Aug. 2021, doi: 10.3390/s21165387.

[51] P. Placidi, L. Gasperini, A. Grassi, M. Cecconi, and A. Scorzoni, “Characterization of Low-Cost Capacitive Soil Moisture Sensors for IoT Networks,” *Sensors*, vol. 20, no. 12, p. 3585, Jun. 2020, doi: 10.3390/s20123585.

[52] P. Suebsombut, A. Sekhari, P. Sureephong, A. Belhi, and A. Bouras, “Field Data Forecasting Using LSTM and Bi-LSTM Approaches,” *Appl. Sci.*, vol. 11, no. 24, p. 11820, Dec. 2021, doi: 10.3390/app112411820.

[53] S. An,F. Yang,Y. Yang,Y. Huang,L. Zhangzhong, X. Wei, J. Yu., “Water Demand Pattern and Irrigation Decision-Making Support Model for Drip-Irrigated Tomato Crop in a Solar Greenhouse,” *Agronomy*, vol. 12, no. 7, p. 1668, Jul. 2022, doi: 10.3390/agronomy12071668.

[54] D. Massa, J. J. Magán, F. F. Montesano, and N. Tzortzakis, “Minimizing water and nutrient losses from soilless cropping in southern Europe,” *Agric. Water Manag.*, vol. 241, p. 106395, Nov. 2020, doi: 10.1016/j.agwat.2020.106395.

[55] C. M. Angelopoulos, G. Filios, S. Nikoletseas, and T. P. Raptis, “Keeping data at the edge of smart irrigation networks: A case study in strawberry greenhouses,” *Comput. Netw.*, vol. 167, p. 107039, Feb. 2020, doi: 10.1016/j.comnet.2019.107039.

[56] K. Tatas, A. Al-Zoubi, N. Christofides, C. Zannettis, M. Chrysostomou, S. Panteli, A. Antoniou., “Reliable IoT-Based Monitoring and Control of Hydroponic Systems,” *Technologies*, vol. 10, no. 1, p. 26, Feb. 2022, doi: 0.3390/technologies10010026.

[57] R. Urbieta Parrazales, M. T. Zagaceta Álvarez, K. A. Aguilar Cruz, R. Palma Orozco, and J. L. Fernández Muñoz, “Implementation of a Fuzzy Logic Controller for the Irrigation of Rose Cultivation in Mexico,” *Agriculture*, vol. 11, no. 7, p. 576, Jun. 2021, doi: 10.3390/agriculture11070576.

[58] N. K. Nawandar and V. R. Satpute, “IoT based low cost and intelligent module for smart irrigation system,” *Comput. Electron. Agric.*, vol. 162, pp. 979–990, Jul. 2019, doi: 10.1016/j.compag.2019.05.027.

[59] H. Benyezza, M. Bouhedda, and S. Rebouh, “Zoning irrigation smart system based on fuzzy control technology and IoT for water and energy saving,” *J. Clean. Prod.*, vol. 302, p. 127001, Jun. 2021, doi: 10.1016/j.jclepro.2021.127001.

[60] J. M. Pearce, “Parametric Open Source Cold-Frame Agrivoltaic Systems,” p. 20, 2021.

[61] T. Zhang, W. Zhou, F. Meng, and Z. Li, “Efficiency Analysis and Improvement of an Intelligent Transportation System for the Application in Greenhouse,” *Electronics*, vol. 8, no. 9, p. 946, Aug. 2019, doi: 10.3390/electronics8090946.

[62] S. Krul, C. Pantos, M. Frangulea, and J. Valente, “Visual SLAM for Indoor Livestock and Farming Using a Small Drone with a Monocular Camera: A Feasibility Study,” *Drones*, vol. 5, no. 2, p. 41, May 2021, doi: 10.3390/drones5020041.

[63] W. Cai, R. Wei, L. Xu, and X. Ding, “A method for modelling greenhouse temperature using gradient boost decision tree,” *Inf. Process. Agric.*, vol. 9, no. 3, pp. 343–354, Sep. 2022, doi: 10.1016/j.inpa.2021.08.004.

An Adaptive Texture Enhancement Algorithm for AR Live Screen Based on Approximate Matching

Panpan Yang, Lingfei Ma, Chao Yin*, Yang Ping
Institute of Data Science and Agricultural Economy
Beijing Academy of Agricultural and Forestry Sciences, Beijing, 100097, China

***Abstract*—In order to improve the visual effect of AR live video detail texture, an adaptive enhancement algorithm based on approximate matching is proposed. According to the local self-similarity of the original image, the best matching block of the initial super-resolution image block is obtained. After extracting its high-frequency information, the improved singular value decomposition (SVD) is used to embed the watermark into the original super-resolution gray image; And through the watermark extraction and reconstruction matrix, the effective high-frequency detail texture information is obtained, then the final super-resolution image is obtained, and the AR live video detail texture enhancement is completed. The experimental results show that the peak signal-to-noise ratio, structure similarity and feature similarity of the algorithm are high, and the reconstruction effect is good; After detail texture enhancement, the detail texture clarity and edge sharpness of AR live picture are better, and the best visual effect is achieved.**

***Keywords*—Approximate matching; AR live; picture details; adaptive; texture enhancement; matrix reconstruction**

I. Introduction

AR, also known as augmented reality, is a fusion technology used between the real and virtual worlds. It can realize the real environment and virtual objects overlap, so that both appear in the same screen. It has 5 core technologies, namely, tracking registration, display, virtual object generation, interaction and merging [1]. AR live broadcast is the product of this technology, which uses cameras and sensors to collect information and data in real environment. After the input processor is analyzed and reconstructed, it is returned to the display terminal such as the camera, And AR live terminal screens can have multiple special effects [2]. AR live images usually contain three categories, such as flower, car screen type images, human-dominated portrait images and scene images that can sense the environment. When these pictures are presented through a display terminal such as a camera, the details and texture of the live pictures will be affected by the network, device itself and other problems [3]. The visual homogeneity in the picture is the detailed texture of the picture, and it also represents a feature of the picture. Texture can give the picture a depth or can be understood as an object of visual presentation. Approximate matching refers to the image which has the highest similarity degree and can be matched with the original image by obtaining the most similar image, and can retain the original image features.

In order to ensure the visual effect of the detailed texture of AR live images, the algorithms based on guiding coefficient weighting, NSST multi-scale and deep learning are proposed respectively in [4], [5] and [6] to improve the visual effect of images. Guide coefficient weighting algorithm in the enhancement process. Mainly through the brightness of the image and non-brightness region to complete the division of the screen processing. NSST multi-scale algorithm is the implementation of low-light picture conversion to complete the screen processing. Deep learning algorithm adopts the strategy of mixing deep learning and image fusion to improve the image quality. In the process of processing, the three algorithms do not extract and reconstruct the detailed texture features, so the clarity of the enhanced texture is poor. Based on this, this paper uses approximate matching algorithm to enhance the detail texture and improve the clarity of detail texture.

II. Approximate Matching Texture Adaptive Enhancement Algorithm

A. Approximate Matching of AR Live Broadcast Screen Detail Texture High Frequency Component Extraction

The pictures in the local neighborhood have similar picture image blocks, which is an obvious feature of the local self-similarity of AR live broadcast pictures. And there are the same scale or different scales in the same picture. Correlation analysis and research are carried out on this feature, and the high frequency detail of similar blocks are extracted and analyzed. The detailed texture of the picture is processed by synthetic method. High frequency detail is obtained from high resolution images, and super- resolution reconstruction is based on local self-similarity. The original live picture I, its corresponding high frequency live picture HF and its corresponding interpolation magnified live picture LR are used as the training set to generate the super- resolution AR live picture HR.

The initial high-resolution image LR is obtained by the contour template interpolation algorithm, and I is amplified after the implementation of processing, and the magnification is α. n order to avoid the missing of the image detail texture, it is necessary to determine the window in LR, which is represented by small window image block e. And formed in the case of pixel (i, j) as the center; If a similar block is retrieved at the same time, belongs to a large window of q, and is in I's position for (x, y), then:

*Corresponding Author.

$$x = i/\alpha, y = j/\alpha \tag{1}$$

In formula: The scaling factor is expressed as α. Through matching window e in the search window f near the scope of similar block search, access to similar blocks n, is $e_1, e_2, ..., e_n$. Screen e block can be solved by weighted average of similar blocks. The solution formula is as follows:

$$\hat{e} = \sum_k w(d) e(d) \tag{2}$$

Expression: the d similar block and weight are described by $e(d)$ and $w(d)$. And the weights meet the following conditions:

$$\sum_d w(d) = 1, 0 \le w(d) \le 1 \tag{3}$$

$w(d)$ can be described by equation (4):

$$w(d) = \frac{1}{Z} \exp\left(-\frac{l(d)}{h^2}\right) \tag{4}$$

In the formula: the pixel intensity of the picture block is represented by h^2. The Euclidean distance and normalization constant of picture blocks e and e_k are $l(d)$ and Z respectively.

$$l(d) = \|e - e_k\|_G^2 \tag{5}$$

In the formula: the controllable coefficient of the picture block is represented by k. The grayscale of the high-frequency picture block is denoted by G. On this basis, Z is solved by formula (6):

$$Z = \sum_k \exp\left(-\frac{l(d)}{h^2}\right) \tag{6}$$

After the self-similarity super-resolution reconstruction of the AR live broadcast screen is completed by the above method, it will ensure the integrity of the detailed texture features and implement the self-similarity reconstruction.

In order to obtain the most similar matching block R in the vicinity of the retrieval window f. After the matching search is completed through e, the high-frequency picture block R_{mn} corresponding to R is obtained, and belongs to HF. The similarity indicator is described by the mean of the sum of absolute values of the difference between two picture blocks. Therefore, the total absolute value of the difference between the two picture blocks is the matching block with the highest degree of similarity [7], then:

$$R = \arg\min |e - f| \tag{7}$$

The calculation formula to obtain the high-frequency picture block of R_{mn} is:

$$R_{mn} = R - R^* C_0 \tag{8}$$

Where the guided filter and convolution are described by C_0 and $*$, respectively. To obtain the sum of $\hat{e}$ and R_{mn}, it is done by formulas (2) and (8). Its calculation formula is:

$$e = \hat{e} + R_{mn} \tag{9}$$

B. *Improved Singular Value Decomposition Detail Texture Enhancement*

The high-frequency picture block R_{mn} obtained above is processed as follows with the improved singular value decomposition:

1) Watermark embedding: Let $W_{pq}(p = q)$ denote the matrix with embedded watermark, and p denote the dimension of the matrix.

a) After the high-frequency picture block R_{mn} is decomposed by the first-order wavelet transform, a new matrix B can be obtained, and it is composed of low-frequency coefficients [8].

b) Use 16×16 blocks to perform processing on B, then $B_{ij} = f_{ij}(x, y)$ represents each block after processing, and $i = 1, 2, ..., m$, $j = 1, 2, ..., n$, $x \ge 1$, $16 \ge y$. The decomposition formula of each block matrix after processing is:

$$B_{ij} = U_{ij} S_{ij} V_{ij}^T \tag{10}$$

Where: U and V are orthogonal matrices.

c) Suppose the largest singular value of S_{ij} is denoted by φ_{ij}. According to the properties of singular value decomposition, φ_{ij} is adjusted, and the adjustment formula is:

$$if\left(W_{ij}=1\right)\begin{cases} if\left(\mathrm{mod}\left(\varphi_{ij},a\right)<a/4\right)\varphi_{ij}= \\ \varphi_{ij}-\left(\mathrm{mod}\left(\varphi_{ij},a\right)-a/4\right); \\ else\varphi_{ij}=\varphi_{ij}-\left(\mathrm{mod}\left(\varphi_{ij},a\right)+3^{*}a/4\right) \end{cases}$$

$$else\begin{cases} if\left(\mathrm{mod}\left(\varphi_{ij},a\right)>3^{*}a/4\right)\varphi_{ij}= \\ \varphi_{ij}-\left(\mathrm{mod}\left(\varphi_{ij},a\right)+5^{*}a/4\right); \\ else\varphi_{ij}=\varphi_{ij}-\left(\mathrm{mod}\left(\varphi_{ij},a\right)+a/4\right) \end{cases} \quad (11)$$

In the formula: $\mathrm{mod}(x,y)$ is used to represent $x \bmod y$. The watermark signal values corresponding to the embedded strength factor and S_{ij} are a and W_{ij}, respectively. The adjusted S_{ij} is represented by $\widehat{S}_{ij}$, and singular value decomposition is performed on all $\widehat{S}_{ij}$, so that each block contains watermark information [9].

d) According to the ordinal position, all the blocks containing the watermark are recombined to form a new matrix R' of low frequency coefficients. Reconstruction method is used to process the changed wavelet coefficients [10] to obtain the super-resolution grayscale image I' with the initial watermark embedded.

2) Watermark extraction

a) I' is processed by first-order wavelet decomposition to obtain a matrix M, which consists of decomposed low-frequency coefficients.

b) Use 16×16 blocks to process M, then $M_{ij}=q_{ij}(x,y)$ represents each block after processing, and $i=1,2,...,m$, $j=1,2,...,n$, $x\geq 1$, $16\geq y$.

$$M_{ij}=U'_{ij}S'_{ij}V'^{T}_{ij} \quad (12)$$

c) If the largest singular value of S'_{ij} is represented by σ'_{ij}, and adjust it, its formula is:

$$iq\left(\mathrm{mod}\left(\varphi'_{ij},s\right)>s/2\right)\left\{W_{ij}=1\right\}else\left\{W_{ij}=0\right\} \quad (13)$$

The above steps are taken, and each block is processed to obtain the restored watermark binary picture B_i.

Referring to the overlapping characteristics of the blocks, it is impossible to directly add the B_i obtained by reconstruction to M_{ij}, and a control strategy should be adopted for the overlap to avoid the interference caused by it [11].Therefore, the window processing is performed on the picture block, which is completed by the Gaussian function. And it is centrally symmetric, and $\bar{R}_{mns}$ is obtained after processing, and its formula is:

$$\bar{R}_{mns}=\bar{R}_{mn}G_{\varphi}\left(\left\|x-x_c\right\|\right) \quad (14)$$

Where $G_{\varphi}\left(\left\|x-x_c\right\|\right)=\frac{1}{2\pi\varphi^2}u^{-\frac{\left\|x-x_c\right\|^2}{2\varphi^2}}$ is the Gaussian window function.

The acquired $\bar{R}_{mns}$ and M_{ij} are combined to form a super-resolution picture M^{i}_{mn}, which contains all the detailed texture features of the high-resolution picture [12], which is:

$$M^{i}_{mn}=M_{ij}+\bar{R}_{mns} \quad (15)$$

C. Algorithm Flow

The implementation of adaptive enhancement algorithm based on approximate matching for AR live-broadcast image details texture is divided into two parts. One is to get the high frequency information of the best matching block based on approximate matching. Secondly, the improved singular value decomposition is used to reconstruct the matrix to obtain the effective high frequency texture information. The detailed texture is synthesized to obtain the enhanced ultra-detailed texture [13-15]. Its enhanced overall flow is shown in Fig. 1.

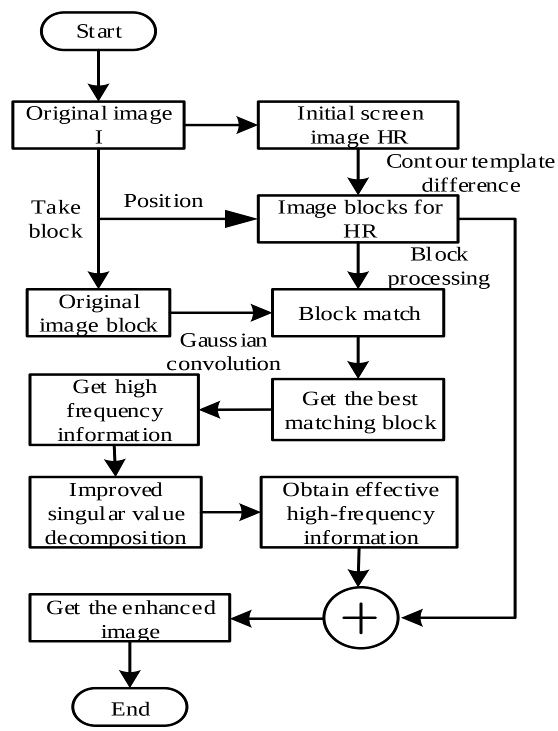


Fig. 1. Algorithmic flow.

III. Test Analysis

In order to test this algorithm for AR live picture details texture enhancement effect. Five live images are randomly selected from the image database of an AR live company in order to verify the effectiveness of the algorithm. The test environment operating system was Windows 10 x 64 8 GB of memory with CPU 3.4 GHz. The application development environment is Matlab 2018a.

In order to test the influence of scaling factor values on approximate matching, the sizes of matching blocks and retrieval windows are set to 2 × 2 and 4 × 4 respectively, and the gray space values (0 ~ 255) are used as features to obtain the normalized mean of the total absolute values of the difference between the two blocks. The smaller the value, the higher the similarity of the matching block. Test the results of the different zoom factor values, as shown in Fig. 2.

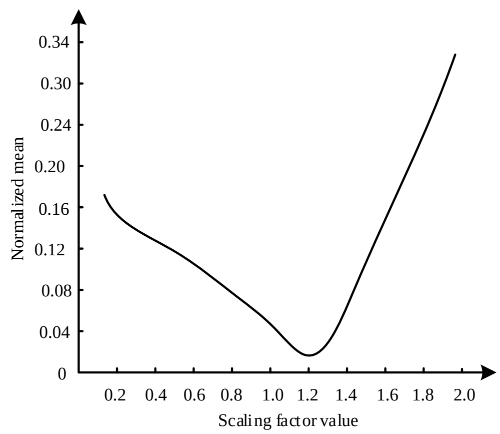


Fig. 2. Normalized mean change results.

According to the test results in Fig. 2, it can be seen that with the increase of the scaling factor value, the normalized mean value first decreases and then increases. When the scaling factor is 1.2, the normalized mean is the smallest. So in this article, determine that the zoom factor has a value of 1.2. In order to more intuitively reflect the impact of the algorithm with the scaling factor changes in image quality. We plotted the changes in the FSIM (feature similarity) metrics to show the picture quality, as shown in Fig. 3.

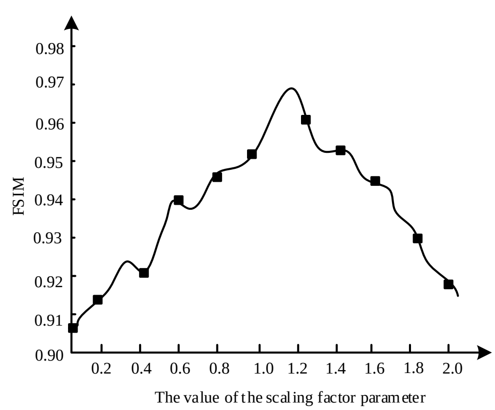


Fig. 3. Influence of Zoom Factor Parameter Value on Screen.

As shown in Fig. 3, the FSIM value is highest when the scale factor parameter is set to a value of 1.2. When the scaling factor parameter value reaches 0.97, the picture quality is the best. Therefore, the best scaling factor for this article is 1.2.

The quality of reconstructed image has some influence on the effect of texture enhancement. In order to measure the quality of the reconstructed images, peak signal-to-noise ratio (PSNR), structural similarity (MSSIM) and feature similarity (FSIM) are used as measurement criteria. The higher the value is, the better the quality is. The calculation formulas are as follows:

$$PSNR = 10\lg\left(\frac{255^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X_1(i,j) - X_2(i,j)^2\right)}\right) \tag{16}$$

$$MSSIM = \frac{1}{W}\sum_{i=1}^{W}(b_i)^{\alpha}(c_i)^{\rho}(h_i)^{\gamma} \tag{17}$$

$$FSIM = \frac{\sum S_L(X_1, X_2)G(X_1, X_2)}{\sum G(X_1, X_2)} \tag{18}$$

Type: original picture and reconstruction picture for X_1 and X_2. The brightness, contrast and structure comparison functions of L block are b_i, c_i and h_i. The corresponding adjustment weights are α, ρ and γ. The number of blocks is W. The phase congruence is described by $S_L(X_1, X_2)$. The similarity measure of gradient features is described by $G(X_1, X_2)$.

In the process of decomposition, the value of the threshold has a great influence on the effect of image details texture enhancement. We choose PSNR and MSSIM as the metrics to get the effect of enhancing the image detail texture under different threshold values. The results are shown in Fig. 4 and Fig. 5.

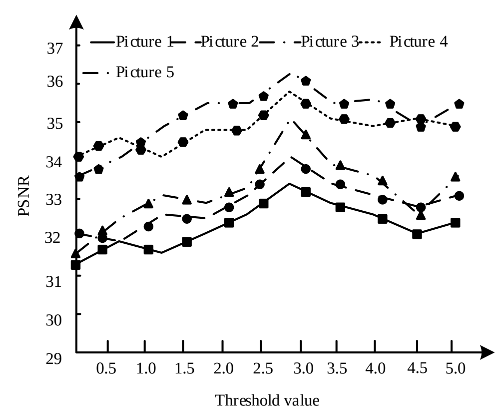


Fig. 4. Picture enhancement PSNR results with different thresholds

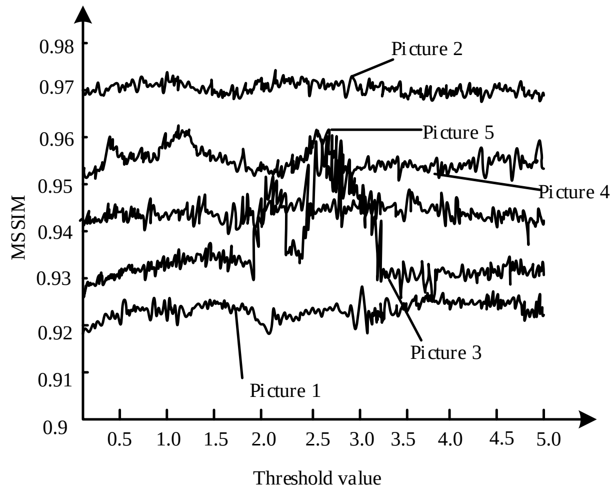


Fig. 5. Picture enhancement MSSIM results with different thresholds

According to Figs. 4 and 5, the PSNR and MSSIM results of the five live images are changed under different thresholds. When the threshold value is 3.0, the PSNR and MSSIM values of all live images after enhancement are the highest, indicating that the effect is best when the threshold value is 3.0. So the threshold value is 3.0 in the following experiment.

The enhanced algorithm based on the weight of the guide coefficient (Document [4] algorithm), the enhanced algorithm based on the multi-scale of NSST (Document [5] algorithm) and the enhanced algorithm based on the deep learning algorithm (Document [6] algorithm) are selected as the comparison algorithm. Through the above formula to obtain the effect of three kinds of algorithm picture reconstruction, as shown in Table I.

According to the test results in Table I, the results of PSNR, MSSIM and FSIM show that the three values of this algorithm are the best, significantly better than the three contrast algorithms. Therefore, it is shown that the reconstruction features of the proposed algorithm are closer to those of the original image and the reconstruction effect is better.

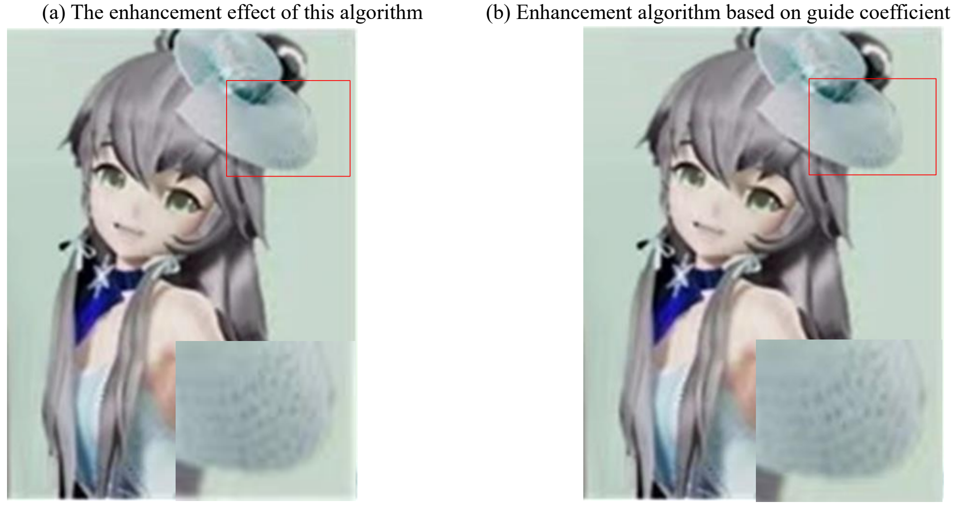
In order to test the subjective effect of the enhanced method, the detailed texture visual effect image of the enhanced image is obtained by three methods. Due to space constraints, only Screen 3 was selected for the test, as shown in Fig. 6.

According to the test results in Fig. 6, it can be seen that after the enhancement, the clarity and edge sharpness of the texture details are better than those of the three contrast algorithms. Experimental results show that the proposed algorithm is better than three contrast methods.

TABLE I. RECONSTRUCTION RESULTS OF THREE ALGORITHMS

Test content	Image number	Algorithm of this paper	Bootstrap weighted enhancement algorithm	Multiscale enhancement algorithm based on NSST	Enhancement algorithm based on deep learning
PSNR/dB	1	28	23	24	23
	2	29	25	26	24
	3	34	30	31	29
	4	32	29	27	28
	5	26	22	23	23
MSSIM	1	0.94	0.89	0.91	0.91
	2	0.91	0.86	0.85	0.87
	3	0.95	0.91	0.90	0.90
	4	0.93	0.87	0.88	0.86
	5	0.93	0.90	0.87	0.89
FSIM	1	0.93	0.89	0.87	0.88
	2	0.92	0.88	0.89	0.88
	3	0.98	0.93	0.92	0.92
	4	0.97	0.91	0.93	0.92
	5	0.94	0.89	0.90	0.91

(a) The enhancement effect of this algorithm (b) Enhancement algorithm based on guide coefficient weighting



(c) NSST based multi-scale enhancement algorithm (d) Enhancement algorithm based on deep Learning

Fig. 6. The visual comparison effect of the three methods.

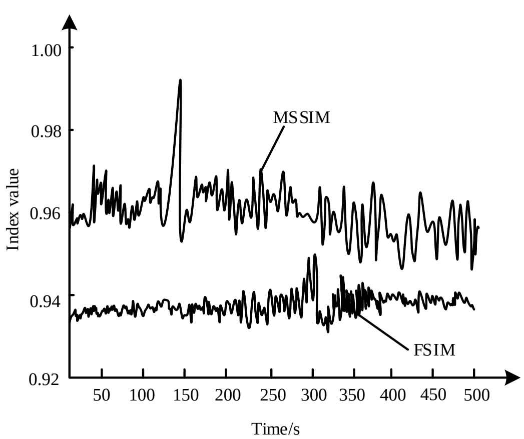


Fig. 7. Changes in MSSIM and FSIM indicators.

The algorithm designed in this paper is to improve the quality of AR live images, but the live images change with time. Therefore, it is necessary to verify the quality of the algorithm. With time as a variable, PSNR, MSSIM, and FSIM as metrics, the test results are shown in Fig. 7 and Fig. 8.

According to the changes of the three indexes in Fig. 7 and Fig. 8, the indexes of MSSIM and FSIM are more than 0.93 and PSNR is more than 26dB, and the indexes are higher. Therefore, the algorithm presented in this paper has good continuity effect and meets the requirements of AR live broadcast.

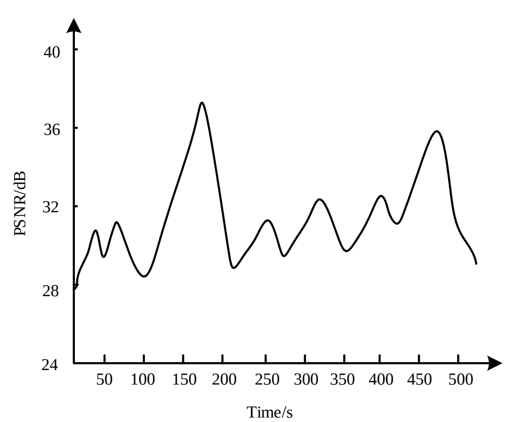


Fig. 8. Changes in PSNR indicators.

IV. Conclusion

AR live broadcast is the product of the rapid development of the Internet. It has been used in many fields, such as education, e-commerce and so on. This paper uses approximate matching algorithm to enhance the image texture details in order to improve the visual effect of AR live images. Self-similarity high frequency matching block is used to obtain and improve the singular value decomposition algorithm to reconstruct the image details and enhance the image texture. Experimental results show that the algorithm has good enhancement effect, and the sharpness and clarity of texture details are improved.

Acknowledgment

The study was supported by Scientific research project of the first-class Municipal Science and Technology Commission: Youth Fund of Beijing Academy of agricultural and Forestry Sciences, "recognition and production state determination of greenhouse vegetables based on image recognition" (qnjj202213).

References

[1] Cao S, Huang L P, Hou B B, et al. Self-Adaptive Non-Local Means Image Denoising Algorithm Based on Fuzzy Edge Complement[J]. Laser & Optoelectronics Progress,2018,55(01):213-218.

[2] Huang Y, Niu B N, Guan H, et al. Image texture based adaptive watermarking algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics,2019,45(12):2403-2414.

[3] Wu Y F, Zhu S C. Image Inpainting Algorithm Based on Matching Adjustment Rule and Gradient Constraint Model[J]. Packaging Engineering,2018,39(13):239-244.

[4] He L F, Zhou G B, Yao B, et al. A haze removal algorithm based on guided coefficient weighted and adaptive image enhancement method[J]. Microelectronics & Computer,2020,37(09):73-77+82.

[5] Wang X H, Zhai Y X, Ma X C. Retinex Low Light Image Enhancement Algorithm Based on Multi-Scale Adaptive NSST[J]. Packaging Engineering,2020,41(03):211-217.

[6] Xu S P, Lin Z Y, Zhang G Z, et al. A Low-Light Image Enhancement Algorithm Using the Hybrid Strategy of Deep Learning and I mage Fusion[J]. Acta Electronica Sinica,2021,49(1):72-76.

[7] Long X, He G T. Image enhancement method based on multi-layer fusion and detail recovery[J]. Application Research of Computers,2020,37(02):584-587.

[8] Lv B K, Wu C M, Tian X P. Fast Image Matching Algorithm Based on Best-Buddies Similarity[J]. Laser & Optoelectronics Progress,2020,57(10):189-195.

[9] Wang X C, Wang L L, Mo B, et al. Image Restoration Algorithm for Rotary Motion Blur Based on Adaptive Gradient Priori Regularization[J]. Acta Armamentarii,2019,40(03):522-529.

[10] Liu C A, Ai Z, Zhao L J. Self-adaptive image feature matching algorithm based on grid motion statistics[J]. Journal of Huazhong University of Science and Technology(Nature Science Edition),2020,48(01):37-40+54.

[11] Zhang C F, Shao L P, Shi J. Improved Criminisi Algorithm Based on Adaptive Gradient Classification and Matching[J]. Journal of Chinese Computer Systems,2018,39(02):379-385.

[12] Qin Z K, Zhang H Y. Adaptive texture image retrieval based on transform domain[J]. Chinese Journal of Liquid Crystals and Displays,2020,35(03):281-286.

[13] Wang H, Zhang S W, Xu K. A Real-Time Adaptive RANSAC Algorithm Applied in Image Matching[J]. Electronics Optics & Control,2020,27(02):90-93+97.

[14] Yin C, Guo X J, Tian X, et al. Contrast and detail enhancement based on structure and texture[J]. Application Research of Computers, 2019, 35(12):318-321.

[15] Yang H F, Li H, Chen K Y, et al. Image feature points extraction and matching method based on improved ORB algorithm[J]. Journal of Graphics,2020,41(04):548-555.

Deep Primary and Secondary Fusion Transformer Based on Internet of Things Technology

Xiaohua Zhang[1]*, Yuping Wu[2], Jianjun Chen[3], Jie Dong[4], Yu Yue[5]
State Grid Jibei Electric Power Company Limited, Beijing, China[1, 2]
State Grid Qinhuangdao Electric Power Supply Company, Qinhuangdao, Hebei, China[3, 5]
State Grid Jibei Electric Power Company Limited, Smart Distribution Network Center, Beijing, China[4]

***Abstract*—As one of the core equipment of power system, power transformer has great fault influence and complex fault causes. In order to ensure the safe and stable operation of power system, its operation state must be monitored and judged. With the increasing maturity of the Internet industry and the continuous development of sensor technology, the emergence of smart grid has contributed to the realization of intelligent transformer on-line monitoring, condition evaluation and fault diagnosis. This paper studies the deep primary and secondary fusion transformer based on the Internet of Things technology, summarizes the development status of intelligent transformer and its existing problems on the basis of relevant literature, and proposes the optimization technology of deep primary and secondary fusion transformer based on the Internet of Things technology to solve its existing problems, and conducts related experiments and fault detection research on the proposed technology. The experimental results show that this optimization technology has good feasibility, good self-state evaluation and fault diagnosis functions. The deep primary and secondary fusion based on the Internet of Things technology proposed in this study can increase the reliability monitoring of intelligent transformer operation, provide a strong technical guarantee for the normal operation of distribution network, and also provide important technical support for the next research of distribution network technology.**

Keywords—Internet of things technology; primary and secondary integration; smart transformer; fault detection

I. Introduction

In recent years, the rapid growth of China's national economy has driven the rapid growth of energy industry, the main industry of economic growth [1, 2]. In order to meet the needs of national economic development, China Power Grid Corporation has successively put forward the construction objectives of "building a strong intelligent network" and "developing UHV technology", and strive to build a "3 / 5" structure with international scale and a network system to realize efficient power supply [3, 4]. In such a "large power grid", the power transformer as the power supply infrastructure of node substation is more important, which directly affects the safety and reliability of network power supply system [5, 6]. In addition to a large number of more functional relay protection and operation control equipment, modern power transformers also require that they can operate continuously in the power grid for a longer and longer time. Therefore, the accidents caused by the failure of transformer parts are completely unavoidable [7, 8]. In case of failure or major accident, even if there is an emergency plan in place in advance, it will not ensure that the power failure does not affect residents' daily life and commercial operation. Therefore, due to the particularity and complexity of transformer operation, it is of great practical significance to effectively monitor the transformer from multiple directions and angles, diagnose whether it is in an abnormal state, whether there is the possibility of error, and how to analyze, judge and use it [9, 10].

Regarding the research on intelligent evaluation and early warning of substation equipment status, some researchers have proposed that the comprehensive evaluation method in China is mainly based on the equipment status evaluation guidelines issued by the State Grid Corporation comprehensive evaluation method. The state quantities used in the method are relatively comprehensive, including family defects, maintenance records, inspection information, online monitoring, live detection and preventive test data, etc., which basically cover all relevant information of the equipment, including a large number of non-stop power supplies. The detection status quantity and the power failure detection status quantity, therefore, the workload of data acquisition in actual operation is relatively large. In addition, the introduction of various artificial intelligence methods into equipment condition evaluation has become a hot topic at home and abroad. The most widely used intelligent evaluation method is the fuzzy comprehensive evaluation method. The core of the fuzzy comprehensive evaluation method is the determination of the membership function and the determination of the weight. In addition, the gray theory, matter element analysis, evidence theory, Bayesian network, support vector machine and other intelligence analytical methods are also widely used in equipment condition evaluation [11]. Some researchers have proposed that the basic condition for power transformer maintenance is accurate, reasonable and efficient power transformer maintenance, and the reliability of evaluation directly affects the reliability of maintenance decision-making. At this stage, the latest research results are mainly focused on the use of intelligent algorithms to analyze DGA data and part of the electric quantity. These are the lack of overall representation of the transformer status. How to simplify system evaluation, maximize the collection of collectible information, and fully reflect the condition of the transformer is a very important step in the process of transformer maintenance. Nowadays, users have higher and higher requirements for reliability, and it is developing dynamically [12]. In summary, there are many research results on power transformers, but there are few research results on deep

*Corresponding Author.

primary and secondary fusion transformers based on the Internet of Things technology.

This article studies the deep primary and secondary fusion transformer based on the Internet of Things technology. After a general understanding of the relevant theories on the basis of relevant literature data, the realization of deep primary and secondary fusion is analyzed, and then conduct experiments on the transformer with deep primary and secondary fusion. This experiment uses the DGA data and electrical test data of the selected transformer as the input feature vector of the diagnosis model,according to the existing research literature, the five characteristic gases are used as the input vector of the multi-parameter information fusion transformer fault diagnosis model. The transformer automatic fault diagnosis model is modeled based on Adaboost_rbf and dsmt，the sample data set is preprocessed by normalization method. Taking the relative volume fractions of the five characteristic gases in the DGA as the input feature vector, the three models of BP neural network, M-RVM, AdaBoost_RBF1 and PSO-IGWO optimized hybrid KELM are used to diagnose transformer faults.It is concluded that the diagnostic accuracy of transformers based on deep primary and secondary fusion is higher than that of transformers based on DGA.

II. Deep Primary and Secondary Fusion Transformer Research

A. Problems with Smart Transformers

1) Integrated design of overall structure: The integrated design of sensors and transformers needs to fully consider issues such as insulation, energy acquisition, communication, interference, heat dissipation, and maintenance. It also needs to meet the conditions of outdoor operation, and complete the research and design of high protection level structures according to actual use scenarios. In addition, it is more difficult to combine control methods such as capacity adjustment, voltage adjustment, three-phase power adjustment, and automatic load switching with the transformer itself.

2) Self-diagnosis and control system mechanism research and development: The intelligent fusion terminal needs to collect information for self-evaluation and self-diagnosis of the transformer, accurately reflect the operating status of the equipment, the appearance of the equipment, and the environmental operation and maintenance status to the platform, and can realize the corresponding automatic adjustment after being approved by the equipment owner or superior management unit as well as autonomous system actions with other equipment (such as distribution transformer clusters, low-voltage switches, etc), reasonable and accurate judgment algorithms and control mechanisms are essential for this project.

3) Wireless communication: In the data transmission part, the collected information includes analog signals and digital signals, and the types of information sources are different. This requires better compatibility of the communication module, and in this project, the data collected by the sensor is mainly collected through wireless communication. Transmitted to the remote monitoring terminal, this poses a challenge to data transmission. First, the transmission distance must be large enough, and secondly, the transmission delay must be small enough. Foreign countries usually use a combination of multiple communication technology methods, and integrate the best communication methods according to the actual needs of condition monitoring.

B. The Realization of Deep One-two Fusion Optimize the Layout of Transformer Structure

1) Transformer monitoring unit: The transformer online monitoring unit has a variety of communication interfaces, which can communicate with the transformer's environmental sensors, noise sensors, power quality management modules, oil chromatographic analysis modules, etc. It also has the ability to interact with the intelligent terminal of the station and the master station. The online monitoring device can also directly control the voltage and capacity adjustment switches of the transformer through the built-in APP capacity and voltage adjustment strategy, and through the built-in algorithm, cut off some unimportant branches when the transformer is overloaded to ensure the safe operation of the transformer.

2) Communication structure: The online monitoring platform communicates with the master station through wireless, optical fiber and other methods, interacts with data through 104 or MQTT, accepts the APP download of the master station, parameter settings, container management, cluster control; and can be integrated with the station area through the terminal 104 or MQTT interactive data; the environment and noise of the transformer are mainly collected through 485, wireless and other methods for data analysis; the component collection of the oil chromatogram is mainly carried out through 485; the communication with the power quality management module is mainly through the 485 method, Fig. 1.

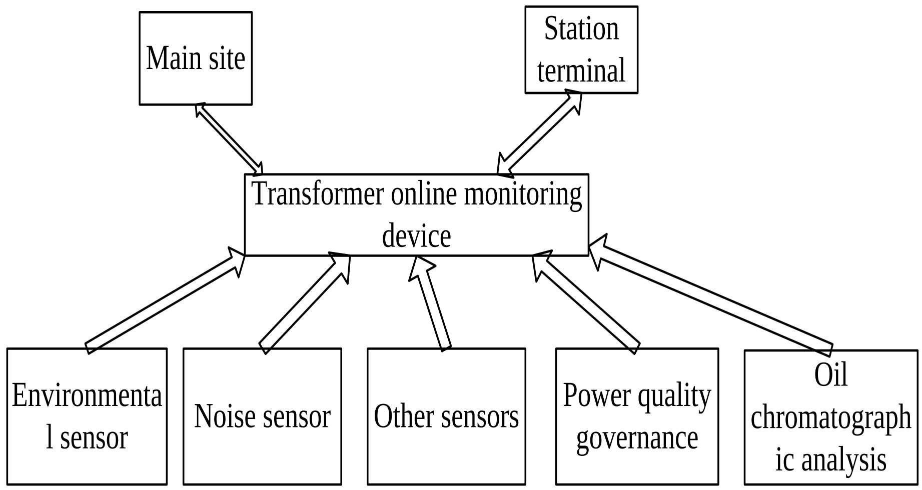


Fig. 1. Communication architecture.

C. Fault Diagnosis

1) Comprehensive application of artificial neural network (ANN) and evidence theory and other methods to construct a multi-information fusion comprehensive state assessment model for transformers. First, the data preprocessing module processes the operating data, constructs the parameter subspace and the fault subspace, selects training samples from the actual operating data, and constructs and trains the sub-network. Based on the evidence theory, construct the identification framework of the diagnosis system, determine the basic probability distribution function of each evidence, and seek to effectively fuse multiple preliminary diagnosis results, including models and methods that can quickly fuse feature-level diagnosis results based on multiple methods. Finally, the evidence is synthesized and diagnosed, so that it can identify the fault correctly and efficiently.

2) Through the creation of state test scenarios, a comprehensive simulation of the various states of the transformer, and the creation of a test data information database. Find the mutual influence relationship between the amount of information sensed by the distribution transformer, establish the transformer condition monitoring system and fault diagnosis model, complete the interactive algorithm between the computing unit and the ontology perception, environmental perception, historical state extraction, interconnection of similar equipment, etc, and realize the status communication and exchange between the equipment owner and the superior management unit, establish the transformer load capacity evaluation parameter system, and evaluate and predict the transformer load capacity. Incorporating systematic data such as transformer body structure, internal faults, technical parameters, cooling methods, power grid operating environment, geographical and meteorological environment, and using principal component analysis algorithm to establish a transformer state evaluation model to obtain a transformer health status score.

D. Panorama Detection

1) Distribution line image monitoring system is suitable for visual monitoring and patrolling of distribution lines. It is a lightweight and convenient installation image monitoring device designed for the visualization of distribution lines. This device can pass 3G/4G/0PGW, optical fiber network, etc, will transmit pictures to the monitoring center regularly, and line maintenance personnel can realize remote picture browsing through computer clients and WeChat clients. The realized functions mainly include: picture capture; timed picture upload; when there is no network or poor signal, the current picture is temporarily stored, and the picture will be resumed after the network is normal; remote manual wake-up capture/small video recording; supporting mobile client, anytime grasp the line status anywhere, and effectively solve the problem of high cost of line channel visualization.

2) Configure high/low voltage current and voltage transformer (built-in), pressure sensor, low oil level / pressure release sensor, temperature sensor, vibration sensor, configure oil chromatography online monitoring device, partial discharge detection device, camera (support picture with 4G / 5G way to upload to the station area).

E. Application of Iot Technology in Deep Primary and Secondary Integration Transformers

1) Global positioning system: Applying the timing characteristics of the GPS to the power system is more important and more extensive than the general positioning characteristics [13]. The power system includes monitoring and protection systems, such as automated microcomputers and safety systems, dispatch automation systems, monitoring systems, fault recorders and accident recorders. All of the above require an accurate time standard to achieve accurate synchronization. In addition, with the development and expansion of the power system, especially the expansion of the power grid, higher requirements are put forward for the convenience and accuracy of the time standard.

2) Radio frequency technology: Radio frequency technology mainly identifies tags. The label usually represents the data information on the device, which is set when the device leaves the factory. The information contained in the label usually refers to data parameters that may have a significant impact [14]. The label and the measured object have a binding relationship. The device information reflected on the label usually includes parameters that have a significant impact on the device, rated voltage, and operating current.

F. Fusion Algorithm

1) The RBF neural network is based on how to manipulate multiple variables and is a three-level network suitable for solving classification problems. A general RBF neural network structure composed of input layer, hidden layer and output layer [15].

The node output returned by the hidden layer of the RBF network can be expressed as:

$$r = \exp(-\frac{1}{2\sigma^2}\|R-T\|^2) \tag{1}$$

Among them, n is the output of the final hidden layer node normalized in the hidden layer, r is the return input vector n, t is the normalized center vector of the return function of the final hidden layer node, and σ is the normalization of the final hidden layer node the output, the normalized linear combination formula of the final hidden layer node is as follows:

$$y = \sum_{k=1}^{h} \varpi r - \theta_j \tag{2}$$

Where r is the output returned by the hidden layer from the node, y is the output of the hidden layer from node j, a is the threshold for returning from the hidden layer to the output plane, and θ is the input threshold for outputting node j from the hidden layer.

2) The structure of DBN is to iterate multiple restricted Boltzmann machines (RBMs), each layer is composed of multiple RBMs, and form a multi-layer deep learning structure by stacking multiple RBMs. These restricted Boltzmann machines can complete bottom-up unsupervised learning. According to the sensing data characteristics of substation equipment, a BP network is set at the top layer as the top-down supervised learning in deep learning [16]. This structure can better obtain the excellent feature expression of data, that is, more abstract and effective high-level features are extracted on the basis of multiple fusion features. In order to realize the evaluation of equipment status, on the basis of feature extraction, a supervised classifier needs to be added at the top of the network to predict and classify the equipment status, so as to determine the early warning and diagnosis information according to the classification results [17]. The basic structure of equipment condition assessment network is shown in Fig. 2.

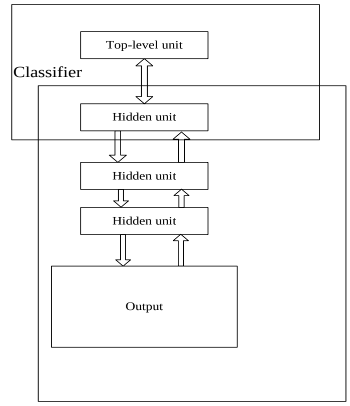


Fig. 2. Basic structure of equipment condition assessment network.

III. Deep Primary and Secondary Fusion Transformer Experiment Based on Internet of Things Technology

A. The Purpose of the Experiment

This article conducts experiments on the deep primary and secondary fusion transformers based on the Internet of Things technology, mainly to test the fault detection of the proposed deep primary and secondary fusion transformers.

B. Experimental Data

The transformer is a complex system, and its fault factors are so complicated that selecting a single DGA data cannot completely and accurately reflect the operating status of the transformer. Different characteristic signals of different levels must be used to synthesize for diagnostic purposes. In order to improve the accuracy of transformer fault diagnosis and more accurately determine the operating status of the transformer, this paper uses the DGA data and electrical test data of the selected transformer as the input feature vector of the diagnosis model [18]. At the same time, according to the existing research literature, the five characteristic gases (CH4, C2H2, C2H4, C2H6, H2), EC 3 ratio (C2H2/CH4, CH4/H2, C2H4/C2H6), Rogers ratio (C2H2/C2H4, CH4 /H2, CzH4/CzH6, C2H6/CH4), Donenberg ratio (C2H2/ C2H4, CH4)/H2, CzH2/CH4, CHs/C2H2) are used as the input vector of the multi-parameter information fusion transformer fault diagnosis model.

C. Failure Model Construction

In the fault diagnosis of power transformer, various state parameters and test characteristic parameters can reflect the fault type and fault location of transformer. It is important to select reliable and effective characteristic parameters to improve the accuracy of transformer fault diagnosis. The selection of characteristic parameters for fault diagnosis should not only meet the practicability of information, but also meet the independence of detection and collection, that is, the characteristic parameters should be independent of each other and reflect the fault characteristics of the transformer as a whole. According to the selection principle of the above characteristic parameters, the transformer fault characteristic parameters are classified into three parts: dissolved gas content in oil (unit: μL /L), electrical test and insulating oil test information, and gas production rate.

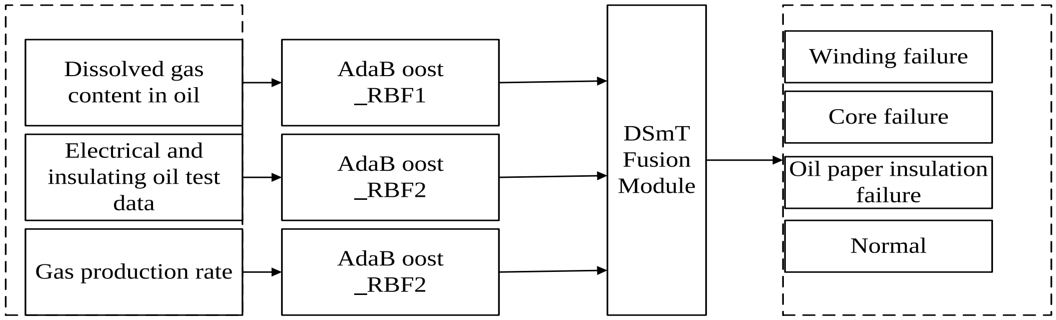


Fig. 3. Transformer fault diagnosis model.

The automatic fault diagnosis model of the transformer is based on Adaboost_rbf and dsmt, and the resulting block diagram after modeling is shown in Fig. 3. In this model, the reference unit of the average volume and fraction of each gas contained in the gas that can be dissolved in the oil is i1. As the core internal current, the oil breakdown trend and the data of the insulating oil test are the diagnostic information unit i2, which contains c2h2, h2 and their total hydrocarbons. The information unit of the absolute value of the gas production rate is the information unit of i3 to diagnose the gas production rate [19]. Combine a classified diagnostic module, use h1, h2, h3 as the input of Adaboost_rbf, and connect each AdaBoost_RBF module in parallel as a first-level diagnostic model system.

The output nodes are diagnostic information modules AdaBoost-RBF1, AdaBoost-RBF2 and AdaBoost-RBF3 corresponding to the basic reliability assignment of the identification framework. In the DSmT fusion module, each reliability assignment is fused to obtain the fused reliability assignment. The final transformer fault diagnosis results are obtained based on the decision rules of maximum reliability allocation.

D. Sample Preprocessing

Due to the large data differences and measurement errors between the transformer oil chromatographic data, the number of samples used only for fault state estimation will affect the final classification performance. In addition, due to different transformer models and peripherals, there are some differences in data [20]. Therefore, in order to effectively eliminate the impact of data differences on fault determination, this article introduces the following normalization methods for preprocessing sample data sets.

$$x_{mew} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

Where is the value after normalization of the sample; x is the initial sample value.

IV. Analysis of Experimental Results

A. Accuracy of Diagnosis Results

This paper conducts algorithm training on the 110kV transformer operation test data collected from the official website of the Power Grid Research Institute. Among the collected samples, there are 50 groups of positive samples and 300 groups of negative samples [21]. Negative samples are classified according to the location where the fault occurs, and are also called winding faults, core performance faults, and petroleum paper electrode performance faults. Finally, the data information of the case sample and the label of the error type of the case sample together constitute a data set of three information modules [22]. Including 320 sets of dga test data, 156 sets of inspection result data and 187 sets of gas rate test data; among them, 320 sets of training sets and 100 sets of dga data tests; 100 sets of training sets and 20 sets of test data. In the air data, there are 100 training sets and 20 total tests.

Set the number of iterations T = 16, the neural network spread factor RBF spread = 0.06, and set the decision limit to ε = 0.3. Table 1 shows the AdaBoost RBF1 training results of the DGA information module. Here, the correct sample rate is the correct test rate of the correct 320 data set diagnostic index and the model's DGA training set.

It can be seen from Fig. 4 that the diagnostic accuracy of AdaBoost_RBF1 is better than that of BF neural network. The diagnostic accuracy of AdaBoost_RBF1 has been increasing and is infinitely close to 100%, while BF neural network has increased at the beginning, but then began to decline.

The electrical test/oil test information module AdaBoost_RBF2 and the gas production rate information module AdaBoost_RBF3 were trained respectively. The iteration parameter T=16 was set, and the spread was 1 and 0.1, respectively. The test sample accuracy of the three AdaBoost_RBF modules, RBF test sample accuracy and SVM test sample accuracy are compared in Table II. It can be seen that the performance of AdaBoost_RBF algorithm is obviously better than RBF and SVM.

TABLE I. Accuracy of Diagnosis Results

	RBF neural network	**AdaBoost_ RBF1**
1	80%	80%
2	81%	81%
4	78%	90%
6	71%	91%
8	70%	93%
10	68%	95%
12	67%	96%
14	60%	98%
16	71%	99%

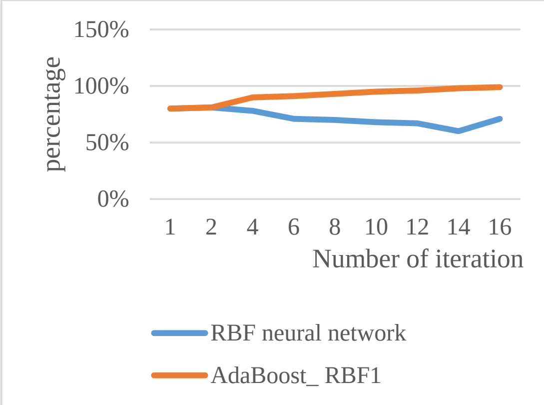


Fig. 4. Accuracy of diagnosis results.

TABLE II. TEST SAMPLE RECOGNITION ACCURACY

The accuracy of test samples under different algorithms	Module class		
	DGA	**Electrical test and oil test**	**Gas production rate**
AdaBoost_RBF	78.33%	93.33%	76.59%
RBF	66.67%	76.67%	51.06%
SVM	63.33%	83.33%	53.19%

B. Comprehensive Diagnosis

Taking the relative volume fractions of the five characteristic gases in the DGA as the input feature vector, the three models of BP neural network, M-RVM, AdaBoost_RBF1 and PSO-IGWO optimized hybrid KELM are used to diagnose transformer faults. The diagnosis results are shown in Table III.

It can be seen from Fig. 5 that the diagnostic accuracy of transformers based on deep primary and secondary fusion is higher than that of transformers based on DGA, which increases the accuracy by about 3%.

According to the above experimental results, the diagnostic accuracy of AdaBoost_RBF1 is better than that of BF neural network. The diagnostic accuracy of transformer based on depth primary and secondary fusion is higher than that of transformer based on DGA. It can be concluded that the diagnostic accuracy of the depth-based primary and secondary fusion transformer is high, the performance of fault detection is good, the accuracy of information detection is improved, the range of data observation is expanded, and the working test performance of the primary and secondary fusion transformer is improved.

TABLE III. COMPREHENSIVE DIAGNOSIS RESULT

	Deep primary and secondary fusion transformer	DGA transformer
BP neural network	76.1%	72.1%
M-RVM	83.5%	81.4%
AdaBoost_ RBF1	89.8%	87.3%
PSO-IGWO optimized hybrid KELM	87.8%	86.3%

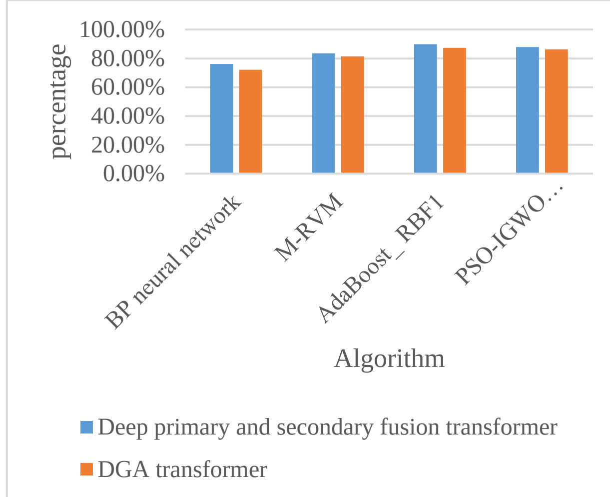


Fig. 5. Comprehensive diagnosis result.

The structure of the model has a direct impact on the effect of diagnosis: (1) Deep learning should fully consider the depth of the model, model parameters and other information. (2) Ensemble learning should not only select homogeneous base classifiers, but also combine different base classifiers to fully synthesize the advantages of each model. (3) Combine the advantages of various intelligent algorithms. Expert system, fuzzy theory, neural network, intelligent optimization computing and other intelligent methods should complement each other, enhance strengths and avoid weaknesses, so as to develop models with superior performance. (4) Model selection and training methods should consider the characteristics of data. To solve the problems of data imbalance and data redundancy with appropriate models and training methods will be more conducive to improving the accuracy of transformer fault diagnosis and guiding the production operation.

V. CONCLUSION

This paper focuses on the research of the deep primary and secondary fusion transformer based on the Internet of Things technology, through optimizing the transformer structure layout and neural network intelligent algorithm realizes the physical fusion and information fusion of primary and secondary equipment. So that it has the ability of panoramic perception of self-state, and realizes the holographic transparent operation of transformers, intelligent research and judgment of situation, active decision-making of hidden dangers, and dynamic autonomous adjustment. This paper verifies the feasibility and effectiveness of fault detection and fusion technology of deep primary and secondary fusion transformer based on Internet of Things technology through theory and experiment, which provides a reference for the subsequent development of intelligent transformers with higher fusion degree. The development of primary and secondary fusion equipment solves a series of complex management problems of the original primary and secondary equipment, but there are also shortcomings in some aspects. Future technology needs to further solve the safety, reliability and efficient fusion problems of primary and secondary fusion equipment. It needs to be further studied in the later period.

ACKNOWLEDGMENT

Research and application of local discharge detection of distribution network switchgear based on electric pulse technology (52053018000M).

REFERENCES

[1] W. U. Zhao, "Research on the Application of Internet of Things Technology to Digital Museum Construction," Acta Geoscientica Sinica, Vol. 38, No. 2, pp. 293-298, 2017.

[2] S. Liu, "Research on the application of internet of things technology in digital sports and community fitness," Boletin Tecnico / Technical Bulletin, Vol. 55, No. 4, pp. 139-145, 2017.

[3] Z. P. Fang, L. Yang, S. Yang, et al. "Transformer-Less Unified Power-Flow Controller Using the Cascade Multilevel Inverter," IEEE Transactions on Power Electronics, Vol. 31, No. 8, pp. 5461-5472, 2016.

[4] K. H. Kim, E. Y. Lee, Energies, et al. "Simultaneous Production of Transformer Insulating Oil and Value-Added Glycerol Carbonates from Soybean Oil by Lipase-Catalyzed Transesterification in Dimethyl Carbonate," Nano Letters, Vol. 2, No. 12, pp. 1349-1352, 2017.

[5] M. Liserre, G. Buticchi, M. Andresen, et al. "The Smart Transformer: Impact on the Electric Grid and Technology Challenges," IEEE Industrial Electronics Magazine, Vol. 10, No. 2, pp. 46-58, 2016.

[6] M. Iwakuma, K. Funaki, H. Kanetaka, et al. "Quench analysis in a 72 kVA superconducting four-winding power transformer," Cryogenics, Vol. 29, No. 11, pp. 1055-1062, 2016.

[7] J. Jiang, G. M. Ma, H. T. Song, et al. "Tracing Acetylene Dissolved in Transformer Oil by Tunable Diode Laser Absorption Spectrum," IEEE Transactions on Dielectrics & Electrical Insulation, Vol. 7, No. 6, pp. 3435-3442, 2017.

[8] J. Li, Q. Zhang, W. Ke, et al. "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," IEEE Transactions on Dielectrics & Electrical Insulation, Vol. 23, No. 2, pp. 1198-1206, 2016.

[9] S. Dutta, S. Hazra, S. Bhattacharya, "A Digital Predictive Current-Mode Controller for a Single-Phase High-Frequency Transformer-Isolated Dual-Active Bridge DC-to-DC Converter," IEEE Transactions on Industrial Electronics, Vol. 63, No. 9, pp. 5943-5952, 2016.

[10] V. F. Pires, J. F. Martins, D. Foito, et al. "A Grid Connected Photovoltaic System With a Multilevel Inverter and a Le-Blanc Transformer," International Journal of Renewable Energy Research, Vol. 2, No. 1, pp. 84-91, 2016.

[11] X. Xi, R. Sioshansi, "A dynamic programming model of energy storage and transformer deployments to relieve distribution constraints," Computational Management Science, Vol. 13, No. 1, pp. 119-146, 2016.

[12] N. Hashemnia, A. Abu-Siada, S. Islam, "Detection of power transformer bushing faults and oil degradation using frequency response analysis," IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 23, No. 1, pp. 222-229, 2016.

[13] Y. Chen, X. Liu, Y. Cui, et al. "A MultiWinding Transformer Cell-to-Cell Active Equalization Method for Lithium-Ion Batteries with Reduced Number of Driving Circuits," IEEE Transactions on Power Electronics, Vol. 31, No. 7, pp. 4916-4929, 2016.

[14] Y. Kim, G. Jung, H. Kim, et al. "Wheel Transformer: A Wheel-Leg Hybrid Robot with Passive Transformable Wheels," IEEE Transactions on Robotics, Vol. 30, No. 6, pp. 1487-1498, 2017.

[15] Y. Shi, T. Wu, M. Cai, et al. "Energy conversion characteristics of a hydropneumatic transformer in a sustainable-energy vehicle," Applied Energy, Vol. 171(Jun.1), pp. 77-85, 2016.

[16] Zou X . The Application of Internet of Things Technology in Building Integration System Design[J]. Computer & Telecommunication, 2015.

[17] Zhang K , Amp I , Company T . Application and Practice of Internet of Things Technology in the Business Integration and Information Connection of the Life Cycle of Power Grid Assets.

[18] dos, Santos, Gustavo, et al. Thermography and artificial intelligence in transformer fault detection[J]. Electrical Engineering, 2018.

[19] Samet H , Marvasti F D . Fault detection in the secondary side of electric arc furnace transformer using the primary side data[J]. European transactions on electrical power, 2014.

[20] Jadhav R V, Lokhande S S , Gohokar V N . Monitoring of transformer parameters using Internet of Things in Smart Grid[C]// International Conference on Computing Communication Control & Automation. IEEE, 2017.

[21] Bai D , Yang Y , Li C . Research on Low Power Sensor Networking Technology of Transformer Equipment Based on Node Energy[J]. Journal of Physics Conference Series, 2019, 1346:012043.

[22] Dong X , Chang Y , Wang Y , et al. Understanding usage of Internet of Things (IOT) systems in China: Cognitive experience and affect experience as moderator[J]. Information Technology & People, 2017, 30(1):117-138.

Research on High Voltage Cable Condition Detection Technology based on Wireless Sensor Network

Yang Zhao[*1], Qing Liu[2], Tong Shang[3], Yingqiang Shang[4], Rong Xia[5], Shuai Shao[6]
State Grid Beijing Powercable Company, Beijing, 100022 China[1, 2, 3, 4]
China Electric Power Research Institute Limited, Wuhan Branch, Wuhan, 430079 China[5]
Center of Jinan Power Supply Company of State Grid Shandong Electric Power Company, Jinan, 250012, China[6]

***Abstract*—The development and progress of modern society cannot be achieved without the support of electric power resources, and at present, electric power is the most important energy source to promote social development and maintain human life. As a key unit under the power distribution and transmission system, the high-voltage cable of the power grid will undertake the task of supplying power resources to the whole power grid. Therefore, based on the transmission line fault diagnosis framework of Wireless Sensor Networks (WSN), a high-voltage cable path condition monitoring scheme using LoRa technology is proposed. Three high-voltage cable condition monitoring periods are proposed according to the difference of high-voltage cable fault rate, and the delay and energy consumption of the high-voltage cable monitoring system are optimized by multi-objective particle swarm algorithm reality. The experimental results show that the proposed high-voltage cable detection technology can switch the working mode according to different environments, and the data communication packet loss rate is less than 5%, while the detection platform has excellent delay performance and energy saving effect. The high-voltage cable status detection solution can effectively solve the problem of blind high-voltage cable channels in high mountain areas of China. The research content has important reference value for the detection of China's power grid circuit system.**

Keywords—Wireless sensor; high-voltage cable; fault diagnosis; particle swarm algorithm

I. INTRODUCTION

Wireless sensor network is a kind of communication and acquisition network composed of nodes, which can realize the detection and acquisition of external physical energy. With the development of information and communication technology in recent years, wireless sensor networks have a wide range of applications in modern medical, industrial manufacturing, smart grid, etc. [1]. Wireless sensor network has the characteristics of bottom power consumption, simple application and easy expandability; therefore, wireless sensor network also has a very wide application in the field of smart grid power data acquisition, cable status monitoring, grid fault diagnosis, etc. [2-3]. With the advent of the information era, smart grid construction has become the new goal of China's power grid development. The United States as the representative of the developed countries on the future development of smart grid development strategy, which will be the wireless sensor network ZigBee protocol as an important communication standard for the construction of smart grid [4]. This shows that wireless sensor networks play an important role in the construction of smart grid. However, the construction of smart grid needs to meet the requirements of delay, bandwidth, reliability, and coverage, etc. The high-voltage grid is in a complex working environment, subject to harsh environment, complex weather, and interference from various electrical equipment signals, which will affect the application of wireless sensor networks [5]. Therefore, in view of the high difficulty in maintenance and repair of transmission lines in remote mountainous areas, a detection scheme for high-voltage cables in mountainous areas is proposed based on WSN technology. Different detection schemes are adopted for cables in different periods, and multi-objective particle swarm optimization algorithm is used to solve the problem, so as to solve the problem of high-voltage cable line fault detection in remote areas. The research content has important research value for the construction and development of smart grid in China.

Wireless sensor networks are widely used in modern medical, industrial manufacturing, smart grid, etc. A lot of work has been done by domestic and foreign researchers to study wireless sensor networks. Tabella et al. investigated the effect of wireless sensor networks (WSN) on oil spill detection and localization in subsea production systems. Four localization algorithms were studied and the detection and localization performance were compared with the (location) perspective chair Varshney's rule (CVR) and Cramér-Rao lower limit (CRLB), respectively. The results show that the components that would lead to leakage in case of failure and their corresponding failure rates [6]. Xf et al. found that the routing survivability in harsh environments is questionable. To address this issue, an environmental fusion multipath routing protocol (EFMRP) was proposed. The results show that EFMRP can significantly improve packet delivery rate and network lifetime in harsh environments [7]. Hu et al. found that a fully distributed time synchronization method based on Gaussian belief propagation would lead to degradation of synchronization accuracy. Thus, a Sequential Belief Propagation-based Distributed Time Synchronization algorithm (SBP-DTS) was proposed to reduce the number of Gaussian mixture components in messages using a weighted expectation maximization (EM) algorithm. Finally, the performance of SBP-DTS was evaluated under asymmetric Gaussian and exponential delay models [8]. Houssein et al. found it a challenging task to determine the location of convergence nodes in LSWSNs, and Harris' hawk's optimization (HHO) algorithm was used to solve this problem. The results show that the adopted approach has advantages in terms of energy consumption and localization errors [9]. Lin et al. proposed a priority-aware packet transmission scheduling (PPTS) framework in a cluster-based IWSN, in which the PPTS policy, optimization theory, and implementation design are systematically considered. The results show that the proposed PPTS policy not only minimizes the transmission delay of high-priority packets but also improves the transmission delay of low-priority packets [10]. Zhan et al. propose an optimization problem that minimizes the weighted sum of the above two costs by optimizing the UAV trajectory and wake-up time allocation as well as the transmit power of all SNs. In addition, a new approach to design the initial UAV trajectory using the multi-traveler problem (MTSP) technique is proposed. The results show that the proposed design achieves a flexible trade-off in cost balance between UAVs and SNs [11]. Wang et al. propose a LEACH-EA protocol based on the LEACH algorithm for the low-power requirements of wireless sensor networks. Experiments show that the LEACH-EA protocol provides significant improvements in network life cycle, total network data transmission and number of cluster heads [12]. Verma et al. found limitations when using the battery power of sensor nodes. A fuzzy logic based effective clustering (FLEC) for homogeneous wireless sensor networks for mobile receivers is proposed. The probability of average energy is used to select the appropriate cluster head and the results show that the proposed FLEC scheme outperforms LEACH, DEEC and LEACH fuzzy protocols [13]. Jaber et al. proposed an "adaptive fully distributed duty cycle for content-centric wireless sensor networks (ADDC-CCWSN) "mechanism. The ADDC-CCWSN aims to reduce the activity of nodes with a high percentage of unsatisfied interest in the PIT. The results show that the proposed approach achieves significant energy efficiency gains while ensuring a high interest satisfaction rate and maintaining almost the same latency [14].

Neural network algorithms have a wide range of applications in the field of smart grid, providing important data support for data collection and grid diagnosis, and researchers at home and abroad have done a lot of work to study them. Yu et al. proposed a deep reinforcement learning algorithm for safe shutdown strategy in order to deal with incomplete distribution grid models to solve the voltage reactive power control problem in a model-free manner. The results show that the proposed algorithm outperforms existing reinforcement learning algorithms [15]. Pan E et al. used hierarchical cluster analysis to analyze grid influencing factors, and the results show that the prediction model is important to achieve accurate calculation of provincial enterprise investment [16]. Ying et al. proposed a coordinated scheduling method for plug-in electric vehicle PV generation microgrid based on extended power prediction, using a clustering algorithm to build the power prediction model was established and the results showed the good performance of the proposed extended coordinated scheduling algorithm [17]. Bhamidipati et al. developed Wide Area Monitoring Systems (WAMSs) in order to monitor the grid in a wide area. And a new wide-area monitoring algorithm was developed to verify each substation in an artificial intelligence framework. The results show that the Kullback-Leibler scatter-based method has fast detection time and the timing error estimation accuracy exceeds the limits provided by the IEEE C37.118.1-2011 standard [18]. Iliadis et al. developed a new algorithm for managing the currents of an islanded power system that enables more stable conventional unit operation and peak demand reduction. The results demonstrate that the proposed algorithm can achieve smoother diesel generator operation [19]. Jyotheeswara et al. proposed a neural network-based maximum power point tracking (MPPT) controller for PEMFC grid-connected systems to extract maximum power from proton exchange membrane fuel cells. The results show that the proposed RBFN-MPPT controller has excellent performance [20].

It can be seen from related domestic research that neural network algorithms have a wide range of applications in the field of smart grid. Optimization of wireless sensor network parameters by neural network algorithm is important to improve the monitoring effect of power grid cables.

II. Construction of High-Voltage Cable Condition Detection Model based on Wireless Sensor Network

A. Mountain High Voltage Cable Path Loss Model Construction

Wireless sensor network is a wireless communication technology, and its low cost and maintenance-free characteristics make WSN technology gradually replace the backward wired monitoring scheme. Therefore, WSN technology will have a very broad application prospect in the field of smart grid. High-voltage cables in mountainous areas are susceptible to seasonal climate and harsh environment, and high-voltage cable faults are mainly concentrated in summer and winter time. To meet the requirements of monitoring high voltage cables in mountainous areas, transmission cable monitoring solutions need to meet the challenges of signal transmission, harsh environmental issues, and service life in mountainous areas. Therefore, based on wireless sensor network technology, an in-line monitoring system for high voltage cables in mountainous areas using LoRa is proposed. The high-voltage line transmission architecture is constructed by setting LoRa and cellular modules in the mountainous environment, and dynamic LoRa packet networking modules are set to guarantee the stable transmission of data signals as well as to improve the efficiency of system resources utilization and reduce system energy consumption. As shown in Fig. 1, the application principle of WSN in the field of smart grid.

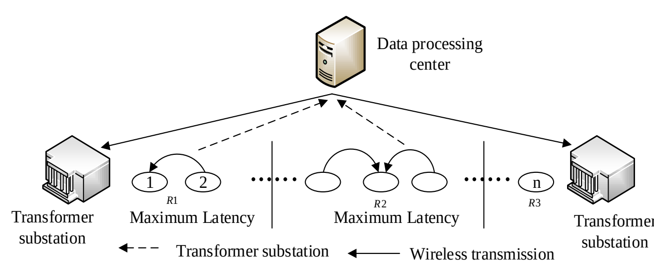


Fig. 1. Application principle of WSN in Smart Grid

According to the analysis of mountain high-voltage cable monitoring data, the probability of cable failure is different at different times, and mountain high-voltage cable monitoring is divided into three monitoring time periods. In the mountain storm season, snowstorm and frost season, when the relevant values are higher than the meteorological safety indicators, the occurrence of natural disasters around the mountain area of high-voltage cables is high, and the monitoring of mountain high-voltage cables is at the high-risk monitoring stage; natural disasters or abnormal communication signal transmission occur in the mountainous grid area, and the monitoring of mountain high-voltage cables is at the fault monitoring stage; cable monitoring data return to normal, and meteorological disasters gradually decrease, and mountain high-voltage cable monitoring is in the low-risk monitoring stage. According to the characteristics of the monitoring period, the fault stage gives priority to feedback data to the monitoring and issue fault maintenance orders; secondly, the high-risk stage feeds data to the monitoring and monitors the grid route fault problem; the low-risk stage only needs to meet the requirements of stable data transmission and cable detection. The signal strength of the transceiver module LoRa at a point in the mountainous area is expressed as the difference between the field strength of the system base station antenna and the field strength loss of the signal reaching that point, as seen in equation (1).

$$E_R = E_T - L_{PS} \tag{1}$$

In equation (1), L_{PS} represents the signal loss in the median path, and E_T represents the field generated by the LoRa antenna position of the transceiver module. During the transmission of the signal in the non-direct path, the bypassing phenomenon will occur, as shown in Fig. 2, then the bypassing path expression is seen in equation (2).

$$\mathrm{d}_D = \sqrt{d_1^2 + (h_m - h_T)^2} + \sqrt{d_2^2 + (h_m - h_R)^2} \tag{2}$$

In equation (2), d_1 indicates the distance between the mountain and the transceiver module, d_D indicates the signal bypass distance, d_2 indicates the distance between the mountain and the antenna, h_R indicates the antenna altitude, h_T indicates the transceiver module altitude, and h_m indicates the mountain altitude.

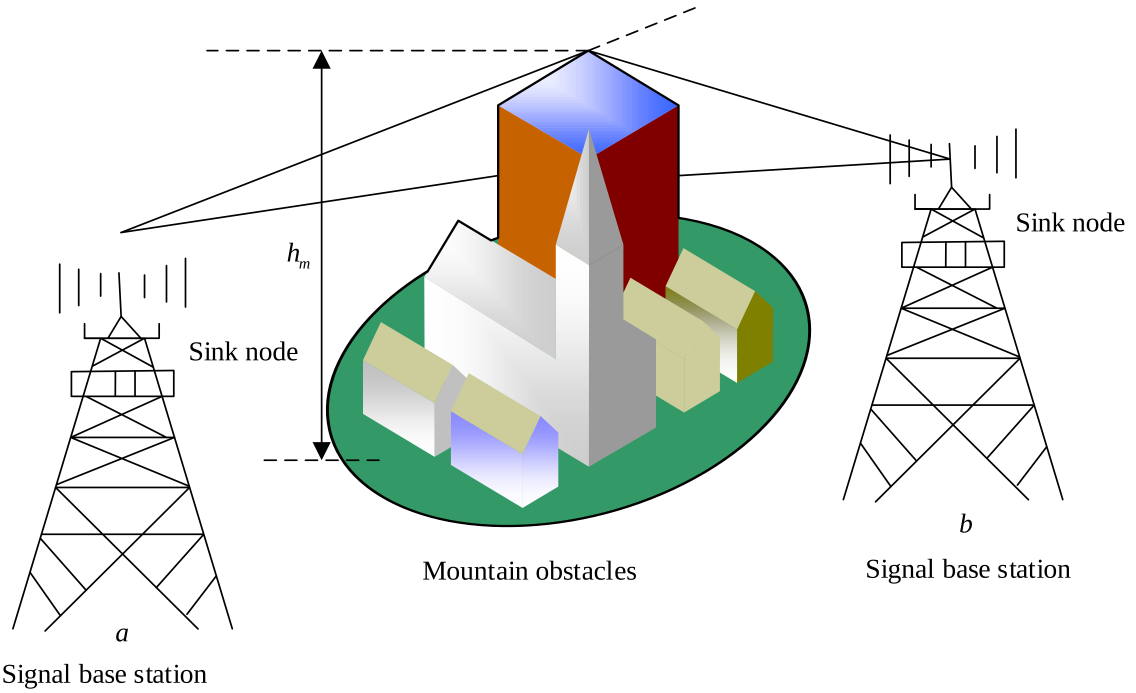


Fig. 2. Radio diffraction principle

The colored building areas in Fig. 2 represent signal blocking areas such as mountains and forests. The signal tower transmitting signal is blocked by the mountain and signal bypassing phenomenon occurs. The transceiver module LoRa signal band is 433-915Mhz, to meet the bypass signal loss requirements, the model applicable to the loss monitoring in this band is Egli model, which can be more accurate to evaluate the terrain field strength in mountainous and hilly areas, etc. The Egli signal loss expression is seen in equation (3).

$$L_{ps} = 88 + 20\lg f + 40\lg d - 20\lg(h_t h_r) - K_h \tag{3}$$

In equation (3), h_r indicates the antenna height, h_t indicates the transceiver module height, f radio frequency,

K_h environmental correction factor, and d indicates the distance between the transceiver and signal antennas.

B. Construction of Time Delay and Energy Consumption Model of High Voltage Cable in Mountainous Areas

The high-voltage cable monitoring model uses LoRa to achieve optimization of dynamic grouping with multiple objectives, and algorithms are used to find the number of cellular wireless modules installed as well as the number, so as to achieve the type of network under different system delay and energy consumption conditions, and to meet the switching of multiple working states of the high-voltage cable monitoring system of smart grid. Network modeling is used in the study as a data directed graph, and transmission towers are located at substation locations, which can directly use LoRa modules to achieve end-to-end data transmission. Therefore, the directed graph then mainly considers the number of transmission towers required by the wireless transmission module. The directional diagram of network data transmission is shown in Fig. 3.

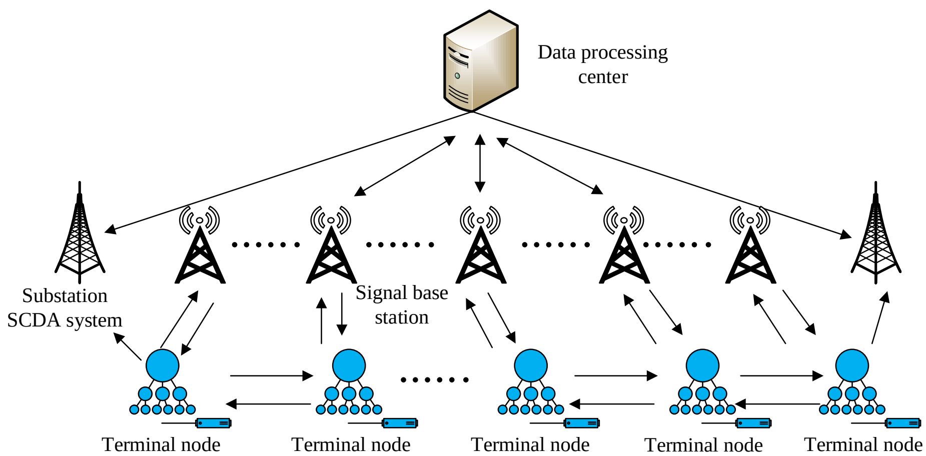


Fig. 3. Directed graph of network data transmission

The cable detection model mainly needs to find the available communication path and the communication link delay is kept minimum and can save the network construction cost, then the communication link delay should be less than or equal to the maximum communication delay, as seen in equation (4).

$$\sum_{(i,j)} D_{i,j,k} M_{i,j,k} \le D_{\max} \quad \forall k \in N \tag{4}$$

In equation (4), $M_{i,j,k}$ denotes the binary decision variable, $D_{i,j,k}$ denotes the node k communication delay, and $D_{\max}$ denotes the system maximum delay requirement. To ensure that the communication link is reused at a reduced cost, then $O_{i,j}$ denotes the binary variable, if the data link (i, j) is used, there is $O_{i,j} = 1$, otherwise $O_{i,j} = 0$, expressed as seen in equation (5).

$$M_{i,j,k}, L_{i,} G_i,\ O_{i,j} \in \{0,1\} \quad \forall\ (i,j) \in p,\ \forall k \in N \tag{5}$$

In equation (5), $L_{i,}$ and G_i are binary variables, where i towers use cellular network, then there is $G_i = 1$, otherwise 0, and if i towers use LoRa, then $L_{i,} = 1$, otherwise 0. The variables $M_{i,j,k}$ and $O_{i,j}$ can be determined by equation (5). Also combine equation (4) with equation (5) to get all network delay functions as seen in equation (6).

$$D(M_{i,j,k}, L_{i,} G,\ O_{i,j}) \quad \forall\ (i,j) \in p,\ \forall k \in N \tag{6}$$

The expression of the network loss function can be found by equation (6) as seen in equation (7).

$$E(G_i, L_i) = \sum_{i=1}^{N} (dG_i + bL_i) + \sum_{(i,j)\in p} C_{i,j} O_{i,j} + L_{ps} \tag{7}$$

In equation (7), d denotes the energy consumption of the cellular radio module, b denotes the energy consumption of the LoRa module, and $C_{i,j}$ denotes the energy consumption at the data link (i, j). The expression of the dynamic model of mountain cable LoRa obtained by Eq. (4) to Eq. (7) is seen in Eq. (8).

$$F\left[D(i,j,k), E(i,j)\right] \ \forall (i,j) \in p, \forall k \in N \tag{8}$$

In the LoRa module, SF is the spreading factor, BW is the signal bandwidth, and CR is the encoding rate. Using the sem tech formula, the transmission time of a LoRa individual node packet over the air can be calculated, and by using the user key control parameters, the LoRa symbol rate is expressed as seen in equation (9).

$$R_s = \frac{BW}{2^{SF}} \tag{9}$$

With the user control parameters and the symbol rate definition information, the LoRa symbol period representation is obtained as seen in equation (10).

$$T_s = 1/R \tag{10}$$

The number of propagation times of LoRa data in the air is the sum of the packet transmission time and the leading code time, while the leading code is calculated as seen in equation (11).

$$T_p = (n_p + 4.25)T_s \tag{11}$$

In equation (11), n_p is the leading code setting length, which is stored in the register location, then the payload transmission time is equation (12).

$$T_1 = \varepsilon * T_s \tag{12}$$

In equation (12), ε denotes the payload symbol, then the LoRa data transmission time can be obtained as the sum of the leading code transmission time and the payload time, as seen in equation (13).

$$T = T_p + T_l \tag{13}$$

In equation (13), T_p is the lead code transmission time T_p, T_l is the payload time. LoRa data transmission process energy generation as seen in Figure 4.

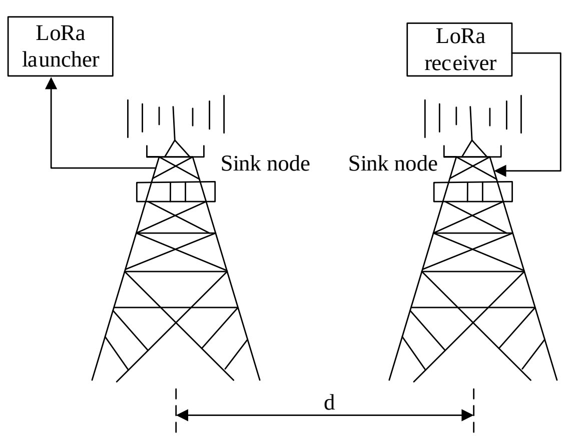


Fig. 4. Lora energy generation in lora data transmission process

Then, in LoRa data transmission, E_{RX} denotes the data transmitting energy consumption, E_{TX} denotes the data receiving energy consumption, d denotes the distance of data transmission process, and n denotes the data size. The total energy consumption of LoRa data transmission is shown in equation (14).

$$\begin{cases} T_{TX}(n,d) = E_{Telec} \times n + \varepsilon_{amp} \times n \times d^k \\ E_{RX(n)} = E_{RX-elec(n)} = E_{\mathrm{Re}lec} \times n \\ E_{LoRa} = E_{TX}(n,d) + E_{RX}(n) \\ E = L_{LoRa} + E_{CC}(n) \end{cases} \tag{14}$$

In equation (14), E_{Telec} is the energy consumption of the device sending data, $E_{\mathrm{Re}lec}$ is the energy consumption of the device receiving data, ε_{amp} is the energy consumption of the device power amplifier unit, $E_{CC}(n)$ is the energy consumption of the cellular module sending data, and k is the signal attenuation index with a value of 4 in the mountain area.

C. Lora Transmission Model Construction based on MOPSO Algorithm

The analysis of the LoRa transmission model reveals that the delay model is influenced by the LoRa transmission data and device transmission parameters, while the energy consumption model is influenced by the data transmission distance, data transmission volume, and transceiver module power. Meanwhile delay and energy consumption are two contradictory target data. To ensure both low latency and good energy consumption for high voltage cable monitoring system, a compromise approach is needed. A neural network algorithm is used to optimize the delay and energy consumption objectives so that the delay and energy consumption targets meet the requirements of the online high-voltage cable detection system. Therefore, the multi-objective particle optimization algorithm (MOPSO) is used to optimize the energy consumption E and T two important parameter objectives, then the mathematical expression of MOPSO for M objectives is seen in equation (15).

$$\begin{cases} \min y = f(x) = \left| f_1(X), f_2(X)..., f_m(X) \right| \\ S.t. \quad x = (x_1, x_2..., x_D) \in X \\ \qquad y = (y_1, y_2..., y_M) \end{cases} \tag{15}$$

In Eq. (15), y denotes the target space, X denotes the decision space, y denotes the target vector, and x denotes the decision vector. In the multi-objective particle swarm algorithm calculation, each particle has a unique solution in the solution space and adjusts its target flight according to the spatial flight experience of its peers and its own experience. The particle trajectory in space has the best flight position, which is the optimal solution for the particle swarm. And the flight position that the whole particle swarm experiences is the optimal solution that the swarm finally searches for.

Define $H_a = (h_{a1,} h_{a2}, ..., h_a D)$ as the D dimensional position of the a particle of the particle swarm and the particle $a = 1, 2, ...s$. The flight speed of particle a is $aV_a = (v_{a1}, v_{a2}, v_{a3} ..., v_{aD})$. The optimal solution of the particle swarm a is $p_a = (p_{a1}, p_{a2}, p_{a3} ..., p_{aD})$, and the global optimal solution of the whole particle population is $p_o = (p_{o1}, p_{o2}, p_{o3} ... p_{oD})$. The particle population counts the velocity expression of the particles in each iterative update as seen in equation (16).

$$v_{ab}(k+1) = \theta v_{ab}(k) + c_1 r_1 (p_{ab} - h_{ab}(k)) + c_2 r_2 (pg_b - h_{ab}(k)) \quad (16)$$

In Eq. (16), denotes the number of particle swarm, a $a = 1, 2, ...s$ and b are the parameter values, $b = 1, 2, 3, ..., D$ and L_b denote the lower limit of the search space, U_b denotes the upper limit of the search space, v_{ab} denotes the swarm flight speed, c_1 and c_2 denote the swarm learning factor, θ denotes the inertia weight, and r_1 and r_2 denote the random number. For each iteration, the position of the particles is expressed as shown in equation (17).

$$h_{ab}(k+1) = h_{ab}(k) + v_{ab}(k+1) \quad (17)$$

In equation (17), h_{ab} is the particle swarm dimensional position, $Dh_{ab} \in [L_b, U_b]$. The flow of multi-objective particle swarm to achieve the objective optimization is shown in Fig. 5.

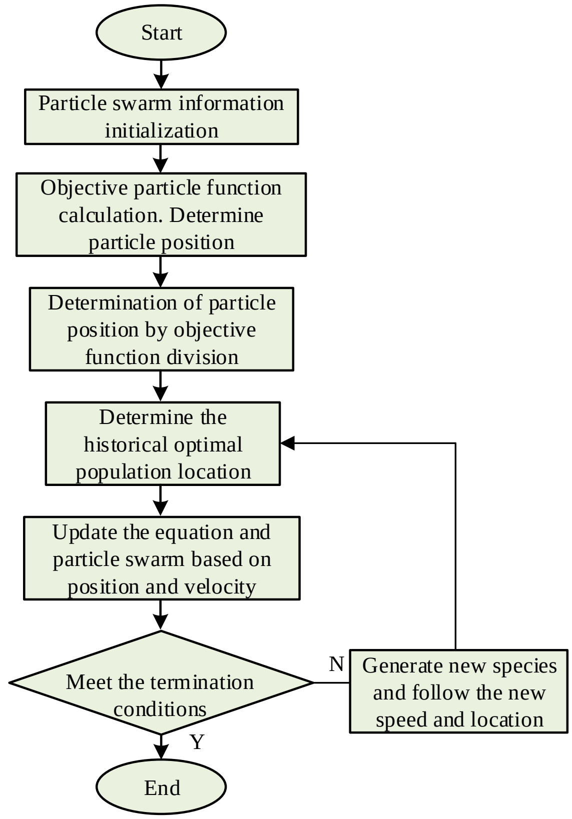


Fig. 5. Multi-objective particle swarm optimization process

III. High Voltage Cable Condition Detection Model Simulation Test

In order to verify the performance of the proposed online high-voltage cable detection model, the online high-voltage cable monitoring network will be simulated and tested using MATLAB with Pocket Tracer simulation platform. The LoRa module used in the experimental test transmits from 1km to 5km range, and the transmission rate ranges from 5kb to 30kb range, and the distance is inversely proportional to the data transmission rate; while the cellular wireless module is used without considering the distance, and the rate is 125m/s. The cellular network has been affected by data access delay, routing conversion, and the delay is taken as the average value of 55ms. 100 transmission poles with a spacing of 1 km, and each inductive pole tower are placed a group of sensors, each group of sensors each time to collect data capacity of a total of 2.78kb.

In the test of the relationship between path loss and the influence of LoRa signal receiving and transmitting segment distance, the experiments are based on the Egli path loss model, and the relationship between path loss and LoRa signal segment is obtained through simulation tests, as shown in Fig. 6.

Fig. 6 shows the results of the relationship between the actual field strength of the path and the theoretical field strength. The real field strength test shows that the LoRa signal becomes weaker with the increasing distance of the field strength, which corresponds to the decreasing trend of field strength with increasing distance of Egli model. However, the actual comparison shows that the decreasing trend of field strength in the experimental test is significantly smaller than the trend of Egli's theoretical calculation. In the case of mountain peaks, the difference between the actual field strength and the theoretical field strength will be monotonically reduced by the increasing distance, but the error value will increase in the case of mountain peaks. Although the Egli model can describe the relationship between field strength and distance for transmission signals in mountainous areas, the model has a large error and needs to be optimized to effectively predict the path loss in mountainous areas. Meanwhile, in online high-voltage cable detection system, delay model and energy consumption model are the key to system optimization. To discuss the performance of LoRa wireless networking, it is necessary to analyze the LoRa bottleneck energy in cellular wireless modules as well as the number of groups in LoRa modules and multi-hop transmission distance between modules. Therefore, the relationship between different wireless transmissions and the maximum system delay during networking is tested in the Pocket Tracer simulation platform, as shown in Fig. 7.

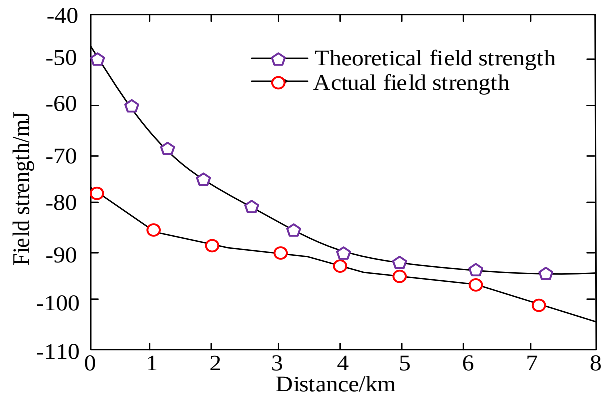


Fig. 6. It is the result of the relationship between the actual field strength and the theoretical field strength of the path

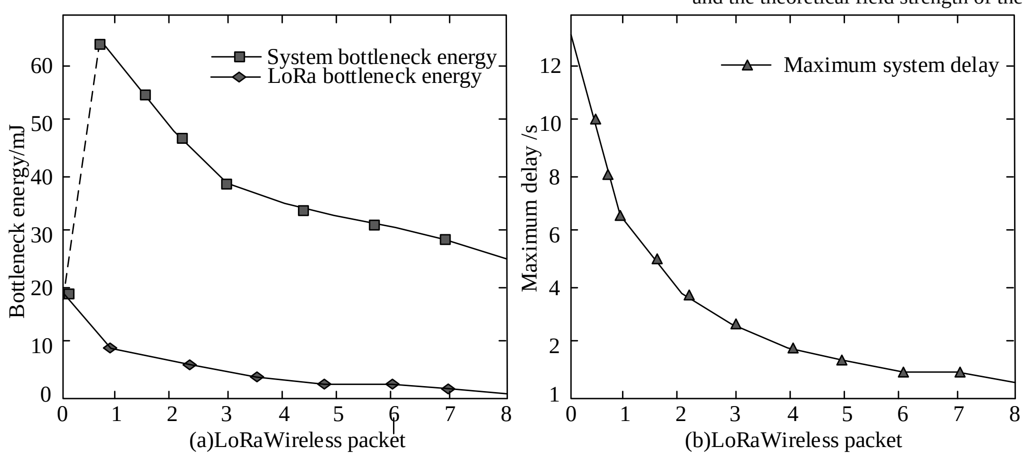


Fig. 7. Results of the relationship between system delay and bottleneck energy consumption

Fig. 7 shows the results of the relationship between the system delay and the bottleneck energy consumption. In Fig. 7(a) the results of the relationship between the number of LoRa wireless groups and the bottleneck energy are shown. When the cable monitoring network belongs to LoRa step-by-step hopping network, the system bottleneck energy is the same as LoRa bottleneck energy, and the terminal node closest to the gateway node transmission network will forward all node data at this moment, and the system bottleneck energy is consumed as the energy of this node; meanwhile, LoRa transmission group expands with the number, this network system takes cellular module for data transmission, and the cellular module energy consumption is much higher than the LoRa module, and the system bottleneck energy at this moment also increases with it. However, along with the expansion of the number of cellular modules, the transmission of data information of individual modules decreases, so the system bottleneck energy under the network will decrease as the number of module groups increases. Also, when the system uses a heterogeneous network, LoRa modules will also consume less bottleneck energy as the number of LoRa groups increases. The bottleneck energy consumed by the LoRa and cellular networks is relatively fixed at this moment, with the LoRa bottleneck energy consumption being 2mJ and the cellular 27mJ. The results of the relationship between the number of wireless groups and the maximum delay are shown in Fig. 7(b).

Due to the increasing number of LoRa wireless packets, the data transmission between individual LoRa modules is reduced, and not only the energy consumption is reduced, but also the maximum delay of data transmission is gradually reduced.

The MATLAB platform is used to implement the simulation test for the multi-objective optimization of the system, and the multi-objective particle swarm algorithm is used to optimize the two objectives of the delay module and the energy consumption module. The simulation test continues to reflect the results of the relationship between energy consumption and delay by bottleneck energy and maximum delay, as seen in Fig. 8.

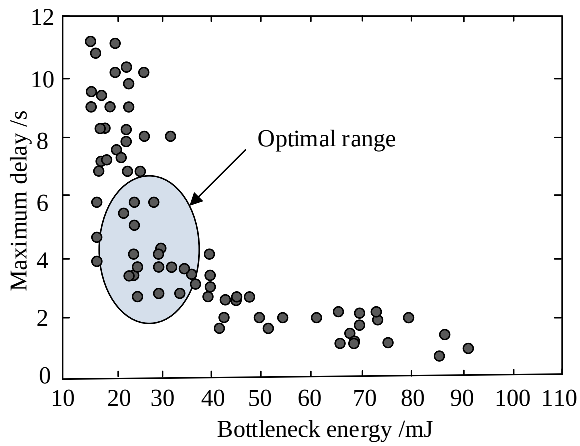


Fig. 8. Results of the relationship between time delay and energy consumption under particle swarm optimization algorithm

The results of the relationship between time delay and energy consumption under the particle swarm algorithm are shown in Fig. 8. The time delay and energy consumption are two conflicting objectives that show an inverse proportional relationship. Optimization of the two models of time delay and energy consumption by the multi-objective particle swarm algorithm does not search for the optimal solution set that matches the objective function. However, in the simulation test, the blue area of the image ellipse is the optimal solution range of the objective. In order to find the best working conditions parameters that meet the system work, it is necessary to further analyze the optimal solution set obtained by the particle swarm algorithm, and through the weighting process, the weights achieve the switching effect of working mode, and the normalization process is carried out to get the results corresponding to the parameters of time delay and energy consumption under different weights ω , as shown in Table I.

TABLE I. SHOWS THE RESULTS OF TIME DELAY, ENERGY CONSUMPTION AND THE RELATIONSHIP BETWEEN PARAMETERS

Weight	Transmission distance/km	Signal bandwidth/KHz	System energy consumption/mJ	System delay/s
0.1	1	125	12	9.13
0.2	1.5	125	14	5.01
0.3	2	125	25	2.72
0.4	2.6	125	32	0.91
0.5	3	125	48	0.53
0.6	3.5	250	53	0.47
0.7	4	125	59	0.36
0.8	4.5	250	71	0.26
0.9	5	250	74	0.16

Table I shows the results of time delay, energy consumption and the relationship between each parameter. At the weight of $\omega = 0.2$, the system has the best balance of energy consumption and time delay, which can be used as the monitoring mode of cable detection system at this moment, and at $\omega = 0.5$, the system has low delay and good signal transmission distance, which can be used as the warning mode of the system. At $\omega = 0.9$, the system has the best delay performance and transmission distance, 0.16s and 5km, respectively, as maintenance mode. The proposed wireless inspection system can effectively meet the work content requirements of cable repair and maintenance, and the energy consumption is reduced by 81% compared with the traditional wired inspection system.

Packet loss rate is an important parameter to evaluate the quality of communication transmission, mainly refers to the ratio of the amount of data lost during the transmission and exchange process of communication data to the actual sent communication data. A higher packet loss rate means a worse quality of communication data transmission. Therefore, in order to ensure that the system communication has a high quality, it is necessary to consider the communication packet loss rate. The packet loss rate is related to the communication distance, packet length and the transmitting frequency of the signal base station. Therefore, the packet loss rate was tested under three detection models, and 2000 16-bit byte packets were sent simultaneously in the three operating modes of the system, as shown in Table II.

TABLE II. SHOWS THE RESULTS OF PACKET LOSS RATE OF THE SYSTEM UNDER VARIOUS WORKING MODES

System working mode	Total transmitted packets	Communication distance /kg	Packet loss rate	Time delay/s
Early warning mode	1950	3	<3%	0.51
	1956	3	<3%	0.52
	1972	3	<3%	0.53
Monitoring mode	1901	1.5	<5%	5.12
	1908	1.5	<5%	5.16
	1910	1.5	<5%	5.19
Maintenance mode	1980	5	<1%	0.16
	1982	5	<1%	1.17
	1986	5	<1%	1.17

Table II shows the results of the packet loss rate situation of the system under each operating mode. The packet loss rate of the system under all three modes tested is controlled within 5%, which meets the relevant requirements of packet loss rate for China's power grid communication construction. In the early warning mode and maintenance mode, the packet loss rate is significantly lower than that in the monitoring mode, mainly because the communication system needs to maintain high frequency signal bandwidth to meet the requirements of low delay in communication in both working environments, and therefore the energy consumption in both working modes is relatively high. It can be concluded that the three working modes will be freely switched according to the environmental conditions where the high-voltage cable is located, maintaining the basic communication and low energy consumption detection requirements in the detection mode, and switching to the warning or maintenance mode under the harsh environmental conditions to guarantee the detection of the system and the exchange of communication data. Finally, compare the wireless cable detection technology studied with the traditional wired cable detection technology, as shown in Table III.

TABLE III. COMPARISON BETWEEN THE PROPOSED CABLE DETECTION TECHNOLOGY AND TRADITIONAL CABLE DETECTION TECHNOLOGY

Comparison type	Traditional cable detection technology	The proposed wireless cable detection technology
Scope of application	Only applicable to places where wired equipment can be built	It can adapt to most places, including high mountains, cold and hot areas
Unit energy consumption	higher	moderate
Self energy level system	Manual automatic control	Intelligent/manual control
Construction requirements	High construction difficulty, not suitable for all places	It can be built in most places and mobile communication coverage areas
Maintain	High maintenance difficulty and cost	Lower overall maintenance and cost
Check the accuracy	No time interval monitoring, and manual monitoring is required for special environment	Adapt to full time and assist manual monitoring, with higher detection accuracy

Table III is a comparison table between the proposed scheme and the traditional cable detection technology. It can be seen from the information in the table that compared with the traditional cable detection technology, the proposed wireless cable detection technology has obvious advantages in the scope of application, construction cost and detectability. At the same time, the cable detection technology based on wireless technology has more advantages than the traditional cable detection technology in terms of system updating and technology upgrading, and is more in line with the construction requirements of the existing smart grid.

IV. CONCLUSION

High-voltage transmission circuit is an important part of the grid system, and the reliability of high-voltage cables directly affects the safe operation of the grid system. Along with the development of interest communication technology, advanced wireless communication technology will provide important technical support for the development of smart grid. To address the problem of insufficient maintenance of traditional wired high-voltage cable detection system in remote mountainous areas, a high-voltage cable path state monitoring scheme using LoRa technology is proposed based on wireless sensor network. And three monitoring modes are proposed according to the fault types of high-voltage cables; the balance of energy consumption and delay of cable communication is achieved by multi-objective particle swarm algorithm. The experimental results show that the system delay, energy consumption and signal bandwidth are different under different weight parameters, and the overall effect of the cable detection system can be improved by selecting the appropriate cable detection model according to the different weight of the cable detection area. When the weight is $\omega = 0.5$, the system has low delay and good signal transmission distance. At this moment, the system adopts warning mode. At the weight $\omega = 0.9$, the system signal and transmission distance are in the best state. At this moment, the system uses maintenance fuzziness for routine cable maintenance. Compared with the traditional cable detection system, the overall energy consumption of the proposed system is reduced by 81%. Finally, in the communication data transmission test, the packet loss rate of data transmission under the three working modes is less than 5%. The system reasonably selects the working model according to the environment and ensures that the system has excellent data signal transmission performance, and meets the cable fault detection requirements. Compared with the traditional wired cable detection technology, the wireless cable detection technology consumes less energy and can adapt to a more demanding detection environment. At the same time, the system uses neural algorithms to balance the system parameters, which makes the overall detection performance more stable and better. Therefore, the high voltage cable detection technology based on wireless sensor network meets the requirements of smart grid construction and ensures the timely and effective detection of high voltage cables. However, there are also shortcomings in this research. The wireless cable detection technology only considers the impact of mountain environment on technology. In the actual technology construction, more demanding environment needs to be considered. At the same time, the wireless technology needs to consider the regional communication coverage during the construction process, and the combination of wired and wireless technology is the most effective method at present.

REFERENCES

[1] H. Xu, "Fastest adaptive estimation algorithms for topological structure errors in smart grid networks", *Computer Communications*, vol. 160, pp. 197-203, 2020,

[2] O. Elsayed, J. Zarate-Roldan, A. Abuellil, et al. "Highly Linear Low-Power Wireless RF Receiver for WSN", *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 27(5), pp. 1007-1016, 2019.

[3] J. C. Wang, M. S. Liao, Y. C. Lee, et al. "On enhancing the energy harvesting performance of the photovoltaic modules using an automatic cooling system and assessing its economic benefits of mitigating greenhouse effects on the environment", *Journal of Power Sources*, vol. 376, pp. 55-65, 2018.

[4] S. Chamanian, S. Baghaee, H. Ulusan, et al. "Implementation of Energy-Neutral Operation on Vibration Energy Harvesting WSN", *IEEE sensors journal*, vol. 19(8), pp. 3092-3099, 2019.

[5] Kumar, Sudhir. "Compartmental Modeling of Opportunistic Signals for Energy Efficient Optimal Clustering in WSN", *IEEE Communications Letters*, vol. 22, pp. 173-176, 2018.

[6] G. Tabella, N. Paltrinieri, V. Cozzani, et al. "Wireless Sensor Networks for Detection and Localization of Subsea Oil Leakages", *IEEE Sensors Journal*, vol. 21, pp. 10890-10904, 2021.

[7] A. Xf, B. Gf, B. Pp, et al. "Environment-fusion multipath routing protocol for wireless sensor networks", *Information Fusion*, 2020, vol.

53, pp. 4-19, 2020.

[8] B. Hu, Z. Sun, J. Liu, “Distributed time synchronization algorithm based on sequential belief propagation in wireless sensor networks”, *Computer Communications*, vol. 176, pp. 119-127, 2021.

[9] E. H. Houssein, M. R. Saad, K. Hussain, et al. “Optimal Sink Node Placement in Large Scale Wireless Sensor Networks Based on Harris’ Hawk Optimization Algorithm”, *IEEE Access,* vol. 8(99), pp. 19381-19397, 2020.

[10] F. Lin, W. Dai, W. Li, et al. “A Framework of Priority-Aware Packet Transmission Scheduling in Cluster-Based Industrial Wireless Sensor Networks”, *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 5596-5606, 2020.

[11] C. Zhan, Y. Zeng. “Aerial-Ground Cost Tradeoff for Multi-UAV-Enabled Data Collection in Wireless Sensor Networks”, *IEEE Transactions on Communications*, vol. 68(3), pp. 1937-1950, 2020.

[12] W. Wang, G. Tong, “Multi-path unequal clustering protocol based on ant colony algorithm in wireless sensor networks”, *IET Networks*, vol. 9(2), pp. 56-63, 2020.

[13] A. Verma, S. Kumar, P. R. Gautam, et al. “Fuzzy Logic Based Effective Clustering of Homogeneous Wireless Sensor Networks for Mobile Sink”, *IEEE Sensors Journal*, vol. 20(10), pp. 5615-5623, 2020.

[14] G. Jaber, R. Kacimi, LA. Grieco, et al. “An adaptive duty-cycle mechanism for energy efficient wireless sensor networks, based on information centric networking design”, *Wireless Networks*, 2020, vol. 26, pp. 791-805, 2020.

[15] W. Wang, N. Yu, Y. Gao, et al. “Safe Off-Policy Deep Reinforcement Learning Algorithm for Volt-VAR Control in Power Distribution Systems”, *IEEE Transactions on Smart Grid*, vol. 11, pp. 3008-3018, 2020.

[16] E. Pan, P. Dong, W. Long, et al. “Provincial Grid Investment Scale Forecasting Based on MLR and RBF Neural Network”, *Mathematical Problems in Engineering*, pp. 1-12, 2019.

[17] H. A. Ying, D. A. Lei, B. Jl, et al. “Power forecasting-based coordination dispatch of PV power generation and electric vehicles charging in microgrid”, *Renewable Energy*, vol. 155, pp. 1191-1210, 2020.

[18] S. Bhamidipati, K. J. Kim, H. Sun, et al. “Artificial-Intelligence-Based Distributed Belief Propagation and Recurrent Neural Network Algorithm for Wide-Area Monitoring Systems”, *IEEE Network*, vol. 34, pp. 64-72, 2020.

[19] P. Iliadis, S. Chapaloglou, A. V. Nesiadis, et al. “Smart energy management algorithm for load smoothing and peak shaving based on load forecasting of an island’s power system”, *Applied Energy,* vol. 238, pp. 627-642, 2019.

[20] R. K. Jyotheeswara, N. Sudhakar, “A new RBFN based MPPT controller for grid-connected PEMFC system with high step-up three-phase IBC”, *International Journal of Hydrogen Energy*, vol. 43, pp. 17835-17848, 2018.

Research on the Application of Improved Decision Tree Algorithm based on Information Entropy in the Financial Management of Colleges and Universities

Huirong Zhao
School of Economics and Management, Jiangsu College of Engineering and Technology, Nantong, 226000, China

***Abstract*—In the era of information technology, the work relies on information technology to generate a huge amount of data and information. Among them, the financial data information of universities is growing exponentially, and the manual method of organizing data and extracting key information can no longer meet the requirements of financial data management of universities. Taking the financial management of higher education institutions as an example, it is difficult to grasp the progress of financial budget execution with frequent and complicated daily expenditure and income problems, and then it is difficult to execute correct decisions in the management. The study uses information entropy as the decision basis of decision tree in the financial management of higher education institutions. The higher the value of information entropy generated in financial management, the higher the prediction accuracy of the decision tree. The metric calculation method is introduced to obtain the information entropy as well as the information gain rate to predict the likelihood of problematic events. The study validates the performance of the improved decision tree with a dataset that achieves a maximum accuracy of 95% in the experiment. With the higher prediction accuracy, for the university financial management system, a decision tree for financial warning is established and the link between the current month's financial expenditure and the warning mechanism is analyzed, and finally the two common decision tree algorithms, (Iterative Dichotomiser3，ID3) ID3 and (Classification and regression tree，CART) CART, are compared with the algorithm proposed in the study. The mean square error and the sum of squared error metrics are used to conclude that the algorithm proposed in the study has better performance. By improving the existing decision tree algorithm, the study proposes a decision tree model based on information entropy, which aims to help decision makers to quickly and accurately distill relevant data and make correct decisions in a large amount of information data for more rational financial management.**

Keywords*—Information entropy; financial management; decision tree; information gain rate; C4.5 algorithm; early warning structure

I. INTRODUCTION

Financial management in colleges and universities is one of the core management works of higher education institutions, and also an important guarantee for normal daily work of institutions. As the enrollment scale of colleges and universities is expanding, the financial data of colleges and universities also usher in an explosive growth. How to mine these data information and fully apply the useful information in the data for effective college financial management and financial decision making is the focus of current intelligent algorithms in the field of college management research [1]. Since the society has entered the information age, computer technology and information technology have been developed rapidly. The development of computer information technology has brought a large amount of data information. The large amount of data can provide more reliable basis in extracting key information, but at the same time the redundant information mixed with it also brings redundant workload.

In the present time when the intelligent technology is popular, the information extraction technology is more in the form of automation, so the construction and improvement of financial information system has become the key research content in the management of universities. With the continuous development and improvement of financial management information systems in universities, the managers' use of traditional financial statements and manual decision-making solutions can no longer meet the needs of university financial management informatization, not to mention the development goals of data and information in the intelligent information era [2]. With the increasing requirements of university management for management effectiveness, reasonable decision quality and decision speed become crucial. The use of advanced technology and intelligent decision-making tools can make more effective management decisions and solve various problems in the financial management of colleges and universities. The financial information system of colleges and universities has a large amount of financial information. The financial information contains many important data, and the relevant information extracted from the data can be used as the judgment basis for the financial management decision of colleges and universities. The current information management of university finance is far from meeting the development requirements at this stage, and the management of finance only stays on the surface basis, focusing only on the accounting situation of financial data and the supervision of data. With the speed of digital campus construction, the needs of university development and financial data are accumulated in large quantities, and the managers need not only to achieve the most basic financial management, but also to control the risks of university development to the minimum through financial management. Therefore, using intelligent algorithms for management decisions can not only provide new management directions for university financial management, but also predict other types of financial problems.

At present, many data processing techniques and intelligent algorithms have been applied to financial management and risk decision making in universities. For example, data warehouse technology, data mining technology, data analysis technology, etc. The study introduces digital and information technology into the financial management of universities to get the improved algorithm of decision tree about information entropy. The improved algorithm can reasonably use a large amount of data generated in the financial system of universities, extract key information for the sustainable development of universities, provide a scientific theoretical basis for the construction of universities, and then control the risks in the financial management of universities for the purpose of financial budget progress management as well as financial early warning management [3]. The research content firstly uses information entropy as the decision basis of decision tree in financial management of universities, and proposes the decision tree algorithm based on information entropy and analyzes it. The higher the value of information entropy generated in financial management indicates the higher the prediction accuracy of the decision tree. Then, the metric is introduced, and the information entropy and the information gain rate are calculated, and a model based on the improved decision tree algorithm is proposed for the application of college financial management, and the likelihood of problematic events is predicted using the model. Finally, the performance of the constructed improved decision tree model is verified by the data set.

II. Related Work

The construction of higher education institutions is the basis of social progress and development. In order to ensure the normal operation and development of colleges and universities, many scholars have conducted relevant research on the management of colleges and universities. Priya E (Priya E 20210 [4] proposed a breast cancer diagnosis model based on deep segmentation of residual network, and applied it in the initial detection stage of breast cancer. In order to detect and classify breast images, the model adopts decision tree classifier for operation. It is proved by relevant experiments that the model has excellent classification performance, and the accuracy reaches 98.86%. Al C (Al C 2021) [5] classifies the students' tendency to choose TVET after completing their studies through multiple decision tree models, and constructs different decision tree algorithm types. The experimental results show that the decision tree has a low misclassification rate in TVET classification. Researchers such as Mao L (Mao L et al. 2021) [6] conducted an evaluation and analysis of entrepreneurship education in higher education institutions, improved and optimized the algorithm on the basis of decision tree and fuzzy algorithm, and constructed an efficient entrepreneurship education evaluation system. Through empirical research, the evaluation system established in the study has a certain degree of effect in practical application. S Ziweritin team (S Ziweritin et al. 2021) [7] through the simulation experiment, the decision tree is trained and tested, and the accuracy of the model reaches 99.86%. Mabuni D and Babu S A (Mabuni D and Babu S A 2021) [8] found that the classification effects obtained in smaller datasets were arranged to be reliable, and therefore performed cross-validation with multiple random folds on the same dataset. The verification method is also tested experimentally, and the experimental results show that the classification accuracy of the decision tree with prime number folding is higher than that of the existing decision tree classification.

Ahmad N (Ahmad N et al. 2019) [9] collected public university leaders' views on the overall financial management and sustainability of institutions through questionnaires. According to the survey results, the leaders interviewed believe that the key to the sustainability of financial management is to make use of university resources. The financial management of the school takes the basic activities of the school as the management, which is of great significance to the financial management of public universities. Researchers such as Asandimitra N (Asandimitra N et al. 2020) [10] designed a conclusive causal relationship study, conducted multiple regression analysis on the research results through relevant statistical software, and found that teachers' financial management behavior has an important relationship with teachers' own financial knowledge and financial literacy. Cultivating teachers' financial self-efficacy is of great help to school financial management. The Mohammad M team (Mohammad M et al. 2021) [11] used the mean and standard deviation of the collected data as a measure of the importance of the research question in their research on the quality and impact of AI on classified hotel financial statements. Through the survey results, it is found that the hotel application AI performs linear regression analysis on financial information, analyzes the important factors affecting financial management, and plays an important role in the sustainability of financial management. Malikov VV scholars (Malikov V V 2020) [12] constructed the algorithm of financial management system in enterprise management. The system first determines the overall financial goals of the enterprise, allocates the corresponding financial management subsystems, and formulates corresponding financial management rules. In practical application, the algorithm of the system successfully avoids the problem of cash gap and optimizes the relationship between creditors. Researchers such as UJ A (UJ A et al. 2020) [13] have developed some data classification models for financial management decision-making, which include the feature selection stage and data classification stage of ant colony feature optimization. Through the validation of multiple algorithms in the same dataset, the proposed model has better performance than other algorithms.

In summary, decision trees are widely used in classification, forecasting and other fields. Financial management needs to make predictive behaviors for the future. However, there are few and immature studies on the combination of the two. Therefore, the research combines the two. A relatively complete research is conducted to help colleges and universities become more scientific in financial management.

III. Application and Improvement of Decision Tree Algorithm in Financial Management

A. Improvement Analysis of Decision Tree Algorithm based on Information Entropy

Decision tree algorithm can classify data purposefully, and find and extract valuable information from the data, thus helping decision makers to provide a choice path [14]. Based on the above advantages, decision tree algorithm is widely used in prediction models. In data mining technology, the good tree structure formed by the decision tree can highly summarize the sample data and correctly identify the old samples or different types of samples. Compared with other classification algorithms, decision tree has a classification rule that is easier to understand. In the classification task, the algorithm runs faster and has higher classification accuracy.

Common decision tree algorithms include ID3 algorithm, C4.5 algorithm and CART algorithm. In the late 1970s, J Ross Quinlan proposed the ID3 algorithm, which aims to reduce the depth of the decision tree and simplify the operation steps of the model. However, ID3 algorithm ignores the study of the number of leaves while reducing the depth of the tree. Therefore, C4.5 algorithm is proposed to improve ID3 algorithm. As a decision tree algorithm based on information entropy, C4.5 algorithm has the advantage that it can use the information gain rate to replace the information entropy as the selection standard, avoiding the defects of ID3 algorithm. As an improved decision tree algorithm, the basic process of the algorithm is: first, select the training set as the root node, and select the appropriate criteria to select the splitting attributes. Secondly, according to the difference of split attribute values, the training set is divided into several subsets and used as the first sub node of the root node. Then, treat these sub nodes as root nodes and repeat the above steps until the final classification conditions are met.

In practical application, when there are many data items, more complex decision trees will be generated, which will lead to uneven data and even loss of attributes, affecting the operation efficiency and classification accuracy of the algorithm [15]. In view of the shortcomings of decision tree algorithm, C4.5 algorithm is proposed to be used in the decision tree. C4.5 algorithm is an improved decision tree algorithm, which mainly operates on the basis of information entropy. When dealing with continuous lost data, tree pruning technology is introduced. This technology not only simplifies decision-making, but also improves classification accuracy. In addition, the research attempts to introduce a measurement method into information entropy. This method can be used for data reference at the same time when classifying, and finally an improved decision tree algorithm based on measurement is obtained. The flow chart of the improved algorithm's decision tree is shown in Fig. 1.

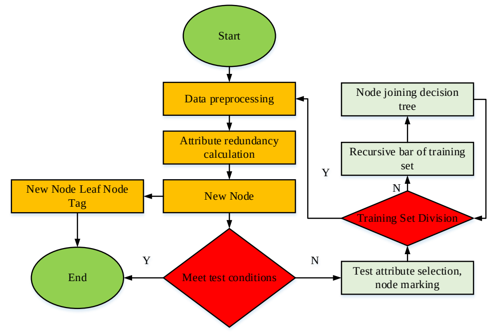


Fig. 1. Decision tree flow chart of improved algorithm

The decision tree model established by introducing the measurement method is essentially a recursive algorithm, which forms the branch conditions of the decision tree and adopts the threshold method. Assuming that $T=\{t_i\}$ the set of $A=\{a_i\}$ samples is taken as the feature space of the samples, the total number of samples is expressed by c, and the attributes of the feature space are expressed by m. The power space of the feature space is represented by $B=\{b|b\in A\}$, the misclassification threshold in the model is represented by β_e, and the cross misclassification threshold is represented by β_c. For the superset, it includes samples of multiple categories. If the sample attribute set threshold value in the superset x is the smallest, and the distance from the other category values in the typical Mahalanobis distance is the smallest, the sample will be classified into this category. Assuming that the sample is misclassified to other samples, and the proportion is greater than the cross misclassification, the two samples will be considered to be the same class and continue to be classified in the lower layer. The classification strategy adopted when the above attribute set is not found is the best classification situation of the model. When all samples are classified or can no longer be classified, the model classification process ends. In the model, the randomness of information entropy is used as a standard to select model samples to test attributes [16]. In the process of building a decision tree, an attribute is required to divide the data, so that the data class values in the child nodes are basically the same. When the data class values in the node are unevenly distributed, it means that the entropy of the node is small. The definition of entropy is shown in formula (1).

$$Entropy(p_1,p_2,\cdots p_n)=p_1\log_{2p1}-p_2\log_{2p2}-\cdots-p_n\log_{2pn} \tag{1}$$

In formula (1), "-" indicates that the logarithm of the score is negative; p represents the fraction in the definition of entropy. In fact, the entropy is positive, as shown in formula (2) in general.

$$Info\{[C_1, C_2, \cdots, C_n]\} = entropy[E_s] \quad (2)$$

In formula (2), it is the expression of the training set E_s; it is used to represent the samples in the training set C. All training sets are grouped, and the spatial features are used to $entropy[E_s]$ reduce the value. Assume that the training set is set with a new expected amount of information, and its expression is shown in formula (3).

$$new\_entropy(E_s, A) = \sum i \notin value_{(A)} = [|E_s| / |E_s|] entropy[E_s] \quad (3)$$

The reduced expectation value of the training set is the information gain of the feature space relative to the training set. If it is beneficial to the training set, the larger attribute of the information gain needs to be removed. The specific expression is shown in formula (4).

$$Gain[E_s, A] = entropy[E_s] - new\_entropy(E_s, A) \quad (4)$$

Equation (4) $Gain[E_s, A]$ is the expression of information gain. Information entropy is the average amount of information obtained after excluding redundant information, which is used to represent the degree of confusion of information. The calculation formula of information entropy is shown in Equation (5).

$$I(S_1, S_2, \cdots, S_m) = -\sum_{i=1}^{m} P_{i \log_2 (P_i)} \quad (5)$$

In formula (5), it is used to represent the sample set S, which is used to represent P_i the probability of any sample in $P_i = S_i / S$, and C_i. The measure of the amount of information is the size of the information gain. When all attributes are classified, the value of the information gain is first calculated, and then the node selected for the construction of the decision tree is determined. big. Therefore, the feature space branching obtains the information gain as shown in Equation (6).

$$Gain(A) = I(s_1, s_2, \cdots, s_m) - E(A) \quad (6)$$

In formula (6), it $I(s_1, s_2, \cdots, s_m)$ represents the expected information entropy of a given sample, and the calculation method of sample expectation is shown in formula (7).

$$E(A) = -\sum_{j=1}^{v} \frac{s_{1j} + \cdots + s_{mj}}{s} I(s_{1j}, s_{2j}, \cdots, s_{mj}) \quad (7)$$

When calculating the information gain, it mainly depends on the attribute value, but in special cases, such calculation is also meaningless. Therefore, the division of nodes plays an important role in the construction of decision trees. The C4.5 algorithm optimizes the node division by means of the information gain rate, that is, the ratio between the information gain and the information entropy. When comparing, the amount of information of sample attributes is not the total amount. When the information gain is divided into subsets, there will be a certain degree of deviation in the value of its variables, which can be reduced by formula (8).

$$SplitInfo(S, v) = \sum_{i=1}^{m} \frac{|s_1|}{|s|} \times \log_2 \frac{|s_1|}{|s|} \quad (8)$$

Therefore, the gain rate expression is specifically as shown in Equation (9).

$$GrainRatio = -\frac{Grain(S, v)}{SplitInfo(S, v)} \quad (9)$$

By introducing information entropy and measurement methods, the C4.5 algorithm assumes a data set T in which different categories are represented by sets C, selects an attribute to divide the data set into multiple subsets, and the values of the attributes are all equal. If they are not the same, the dataset is divided into the same number of subsets. The probability of category selection occurring is shown in formula (10).

$$P(C_i) = |C_j| / |T| \quad (10)$$

Because the instances of the subset all take the value of the attribute, the probability of the occurrence of the attribute value and the probability of having the category condition are shown in formula (11).

$$\begin{cases} P(v_i) = T_i / T \\ P(C_j | V_j) = c_{jv} / |T_i| \end{cases} \quad (11)$$

In formula (11), the probability of occurrence $P(C_j | V_j)$ with C_j category conditions is expressed, and the probability calculation formula can be obtained to further obtain the calculation of category information entropy and category condition information entropy, as shown in formula (12).

$$\begin{cases} Info(C) = -\sum_j p(C_j) \log p(C_j) = -\sum_{j=1}^{k} \frac{P(C_i)}{|T|} \log \frac{P(C_i)}{|T|} = Info(T) \\ Info(\frac{c}{v}) = -\sum_j p(v_j) \sum_1 p(\frac{c_j}{v_j}) \log p(\frac{c_j}{v_j}) = -\sum_{i=1}^{n} \frac{|T_i|}{|V_i|} Info(T_i) = Info(T) \end{cases} \quad (12)$$

Finally, the information entropy of the attribute is obtained according to formula (8), as shown in formula (13).

$$Info(V) = -\sum_i p(v_i) \log p(v_i) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \log \frac{|T_i|}{|T|} = split\_Info(v) \quad (13)$$

The attribute information gain rate is obtained according to formula (9), and the specific expression is shown in formula (14).

$$gai\_ration(V) = n\frac{gain(v)}{split\_Info(v)} \quad (14)$$

B. The Application of the Improved Decision Tree Algorithm in the Financial Management Model of Colleges and Universities

The above research method introduces the decision tree algorithm based on information entropy, and obtains the specific calculation method of information gain rate. This calculation method is very important in practical application. The size of the decision tree can be judged through the calculation of information entropy. Part 2.2 will give an overall description of the financial management of colleges and universities, and apply the decision tree to it. In college financial management, college budget management and budget performance management play an important role. The budget management of colleges and universities can make detailed use plans for financial income and use according to financial and teaching contents [17]. The flow chart of budget performance management is shown in Fig. 2.

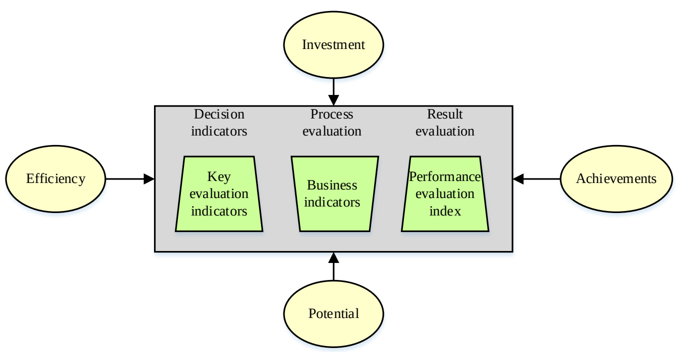


Fig. 2. Flow chart of budget performance management

The data in the process of decision tree calculation needs to be processed, so it is necessary to select a more appropriate algorithm to build the decision tree model. After the decision tree is built, it is pruned, that is, the decision tree is optimized. The main purpose is to eliminate isolated nodes and noise, cut meaningless branches, and make the decision tree have the ability to correctly classify [18-20]. The pruning idea in the algorithm firstly sets the node of the tree as the number t of $n(t)$ samples that reach the node; to $e(t)$ represent the number of samples that can reach the node but does not belong to the node; the ratio of error samples at the node is expressed as $r(t)$. The probability of repeating the experiment multiple times is the probability of occurrence $e(t)$, and the confidence interval of the wrong sample ratio can also be obtained. The confidence interval is expressed in CF, and the value of the confidence interval can control the degree of pruning. The higher the value of the confidence interval, the lower the degree of pruning. In C4.5 algorithm, the default value of confidence interval is 0.25, and the samples obey binomial distribution. The calculation process of the decision tree algorithm is integrated and calculated in the Clementine platform. The decision tree 4.5 algorithm is used to extract, process and mine the database data, and then establish a decision tree of financial budget to achieve early warning analysis of financial budget. The overall framework of early warning analysis is shown in Fig. 3.

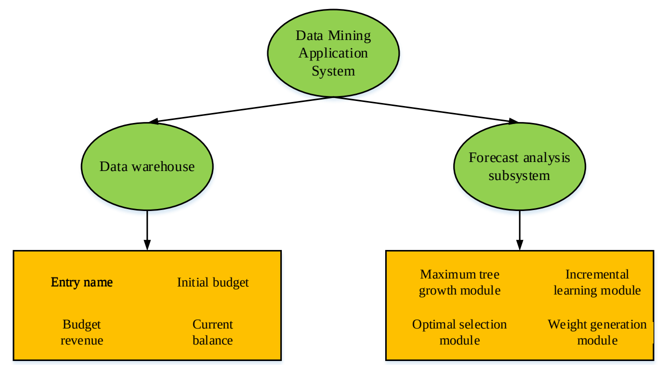


Fig. 3. Overall framework of early warning analysis

The data warehouse in Fig. 3 includes a lot of information, such as information of various departments of the university, opening balance, income and expenditure data, balance data, and so on. There are many problems in many data, such as clutter, duplication and incompleteness, so cleaning the data is a necessary step of the system. For data processing, there are methods of filling and deleting. When multiple attributes are missing in the sample, and the number of missing attribute values is more than one-tenth of the total number of attributes, the sample data will be deleted; when the mean values of the sample attributes are the same, then fill it. In the financial management early warning analysis sub-structure system, it also includes a maximum tree growth module, an incremental learning module, a pruning optimization module, and a weight generation module. Its early warning program first selects the said information from the database for preprocessing, and saves the data of the financial budget report to the data warehouse; the data warehouse is an organization method for storing data. Then, the data is classified by the C4.5 algorithms to obtain new knowledge; the new knowledge generated includes the weight generation module, which uses the weight distribution method to assign corresponding weights to each node on the decision tree value, and an incremental learning module. The incremental learning module combines incremental learning with the C4.5 algorithms, and uses incremental data and learning methods to modify the generated decision tree; output to help decision makers make corresponding countermeasures. The details are shown in Fig. 4.

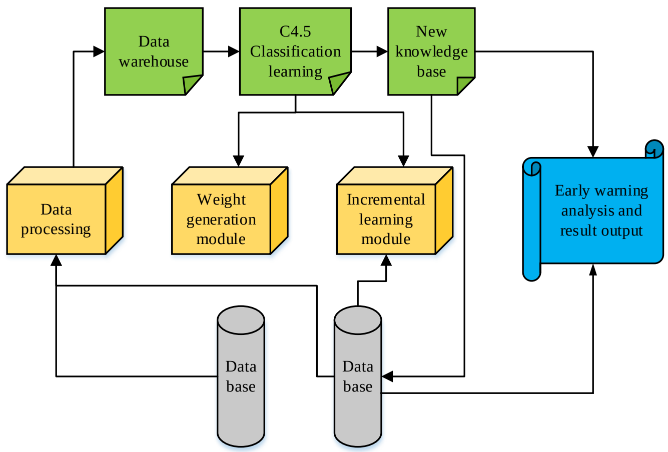


Fig. 4. Data flow chart of early warning analysis system

The construction of financial budget data warehouse is established from the three-level model of determining data source, data preprocessing and data warehouse. The data sources are all from the financial budget database of universities, and the data warehouse is studied according to the theme of decision tree; data preprocessing cleans the data, eliminates noise and irrelevant attributes, and converts the data; the three-level model of the data warehouse is established. The three-level model of the data warehouse, the first is the conceptual model, which determines the boundary of the financial budget model, extracts the data information required by the financial budget, determines the main performance indicators of the model, and defines the relevant data dimensions. The second is the logical model. The logical model is designed according to the data table division, granularity, financial budget execution progress and relational model. The third is the establishment of the physical model, the process of designing the physical database with the data organization form stored in the data warehouse as the main body.

IV. Financial Management Application Analysis of Improved Decision Tree Algorithm Based on Information Entropy

A. *Analysis of the Application Effect of Financial Management based on Improved Algorithm*

The research selects representative universities as the experimental object, and analyzes the data according to the accumulated data of the financial department of universities. In order to make the data mining information have the characteristics of high quality and high informatization, the selected samples should be small and representative. Table I shows the experimental data of mining 2022 college financial data after preprocessing.

TABLE I. College 2021 Financial Data Preprocessing Results

Project	Contrast ratio	Uniformity	Relevance	Entropy
T1	379.0191	19.3456	-0.0001	199.2554
	369.2883	19.2463	0.0000	209.5737
	249.2672	19.6178	-0.0001	179.8338
	489.4266	29.6728	0.0000	239.8373
T2	29.3567	9.2728	-0.0038	39.2628
	19.3638	9.2627	-0.0038	39.2727
	19.7378	9.2727	-0.3385	39.8262
	19.3665	9.2636	-0.4859	39.6528
T3	1549.2669	0.2627	-0.5049	129.3565
	1702.6229	0.2739	-0.4860	119.3743
	2489.2529	0.7283	-0.5050	209.2728
	1289.2536	0.2738	-0.4960	109.2727
T4	1039.2738	1.2738	-0.4950	18.7279
	501.2738	2.3730	-0.4951	18.2829
	203.3637	3.3738	-0.4489	8.1930
	929.2727	3.3829	-0.4960	10.8274
T5	1039.2627	1.2737	-0.1111	29.2833
	282.2719	1.1828	-0.5050	36.8189
	282.2728	0.8119	-0.5105	31.8279
	397.8192	2.2829	-0.1111	31.2828

The research uses the information entropy reduction of the C4.5 algorithms to determine the optimal branch variable and segmentation threshold of the decision tree. In the world of information data, the size of the entropy value can represent how much information is transmitted. If the entropy value is higher, it means that the transmission information is more; on the contrary, the transmission information is less, the information entropy value is smaller. In financial management, it can be divided by payment method to indicate the entropy value of the event. In ordinary budget indicators, the payment methods are divided into two types: direct payment and authorized payment, and the two payment methods are guaranteed to have equal probability of opportunity, which makes it difficult to predict the result of the next financial payment method, and the two results are mutually exclusive. Independently, the entropy in this form is one bit. When the budget index amount is relatively large and the university only allows direct payment, the entropy value of the event division method is 0. Algorithm C4.5 takes the decreasing speed of information entropy as the basis for judging the progress of budget execution, and the decision tree for getting early warning of budget execution progress is shown in Fig. 4.

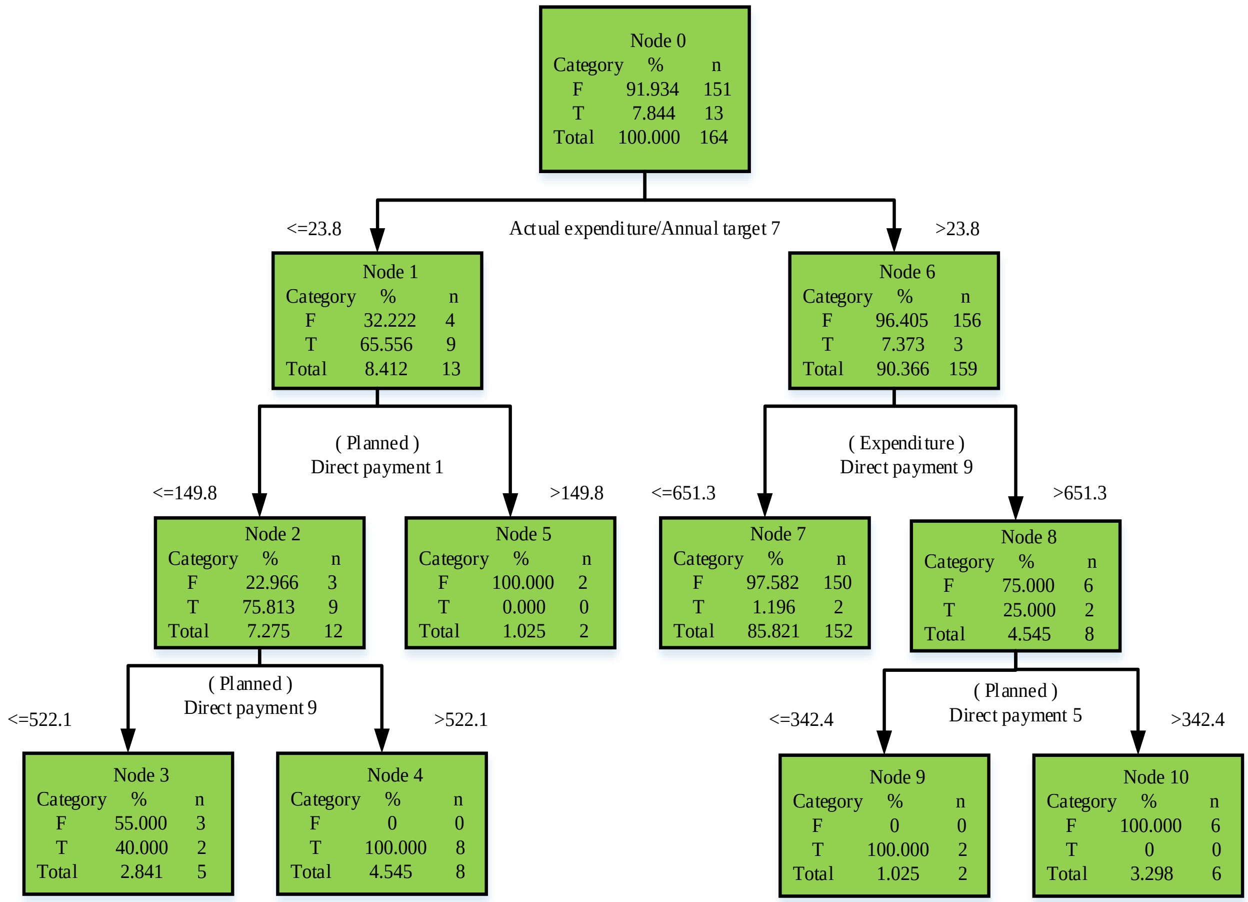


Fig. 5. Decision tree for budget implementation progress alert

In Fig. 5, 23.8% is the cut-off point for the payment method of the decision tree. A Situation: if the budget implementation progress of the project in July is not more than 23.8%, and the direct payment amount in January is not less than 150000 yuan, no budget is required; B Situation: For the same project, if the budget implementation progress is no more than 23.8%, the direct payment amount in January is less than 150000 yuan, and the authorized payment amount in September is no less than 522000 yuan. The decision tree needs to give an alert; In case b, when the authorized payment amount in September is less than 522000 yuan, it is necessary to judge according to the authorized payment amount in January. If it is not less than 8800 yuan, there is no need for early warning. On the contrary, if it is less than 8800 yuan, there is a need for early warning. By analogy, we can pay more attention to the financial budget of the current month according to the results and budget implementation progress. The study obtains an early warning decision tree about the execution progress of the budget, and puts the data in the three test sets into the decision tree. The purpose is to compare the prediction result with the real value to verify the prediction accuracy of the decision tree in financial management. A decision tree was built on some data in test set 2, and the budgets of 100 projects were predicted and analyzed. In the judgment of the decision tree, the predicted results of 95 projects were the same as the actual results, and the predicted results of only five projects were the same as the actual results. Therefore, the decision tree built on the basis of information entropy has a higher prediction effect. Different test sets are used to verify the prediction accuracy of decision tree, as shown in Fig. 6.

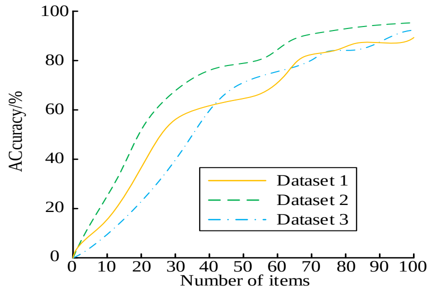


Fig. 6. Verification of prediction accuracy of decision tree with different test sets

Through the verification of the prediction accuracy of the decision tree from three data sets, it is found that the decision tree based on information entropy has a higher prediction accuracy, which verifies that the decision tree based on information entropy is applicable to a variety of data. Therefore, the analysis results of the decision tree constructed in Fig. 5, which is used in colleges and universities, are reliable.

B. *Performance Comparison of Improved Decision Tree Algorithm based on Information Entropy*

The experiment uses the ID3 algorithm, the CART algorithm and the improved C4.5 algorithms to compare to verify the superiority of the algorithm performance. The performance experiment of the algorithm uses the relevant error indicators as the judgment basis. The mean square error results of the three algorithms in the model are shown in Fig. 6.

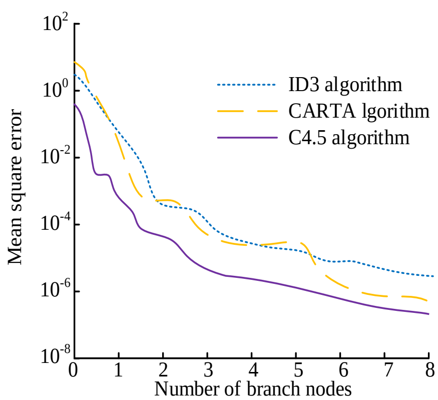


Fig. 7. Mean square error results of three algorithms in the model

From Fig. 7 it is seen that the mean square error of the C4.5 algorithms is lower among the three models, and when the number of branch nodes is small, the mean square error of the algorithm is still higher, and the mean square error gradually decreases with the increase of the number of nodes, after a certain number of nodes, the mean square error value begins to converge, and the change range is small.

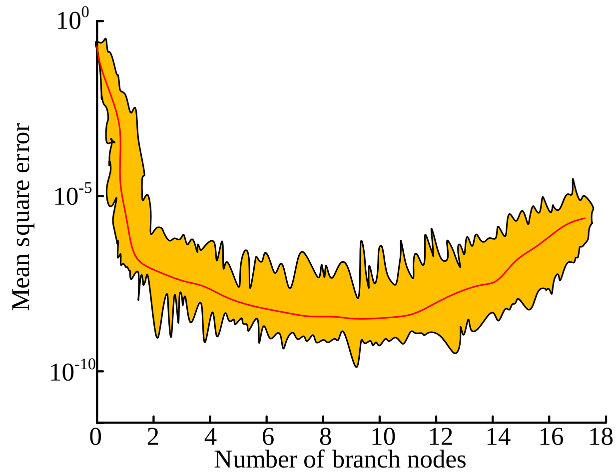
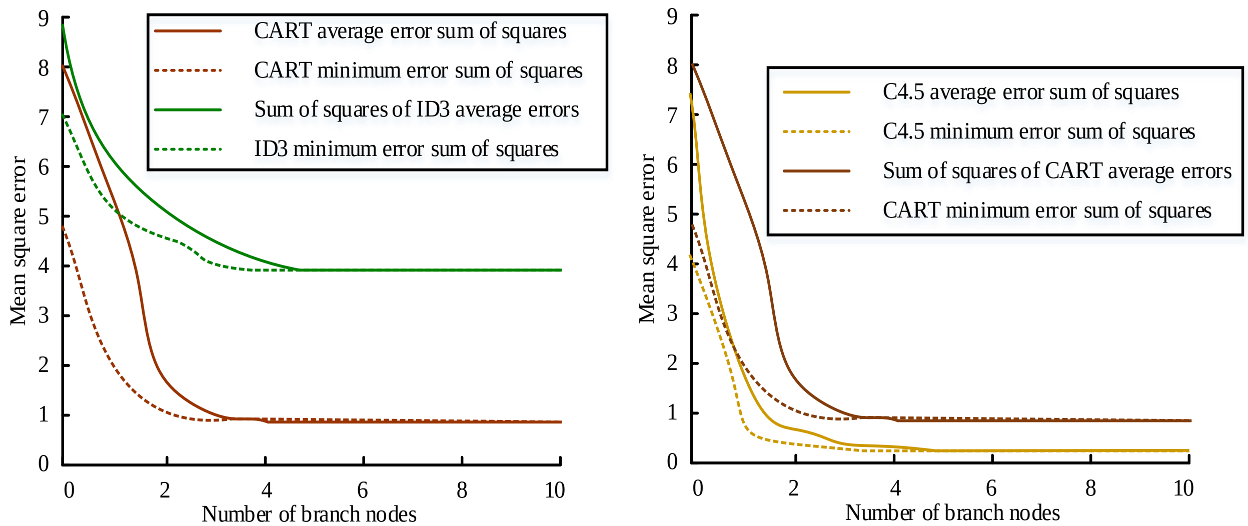


Fig. 8. Mean square error result of C4.5 algorithm

Fig. 8 shows the mean square error result of C4.5 algorithms. From Fig. 8 it is seen that when the number of nodes in the decision tree is small, its mean square error is large, and the increase in the number of nodes makes the mean square error of the decision tree change significantly. When the number of nodes is about three, the mean square error begins to converge gradually, and the variation range is small; but when the number of nodes is about 12, the mean square error increases gradually. It shows that the increasing number of nodes will lead to the increase of algorithm operation, and the decrease of its efficiency and prediction accuracy. The algorithm is further compared. The error square of the algorithm is shown in Fig. 9. The sum of the error squares of the two algorithms is further compared.

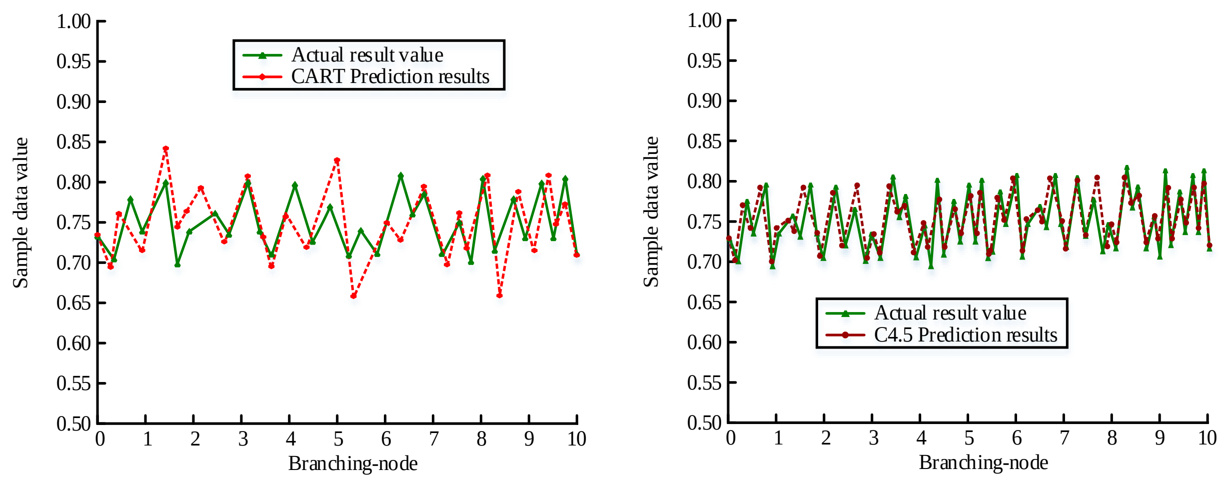
Fig. 9 shows the sum of error squares of ID3 algorithm and CART algorithm, and the sum of error squares of CART algorithm and C4.5 algorithms. It can be seen from Fig. 9(a) that the sum of error squares of CART algorithm is smaller than that of ID3 algorithm as a whole. When the number of branch nodes is 1-3, the error square changes greatly, and the sum of error squares changes from eight to about one. Since the number of nodes is four, the convergence speed of the average error square and the minimum error square of the algorithm decreases significantly, and the convergence tends to be smooth. The sum of squares of the CART algorithm converges to about 0.9. From Fig. 9 (b), it can be concluded that the sum of squares of errors of C4.5 algorithms is smaller than CART algorithm as a whole. When the number of branch nodes is no more than three, the error square of C4.5 algorithm changes greatly, and the convergence speed is fast. The sum of error squares changes from 7.5 to about 0.5. When the number of self-nodes is no less than four, the convergence speed of the average error square and the minimum error square of the algorithm decreases significantly, and the convergence tends to be smooth. The sum of squares of C4.5 algorithm converges to about 0.17.



(a) The result of the sum of squares of errors of ID3 algorithm and CART algorithm

(b) Sum of squares of errors of CART algorithm and C4.5 algorithm

Fig. 9. The result of the sum of squares of errors of ID3 algorithm and CART algorithm, and the result of the sum of squares of errors of CART algorithm and C4.5 algorithm



(a) Prediction results of CART algorithm on test samples

(b) C4.5 Prediction results of algorithm on test samples

Fig. 10. Prediction results of CART algorithm on test samples and C4.5 algorithm on test samples

Fig. 10 shows the prediction results of CART algorithm on test samples and C4.5 algorithm on test samples. It can be seen from Fig. 10(a) that there is a certain degree of difference between the predicted results of CART algorithm and the actual results. The actual data results fluctuate between 0.70 and 0.80, while the predicted results of CART algorithm vary between 0.70 and 0.85. It can be seen from Fig. 10(b) that the improved prediction results have a high coincidence with the actual results. Therefore, the improved C4.5 algorithm has a high accuracy, and there is no significant error between the prediction results and the actual results.

Fig. 11 is a comparison chart of prediction model accuracy under three different algorithms. In Fig. 11, the predicted average value of CART model is 89.96%. The predicted average value of ID3 model is 92.56%. Decision tree algorithm C4.5 based on information entropy has a prediction accuracy of 95.1% in the model. By comparing the accuracy values of the prediction models under three different algorithms, it can be seen that the model proposed in this study has better prediction accuracy, so it can be used in actual prediction.

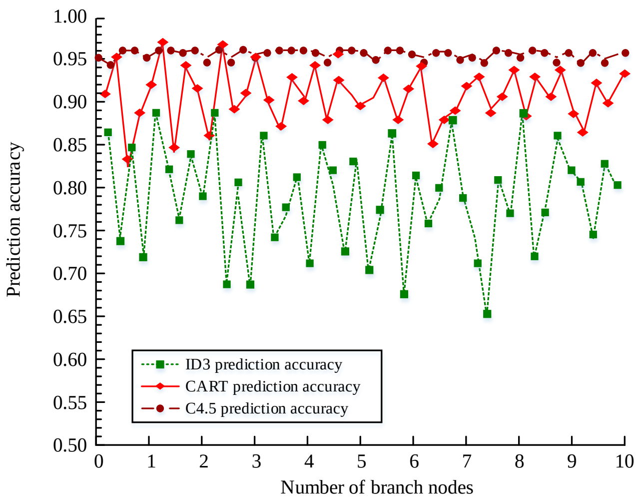


Fig. 11. Comparison chart of prediction model accuracy under three different algorithms

In conclusion, the improved decision tree algorithm model built by this research can have better prediction accuracy. Applying this model to the analysis of financial data in colleges and universities can obtain the financial implementation progress. The improved decision tree algorithm model can give an early warning according to the monthly financial income and expenditure of colleges and universities, so as to remind managers of the financial situation of colleges and universities in the current month.

V. CONCLUSION

The development of information industry has brought great impact to society, making information everywhere. The financial management in higher education institutions, due to the tedious work, generates a lot of information and is difficult to sort out, which affects the leaders' decision on the follow-up work. Therefore, in order to extract relevant data from redundant information data and facilitate making correct decisions, a decision tree algorithm based on information entropy is proposed. The algorithm uses entropy value and information gain rate to predict the probability of event occurrence, which can achieve different early warning effects for different events. The research verifies the prediction performance of the decision tree through the test set. In the three test sets, the precision measurement of the algorithm prediction under the support of information entropy reaches 95%. Then, the current financial data of colleges and universities are analyzed, and the financial implementation progress is obtained. Colleges and universities should be cautious about their financial expenditure in the month, otherwise the decision tree will issue a financial warning. Finally, to verify the performance of the algorithm, compared with ID3 algorithm and CART algorithm, the decision tree algorithm based on information entropy performs best by comparing the mean square error, the sum of squares of errors and the accuracy.

VI. FUTURE WORK

Although the C4.5 algorithm in the traditional decision tree algorithm has been improved in this research, the decision tree algorithm based on information entropy is proposed and analyzed with information entropy as the decision basis of the decision tree. At the same time, the measurement method is introduced to calculate the information entropy and information gain rate, and then an application model of university financial management based on the improved decision tree algorithm is proposed. However, there are still shortcomings in this research. The calculation of the information gain rate of the decision tree established on the basis of information entropy is more complex, which increases the difficulty of the algorithm. In the experiment, there are few sample data, resulting in some errors in accuracy. In the subsequent research, we can find a more simple calculation method and add more data samples from more colleges to improve the algorithm efficiency and prediction accuracy.

REFERENCES

[1] B.J. Jang, L.H. Kim, K. Y.Kim, “The Effects of the Creating of Shared Value of Hospitality Industry Companies on Corporate Trust and Non-financial Management Performance: Focused on Travel Agencies.” Journal of Tourism and Leisure Research, Vol. 33, No. 1, pp. 197-216, 2021.

[2] S. Cohen, F. Manes-Rossi, I. Brusca, E. Caperchione. “Guest editorial Happy endings and successful stories in public sector financial management: a lesson drawing perspective.” International Journal of Public Sector Management, Vol. 34, NO. 4, pp. 393- 406, 2021.

[3] Z. Zhu, J. Liu, Q. Fan, Y. Liu, Y. Wu, J. Zhou, F. He. “Comparative analysis and research on composition of mulberry-sourced medicinal materials by fingerprint division total statistical moment method and information entropy.” China journal of Chinese materia medica, Vol. 46, No.10, pp. 2547-2555, 2021.

[4] E. Priya, “Resnet Based Feature Extraction with Decision Tree Classifier for Classificaton of Mammogram Images.” Turkish Journal of Computer and Mathematics Education (TURCOMAT), Vol. 12, NO. 2, pp. 1147-1153, 2021.

[5] C. Al, “Application of Decision Tree in Classifying Secondary School Students’ Tendencies to Choose TVET in Malaysia.” Turkish Journal of Computer and Mathematics Education (TURCOMAT), Vol. 12, NO. 3, pp. 3002-3012, 2021.

[6] L. Mao, W. Zhang, “Analysis of entrepreneurship education in colleges and based on improved decision tree algorithm and fuzzy mathematics.” Journal of Intelligent and Fuzzy Systems, Vol. 40, NO. 2, pp. 2095-2107, 2021.

[7] S. Ziweritin, I.A. Ibiam, T.A. Oyeniran, G.E. Okol. “KNN and Decision Tree Model to Predict Values in Amount of One Pound Table.” International Journal on Computer Science and Engineering, Vol. 9, NO. 7, pp. 17-21, 2021.

[8] D. Mabuni, S. A. Babu, “High Accurate and a Variant of k-fold Cross Validation Technique for Predicting the Decision Tree Classifier Accuracy.” International Journal of Innovative Technology and Exploring Engineering, Vol. 10, no. 2, pp. 105-110, 2021.

[9] N. Ahmad, S. Ismail, S. A.Siraj, “Financial sustainability of Malaysian public universities: officers’ perceptions.” The International Journal of Educational Management, Vol. 33, No. 2, pp. 317-334, 2019.

[10] N.Asandimitra, A. Kautsar, “The Influence Of Financial Information, Financial Self Efficacy, And Emotional Intelligence To Financial Management Behavior Of Female Lecturer.” Humanities & Social Sciences Reviews, Vol. 7, No. 6, pp. 1112-1124, 2020.

[11] M. Mohammad, A. Saleh, O.A.A. Jawabreh, R.A. Om. "Artificial Intelligence (Ai) And The Impact of Enhancing the Consistency and Interpretation of Financial Statement in The Classified Hotels in Aqaba, Jordan." Academy of Strategic Management Journal, Vol. 20, No. S(3), pp. 1-18, 2021.

[12] V. V. Malikov, "The Algorithm for Building a Financial Management System at the Enterprise." Business Inform, Vol. 3, No. 506, pp. 344-348, 2020.

[13] A. UJ, E. NMB, C. KS, D. SKL, "Financial crisis prediction model using ant colony optimization - ScienceDirect." International Journal of Information Management, Vol. 50, pp. 538-556, 2020.

[14] J. Hou, "Online teaching quality evaluation model based on support vector machine and decision tree." Journal of Intelligent and Fuzzy Systems, Vol. 40, No. 2, pp. 2193-2203, 2021.

[15] M.N. Rabbani, A. Yusuf, D. Rolliawati, "Komparasi Model Prediksi Daftar Ulang Calon Mahasiswa Baru Menggunakan Metode Decision Tree Dan Adaboost." Jurnal Sisfokom (Sistem Informasi dan Komputer), Vol. 10, No. 1, pp. 18-24, 2021.

[16] D.V. Krishnaveni, S.A. Kumar, J. Sudarvel, P. E. aswaran. "An Intelligent Sale Prediction Classification of Jute Products Using Decision Tree Clustering & Statistical Model." Türk Fizyoterapi ve Rehabilitasyon Dergisi/Turkish Journal of Physiotherapy and Rehabilitation, Vol. 32, NO. 3, pp. 1033-1039, 2021.

[17] N. Kolenda, N. Matviichuk, L. Cherchyk, H. Nakonechna, K. Amit. "Strategic Approach to The Management of Financial Security of Enterprise." Financial and Credit Activity Problems of Theory and Practice, Vol. 4, No. 35, pp. 79-86, 2020.

[18] I. Podhorska, J. Vrbka, G. Lazaroiu, M. Kovacova. "Innovations in Financial Management: Recursive Prediction Model Based on Decision Trees." Marketing and Management of Innovations Vol. 3, pp. 276-292, 2020.

[19] J. Tyndall, "Prairie and tree planting tool—PT2 (1.0): a conservation decision support tool for Iowa, USA." Agroforestry Systems, Vol. 96, No. 1, pp. 49-64, 2022.

[20] E.B. Hicham, K. Salah-Ddine, "Financial News Analysis for Moroccan Stock Trend Predictions." Test Engineering and Management, Vol. 82, No. 1, pp. 1712-1717, 2020.

A Semantic NoSQL Application Program Interface for Big Data

K. ElDahshan[1], E. K. Elsayed[2], H. Mancy[3], A.AbuBakr[4]
Department of Mathematics-Faculty of Science, Al-Azhar University, Cairo, Egypt[1]
Department of Mathematics-Faculty of Science (Girls), Al-Azhar University, Cairo, Egypt[2,3,4]
Computer Science Institute, Canadian International College, Cairo, Egypt[2]

***Abstract*—Complexity, heterogeneity, schemaless-ness, data visualization, and extraction of consistent knowledge from Big Data are the biggest challenges in NoSQL databases. This paper presents a general semantic NoSQL Application Program Interface that integrates and converts NoSQL databases to semantic representation. The generated knowledge base is suitable for visualization and knowledge extraction from different Big Data sources. The authors use a case study of the COVID-19 pandemic prediction and other weather occurrences in various parts of the world to illustrate the suggested API. The Authors find a correlation between COVID-19 spread and deteriorating weather. According to the experimental findings, the API's performance is enough for heterogeneous Big Data.**

Keywords—NoSQL database; formatting; semantic technology; data integration; pandemic prediction

I. Introduction

Large data sets have their roots in the 1960s and 1970s when the world of data was just getting started with the creation of the first data centers and the development of the relational database. Nowadays, the data grows at sky-high rates. The world has multiple Big Data sources with different structures like sensors, scientific experiments, and social networks. We produce and collect more data every minute, and we need to be able to process them as soon as possible. Everything depends on timing, including stock trading, tracking the growth of epidemics, and traffic monitoring. A minor misunderstanding could lead to both financial loss and fatalities. The high level of adoption of Big Data technology is influenced by the quick and continual growth in data volumes.

A. Research Problem

Researchers face many problems when dealing with data from multiple sources. Because the structure of the data is different it makes it difficult to process the data and extract knowledge. Most researchers are now trying to store the data in a semantic form to make it easier to process. To do this they need to collect data from different data stores and preprocess these data to present them in semantic form. The preprocessing phase takes much time and effort which we need to solve the real problems we face and take better decisions based on the knowledge extracted from the data.

This research is funded by the Academy of Scientific Research and Technology (ASRT), Cairo, Egypt, project titled “Coronavirus Prevalence Prediction Model” (Project ID: 6641).

B. Research Objectives

This article studies many semantic Big Data frameworks and their limitations. Then it proposes a semantic NoSQL application program interface (API). The proposed API can read data from multiple NoSQL databases and convert them to Ontology. By using the proposed API, it will be possible to apply semantic queries on different data stores. The proposed API can also be used to integrate different data stores into a defined format.

C. Organization

This paper is organized in the following way to fulfill the research goals: Section 2 reviews a background about Big Data, Ontology, and NoSQL databases. Section 3 examines the history of earlier similar works. Section 4 Shows the proposed API architecture. The implementation of our proposed API is described in Section 5. Section 6 provides a case study of the API to analyze COVID-19 spread and weather behavior. The impact of utilizing the suggested API is described in Section 7. The paper's conclusion is covered in Section 8, which also analyses the research's significant contribution and limitations.

II. Background

A. Big Data

The term Big Data refers to a concept mostly used to classify large amounts of data. Despite the wide agreement on the promises and prospects of Big Data, there is no standard definition for it at this time.

The Institute of McKinsey Global defined the term Big Data as a “very Big Data set that can’t be stored or managed with database software tools” [1].

Gartner [2] defined Big Data as "a high-volume, high-variety, and/or high-velocity information asset that necessitates cost-effective, novel types of data processing that provide better insight, decision-making, and process automation."

The term "Big Data" was defined by Deepak Gupta and Rinkle Rani [3] as ‘Big data refers to large datasets which require non-traditional scalable solutions for data gathering, storage, management, analysis, and visualization, to extract actionable insights that could have an impact on every area of human life’. Big data characteristics are portrayed as extensions of ’V’s [3][4]. Fig. 1 presents the Big Data characteristics.

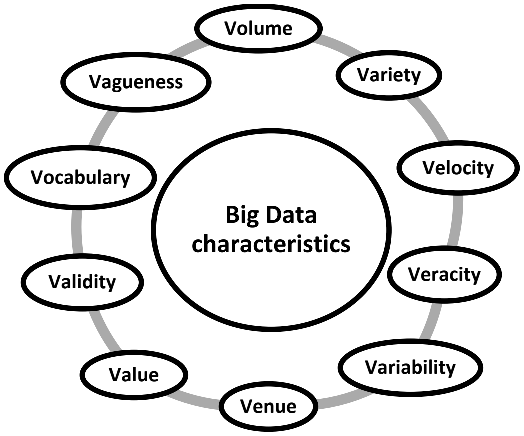


Fig. 1. Big Data characteristics

B. Ontology

An Ontology consists of O = A, C, I, P, R, T. Where A is the set of axioms; C is the set of concepts or classes; I is the set of instances; P is the set of properties of the concepts describing various features; R is the set of relationships between concepts and T is the set of hierarchical relationships among concepts that are called taxonomies [5].

The components of Ontology are as follows:[6]

- *Individuals*: "The ontology population is represented by individuals, which are objects. They are instances of classes."
- *Classes*: "Classes are particular categories of objects or things that represent a collection of instances or specify a particular category of entities. They are frequently referred to as concepts or entity types. Classes can be used to classify individuals, other classes, or a combination of both."
- *Attributes*: "Properties associated with objects or classes are called attributes. They include statements about Datatypes and their DataValues, characteristics, features, or parameters of individuals and classes."
- *Relations*:" The numerous links that indicate how two individuals or classes are related. Additionally called associations, roles, relationship types, and object properties".
- *Axioms*:" Logical rules and assertions that together create the general theory that describes the relationship between the ontology elements. They represent formal definitions of the ontology knowledge."

C. NoSQL Databases

NoSQL (Not Only SQL) is a database that is a non-relational distributed database system. It facilitates the rapid structuring of data analysis with large volumes of data and a variety of data types. NoSQL is also referred to as a cloud database, a non-relational database. NoSQL databases are not based on tables and do not often employ structured query language to manipulate data. Receive and append operations are frequently highly optimized in NoSQL database systems. When working with large amounts of data and this data structure does not require a relational model, NoSQL databases come in handy [4].

When compared to relational databases, NoSQL databases are more scalable, diverse, simple to use, flexible, and give better performance. MongoDB is now the most popular NoSQL database, with Apache Cassandra, Redis, and HBase following closely behind. Neo4j is the most popular NoSQL graph database and the most common cloud database is Amazon DynamoDB [7]. NoSQL data models allow related data to be stored in a nested data structure [8].

Based on their data model, NoSQL databases are classified into a variety of types[5]. The main types of NoSQL databases are wide-column databases, document databases, key-value databases, and graph databases. They provide flexible schemas and they are very scalable to large amounts of data and high user loads. Table I provides a comparative study of the four types of NoSQL databases.

TABLE I. A COMPARATIVE STUDY OF THE FOUR TYPES OF NOSQL DATABASES

NoSQL database types	**Document database**	**Key-value store**	**Wide-column store**	**Graph store**
Data Storage	It stores data in JSON, BSON, or XML documents	It stores data as an attribute name (or "key") combined with its value	It stores data in tables, rows, and dynamic columns	It stores data in nodes and edges
Use Cases	It is great for a wide variety of use cases and can be used as a general-purpose database	It is the best to use when you need to store large amounts of data but you don't need to perform complex queries to retrieve it such as storing user preferences or caching	Common use cases for wide-column stores include storing Internet of Things data and user profile data	It is the best to use when you need to traverse relationships to look for patterns such as fraud detection, social networks, and recommendation engines
Performance and scalability	High	High	High	Very high
Flexibility	High	High	Moderate	High
Complexity	Low	Very Low	Low	High
Examples	MongoDB, CouchDB	Redis , DynamoDB , Voldemort	Hbase, Big Table, Cassandra	Neo4j, HyperGraph, InfiniteGraph

III. Related Work

Many researchers studied the relationship between semantic technology and Big Data technology. They tried to connect knowledge management systems with NoSQL database management systems to apply semantic queries on Big Data.

Bansal S and Kagemann S proposed Semantic Extract-Transform-Load (ETL) framework that uses semantic technologies to integrate and publish data from multiple sources. The Extract-Transform-Load (ETL) process refers to a process in data warehousing that extracts data from outside sources, transforms it to fit operational needs, which can include quality checks, and loads it into the end target database. The authors extracted data from different sources in flat file formats such as CSV. The proposed semantic ETL framework first creates a semantic model of the datasets being integrated, and then it creates semantically linked data that adheres to the data model. A semantic data model and semantically linked data (RDF triples) are produced using semantic technologies and stored in a data warehouse during the transform phase of an ETL process. The transformation phase will involve a manual process of analyzing the datasets, the schema, and their purpose. Based on the findings, the schema will have to be mapped to an existing domain-specific Ontology or Ontology will have to be created from scratch [9].

Hanen Abbes and Faiez Gargouri implemented a tool to generate Ontology from MongoDB databases. It proposed transformation rules from MongoDB to OWL Ontology. This work is done in five main steps. First is the creation of the Ontology skeleton by defining Ontology classes and detecting the relationships between them. Second, learn object properties and datatype properties. Third, identify Individuals. Forth, deduce class axioms, property axioms, and constraints. Finally, enrich the Ontology with class definition operators [10].

Mahmudul Hassan and Srividya K. Bansal proposed a solution to execute SPARQL query as SQL query using Apache Spark on large-scale RDF data stored in NoSQL databases such as HBase and Cassandra. It translated the SPARQL query to SPARK SQL for both HBase and Cassandra storage schemas. It first converted RDF data to store it in HBase and Cassandra and then proposed an algorithm to convert SPARQL query to SPARK SQL using the in-memory data processing engine Spark. The main purpose of this paper is to execute a query on RDF data (semantic data) with a large volume. To achieve that, it stored the RDF data in a NoSQL database. However, the algorithm can be used to perform SPARQL queries on data already stored in HBase or Cassandra [11].

K. ElDahshan, E. K. Elsayed, and H. Mancy developed a semantic dashboard using java JDK to connect to HBase and built a universal knowledge base using Protégé. It converted the SPARQL query to spark SQL using Sempala. The conversion is done using the algebra tree [12].

S. Mhammedi, H. El Massari, and N. Gherabi proposed an approach to automatically learn OWL Ontology from data in the Couchbase database by applying six mapping rules. The mapping rules are learning classes, learning object properties from the embedded document, learning datatype properties, transforming all data values of fields in each document to individuals, learning property restrictions, and learning class hierarchies [13].

All of them except [11,12] do not handle Big Data. All of them allow the conversion of only one NoSQL database type. The proposed API handles Big Data, accepts any type of NoSQL database, allows schema conversion, and the proposed API is platform-independent. Table II shows the comparison among related works through data sources, NoSQL type, tools used, and limitations.

TABLE II. A Comparative Study of the Related Works

Research	Data Source	NoSQL Type	Tools	Limitations
[9]	Data from multiple sources in flat file formats such as CSV	--	Semantic Extract-Transform-Load (ETL) framework which generates a semantic model of the datasets under integration and then generates semantically linked data	The framework can not deal with NoSQL database stores
[10]	MongoDB	Document database	A tool to generate Ontology from MongoDB implemented by the JAVA programming language	The tool can deal with only one type of NoSQL database store
[11]	HBase and Cassandra	Wide-column	A query compiler that is written in Flex and Bison that translates SPARQL to Spark SQL to execute a query on RDF data stored in HBase and Cassandra	The tool can deal with semantic data in RDF format stored in a wide-column database
[12]	HBase	Wide-column	A semantic dashboard using java JDK to connect to HBase and built a universal knowledge base using Protégé	The tool can deal with only one type of NoSQL database stores
[13]	Couchbase	Document database	An approach to learning Ontology from the Couchbase database using mapping rules. This work is done by the JAVA programming language and OWL API	The tool can deal with only one type of NoSQL database stores

IV. The Proposed Semantic Integration Application Program Interface Architecture

Researchers and academics regard data access to be unattainable in many situations for a variety of reasons. These include the following: a profusion of data, non-computerization of processes, heterogeneity, data duplication, and the presence of a lot of isolated data in databases that can only be accessed in a specific context. These traits typically lead to low-quality information, which makes it challenging for researchers to organize and evaluate them during the decision-making process [14]. Data integration is the process of making it possible for people to access, deliver, and utilize data from several sources and huge businesses while preserving its integrity and quality. Real-time updates to data saved in one source can also be mirrored in other sources thanks to this [15].

Although Ontology development is not a new field of study, Big Data faces new difficulties due to its features (velocity, variety, and volume). Therefore, our API takes into account the properties of Big Data for Big Data Ontology creation. The main idea is to create Ontology from a large number of diverse sources. The authors' goal is to make the process of automatically generating Ontology and importing Big Data as simple as possible for the user by providing an independent application program interface. The research's additional significant contributions include (1) the ability to integrate different sources of NoSQL, (2) the ability to support the Ontology creation-based Big Data access layer, (3) the ability to re-engineer and combine different Ontologies, and (4) the capability to easily build a semantic query over Big Data. Fig. 2 shows the proposed API overview.

The structure of the proposed API is as follows:

- *Connection Phase*: In this phase, the authors build a Big Data access layer that allows users to connect to any NoSQL database.
- Integration and Extraction *Phase*: In this phase, the API reads and extracts every record in the database and puts them in a Python dictionary.
- *Conversion phase*: In this phase, the API converts the dictionary to an XML file.
- *OWL Creation phase*: In this phase, the API creates OWL from the XML file.
- *Semantic and visualization phase*: In this phase, the API builds semantic rules and visualizes the data in a graph.

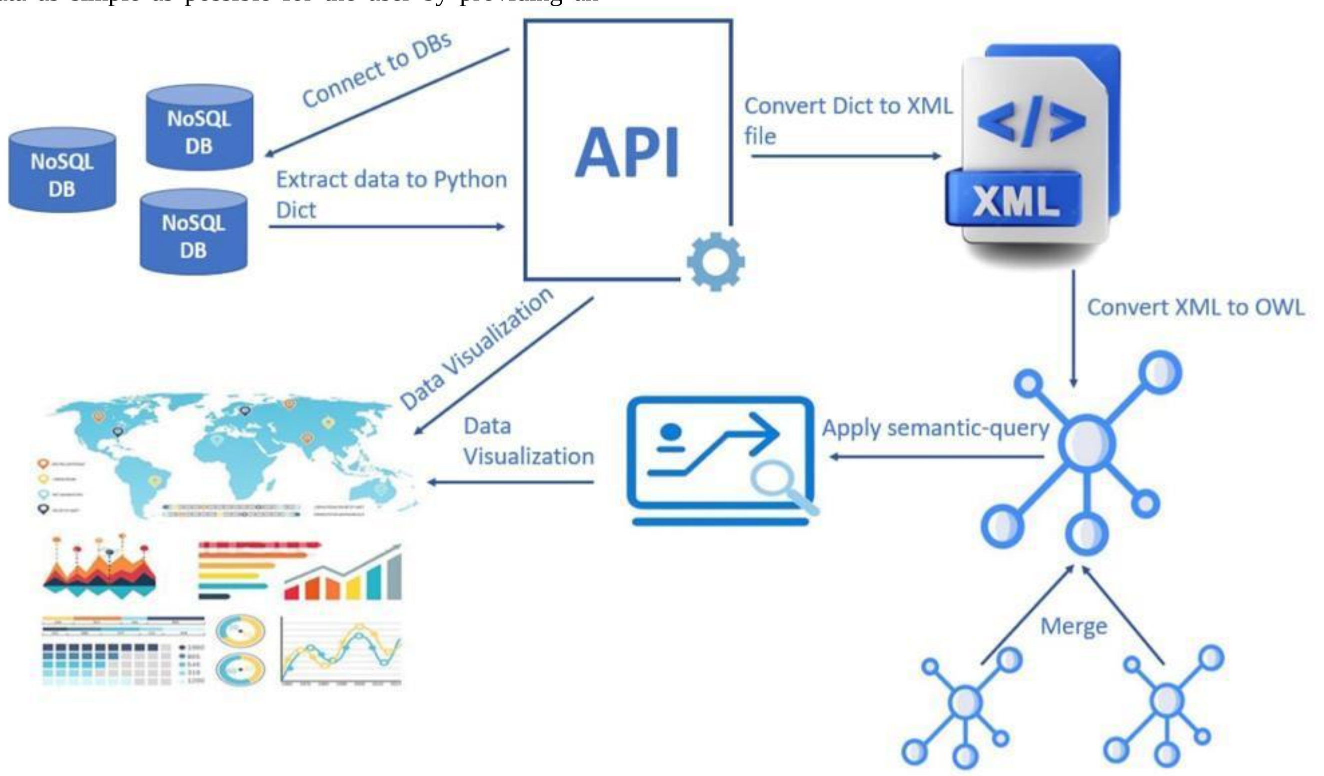


Fig. 2. The proposed API Overview

VI. Proposed Semantic Integration Application Program Interface Implementation

The API is created with Python 3.8.5 on Windows 10 operating system with Intel(R) Core(TM) i7-6500U CPU and 8GB RAM. The proposed API asks the user to enter the databases' names to be converted. The API scans the databases and retrieves every record in them with the same structure as it was in the database. After getting every record, it will be converted to XML. The API can handle data with any structure. Every record will be retrieved as a dictionary data type in Python. Then it will be converted to a JSON array. The JSON array will be converted to XML using the 'dicttoxml' package. The API can handle documents with complex structures and large sizes. For every collection in MongoDB and every table in HBase, a root class will be created. A subclass will be created to represent each attribute in the records. If the document in the MongoDB collection has embedded objects, other subclasses will be created to represent the attributes of the embedded object. After the conversion is completed, the XML is converted to OWL Ontology using DTD2OWL. Fig. 3 explains the proposed API workflow. Algorithm 1 demonstrates the Semantic NoSQL API execution steps.

Algorithm 1: Semantic NoSQL API

```
Initialize
Connect to the database
Scan every record in the database
For (every record) do
    | Put the record in Python dictionary
    | Convert the dictionary to XML
    | Save the output to a file
    | End
Convert XML to OWL
Build semantic Query
Visualize the result
```

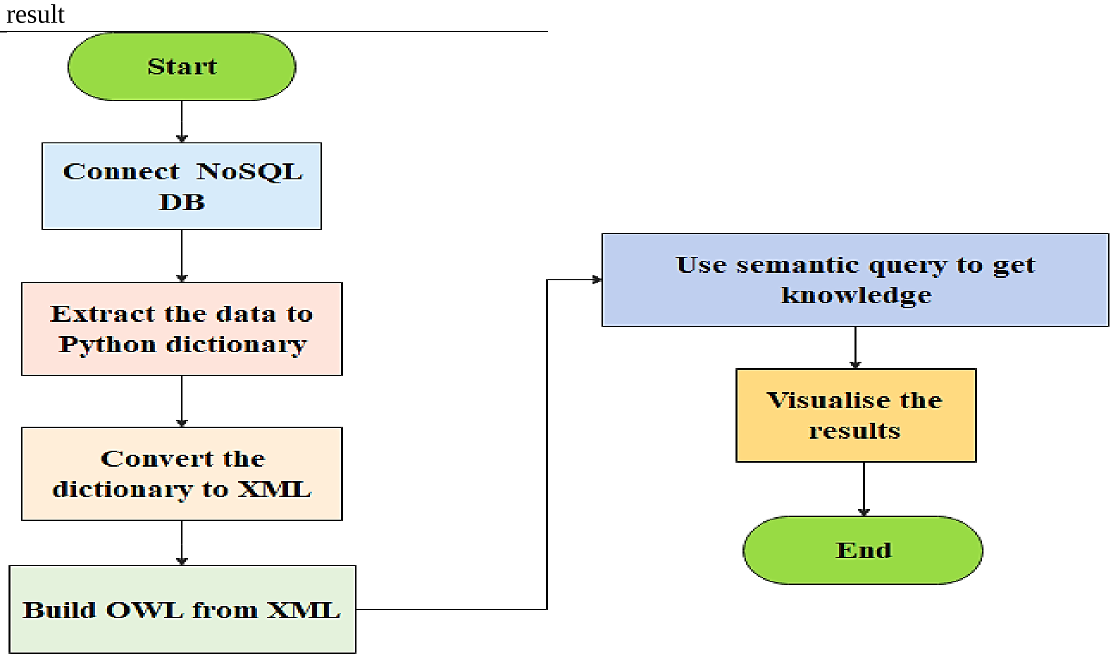


Fig. 3. The proposed API workflow

The phases of the proposed Onto-NoSQL API are implemented as follows:

A. Connection Phase

To access a NoSQL database, the author used the pymongo package to connect to MongoDB as a document database type, the happybase package to connect to HBase as a Wide-Column database type, the Redis-py package to connect to Redis as a Key-value database, and the Neo4j Python Driver to connect to Neo4j as a Graph database type. Algorithm 2 represents the connection code to MongoDB. Algorithm 3 represents the connection code to Hbase.

This paper focuses on both document database stores and column-oriented database stores as they are the most used NoSQL database stores. The most popular document store is MongoDB [6]. MongoDB is very popular because it allows multiple data types to be used. The database consists of collections. The collection stores the data as JSON documents. Schema is not needed to store the data in MongoDB. The data stored can be a string, number, date, array, or object. MongoDB can store data from multiple sources in different formats.

Algorithm 2: Connect to MongoDB

```
from pymongo import MongoClient
client = MongoClient()
#enter database name
database= input('Enter DB name:')
db = client[database]
```

One of the most popular column-oriented database stores is HBase. HBase is an open-source management system that is a versioned and distributed database based on Google's BigTable.

This system is column-oriented and built on top of HDFS, which speeds up read and write operations across Big Data sets. Application programming interfaces (APIs) such as Thrift and Java provide access to HBase. There are no query or scripting languages specific to these APIs. HBase is reliant on a ZooKeeper instance by default [16].

Algorithm 3: Connect to HBase

```
import happybase as hb
conn= hb.Connection(HostName,PortNumber)
conn.open()
```

B. *Integration and Extraction Phase*

The proposed API integrates MongoDB and HBase databases. It extracts every record in the database and puts them in a Python dictionary. Algorithm 4 reads records from MongoDB. Algorithm 5 reads records from HBase.

Algorithm 4: Read records from MongoDB

```
collection_names = db.collection_names()
For (every collection ) do
    #retrieve documents in each collection
    docs = db[collection_name].find()
    End
```

Algorithm 5: Read records from HBase

```
table = conn.table('table_name')
For (every key, row) do
    # retrieve records
    row = str(row)
    row =json.loads(row)
    End
```

C. *Conversion Phase*

The proposed API converts the dictionary to an XML file. Algorithm 6 converts the dictionary generated from MongoDB records to an XML file. Algorithm 7 converts the dictionary generated from HBase records to an XML file.

Algorithm 6: Convert dictionary generated from MongoDB to XML file

```
#create XML file
f = open(database+'/'+collection_name+'.xml', 'wb')
For (every doc) do
    #convert object id to string
    doc_sanitized = json.loads(json_util.dumps(doc))
    #convert to XML
    xml = dicttoxml.dicttoxml(doc_sanitized)
    f.write(xml)
    End
f.close()
```

Algorithm 7: Convert dictionary generated from HBase to XML file

```
#create XML file
f = open('table_name.xml', 'wb')
For (every key, row) do
    xml = dicttoxml(row)
    f.write(xml)
    End
f.close()
```

The interface of the API was created using PHP (Laravel Framework). First, we need to install the Process component that executes commands in sub-processes. This can be done using the composer to install the package. Second, use the Process class that enables Laravel to run a script. Third, create a new instance of the Process class, which takes three parameters the name of the script, the script itself, and the arguments passed to the script. An HTML form was created to read the data from the user. The user chooses the NoSQL database engine (MongoDB or HBase) and writes the name of the database (in the case of choosing MongoDB) or the table (in the case of choosing HBase) to be converted to XML. After the script is executed, the user gets a message that the database (in the case of using MongoDB) or the table (in the case of using HBase) is converted successfully. The XML files are located in the public folder in the Laravel project. Fig. 4, 5, and 6 represent the screens of the proposed API. Fig. 7 is a snapshot from the generated XML file.

Fig. 4. The API screens- the input form

Fig. 5. The API screens- choose NoSQL data store

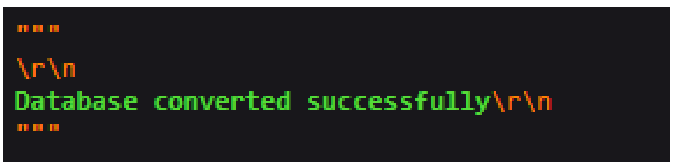


Fig. 6. The API screens- the success message after the database is converted


```
<?xml version="1.0" encoding="UTF-8"
?><root><doc><_id type="dict"><key name="$oid"
type="str">61c481b0593caffbcd8229d9</key></_id><Co
nfirmed type="int">0</Confirmed><Country
type="str">Afghanistan</Country><Date
type="str">2020-01-22</Date><Deaths
type="int">0</Deaths><Recovered
type="int">0</Recovered></doc>
<doc><_id type="dict"><key name="$oid"
type="str">61c481b0593caffbcd8229da</key></_id><Co
nfirmed type="int">0</Confirmed><Country
type="str">Afghanistan</Country><Date
type="str">2020-01-25</Date><Deaths
type="int">0</Deaths><Recovered
type="int">0</Recovered></doc>
<doc><_id type="dict"><key name="$oid"
type="str">61c481b0593caffbcd822c93</key></_id><Co
nfirmed type="int">157795</Confirmed><Country
type="str">Afghanistan</Country><Date
type="str">2021-12-20</Date><Deaths
type="int">7335</Deaths><Recovered
type="int">0</Recovered></doc></root>
```


Fig. 7. A snapshot of the generated XML file

D. OWL Creation Phase

The proposed API converts XML format to OWL. The Authors implement the DTD2OWL2 Method in python [17].

E. Semantic and Visualization Phase

The proposed API builds a semantic query using rdflib [18] and visualization using the Plotly package in Python.

VII. Case Study

The authors apply API to COVID-19 and weather data. Data was collected from various sources regarding the spread of Covid19. The number of cases was collected from the WHO. The weather data was collected from National Centers for Environmental Information website [19]. The authors used the number of confirmed cases, population, and weather data to predict a new pattern of confirmed cases and to find a relation between it and the other factors. The data was preprocessed and converted to JSON format. MongoDB was used to store the confirmed case data. HBase was used to store the weather data. By converting the database to OWL, it is possible to apply semantic queries and extract more knowledge from the data. It was possible to visualize the data and find a relation between them despite being stored in different databases with different structures. Fig. 8, 9, and 10 visualize the data on the world map.

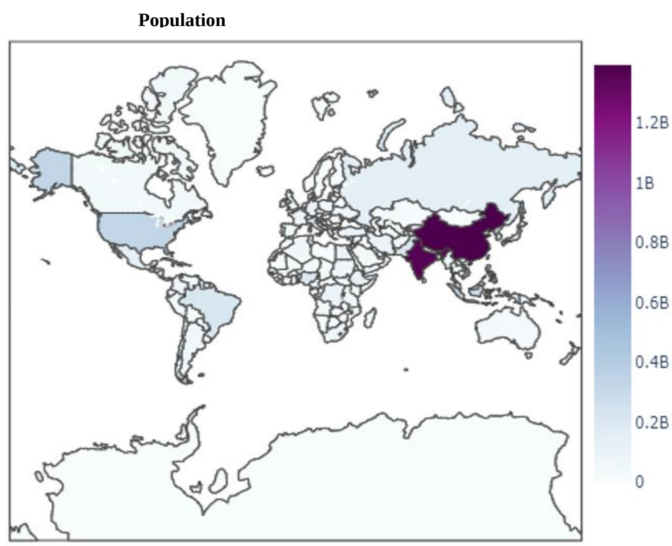


Fig. 8. The population of world countries

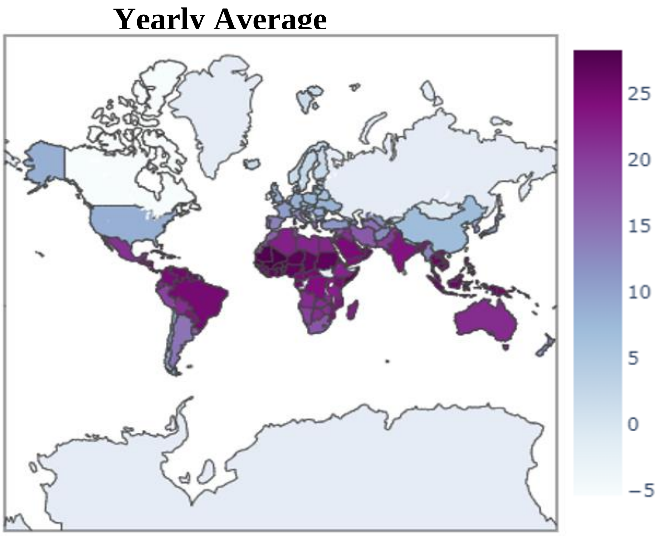


Fig. 9. The yearly average temperature of the world

The Covid-19 data was used to implement the proposed API. These data are various, change rapidly, and have large sizes. The data type can be text describing the case or symptoms, numbers describing the number of infected cases, images of the X-ray performed on the patients, or geographical data describing the spread of the virus. All these data can be stored in NoSQL databases. The user can work with one type of stored data to predict the behavior of the virus or even diagnose a patient based on the X-ray performed on his chest. The user needs to integrate these data from multiple sources to achieve his goal. The data can be integrated to extract more knowledge from them and help countries to prepare and deal with the spread of the virus. NoSQL databases can help to improve dealing with such important and various data to make better decisions. By converting the NoSQL database to Ontology, the user can discover new rules and relationships.

VIII. Analysis of Results

The used data were in the form of text and integers that represent the number of confirmed cases of Covid19 around the world. They also represent other factors like each country's temperature and population. The interval of the collected data was over 500 days from 22nd Jan 2020 to 31st Aug 2021. The confirmed case data and the population of every country in the world were stored in MongoDB, whereas the weather data were stored in HBase. By studying Fig. 10, it may be possible to see the spread of the virus around the world. By studying other factors, it could be possible to find a pattern or even a factor that deeply affects the infection rate, which will help us find a way to decrease the infection rate.

In Fig. 11 you could see the relation between the temperature represented by the blue curve and the number of confirmed cases represented by the red curve. The data were scaled in the range of 0 to 1 using the min-max scaler

As stated before, these data were stored in different NoSQL databases with different types. By using the proposed API, it was possible to connect to these databases and retrieve data from them. Four countries from around the globe were chosen to study the effect of temperature on the spread rate of the virus. The X-axis represents the number of days while the Y-axis represents the number of confirmed cases and temperature degrees every day.

By choosing a small interval of time such as in Fig. 12, it could be concluded that the infection rate is inversely proportional to the temperature degree. Of course, other factors affect the infection rate such as the population of the country under study (Fig. 8 demonstrates the population of world countries), the educational level of most of the population that affects their behavior, or the economy of the country that affects the medical care system.

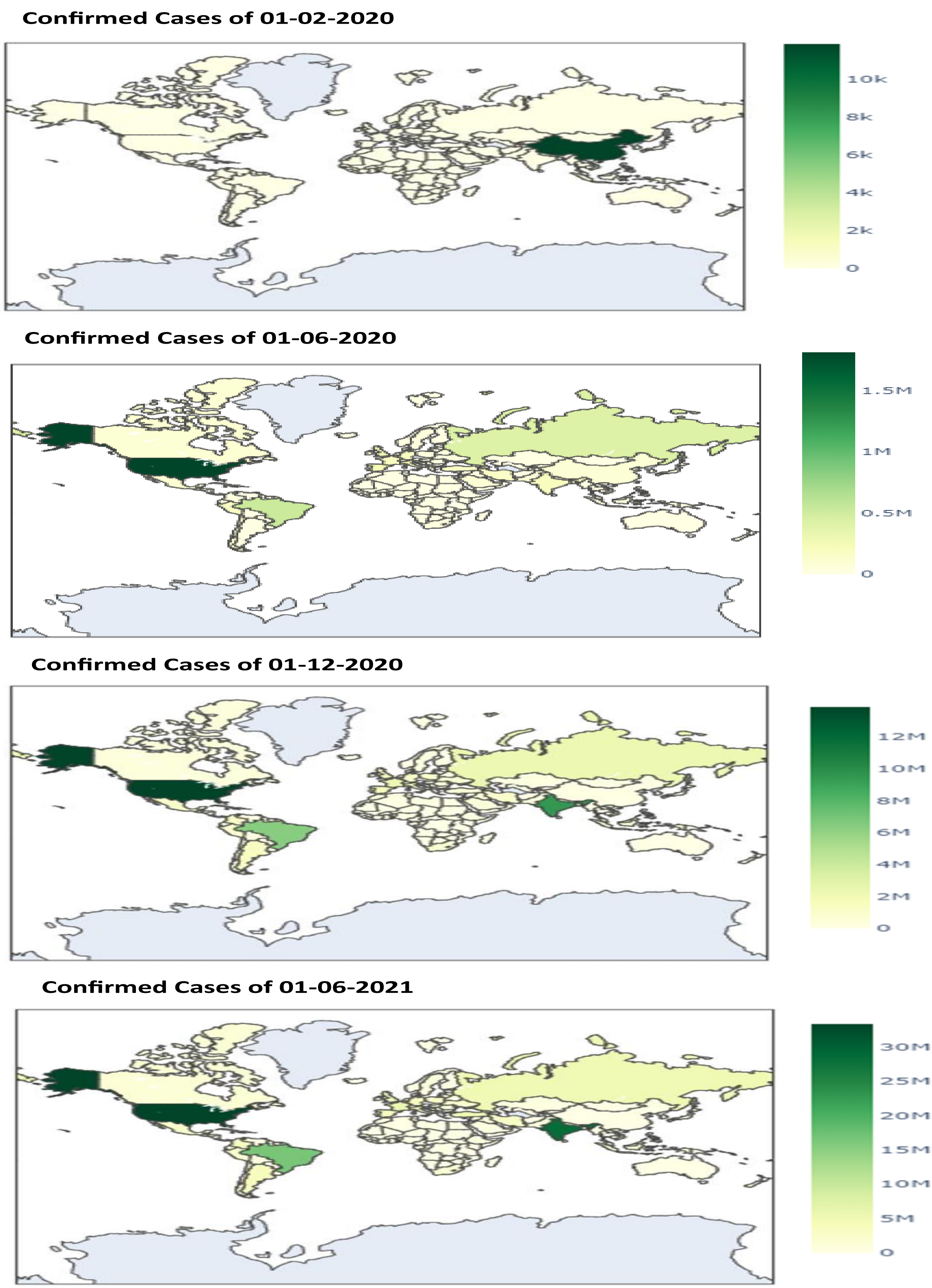


Fig. 10. The spread of confirmed cases over the world

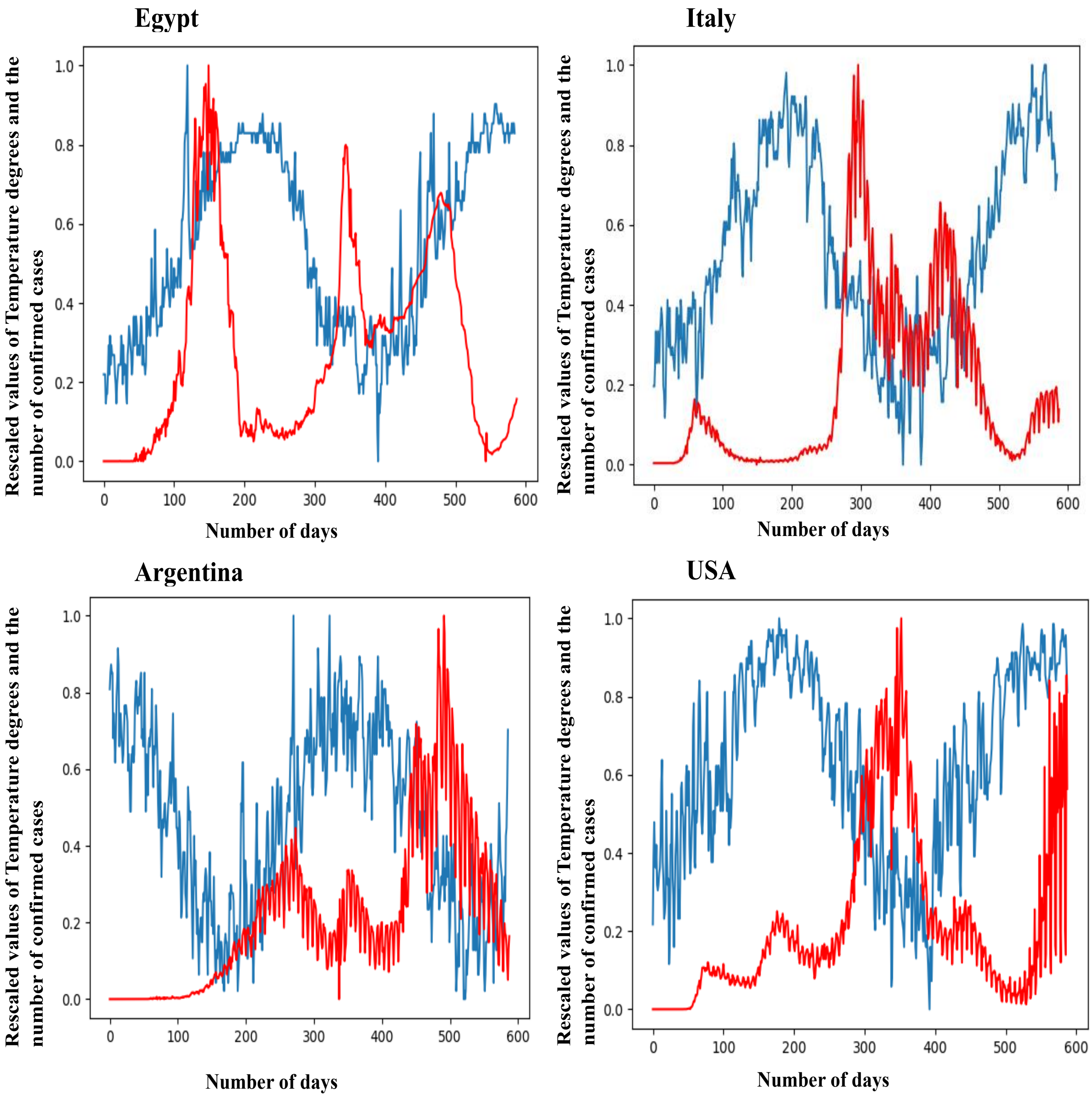


Fig. 11. The relation between temperature (blue) and number of confirmed cases (red) (The X-axis represents the number of days)

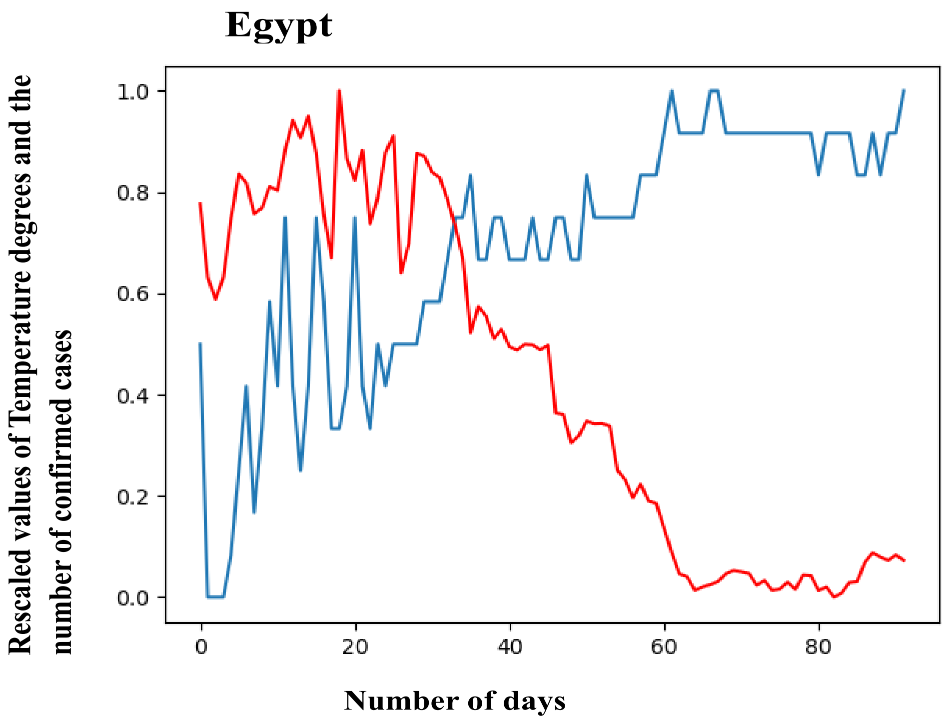


Fig. 12. The confirmed cases in Egypt during June, July, and August 2020 are in red, and the temperature is in blue (The X-axis represents the number of days)

IX. Conclusion

This paper presented Big Data concepts, technologies, and challenges. It studied NoSQL databases, their types, and the relationship between NoSQL Databases and Ontology. A proposed API was developed to integrate multiple NoSQL databases and generate OWL Ontology from them. By using the proposed API, it is possible to apply semantic queries, find new rules, and extract more knowledge. Also, it can be used to integrate data from different data stores into a unified defined format. Semantic data visualization has shown the relation between different data and discovered new patterns. One can say that the temperature is a factor in the virus's spread. It spreads faster in lower temperatures like in the USA, Russia, and some European countries (Fig. 9 demonstrates the yearly average temperature of the world). It also spreads faster in countries with a high population like India.

References

[1] Manyika J, Chui M, Brown B, et al. *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute Report, May 2011.

[2] Gartner. *Gartner IT Glossary: Big Data definition*. Available from: http://www.gartner.com/it-glossary/big-data [Accessed 22-2-2021].

[3] Gupta D, Rani R. A study of big data evolution and research challenges, *Journal of Information Science*. 2019; 45(3) :322–340.

[4] Moorthy J, Lahiri R, Biswas N, Sanyal D, Ranjan J, Nanath K, et al. Big Data: Prospects and Challenges. *VIKALPA The Journal for Decision Makers*. 2015;40(1):74-96 Available from: DOI:10.1177/0256090915575450 .

[5] ElDahshan K, Elsayed E.K, and Mancy H. Enhancement Semantic Prediction Big Data Method for COVID-19: Onto-NoSQL. *IAENG International Journal of Computer Science*.2020;47(4):613-622

[6] Khadir A, Aliane H, Guessoum A. Ontology learning: Grand tour and challenges. *Computer Science Review*. 2021;39(100339). Available from: https://doi.org/10.1016/j.cosrev.2020.100339.

[7] DB engines, Available from: https://db-engines.com/en/ranking [Accessed 8-12-2021].

[8] MongoDB, Inc. *What is NoSQL?*. Available from: https://www.mongodb.com/nosql-explained [Accessed 22-2-2021].

[9] Bansal S, Kagemann S. Semantic Extract-Transform-Load framework for Big Data Integration. *Computer*. 2015 Mar 1;48(3):42-50. https://doi.org/10.1109/MC.2015.76.

[10] Abbes H, Gargouri F. Big Data Integration: a MongoDB Database and Modular Ontologies based Approach. *Procedia Computer Science*. 2016; 96:446-455. Available from: https://doi.org/10.1016/j.procs.2016.08.099

[11] Hassan M, Bansal S. Semantic Data Querying over NoSQL Databases with Apache Spark. In: 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA. IEEE; 2018. P.364-371, doi:10.1109/IRI.2018.00061

[12] ElDahshan K, Elsayed E Mancy H. Semantic Smart World Framework. *Applied Computational Intelligence and Soft Computing*. 2020, Article ID 8081578. Available from: https://doi.org/10.1155/2020/8081578

[13] Mhammedi, S., El Massari, H., Gherabi, N. Cb2Onto: OWL Ontology Learning Approach from Couchbase. In: Gherabi, N., Kacprzyk, J. (eds) *Intelligent Systems in Big Data, Semantic Web and Machine Learning*. Advances in Intelligent Systems and Computing, vol 1344. Springer, Cham. Available from: https://doi.org/10.1007/978-3-030-72588-4_7

[14] Lima V, Bernardi F, Domingues M, Kritski A, Rijo R P, Alves D. A computational infrastructure for semantic data integration towards a patient-centered database for Tuberculosis care. *Procedia Computer Science*. 2022;196:434-438. Available from: DOI:10.1016/j.procs.2021.12.033

[15] Ahmed J, Ahmed M. Semantic Web Approach of Integrating Big Data. *International Journal of Computer Sciences and Engineering*. Sep 2018;6(9):529-532. Available from: DOI:10.26438/ijcse/v6i9.529532

[16] Khan N, Yaqoob I, et al. Big Data: Survey, Technologies, Opportunities, and Challenges *The Scientific World Journal*. 2014, Article ID 712826. Available from: http://dx.doi.org/10.1155/2014/712826

[17] Hacherouf M, Bahloul S.N, Cruz C. Transforming XML documents to OWL ontologies: A survey. *Journal of Information Science*.2015; 41(2):242-259.Available from: DOI:10.1177/0165551514565972.

[18] *rdflib documentation*. Available from: https://rdflib.readthedocs.io/en/stable/ [Accessed 15-3-2022]

[19] National Centers for Environmental Information. Available from: https://www.ncei.noaa.gov/ [Accessed 20-12-2021]

Big Data Application in Forecasting Financial Investment of e-Commerce Industry for Sustainability

Yanfeng Zhang
Business School, Sias University, Xinzheng, 451150, China

***Abstract*—With the rapid development of e-commerce, financial investment forecasting in the e-commerce industry has gradually become a concern of relevant personnel. Based on DBN, the study proposes a PVD prediction model. For training and test sample sets, the PLR_VIP algorithm is calculated and min-max normalization is applied to the original financial time series. To determine appropriate network parameters, the DBN network is trained and tested, and then Elliott wave patterns are predicted based on financial time series. The experimental results show that the MSE of the PVD model is 0.4015 and the prediction accuracy is 70.21%, indicating that it can efficiently and accurately identify the Elliott wave pattern of financial time series. Comparing the prediction results of the PVD model with the other five models, the values of the four evaluation indicators of PVD are the lowest among all models, which are 0.6336, 0.4015, 0.9052, and 29.79%, respectively. Compared with the training error changes of other models, it can be seen that the error curve of the DBN network is smoother and the training error is smaller. It shows that it has higher stability, faster convergence speed, higher reliability and accuracy, and shows excellent prediction performance, which is significantly better than other models. Experiments show that under the background of sustainable development, the PVD forecasting model proposed in the study performs well in financial investment forecasting, which provides a reference for the development of financial investment forecasting in the e-commerce industry.**

Keywords—Stainable development; big data; e-commerce industry; financial investment forecast; deep belief network

I. Introduction

The financial investment market plays an important role in today's social and economic field, and the direction of the financial market will directly affect the core competitiveness of enterprises [1]. The academic community continues to study the changing laws of the financial market, and find reliable means that can describe and predict the changing laws and future trends of the financial market [2]. Neural networks have been developing continuously in recent years, and a large number of related technologies have been applied to various fields, including forecasting the financial investment market [3]. e-Commerce is a new type of networked economic activity. In recent years, it has become an important means for many countries to enhance their economic strength and gain the advantages of global resource allocation. However, the e-commerce industry in China currently has problems such as waste of resources and inconsistent supply and demand, so it is necessary to use effective forecasting methods to forecast the financial investment market in the e-commerce industry [4]. To improve the development ability of financial investment in China's e-commerce industry, the research proposes a PVD (Piecewise Linear Representation VIP DBN) prediction model based on Deep Belief Network (DBN), which uses a deep learning model to model financial time series, and innovatively combines DBN and piecewise linear representation based on very important points (PLR_VIP). It can discriminate the Elliott wave pattern of financial time series and then predict the trend of the financial investment market in the e-commerce industry.

There are two innovation points in the research. First, the financial investment forecast dataset is built against the background of big data; secondly, intelligent algorithm is introduced into financial forecasting to realize automatic forecasting of financial investment. Compared with previous studies, the use of intelligent algorithms and big data technology can more effectively achieve its change prediction, and can more easily update data with the help of Internet technology.

II. Related Work

With the related research and development of big data, deep learning related technologies have been applied to the field of financial forecasting, and many scholars at home and abroad have conducted research on it. Kare et al. used Singular Spectrum Analysis (SSA) and Multi-channel Singular Spectrum Analysis (MSSA) to identify important deterministic cycles in residential property prices, private non-financial sector credit shares. Prediction test results show that the inclusion of financial big data significantly improves the prediction accuracy of financial cycle components [5]. Scholars such as Kelotra proposed a stock market forecasting system that effectively predicts the state of the stock market. First, the calculation of technical indicators is performed on the data from the real-time stock market, indicating that the necessary features are obtained by clustering using Sparse Fuzzy C-means (Sparse-FCM), and then feature selection is performed. The selected features are given to the Deep-ConvLSTM model to perform accurate predictions. Evaluation Based on evaluation metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), the proposed stock market forecasting model obtained the minimum MSE and RMSE of 7.2487 and 2.6923, which showed the effectiveness of the method in stock market forecasting [6]. Chen et al. proposed a new stock price prediction model that combines a Convolutional Neural Network (CNN), a Bidirectional Long Short-Term Memory (BiLSTM) network, and an Efficient Channel Attention (ECA). CNNs are used to extract deep features from stock data to reduce the effects of high noise and nonlinearity. BiLSTM network is used to predict stock prices based on the extracted deep features. At the same time, a novel ECA module is introduced into the network model, which further improves the sensitivity of the network to important features and key information. The results demonstrate the effectiveness and feasibility of the proposed CNN-BILSTM-ECA network model [7]. Lei et al. constructed investor attention factors through Baidu search antecedent index, and then combined other transaction information such as transaction volume, trend indicators, and quotation rate of change as input indicators, and finally used a deep temporal convolutional network (TCN) learning model to predict Volatility under high frequency financial data. Compared with the traditional econometric model, the multi-step forecast results of the TCN model have higher stability, provide a more accurate and reliable method for volatility forecasting of big data, and enrich the index system of volatility forecasting [8]. Scholars such as Yh proposed a new financial data prediction model based on Variational Mode Decomposition (VMD). In this model, the parameters of VMD are optimized by genetic algorithm. VMD then decomposes the data series into long-term and short-term trends. Finally, a long short-term memory (LSTM) network is used to predict future data, and the data input generated by the VMD is used. Experimental results show that the model has high accuracy in the advance forecasting of financial time series and outperforms the baseline model [9].

Researchers such as Bukhari AH proposed a novel hybrid model with fractional derivative strength and its dynamic characteristics of deep learning, long short-term memory (LSTM) networks to predict changes in financial markets; where the ARFIMA model-based filter has a linear trend in the data that outperforms the Autoregressive Integrated Moving Average model (ARIMA) model, and passes the residuals to the LSTM model, which has the help of an external dependent variable The residual value that captures the nonlinearity below. This model not only alleviates the volatility problem, but also overcomes the overfitting problem of neural networks [10]. Luo et al. propose a hybrid model that combines ensemble empirical mode decomposition (EEMD), autoregressive integral moving average (ARIMA), and Taylor expansion using tracking differentiators to forecast financial time series. Specifically, financial time series are decomposed into subsequences by EEMD. Then, the linear part of each subsequence is predicted by the linear ARIMA model, while the nonlinear part is predicted by the nonlinear Taylor expansion model. Combine predictions from linear and nonlinear models into predictions for each subseries. The final prediction result is obtained by combining the predicted values of all subsequences [11]. Scholars such as Darley Olufunke use the ARIMA model to analyze and model the price trend, and propose a model suitable for prediction. The results show that the ARIMA (6, 1, 12) model is the most suitable, based on a combination of the number of significant coefficients and volatility values, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). After two months of experiments, the results show that the prediction accuracy gradually decreases as the number of days in the test period increases [12].

The study proposes a PVD prediction model based on DBN, which performs PLR_VIP algorithm calculation and min-max normalization on the original financial time series to obtain training sample sets and test sample sets, trains and tests the DBN network to obtain appropriate network parameters, and then predicts Elliott wave patterns for financial time series.

III. PVD Financial Investment Prediction Model Based on Deep Belief Network

A. *Fusion of Deep Belief Network and Elliott Wave Model*

DBN is a model of deep learning technology, which is composed of stacking units. The training process of DBN includes two stages: pre-training and fine-tuning. The pre-training adopts an unsupervised greedy layer-by-layer learning strategy to train each Restricted Boltzmann Machine (RBM) from the bottom up, which can feature information in the original data and is extracted layer by layer to abstract the high-level feature representation of the data [13]. Fine-tuning adopts a supervised learning algorithm to adjust the parameters of the network to complete classification and recognition. RBM is a generative model with strong unsupervised learning ability. It consists of two parts: visible layer and hidden layer. Units between different layers are connected by full connection [14]. The structural framework of the DBN network is shown in Fig. 1.

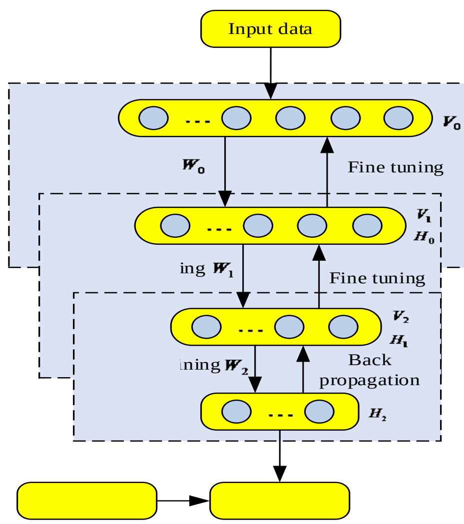


Fig. 1. Structure framework of DBN network

The visible layer consists n of neurons, $v=\left(v_1,v_2,\cdots v_n\right)$, where $v\in\left\{0,1\right\}^n$; the hidden layer consists m of neurons, $h=\left(h_1,h_2,\cdots h_m\right)$, where $h\in\left\{0,1\right\}^m$. A value of 1 indicates that the neuron is activated, otherwise it is not activated. The parameter $\theta=\left\{w,b,a\right\}$ determines the RBM model, in which w is the weight between the visible layer unit and the hidden layer unit, a representing the offset of the visible layer unit and the offset of the b hidden layer unit. The structure of the RBM is shown in Fig. 2.

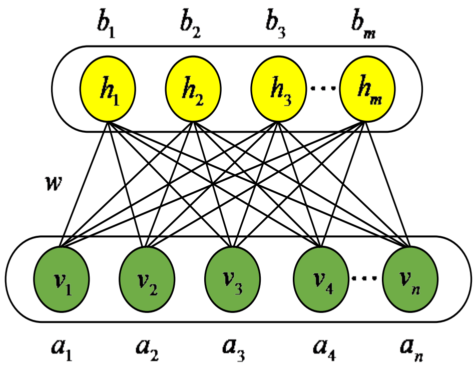


Fig. 2. Structure of RBM

The joint probability distribution of the visible layer and hidden layer units of RBM is shown in formula (1).

$$P(v,h)=\frac{1}{N}e^{-E(v,h)} \quad (1)$$

In formula (1), $N=\sum_{v,h}e^{-E(v,h)}$ is the normalization factor, which $E(v,h)$ is the energy function. It is assumed that the visible layer and the hidden layer of the RBM model obey the Bernoulli distribution, and the energy function is defined as shown in formula (2).

$$E(v,h)=-\sum_{j=1}^{n}a_j v_j-\sum_{i=1}^{m}b_i h_i-\sum_{i=1}^{m}\sum_{j=1}^{n}w_{ij}v_j h_i \quad (2)$$

Exponentiate and regularize the energy function to obtain the joint probability distribution formula, and sum up the states of all neurons in the hidden layer and the visible layer respectively to obtain the edge probability distribution of the visible layer and the hidden layer, as shown in Equation (3).

$$\begin{cases} P(v)=\sum_{h}P(v,h)=\frac{1}{N}\sum_{h}e^{-E(v,h)} \\ P(h)=\sum_{v}P(v,h)=\frac{1}{N}\sum_{v}e^{-E(v,h)} \end{cases} \quad (3)$$

After the state of the visible layer unit is obtained, i the conditional probability that the first unit of the hidden layer is activated can be calculated as shown in equation (4).

$$P(h_i=1|v)=\sigma\left(b_i+\sum_{j=1}^{n}w_{ij}v_j\right) \quad (4)$$

After the state of the hidden layer unit is obtained, j the conditional probability that the first unit of the visible layer is activated can be calculated as shown in formula (5).

$$P(v_j=1|h)=\sigma\left(a_j+\sum_{i=1}^{m}w_{ji}v_i\right) \quad (5)$$

where σ represents the sigmoid function, $\sigma(x)=\frac{1}{1+e^{-x}}$. The mean square error is used to measure the difference between the reconstructed output of the hidden layer to the visible layer and the input of the RBM.

Set a sample set $v=\{v_0,v_1,\cdots,v_K\}$ and train the RBM model. In order to make the joint probability distribution fit the distribution of the input samples as much as possible and make the system energy stable, it is necessary to find a suitable parameter $\hat{\theta}=\{w,b,a\}$. The maximization likelihood function is used $L(\theta)$, which is equivalent to maximizing the log-likelihood function $\ln L(\theta)$, as shown in equation (6).

$$\hat{\theta}=\arg\max_{\theta}L(\theta)=\arg\max_{\theta}\sum_{t=0}^{K}\ln P(v_t|\theta) \quad (6)$$

The gradient ascent method is used to update the parameters as shown in equation (7).

$$\theta^{*}=\theta+\eta\frac{\partial\ln P(v)}{\partial\theta} \quad (7)$$

Equation (7) η represents the learning rate, η >0. The log-likelihood function for a single sample is shown in Equation (8).

$$\ln P(v_0)=\ln\frac{1}{N}\sum_{h}e^{-E(v_0,h)}=\ln\sum_{h}e^{-E(v_0,h)}-\ln\sum_{v,h}e^{-E(v,h)} \quad (8)$$

The gradient formula of this sample is shown in formula (9).

$$\frac{\partial\ln P(v_0)}{\partial\theta}=-\sum_{h}P(h|v_0)\frac{\partial E(v_0,h)}{\partial\theta}+\sum_{v,h}P(v,h)\frac{\partial E(v,h)}{\partial\theta} \quad (9)$$

The joint probability distribution cannot be directly calculated, and it needs to be approximated by the sampling method, which $P(v,h)$ can be expressed as $P(h|v)P(v)$, and substitute it into the gradient formula, as shown in formula (10).

$$\frac{\partial\ln P(v_0)}{\partial\theta}=-\sum_{h}P(h|v_0)\frac{\partial E(v_0,h)}{\partial\theta}+\sum_{v}P(v)\sum_{h}P(h|v)\frac{\partial E(v,h)}{\partial\theta} \quad (10)$$

The update formula of the three parameters of the RBM is obtained. The traditional parameter update formula is more complicated and requires a large amount of calculation. Therefore, the Contrastive Divergence (CD) algorithm is used to update the parameters, usually only one step is needed to achieve better results. Using the k-CD algorithm, set a training set v, $\forall v\in v$ take the initial value v^0, and perform k step Gibbs sampling. When the first t step ($t=1,2,\cdots,k$) is reached, it is obtained by sampling, $P(h|v^{(t-1)})$ obtained $h^{(t-1)}$ by $P(v|h^{(t-1)})$ sampling $v^{(t-1)}$, and obtained after k

sub-Gibbs sampling, $v^{(k)}$ and an approximation of the parameters can be obtained. The gradient formula is as formula (11).

$$\begin{cases} \frac{\partial \ln P(v)}{\partial w_{ij}} \approx P\left(h_i = 1 \middle| v^0\right) v_j^{\,0} - P\left(h_i = 1 \middle| v^k\right) v_j^{\,k} \\ \frac{\partial \ln P(v)}{\partial b_j} \approx v_j^{\,0} - v_j^{\,k} \\ \frac{\partial \ln P(v)}{\partial c_i} \approx P\left(h_i = 1 \middle| v^0\right) - P\left(h_i = 1 \middle| v^k\right) \end{cases} \tag{11}$$

The update formula of the parameters can be estimated by Equation (12).

$$\begin{cases} w_{ij}^{\,*} \approx w_{ij} + \eta\left(P\left(h_i = 1 \middle| v^0\right) v_j^{\,0} - P\left(h_i = 1 \middle| v^k\right) v_j^{\,k}\right) \\ b_j^{\,*} \approx b_j + \eta\left(v_j^{\,0} - v_j^{\,k}\right) \\ c_i^{\,*} \approx c_i + \eta\left(P\left(h_i = 1 \middle| v^0\right) - P\left(h_i = 1 \middle| v^k\right)\right) \end{cases} \tag{12}$$

Set up a l DBN network with one hidden layer and calculate its joint probability distribution as Equation (13).

$$P(v, h_1, h_2, \cdots, h_l) = P(v|h_1) P(h_1|h_2) \cdots P(h_{l-2}|h_{l-1}) P(h_{l-1}, h_l) \tag{13}$$

Equation (13) $P(h_{l-1}, h_l)$ represents the probability distribution of the top layer of the RBM, which represents $P(h_k|h_{k+1})$ the conditional probability ($k = 0,1,2,\cdots,l-1$, where $h_0 = v$) of the hidden layer of h_k each RBM layer when the visible layer is known. h_{k+1} For a n training sample set with one sample $X = \{x_0, x_1, \cdots, x_n\}$, the data label set of the sample is $Y = \{y_0, y_1, \cdots, y_n\}$. The input of the pre-training process X does not require data labels. The k-CD algorithm is used to train the network parameters unsupervised, and the output value obtained by each layer of RBM is input into the next layer of RBM, and the feature vector set of the sample is obtained layer by layer. The pre-training process adjusts the RBM model of each layer to ensure the optimal output of this layer, but cannot guarantee the optimal output of the entire DBN. Therefore, the DBN network parameters need to be adjusted, and the network parameters Y are supervised top-down using the sample data label set [15].

There are two types of Elliott wave cycle modes: driving and adjusting. An impulsive wave is a homeopathic wave that pushes the market to change in the direction of the main trend. It is generally composed of five sub-waves with a simple form, and can be divided into conventional impulsive waves and unconventional impulsive waves. A regular motive wave consists of five waves 1-2-3-4-5 directed upwards or downwards. After the fifth wave ends, there will be a movement in the opposite direction of the trend. This movement is a corrective wave, which is a partial retracement of the driving relative to the previous driving wave of the same level. The structure of corrective waves is more complex than that of driving waves, mainly including three forms: zigzag, flat and triangle [16-17].

B. 2.2 Establishment of PVD Prediction Model based on Deep Belief Network

When analyzing and forecasting financial markets, financial time series are often used to extract financial transaction data. Time series is a collection of measured values obtained over time. The core is to use reasonable technical methods to ignore the small fluctuations of the time series to obtain the concept of data sawtooth, and to identify the corresponding time series on different time scales pattern [18]. Time variables and data variables together make up each data unit of the time series. The research uses deep learning combined with Elliott wave theory to establish a financial market prediction model, extracts transaction price information and abstracts it as a time series, extracts features through deep learning model and obtains the Elliott wave pattern corresponding to the original time series [19]. Since the lengths of the time series corresponding to the same wave pattern are different, but the length of the input sequence needs to remain the same, the length of the input data needs to be fixed, and the original data is preprocessed before input, that is, the data is processed Dimensionality reduction and compression, discarding parts that have less impact on the global. The research uses the PLR_VIP algorithm to process the original data [20].

The basic principle of piecewise linear is to set a n dimensional time series $S = \{s_1, s_2, \cdots, s_n\}$ and extract a new series S of k ($k << n$) pieces that construct an approximate series according to it $S^* = \{s^*_1, s^*_2, \cdots, s^*_{k+1}\}$. Using the PLR method, the adjacent continuous line segments can be approximated as the original time series. The selection of segment points has different degrees of influence on the time series structure, and the degree of influence can be expressed as importance. Select important points on the original time series, keep important points and discard unimportant points. In order to fix the step size of the time series, the segmentation condition is set to a fixed number of segments. When the interval of the time series is fixed, the importance of a point is measured by the distance of the point from the endpoint of the region where the point is located. The raw time series is preprocessed with vertical distance as a measure. An example of the PLR_VIP algorithm is shown in Fig. 3.

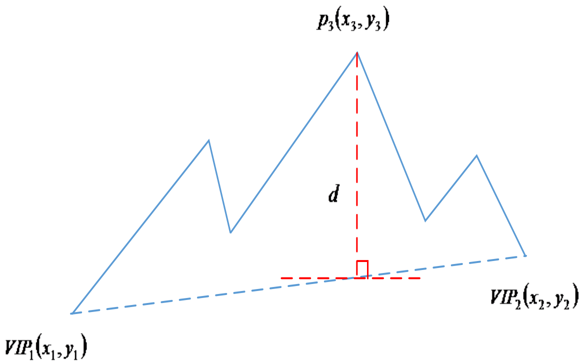


Fig. 3. PLR_VIP algorithm example

The time series shown in Fig. 3 takes the VIP_1 sum as the end VIP_2 point, and the vertical distance between the d calculation point and the two end points is p_3 shown in formula (14).

$$d = \left| \left(y_1 + (y_2 - y_1) \frac{x_3 - x_1}{x_2 - x_1} \right) - y_3 \right| \quad (14)$$

VIP_1 sum VIP_2 is two important points, find the point with the farthest vertical distance from the two endpoints p_3 as the next important point VIP_3, find the point with the farthest vertical distance in the area with the VIP_1 sum VIP_3 as the endpoint and the VIP_2 sum VIP_3 as the endpoint respectively and mark it For important points, the above process is iterated until the number of important points reaches a predetermined number. During the training process of the network, the data of the order of magnitude will mask the contribution of the data of the smaller order of magnitude to the network, thus affecting the accuracy of the network, and the cardinality of the sample and the average value are quite different, which will reduce the convergence speed of the network. Therefore, standardize the sample data, use the min-max normalization method to map the data to the interval [0, 1], and x perform min-max normalization on the first i data of the sample as shown in formula (15). x_i

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (15)$$

The study designed eight types of Elliott wave patterns, each of which consists of eight sub-waves, as shown in Fig. 4.

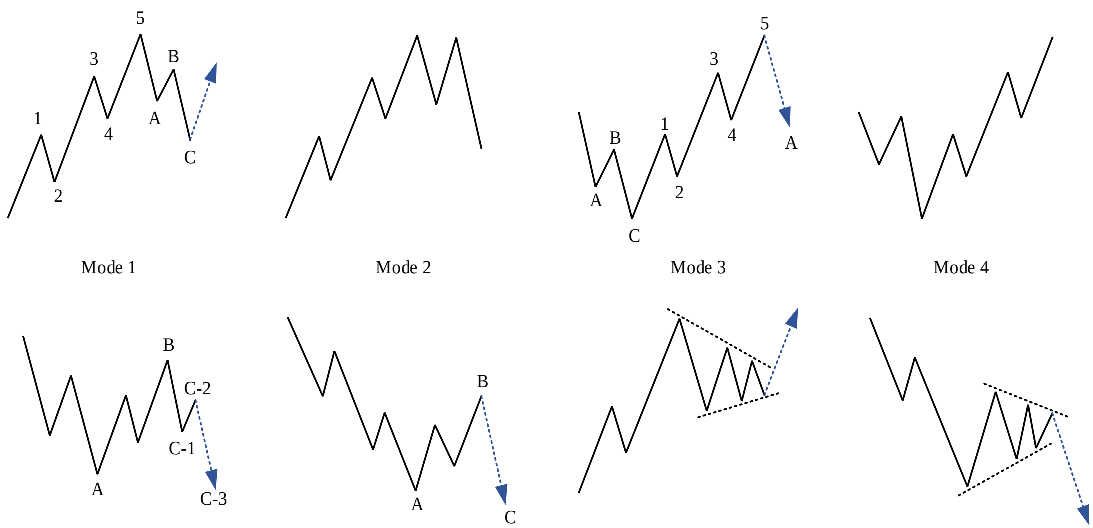


Fig. 4. 8 Elliott Wave Model

Pattern 1 consists of driving waves and zigzag waves. When the financial time series is identified as pattern 1, it means that the market is about to enter a rising stage, and there may be a new round of bull or bear market upward adjustment. Wave 3 is the key sub-wave of the bull market, and wave C is the key sub-wave of the bear market, both of which are the strongest waves under normal conditions. Pattern 2 consists of motive waves and flat correction waves. The types of correction waves are different from those of pattern 1, including spread and trend. When the financial time series is identified as mode 2, it means that the market has entered a rising period and will usher in a trading opportunity. The attribute of the market rising is the same as that of mode 1. Pattern 3 consists of ZigZag and Motivation. When a financial time series is identified as pattern 3, it indicates that the market will enter a downward phase. The falling phase may result from a pullback from the previous driving wave, or it may usher in a new bear market. Pattern 4 is composed of a flat adjustment wave and a driving wave. The type of adjustment wave is different from that of pattern 3. The flat adjustment wave of pattern 4 includes spread and trend. When the financial time series is identified as pattern 4, it means that the market will enter an adjustment phase, and the property of the market falling is the same as that of pattern 3. Pattern 5 represents part of a flat correction, its role is to predict the main decline in the market, and its end indicates the beginning of the main decline. Pattern 6 is an inversion of Pattern 1 and appears as part of a ZigZag. When the financial time series is identified as pattern 6, it means that the financial market is about to enter the main down stage. The purpose of model prediction is to judge the future trend of the financial market, so there is no need to consider whether the triangle is an unconventional driving wave or a corrective wave. When the financial time series completes pattern 7, the financial market will start to rise, and after pattern 8 it will start to fall.

The PVD prediction model proposed by the research mainly includes three parts: PLR_VIP algorithm, min-max normalization processing and DBN network. PVD forecasting models can identify and judge Elliott wave patterns in financial time series. The PLR_VIP algorithm can retain the important points of the original financial time series and standardize the step size of the financial time series. The min-max normalization method can standardize the range of data changes and improve the accuracy and convergence speed of the network. Perform PLR_VIP algorithm calculation and min-max normalization on the original financial time series to obtain a training sample set and a test sample set, train and test the DBN network to obtain appropriate network parameters, and then predict the Elliott wave pattern of the financial time series. The specific process of the PVD prediction model is shown in Fig. 5.

To test the prediction accuracy and performance of the proposed model, four performance indicators are selected as evaluation criteria. The lower the indicator value, the more accurate the prediction results of the prediction model are, as shown in Table I.

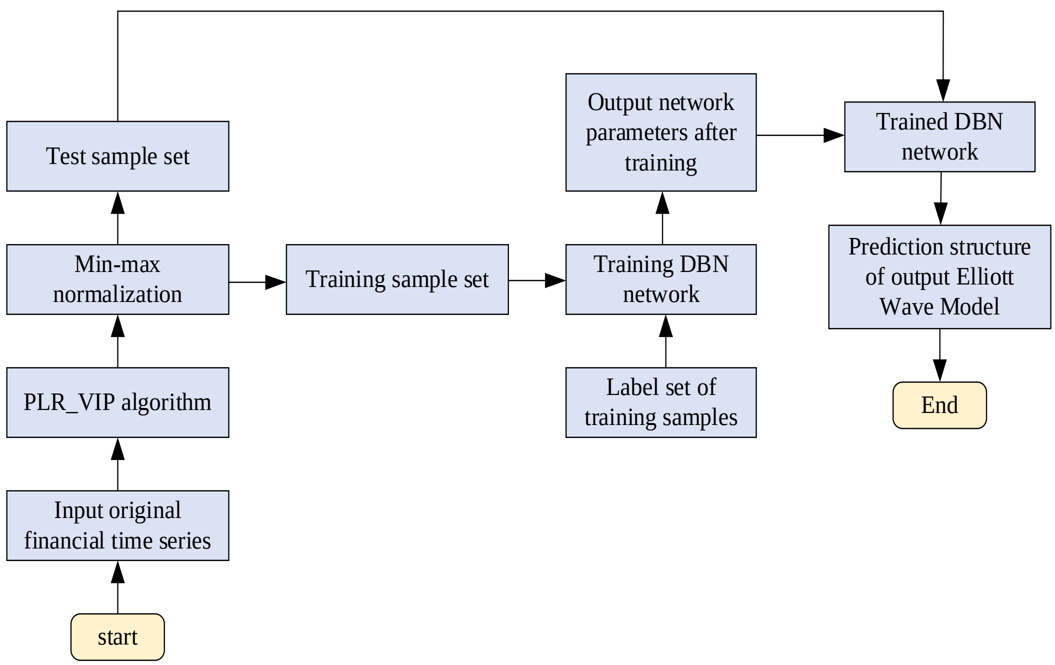


Fig. 5. Specific process of PVD prediction model

TABLE I. PERFORMANCE INDEX

Performance index	Calculation formula
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(T_i - P_i\right)^2}$
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(T_i - P_i\right)^2$
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\|T_i - P_i\right\|$
ER	$ER = \frac{\sum\left\{1 \middle\| Model_P \neq Model_T\right\}}{\sum\left\{1 \middle\| Model_P = Model_P\right\}} \times 100\%$

Table I shows that root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) and relative error (ER) are used to evaluate the prediction performance of the model.

The study selected some data from the historical price data of more than 10 trading varieties in the global stock index, foreign exchange market and bulk commodity markets for the experiment. A total of 6000 relevant financial data were collected in this experiment, 4829 of which were from Internet data and 1171 from books, newspapers and other tools. Take 20% of all data as the test set and the remaining 80% as the training set.

IV. Experimental Analysis of PVD Prediction Model Based on Deep Belief Network

After preprocessing through the PLR_VIP algorithm and the min-max normalization method, the original financial time series sample data of the DBN network for training and testing are obtained. The obtained sample data is input into the DBN network for iterative training. As the number of iterations increases, the training error of the DBN network is shown in Fig. 6.

Using the PVD forecast model to classify and predict the Elliott wave pattern of the financial time series, the classification results show that the MSE is 0.4015, and 70.21% of the financial time series in the sample have obtained the correct classification result, that is, the prediction accuracy of the model is 70.21%. It shows that the PVD forecasting model can effectively identify the Elliott wave pattern of financial time series, and can more accurately predict the trend of the financial market. Taking the S&P 500 index, the US dollar index and WTT crude oil as examples, the prediction effect of the model is shown in Fig. 7.

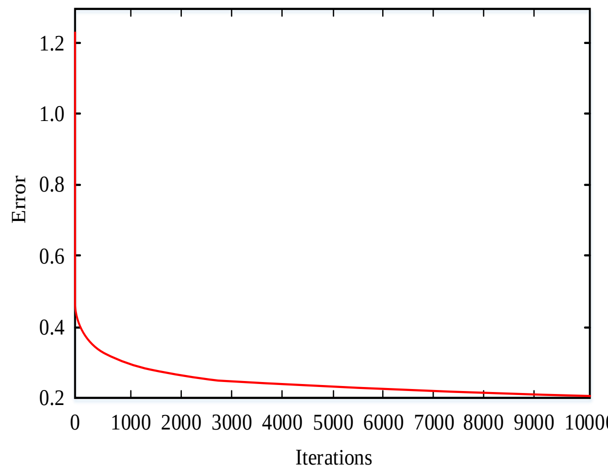


Fig. 6. Training error of DBN network

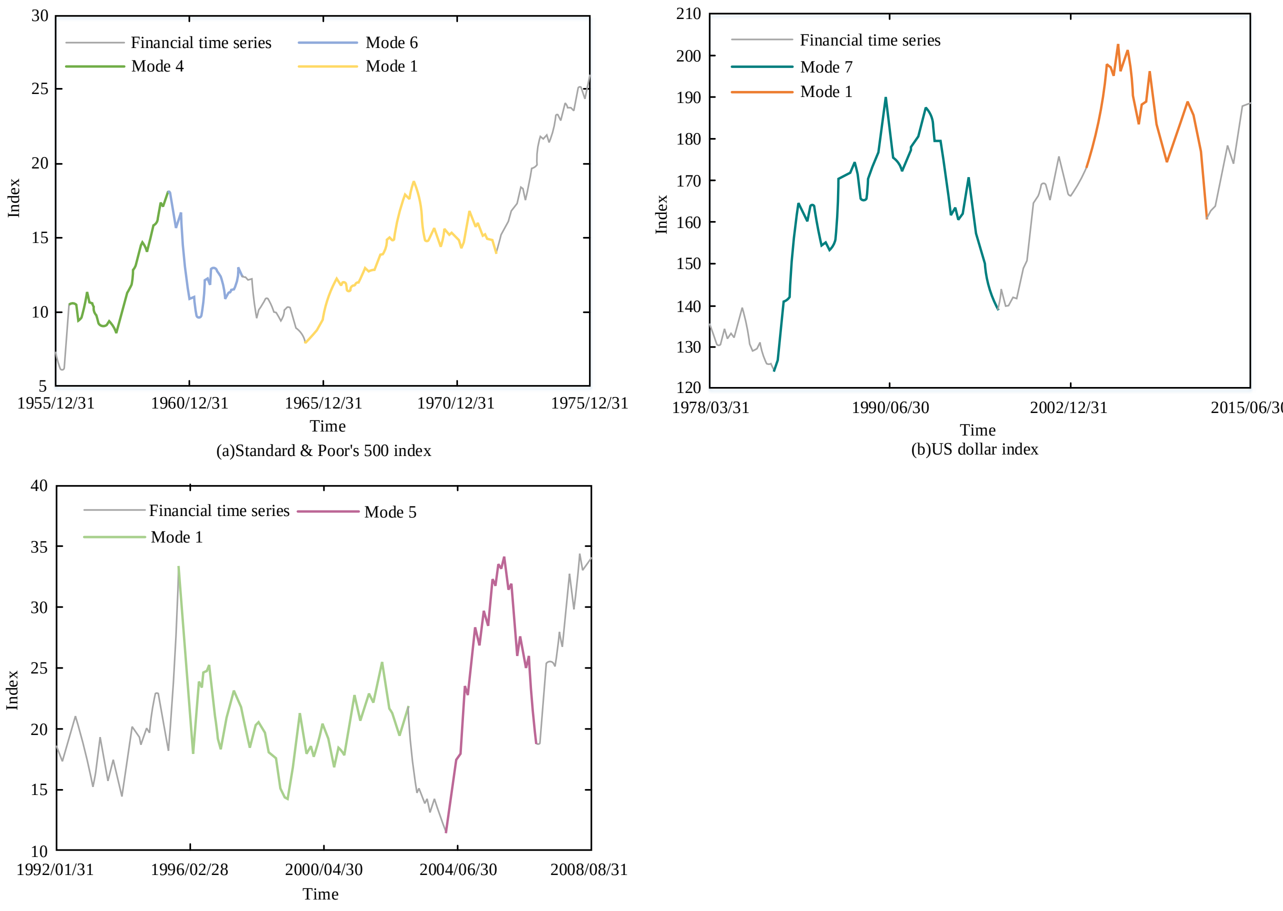


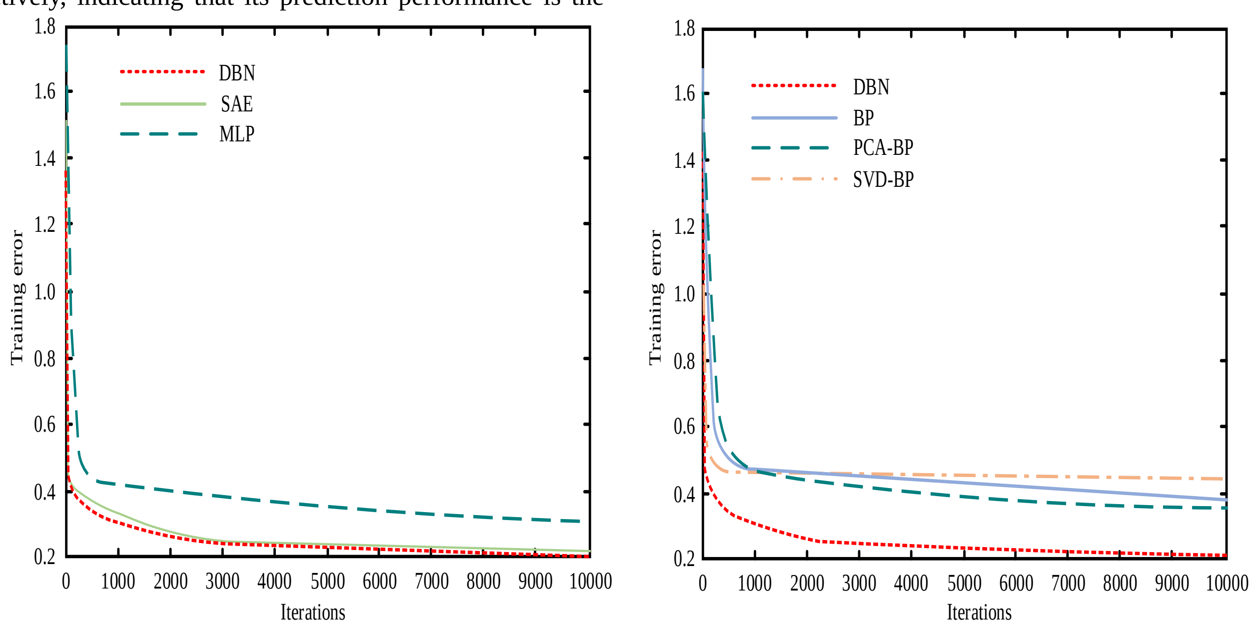
Fig. 7. Prediction effect of the model

The S&P 500 index has a monthly cycle, and its price data is shown in Fig. 7(a). The PVD model accurately discriminates the financial time series between June 1956 and February 1960 as mode 4, and the 1960-2. The financial time series between March 1965 and May 1972 is identified as mode 6, and the financial time series between March 1965 and May 1972 is identified as mode 1. The U.S. dollar index takes a quarterly cycle, and its price data is shown in Fig. 7(b). The PVD model accurately discriminates the financial time series from June 1982 to December 1997 as mode 2, and from December 2003 to 2012. The financial time series June of 2008 is identified as mode 7. WTT crude oil has a monthly cycle, and its price data is shown in Fig. 7(c). The PVD model discriminates the financial time series from September 1995 to September 2002 as mode 5. The financial time series between months is identified as mode 1. A comparative experiment was carried out on the PVD prediction model and the other five models, and the validity of the six prediction models was verified, but the prediction effect of each model was different. The performance comparison of the six models is shown in Table II.

TABLE II. PERFORMANCE COMPARISON OF MODELS

Model	Deep network model			Shallow network model		
	PVD	SAE	MLP	BP	SVD-BP	PCA-BP
RMSE	0.6336	0.6653	0.8367	0.8634	0.9312	0.8314
MSE	0.4015	0.4426	0.7001	0.7455	0.8671	0.6912
MAE	0.9052	0.9648	1.4671	1.4702	1.6310	1.3064
IS	29.79%	35.86%	44.35%	48.49%	78.23%	56.85%

It can be seen from Table II that the deep network model performs better than the shallow network model on the four evaluation indicators, and the mean values of RMSE, MSE, MAE and ER of the deep network model are 0.7119, 0.5147, 1.1124 and 36.67%, respectively. The mean values of RMSE, MSE, MAE and ER of the shallow neural network are 0.8753, 0.7679, 1.469 and 61.19%, respectively. The prediction effect of the deep network model is significantly better than that of the shallow network model. Compared with the four indicators of the shallow network model, the deep network models are reduced by 18.67%, 32.97%, 24.27% and 40.07%, respectively. Among the six prediction models, the values of the four evaluation indicators of PVD are the lowest among all models, which are 0.6336, 0.4015, 0.9052 and 29.79%, respectively, indicating that its prediction performance is the best and significantly better than other models. The values of RMSE, MSE, MAE and ER of the SVD-BP model are the highest, indicating that its prediction effect is the worst and the prediction accuracy is the lowest. Although the training error of the PCA-BP model is smaller than that of the BP network model, its error rate is higher than that of the BP network model, indicating that the algorithm has over-fitting; weaker ability. The main reason why the prediction accuracy of PVD model is significantly higher than that of other models is that the research has introduced in-depth learning into the prediction model, so it can achieve real-time analysis with the help of financial data, and the prediction results will be more accurate. The variation of training error between the deep belief network in the PVD model and other models is shown in Fig. 8.



(a)Comparison of training error with depth network model

(b)Comparison of training error with shallow network model

Fig. 8. Comparison of training error changes

It can be seen from Fig. 8(a) that the error curves of the DBN network and the SAE network are not much different, but the error curve of the DBN is smoother, indicating that the DBN has better stability. The error curve of the MLP network is very flat, indicating that it has higher stability and faster convergence speed, but its training error is higher than that of the DBN network, which is caused by the network falling into a local optimum. It has higher reliability and accuracy, and can effectively improve the local optimal problem caused by the BP algorithm. Compared with the shallow network model, the DBN network has faster convergence speed and higher accuracy. In the classification training process, compared with other models, the DBN network has the fastest convergence speed, the smallest training error and good stability. In general, the PVD forecasting model proposed in this study has high reliability and stability, and can efficiently and accurately identify Elliott waves of various patterns in financial time series. Compared with the other five models, the PVD model has higher stability and accuracy, faster convergence speed due to the use of DBN network, and also has excellent performance in feature extraction. The PVD model improves the prediction performance of the traditional BP network and multi-level classifier, reduces the error of the deep network model, and improves the accuracy of the model with high-frequency trading as the decision criterion. Experiments show that under the background of sustainable development, the PVD forecasting model proposed in the study performs well in financial investment forecasting, which provides a reference for the development of financial investment forecasting in the e-commerce industry.

V. CONCLUSION

With the rapid development of the e-commerce industry, the financial investment forecast of the e-commerce industry has gradually become a concern of people, and how to do a good job of financial investment forecast in the context of sustainable development is the focus of the industry. Researched and proposed a PVD prediction model based on DBN, and used the deep learning model to model the financial time series to achieve financial investment prediction. The experimental results show that the MSE of the PVD model is 0.4015 and the prediction accuracy is 70.21%, indicating that it can efficiently and accurately identify the Elliott wave pattern of financial time series. The PVD model is compared with the other five models. Among all the prediction models, the values of the four evaluation indicators of PVD are the lowest among all models, which are 0.6336, 0.4015, 0.9052 and 29.79% respectively, indicating that its prediction effect is the best, with the highest prediction accuracy, with good accuracy and stability, and significantly outperforms other models. Compared with the training error changes of other models, it can be seen that the error curve of the DBN network is smoother and the training error is smaller, indicating that it has higher stability, faster convergence speed, and higher reliability and accuracy. Experiment shows that under the background of sustainable development, the PVD prediction model proposed in the study has excellent stability and reliability in financial investment prediction, and can effectively predict the trend of financial investment. However, the research fails to realize the fractal level concept of Elliott wave theory. If the neural network can be combined with the fractal level concept, the prediction performance and prediction accuracy of the model will be effectively improved. Therefore, in the follow-up work, the research will try to combine the concept of neural network and fractal level, propose a more perfect prediction model, and hope to achieve accurate prediction of financial investment in e-commerce industry, and provide strategies for the development of China's economic market.

REFERENCES

[1] A. Ry, Y. B. Lin and D. Yzc, et al., “Big data analytics for financial Market volatility forecast based on support vector machine – ScienceDirect,” International Journal of Information Management, Vol. 50, pp. 452-462, 2020.

[2] F. Wang, M. Li and Y. Mei, et al., “Time Series Data Mining: A Case Study with Big Data Analytics Approach,” IEEE Access, Vol. 8, pp. 14322-14328, 2020.

[3] P. Huang, “Big data application in exchange rate financial prediction platform based on FPGA and human-computer interaction – ScienceDirect,” Microprocessors and Microsystems, Vol. 80, pp. 1-6, 2020.

[4] S. Puiu, “E-Commerce and the Factors Affecting Its Development in the Age of Digital Technology: Empirical Evidence at EU-27 Level,” Sustainability, Vol. 14, No. 1, pp. 1-17, 2021.

[5] M. Kare and M. Porada-Rochoń, “Forecasting financial cycles: can big data help,” Technological and Economic Development of Economy, Vol. 26, No. 5, pp. 1-15, 2020.

[6] A. Kelotra, P. Pandey, “Stock Market Prediction Using Optimized Deep-ConvLSTM Model,” Big Data, Vol. 8, No. 1, pp. 5-24, 2020.

[7] Y. Chen, R. Fang, T. Liang, et al., “Stock Price Forecast Based on CNN-BiLSTM-ECA Model,” Scientific Programming, Vol. 5, pp. 1-20, 2021.

[8] B. Lei, B. Zhang, Y. Song, “Volatility Forecasting for High-Frequency Financial Data Based on Web Search Index and Deep Learning Model,” Vol. 9, No. 4, pp. 1-17, 2021.

[9] A. Yh, B. Yg, G. Yan, et al., “A new financial data forecasting model using genetic algorithm and long short-term memory network – ScienceDirect,” Neurocomputing, Vol. 425, pp. 207-218, 2021.

[10] A. H. Bukhari, M. Raja, M. Sulaiman, et al., “Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting,” IEEE Access, Vol. 8, pp. 71326-71338, 2020.

[11] Z. Luo, W. Guo, Q. Liu, et al., “A hybrid model for financial time-series forecasting based on mixed methodologies,” Expert Systems, Vol. 38, No. 2, pp. 1-16, 2020.

[12] G. DarleyOlufunke, “Price Analysis and Forecasting for Bitcoin Using Auto Regressive Integrated Moving Average Model,” Annals of Science and Technology, Vol. 6, No. 2, pp. 47-56, 2021.

[13] Y. Wang, Y. Guo, “Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost,” China Communications, Vol. 17, No. 3, pp. 205-221, 2020.

[14] W. Xie, N. Metawa, “Financial stock market forecasting based on intelligent support vector machine regression model,” Journal of Intelligent and Fuzzy Systems, No. 2, pp. 1-10, 2021.

[15] GeWenbo, LalbakhshPooia, IsaiLeigh, et al., “Neural Network–Based Financial Volatility Forecasting: A Systematic Review”, ACM Computing Surveys (CSUR), Vol. 55, No. 1, pp. 1-30, 2022.

[16] F. Kamalov, I. Gurrib, K. Rajab, “Financial Forecasting with Machine Learning: Price Vs Return”, Journal of Computer Science, Vol. 17, No. 3, pp. 251-264, 2021.

[17] I. E. Livieris, E. Pintelas, P. Pintelas, “A CNN–LSTM model for gold price time-series forecasting,” Neural Computing and Applications, Vol. 32, pp. 17351-17360, 2020.

[18] D. K. Mademlis, N. Dritsakis, “Volatility Forecasting using Hybrid GARCH Neural Network Models: The Case of the Italian Stock Market,” International Journal of Economics and Financial Issues, Vol. 11, No. 1,

pp. 49-60, 2021.

[19] R. Lee, “Chaotic Type-2 Transient-Fuzzy Deep Neuro-Oscillatory Network (CT2TFDNN) for Worldwide Financial Prediction,” IEEE Transactions on Fuzzy Systems, Vol. 28, No. 4, pp. 731-745, 2020.

[20] M. Lu, “A monetary policy prediction model based on deep learning”, Neural Computing and Applications, Vol. 32, No. 10, pp. 5649-5668, 2020.

Embedded Monitoring Method of Greenhouse Environment based on Wireless Sensor Network

Weixue Liu
Chifeng University, Teaching Staff Development Center, Chifeng, 024000, China

***Abstract*—Aiming at the problem of low accuracy of greenhouse environmental monitoring, an embedded monitoring method of greenhouse environment based on wireless sensor network is studied. The embedded microprocessor S3C2410 is selected as the control chip of greenhouse environmental monitoring. The wireless sensor network is composed of wireless detection node, sink node and remote-control terminal. The temperature and humidity sensor, CO_2 concentration sensor and light intensity sensor are set as the detection nodes of the wireless sensor network. Each detection sensor node is configured with wireless communication module to form a wireless sensor network to realize data communication of environmental monitoring. The distribution map method is selected to eliminate the missing data of greenhouse environmental monitoring data collected by wireless sensor network, and the weighted average method is used to fuse the monitoring data after eliminating the missing data to obtain the final greenhouse environmental monitoring results. The experimental results show that this method can effectively detect the environmental data such as temperature, humidity and light intensity of greenhouse, and the relative error of each data is less than 1%.**

Keywords*—Control chip; Detection node; embedded; greenhouse environmental monitoring; missing data; wireless sensor networks

I. Introduction

Facility agriculture is a modern form of agriculture that uses engineering techniques for the efficient production of plants and animals under relatively controlled environmental conditions. Facility agriculture covers facility farming, facility breeding and facility edibles. In addition to this, facility agriculture can improve or create environmental meteorological factors on a local scale to provide good environmental conditions for plant growth and efficient production. Facility agriculture is represented by the intelligent greenhouse, which embodies the direction of modern agriculture. Through greenhouse monitoring methods, various greenhouse environmental factors affecting plant growth can be obtained [1], which plays an important role in improving the yield and efficiency of crops in greenhouses. Among them, how to obtain accurate and reliable greenhouse environmental information is the basis for realising intelligent greenhouses, so information transmission is an indispensable and important link. According to different information transmission methods, greenhouse environmental information can be transmitted via wired communication or wireless communication [2].

Wiring is an unavoidable problem in wired communications. Due to the increasing complexity of the overall structure of modern greenhouses, the use of wired communication will significantly increase the complexity of design, installation and maintenance. Once a node is damaged, it can lead to the collapse of the entire communication network. Compared to wired communication, wireless communication offers better prospects for application. Infinite communication not only requires no wiring and is easy to maintain at a later stage, but also has the advantage of being able to delete measurement nodes at will [3].

In greenhouse environmental monitoring, the monitoring of various greenhouse environmental factors that affect plant growth is the most critical. The accurate monitoring of various greenhouse environmental parameters can effectively automate and efficiently produce the greenhouse environment [4]. Because facility agriculture gets rid of the constraints of natural climate conditions in traditional agricultural production and overcomes the constraints that traditional agriculture is difficult to solve, it is necessary to monitor greenhouse environmental factors such as temperature, humidity, light intensity and carbon dioxide concentration. China's greenhouse environment has the characteristics of high humidity, high acidity, less infrastructure, many kinds of crops and dynamic changes [5]. Through investigation and analysis, greenhouse monitoring needs mainly include: real-time data acquisition, display and processing of environmental factors such as atmospheric temperature, humidity, light intensity and CO_2 concentration. Traditional sensors are connected by wires, which are very troublesome to wire a large number of wires in greenhouses and are not conducive to agricultural cultivation. Therefore, the application of wireless transmission technology in greenhouses helps to solve the limitations of wired transmission [6]. On the other hand, many agricultural monitoring systems use microcomputers as central controllers, which occupy large space, are costly, and have high environmental requirements [7].

Wireless sensor networks are a research hotspot in the field of information technology in recent years, integrating technologies from many fields, such as sensing technology, computer science, signal and information processing, and communication. The fact that wireless sensors are so obviously interdisciplinary [8] has led researchers formerly working in some specialized fields to start intervening in this field from different perspectives and studying wireless sensor networks from different aspects.

With the development of wireless communication technology, the information transmission method of wireless sensor network began to be applied to the remote monitoring of greenhouse environment. Wireless sensor networks have the characteristics of wide coverage and low influence by the environment, and can be applied to distributed remote data collection and monitoring. Wireless sensor networks transmit data in the form of short messages, which have the advantages of flexible node deployment and small regional restrictions [9]. As long as there is wireless sensor network signal coverage, nodes can be deployed to collect data. As a new information acquisition and processing technology, wireless sensor network has gradually penetrated into the field of agriculture with its advantages of low power consumption, low cost and high reliability. Greenhouse is a relatively closed environment with limited self-regulation ability. In order to meet the requirements of agricultural production, it is necessary to artificially regulate various environmental parameters of greenhouse in order to create a more suitable environment for crop growth [10]. Monitoring greenhouse environmental information has become a key means to realize the automation and efficiency of greenhouse production. The traditional greenhouse monitoring method is based on wired communication, which has a series of problems, such as complex wiring, difficult maintenance, unable to flexibly deploy sensor nodes and so on, which limits the popularization and application of greenhouse detection methods to a certain extent. With the rapid development of modern information technology, WiFi, Bluetooth, UWB, RFID, ZigBee and other wireless communication technologies have emerged one after another. Among them, WiFi and Bluetooth cannot be popularized and applied on a large scale in the field of greenhouse monitoring due to their high cost and high-power consumption [11].

The history of embedded chips is inevitable, because embedded chips have many advantages. Non-embedded computing chips have good real-time performance, but embedded chip solutions have better advantages in terms of development cost, hardware resources and performance, in line with the development trend. Currently, embedded controllers are used almost everywhere [12]. Mobile devices, represented by cell phones, are the fastest growing embedded industry in recent years. At present, the application of embedded technology has penetrated into various fields such as industry, agriculture, education, national defense, scientific research and daily life. It has played an extremely important role in promoting technological transformation, product upgrading, accelerating the automation process and improving productivity in various industries.

There are many studies on greenhouse environmental monitoring. Pisanu et al. designed a low-cost electronic platform for greenhouse environmental monitoring based on the perspective of agriculture 4.0[13]. This platform can effectively monitor the greenhouse environment, but it has the drawback that the monitoring data cannot be transmitted to the control platform in time; Khan et al. applied the Internet of Things technology to greenhouse plant environment monitoring [14]. The accuracy of greenhouse environment monitoring is low because the sensors set cannot meet the needs of greenhouse environment monitoring.

For the shortcomings of the above two methods, an embedded greenhouse environment monitoring method based on wireless sensor network is studied. As a new information collection and processing technology, ZigBee-based wireless sensor network has the characteristics of large node size, small size, low cost and ad hoc network. Therefore, it has a broad application prospect in the field of agricultural environment monitoring. This research collects greenhouse environment information through temperature sensors, humidity sensors and other sensors, and transmits the collected information to remote monitoring center through ZigBee wireless sensor network, which is convenient for greenhouse managers to monitor greenhouse environment remotely and in real time and provide a good environment for crop growth.

II. Materials and Methods

A. *Overall Structure of Embedded Greenhouse Environmental Monitoring based on Wireless Sensor Network*

The overall structure of the embedded greenhouse environmental monitoring method based on wireless sensor network is shown in Fig. 1.

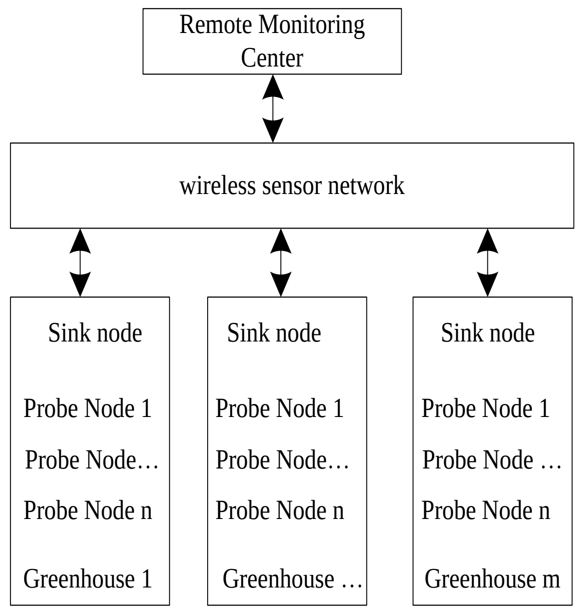


Fig. 1. The overall structure of the embedded greenhouse environmental monitoring method

As can be seen from Fig. 1, the wireless sensor network applied to embedded greenhouse environment monitoring mainly consists of wireless detection nodes, sink nodes and remote control terminals. Each detection sensor node is equipped with a wireless communication module and organized into a wireless sensor network, thus enabling mutual communication and greenhouse environment monitoring. The transmission of information in the network is achieved by "multi-hop" routing, which helps to improve the wireless communication distance and reliability. Each greenhouse has a sink node, which is equipped with a wireless communication module to realize data transmission for remote monitoring of the greenhouse environment network.

B. *Wireless Sensor Communication Module*

The functional structure of wireless sensor network for monitoring greenhouse environment is shown in Fig. 2.

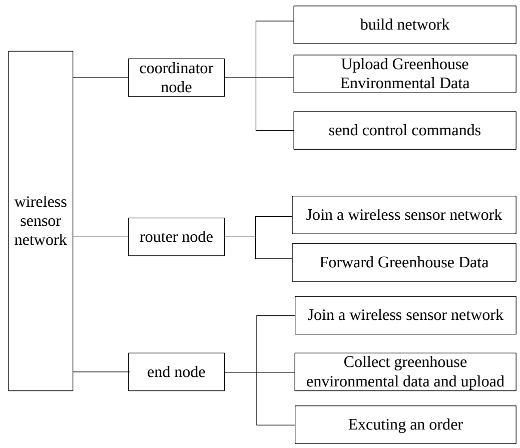


Fig. 2. Functional structure diagram of wireless sensor network

In wireless sensor networks, end nodes are responsible for collecting greenhouse environmental data and sending it to routing nodes. The routing node aggregates the data and sends it to the coordinator node, which then sends the data to the server using the router node [15]. The coordinator node receives action instructions from the server and transmits them to the routing node, which then sends the data in the form of broadcasts to the corresponding terminals for control operations. Throughout the network communication process, each node determines whether the command executor is the node based on its own 16-bit short address. If it is the node, it performs the corresponding execution action; otherwise, it ignores the command.

1) Communication chip: The CC2530 chip from TI was selected as the wireless communication module. 8051 CPU used in the CC2530 chip is a single-cycle 8051-compatible core that integrates 8KB of SRAM, Flash memory and an XREG/SFR register. The CC2530-F256 chip of the CC2530 series is selected as the communication chip for greenhouse environment monitoring. The CC2530 chip is very suitable for ultra-low-power greenhouse environment monitoring. The circuit composition of the CC2530 chip includes clock circuit and reset circuit. The clock circuit frequency of the CC2530 chip is composed of two borderless when the communication module is operating normally, the function of Y1 (33.549 KHz) crystal oscillator is to provide low-power mode for the node; the function of Y2 (33MHz) crystal oscillator is to provide clock for the node; the reset circuit is composed of resistor and on/off key. When the on/off key is pressed, the wireless sensor network will return to the initial state. The RF part of the CC2530 chip is responsible for transmitting and loading RF signals. The enhanced 8051 CPU core integrated in the chip is responsible for controlling the RF, peripherals and running the ZigBee protocol stack.

2) Building zigbee wireless sensor networks: When building a ZigBee wireless sensor network, the coordinator needs to be powered on while the communication module initializes the protocol stack and hardware. The coordinator sends a BEACON frame from the first channel. If BEACON responds, it can be concluded that there is another coordinator on this channel. If this occurs, it indicates that the channel is occupied and the coordinator sends BEACON frames to the next channel. This is repeated until no corresponding event for a BEACON frame occurs in a channel. After an empty channel is found, a wireless sensor network can be established. Once the network is established, the coordinator can connect to the device and then the coordinator assigns the network address. The coordinator can then read the data information and then send the transmitted or collected data information to the main control computer. When the coordinator receives an ASSOCIATION_REQ network access request from a ZigBee device, it will send the BEACON signal in response to the node's request [16] and then send the request signal ADDRESS_RQE; if the coordinator receives it, it will send the ADDRESS_ASSIGN signal and wait for an acknowledgement; if the coordinator receives ADDRESS_ASSIGN, the address list will be updated automatically. After power-up, the node will scan the channel to find an available coordinator in the network. After monitoring the beacon signal, the node will send the address assignment request signal ADDRESS_REQ, and then the node will search for the ADDRESS_ASSING signal. If it does not find one, it will continue to send the ADDRESS_REQ signal until it receives the ADDRESS_ASSING signal. After receiving the signal, the node will send ADDRESS_AKC to the coordinator to confirm the signal address.

3) CSMA / CA protocol and network interception strategy: Since the commonly used CSMA / CD protocol may be difficult for the adapter to detect whether there is channel conflict for wireless products, this wireless sensor network adopts a new protocol, Carrier Sense Multiple Access with Collision Avoidance (CSMA / CA). CSMA /CA protocol solves the conflict problem caused by multiple wireless devices transmitting data at the same time, and brings convenience to the construction of ZigBee wireless sensor network. CSMA / CA protocol needs to complete the following two tasks: the first is to find the idle channel (carrier listening); the second is to wait for random time (avoiding collision).

a) IEEE802. 15.4 CSMA / CA classification: In IEEE802 15.4 in the protocol, there are two ways of multiple access / collision avoidance mechanism for carrier frame: CSMA / CA with time slot and CSMA / CA without time slot. Generally speaking, CSMA / CA with time slot is used for beacon trigger mode, and CSMA / CA without time slot is used for beacon trigger mode. The two algorithms used in CSMA /CA are based on backoff, and the time unit is the backoff period.

b) CSMA / CA without beacon: The operation modes of the two protocols of IEEE802. 15.4 are similar, but the backoff time is different. In the network, if there is no super-frame structure, there will be no beacon (time slot) and no network coordinator for signal use. The sensor needs to wait for random time to listen to the activity of the channel. If the channel is idle, it can send information. Otherwise, the device needs to wait for random time again.

c) CSMA / CA with beacon: CSMA / CA with beacon limits the space occupied by the backoff cycle boundary, and the backoff cycle boundary must be at the super-frame cycle boundary, so as to ensure normal operation.

d) IEEE802. 15.4 CSMA / CA process: CSMA / CA protocol is mainly used for data transmission of ZigBee devices to avoid mutual interference of signals sent by different ZigBee devices. When the ZigBee device transmits data, it needs to find an idle channel, and it will send a signal for channel competition. If an idle channel is found, the ZigBee device will send a backoff delay time, which is used to prevent other ZigBee devices from occupying this channel at the moment and avoid interference caused by the collision of signals sent by the two devices. However, if the channel found by ZigBee device is busy, the device needs to listen to the channel again and repeat the above process until it finds a free channel.

C. Embedded Control Chip

S3C2410, a 32-bit RISC embedded microprocessor based on ARM920T core, is selected as the embedded chip of greenhouse environmental monitoring. The ARM920T core consists of ARM9TDMI, memory management unit (MMU) and cache. The core board integrates Samsung's S3C2410 processor (32-bit ARM920T core), 16M FALSH and 64M SDRAM, JTAG2ICE debugging interface, etc. The resources of the core board include: 2 UARTS, 1 USB port, data cable and LCD interface. The backplane includes: power circuit, various interface circuits, touch screen, keypad, USB storage circuit, etc. The structure diagram of embedded control chip is shown in Fig. 3.

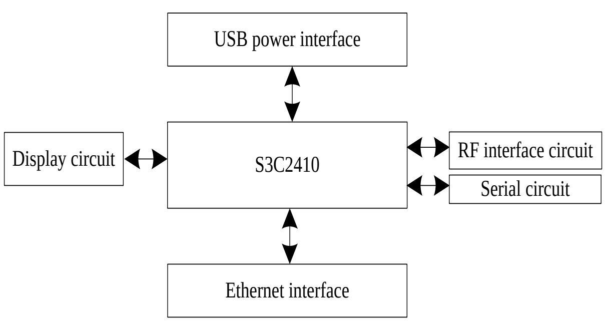


Fig. 3. Embedded control chip hardware structure

The power supply circuit provides the required power supply for each part of the greenhouse monitoring controller [17], and the selected chips are LT1764 and MIC29302. Among them, LT1764 is divided into adjustable and non-adjustable, providing 3.3V and 5.0V voltage respectively; MIC29302 provides a voltage of 4.2V. The voltage of 5.0V is mainly used for LCD circuit and USB circuit, the voltage of 4.2V is used for GPRS circuit, and the other circuits use the voltage of 3.3V. The power supply of the core board is 3.3V, which also comes from the backplane, while the S3C2410 core voltage and I / O voltage require 1.8V and 3.3V respectively. Therefore, AIC1117A is used on the core board to reduce the 3.3V voltage provided by the backplane to 1.8V and provide it to the processor core.

The function of serial interface is to act as an encoder converter between CPU and serial equipment. There are three kinds of asynchronous serial receiving devices provided by the S3C2410 general-purpose serial port and the asynchronous serial port. Each serial port contains a baud rate generator, a transmitter, a receiver and a control unit. The greenhouse controller uses the two serial port resources provided by the core board. According to the design requirements, the two serial ports are set as the standard RS2232 serial port. In addition to the function of ordinary serial port, serial port 1 also undertakes the task of wireless sensor network communication and acts as the debugging terminal of embedded control chip. Since the core board provides TTL level, MAX3232 chip is used to convert the voltage into a voltage conforming to RS2232 standard.

D. Greenhouse Environmental Monitoring Sensor

As a monitoring device to help people understand the external environment, sensors can detect and perceive the surrounding environmental information, and express various abstract environmental factors in the environment in the form of electrical signals according to certain rules. At the same time, as a machine, there is no subjective feeling, and the perception of the external environment can be expressed more accurately. Therefore, sensors are widely used in current industrial production. In the process of sensor selection, many factors need to be considered, including measurement range, working temperature, measurement accuracy, cost and matching degree with the chip. At the same time, combined with the particularity of greenhouse environment, the appropriate device can be selected. The corresponding sensor is selected according to the environmental parameters to be collected for greenhouse environmental monitoring, and the sensor is selected considering the cost, sensor accuracy and matching degree with the main control module.

1) Temperature and humidity sensor: Among AM2302 air temperature and humidity sensor is selected as the sensor for temperature and humidity collection in greenhouse environmental monitoring. It is characterized by not only capacitive humidity sensing elements, but also high-precision temperature sensing elements. Each element is calibrated with specific temperature and humidity, and can be used only after passing the inspection. AM2302 air temperature and humidity sensor has an 8-bit single chip microcomputer, so the response speed is relatively fast and the anti-interference is also very good. Because of its small size, it has low energy consumption. In addition, it has a series of advantages of single bus interface, so it can be widely used in meteorological stations and other places where temperature and humidity monitoring is needed. Due to the integrated packaging, the signal strength and anti-interference ability of the air temperature and humidity sensor are greatly enhanced [18]. The chip provides a two-wire serial interface similar to I2C bus, which is connected with the input interface of the host through bidirectional serial data line data and serial clock input SCK.

It not only inherits the characteristics of simple use of I2C interface, but also supports CRC verification of data transmission results to ensure the reliability of data transmission.

The air temperature and humidity sensor use single bus to communicate with MCU, and only one I / O port is needed to realize data transmission. The temperature and humidity data inside the sensor is transmitted to the single chip microcomputer at one time through the pin. The data is composed of 40bit, and the data is divided into decimal part and integer part. The transmission format is: 8bit humidity integer data + 8bit humidity decimal data + 8bit temperature integer data + 8bit temperature decimal data + 8bit verification data. The check data is the last eight bits of the sum of the first four bytes. The data output by the air temperature and humidity sensor is eight-bit binary data, which represents the real temperature and humidity.

When the bus is idle, it is high level. When the embedded control chip sends the start signal to AM2302, it pulls down the bus for at least 18ms to ensure that the AM2302 chip can detect the start signal. After detecting the start signal, AM2302 will convert from low-speed mode to high-speed mode. Then, the bus will be pulled up by the pull-up resistance for 20-40us, to read the response signal of AM2302, wait for it to pull down the data line and keep it for 80us, and then pull it up for 80us, finally start data transmission. The data of each bit starts with a 50us low-level slot. Whether the data bit is “0” or “1” is determined by the length of the high-level. If the high-level is 26-28us, the data “0” is sent, and if the high-level is 70us, the data “1” is sent. After all data is transmitted, AM2302 pulls down the bus by 50us, and then pulls up the bus through the pull-up resistor to enter the idle state.

2) CO2 Concentration Sensor: MH-Z14CO2 concentration sensor is used as the concentration acquisition sensor for greenhouse environmental monitoring. MH-Z14CO2 concentration sensor integrates the A / D conversion module and comparator. The conductivity of the module can be enhanced with the increase of gas concentration. The A / D conversion circuit in the module can convert this change into voltage change. The single chip microcomputer can directly read the voltage value or digital switching value. The module directly converts the gas concentration into the corresponding voltage value. The single chip microcomputer collects and reads the voltage through A/ D, and then obtains the gas concentration value in proportion. The detected concentration is 100-10000ppm. MH-Z14 sensor has a long service life, which reduces the cost to a certain extent. At the same time, it also has the performance of temperature compensation to ensure the measurement accuracy. Based on the above advantages, MH-Z14 sensor is widely used in the production environmental monitoring of industrial and agricultural equipment.

3) Light intensity sensor: The regulation of light in the greenhouse environment is mainly to increase the light intensity in the warm room. The demand for light intensity varies greatly among different crops. Generally, when the light intensity in the warm room decreases to 1000 lux, the crops in the warm room should be supplemented with light. According to the requirements of greenhouse crops for light intensity [19], the illuminance sensor BH1750fvi produced by Rohm company is selected. The high resolution of BH1750 can detect a wide range of light intensity changes (1lx-65535lx), and the visible light in 400nm-700nm band has excellent sensitivity. Logic+I2C Interface is the interface between light intensity calculation and I2C bus. PD is a photodiode, and its sensitivity is similar to that of human eyes; AMP integrated operational amplifier can convert PD current into PD voltage; ADC analog-to-digital conversion obtains 16-bit digital data. The ADDR port of register accumulator of BHT1750 illuminance sensor is grounded, and the two serial ports of clock end SCL and data end SDA are connected with CC2530. The working principle of the light intensity sensor is to convert the light intensity value into voltage value. Firstly, it initializes the program, then configures the input mode, collects data and carries out A/ D conversion. After the conversion, the single chip microcomputer reads the conversion directly.

E. Sensor Data Fusion Method

The temperature in the greenhouse is described by multiple measurement points. The temperature in the greenhouse changes with space. At the same time, the temperatures measured in different places in the greenhouse are also different. In order to control the temperature, humidity and other factors in the greenhouse, we must give an equivalent value to describe the temperature, humidity and other states of the current greenhouse consistently. Therefore, we use data fusion method to determine the effective value, which is used to control the greenhouse, such as rolling shutter, heating equipment, ventilation equipment, etc. Data fusion makes full use of multiple sensor resources. Through the reasonable control and use of these sensors and their observation information, the redundant or complementary information of multiple sensors in space or time is combined according to some criteria, so as to obtain better performance than the subset of its components.

After using wireless sensor network to collect greenhouse environmental monitoring data, we should first eliminate the missing data. There are many methods to eliminate the missing data, and select the distribution map method to eliminate the missing data.

The main parameters used in the distribution map are: median X_M, upper quartile F_u lower quartile F_l, quartile dispersion dF and elimination point ρ. Their definitions and usage are as follows:

Firstly, it can carry out multiple independent measurements on a tested object, and arrange the measured data in the order from small to large:

X_1. X_2,..., X_N, then X_1 is called the lower limit and X_N is called the upper limit.

Secondly, the median value X_M is defined:

$$X_M = \frac{(N+1)X}{2} \quad (1)$$

$$X_M = \frac{\left(\frac{N+1}{2} + \frac{N}{2}\right)X}{2} \quad (2)$$

Thirdly, in the partition map method, if the median of interval $[X_M, X_N]$ is F_u, F_u is the upper quartile, and the lower quartile F_I is the median of interval $[X_1, X_M]$

Fourthly, the dispersion dF of the quartile is defined as:

$$dF = F_u - F_I \quad (3)$$

It is stipulated that the data whose distance from the median is greater than βdF is divorce data, that is, the judgment interval of invalid data is:

$$X_i - XM > \beta dF \quad (4)$$

Where, β is a constant, and its size is determined according to the accuracy requirements of greenhouse monitoring. The elimination point is defined as:

$$\rho_1 = \frac{(F_I - \beta)dF}{2} \quad (5)$$

$$\rho_2 = \frac{(F_u - \beta)dF}{2} \quad (6)$$

The measurement data outside the interval $[\rho_1, \rho_2]$ is considered as the elimination point.

After the above steps, the remaining data are consistency measurement data. We can fit the consistency data. The weighted average method is selected to fuse the data collected by the sensor. Since many sensors are placed in the greenhouse, the greenhouse data collected by each sensor is different [20], so it is necessary to fuse the data collected by wireless sensors to give users an effective value to guide users to operate the greenhouse. The problem is described as follows: there are n temperature wireless sensors in the greenhouse, and the greenhouse data collected by each wireless sensor at t time is t_1, t_2, $\cdots$, t_n. Firstly, the distribution diagram method is used to eliminate the missing data, and the data are sorted to obtain the consistent data columns t_e, $\cdots$, t2 t_p. Weight is represented by w_i. w_i is usually selected as $\frac{1}{p-e}$, and the average value of the temperature in the greenhouse is calculated by the weighted average method:

$$\bar{T} = \sum_{i=e}^{p} \frac{t_i}{p-e} \quad (7)$$

The weighted average method is simple, easy to implement and occupies less resources. When the wireless sensor network is used to monitor the embedded greenhouse environment, the weighted average method is used to fuse the data collected by the sensor to obtain the final greenhouse monitoring parameters such as ambient temperature and ambient humidity.

III. Results

In order to verify the effectiveness of the embedded environmental monitoring method of greenhouse based on wireless sensor network, the method in this paper is applied to the greenhouse environmental monitoring of an agricultural production. The average communication distance between the deployed nodes of the wireless sensor network is about 20m, and the nearest distance between the monitoring node and the gateway node is about 100m. After installation and operation, after the gateway is started, the average time required for node binding and ad hoc network establishment is less than 1min. The sampling frequency setting scheme of sensor nodes is: air temperature and humidity for 2min, soil temperature and humidity for 10min, light intensity for 3min and CO2 concentration for 30min. After completing data collection and transmission, each node will automatically enter the sleep state until the next sampling cycle wakes up.

In order to verify whether the communication stability between nodes of wireless sensor network is good and whether the communication quality is qualified, the packet loss rate of wireless sensor network is tested. Packet loss rate refers to the proportion of data packets not received in a certain period of time in the total data packets sent. If the packet loss rate is too high, it indicates that a large number of temperature and humidity, smoke concentration and light intensity data have not been collected, which will lead to the control node cannot accurately turn on and off the relevant equipment, cannot accurately monitor the environment in the shed, resulting in measurement failure. To ensure the reasonable distance between network nodes, it needs to try to ensure a low packet loss rate and enhance the accuracy of measurement. The greenhouse in the study area is selected as the test site. The coordinator is placed in the monitoring room, adjacent to the upper computer, and the terminal equipment is placed in different positions in the greenhouse. The transmit power of CC2530 processor is -3dBm; the total number of single packet test is set to 2000, and the packet sending speed is one group per second, with 5 packets per group. This speed fully meets the needs of the actual project. After fixing the coordinator position, it can send the same number of data packets at the terminal nodes in different positions of the greenhouse, to measure the number of received data packets, and calculate the packet loss rate. The calculation of packet loss rate is shown in the following formula:

$$A = \frac{B - C}{B} \times 100\% \quad (8)$$

In formula (8), B and C represent the transmitted data packet and received data packet of wireless sensor network respectively.

Using the method in this paper to collect greenhouse environment data, the communication performance of wireless sensor network is shown in Table I.

TABLE I. WIRELESS SENSOR NETWORK COMMUNICATION PERFORMANCE

Sensor terminal serial number	Distance from gateway node	number of packets sent	Number of received packets
1	5	2000	1999
2	10	2000	1999
3	15	2000	1999
4	20	2000	1997
5	25	2000	1995
6	30	2000	1991
7	35	2000	1987
8	40	2000	1985
9	45	2000	1981
10	50	2000	1976
11	55	2000	1961
12	60	2000	1958

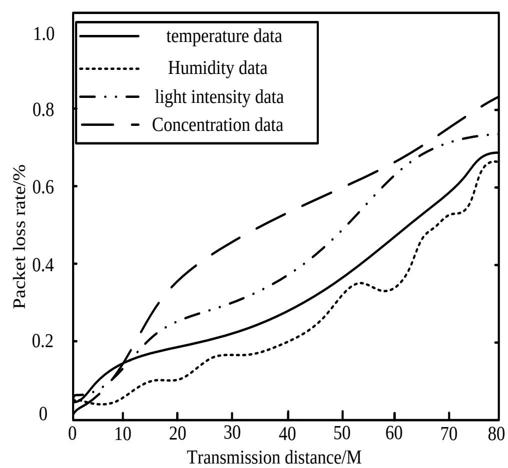


Fig. 4. Statistics of packet loss rate

The data of greenhouse environment monitored by wireless sensor network used in this paper are statistically analyzed and the packet loss rate results of collecting different types of greenhouse environmental data are shown in Fig. 4.

According to the experimental results in Fig. 4, the wireless sensor network is greatly affected by the transmission distance. Within the short range of 10 meters, the data packet loss rate maintains a low value within 0.1%. However, when the distance between the sensor terminal and the coordinator is more than 10 meters, the packet loss rate increases sharply, which is caused by the great irregularity of the communication link of the low-power module. Through the experimental test, even when the packet loss rate of 80m is the largest, the packet loss rate of ZigBee wireless communication is still less than 1%, which meets the needs of greenhouse environmental monitoring.

In the greenhouse monitoring area, each monitoring point is equipped with three sensors. The method in this paper is used to fuse the greenhouse environmental data collected by the three sensors to improve the accuracy of greenhouse environmental data collection. The temperature data of different monitoring points in the greenhouse are collected by this method, and the results are shown in Fig. 5.

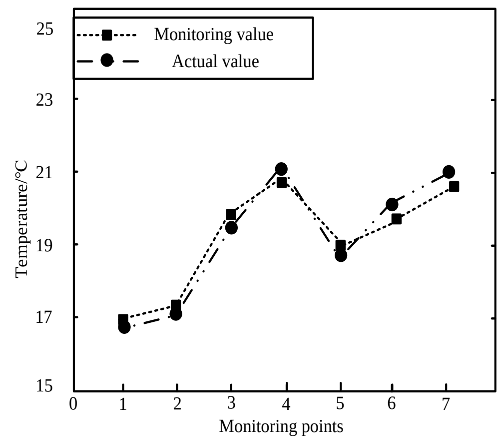


Fig. 5. Temperature data collection results

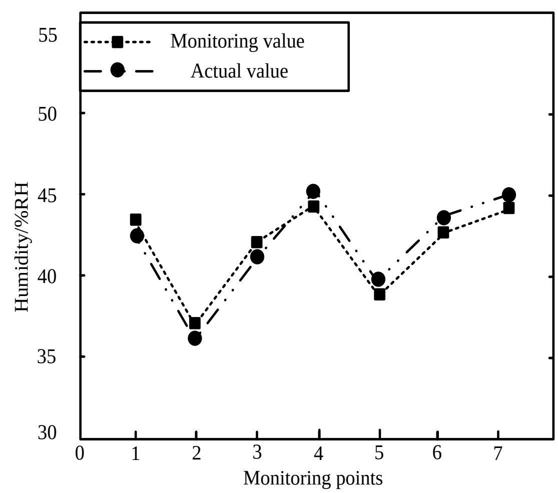


Fig. 6. Collection results of ambient humidity

The humidity data of different monitoring points in the greenhouse are collected by the method in this paper, and the results are shown in Fig. 6.

The experimental results in Fig. 5 and Fig. 6 show that the method of this paper is used to monitor the greenhouse environment, and the collected temperature and humidity data of the greenhouse environment is close to the actual environmental temperature and humidity data of the greenhouse within a reasonable range of error, which indicates that the data collected by the temperature and humidity sensor is accurate and reliable, and the sensor used in this method can effectively monitor the temperature and humidity data of the greenhouse environment. The error is within a reasonable range, which indicates that the data collected by the temperature and humidity sensor is accurate and reliable.

During the experiment, a full spectrum LED lamp is set in the greenhouse to gradually shorten the distance between the full spectrum LED lamp and the lighting intensity measuring equipment, and increase the lighting intensity. The average value of five groups of data is taken for each measurement and compared with the data measured by the digital illuminance meter. The comparison of light intensity data is shown in Fig. 7.

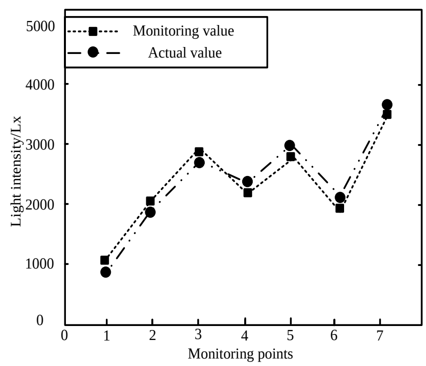


Fig. 7. Comparison of light intensity data

According to the experimental results in Fig. 7, the difference between the greenhouse light intensity data monitored by the method in this paper and the actual light intensity data of greenhouse is small. From the above data, the designed light intensity acquisition circuit is very close to the actual light intensity value, and the error is within a reasonable range. The data collected by the light sensor is accurate and reliable.

The CO_2 concentration data collected at different monitoring points by the method in this paper are counted, and the collected CO_2 concentration data are compared with the actual CO_2 concentration data. The comparison results are shown in Fig. 8.

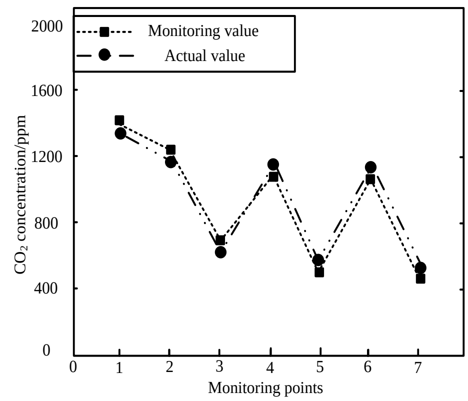


Fig. 8. Comparison of CO2 concentration data

The experimental results in Fig. 8 verify that the method in this paper can effectively collect CO_2 concentration data, and the collected CO2 concentration data is very close to the actual CO2 concentration data of greenhouse environment. The experimental results in Fig. 8 verify that the method in this paper can effectively collect the CO_2 concentration data of greenhouse environment by using wireless sensor network. At the same time, the collected CO_2 concentration data is in good agreement with the actual CO_2 concentration data, indicating that the method in this paper has high accuracy in collecting CO_2 concentration data.

The relative errors between the temperature data, humidity data, light intensity data and CO_2 concentration data collected by the method in this paper and the actual data are counted. The statistical results are shown in Table II.

TABLE II. COMPARISON OF RELATIVE ERRORS OF ENVIRONMENTAL MONITORING

Monitoring points	Temperature/%	Humidity/%	Light intensity/%	CO_2 concentration/%
1	0.25	0.34	0.29	0.74
2	0.34	0.47	0.43	0.56
3	0.28	0.29	0.28	0.42
4	0.13	0.36	0.16	0.61
5	0.25	0.48	0.34	0.53
6	0.34	0.35	0.28	0.48
7	0.29	0.48	0.17	0.34
8	0.34	0.29	0.26	0.52
9	0.19	0.35	0.23	0.64
10	0.52	0.41	0.45	0.51

From the experimental results in Table II, it can be seen that the method in this paper uses wireless sensor network to collect greenhouse environment data, and the relative error between the collected data and the actual data is less than 1%.

The experimental results show that the method can achieve effective monitoring of the greenhouse environment with high monitoring accuracy. According to the collected greenhouse environment data of each monitoring point, the greenhouse environment can meet the growth needs of crops. Through experimental verification, the research method can receive and display environmental data such as temperature and humidity, light intensity, and CO_2 concentration collected by sensor nodes in real time, and can view the real-time operation status of each node. During the test period, the sensors operated stably and all functions worked normally, and played an important role in the daily management of the greenhouse environment, achieving the expected results.

IV. Discussion

Existing Research：wireless sensor networks have brought a revolution in information sensing due to their low power consumption, low cost, distributed and self-organizing features. Wireless sensor network consists of a large number of inexpensive miniature sensor nodes deployed in the monitoring area, which is a multi-hop self-organized network formed by wireless communication. ZigBee is a local area network protocol based on IEEE802.15.4 standard, which has the features of low cost, simplicity, short distance, automatic networking of nodes, low power consumption and low data transmission rate, and can be applied to various applications by embedded devices in automatic control field. Currently, many scholars have applied wireless sensors in greenhouse environmental monitoring. Benahmed K et al. proposed an algorithm based on wireless sensor network technology for monitoring microclimate in greenhouses and developed a linear equation model for optimizing plant production and material costs [21]. Given that subsurface soil sensing uses several types of sensors and devices that can help to monitor soil fertility in real time, Ghosh D et al. investigated the existing subsurface sensing technologies used to monitor soil fertility in agriculture using underground IoT[22]. Abbasikesbi R et al. proposed a wireless sensor network based on the proposed algorithm to improve tomato crops in greenhouses. The developed sensor nodes have low power consumption and low cost to monitor parameters such as temperature, humidity, CO, CO_2 and light intensity[23]. In all of the above studies, wireless sensors were used to monitor the greenhouse environment, but the accuracy of greenhouse environment monitoring was low because the sensors set up did not meet the needs of greenhouse environment monitoring.

Research results of this paper：Results of this study: Based on the previous research, this paper proposes an embedded greenhouse environment monitoring method based on ZigBee wireless sensor network. The ZigBee-based wireless sensor network has the features of large node size, small size, low cost and ad hoc network, which can not only realize the rapid automatic networking of sensor nodes, but also has the ability to collect, transmit, store and process various greenhouse environmental data in real time. The ZigBee-based wireless sensor network has the advantages of low cost, low power consumption, no wiring, flexible networking and friendly human-machine interface, which makes a useful exploration for the application of wireless sensor network and embedded technology in the field of greenhouse management. This research collects greenhouse environment information through temperature sensors, humidity sensors and other sensors, and transmits the collected information to remote monitoring center through ZigBee wireless sensor network, which is convenient for greenhouse managers to remotely monitor greenhouse environment in real time and provide a good environment for crop growth.

Research limitations: Limitations of this research: Although the embedded greenhouse environment monitoring method based on ZigBee wireless sensor network used in this research can overcome some defects of traditional greenhouses in structure and performance and help to establish large-scale intelligent greenhouse clusters, there are still some shortcomings. In the actual greenhouse environment monitoring, the monitoring error is difficult to be completely eliminated because of the influence of many factors. Besides, the interaction between different influencing factors should be considered, and in addition to the monitoring of individual factors, joint analysis should be conducted in subsequent studies.

Future work: The Future of Work: Today's Internet applications are penetrating into the embedded space. 5G technology is enabling the rapid spread of smartphones. Due to the rapid development of embedded technology, web servers can run on single-chip microcomputers with high-capacity flash memory. Embedded technology will develop rapidly as various electrical devices become intelligent and Internet-enabled. With the application of these intelligent technologies in the field of environmental monitoring, the level of environmental monitoring will be greatly improved. Today, with the development of embedded intelligent technology and "5G technology", the greenhouse monitoring based on wireless sensor and network technology network has a better technical design basis than the traditional greenhouse monitoring method. The future direction of sensor technology research will be supported by 5G technology to get more development.

V. Conclusion

Through greenhouse environment monitoring, the indoor environment can be automatically monitored to provide suitable greenhouse environment for crops according to the requirements of different crops, thus meeting the needs of scientific research and production and accumulating rich data for intelligent agriculture experts to develop new technologies. The embedded greenhouse environment monitoring method based on wireless sensor network overcomes some defects of traditional greenhouses in structure and performance, and also has some significance for establishing large-scale intelligent greenhouse clusters. Through this research method, the parameters of temperature, humidity and light intensity in different areas of the greenhouse are obtained in real time, so that the crops can achieve a suitable growth environment and provide a convenient, efficient, adaptable, simple and practical monitoring scheme for greenhouse managers. The experiments show that the method is effective in monitoring temperature, humidity and other parameters and has high applicability.

ACKNOWLEDGMENT

The research is supported by: Project approved by Inner Mongolia Education Department: Research on discipline and specialty Integration construction of local universities under the background of "Double First-class" construction No. NJSY22136.

REFERENCES

[1] J. Wang, M. Chen, J. Zhou, and P. Li, "Data communication mechanism for greenhouse environmental monitoring and control: an agent-based iota system", *Information Processing in Agriculture*, vol. 7, no. 3, pp. 12, 2020.

[2] Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, and S. Bhansali, "Review-machine learning techniques in wireless sensor network-based precision agriculture", *Journal of The Electrochemical Society*, vol. 167, no. 3, pp. 037522, 2020.

[3] J. He, J. Lee, T. Song, H. Li, and K. Wang, "Delay-tolerant indoor optical wireless communication systems based on attention-augmented recurrent neural network", *Journal of Lightwave Technology*, pp. 1-7, 2020.

[4] C. Zhao, B. Gao, L. Wang, W Huang, and S Cui, "Spatial patterns of net greenhouse gas balance and intensity in Chinese orchard system", *Science of The Total Environment*, vol. 779, no. 6, pp. 146250, 2021.

[5] B. Hua, Z. Li, W. Gao, H. Feng, and G. Liu, "Soil amendment in plastic greenhouse using modified biochar: soil bacterial diversity responses and microbial biomass carbon and nitrogen", *Biotechnology Letters*, 43(FEB.1PT.A), pp. 1-12, 2021.

[6] M. L. Gu, and M. Li, "Optimization Design of Wireless Energy Transmission System Based on Improved Transmission Coil", *Computer Simulation*, vol. 37, no. 1, pp. 284-289, 2020.

[7] R. Anzum, and J. Naeem, "Leveraging lorawan technology for smart agricultural monitoring of malaysian palm oil plantation", *IOP Conference Series: Earth and Environmental Science*, vol. 756, no. 1, pp. 012052, 2021.

[8] L. Cheng, M. Xue, Y. Wang, Y. Wang, & Y. Bi, "A robust tracking algorithm based on modified generalized probability data as-sociation for wireless sensor network", *IEEE Transactions on Industrial Electronics*, vol. 111, no. 2, pp. 1-7, 2021.

[9] E. B. Priyanka, T. K. Devi, P. Sakthivel, and A. S. Sagayaraj, "Gate diffusion input (GDI) codes involved viterbi decoders in wireless sensor network for enhancing qos service", *Analog Integrated Circuits and Signal Processing*, vol. 111, no. 2, pp. 287-298, 2022.

[10] K. Ravindra, T. Singh, and S. Mor, "Covid-19 pandemic and sudden rise in crop residue burning in india: issues and prospects for sustainable crop residue management", *Environmental Science and Pollution Research*, vol. 29, no. 2, pp. 3155-3161, 2021.

[11] P. F. Han, N. Zeng, B. Yao, W. J. Zhou, L. Q. Chen, & Q. Shao, "Preface to special topic on atmospheric greenhouse gas measurement and application in China", *Advances in Atmospheric Sciences*, vol. 37, no. 6, pp. 3-4, 2020.

[12] N. Sarma, P. M. Tuohy, & S. Djurovic, "Stator electrical fault detection in dfigs using wide-band analysis of the embedded signals from the controllers", *IEEE Transactions on Energy Conversion*, pp. 1-6, 2020

[13] T. Pisanu, S. Garau, P. Ortu, L. Schirru, and C. Macciò, "Prototype of a low-cost electronic platform for real time greenhouse environmental monitoring: an agriculture 4.0 perspective", *Electronics*, vol. 9, no. 5, pp. 726, 2020.

[14] F. A. Khan, A. A. Ibrahim, & A. M. Zeki, "Environmental monitoring and disease detection of plants in smart greenhouse using internet of things", *Journal of Physics Communications*, vol. 4, no. 5, pp. 055008, 2020.

[15] D. Liu, J. Cui, J. Zhang, C. Y. Yang, and L. Hanzo, "Deep reinforcement learning aided routing in aeronautical ad hoc networks", *IEEE Transactions on Vehicular Technology*, PP 1-8. 2021.

[16] X. Sun, and S. Cao, "A routing and wavelength assignment algorithm based on two types of leo constellations in optical satellite networks", *Journal of Lightwave Technology*, pp. 1-8, 2020.

[17] M. G. Chen, R. Zheng, Y. Fu, J. Qi, and W. Yao, "An analog control strategy with multiplier-less power calculation circuit for flyback micro-inverter", *IEEE Transactions on Power Electronics*, pp. 1-7, 2021.

[18] Y. Zhao, Z. R. Zhang, S. C. Yan, & R. J. Tong, "High-sensitivity temperature sensor based on reflective solc-like filter with cascaded polarization maintaining fibers", *IEEE Transactions on Instrumentation and Measurement*, vol. 9, no. 9, pp. 1-7, 2021.

[19] Q. Shang, C. Cheng, J. Wang, K. Luo, & X. Yang, "Net global warming potential, greenhouse gas intensity and carbon footprint as affected by different tillage systems from Chinese double-cropping paddy fields", *Soil and Tillage Research*, vol. 288, no. 5, pp. 104947, 2021

[20] H. H. Lee, S. U. Kim, H. R. Han, D. Y. Hur, & C. O. Hong, "Mitigation of global warming potential and greenhouse gas intensity in arable soil with green manure as source of nitrogen", *Environmental Pollution*, vol. 288, no. 2, pp. 117724, 2021.

[21] Benahmed K, Touhami A, Parra L, et al. An intelligent monitoring of greenhouse using wireless sensor networks. *SMART STRUCTURES AND SYSTEMS*, vol. 26, no. 1, pp. 117-134, 2020.

[22] Ghosh D, Anand A, Gautam S S, et al. Soil Fertility Monitoring With Internet of Underground Things: A Survey. *IEEE Micro*, vol. 42, no. 1, pp. 8-16, 2022.

[23] Abbasikesbi R, Nikfarjam A, Nemati M. Developed wireless sensor network to supervise the essential parameters in greenhouses for internet of things applications. *IET Circuits Devices & Systems*, vol. 8, no. 14, pp. 1258-1264, 2020.

Research on the Application of Virtual Technology-based Posture Detection Device in Swimming Teaching

Hongming Guo [1], Jingang Fan[2, 3*]
College of Physical Education, Putian University, Putian, 351100, China[1]
Department of Physical Education, Sichuan International Studies University, Chongqing, 400031, China[2]
Graduate School, JoséRizal University, Manila, 1552, Philippines[3]

***Abstract*—With the socio-economic development, the national demand for playing leisure sports has increased, and swimming is one of the popular choices. To help swimming beginners understand the correct swimming posture more quickly and directly, hybrid neural network algorithms based on sliding window detection and deep residual networks are designed in this study, and two corresponding virtual image classification models of swimmer's posture are designed based on these algorithms. In order to reduce the noise of the input data and reduce the cost of data collection, the virtual reality technology is used to convert the swimmer's swimming pose image into the image model in the virtual reality space as the input data of the algorithm. The performance test experimental results show that the classification accuracy of the swimmer pose recognition models based on PTP-CNN algorithm and SW-CNN algorithm designed in this research are 97.48% and 96.72% respectively on the test set, which are much higher than other comparison models, and the model built based on PTP-CNN algorithm has the fastest computation speed. The results of this research can be applied to assist participants in swimming pose recognition in teaching beginner swimmers.**

Keywords—Virtual reality; pose recognition; swimming; neural network; sliding window

I. Introduction

Swimming has both leisure and physical exercise attributes, making it a better form of exercise for adults with a certain financial base [1]. In fact, the middle class population in China has grown significantly in the last decade, which has directly led to an increase in the number of people learning to swim in their spare time. However, it is difficult for beginners to distinguish between accurate and incorrect common swimming positions, but the ability to accurately judge and master scientific swimming positions is important to learn to swim and reduce the risk of personal safety during exercise [2]. In the past, beginners often need to spend a lot of time and energy on learning to identify and practice various correct swimming postures, which is inefficient, and may even make learners lose interest in swimming [3]. At the same time, the emergence of high and new technologies, including virtual reality and artificial intelligence, provides new ideas for automating traditional physical education teaching tasks. These technologies can free teachers from repetitive low-end teaching tasks, and the teaching efficiency and error rate are far lower than those of the teaching system based on traditional machine learning algorithms [4]. Aiming at this problem, this research tries to use virtual reality technology and artificial intelligence technology to design a model that can intelligently judge whether the swimmer's posture in the swimming image is correct, so as to provide timely and accurate feedback for beginners to learn the project. It is assumed that it may play a role in shortening the learning cycle of swimming learners and reducing the learning difficulty, this is also the contribution and value of this paper to the field of swimming teaching.

*Corresponding Author.

II. Related Work

A large number of related studies have been conducted by experts in different fields about virtual reality technology, bit-posture detection technology, and smart sports teaching involved in this research. Elbamby M S et al. discussed the challenges faced in achieving ultra-reliable and low-latency virtual reality by addressing the shortcomings of virtual reality technology in terms of high throughput, low latency, and reliable communication, thus finding that the use of millimeter wave communication, edge computing, and Intelligent networks with active caching techniques can solve the above problems of virtual reality technology applications to a certain extent [5]. Chen M et al. studied the resource management problem of networks with wireless virtual reality users communicating over small cell networks and proposed a virtual reality model based on multi-attribute utility theory, and simulation test results showed that the model has faster convergence than the traditional model and provides lower latency of virtual reality services [6]. Thies J et al. proposed an image processing algorithm based on virtual reality technology to build a more realistic virtual meeting environment in video teleconferencing [7]. Du et al. found that virtual reality technology has been gradually and widely used in the architecture, engineering, construction, and facilities management industries in recent years, because it can improve the workflow of teams by enhancing the consensus of team members. However, the current virtual reality application suffers from the difficulty of generating a virtual reality dialogue environment from input data. Therefore, the research team designed a synchronous communication system based on virtual reality technology, and the test results showed that the system has higher utilization value and communication efficiency than traditional virtual reality communication systems [8].

Cga B et al. found that obtaining accurate information about students' posture during learning is important for assessing students' learning status and improving teaching methods, but the presence of various occlusion elements in the educational environment makes it more difficult to carry out current student posture detection. So this study proposes a new posture detection method based on a single-stage object detector, which uses an adaptive fusion mechanism to learn complementary. This method uses an adaptive fusion mechanism to learn complementary spatial features to make the feature extractor more discriminative. Experimental results show that the student pose detection capability of the method is significantly better than other single-stage target detection methods on a real classroom pose dataset [9]. Lin G et al. found that fruit detection is necessary for automatic guava harvesting under real outdoor conditions, and that the branch-related pose of the fruit is also essential to guide the robot to approach and separate the target fruit without colliding with its parent branch. In order to perform automatic, collision-free harvesting, this research team designed a fruit detection and pose estimation method based on sensors with four channels of red, green, blue, and depth data images from cell phones. Quantitative experimental results show that the method has an accuracy and recall of 0.983 and 0.948 for guava fruit detection, respectively, which can be applied to improve the control of automatic guava harvesting machinery [10]. Tang L's research team found that it is difficult to detect the bit pose of small targets under poor imaging conditions such as severe occlusion and low resolution, and proposed a bit pose detection algorithm that combines merged region of interest pools and local static. Therefore, a pose detection algorithm combining merged region of interest pool and local static learning is proposed. Experimental results show that the method outperforms existing methods for pose recognition [11]. Zhou L et al. found that existing fabrication methods cannot produce flexible sensors that match the shape of soft robots. Therefore, a new 3D printed multi-functional inductive flexible stretchable liquid metal sensor was designed using the posture detection algorithm, and the test results showed that this sensor has a higher perception of deformation than the conventional sensors, which also means that the posture detection algorithm improves the manufacturing accuracy of the sensor components [12].

In summary, it can be found that virtual reality and posture detection techniques have been increasingly applied to industrial and educational fields, and some economically valuable and educational results have been achieved. However, studies combining virtual reality and posture detection technology to build a swimming stance recognition model with high recognition accuracy are still rare, so it is really necessary to use virtual reality technology to reduce the noise of training images and improve the data scaling effect when building this teaching aid model, while the posture detection work can help swimming beginners to learn swimming posture better, which is the reason for conducting this study.

III. Design of Swimming Position Detection Algorithm for Swimming Scene

A. Design of Bit Pose Detection Algorithm Incorporating Sliding Window Detection and Convolutional Neural Network

The dataset used for training and testing the pose detection algorithm in this study is a virtual model image of a swimming scene and its swimmer's swimming pose constructed by virtual reality technology. Before designing the pose detection algorithm, it is necessary to design the pre-processing process of the dataset [13-15]. The preprocessing of virtual image of swimmer's pose mainly includes image cropping, grasping rectangle frame extraction, image normalization and homogenization [16]. In order to reduce the content of irrelevant information in the image and the computational effort of the algorithm, the virtual model image needs to be cropped to a uniform size (640 480 size is chosen here) [17]. The purpose of homogenization and normalization of the virtual model images is to eliminate the negative impact of different inter-feature scales in the image data on the training results. The range of RGB (Red Green Blue) channel values in the image data is [0,255], and the normalization operation can be completed according to equation (1) to compress the range of values to [-1,1].

$$I_{i,j,k}^{(n)'} = \frac{2I_{i,j,k}^{(n)} - 255}{255} \tag{1}$$

$I_{i,j,k}^{(n)}$ represents the pixel values of row, column and channel numbers i , j and k positions in the n image before normalization. $I_{i,j,k}^{(n)'}$ represents the pixel values of row, column and channel numbers i , j and k positions in the n image after normalization. When constructing a virtual model of the swimmer using virtual technology, the original physical image is a depth map, and the depth value is used to describe the distance between the camera and the photographed object. But the index is mainly distributed in the middle part of the image, which needs to be mapped and processed according to the camera desktop distance and the range distance, see equation (2).

$$d = \begin{cases} 1.4m, d < 0.8m \\ 0.8m, d > 1.4m \\ d, d \in [0.8m, 1.4m] \end{cases} \tag{2}$$

d is the depth value after the mapping process, and m is the ratio of the camera range distance to the camera desktop distance. Then use equations (3) and (4) to normalize and homogenize d .

$$d_i = \frac{d_i - d_{\min}}{d_{\max} - d_{\min}} \tag{3}$$

$$d_i = d_i - \frac{\sum_{i=0}^{n} d_i}{n} \quad (4)$$

In equation (3), d_i is the depth value of the i row, $d_{\min}$ and $d_{\max}$ are the minimum and maximum values of the depth value of the i row, and n is the number of pixels of the i row. This time, the sliding window method is chosen to generate the rectangle to be selected. However, the number of selected rectangles generated by the image is too large, which makes the algorithm computationally inefficient, so it is necessary to filter out some of the rectangles. First of all, the background image can be removed using the background difference method, and then the image is divided into a grid of specified distances, and the detection frame containing the smallest object is set as the detection range of the sliding window, and the center point of the rectangular frame can only be in the intersection of the restricted range [18]. According to the size statistics of the dataset, it is need to set the pixel point taking steps for the rotation angle θ, length h, and width w of the grasping rectangle, and the range of values which are shown in equations (5), (6), and (7).

$$\theta = 0° : 15° : 180° \quad (5)$$

$$h = 10 : 10 : 90 \quad (6)$$

$$w = \min(10, h-50) : 10 : \max(90, h+50) \quad (7)$$

Equation (5) represents the interval θ in the range $0°$ to $180°$ taken by $15°$, and the meanings of equation (6) and equation (7) can be analogously derived in turn. Since convolutional neural network can generalize and extract features in images with high recognition accuracy, this study chooses it to build a swimming pose detection model based on the mature AlexNet, and the core hierarchy in AlexNet is designed below. In general, if there is an input image of size $n \times n \times n_c$ and its convolutional kernel size is $f \times f \times n_c$. n_c and f are the number of image channels and convolutional kernel size respectively, and the convolutional step parameter is s, then the size of the output value of the convolutional layer can be obtained by equation (8).

$$((n+2p-f)/s+1) \times ((n+2p-f)/s+1) \times m_c \quad (8)$$

p and m_c represent the size of the edge filling layer and the total number of convolutional kernels used, respectively. The activation functions of the convolutional layers are used to perform nonlinear mapping of the output elements of each layer. The commonly used activation functions are relu function, elu function, tanh function, and sigmoid function. However, for neural networks with more layers, using the tanh function and sigmoid function as the activation function may cause the gradient disappearance problem and reduce the overall convergence speed of the algorithm, so the relu function and elu function are more widely used. The function graphs of the above four activation functions are shown in Fig. 1.

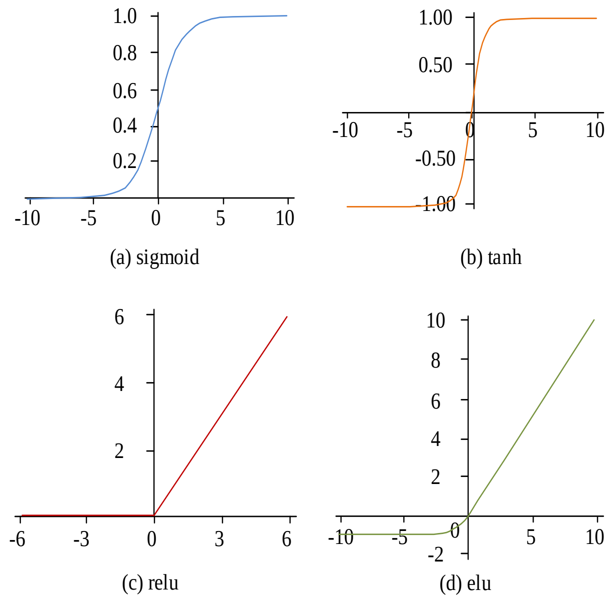


Fig. 1. Function curve of common activation functions

The role of pooling layer in AlexNet is to extract the high-latitude features in the images and reduce the size of the dataset. Depending on the processing method, there are two types of pooling layers, average pooling and maximum pooling. The maximum pooling method is chosen to construct the pooling layer in AlexNet in this study, and the calculation method of maximum pooling is shown in Fig. 2.

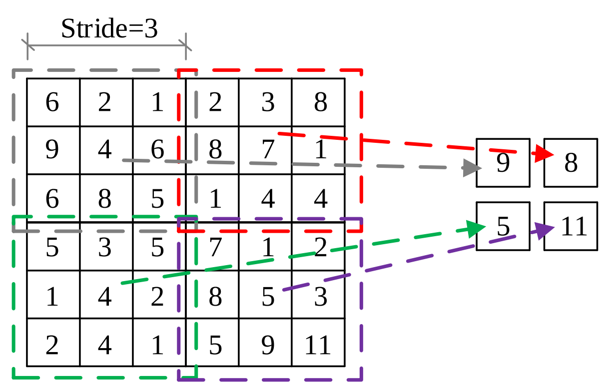


Fig. 2. Schematic diagram of the calculation rules for maximum pooling

According to the content above, based on the simplified AlexNet, a neural network is designed to judge whether the swimmer's swimming posture in the generated rectangular box is accurate. The input is the preprocessed RGD (Red Green Depth) three channel data, and the computational flow of the algorithm is shown in Fig. 3.

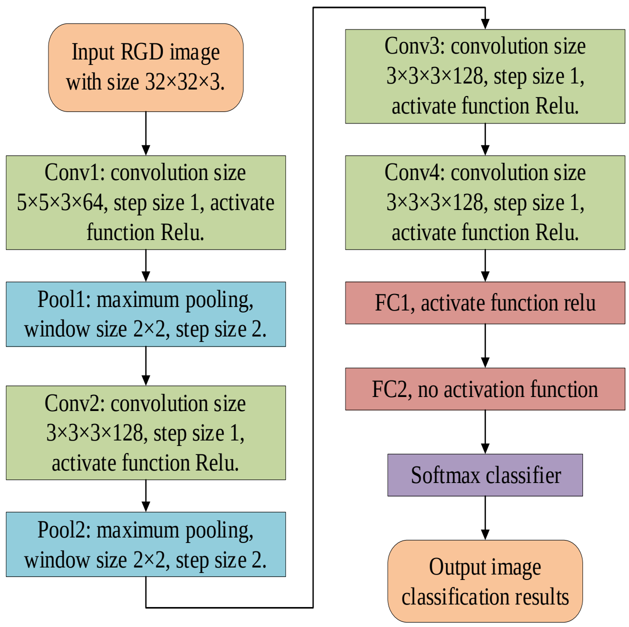


Fig. 3. Flow chart of the bit pose detection algorithm combining sliding window detection and convolutional neural network

As shown in Fig. 3, the algorithm contains four convolutional layers, two maximum pooling layers and two fully connected layers. The size of the convolutional kernel is 5×5, and the output data size is $16\times16\times64$ after the maximum pooling process, and then the output data size is $8\times8\times128$ after three convolutional layers with the convolutional kernel size of 3×3 and the maximum pooling layer. The size of the output data is adjusted to 2. Finally, the data is input to the softmax classifier for binary classification, and the judgment result of whether the object pose of the rectangular box in the input image data is correct is output.

B. Design of Swimming Pose Detection Algorithm based on Deep Residual Network

The pose detection algorithm designed in this study, which combines sliding window detection and convolutional neural network (SW-CNN), has many layers of networks and complex processing, so it may have a problem of slow detection speed in actual use, which can't meet the timeliness requirements of its application in swimming teaching. Therefore, in order to further improve the recognition accuracy of virtual swimming stance images, a virtual model recognition algorithm of swimmer's stance that integrates end-to-end idea and convolutional neural network (PTP-CNN) is proposed. Unlike the SW-CNN algorithm, the former trained neural network can directly reflect the RGB image and its corresponding standard swimming pose mapping relationship, so the input data of the PTP-CNN algorithm in the training phase is the image and the corresponding standard swimming pose information in the figure. In addition, since the PTP-CNN algorithm removes the sliding window detection link, and the learning ability of the neural network with the requirement of diversity of input image data is increased. Therefore, migration learning and data augmentation, deep residual network methods are needed to learn swimming pose features.

Due to the inefficiency of constructing virtual models based on physical images of swimmers' swimming stance and the complexity of computational processing, the data set that can be used for training the algorithm is often small in number and cannot well meet the training requirements of the algorithm, and overfitting may even occur. Therefore, before designing the PTP-CNN algorithm, it is necessary to expand the dataset and increase the size of the dataset appropriately through data augmentation techniques to make up for the lack of data and the problem that the data diversity and feature information are relatively single, so that the neural network can learn to master more important image features. In this study, considering the type of task of the algorithm, we choose to use data augmentation methods that do not change the features themselves but only the pixel distribution to process the dataset by flipping transformation, rotation transformation, translation transformation, contrast transformation, color transformation, sharpening, adding noise disturbance, adding blurred information, etc. The program selects one or more measures to process the original dataset in a random way.

Even with data augmentation, the total number of samples in the dataset is still relatively limited, which makes it difficult and inefficient to obtain more data. And using the existing dataset directly for training may lead to overfitting of the model. In addition, training the whole neural network algorithm from scratch will consume a lot of computer and time resources. Therefore, the PTP-CNN algorithm is built using the migration learning approach. There are three common applications of transfer learning, as a feature extractor for the algorithm to be trained, as a pre-trained model for the algorithm to be trained, and as an aid for fine-tuning the model parameters. The most central step in the algorithm training here is to improve the image feature extraction capability of the algorithm, therefore, the convolutional neural network trained with Imagenet dataset is selected as the image feature extractor in PTP-CNN algorithm. Then the fully connected layer is set after the convolutional neural network, and the output of the last layer is changed to the position and stance parameters that measure the swimmer's stance. That is, the parameters of the layers within the migrated convolutional neural network remain unchanged, and only the parameters of the fully connected layer behind are fine-tuned to achieve migrated learning.

When the neural network has too many levels, its recognition accuracy will be degraded, and some scholars have proposed the use of central normalization and normalized initialization methods, which can solve the problem to some extent, but the effectiveness of these methods is also poor if the neural network has too many levels, and the residual neural network is more effective in dealing with the problem. In this study, a residual network module constructed by constant mapping is introduced into the ResNet network of PTP-CNN algorithm to solve the degradation problem of deep convolutional neural networks.

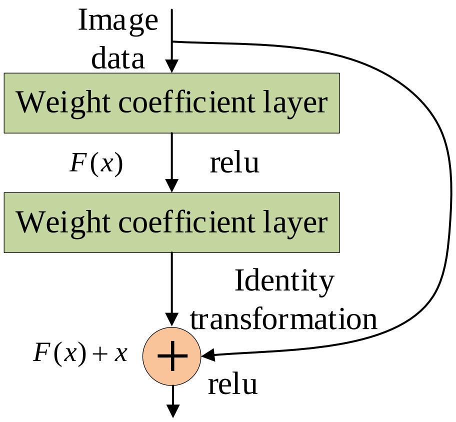


Fig. 4. Structure of residual neural network based on constant mapping

As shown in Fig. 4, the output of the residual module is obtained by superimposing the output of multiple convolutional layers and then performing ReLU processing, as shown in Equations (9) and (10).

$$F = W_2 \times relu(W_1 x) \tag{9}$$

$$y = F(x, W_1) + x \tag{10}$$

In equation (10) $F(x,W_1)$ is the residual mapping function constructed based on the input x , residual matching weights W_1 , W_2, and y is the output data after the resultant residual processing. Here ResNet-50 containing 50 computational layers is used as the basis for forming the PTP-CNN algorithm, and the image data size of the input network needs to be adjusted to 224×224 , then the convolutional kernel and input feature data of each core level in the network are shown in Table I.

TABLE I. CONVOLUTIONAL KERNEL AND INPUT FEATURE PARAMETERS AT THE CORE LEVEL IN THE RESNET-50 NEURAL NETWORK

Layer Name Type	Output Data Size	Number of computing layers	Convolution kernel or pooling window size	Number of convolution kernels
C1	112×112	1	7×7	64
P1	112×112	1	3×3 (maximum pooling)	/
C2_X	56×56	3	1×1, 3×3, 1×1	64, 64, 256
C3_X	28×28	4	1×1, 3×3, 1×1	128, 128, 512
C4_X	14×14	6	1×1, 3×3, 1×1	256, 256, 1024
C5_X	7×7	3	1×1, 3×3, 1×1	512, 512, 2048
P5	1×1	4 (average pooling, fully connected, softmax)	3×3	/

After completing the above preparatory work, we started to formally design the virtual model recognition model of swimmer's stance based on PTP-CNN algorithm, the input image in this model needs to be pre-processed and data augmented, the feature extractor in the model is ResNet-50 convolutional neural network trained by ImageNet dataset, followed by docking pressure leveling layer, deletion layer, SoftMax, and finally, the overall structure of the model is shown in Fig. 5.

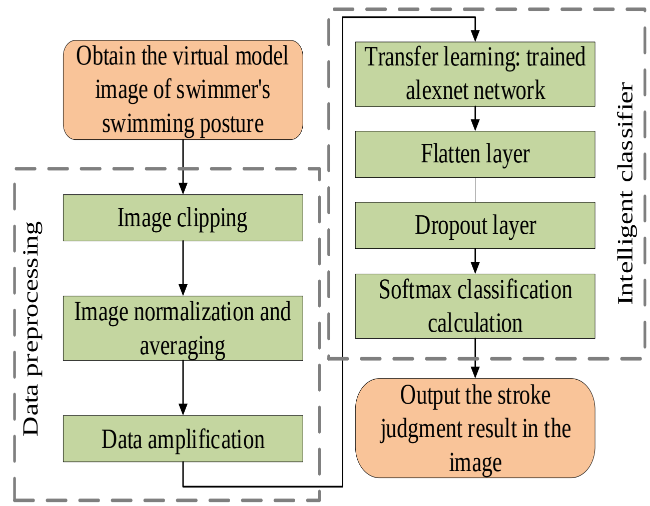


Fig. 5. Structure of virtual image recognition model of swimmer's position pose based on PTP-CNN algorithm

As shown in Fig. 5, the last layer of the standard ResNet network is the SoftMax classifier. But, in order to downscale the dimensionality of the data by one dimension for classification, a flattening layer needs to be added before, and a deletion layer is subsequently added to prevent overfitting of the model. Finally, the PTP-CNN algorithm is trained by using the cross-entropy loss function $loss(x)$, which is given in equation (11).

$$loss(x) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \ln y_{ip} + (1-y_i)\ln(1-y_{ip})] \tag{11}$$

In equation (11), y_i and y_{ip} are the label value and algorithm prediction of the image i , respectively.

IV. PERFORMANCE VERIFICATION EXPERIMENT OF SWIMMING POSTURE DETECTION ALGORITHM

A. *Model Parameter Setting and Performance Verification Experimental Design*

The experiments were designed to verify the recognition performance of the two virtual image recognition models of swimmer's stance designed in this study, and to build a comparative classification model based on the common SVM (Support Vector Machine) algorithm and AlexNet neural network. The data set used for model training is the images of different swimming stances purchased by the research team from a third-party data agency in China, and the images are labeled as "correct stance" and "wrong stance". The data set contains 1843 images, and after data expansion, it reaches 15442 images, which are divided into training set and test set (containing 4632 images) according to the ratio of 7:3. The model based on SW-CNN algorithm also uses cross-entropy loss function and Adam optimization method, and the learning rate is determined to be 0.001 after multiple debugging, and the exponential decay rates of β_1 and β_2 for first-order and second-order moment estimation are 0.9 and 0.998, respectively, and the batch size is 64. The model based on PTP-CNN algorithm is also optimized using Adam, and the learning rate is determined to be. The exponential decay rates of β_1 and β_2 are 0.88 and 0.999 for first-order and second-order moment estimation, respectively, and the batch size is set to 32. The common evaluation indexes of image classification accuracy are Accuracy, Precision, Recall, and F1 value, which are calculated in equations (12) to (15).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

$$\mathrm{Re}\,call = \frac{TP}{TP+FN} \tag{14}$$

$$F1 = \frac{2*Precision*\mathrm{Re}\,call}{(Precision+\mathrm{Re}\,call)} \tag{15}$$

In Eqs. (12)~(15), TP represents the number of samples that are actually positive classes are predicted to be positive classes, referred to as the true positive number, and so on TN, FP, FN are the true negative number, false positive number, and false negative number, respectively. Considering the importance of identifying correct and incorrect postures in swimming teaching, the accuracy rate was chosen as the performance evaluation index of each model.

B. Analysis of Experimental Results

After completing all the performance test experiments, the Excel software and SPSS23.0 software were used to count the changes of the loss function values during the training process for the virtual image recognition models of swimmer's position pose constructed based on SW-CNN algorithm and PTP-CNN algorithm designed in this study, and the comparison models constructed based on SVM algorithm and AlexNet algorithm, as shown in Fig. 6.

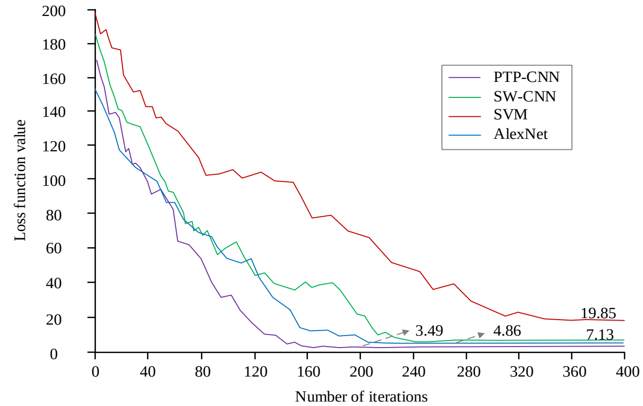


Fig. 6. Loss function value change curve of virtual image recognition model for each swimmer's position pose during training

In Fig. 6, the horizontal axis represents the number of iterations of the core algorithm in each model, the vertical axis represents the loss function value of each algorithm, and different colors represent different virtual image recognition models of swimmer's position pose. As we can see in Fig. 6, the model built based on PTP-CNN algorithm has the fastest convergence speed during the training process and the smallest loss function value after convergence, while the model built based on SVM algorithm has the slowest convergence speed and the largest loss function value after convergence. Specifically, when all the models finished training (i.e., the loss function values converged), the loss function values of the virtual image recognition models constructed based on the PTP-CNN algorithm, AlexNet algorithm, SW-CNN algorithm, and SVM algorithm were 3.49, 4.86, 7.13, and 19.85, respectively. The detection accuracy of each model on the test set is analyzed below. The statistical results are shown in Fig. 7.

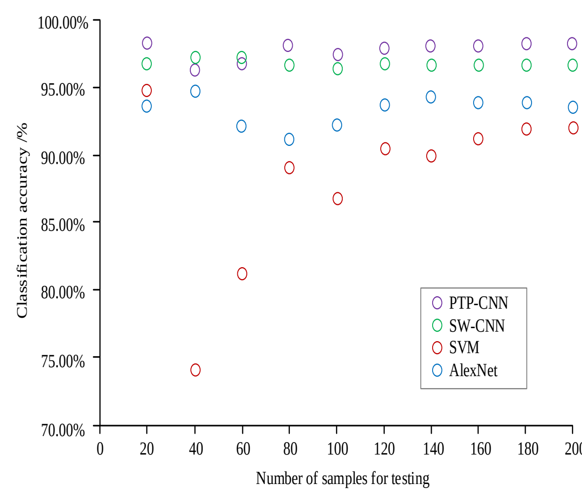


Fig. 7. Recognition accuracy of virtual image recognition models for each swimmer's position pose on the test set

The horizontal axis in Fig. 7 represents the number of samples used to test each model, and the vertical axis represents the classification accuracy of each model on image data in %, with two valid digits retained. Since the accuracy of each model changes slightly after the number of samples exceeds 200, only the test results before the number of test samples is not higher than 200 are retained here. It can be seen from Fig. 7 that the classification accuracy fluctuation degree of the model changes with the number of test samples. The descending sorting results are SW-CNN, PTP-CNN, AlexNet and SVM. However, from the perspective of the overall classification accuracy, the descending sorting results are PTP-CNN, SW-CNN, AlexNet and SVM. When the number of test samples is 200, the classification accuracy of each algorithm is 97.48%, 96.72%, 93.60% and 91.39%, respectively. The image recognition speed of each trained model with different test samples is then analyzed and shown in Fig. 8.

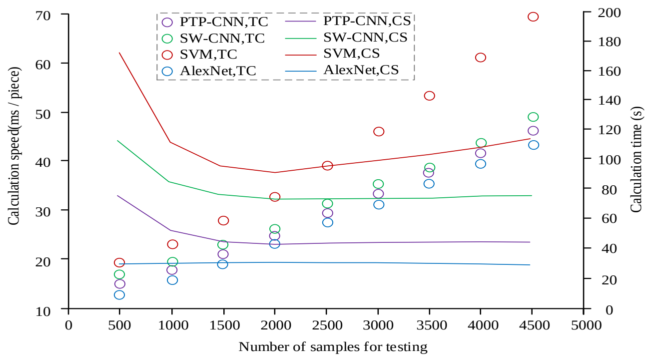


Fig. 8. Computational speed of the virtual image recognition model for each swimmer's position pose on different number of test samples

In Fig. 8, the horizontal axis is the number of samples used to test the model, and the test step size is 500 samples in line with the scale. The circle and "TC" represent the computation time, and the data line and "CS" represent the computation speed. As can be seen in Fig. 8, when the training samples are small, the computational speed of each model is high because the algorithm takes a relatively constant time to start. From the model perspective, the model based on the SVM algorithm has the slowest computation speed due to the high computational complexity of its own algorithm, while the computation speed of the other three models remains roughly stable as the number of samples grows. When the number of test samples is 4500, the computation time of the models built based on PTP-CNN algorithm, AlexNet algorithm, SW-CNN algorithm and SVM algorithm is 24 ms/sheet, 19 ms/sheet, 34 ms/sheet and 45 ms/sheet, respectively. Finally, the computer memory occupancy of each algorithm during operation is analyzed in Fig. 9. It is need to note that five parallel experiments were conducted for each experimental scheme to ensure the reliability of the test results.

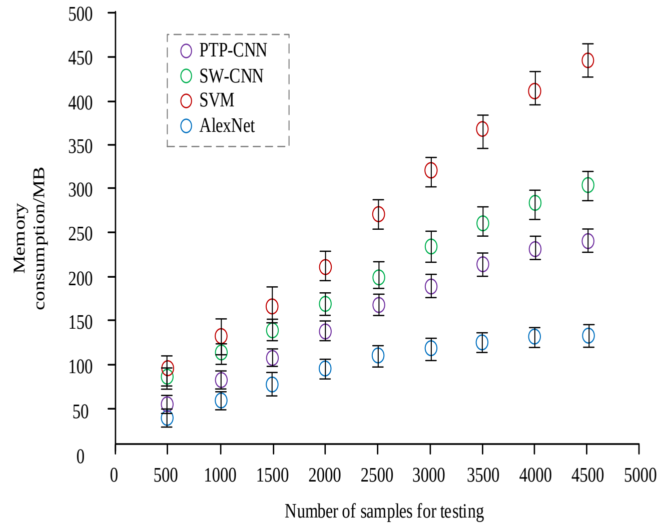


Fig. 9. Memory consumption of the virtual image recognition model computed for each swimmer's position pose

In Fig. 9, the horizontal axis is the number of test samples, the vertical axis is the computational memory consumption in MB, and the vertical coordinate of the center point of the circle in the figure is the mean computational memory consumption of multiple experiments, and the upper and lower limits represent the upper and lower deviation of memory consumption. As we can see in Fig. 9, the computational memory consumption of each model tends to increase as the number of test samples increases, but the memory consumption of the model built by the SVM algorithm grows the fastest, the memory consumption of the model built by the AlexNet algorithm grows the slowest, and the model built by the PTP-CNN algorithm grows the second. When the number of test samples is 4500, the memory consumption of the models constructed based on the PTP-CNN algorithm, AlexNet algorithm, SW-CNN algorithm, and SVM algorithm are 225MB, 105MB, 269MB, and 438MB, respectively.

V. Discussion

For the purpose of helping swimmers learn swimming postures by themselves, this research uses sliding window and depth residual network to build a hybrid neural network algorithm, and based on this algorithm, two corresponding virtual image classification models of swimmers' postures are designed. At the same time, in order to reduce the noise of the input data and reduce the cost of data collection, virtual reality technology is used to convert the swimming pose image of swimmers into an image model in the virtual reality space as the input data of the algorithm. The performance test experiment results show that the model based on PTP-CNN algorithm designed in this study is the first to complete convergence in the training phase, and the loss function value after convergence is also significantly lower than all the comparison models, which is 3.49. The research results of X. Tong et al. also show that the convergence speed of the pose detection neural network incorporating sliding window technology in the training phase is significantly faster than that of the neural network before improvement [19]. From the perspective of the pose recognition accuracy index of the test set, the classification accuracy of the model built based on the PTP-CNN algorithm designed in this study is 97.48%, which is the highest among all the comparison models, while the classification accuracy of the recognition model built based on the SW-CNN, AlexNet and SVM algorithms is 96.72%, 93.60% and 91.39%, respectively. The research results of J. M. Liang et al. also show that the use of virtual reality technology to convert teaching material images into virtual reality models is conducive to improving the work progress of the teaching assistance system [20]. From the perspective of computing efficiency, when the number of test samples is 4500, the computing speed and memory consumption of the model based on PTP-CNN, SW-CNN, AlexNet and SVM algorithms are respectively 24 ms/piece, 34 ms/piece, 19 ms/piece, 45 ms/piece and 225 MB, 269MB, 105 MB, 438 MB. This is mainly because the deep residual network consumes more computing resources than the general convolutional neural network, slowing down the computing speed of the entire model.

VI. Conclusion

To solve the problem of swimming teaching, beginners' learning progress of swimming skills is affected due to insufficient knowledge of correct swimming posture. This study uses virtual reality technology to convert physical images of swimmers' swimming posture and position into models in virtual space, and constructs two hybrid convolutional neural network-based image classification models for virtual models of swimmers' position posture. The performance test results show that the model based on PTP-CNN algorithm designed in this study has the fastest convergence speed in the training phase, and the loss function value after convergence is the smallest than the comparison model, 3.49. The classification accuracy of the models based on PTP-CNN, SW-CNN, AlexNet and SVM algorithms on the test set is 97.48%, 96.72%, 93.60% and 91.39% respectively. When the number of test samples is 4500, the calculation speed and memory consumption are 24 ms/piece, 34 ms/piece, 19 ms/piece, 45 ms/piece and 225 MB, 269MB, 105 MB and 438 MB, respectively. The experimental data shows that the model based on PTP-CNN algorithm and SW-CNN algorithm designed in this study is higher than the common neural network algorithm and machine learning algorithm in the virtual image recognition accuracy of swimmer's swimming posture, but the former has faster computation speed and consumes less resource for computation. However, due to the limitation of experimental conditions, further classification and recognition of swimmer's swimming posture by joints were not carried out, which could help learners to use the recognition model in more cases, and this is also a point for improvement in the subsequent research.

Acknowledgment

This research is supported by: Sichuan International Studies University used the teaching reform project (No. JY2296285) of 2022: college students used PDCA to overcome sports acquired helplessness; Research project of Sichuan International Studies University, Investigation and analysis of college sports coaches and athletes' Sense of Gain in the new era (No. sisu2018074).

References

[1] Y. Zhang, S. Mi, J. Wu, and X. Geng. "Simultaneous 3D hand detection and pose estimation using single depth images," Pattern Recognition Letters, vol. 140, no. 9, pp. 43-48, 2020.

[2] R. Jin, J. Jiang, Y. Qi, D. Lin and T. Song. "Drone detection and pose estimation using relational graph networks," Sensors, vol. 19, no. 6, pp. 1479-1498, 2019.

[3] Hu Z, Xing Y, Lv C, et al. "Deep convolutional neural network-based bernoulli heatmap for head pose estimation," Neurocomputing, vol. 436, no. 5, pp. 198-209, 2021.

[4] Py A, Jy A, Gl B, et al. "Graph neural network for 6D object pose estimation," Knowledge-Based Systems, vol. 218, no. 4, pp. 106839.1-106839.9, 2021.

[5] M. S. Elbamby, C. Perfecto, M. Bennis and K. Doppler. "Towards low-latency and ultra-reliable virtual reality," IEEE Network, vol. 32, no. 2, pp. 78-84, 2018.

[6] M. Chen, W. Saad and C. Yin. "Virtual reality over wireless networks: quality-of-service model and learning-based resource management," IEEE Transactions on Communications, vol. 66, no. 11, pp. 5621-5635, 2018.

[7] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, M. Nießner. "FaceVR: Real-time facial reenactment and eye gaze control in virtual reality," ACM Transactions on Graphics, vol. 37, no. 2, pp. 1-15, 2018.

[8] J. Du, Z. Zou, Y. Shi, D. Zhao. "Zero latency: real-time synchronization of BIM data in virtual reality for collaborative decision-making," Automation in Construction, vol. 85, pp. 51-64, 2018.

[9] C. Cao, S. Ye, H. Tian and Y. Yan. "Multi-scale single-stage pose detection with adaptive sample training in the classroom scene," Knowledge-Based Systems, vol. 222, no. 6, pp. 107008-107013, 2021.

[10] G. Lin, Y. Tang, X. Zou, J. Xiong and J. Li. "Guava detection and pose estimation using a low-cost RGB-D sensor in the field," Sensors, vol. 19, no. 2, 2019.

[11] L. Tang, C. Gao, X. Chen and Y. Zhao. "Pose detection in complex classroom environment based on improved Faster R-CNN," IET Image Processing, vol. 13, no. 3, pp. 451-457, 2019.

[12] L. Zhou, Q. Gao and J. F. Zhan. "3D printed wearable sensors with liquid metals for the pose detection of snakelike soft robots," ACS Applied Materials & Interfaces, vol. 10, no. 27, pp. 23208-23217, 2018.

[13] X. Li, Z. Fan, Y. Liu and Q. Dai. "3D pose detection of closely interactive humans using multi-view cameras," Sensors, vol. 19, no. 12, pp. 2831-2836. 2019.

[14] Q. Gao, J. Liu, Z. Ju and X. Zhang. "Dual-hand detection for human-robot interaction by a parallel network based on hand detection and body pose estimation," IEEE Transactions on Industrial Electronics, vol. 66, no. 12, pp. 9663-9672, 2019.

[15] Hu W and Guan Y. "Landmark-free head pose estimation using fusion inception deep neural network," Journal of Electronic Imaging, vol. 29, no. 4, pp. 43030.1-43030.11, 2020.

[16] C. Chen, T. Wang, D. Li and J. Hong. "Repetitive assembly action recognition based on object detection and pose estimation – ScienceDirect," Journal of Manufacturing Systems, vol. 55, no. 5, pp. 325-333, 2020.

[17] S. Qi, S. Li and J. Zhang. "Designing a teaching assistant system for physical education using web technology," Mobile Information Systems, no. 6, pp. 1-11, 2021.

[18] L. Zhao and Y. Zhao. “The construction of the fusion and symbiosis path of infant sports development based on intelligent environment,” Mathematical Problems in Engineering, 2021, 2021(3):1-9.

[19] X. Tong, R. Li, L. Ge, L. Zhao, K. Wang. “A new edge patch with rotation invariance for object detection and pose estimation,” Sensors vol. 20, no. 3, pp. 1-17, 2020.

[20] J. M. Liang, W. C. Su, Y. L. Chen, S. L. Wu, J. J. Chen. “Smart interactive education system based on wearable devices,” Sensors, vol. 19, no. 15, pp. 3260-3267, 2019.

Dynamic Force-directed Graph with Weighted Nodes for Scholar Network Visualization

Khalid Al-Walid Mohd. Aris[1], Chitra Ramasamy[2], Teh Noranis Mohd. Aris[3], Maslina Zolkepli[4]
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia,
Serdang, Malaysia

***Abstract*—Through the growth of portals and venues to publish academic publications, the number of academic publications is growing exponentially in recent years. An effective exploration and fast navigation in the collection of academic publications become an urgent need to help academic researchers find publications related to their research and the surrounding community. A scholar network visualization approach is proposed to help users to explore a large number of academic publications concerning the strength of the relationship between each publication. The approach is realized by creating a web-based interface using D3 JavaScript algorithm that allows the visualization to focus on how data are connected to each other more accurately than the conventional lines of data seen in traditional data representation. The proposed approach visualizes data by incorporating a force-directed graph with weighted nodes and vertices to give more descriptive information of millions of raw data such as author names, publication title, publication year, publication venue and number of citations from the scholar network dataset. By introducing a weighted relationship in the network visualization, the proposed approach can give a more insightful detail of each publication such as a highly cited publication by looking at and exploring the generated interactive graph. The proposal is targeted to be incorporated into a larger-scale scholar network analytical dashboard that can offer various visualization approaches under one flagship application.**

Keywords—Force-directed graph; weighted network; citation network; D3 algorithm

I. Introduction

The existence of numerous academic social networking websites such as Google Scholar and ScienceDirect has accommodated scholars to publish their scientific publications to the public effortlessly. The purpose of this platform is to acknowledge people about a specific topic in certain disciplines.

A concern regarding academic social networking websites is how to handle the flood of information offered by the websites. People can no longer rely on traditional ways to deal with the outgrowth of scientific publications. Traditional searching and browsing functions at academic social networking websites have become outdated as more time is needed to browse through each publication to see their relevancy.

The overload of information could lead to a lack of efficiency and a lengthy period of time spent searching for valuable information. It will lead to failure in receiving a full, in-depth overview of the desired topics and domains. Therefore, a mechanism to efficiently handle the flood of information needs to be introduced, as it would speed up the process of searching and understanding the scientific network and communities in one specific discipline. An efficient search and analysis of academic networks can also help non-specialists from other disciplines quickly find existing networks that they are interested in.

In general, on academic social networking websites, the browsing function has a fixed classification algorithm that cannot provide a user with desired topics and domains. Some non-specialist users are not able to understand the jargon used in domains unfamiliar to them. Fortunately, humans are intensely visual creatures. Normally, people can read a pattern of the growth of diseases by looking at a chart, and even children can describe a bar chart and extract information from it. For that reason, an efficient scholar network visualisation approach can be an alternative way for users to replace traditional searching and browsing functionalities on academic social networking websites.

The scholar network visualization approach is considered part of data analytics and visualization, which has become a highly active field of research in recent times due to the information overload all around us. Data visualization, which deals with brain psycho-visual vision and cognitive capacities, is a privileged tool to analyze one's environment. Network visualization research can be defined by the techniques that allow humans to visualize data through a network graph that presents a network of connected entities and nodes visually.

Many visualization tools have been introduced in recent years [1,2]. They offer many useful functions, such as data processing and visual analytics. Therefore, it has simplified the process of data visualization for users, whether they have any programming knowledge or not. The tools give users the capability to transform the data into interactive charts that are more understandable and readable by everyone. Data visualization is commonly utilized in business intelligence, scientific visualization, and analytical analysis. There are two types of visualization tools: visualization tools with programming languages and visualization tools without programming languages.

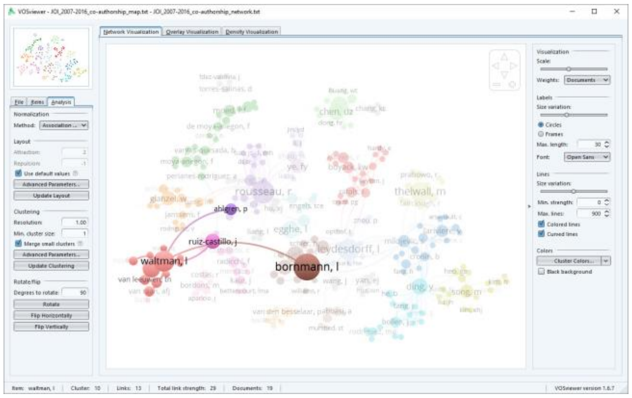


Fig. 1. Network visualization generated in VOSviewer

Tableau [3] is a software that is used mainly in business intelligence and analytics. It supports various file formats, such as txt, xlsx, csv and json. Data can also be imported from online servers such as MySQL and Oracle. Tableau can generate a suitable graph automatically by extracting the header of each variable in our dataset. Users can also use the drag-and-drop feature to add rows and columns and select a chart type. A web-based application called Infogram [4] can complete data visualization quickly; first-time users just need to register, and they can upload their own data files in various formats, such as xlsx or csv to the website. Users can also import data from Google Drive, Dropbox, OneDrive, or a JSON feed. One of the disadvantages of Infogram is data privacy. An open-source JavaScript library, D3.js [5] combines HTML and CSS methods. On D3.js official website, it provides plenty of examples with the source code to inspire users to create their own data visualization. All the graphs generated will be in svg format.

R programming also provides a package called Ggplot2 [6], which is an open-source package to visualize by generating charts. Compared to basic R graphs, the Ggplot2 package allows the user to edit the plotting component of the graph. Ggplot2 also has its own repository on Github that provides the user with an annual case study competition to show their skills. Users have a chance to use the package, and in return, they can contribute codes back to ggplot2.

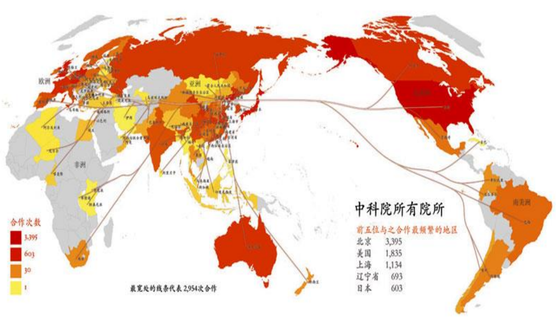


Fig. 2. Chinese Academy of Science Co-author network generated in Sci2

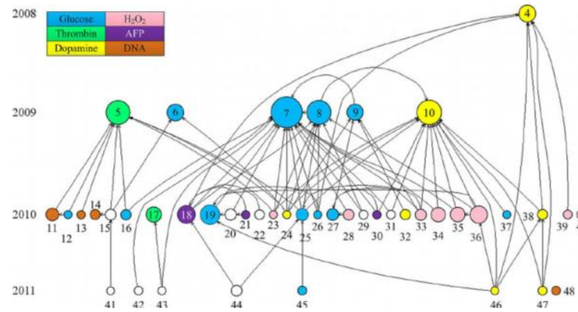


Fig. 3. Mapping papers with highest local citation score generated in HistCite

Another free visualization and analysis tool called VOSviewer [7] is able to construct and visualize bibliometric networks. It visualizes scholarly data into bibliometric networks by clustering solutions. Users can visualize their data by importing the data files from Web of Science (WOS), Pajek, and Graph Modelling Language (GML). The networks can be saved as a bitmap file or in vector format. Fig. 1 shows network visualization generated in VOSviewer. Sci2 [8] that represents The Science of Science is an open-source tool that supports temporal, geospatial, topical, and network studies. It also generates different kinds of networks. The network that generates from small datasets can be explored interactively and the network from large datasets can be rendered in Postscript files that users can convert. Fig. 2 shows a co-authorship network from the Chinese Academy of Science generated in Sci2.

HistCite [9] is used to visualize scholarly data and bibliometric analysis, including the productive authors, the scale of journals, the frequency of words, the types of documents, and the ranking of institutions. A bibliography's dataset will be converted into time-based networks called historiographs by HistCite. The historiograph assists the user in understanding the subject's main publishing events as well as the impact of the chronology on networks. Fig. 3 shows the mapping of 45 papers with the highest local citation score generated in HistCite.

BibExcel [10] is used to do multiple types of bibliometric analysis, such as citation analysis, cluster analysis, and co-citation analysis. The system allows users to select a catalogue from their data and add it as a variable in the data matrix of output files. Users can also export the files that include the data matrix and import them into other visualization tools such as Gephi, Pajek, and VOSviewer to continue their analysis. Fig. 4 shows the mapping science using BibExcel and Pajek.

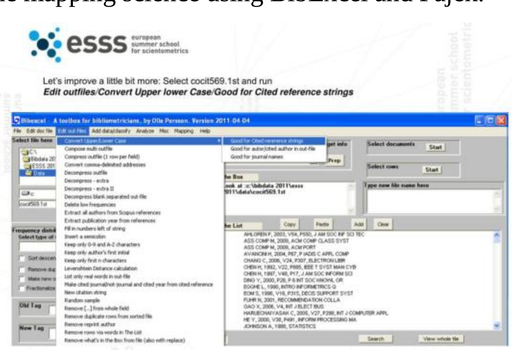


Fig. 4. Mapping science using BibExcel and Pajek

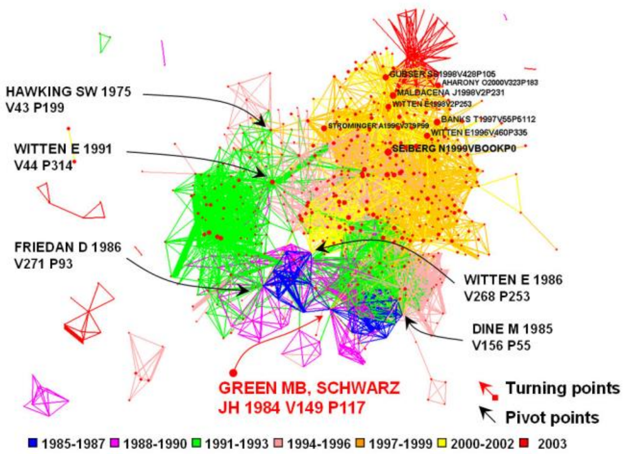


Fig. 5. Visualizing patterns and trends in scientific literature using CiteSpace

A Java application called CiteSpace [11] facilitates the user by detecting, visualizing, and analyzing increasing trends and critical changes in scientific literature. It combines information visualization methods and bibliometrics with the algorithm of data mining to read the pattern in citation data. Fig. 5 shows the visualizing patterns and trends visualised in the scientific literature using CiteSpace. A fuzzy-based clustering visualization approach, Bibliographic Big Data Visualization [12] offers a hybrid fuzzy clustering-based visualization by applying the Fruchterman-Reingold algorithm. The visualization can divide the nodes into soft clusters, but they lack the strength of the connection between the nodes. Fig. 6 shows the fuzzy clustering in Bibliographic Big Data Visualization.

By implementing query optimization and the spectral centrality measure [13], an improved scholar data visualisation was proposed, in which the scholar data is visualised in a network diagram using the centrality measure for better and faster decision making. By using the concept of a word cloud, the visualization offers a weighted network visualization. Fig. 7 shows the enhanced bibliographic data retrieval using query optimization and the spectral centrality measure. Table I compares existing scholar network visualisation approaches, including their functions and limitations.

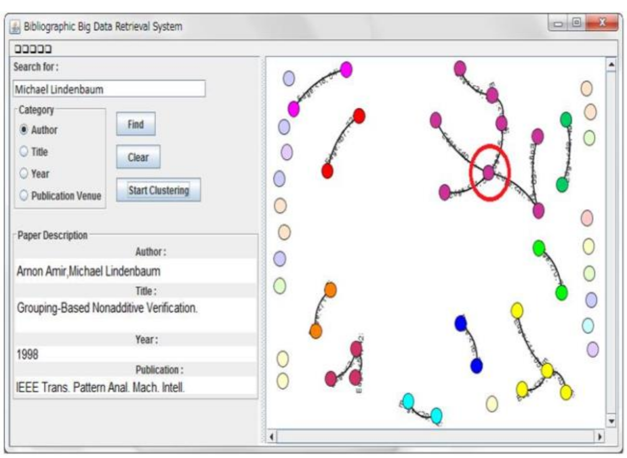


Fig. 6. Fuzzy Clustering in Bibliographic Big Data Visualization System

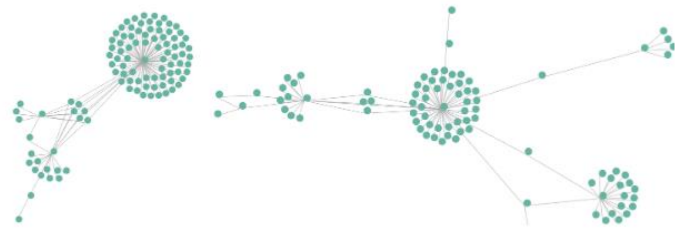


Fig. 7. Enhanced Bibliographic Data Retrieval Using Query Optimization and Spectral Centrality Measure

TABLE I. COMPARISON OF EXISTING SCHOLAR NETWORK VISUALIZATION APPROACHES

Name	Main function	Data format	Platform	Limitation
Tableau[3]	General data analysis	.txt .csv .xlsx	Windows Mac OS	Not programmable to improve to algorithm
Infogram[4]	General data analysis	JSON	Windows Mac OS	Not programmable to improve to algorithm
D3[5]	General data analysis	JSON	Windows Mac OS	Limited data size
Ggplot2[6]	Chart visualization	.csv	Windows Mac OS	Slow to create graphics
VOS Viewer[7]	Citation analysis	WOS Pajek,GML	Windows Mac OS	Only support node network diagram and heat map
Sci2[8]	Network analysis	.txt .csv	Windows Mac OS X Linux	High memory footprint when process large datasets.
HistCite[9]	Static analysis	WOS	Windows but only on IE	Only support data from WOS.
Bib Excel[10]	Process scholar database	WOS Med-line	Wind ows	Not easy to use without its help document
Cite Space[11]	Co-citation analysis	WOS Pub-Med & arXiv	Windows	Cannot delete irrelevant node

Bibliographic Big Data Visualization [12]	Fuzzy Citation Network analysis	AMiner	Java	Takes more than 2 minutes to produce visualization result due to clustering process.
Bibliographic Data Retrieval Using Spectral Centrality Measure[13]	Hybrid Clustering Citation Network analysis	AMiner	Python MongoDB	Only supports JSON/XML format dataset

Most recently, NetV.js [14] high-efficiency visualization approach was introduced for large-scale graphs. It is an open-source JavaScript library that supports the fast visualization of large-scale graph data at an interactive frame rate with a commodity computer. It consists of the Graph Model Manager, the Rendering Engine, and the Interaction Manager. While D3.js library can support up to 20,000 nodes and 400,000 edges, NetV.js can support up to 50 thousand nodes and 1 million edges. For the scholar network dataset used in this study, D3.js is sufficient to produce the visualization as the dataset only has 800 nodes, but to produce large scale graphs, NetV.js is more suitable to be used as the visualization approach.

Another recent approach for visualizing large real-world (social) network data on a high-resolution tiled display system was introduced on a tiled display system consisting of multiple screens [15]. The high resolution tiled display approach used GPUs to ensure an interactive setting with real-time visualization. GPUs are gaining popularity for large-scale datasets because they can process visualization much faster.

Section II describes the scholar network dataset from AMiner and the Fruchterman-Reingold force-directed graph applied in this study. The application of Fruchterman-Reingold to the scholar network dataset and color scheme for graph nodes and vertices visualization is presented in Section III. Section IV discusses the scholar network visualization produced from the analytics, and the research conducted in this study is summarized in Section V.

II. Materials and Method

A. Scholar Network Dataset from AMiner

This section describes the dataset used in the proposed approach. The dataset is acquired from the AMiner website [16-21]. AMiner is a free online web service used to index, search, and mine big scientific data. Data acquired from AMiner is suitable for data analytics operations on academic publication information to identify connections between researchers, conferences, and publications. Some of the insights that can be produced are expert findings, geographic search, trend analysis, reviewer and examiner recommendation, association search, course search, academic performance evaluation, and research domain modeling.

The Scholar network dataset from AMiner consists of eight attributes. Table II shows the data schema of the Scholar network dataset. The citation data is extracted from DBLP, ACM, MAG, and other sources. The dataset attributes include publication id, publication title, publication authors, publication venue, published year, citation number, citing publications' id, and abstract.

B. Fruchterman-Reingold Algorithm for Vertices and Edges Visualization

Force-directed graph [22-25] is used to visualize the scholar network as it provides the ability to convey the relationship between data, the weightage of the relationship, and the flow often brings out the untold insights into the limelight.

The advantages of a force-directed graph include its flexibility to adapt to increasing criteria, its intuitiveness to make a graph easy to be predicted and understood, and its simplicity in terms of fast implementation using minimal lines of code. The interactivity a force-directed graph can offer is also a big advantage as users prefer to interact with the interface for a deeper understanding of the visualization. Lastly, the force-directed graph has a strong theoretical foundation due to its usefulness in multiple fields such as physics and statistics.

In the proposed study, the Fruchterman-Reingold [26] algorithm is selected to become the visualization approach for the scholar network. The Fruchterman-Reingold algorithm offers a dynamic force-directed graph suitable for edge crossing reduction and planar graph drawing. The algorithm introduces two principles, which are the vertices connected by an edge should be drawn near each other and the vertices should not be drawn too close to each other.

TABLE II. Data Schema of the Scholar Network Dataset

Field Name	Description	Example
Id	Publication ID	013ea675-bb58-42f8-a423-f5534546b2b1
Title	Publication title	Prediction of consensus binding mode geometries for related chemical series of positive allosteric modulators of adenosine and muscarinic acetylcholine receptors
Authors	Publication authors	["Leon A. Sakkal", "Kyle Z. Rajkowski", "Roger S. Armen"]
Venue	Publication venue	Journal of Computational Chemistry
Year	Published year	2017
Citation	Citation number	0
Reference	Citing publications' id	["4f4f200c-0764-4fef-9718-b8bccf303dba", "aa699fbffabe-40e4-bd68-46eaf333f7b1"]

Suppose f_a and f_r are the attractive and repulsive forces respectively, with d as the distance between the two vertices and k as the radius of the empty area around a vertex, then

$$f_a(d) = d^2 k \quad (1)$$

$$f_r(d) = -k^2 \quad (2)$$

Given a graph $G = (V, E)$, the combined force applied on vertex v is:

$$(v) = \sum_{(u,v)\in E} f_{a,uv} + \sum_{(u,v)\in V*V} f_{r,\ uv} \quad (3)$$

Fig. 8 shows the general flow of the Fruchterman-Reingold algorithm. In Fruchterman-Reingold, each node applies a repellant force on other nodes that are inversely proportional to the distance between those nodes, and each arc applies an attractive force on its endpoints proportional to the square of the distance between those nodes. Therefore, as linked nodes grow more distant from one another, the attractive force activates quickly and the repellant force drops off, so linked nodes will have the tendency to get back closer to one another. Similarly, as the nodes get increasingly close, the repellant force activates rapidly while the attractive force ceases, and the nodes will be pushed away from each other. Only when the nodes are at a well-adjusted distance from one another, the forces begin to balance; therefore the nodes will slowly stop moving. To keep track of the forces on each node, a *Δx* and *Δy* value for each node is maintained, where they store the gain forces on that node along the *x* and *y* axes. The algorithm is constantly tracking the location of each node since it's possible that a node might be repelled entirely vertically, in which case it will have a strong force in the *y* direction but no force in the *x* direction, or horizontally, where strong force in the *x* direction, no force in the *y* direction. The gain forces in each direction beginning at zero but will be adjusted by the interactions of each node with each other node. Fig. 9 depicts the pseudocode of the Fruchterman-Reingold algorithm.

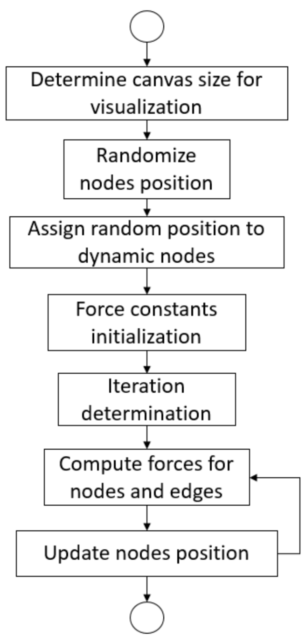


Fig. 8. Flow diagram of the Fruchterman-Reingold algorithm

```
area:= W * L; {W and L are the width and length of the frame}
G := (V, E); {the vertices are assigned random initial positions}
k := sqrt(area/|V|);
function f_a(x) := begin return x^2/k end;
function f_r(x) := begin return k^2/x end;
for i := 1 to iterations do begin
    {calculate repulsive forces}
    for v in V do begin
        {each vertex has two vectors: .pos and .disp
        v.disp := 0;
        for u in V do
            if (u ≠ v) then begin
            {δ is the difference vector between the positions of the two vertices}
                δ := v.pos − u.pos;
                v.disp := v.disp + (δ/|δ|) * f_r(|δ|)
            end
    end
    {calculate attractive forces}
    for e in E do begin
        {each edges is an ordered pair of vertices .vand.u}
        δ := e.v.pos − e.u.pos;
        e.v.disp := e.v.disp − (δ/|δ|) * f_a(|δ|);
        e.u.disp := e.u.disp + (δ/|δ|) * f_a(|δ|)
    end
    {limit max displacement to temperature t and prevent from displacement
outside frame}
    for v in V do begin
        v.pos := v.pos + (v.disp/|v.disp|) * min(v.disp, t);
        v.pos.x := min(W/2, max(−W/2, v.pos.x));
        v.pos.y := min(L/2, max(−L/2, v.pos.y))
    end
    {reduce the temperature as the layout approaches a better configuration}
    t := cool(t)
end
```

Fig. 9. Pseudocode of the Fruchterman-Reingold algorithm

The Fruchterman-Reingold algorithm is applied in the experiment through a plugin in D3.js [5].

III. Results

A. Data Preprocessing and Exploration on the Scholar Network Dataset

The initial data consists of academic publications from 1936 to 2018. To ensure that the visualization process is fast and the graph produced is manageable, data earlier than 2010 is excluded from the experiment. Only data from 2010 and above will be visualized in the final visualization. The initial attributes of the dataset include publication id, publication title, authors' name, publication venue, published year, citation number, citing publication ID, and abstract. In the experiment, only 5 attributes are included: the authors' name, publication ID, title, number of citations, and year of publication. From the scholar dataset exploration, there is an increasing number of academic publications from year to year. Fig. 10 shows the histogram of publications from 2010 to early 2018.

A network graph has two key data elements, nodes/vertices, and links/edges. All nodes must have unique identifiers. In each node, it is possible to add as many custom variables as necessary. Links must have a valid node id as a source and a target, and they can be text or numbers. Fig. 11 shows a snapshot of the cleaned dataset ready to be visualized in the D3 algorithm. The intention is to develop a node-to-node relationship to emphasize the relationship between authors and their publications. Every node is connected to the target node with the same relationship. Another feature of the proposed approach is that every node will have a different color based on the year it was published to the public, and the radius of the node will correspond with the number of citations in every academic publication. If the user enters the author's name in the space provided, it will highlight the other nodes that are related to it. If users hover the mouse over one of the nodes, they can see the details for every academic publication they want. The information will be displayed on the left side of the graph.

B. *Color Scheme for Graph Nodes and Vertices Visualization*

The color palette for the nodes and vertices was chosen according to the Web Content Accessibility Guidelines (WCAG) [27] which suggests the minimum contrast ratio between text or image and background is 4.5:1. Table III describes the color ratio for every color used in the force-directed graph visualization. Ten colors are chosen to represent 10 different clusters of the scholar network to be visualized in the graph. If more than 10 network clusters exist, the same color will be repeated in other clusters.

TABLE III. COLOR RATIO FOR EVERY COLOR IN THE FORCE-DIRECTED GRAPH

Color Code	Contrast Ratio	Color
#ffffff	18.37	
#ffb646	10.52	
#ff863d	7.63	
#ff8882	7.57	
#00aa9f	6.34	
#1d9c3d	5.14	
#ff352e	5.07	
#c06c30	4.74	
#9262f8	4.65	
#0781df	4.56	

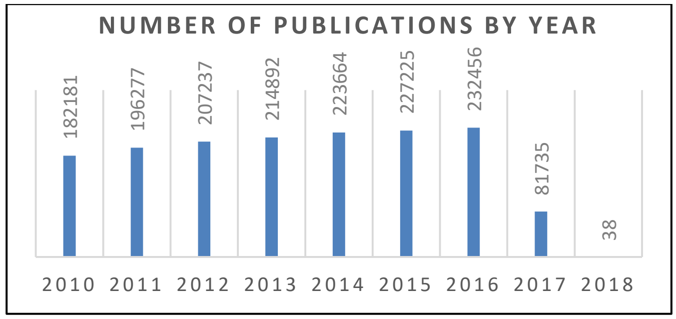


Fig. 10. Number of academic publications from 2010 to 2018 from AMiner

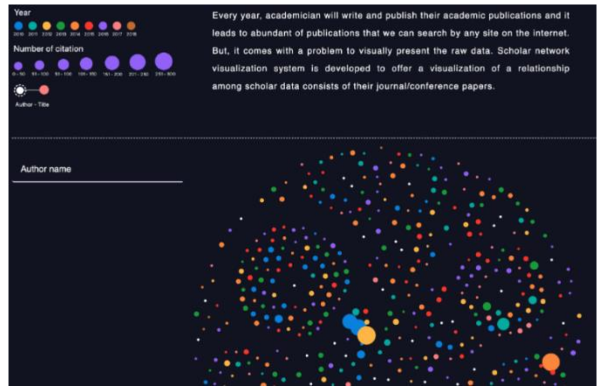


Fig. 11. The visualization of the Scholar Network Visualization approach (Initial Graph)

IV. DISCUSSIONS

This section discusses the scholar network visualization produced using the D3 library based on the experiment done on the scholar network dataset.

The Fruchterman-Reingold algorithm analyses the scholar network dataset to produce a dynamic force-directed graph visualization, and the visualization is created using the D3 algorithm [28]. D3 is a JavaScript library for manipulating documents based on data. It can bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards offers the full capabilities of modern browsers without tying the data to a proprietary framework, combining powerful visualization components with a data-driven approach to DOM manipulation. It is an increasingly popular approach to data analytics visualization as it can produce sophisticated data visualization that is fast, interactive, and shareable across many platforms.

The graph produced contains nodes linked by lines that represent the relationship between the nodes. D3 implements the Fruchterman-Reingold algorithm to give the user more control over the layout. It implements three primary forces upon the nodes at each tick:

- The sum of the forces acting on each node by all other nodes
- The force pushing and pulling between two linked nodes
- The force pulling each node to a focal point, usually the center of the user-defined space.

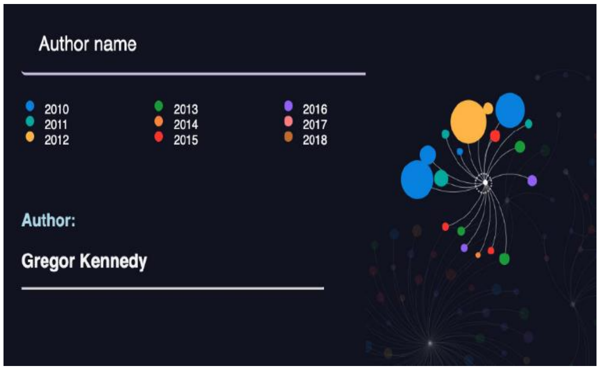


Fig. 12. The highlighted nodes after the user search for “Gregor Kennedy”

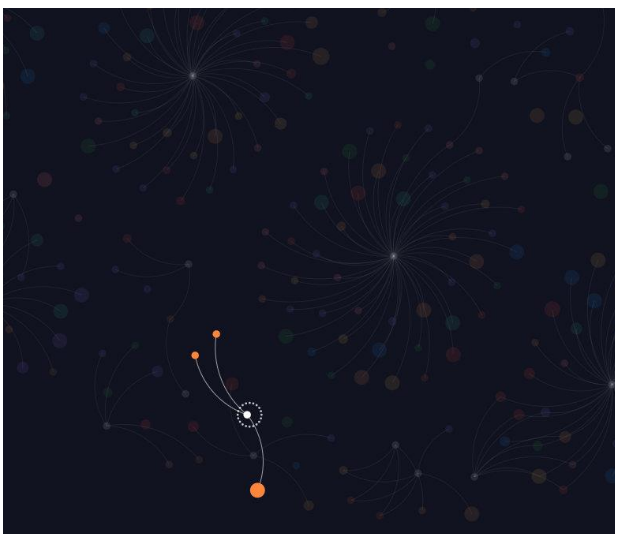


Fig. 13. Visualization of academic publications from author name “Maslina Zolkepli”

Fig. 14. Visualization of academic publications from author name “Mahdi Abavisani”

Fig. 12 shows the live interface of the D3 algorithm. Initially, when no keyword is supplied by the user, the generated graph consists of all the data that exists in the scholar network dataset. This is feasible when the dataset has between 50 and 500 nodes. If the number of nodes gets larger than 500, the graph becomes too crowded and difficult to comprehend and understand. To ensure that the graph is manageable, data cleaning and pre-processing are required to ensure only recent data is included in the dataset.

Fig. 13 shows the live interface when the keyword "Gregor Kennedy" is supplied by the user. Based on the keyword, a new graph is generated with only data that is related to "Gregor Kennedy". The nodes are weighted according to the number of citations. The larger the node, the higher the number of citations each publication has. Each node is also linked to other nodes that show the link between a publication and its citations. Fig. 14 shows the scholar network visualization result for the keyword "Maslina Zolkepli". Only vertices that contain the information related to the supplied keyword will be highlighted, and users can focus on the highlighted note. Unrelated vertices will be disabled, thus saving resources and time to produce the visualization and speeding up the process for further actions. Fig. 15 shows the scholar network visualization result for the keyword "Mahdi Abavisani". When the mouse cursor is placed on the node, the description of the publication is displayed in the left-hand corner of the page.

V. Conclusion

The abundance of scholarly data that is available right now brings a variety of opportunities and challenges for scholarly data analysis. Users are more aware of the significance of applying visualization technologies to different datasets to comprehend the science itself. Thus, the scholar network visualization approach plays a role in addressing the problems that arise from large volumes of diverse and important data.

In this paper, a scholarly network visualization approach is proposed by incorporating weighted nodes and vertices in a dynamic force-directed graph generation. By offering weighted nodes and vertices, users can get an informative view of the visualization. The network graph is dynamic and responds to the user’s action in order to focus on several important nodes as requested by the user. The nodes can be further explored by clicking on them, and the related information will be displayed. The proposed approach is expected to increase the significance of data visualization and highlight some insights for people.

Some of the suggestions for the improvement of the scholar network visualization approach in the future are that it should be able to categorize visualization into specific fields and domains to decrease the visualization complexity. It also should be able to use various visualization techniques that can handle large-scale graphs, such as NetV.js and network visualization using a tiled display system, to deal with the ever-increasing complexity of the data. By exploring more ways to visualize data, the scope of the data can also be increased to show more relationships between the data in the best way possible.

REFERENCES

[1] J.M. Brunetti, S. Auer, R. Garcia, J. Klimek, M. Necasky, “Formal Linked Data Visualization Model,” in Proc. Intl. Conf. on Inf. Integration and Web-based Appl. & Svcs, ACM, New York, NY, USA, 2013, pp. 309-318.

[2] F. Desimoni, L. Po, “Empirical evaluation of Linked Data visualization tools,” Future Generation Computer Systems, vol. 112, pp. 258-282, 2020.

[3] J. Hoelscher, A. Mortimer, “Using Tableau to visualize data and drive decision-making,” Journal of Accounting EducationI, vol. 44, pp. 49-59, 2018.

[4] F. Khouzam, N. Sharaf, M. Saad, C. Sabty, S. Abdennadher, “Automatic Infogram Generation for Online Journalism,” in 23rd International Conference Information Visualisation (IV) 2019, IEEE,Paris, France, 2019, pp. 56-6.

[5] A.A Khade, “Performing Customer Behavior Analysis using Big Data Analytics,” Procedia Computer Science, vol. 79, pp. 986-992, 2016.

[6] H. Wickham, “ggplot2: Elegant Graphics for Data Analysis (Use R!),” 2nd ed., New York, USA: Springer-Verlag, 2016.

[7] L. Xie, Z. Chen, H. Wang, C. Zheng, J. Jiang, “Bibliometric and Visualized Analysis of Scientific Publications on Atlantoaxial Spine Surgery Based on Web of Science and VOSviewer,” World Neurosurgery, vol. 137, pp. 435-442, 2020.

[8] K. Börner, “Plug-and-Play Macroscopes: Network Workbench (NWB), Science of Science Tool (Sci2), and Epidemiology Tool (Epic),” in Encyclopedia of Social Network Analysis and Mining, pp. 1280-1290, 2014.

[9] E. Garfield, “From the science of science to Scientometrics visualizing the history of science with HistCite software,” Journal of Informetrics, vol. 3, no.3, pp. 173-179, 2009.

[10] O. Persson, R. Danell, J. Wiborg Schneider, “How to use Bibexcel for various types of bibliometric analysis,” in Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, ed. F. Åström, R. Danell, B. Larsen, J. Schneider, 2009, pp. 9–24.

[11] C. Chen, F. I. Sanjuan, J. L. Hou, “The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis,” J. Assoc. Inf. Sci. Technol., vol. 61, pp. 1386-1409, 2010.

[12] M. Zolkepli, F. Dong, K. Hirota, “Visualizing Fuzzy Relationship in Bibliographic Big Da-ta using hybrid approach combining fuzzy c-means and Newman-Girvan algorithm,” Journal of Advanced Computational Intelligence and Intelligent Informatics(JACIII), vol. 18, no.6, pp. 896-907, 2014.

[13] C. Ramasamy, M. Zolkepli, “Enhanced Bibliographic Data Retrieval and Visualization Using Query Optimization and Spectral Centrality Measure,” Journal of Advanced Research in Dynamical and Control Systems(JARDCS) vol.11, no.3, pp. 1734-1742, 2019.

[14] D. Han, J. Pan, X. Zhao, W. Chen, “NetV.js: A web-based library for high-efficiency visualization of large-scale graphs and networks,” Visual Informatics, vol. 5, no. 1, pp. 61-66, 2021.

[15] G.G. Brinkmann, K.F.D. Rietveld, F.J. Verbeek, F.W. Takes, “Real-time interactive visualization of large networks on a tiled display system,” Displays, vol. 73,pp. 102164, 2022.

[16] J. Tang, J., A. C. M. Fong, B. Wang, J. Zhang, “A Unified Probabilistic Framework for Name Disambiguation in Digital Library,” IEEE Transaction on Knowledge and Data Engineering (TKDE), vol. 24, no. 66, pp. 975-987, 2012.

[17] J. Tang, D. Zhang, L. Yao, “Social Network Extraction of Academic Researchers,” in Proceedings of 2007 IEEE International Conference on Data Mining(ICDM'2007), pp. 292-301, 2007.

[18] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, Z. Su, “ArnetMiner: Extraction and Mining of Academic Social Networks,” in Proceedings of the 14th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (SIGKDD'2008), pp. 990-998, 2008.

[19] J. Tang, J. Zhang, R. Jin, Z. Yang, K. Cai, L. Zhang, Z. Su, “Topic Level Expertise Search over Heterogeneous Networks,” Machine Learning Journal, vol. 82, no. 2, pp. 211-237, 2011.

[20] J. Tang, L. Yao, D. Zhang, J. Zhang, “A Combination Approach to Web User Profiling,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 5, no. 1, 2010.

[21] H. Wan, Y. Zhang, J. Zhang, J. Tang, “AMiner: Search and Mining of Academic Social Networks,” Data Intelligence, vol.1, no.1, pp. 58–76, 2019.

[22] J. Lu, Y.W Si, “Clustering-based force-directed algorithms for 3D graph,” The Journal of Supercomputing, vol. 76, no. 6, pp. 9654–9715, 2020.

[23] S.H. Cheong, Y.W. Si, R.K. Wong, “Online force-directed algorithms for visualization of dynamic graphs,” Information Sciences, vol. 556, pp. 223-255, 2021.

[24] R. Tamassia, “Handbook of Graph Drawing and Visualization,” 1st ed., London, England: Chapman & Hall/CRC, 2016.

[25] D. L. Reingold “Chapter 4 - Installation, orientation, and layout,” in Analyzing Social Media Networks with NodeXL, 2nd ed., Cambridge, MA, USA: Morgan Kaufmann, 2020, pp. 55-66.

[26] T.M. Fruchterman, E.M. Reingold, “Graph drawing by force-directed placement,” Software: Practice and Experience, vol. 21, no. 11, pp. 1129–1164, 1991.

[27] S.H. Li, D.C. Yen, W.H. Lu, T.L. Lin, “Migrating from WCAG 1.0 to WCAG 2.0 – A comparative study based on Web Content Accessibility Guidelines in Taiwan,” Computers in Human Behavior, vol. 28, no. 1, pp. 87-96, 2012.

[28] R. W. Milton, “Geospatial Computing: Architectures and Algorithms for Mapping Applications,” Ph. D. dissertation, The Bartlett Centre for Advanced Spatial Analysis, University College London, London, England 2019. [Online]. Available: https://discovery.ucl.ac.uk/id/eprint/10072340/.

The Role of Machine Learning in Remote Sensing for Agriculture Drought Monitoring: A Systematic Review

Aries Suharso[1], Yeni Hediyeni[2], Suria Darma Tarigan[3], Yandra arkeman[4]
Department of Computer Science, IPB University, Bogor, Indonesia[1, 2]
Informatics of Computer Science, University of Singaperbangsa Karawang[1]
Department of Soil and Land Resource, IPB University, Bogor, Indonesia[3]
Department of Agro-Industrial Technology, IPB University, Bogor, Indonesia[4]

***Abstract*—Agricultural drought is still difficult to anticipate even though there have been developments in remote sensing technology, especially satellite imagery that is useful for farmers in monitoring crop conditions. The availability of open and free satellite imagery still has a weakness, namely the level of resolution is low and coarse with atmospheric disturbances in the form of cloud cover, as well as the location and period for taking images that are different from the presence of weather stations on Earth. This problem is a challenge for researchers trying to monitoring agricultural drought conditions through satellite imagery. One approach that has recently used is high computational techniques through machine learning, which is able to predict satellite image data according to the conditions of mapping land types and plants in the field. Furthermore, using time series data from satellite imagery, a predictive model of crop cycles can be regarding future crop drought conditions. So, through this technology, we can encourage farmers to make decisions to anticipate the dangers of agricultural drought. Unfortunately, exploration of the use of machine learning for classification and prediction of agricultural drought conditions has not conducted, and the existing methods can still improve. This review aims to present a comprehensive overview of methods that used to monitor agricultural drought using remote sensing and machine learning, which are the subjects of future research.**

Keywords—Drought monitoring; exploration of the use of machine learning; Landsat imagery; remote sensing

GLOSSARY

Term	Description
AMSR-E	Advanced Microwave Scanning Radiometer 2
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
AWS	Autonomous Weather Stations
BRT	Boosted Regression Trees
CDR	Climate Data Record
CHOMPS	CICS High-Resolution Optimal Interpolation Microwave Precipitation from Satellite
CMAP	CPC Merge Analysis of Precipitation
DEM	Digital Elevation Model
DFNN	Deep Forward Neural Network
DT	Decision Tree
ERT	Extreme Regression Tree
ESA-CCI	European Space Agency - Climate Change Initiative
ESTARFM	Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model
EVI	Enhanced Vegetation Index
GA	Genetic Algorithm
GAM	General Additive Model
GDEM	Global Digital Elevation Model
GLDAS-2	Global Land Data Assimilation System Version-2
GMDH	Group Method of Data Handling
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement
GRACE	Gravity Recovery and Climate Experiment
HSMDI	High Soil Moisture Drought Index
IMERG	Integrated Multi-satellitE Retrievals for GPM
ISMN	International Soil Moisture Network
KKN	K-nearest neighbors algorithm
Landsat ETM	Landsat + Enhanced Thematic Mapper
LST	Land Surface Temperature
M5P	is a reconstruction of Quinlan's M5 algorithm for inducing trees of regression models.
MERRA-2	Modern-Era Retrospective analysis for Research and Applications
MIDI	Microwave Integrated Drought Index
MLP	Multi-Layer Preceptron
MODIS	Moderate Resolution Imaging Spectroradiometer
MCD43C4	MODIS Product
MOD11C1	MODIS Product
MOD13A3	MODIS Product
MYD11C3	MODIS Product
MYD13C2	MODIS Product
MCD12Q1	MODIS Product provides global land cover types at yearly intervals (2001-2016) derived from six different classification schemes
MCD43A4	MODIS Product contains 16 days of data provided in a level-3 gridded data set in Sinusoidal projection
MOD09A1	MODIS Product provides an estimate of the surface spectral reflectance of Terra MODIS bands 1-7 at 500m resolution and corrected for atmospheric conditions such as gasses, aerosols, and Rayleigh scattering
MOD11A2	MODIS Product provides an average 8-day land surface temperature (LST) in a 1200 x 1200 kilometer grid
MOD16A2	MODIS Product provide Evapotranspiration/Latent Heat Flux product is an 8-day composite product produced at 500 meter pixel resolution
NDVI	Normalized Difference Vegetation Index

NOAA	National Oceanic and Atmospheric Administration
ORLIKE-OWA	ORLIKE-Ordered Weighted Averaged (OWA)
ORNESS-OWA	ORNESS-Ordered Weighted Averaged (OWA)
PCI	Precipitation Condition Index
PERSIANN	Precipitation Estimation From Remotely Sensed Information using Artificial Neural Networks
RCI	Rainfall Condition Index
RF	Random Forest
RFE	Recursive Feature Elimination
SMAP	Soil Moisture Active Passive
SMDI	Soil Moisture Deficit Index
SPEI	Standardized Precipitation Evaporation Index
SPI	Standardized Precipitation Index
SRTM	Shuttle Radar Topography Mission
SVM	Support Vektor Machine
SVR	Support Vektor Regression
SWDI	Soil Water Deficit Index
TAMSAT	Tropical Applications of Meteorology using Satellite data
TCI	Temperature Condition Index
TRMM (3B43)	Tropical Rainfall Measuring Mission
UAV	Unmanned Aerial Vehicle
VCI	Vegetation Condition Index
VSDI	Shortwave Infrared Drought Index
VSWI	Vegetation Supply Water Index
VTCI	Vegetation Temperature Condition Index

I. Introduction

One of the problems of rainfed agriculture productivity is prolonged drought, lack of rainfall, and lack of water supply in the soil during the vegetative growth phase [1], [2], [3], [4]. In addition, high temperatures during the ripening phase can reduce the conversion yield of sucrose to fructose and glucose [5]. Climate change can also cause diseases and pests [6]. Therefore, it is essential to monitor drought conditions to schedule appropriate irrigation based on the response of plants to drought at various stages of vegetation [7], [8].

However, measuring plant response to drought is very difficult and complex [9], [10], [11], [12], [13], [14]. Detecting and integrating crop water deficits is still complex based on single plant responses [15]. Until 2017 [16] grouped four methods to monitor plant response to drought, namely, (1) Groundwater measurement; (2) Groundwater balanced approach; (3) Plant-based approach; (4) Remote sensing methods. The approach (4) remote sensing is based on the spectral index of vegetation obtained from the Unmanned Aircraft Systems (UAS) hyperspectral sensor, which is the best considering the cost of the sensor is not expensive; the determination of leaf moisture status indicators and plant stomata conductance is high. Non-destructive and non-labor intensive is suitable for automation. The remote sensing method can be adopted as an irrigation scheduling decision [17].

The fact there is an abundance of free Landsat satellite data with open access globally by the US Geological Survey (USGS) starting in 2008 [18] on the Earth Resources Observation and Science (EROS) Center website has attracted researchers from various countries to apply it as a producer of land use land cover (LULC) maps in their respective regions [19], [20]. However, constructing medium and high-resolution land cover maps in cloud-prone areas is still challenging due to infrequent satellite visits and the lack of cloud-free data. It is both an opportunity and a challenge for researchers to accurately map plant droughtes with hyperspectral indices through machine learning classification methods for persistent cloud areas with high temporal dynamics of land cover types that require further investigation. Overall, there have been numerous former studies showing that the use of remote sensing to monitor drought has increased significantly in recent times. Still, the application of machine learning to remote sensing for drought monitoring has not been well-diversified, so there are still numerous exploration gaps that show that its application has not been thoroughly assessed or utilized for drought monitoring purposes.

As a result, in this article we attempt to conduct a systematic review utilizing the meta-analysis method of prior studies using machine learning techniques in remote sensing for agricultural drought monitoring. Meta-analysis methods and systematic reviews can aid in the creation of evaluations that are clearer and more succinct [21]. If there are more studies on similar subjects, the advantages of systematic reviews can be further extended [22]. Systematic reviews can help scientists uncover factors faster, lessen data bias, more accurately define variables, spot trends that previous researchers might have missed, and choose the direction of future study topics [23]. Additionally, systematic reviews can assist researchers in comparing, debating, and choosing from the larger body of literature in order to obtain more trustworthy results [24].

II. Related Work

The use of machine learning techniques to categorize satellite imaging data in remote sensing applications has gained popularity in recent years. On this subject, several research studies have been released, some of which are listed in the paragraphs below.

- Various formalisms are used in applications of machine learning and signal/image processing, including classification and clustering, regression and function approximation, image coding, recovery and enhancement, source separation, data aggregation, and feature selection and extraction [25].
- Machine learning techniques have recently been used in various ways to process data from multispectral and hyperspectral remote sensing [26].
- Using the input data from the satellites Spot5, Sentinel1, and Sentinel2, a Symbolic Machine Learning (SML) classifier with spatial generalization treatment, random theme noise, and spatial displacement noise was created. It made use of multiple Maximum Likelihood Supervised Algorithms, Logistic Regression, Linear Discriminant Analysis, Naive Bayes, Decision Trees, Random Forests, and Support Vector Machines [27].

- Training data needs, user-defined parameter selection and optimization, impact and attenuation feature space, computing costs, and choice of k-nearest neighbor algorithms, enhanced DT, single decision tree (DT), Random Forest, and somewhat mature support vector machine (k-NN) approaches are all taken into account [28].

These studies have shown how well machine learning algorithms work for categorizing remote sensing images and the possibility for further improving the precision and effectiveness of these algorithms through on-going study and development. Previous studies on machine learning in remote sensing have concentrated on a range of methods, such as deep learning algorithms and other supervised and unsupervised approaches, and have investigated their application to various types of remote sensing data and application domains. Overall, applying machine learning to remote sensing has the potential to dramatically increase this field's capabilities and open up a number of new and enhanced applications for satellite data.

III. Materials and Methods

The aim of this work is expected to be able to answer the following four research questions (*RQ*):

- RQ1: What publications are the main targets of machine learning based remote sensing drought monitoring?
- RQ2: What kind of environment observed for drought monitoring? What types of remote sensing data have been used?
- RQ3: Which is the most widely used and most accurate machine learning algorithm for the drought monitoring approach?
- RQ4: How does machine learning play a role in drought monitoring?

After determining the research question (*RQ*) of interest, selecting a candidate paper, and performing data extraction, the last step of a systematic literature study is to synthesize the results. For each *RQ*, the inclusion results are classified into categories corresponding to the *RQ*, and the results are presented in graphs or tables. Furthermore, the results are discussed using various evaluation approaches. Finally, the narrative summary describes the main findings of the systematic literature study.

In this work, we collect and determine the most relevant literature for this particular study with the PRISMA method [29] search strategy to provide a comprehensive and systematic review of relevant previous studies related to the role of machine learning algorithms in remote sensing for drought monitoring on crop land cover maps. Food, semi-arid plantation is suspected to experience drought. A recent search was conducted on Harzing's Publish or Perish search engine with open data sources Google Schoolar and Crossref based on the title text "Drought monitoring" with keywords "remote sensing" and "machine learning" in the publication period between 2010 and 2021.

This study eliminates research that does not use remote sensing and machine learning approaches from the collection of articles obtained. Each article is rated based on the use of remote sensing databases, machine learning methods, accuracy of results, and year of publication. There are about 1147 articles on remote sensing drought monitoring published from 2010 to 2021 (Fig. 1). The search for literature was conducted on July 6, 2022, through the search engine Harzing's Publish or Perish on two open-source articles, namely Google scholar and Crossref with the context of the article title "drought monitoring" and the keywords "remote sensing" and "machine learning," with the limitation of the publication period between 2010 and 2021.

The literature search selection process in Fig. 1 is conducted according to the PRISMA concept, as follows:

1) Identification: initial search obtained 147 articles from open-source Google Scholar and one thousand articles from open-source Crossref. Our next step is to limit the selected articles based on the number of citations in each article to at least twenty citations. This is done to select articles that have referenced popularity by researchers. The results of this limit of twenty citations selected thirty-four articles from the open-source Google Scholar and 171 articles from the open source Crossref, so that the initial number of identified article data containing the context of the article title "drought monitoring" and the keywords "remote sensing" and "machine learning" was as much as 205 articles.

2) Screening: 205 articles from the previous stage (Identification) were checked for duplication of articles, and it turned out that there were 11 related articles, so that they were obtained (n = 194). The process at this stage is conducted on the Microsoft Excel application. Next is the excluded process, namely, discarding a number of articles that do not contain relevant text related to "remote sensing" and "machine learning" in the Abstract section. The results excluded at this stage are n = 166, so the remaining n = 28 articles.

3) Eligibility: at this stage, the articles are examined in full text with the aim of finding research articles that consistently apply machine learning algorithms and the studies carried out contain quantitative analysis or accuracy values. The results are discarded (n = 8 papers without the use of machine learning algorithms); (n = 5 types of paper reviews); (n = 3 papers without quantitative analysis or accuracy scores), leaving (n = 12) articles using machine learning algorithms. The process at this stage is conducted on the Zotero and Mendeley application.

4) Included: from n = 12 selected articles containing the context of "drought monitoring", "remote sensing", and "machine learning", with the type of research article based on observation or experimentation, not a review article. This is done because of a systematic review and meta-analysis, not a narrative review. Furthermore, the selected articles are used as a reference for the main systematic review or meta-analysis.

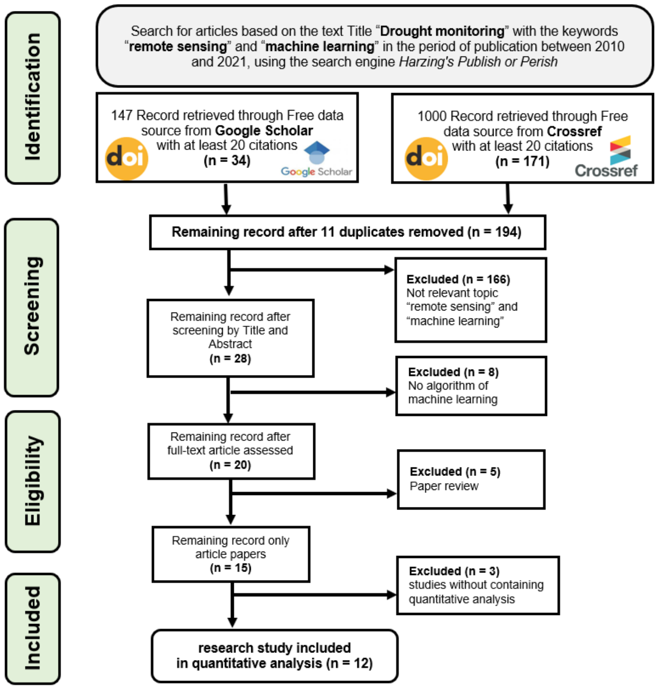


Fig. 1. PRISMA workflow diagram for new systematic review which included search of free database.

IV. Result and Discussion

The inclusion of the PRISMA strategy brief yielded the results of twelve articles that were then analyzed in depth for the content of a meta-analysis that could answer four research questions (*RQ*s).

In response to *RQ*1, Fig. 2 demonstrates that out of a total of 194 papers, the publishers who publish the most scientific journals mention remote sensing-based drought monitoring. The breakdown is as follows: MDPI 50% (*n* = 97), Elsevier 30% (*n* = 57), Taylor & Francis 10% (*n* = 20), IEEE 3% (*n* = 7), Springer and Wiley both 2% (*n* = 4), and the remaining 3% from various publishers (*n* = 5).

Details of the names of the candidate publication journals are listed in Table I.

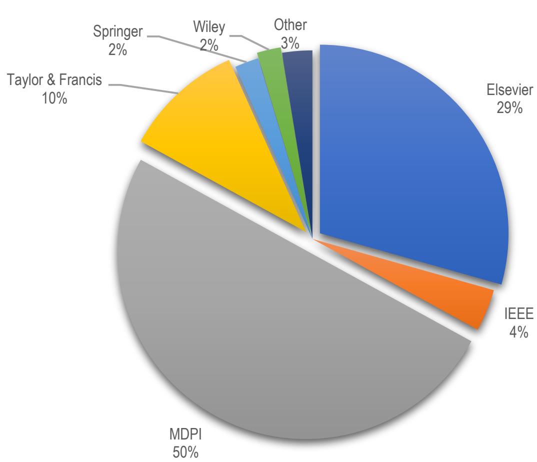


Fig. 2. Publication sources of selected study works.

TABLE I. Publication Source Selected Papers

Journal Name	Publisher	Total
Remote Sensing	MDPI	97
Remote Sensing of Environment	Elsevier	42
International Journal of Remote Sensing	Taylor & Francis	13
GIScience & Remote Sensing	Taylor & Francis	5
Journal of Applied Remote Sensing	Other	5
Agricultural and forest meteorology	Elsevier	4
Water Resources Research	Wiley AGU	4
Environmental monitoring and assessment	Springer	4
IEEE Geoscience and Remote Sensing Letters	IEEE	3
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	IEEE	3
International Journal of Applied Earth Observation and Geoinformation	Elsevier	3
ISPRS Journal of Photogrammetry and Remote Sensing	Elsevier	3
Computers and Electronics in Agriculture	Elsevier	2
Journal of Hydrology	Elsevier	2
Remote Sensing Letters	Taylor & Francis	2
Science of The Total Environment	Elsevier	2

The list of journal names in Table I can be used as a reference source. It is remarkably interesting to observe that all these journals are indexed in the Journal Citation Report, mostly in the *Q*1 and *Q*2 quartiles.

Fig. 3 presents the trend in the number of articles published per year from 2010 to 2019. This graph shows that there has been a significant increase in the number of publications in the area of Remote Sensing for Drought monitoring. Since 2010, this growth has followed a linear trend. Although the number of selected papers is not too many, it does not rule out the possibility of many publications at the end of 2021.

In order to respond to the *RQ*2 questions, we looked through the chosen articles and then searched for metadata pertaining to each paper's research location and the environmental state of the area covered. Table II lists the location, the surrounding environment, and remote sensing data for observations of regions thought to be experiencing drought conditions. We also complete the dryness index that was utilized in each chosen publication.

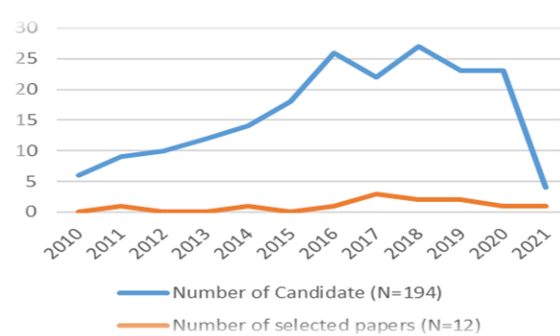


Fig. 3. Publication trends throughout the years 2010 – 2021.

TABLE II. ENVIRONMENT AND REMOTE SENSING DATA

Ref.	Environment	Data	Index	Validation
[30]	Basin area (Iran)	NOAA-AVHRR, Landsat ETM	VCI, NDVI, AVI	NDVI from Landsat+ETM 18 years (1982 - 1999)
[31]	Climate (China)	GRACE	TWSA	TWSC, SWS, SMS, GWS. Fifty-five stations in Yunnan and Guizhou (1950 – 2012)
[32]	Corn and Soybean (USA)	NDVI (MOD16A2 ET and MOD13A3); LST (MOD11A2 and MOD09A1); TRMM 3B43	LST, NDVI, NDWI, NMDI, ET, and TRMM	SPI at 54 stations (28 stations in the arid region and twenty-six stations in the humid region) from 1975 to 2012. NDMC
[33]	Sierra Nevada Forest Tree (California, USA)	MODIS Terra and Aqua observations (MCD43A4, collection 5); DEM	NDVI, EVI, NDWI	reserved validation dataset from USDA Forest Service (USFS) Aerial Detection Surveys (ADS)
[34]	Nineteen percent rice paddies and 64% forests (South Korea)	TRMM 3B43, GPM IMERG , MCD43C4 , MYD11C3 , MYD13C2	SPI and SPEI from ASOS	SPI and SPEI calculated from 61 ASOS weather stations, with 3-, 6-, 9-, and 12-month time scales.
[35]	crop yield and land cover (Korea)	AMSR-E, MODIS,TRMM	High resolution Soil Moisture Drought Index (HSMDI	SPIs for March to November (2003–2011); twenty-nine stations (1973 to 2011
[36]	agriculture (East Asia)	ESA-CCI for soil moisture; MOD11C1 for LST and NDVI; TRMM 3B42 for precipitation	PCI, TCI, VCI, SDCI, SMCI, MIDI, VSDI; Madden–Julian Oscillation (MJO) Index;	Three satellite-based drought indices SDCI, MIDI, and VSDI
[37]	three distinct climatic regions including the mountainous area (Iran)	GPCP, CMAP, CHOMPS, PERSIANN-CDR, TRMM, MERRA-2 and GLDAS-2	nonparametric-SPI; ORNESS-OWA; ORLIKE-OWA; K-nearest neighbors' algorithm (KNN)	Precipitation data for twenty-four stations (1981 - 2011); the Fars Meteorological Organization and Fars Regional Water Organization
[38]	pasture (Kenya)	MODIS and TAMSAT	NDVI, VCI, RFE, RCI, SPI	Precipitation data from TAMSAT
[39]	terrain mountains, plains, basins, valleys, and River (China)	Vegetation index product (MOD13A3), surface temperature product (MOD11A2), land use product (MCD12Q1), and TRMM, SRTM-DEM.	NDVI, EVI, LST, TCI, CI, SPEI, AWC, VSWI, Percentage of precipitation anomaly and TRMM-Z index	_Fifteen major meteorological stations and nine agricultural meteorological stations in Henan Province, (http://data.cma.cn/). _soil Available Water Capacity (AWC), (http://globalchange.bnu.edu.cn/).
[40]	Bare land, Woodland, Water, and Winter wheat (China)	MODIS NDVI, MODIS LST, Sentinel-2 NDVI, Sentinel-2 biophysical, ASTER GDEM	VCTI	Daily precipitation data in eighteen selected counties of the Guanzhong Plain
[41]	Darling River Basin (Australia)	SMAP, GLDAS, Soil attribute product, GPM, ISMN.	SWDI, SMDI	ISMN provides in situ Soil Moisture (SM) measurements of 1400 stations and thirty-five international SM networks available from 1952 to the present. https:// ismn.geo.tuwien.ac. at

Data are gathered from Table II. It turns out that the majority of studies employ MODIS satellite data products to gauge the extent of drought in different types of ecosystems. However, some studies use validation data received from data centers, while the majority of research is based on observations from ground observation stations.

On the basis of the metadata analysis of the selected papers shown in Table III, *RQ*3 may be addressed by stating that the following categories can be used to categorize the application of machine learning in remote sensing for agricultural drought monitoring. The potency of each machine learning method is shown in Table III. There is evidence that ANN (91.00 percent in [31]), BRT (93 percent in [32]), GA (95.73 percent in [37]), and RF are algorithms with accuracy values of more than 90 percent (93 percent in [32] and 96.30 percent in [33]).

Although the GA method competes with RF among other algorithms for the second-best accuracy value, in actuality, researchers frequently choose for the RF approach. This could be the result of a number of problematic situations and different facts.

Responding to *RQ*4 based on an analysis of the metadata of a few articles as indicated in Table IV. The following four categories describe how machine learning is used in remote sensing to monitor agricultural drought. First, for prediction NDVI [30], waves [31], unmeasured area [34], drought [36], np-SPI [37], vegetation condition [38], and vegetation temperature [40]. Second, detect tree death [33]. Third, it measures degree of correlation (sixteen drought factors [32], various hazard factors [39]). Fourth, down-scaling (AMSR-E and TRMM [35], SMAP-SM [41]).

TABLE III. MACHINE LEARNING ALGORITHM FOR THE DROUGHT MONITORING APPROACH

Algorithm	Ref.	Accuracy			
ANFIS	[37]	85.72			
ANN	[30], [31], [38]	79.00	**91.00**	83.00	
BRT	[32]	**93.00**			
CUBIST	[32]	60.00			
DFNN	[39]	85.60			
DT	[34]	15.92			
ERT	[34]	32.01			
ESTARFM-SVM	[40]	83.00			
GA-ORNESS-OWA	[37]	**95.73**			
GAM	[38]	86.00			
GMDH	[37]	88.21			
KKN	[37]	89.68			
M5P model tree	[37]	89.78			
MLP	[37]	90.47			
PERSIANN	[41]	80.00			
RF	[32], [33], [35], [36]	**93.00**	**96.30**	69.00	70.00
SVM	[40]	83.00			
SVR	[37]	83.57			

TABLE IV. ROLE OF MACHINE LEARNING IN REMOTE SENSING FOR AGRICULTURE DROUGHT MONITORING

Ref.	Algorithm	Role
[30]	ANN	Forecast of NDVI
[31]	ANN	Forecasting future waves
[32]	RF; BRT; Cubist	Model of the relationship
[33]	RF	Detection trees mortality
[34]	DT; RF; ERT;	Models for decision making drought in unmeasured areas
[35]	RF	Downscale AMSR-E soil moisture and TRMM precipitation
[36]	RF	Developed drought prediction models
[37]	KNN; MLP; ANFIS; M5P; GMDH; SVR; GA	Estimating np-SPI based on remotely sensed data
[38]	GAM; ANN	To predict vegetation conditions
[39]	DFNN	To construct models by considering a number of various hazard factors
[40]	SVM; ESTARFM	Developing a fused vegetation temperature condition index (VTCI)
[41]	PERSIANN	Downscaled SMAP-SM as well as GLDAS-SM against the in-situ SM

Our review of the two literature with the highest yields [33], [37] showed that the results of the CRF model analysis [33] found that baseline summer NDVI or EVI was one of the key variables to differentiate significant tree mortality. Higher ground may be denser or have higher biomass, resulting in more opportunities to obtain more water during prolonged dry seasons, and thus more resistant to stressors and mortality. The model also reveals that altitude also plays a significant role in the vulnerability of the Sierra Nevada Forest to surviving drought conditions. Altitude affects the local climate and water availability, and thus affects the distribution and drought tolerance of forest types. Vegetation index Z-scores, such as NDVI, proved to be another important variable for detecting tree mortality. The NDVI z-score in a given year represents the cumulative impact of drought on vegetation activity, while the NDWI z-score shows reduced water content. Single-dated mid-resolution imagery from MODIS and VIIRS is limited for monitoring forest health and detecting mortality, particularly at finer scales, but higher temporal frequencies, e.g., daily coverage, are a major advantage for monitoring forest health and potential forecasting capabilities across the globe (big landscape).

Meanwhile, the second-best body of research [37] demonstrates that the ORNESS-OWA fusion approach considerably enhances estimates compared to other models and that ORNESS-OWA performs better for long-term timelines than for short-term estimates. CHOMPS, GPCP, CMAP, PERSIANN-CDR, TRMM, GLDAS-2, and MERRA-2 are some examples of remote sensing precipitation products that can be used to estimate np-SPI. Three sophisticated data fusion approaches (ORNESS-OWA, ORLIKE-OWA, and KNN) are also tested against ground-based np-SPI estimations.

Even though they both employ various methodologies and observational settings, each with their own set of limitations, the two literatures produce the finest results to date. However, further research is still required to achieve results with higher spatial and temporal resolution, wider coverage, and more cost-effective operation. Future study will likely integrate satellite imagery data with field camera and aerial photography in order to provide a more comprehensive strategy that takes into account all factors, including the plant cycle. This will undoubtedly present new challenges.

V. CONCLUSION

This systematic review has provided sophisticated quantitative and qualitative analysis in this fast-growing field. Since 2010, more than 1147 journal and conference papers were found, and this trend is expected to continue in the future. A selection of the 12 most cited papers was undertaken to obtain an in-depth view of the state of the research. Drought monitoring based on remote sensing is a very active area of research with a significant impact on enhancing global sustainability and optimizing natural resources. This is supported by sensor observation technology with open access to satellite data and advances in digital machine learning computational techniques. The role of Machine learning methods has proven to be effective in prediction, detection, correlation and downscaling tasks when processing satellite imagery data.

ACKNOWLEDGMENT

This work was supported by the Infrormatics Program Study, Faculty of Computer Science, University of Singaperbangsa Karawang with Computer Science Department, Faculty of Mathematics and Natural Science, IPB Bogor University.

REFERENCES

[1] J. A. O. Reyes, D. E. Casas, J. L. Gandia, and E. F. Delfin, Drought impact on sugarcane production, vol. 35, no. June. 2021.

[2] T. Ji, G. Li, H. Yang, R. Liu, and T. He, "Comprehensive drought index as an indicator for use in drought monitoring integrating multi-source remote sensing data: A case study covering the Sichuan-Chongqing region," Int. J. Remote Sens., vol. 39, no. 3, pp. 786–809, 2018, doi: 10.1080/01431161.2017.1392635.

[3] Y. D. Giroh and A. A. Girei, "Analysis of the Factors affecting Sugarcane (Saccharum officinarum) Production under the Out growers Scheme in Numan Local Government Area Adamawa State, Nigeria," J. Educ. Pract., vol. 3, no. 8, pp. 195–201, 2012, [Online]. Available: www.iiste.org

[4] M. Â. C. C. de Carvalho et al., "Drought monitoring based on remote sensing in a grain-producing region in the cerrado–amazon transition, brazil," Water (Switzerland), vol. 12, no. 12, pp. 1–16, 2020, doi: 10.3390/w12123366.

[5] A. K. Mishra and V. P. Singh, "Drought modeling - A review," J. Hydrol., vol. 403, no. 1–2, pp. 157–175, 2011, doi: 10.1016/j.jhydrol.2011.03.049.

[6] Y. R. Li and L. T. Yang, "Sugarcane Agriculture and Sugar Industry in China," Sugar Tech, vol. 17, no. 1, pp. 1–8, 2015, doi: 10.1007/s12355-014-0342-1.

[7] S. Geerts et al., "Simulating yield response of quinoa to water availability with aquacrop," Agron. J., vol. 101, no. 3, pp. 499–508, 2009, doi: 10.2134/agronj2008.0137s.

[8] J. R. Mahan, A. W. Young, and P. Payton, "Deficit irrigation in a production setting: Canopy temperature as an adjunct to ET estimates," Irrig. Sci., vol. 30, no. 2, pp. 127–137, 2012, doi: 10.1007/s00271-011-0269-1.

[9] M. M. Chaves et al., "How plants cope with water stress in the field. Photosynthesis and growth," Ann. Bot., vol. 89, no. SPEC. ISS., pp. 907–916, 2002, doi: 10.1093/aob/mcf105.

[10] P. Steduto, T. C. Hsiao, E. Fereres, and D. Raes, Crop yield response to water. FAO IRRIGATION AND DRAINAGE PAPER 66, 2012.

[11] S. Mahajan and N. Tuteja, "Cold, salinity and drought stresses: An overview," Arch. Biochem. Biophys., vol. 444, no. 2, pp. 139–158, 2005, doi: 10.1016/j.abb.2005.10.018.

[12] S. B. Idso, R. D. Jackson, P. J. Pinter, R. J. Reginato, and J. L. Hatfield, "Normalizing the stress-degree-day parameter for environmental variability," Agric. Meteorol., vol. 24, no. C, pp. 45–55, 1981, doi: 10.1016/0002-1571(81)90032-7.

[13] H. G. Jones and P. Schofield, "Principles of thermal remote sensing.," Heat Capacit. Mapp. Mission Anthol., pp. 7–14, 1982.

[14] E.-D. Schulze, E. Beck, and K. Muller-Hohenstein, Plant Ecology. Springer International Publishing, 2005. doi: 10.1016/j.ecoinf.2020.101136.

[15] H. G. Jones, "Remote detection of crop water 'stress' and distinguishing it from other stresses," Acta Hortic., vol. 922, pp. 23–34, 2011, doi: 10.17660/ActaHortic.2011.922.2.

[16] S. O. Ihuoma and C. A. Madramootoo, "Recent advances in crop water stress detection," Comput. Electron. Agric., vol. 141, pp. 267–275, 2017, doi: 10.1016/j.compag.2017.07.026.

[17] S. Park, D. Ryu, S. Fuentes, H. Chung, M. O'connell, and J. Kim, "Dependence of cwsi-based plant water stress estimation with diurnal acquisition times in a nectarine orchard," Remote Sens., vol. 13, no. 14, pp. 1–13, 2021, doi: 10.3390/rs13142775.

[18] Z. Zhu and C. E. Woodcock, "Continuous change detection and classification of land cover using all available Landsat data," Remote Sens. Environ., vol. 144, pp. 152–171, 2014, doi: 10.1016/j.rse.2014.01.011.

[19] S. E. Franklin, O. S. Ahmed, M. A. Wulder, J. C. White, T. Hermosilla, and N. C. Coops, "Large Area Mapping of Annual Land Cover Dynamics Using Multitemporal Change Detection and Classification of Landsat Time Series Data," Can. J. Remote Sens., vol. 41, no. 4, pp. 293–314, 2015, doi: 10.1080/07038992.2015.1089401.

[20] C. D. Man, T. T. Nguyen, H. Q. Bui, K. Lasko, and T. N. T. Nguyen, "Improvement of land-cover classification over frequently cloud-covered areas using landsat 8 time-series composites and an ensemble of supervised classifiers," Int. J. Remote Sens., vol. 39, no. 4, pp. 1243–1255, 2018, doi: 10.1080/01431161.2017.1399477.

[21] J. L. Peters, A. J. Sutton, D. R. Jones, K. R. Abrams, and L. Rushton, "The contribution of systematic review and meta-analysis methods to human health risk assessment: Neurobehavioral effects of manganese," Hum. Ecol. Risk Assess., vol. 14, no. 6, pp. 1250–1272, 2008, doi: 10.1080/10807030802494592.

[22] C. N. Cook, H. P. Possingham, and R. A. Fuller, "Contribution of systematic reviews to management decisions," Conserv. Biol., vol. 27, no. 5, pp. 902–915, 2013, doi: 10.1111/cobi.12114.

[23] E. C. O'Hagan, S. Matalon, and L. A. Riesenberg, "Systematic reviews of the literature: A better way of addressing basic science controversies," Am. J. Physiol. - Lung Cell. Mol. Physiol., vol. 314, no. 3, pp. L439–L442, 2018, doi: 10.1152/ajplung.00544.2017.

[24] W. Canon-Montanez and A. L. Rodriguez-Acelas, "Contribution of Systematic Reviews and Meta-analyses to Nursing Education, Research, and Practice," Aquichan, vol. 21, no. 4, p. e2143, 2021, doi: 10.5294/aqui.2021.21.4.3.

[25] D. Tuia and G. Camps-Valls, "RECENT ADVANCES IN REMOTE SENSING IMAGE PROCESSING," IEEE Int. Conf. Image Process., vol. 16, pp. 3661–3664, 2009, doi: 10.1117/12.913262.

[26] H. Z. M. Shafri, "Machine Learning in Hyperspectral and Multispectral Remote Sensing Data Analysis," pp. 3–9, 2017, doi: 10.1142/9789813206823_0001.

[27] P. Martino, S. Vasileios, and J. A. Maria, Benchmarking of the Symbolic Machine Learning classifier with state of the art image classification methods Application to remote sensing imagery. 2015. doi: 10.2788/638672.

[28] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: an applied review," Int. J. Remote Sens., vol. 39, no. 9, pp. 2784–2817, 2018, doi: 10.1080/01431161.2018.1433343.

[29] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, 2021, doi: 10.1136/bmj.n71.

[30] A. F. Marj and A. M. J. Meijerink, "Agricultural drought forecasting using satellite images, climate indices and artificial neural network," Int. J. Remote Sens., vol. 32, no. 24, pp. 9707–9719, 2011, doi: 10.1080/01431161.2011.575896.

[31] D. Long et al., "Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data," Remote Sens. Environ., vol. 155, pp. 145–160, 2014, doi: 10.1016/j.rse.2014.08.006.

[32] S. Park, J. Im, E. Jang, and J. Rhee, "Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions," Agric. For. Meteorol., vol. 216, pp. 157–169, 2016, doi: 10.1016/j.agrformet.2015.10.011.

[33] S. Byer and Y. Jin, "Detecting drought-induced tree mortality in Sierra Nevada forests with time series of satellite data," Remote Sens., vol. 9, no. 9, pp. 14–17, 2017, doi: 10.3390/rs9090929.

[34] J. Rhee and J. Im, "Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data," Agric. For. Meteorol., vol. 237–238, pp. 105–122, 2017, doi: 10.1016/j.agrformet.2017.02.011.

[35] S. Park, J. Im, S. Park, and J. Rhee, "Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula," Agric. For. Meteorol., vol. 237–238, pp. 257–269, 2017, doi: 10.1016/j.agrformet.2017.02.022.

[36] S. Park, E. Seo, D. Kang, J. Im, and M. I. Lee, "Prediction of drought on pentad scale using remote sensing data and MJO index through random forest over East Asia," Remote Sens., vol. 10, no. 11, pp. 1–18, 2018, doi: 10.3390/rs10111811.

[37] M. R. Alizadeh and M. R. Nikoo, "A fusion-based methodology for meteorological drought estimation using remote sensing data," Remote Sens. Environ., vol. 211, no. March, pp. 229–247, 2018, doi: 10.1016/j.rse.2018.04.001.

[38] C. Adede, R. Oboko, P. W. Wagacha, and C. Atzberger, "A mixed model approach to vegetation condition prediction using Artificial

Neural Networks (ANN): Case of Kenya's operational drought monitoring," Remote Sens., vol. 11, no. 9, pp. 1–18, 2019, doi: 10.3390/rs11091099.

[39] R. Shen, A. Huang, B. Li, and J. Guo, "Construction of a drought monitoring model using deep learning based on multi-source remote sensing data," Int. J. Appl. Earth Obs. Geoinf., vol. 79, no. 219, pp. 48–57, 2019, doi: 10.1016/j.jag.2019.03.006.

[40] X. Zhou, P. Wang, K. Tansey, S. Zhang, H. Li, and L. Wang, "Developing a fused vegetation temperature condition index for drought monitoring at field scales using Sentinel-2 and MODIS imagery," Comput. Electron. Agric., vol. 168, no. 17, p. 105144, 2020, doi: 10.1016/j.compag.2019.105144.

[41] B. Fang, P. Kansara, C. Dandridge, and V. Lakshmi, "Drought monitoring using high spatial resolution soil moisture data over Australia in 2015–2019," J. Hydrol., vol. 594, no. July 2020, p. 125960, 2021, doi: 10.1016/j.jhydrol.2021.125960.

Research on Intellectual Dichotomiser 3 Decision Tree Algorithm Model for Financial Analysis of Colleges and Universities

Sujuan Guo
Smart Finance Industry College, Shandong Institute of Commerce and Technology, Jinan, 250103, China

***Abstract*—The rapid development of college information construction has promoted the processing and analysis of a large number of data in the college system. Decision tree algorithm is often used in the field of financial data analysis, but it has a bias in the selection of attributes. Aiming at the defects of the decision tree algorithm in attribute selection, ID3 algorithm in the decision tree algorithm is selected for weighted improvement, and it is optimized based on Synthetic Minority Oversampling Technique (SMOTE) algorithm and Bagging algorithm to balance the positive and negative data of its training samples, thus obtaining the DSB-ID3 financial analysis model. Using this model to analyze the financial data of a university, its G value and F value are both about 0.78, the recognition accuracy rate for normal samples is 0.7345, and the total recognition accuracy rate is 0.7893, which are the highest among the four models. Compared with other models, model designed in this study has significantly improved classification performance, and its distribution is the most centralized, showing superior stability. The experimental results show that the classification effect of model designed in this study is the best, and it shows superior accuracy and stability in the analysis of financial data. Its superior classification performance shows the potential of decision tree algorithm optimization and the feasibility and necessity of improving it. From the experimental data, it can be seen that the service life and parameters of the model designed in this study are obviously better than those commonly used in the financial analysis industry of colleges and universities. It can be seen from the overall analysis that this model provides a practical reference for the application of decision tree optimization in college financial analysis, and greatly improves the accuracy of financial system data analysis.**

Keywords*—Financial analysis; ID3; Decision tree; model; colleges and universities; machine learning

I. Introduction

With the gradual deepening of higher education reform and the introduction of higher education into the market as an industry, the financial management of colleges and universities has also changed. The financial analysis and management work in colleges and universities has gradually shifted from accounting to analytical, but the existing financial analysis and management system cannot meet the requirements of precision and real-time [1-2]. With the rapid development of information technology, the scale of financial informatization in colleges and universities is also expanding. The college information system can easily collect relevant financial data [3-4]. However, massive data brings management difficulties to data managers. Lack of data mining means and tools will result in a huge waste of data resources, and financial analysis cannot be carried out in a normal and orderly manner. Therefore, computer technology is integrated into financial analysis to effectively analyze and manage data, so as to improve the application level of financial data, optimize the financial structure of the school, and promote the healthy development of the financial management system [5-6]. Decision tree algorithm is often used in the field of data analysis. Research will also establish a university financial analysis model based on decision tree. However, the decision tree has the defect of attribute selection, so this research will select ID3 (intelligent dichotomy 3) algorithm in the decision tree to improve, and based on DSRR (Differential Sampling Rate Resampling, DSRR) technology, SMOTE (Synthetic Minor Over Sampling), so as to improve its noise resistance and balance its attribute selection, which is also the main contribution of this research.

II. Related Work

Scholars such as Febriantono MA proposed a C5.0 deciMany scholars have done a lot of research on decision tree algorithm in financial analysis and other fields. Luo X et al. proposed an initialization method of dendrite neuron model (DNM) based on decision tree to trim those neurons that contribute less to the network output, aiming at the problem of insufficient calculation accuracy of decision tree model. This method can reduce the number of dendrites in DNM and improve the training efficiency, and can select appropriate initialization weight and neuron threshold, thus enhancing the classification accuracy of the algorithm. Experiments show that this method is significantly superior to the original DNM model with the lowest complexity and the fastest training speed, without loss of accuracy [7]. Podhorska I et al. believed that the non-linear data calculation ability of the decision tree model was insufficient, so they combined BP neural network with ID3 decision tree model to build an improved decision tree model, and applied it to predict financial distress. They also established a prediction model for enterprise financial distress, mined and analyzed enterprise financial data, and used it to predict enterprise financial development. The test results showed that the financial data prediction accuracy of the improved algorithm is significantly higher than that of the decision tree model before improvement [8]. Sion tree algorithm based on cost sensitivity, established a model, learned to use the Metacost Method (MM) to get the minimum cost model, and used it to solve multi-class imbalance problems. The performance of the algorithm is better than the C4.5 and ID3 algorithms [9]. Maulana MF et al. used two model classifications for the evaluation of students' learning time, namely Logistic Model Tree (LMT) and Decision Tree J48 Algorithm (DT J48), using the model to predict the effect of students' learning time and find out the influencing factors. The prediction accuracy of LMT is 71% higher than that of DT J48 [10]. Xia Y and other scholars proposed a SurvXGBoost model based on the Gradient Boosting Decision Tree (GBDT) method, and applied it to the consumer loan dataset. The performance is excellent [11].

Researchers such as Balta M proposed a three-level fuzzy decision tree model for Vehicle Ad-Hoc Networks (VANET) system based on Software Defined Networks (SDN). The physical structure of urban intersections often lead to traffic jams, delays, and accidents. Previous studies usually choose the VANET architecture, which can communicate between vehicles and road edge devices. SDN has also become a common research method in this domain, it can solve the performance, programming, scalability and security problems in traditional networks. The proposed model utilizes the architectural recommendations of both SDN and VANET network paradigms to optimize traffic management problems based on three-level fuzzy decision tree algorithm classification [12]. Qiu T et al. proposed a Directed Edge Weight Prediction (DEWP) model based on decision tree integration. It extends the local similarity index to the Directed Weighted Network (DWN), and extracts the similarity index to construct a mixed regression model [13]. In order to explain the underpricing event of Initial Public Offering (IPO) during the epidemic, Keuangan J and other scholars built a decision tree algorithm model and tested the latest 45 action data using more than 100 previous action data. The model can interpret IPO performance according to specific classification scope. [14]. Researchers such as L Xue proposed a Privacy-Preserving Decision Tree Classification (PDTC). First, customize an encryption primitive and a secret sharing technology to design a new secure bilateral comparison protocol, in which the digital inputs of each party are compared privately. Then, based on the protocol, PDTC is constructed using decision tree structure. The customer's input and the service provider's model parameters are hidden from the counterparty, and the classification results are only displayed to the customer. The results show that PDTC achieves ideal safety [15]. Ziweritin S and other scholars established the K-nearest neighbor and decision tree model to forecast the numeric value in the one pound table under a given interest rate and life. Cross validation was performed using the R-squared test to detect fit and summarize model function on the test data set. The model is robust and competitive, and has reference value for financial analysis and its application [16].

In addition, in terms of college financial management, Natawibawa IWY team applied multiple linear regression method to analyze the financial management system of state universities, and the research results show that the availability of financial reports has a significant impact on the transparency of college financial management [17]. The Min Du team has carried out informatization reform on the traditional financial management model of colleges and universities, and the research results show that with the help of informatization, the financial management level of colleges and universities has been effectively improved [18]. The Na Sun team started from the financial position i and staffing of colleges and universities, analyzed the current situation of colleges and universities emphasizing accounting over management, and proposed strategic management and post balance strategy [19]. It can be seen that in the application of decision tree algorithm, there have been some precedents in applying decision tree algorithm to the economic and financial fields. At the same time, in the financial management of colleges and universities, there is still a lack of appropriate information technology to provide a technical basis for the development of financial management of colleges and universities. Therefore, in order to avoid the waste of financial data resources, the research provides data mining means and tools so that the financial analysis work of colleges and universities can be carried out normally and orderly. The research is oriented to college financial analysis, and the ID3 algorithm is weighted. Based on the DSRR technology, SMOTE algorithm and Bagging algorithm, ID3 is optimized, and the DSB-ID3 model is established to overcome the defect that the ID3 algorithm tends to select attributes with more values but not necessarily optimal. This technology can monitor the financial situation of colleges and universities in real time from a dynamic perspective, carry out meticulous financial management according to the financial characteristics of colleges and universities, and improve the efficiency and quality of financial management of colleges and universities.

III. ID3 Decision Tree Optimization Algorithm Model for University Financial Analysis

A. *Improved ID3 Algorithm based on Attribute Selection Weighting*

The ID3 algorithm is a decision tree algorithm based on information entropy. It introduces information theory, uses information entropy as a criterion for selecting test attributes, divides the training set, constructs a decision tree to predict according to the test attributes, and divides the instance space [20-21]. To build a decision tree, first measure the attributes of each node using the maximum information gain from top to bottom, and obtain the corresponding attribute values, which are used to segment the target object. This step is repeated until all objects in the node are judged to be of the same type according to the classification criteria. The features with more attribute values in the decision tree have a greater impact on the calculation of information gain, however, features with more values are uncertain optimal attributes, and the algorithm has poor noise resistance, making it difficult to control the positive and negative examples in data set, so the research improves the algorithm by weighting. The main principle of the information theory introduced by the algorithm is to regard the communication process as a process of transmitting information in an environment of random interference. In order to obtain mutual information, the posterior entropy is first calculated. When the channel receiver receives the output symbol $V_j = \{V_1, V_2, \cdots V_m\}$, the input symbol is calculated.

U The information measure of $P\left(U\left|V_j\right.\right)$ is as formula (1).

$$P\left(U\left|V_j\right.\right)=\sum_j P\left(V_j\right)\sum_j P\left(U\left|V_j\right.\right)\lg\frac{1}{P\left(U\left|V_j\right.\right)} \tag{1}$$

Before receiving the symbol set , V the average uncertainty of the $D\left(U\right)$ input symbol set is expressed as U , after receiving the symbol set , V the average uncertainty of the $D\left(U\left|V\right.\right)$ input symbol set is expressed as U , the relationship between the two is shown in Eq.

$$I\left(U,V\right)=D\left(U\right)-D\left(U\left|V\right.\right) \tag{2}$$

In formula (2), $I\left(U,V\right)$ is U the average mutual information between and , which represents V the amount of information V about the obtained after receiving the symbol set U . In order to avoid the ID3 algorithm to select attributes with many values but not necessarily optimal as test attributes, the research introduces risk weights. Given the user's risk weights for uncertain knowledge $\max\left(-a_1,-a_2,\cdots,-a_i,\cdots-a_n\right)\le b\le 0$, a_i it represents the conditional probability of attributes, which b is a dynamic variable whose value Depends on the conditional probability of the attribute obtained each loop. If there are several conditional probabilities in the calculation process $a_1,a_i,\cdots,a_n$, it is $-b$ equal to the minimum value among them. Introducing risk weights into the algorithm, the calculation of conditional entropy is as formula (3).

$$D_b\left(U\left|V\right.\right)=\sum_{j=1}^{n}\left(P\left(V_j\right)+b\right)\sum_{i=1}^{n}P\left(U_i\left|V_j\right.\right)\lg\frac{1}{P\left(U_i\left|V_j\right.\right)} \tag{3}$$

The corresponding mutual information calculation is shown in formula (4).

$$I_b\left(U,V\right)=D\left(U\right)-D_b\left(U\left|V\right.\right) \tag{4}$$

Equations (3) and (4) are used as criteria for selecting test attributes when constructing a decision tree. The research uses the AHP method to assign weights, compares and judges the importance of indicators at the same level, and constructs a judgment matrix to represent the judgment value of their relative importance. In the measurement process, the scale of the ninth is leaded into form a judgment matrix G . Each element in the g_{ij} matrix represents G the comparison value of the relative importance of the g_{ij} row index C_i to each column index, indicating the relative importance of C_j a certain index and another index, and its scale is shown in Table I.

TABLE I. RELATIVE IMPORTANCE BETWEEN TWO INDICATORS

Ratio of index 1 to index 2	extremely important	Important	Slightly important	Identical	Slightly unimportant	Unimportant	Extremely unimportant
Indicator 1 Evaluation	7	5	3	1	1/3	1/5	1/7
Intermediate value	Take 6, 4, 2, 1/2, 1/4, 1/6 as the intermediate value						

The principle of the AHP method used in the research is to judge the matrix, obtain the matrix sorting vector according to the sorting rules, and use the obtained results to calculate the weight coefficients of each index. Find the product of the elements of each row of the judgment matrix as in formula (5).

$$P_i=\prod_{j=1}^{n}c_{ij},i=1,2,\cdots,n \tag{5}$$

Calculate the square root value of each row P_i as n formula (6).

$$\overline{p}_i=\sqrt[n]{P_i},i=1,2,\cdots,n \tag{6}$$

Formula (6) is n the order of the matrix. The vector is normalized as in formula (7).

$$p_i=\frac{\overline{p}_i}{\sum_{j=1}^{n}\overline{p}_j} \tag{7}$$

p_j is the weight coefficient value of each indicator of demand. The maximum eigenroot of the judgment matrix is calculated as in Equation (8).

$$\lambda_{\max}=\sum_{i=1}^{n}\frac{\left(Gw\right)_i}{nw_i} \tag{8}$$

Compared with other methods for determining the weight coefficient of indicators, the AHP method can pass the consistency test and maintain the consistency of thinking logic. When judging the importance of indicators, when there are more than 3 indicators, the coordination among the judgments is the logical consistency of thinking. Let the maximum eigenroot of the $\lambda_{\max}$ judgment matrix be G , when the matrix G passes the consistency, $\lambda_{\max} = n$, otherwise $\lambda_{\max} \neq n$. To test G the consistency, the consistency index of the judgment matrix is established as formula (9).

$$CI = \frac{\lambda_{\max} - n}{n-1} \quad (9)$$

When the judgment matrix passes the consistency check, $CI = 0$. The average randomness index of the judgment matrix is introduced RI to measure whether the matrix passes the consistency test. For the judgment matrix whose order is between 1 and 9, its RI values are shown in Table II.

TABLE II. AVERAGE RANDOM INDEX RI OF JUDGMENT MATRIX

Order	1	2	3	4	5	6	7
RI	0	0	0.65	0.98	1.15	1.34	1.47

The consistency ratio of the judgment matrix is CR shown in formula (10).

$$CR = \frac{CI}{RI} \quad (10)$$

$CR < 0.1$ At that time, the matrix is consistent, and the relative importance calculated using this matrix is reasonable. When this condition is not met, the matrix needs to be modified until it passes the consistency check. The comprehensive formula of multiple evaluation opinions is shown in formula (11).

$$Y_j = \frac{\sum_{i=1}^{N} \left(W_{ij} \times S_{ij} \right)}{\sum_{i=1}^{N} S_{ij}} \quad (11)$$

In formula (11), N represents the number of evaluation opinions, generally $N \geq 5$. Y_j Indicates N the calculation weight value of the W_{ij} evaluation opinion for the first indicator, indicates i the j calculation j weight value of the evaluation opinion for the first indicator, and indicates the familiarity coefficient of the S_{ij} evaluation opinion i to the first j indicator, generally a dynamic value between 0 and 4, and the evaluation The familiarity of the opinion i to the first j indicator is proportional, and the same evaluation opinion has different familiarity coefficients for different evaluation indicators. After the weights are set, the evaluation results are dynamically generated by expert evaluation, so that the results can change with the environmental conditions.

B. DSB-ID3 Model Optimized by Three Algorithms

When the ID3 algorithm is used alone, there is a defect in the classifier itself. When selecting the relevant classification attributes, it will not select the optimal attributes, but select attributes with more values [22-23]. On the basis of the weighted improvement of the algorithm, the random oversampling technology (Over Sample, OS) is used to optimize the algorithm, so that the number of samples in the two data sets is basically equal, so that the sample balance can be better obtained. The classification performance can reduce the deviation of sample attributes to a certain extent. The specific process of the OS-ID3 model is shown in Fig. 1.

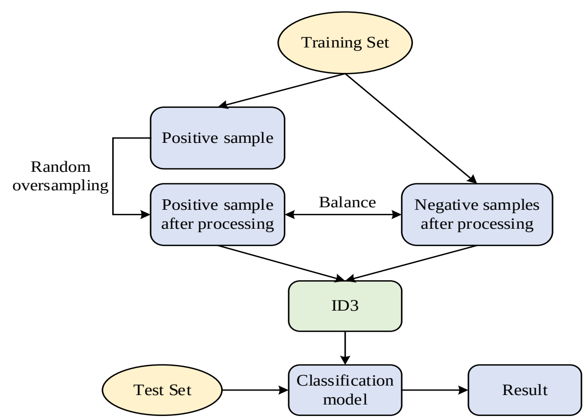


Fig. 1. Specific process of OS-ID3 model

The algorithm has a simple structure and fast calculation speed, which changes the state of imbalance between classes and makes up for the classification defects of the ID3 algorithm to a certain extent. Overfitting phenomenon. The algorithm is optimized using the Over-Under Sample (OUS) technique. When processing samples, the oversampling technique is used to add the quantity of positive samples, and the undersampling technique is used to reduce the quantity of negative samples, so as to balance the two types of samples. The specific process of the OUS-ID3 model is shown in Fig. 2.

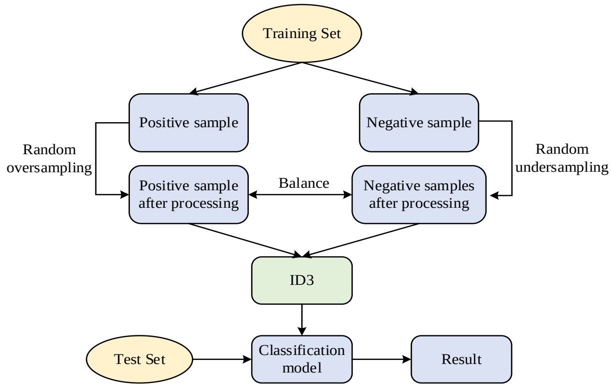


Fig. 2. Specific process of OUS-ID3 model

The structure and steps of OUS-ID3 algorithm are simple, and the calculation speed is fast. To some extent, it avoids the problem that too many positive samples lead to over fitting of the classifier. However, when the quantity of negative samples decreases, some important information will be ignored and the function of the classifier will be degraded. In view of the shortcomings of simple sampling technology, DSRR technology gives different sampling ratios to different classifiers according to the algorithm set internally in the sampling process, so that the number of sample data trained by each classifier is different. Statistically calculate the number of positive samples P_S and negative samples N_S, and calculate the difference between the two types of samples as shown in formula (12).

$$Value = N_S - P_S \tag{12}$$

Assuming that the number of classifiers is M, the sampling coefficient is calculated as shown in formula (13).

$$Coefficient = \frac{Value}{M} \tag{13}$$

Assuming i the sampling rate R of the the classifier, the number of positive samples of the the classifier is calculated i as shown in formula (14).

$$P_{S_i} = P_S + Coefficient \times R_i \tag{14}$$

In formula (14) $1 \le i \le M$. Aiming at the defects of random sampling, the SMOTE algorithm can effectively improve it. The principle of this algorithm is to assume that two positive samples close to each other are also positive samples between them. If the sampling rate is α, each positive class sample point is x_i, its positive class neighbor point is $y_{ij}\left(j=1,2,\cdots,l\right)$, $l = x * \alpha$ is the number of positive class neighbor points, and find all positive class neighbor points. A new negative class sample is generated q_j as shown in equation (15).

$$q_j = x_i + rand\left(0,1\right) \times \left(y_{ij} - x_i\right) \tag{15}$$

formula (15), it means that the $rand\left(0,1\right)$ value is randomly selected q_i in the interval , which will be $\left(0,1\right)$ mixed with the original positive sample to form a minority class sample. According to the difference in the quantity of samples, the balance rate of the two types of samples can be obtained. The sampling principle of the algorithm is shown in Fig. 3.

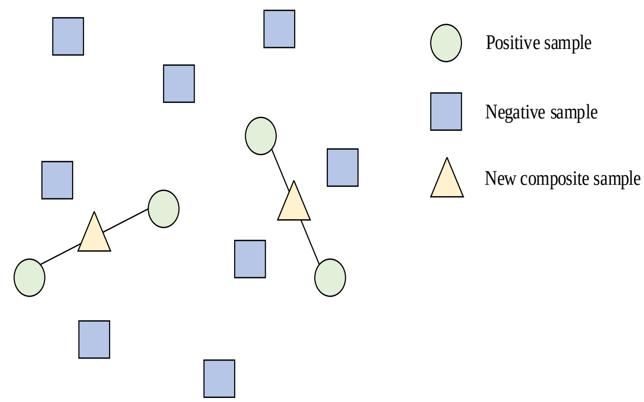


Fig. 3. Sampling principle of SMOTE algorithm

On the premise of retaining the original information and distribution of the minority class, the SMOTE algorithm constructs new samples without repetition, and can overcome the shortcomings of simple oversampling while completing the same task. The Bagging algorithm draws on the replacement sampling method in probability theory. The principle is that after each extraction, the extracted samples are put back into the original data set, so the sample sets extracted each time are the same size, the extracted samples are trained to get the same classifier with different parameters, and the prediction results are obtained. When building a model, small changes in the training samples will lead to significant changes in the classification model, resulting in large differences in the classification results. By training the ID3 algorithm through the Bagging algorithm, multiple different classifiers will be obtained. Different classifiers get different results because the built-in parameters of the ID3 algorithm have changed, resulting in changes in the internal operation of the model and changes in the results obtained. But for multiple classifiers, the majority weighted voting method does not affect the final result due to the result of a single classifier, which strengthens the generalization performance of the ensemble classifier. Assume the training set $T = \left\{\left(x_i, y_i\right), \left(x_2, y_2\right), \cdots, \left(x_n, y_n\right)\right\}$, use the single-classification algorithm to perform T operations on the training set z, extract different training subsets from it, and use it as a sample for each training to obtain T z a classifier with different results. The test set is classified, the z results are weighted and voted, and the class that appears multiple times in the classification result is used as the final result.

Randomly select T' a subset of training samples, put them into a given single-classification algorithm for training, and obtain a model h_t, and loop through the steps z to obtain a set $\{h_1, h_2, \cdots, h_t\}$. Each model h_t gets a classifier, and the unknown samples are classified to get a result, and the category to which they belong is judged according to the principle of getting the most votes. Bagging algorithm overcomes the instability of classifier classification prediction to a large extent, and can reduce a lot of time and cost in the operation process.

The ID3 algorithm model based on DSRR, SMOTE and Bagging uses DSRR technology, SMOTE sampling method and Bagging algorithm combined with ID3 decision tree algorithm to form an integrated classification model based on ID3 classifier. The DSRR technique and the SMOTE sampling method are effectively combined to generate minority class samples according to the sampling rate. The SMOTE sampling method has superior sampling characteristics, and no overfitting occurs during the sampling process, which increases the number of minority class samples and narrows the gap with the majority class samples. On the basis of retaining the original information of the samples, DSRR technology uses the changed sampling rate to establish a sub-classifier with the difference in the number of samples of the majority class and the minority class. The Bagging algorithm combines DSRR technology to generate majority class samples that match the minority class samples. The combination of the two highlights the technical superiority. After removing the weak classifiers, they are integrated into DSB-ID3 (DSRR, SMOTE and Bagging ID3) combined classifier to improve the accuracy of the algorithm based on unbalanced characteristics. The DSB-ID3 model is shown in Fig. 4.

In Fig. 4, the model determines the sampling rate of two samples through DSRR technology, and uses the SMOTE algorithm in the training set to generate several non-repetitive samples according to the determined sampling rate. From the generated samples, DSRR technology is used to extract n samples with different numbers $[X_1, X_2, \cdots X_n]$, and they are combined with the samples extracted by the Bagging algorithm according to the sampling rate X_f to form n new sample training sets. If the sampling rate is the same, the quantity of two samples is the same, and if the sampling rate is different, the number of two samples is different. The ID3 algorithm is used to train n sample training sets, and n ID3 sub-classifiers with different numbers are obtained $[K_1, K_2, \cdots, K_n]$. According to the training results of each sub-classifier, extract 20 classifiers with the best classification effect from the results, obtain $20n$ sub-classifiers, calculate the mean of the results obtained by the $20n$ sub-classifiers, and compare the results of the sub-classifiers with the obtained mean. Compare, if it is greater than the mean, keep the sub-classifier, otherwise remove the sub-classifier.

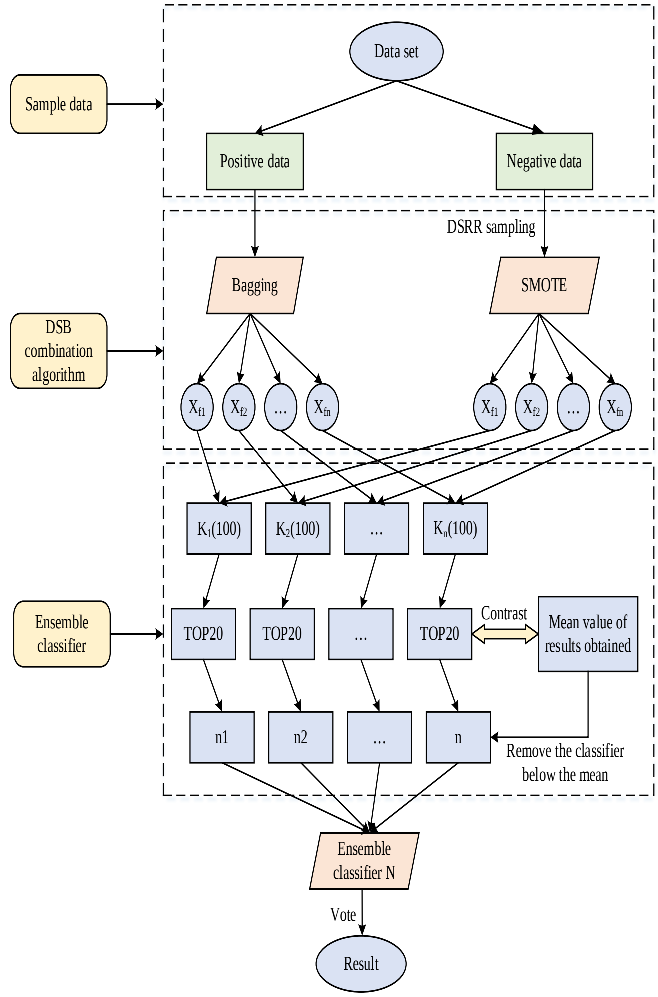
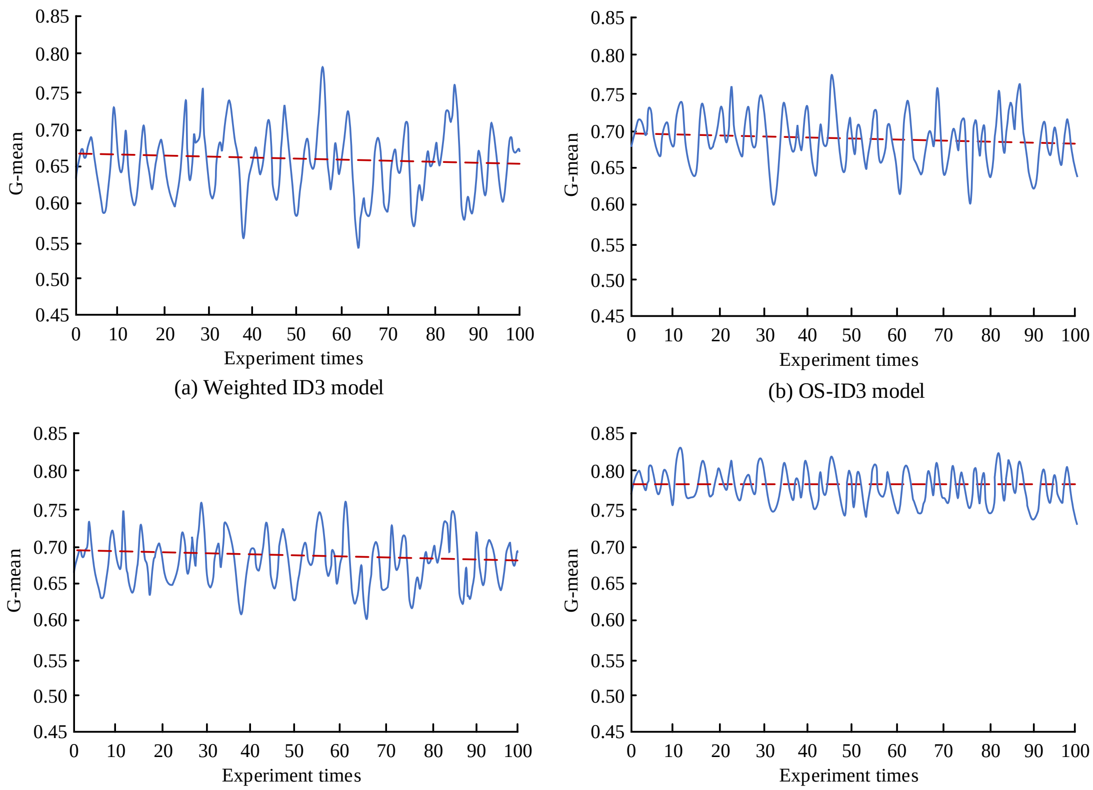


Fig. 4. DSB-ID3 model

The classifiers with different numbers are obtained according to different sampling rates $(n_1, n_2, \cdots, n)$, the remaining classifiers are integrated to obtain the combined classifier N, and different weights are given according to the quantity of remaining classifiers under different sampling rates, and finally the results are obtained by voting. The DSB-ID3 model overcomes the shortcomings of simple sampling technology, and assigns different sampling rates to different classifiers according to the internal algorithm, which makes the number of sample data of the classifiers different. The model also takes advantage of SMOTE sampling, increases the scale of sample data and avoids repeated results, and uses Bagging technology to significantly improve stability. The obtained negative samples are independent of each other and run fast.

IV. Experimental Analysis of DSB-ID3 Model for Financial Analysis of Colleges and Universities

The experimental data comes from the financial analysis report of 2021 of a university, which includes the daily operating expenses and income of the university, the current assets and liabilities of the university, and the change information of the cash flow of the university. This research mainly selects the student payment information in this report for experimental analysis. Student payment information includes student basic information data, student payment information data, receivables information data and fee reduction and exemption information data. The dimensions of defining information include the dimension of student information, the dimension of academic year and semester, the dimension of payable fees, and the dimension of reduced fees. The study used G-mean and F-value as the evaluation index of classification effect. The calculation of the G value is measured by the geometric mean of the classification accuracy of positive and negative samples. In order to reach the maximum value, the classification accuracy of positive and negative samples needs to be larger. The F-value is a combination of precision and recall, where a parameter is introduced to ω reflect the relative importance of precision and recall. The experiment is realized by Matlab and decision tree tools. The DSB-ID3 model and other models are run 100 times in the same environment, and each training set and test set are randomly divided and samples are randomly selected. The G value results of each model are shown in Fig. 5.



(a) Weighted ID3 model

(b) OS-ID3 model

(c) OUS-ID3 model

(d) DSB-ID3 model

Fig. 5. G-mean results of each model

As shown in Fig. 5, with the increase of experimental time, the G value of each model fluctuates repeatedly, but the G value of Weighted ID3 model, OS-ID3 model, and OUS-ID3 model, except for DSB-ID3 model, shows a general downward trend, and their linear regression slopes are 1.74×10^{-4}、1.32×10^{-4} and 1.25×10^{-4} However, the G-value curve of DSB-ID3 model has no overall downward trend. In Fig. 5, the G value of the weighted ID3 model is around 0.66, which is the lowest among all models, and its distribution is also the most dispersed. The G value of the OS-ID3 model is about 0.69, which is improved compared to the weighted ID3 model, but its distribution is also relatively scattered, which is not much different from the weighted ID3 model. The line graph of the OUS-ID3 model is more concentrated than the previous two models, but its G value is basically not improved. The G value of the DSB-ID3 model is about 0.78, which is a significant improvement compared to other models, and the polyline distribution is concentrated. Among the four models, the weighted ID3 model has the worst classification effect, the OS-ID3 model and the OUS-ID3 model have similar classification effects, and the DSB-ID3 model has the best classification effect, and the classification results are accurate and stable. The F value results of each model are shown in Fig. 6.

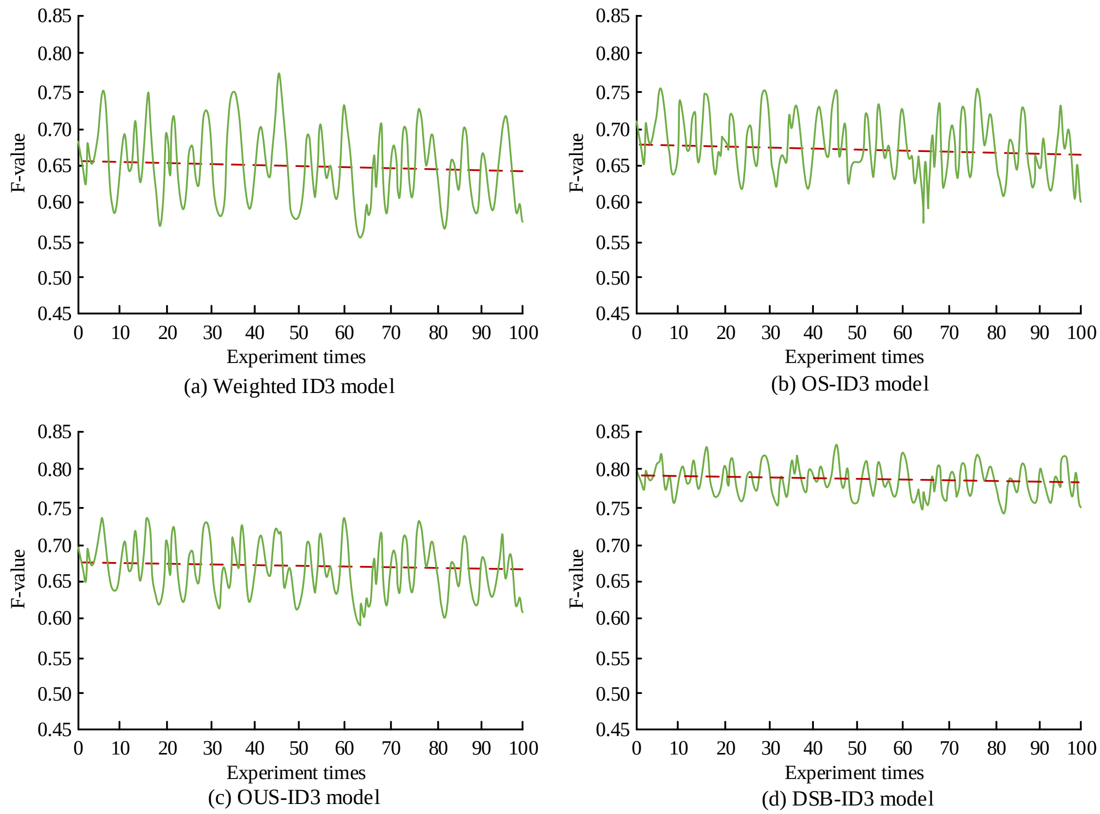


Fig. 6. F-value results of each model

As shown in Fig. 6, the F value of each model fluctuates repeatedly with the increase of experimental time, and the change rule of each model is generally similar to that in Fig. 5. Specifically, the F values of Weighted ID3 model, OS-ID3 model, OUS-ID3 model and DSB-ID3 model show a downward trend in general, and their linear regression slopes are 0.82×10^{-4}、0.94×10^{-4} and $0.71 \times 10\text{-}^{4}$、$0.35 \times 10^{-4}$, it can be seen that the overall descending speed of DSB-ID3 model is the slowest. In Fig. 6, the F value of the weighted ID3 algorithm is about 0.65, and its distribution is as scattered as the G value. Due to the improvement of the balance of sampling rate, the OS-ID3 model has a slight improvement in the F value, which fluctuates around 0.67. Compared with the OS-ID3 model, the F value of the OUS-ID3 model is basically not improved, still fluctuating around 0.67, and the distribution is relatively scattered. The F value of the DBS-ID3 model is around 0.78, which is significantly improved compared to other models, and its distribution is also the most concentrated, showing excellent stability. Among the four models, the weighted ID3 model has the worst classification effect, the OS-ID3 model and the OUS-ID3 model have similar classification effects, and the DSB-ID3 model has the best classification effect. The classification effect of the combination on the data is significantly improved. The mean values of G and F values for each model are shown in Fig. 7.

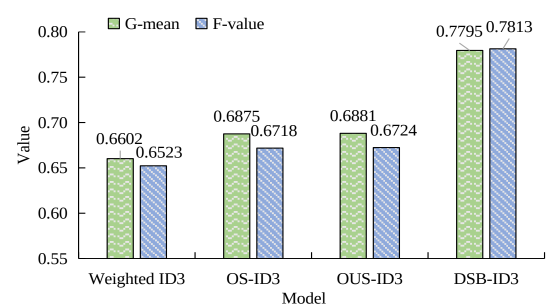


Fig. 7. Mean value of G-mean and F-value of each model

In Fig. 7, the mean G value of the weighted ID3 model is 0.6602, and the mean F value is 0.6523; the mean G value of the OS-ID3 model is 0.6875, and the mean F value is 0.6718; the mean G value of the OUS-ID3 model is 0.6881, and the mean F Value The mean value of the value is 0.6724; the mean value of the G value of the DSB-ID3 model is 0.7795, and the mean value of the F value is 0.7813. The G value and F value of each model are not much different. With the further improvement and optimization of ID3, the G value and F value of the model continue to increase, and its classification effect is also significantly improved. The G value and F value of the DSB-ID3 model largest, and its classification effect is also the best. The recognition accuracy of each model to the data is shown in Fig. 8.

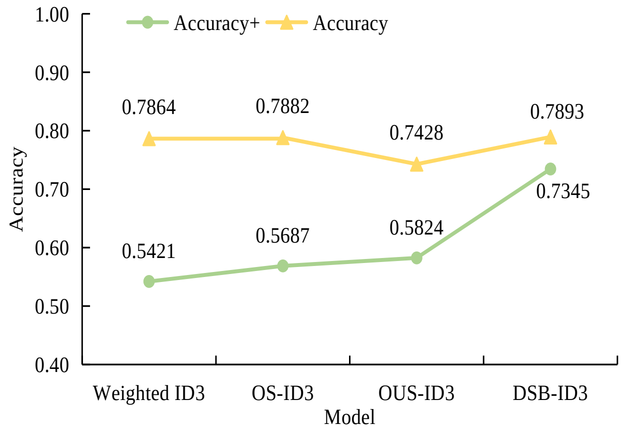


Fig. 8. Data recognition accuracy of each model

In Fig. 8, Accuracy+ represents the recognition accuracy of the model for positive data, and Accuracy represents the recognition accuracy of the model for all data. The accuracy of the weighted ID3 model for the positive data is 0.5421, and the total accuracy is 0.7864; the accuracy of the OS-ID3 model for the positive data is 0.5687, and the total accuracy is 0.7882; The accuracy of class data is 0.5824, and the total accuracy is 0.7428; the accuracy of DSB-ID3 model for positive class data is 0.7345, and the total accuracy is 0.7893. It can be seen that among the four models, DSB-ID3 model has the highest recognition accuracy and total recognition accuracy for regular data, indicating that its classification effect is the best, and it shows superior accuracy and stability in the analysis of financial data.The final student classification accuracy is shown in Fig. 9.

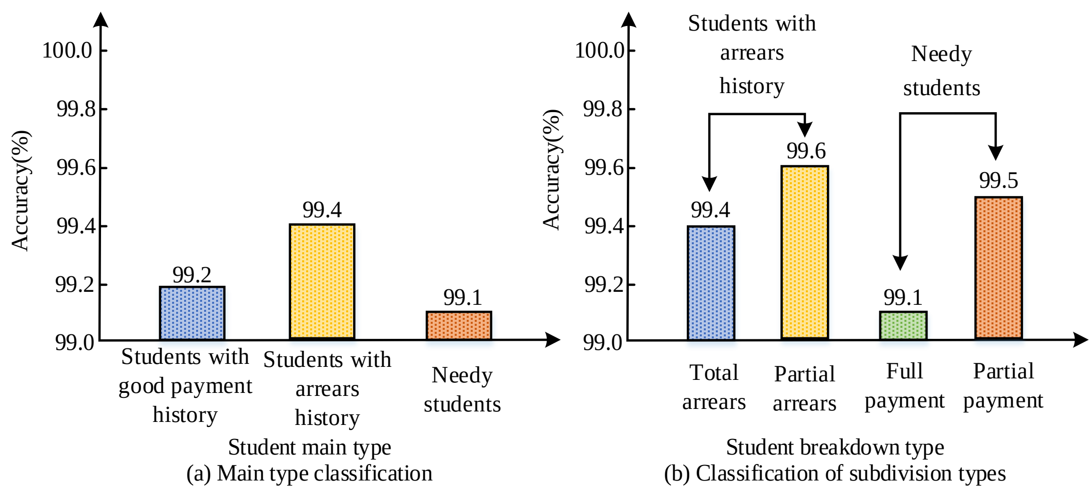


Fig. 9. Student classification accuracy

It can be seen that, for different types of students, the model has a classification accuracy of 99.2% for students with good payment history, 99.3% for students with arrears history, and 99.1% for poor students, all above 99.0%. At the same time, in terms of the classification of historical students in arrears, the classification accuracy reached 99.4% for all students in arrears and 99.6% for some students in arrears. In terms of poor students, the classification accuracy of all students who have paid all the fees has reached 99.1%, and the classification accuracy of students who have paid part of the fees has reached 99.5%. It can be seen that the model designed in this study can obtain good classification effect in practical application, has feasibility and practicality, and has high operational accuracy.

V. Discussion

With the development of management informatization of higher education structure, the demand for financial intelligent management in colleges and universities is growing day by day. The main goal of financial work has also changed from the original accounting type to the analytical type. ID3 algorithm itself has a significant disadvantage, that is, it tends to choose more values rather than the optimal attribute. Therefore, this research introduces risk weight to improve ID3 algorithm, and designs DSB-ID3 model based on SMOTE algorithm and Bagging algorithm. The analysis results show that with the increase of experimental time, the G value of each model fluctuates repeatedly, but except for DSB-ID3 model, the G value of weighted ID3 model, OS-ID3 model and OUS-ID3 model shows a general downward trend, and the linear regression slope is 1.74 respectively $\times 10^{-4}$、1.32 $\times 10^{-4}$ and 1.25 $\times 10^{-4}$。 The F value of each model fluctuates repeatedly with the increase of experimental time, and the change rule of each model is roughly similar to that in Fig. 5. Specifically, F values of weighted ID3 model, OS-ID3 model, OUS-ID3 model and DSB-ID3 model show a downward trend in general, and their linear regression slopes are 0.82 respectively $\times 10^{-4}$、0.94 $\times 10^{-4}$ and 0.71 $\times 10^{-4}$, 0.35 $\times 10^{-4}$, it can be seen that the overall decline speed of DSB-ID3 model is the slowest. It can be seen that using DSB-ID3 model to analyze college financial data, its G value and F value are the highest among the four models, with the best classification effect. Compared with other models, it has significantly improved, and its distribution is the most centralized. The research results of Apalkova V and others also show that the classification ability of the decision tree model can be improved to a certain extent after using SMOTE algorithm to improve the decision tree model [24]. Among the four models, DSB-ID3 model has the highest accuracy, with a recognition accuracy of 0.7345 for positive data and a total recognition accuracy of 0.7893. In the final classification accuracy rate of students, the classification accuracy rate of this model for different types of students exceeds 99%. The classification effect of DSB-ID3 model is the best, showing superior accuracy and stability in financial data analysis. The research results of Podhorska I also show that the classification accuracy of the improved integrated algorithm using SMOTE algorithm is higher than that of the original algorithm [25].

VI. Conclusion

In order to overcome the disadvantage that ID3 algorithm tends to select more values but not necessarily the optimal attribute, this study introduces risk weight to improve ID3 algorithm and designs DSB-ID3 model. Using this model to analyze the financial data of colleges and universities, its G value and F value are the highest among the four models, about 0.78. The classification effect is the best. Compared with other models, it has significantly improved, and its distribution is the most centralized. For superior stability. Among the four models, DSB-ID3 model has the highest accuracy, with a recognition accuracy of 0.7345 for positive data and a total recognition accuracy of 0.7893. In the final classification accuracy rate of students, the classification accuracy rate of this model for different types of students exceeds 99%. The classification effect of DSB-ID3 model is the best, showing superior accuracy and stability in financial data analysis. The DSB-ID3 model of financial analysis in colleges and universities shows excellent classification performance in financial data analysis. The accuracy and stability of its classification effect provide practical reference for the application of financial analysis in colleges and universities. This research uses the annual financial report data of the real university to carry out the research, which ensures the rationality of the research results. However, the model still has many shortcomings in practical application. For example, this study only uses student payment data for experiments, and the application of school financial reports is relatively low.

Acknowledgment

The research is supported by the the 2022 Shandong Provincial Educational Science Planning and Innovation Literacy Special Project ' Research on the Integration Path of Innovation and Entrepreneurship Education and Professional Education for Finance and Economics major in Higher Vocational Colleges' (Project No.: 2022CYB356, Project Leader: Guo Sujuan) and 2021 General subject of the China Vocational Technology Education Society ' Research on the Digital Course Development Model of the Smart Finance Professional Group in the New Professional Catalogue of Higher Vocational Education'(Project No.: 2021B067, Project Leader: Guo Sujuan).

References

[1] A. Luthfiarta, J. Zeniarja, E. Faisal, W. Wicaksono, “Prediction on Deposit Subscription of Customer based on Bank Telemarketing using Decision Tree with Entropy Comparison,” Journal of Applied Intelligent System, Vol. 4, No. 2, pp. 57-66, 2020.

[2] J. Tian, Y. Wang, W. Cui, K. Zhao, “Simulation analysis of financial stock market based on machine learning and GARCH model,” Journal of Intelligent and Fuzzy Systems, Vol. 40, No. 2, pp. 2277-2287, 2021.

[3] T. Mahara, I. Naim, “Framework to identify a set of univariate time series forecasting techniques to aid in business decision making,” International Journal of Intelligent Enterprise, Vol. 7, No. 4, pp. 423-443, 2020.

[4] Z. Wu, “Using Machine Learning Approach to Evaluate the Excessive Financialization Risks of Trading Enterprises,” Computational Economics, Vol. 59, No. 4, pp. 1607-1625, 2022.

[5] Y. Li, Y. Pan, “A novel ensemble deep learning model for stock prediction based on stock prices and news,” International Journal of Data Science and Analytics, Vol. 13, No. 2, pp. 139-149, 2021.

[6] J. Wyrobek, “Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture,” Procedia Computer Science Vol. 176, pp. 3037-3046, 2020.

[7] X. Luo, X. Wen, M. C. Zhou, A. Abusorrah, L. Huang, “Decision-Tree-Initialized Dendritic Neuron Model for Fast and Accurate Data Classification,” IEEE Transactions on Neural Networks and Learning Systems, Vol. 33, No. 9, 4173-4183, 2021.

[8] I. Podhorska, J. Vrbka, G. Lazaroiu, M. Kovacova, “Innovations in Financial Management: Recursive Prediction Model Based on Decision Trees,” Marketing and Management of Innovations Vol. 3, pp. 276-292, 2020.

[9] M. A. Febriantono, S. H. Pramono, R. Rahmadwati, G.Naghdy, “Classification of multiclass imbalanced data using cost-sensitive decision tree C5.0.” Vol. 9, No. 1, pp. 65-72, 2020

[10] M. F. Maulana, M. Defriani, “Logistic Model Tree and Decision Tree J48 Algorithms for Predicting the Length of Study Period,” PIKSEL Penelitian Ilmu Komputer Sistem Embedded and Logic, Vol. 8, No. 1, pp. 39-48, 2020.

[11] Y. Xia, L. He, Y. Li, Y. Xu, “A Dynamic Credit Scoring Model Based On Survival Gradient Boosting Decision Tree Approach,” Technological and Economic Development of Economy, Vol. 27, No. 1, pp. 1-24, 2020.

[12] M. Balta, İbrahim Özçelik. “A 3-stage fuzzy-decision tree model for traffic signal optimization in urban city via a SDN based VANET architecture,” Future Generation Computer Systems Vol. 104, pp. 142-158, 2020.

[13] T. Qiu, M. Zhang, X. Liu, J. Liu, W. Zhao, “A Directed Edge Weight Prediction Model Using Decision Tree Ensembles in Industrial Internet of Things,” IEEE Transactions on Industrial Informatics, Vol. 17, No. 3, pp. 2160-2168, 2020.

[14] J. Keuangan, P. Dan, A. Muditomo, S. Broto, “IPO Performance Prediction during Covid-19 Pandemic in Indonesian Using Decision Tree Algorithm,” Jurnal Keuangan dan Perbankan, Vol. 25, No. 1, pp. 132-143, 2021.

[15] L. Xue, D. Liu, C. Huang, X. Lin, X. S.Shen, “Secure and Privacy-Preserving Decision Tree Classification with Lower Complexity,” Journal of Communications and Information Networks, Vol. 5, No. 1, pp. 16-25, 2020.

[16] S. Ziweritin, I. A. Ibiam, T. A. Oyeniran, G.E. Oko, “KNN and Decision Tree Model to Predict Values in Amount of One Pound Table,” International Journal on Computer Science and Engineering, Vol. 9, No. 7, pp. 17-21, 2021.

[17] Natawibawa I W Y, Mulya I M O, Yoh W H. Transparency and Accountability As Determinants in the Financial Management of Universities: a Study on State Universities in Malang City. Jurnal Tata Kelola Dan Akuntabilitas Keuangan Negara, 2019, 5(1): 57-72.

[18] Min Du. Management Research of College Finance under Informatization Background//.Proceedings of 2018 4th International Conference on Education & Training,Management and Humanities

Science(ETMHS 2018).Clausius Scientific Press,2018:70-74.

[19] Na Sun. Discussion on the Application of Management Accounting in University Finance//.Proceedings of 2019 5th International Workshop on Education,Development and Social Sciences(IWEDSS 2019).Francis Academic Press,2019:164-168.

[20] Atanasijevic J, Milosevic D. Upgrading the business intelligence system by implementing the decision tree model in the R software package. Studies in Informatics and Control, 2020, 29(2):243-254.

[21] Punitha S, Jeyakarthic M. Enhanced Particle Swarm Optimization with Decision Tree based Prediction Model for Stock Market Directions. Journal of Advanced Research in Dynamical and Control Systems, 2020, 12(5):1432-1442.

[22] Feofanov A, Egorov M. IMPROVING THE ENTERPRISE PERFORMANCE ON THE BASIS OF APPLYING DECISION TREE METHOD. Automation and Modeling in Design and Management of, 2021(1):29-34.

[23] Syamala M, Nalini N. A Filter Based Improved Decision Tree Sentiment Classification Model for RealTime Amazon Product Review Data. International Journal of Intelligent Engineering and Systems, 2020, 13(1):191-202.

[24] Apalkova V, Tsyganov S, Meshko N, Tsyganova N，Apalkov S. Application of decision tree model for prediction of immigration policy in different countries of the world. Problems and Perspectives in Management, 2021, 19(3):513-532.

[25] Podhorska I, Vrbka J, Lazaroiu G, Kovacova M. Innovations in Financial Management: Recursive Prediction Model Based on Decision Trees. Marketing and Management of Innovations, 2020(3):276-292.

Clustering-based Automated Requirement Trace Retrieval

Nejood Hashim Al-walidi[1], Shahira Shaaban Azab[2], Abdelaziz Khamis[3], Nagy Ramadan Darwish[4]
Department of Information Systems and Technology[1, 3, 4]
Department of Computer Science[2]
Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt

***Abstract*—The benefits of requirement traceability are well known and documented. The traceability links between requirements and code are fundamental in supporting different activities in the software development process, including change management and software maintenance. These links can be obtained using manual or automatic means. Manual trace retrieval is a time-consuming task. Automatic trace retrieval can be performed via various tools such as Information retrieval or machine learning techniques. Meanwhile, a big concern associated with automated trace retrieval is the low precision problem primarily caused by the term mismatches across documents to be traced. This study proposes an approach that addresses the term mismatch problem to obtain the greatest improvements in the trace retrieval accuracy. The proposed approach uses clustering in the automated trace retrieval process and performs an experimental evaluation against previous benchmarks. The results show that the proposed approach improves the trace retrieval precision.**

***Keywords*—Requirements traceability; information retrieval; term mismatch problem; trace retrieval; TraceLab; clustering**

I. Introduction

Many software development standards have been proposed in response to the high rate of software project failures. These standards include SEI's CMMI and IEEE's JSTD-016. A common feature of these standards is that they all impose requirement traceability (RT) practices on the software development process [1]. RT is an important quality factor of software development that intends to ensure a continued alignment between stakeholder requirements and various outputs of the system development process. In addition [2], RT makes it easy to determine what software artifacts must be updated to fulfill a change request made during the maintenance phase of the software project.

In 2014, members of the Center of Excellence for Software traceability (CoEST) established a roadmap for advancing the state of practice in software traceability and presented a focused research agenda for software traceability. They identified seven broad research focus areas and outlined the specific research needed in each of these areas. Under the umbrella of one of these research areas, that is, *Creating & Maintaining Traces*, we focus herein on trace retrieval, which is concerned with dynamically generating trace links between source and target artifacts [3]. Researchers utilized different algorithms to infer the trace links between related artifacts based on the assumption that related artifacts contain related terms [4]. The underlying concept behind this was that these algorithms could be used to estimate the similarity between two documents.

A big concern associated with trace retrieval research is the low precision problem. Precision is the percentage of correct traces over all retrieved traces. "Low precision" indicates that many false traces have been incorrectly retrieved, and the user must manually evaluate the retrieved links to identify the correct traces. This leads to an emphasis on precision. Accordingly, some researchers examined ideas on how to increase it [4] [5]; however, these ideas resulted in only a minor increase in precision. The problem is primarily caused by term mismatches across documents to be traced. The reason for choosing the proposed approach is to enhance trace retrieval precision. Therefore, the authors formulated the problem by choosing an intelligent solution based on unsupervised learning using clustering. The datasets have a limited number of labels. These labels are provided for testing only. So, the suggestion for addressing this problem is by choosing an intelligent solution based on unsupervised learning to find similarities in the data point and group similar data points together to enhance trace retrieval precision.

In this work, to address the term mismatch problem in automated trace retrieval, we follow the proposed research direction toward achieving automated trace retrieval, that is, to develop intelligent tracing solutions, "which are not constrained by the terms in source and target artifacts, but which understand domain-specific concepts, and can reason intelligently about relationships between artifacts" [3]. We lay down the foundations needed to use clustering in automated trace retrieval between *source code* and *requirements*.

Clustering, which relies on unsupervised machine learning, is the task of grouping a set of objects in such a way that objects in the same group, called a cluster, are more similar in some sense to each other than to those in other groups or clusters [6]. In this study, we used K-means++, Hierarchical, Gaussian mixture model (GMM), and Density-based Spatial Clustering of Application with Noise (DBSCAN) algorithms. Clustering algorithms can be summarized in three steps:

1) *Inputs*: dataset after preprocessing and other steps related to the dataset type.
2) *Applying* suitable models.

A. *Outputs: evaluating the results after the clustering processes; for this, we decided to use clustering to cluster the dataset into two: clusters 0 and 1. Cluster 1 means the* *items are similar. Cluster 0 means the items are different. All clustering steps are described in Section IV. We present an overview of these methods below.*

The K-means is a partition-based iterative clustering algorithm, in which each cluster is characterized by its center point. It is widely used because of its implementation simplicity and effectiveness. It assumes that the number of clusters represented by "K" in K-means is already known. This algorithm aims to minimize the distance inside the same cluster and maximize the distance between the clusters [7].

Hierarchical clustering divides the data set at different levels to form a tree-shaped cluster structure that naturally defines clusters by branches in the hierarchical tree. It makes a few assumptions regarding the overall data point distribution; hence, it is suitable for datasets of many different shapes. Hierarchical clustering has two main implementations: agglomerative and divisive clustering [8].

In model-based clustering, each cluster is considered as a generative model with mean and variance. Instances arise from a distribution that is a mixture of several components. Gaussian (normal) distributions are the most used representation in model-based clustering. The mixture model is the GMM, whose components are Gaussian distributions with different means and variances [9].

The density-based clustering algorithm, called DBSCAN, recognizes arbitrary-shaped clusters under a high noise level of the studied data. The efficiency of DBSCAN depends on the parameter values set up at the initial step of the algorithm usage [10].

In this paper, we adopted four widely used metrics to evaluate the new proposed approach, precision, recall, F-score, and accuracy. Both precision and recall are used to assess the effectiveness of the requirement tracing tools [11]. **Precision** is defined as the percentage of correct retrieved candidate links and measures the fraction of retrieved documents which are relevant [12]. **Recall** is the percentage of the found correct links [13]. **F-score** is used to provide a balance between recall and precision [14]. It is the harmonic mean of recall and precision [15] and used to aggregate both measures into a single value [16]. Finally, **accuracy** measures are the rate of normal and outlier values correctly classified between the total number of classifications [17].

The remainder of this paper is structured as follows: Section II describes the related works in the trace retrieval field; Section III presents the proposed approach; Section IV explains the obtained results; Section V includes a result evaluation; and Section VI provides the conclusions and future work.

II. Related Works

This section reviews some methods used to solve the term mismatch problem in automated requirement trace retrieval.

In the paper [19] titled "Automatic traceability link recovery via active learning", the authors proposed a new traceability link recovery approach based on active learning (AL). They evaluated their approach on seven datasets used in traceability and compared them with an information retrieval (IR)-based approach and a state-of-the-art machine learning approach, called traditional supervised learning. The used datasets are available from the CoEST website (http://www.CoEST.org). The results showed that the AL-based approach outperforms the other two in terms of the F-score.

In the paper [20] titled "Traceability Transformed: Generating more Accurate Links with Pre-Trained BERT Models", the authors proposed a framework, called Trace BERT (T-BERT), to create trace links between the source code and the natural language artifacts. They then applied the T-BERT framework to recover links between issues and commits in open-source projects. The evaluation results as regards the accuracy and efficiency of three BERT architectures indicated that the Single-BERT architecture generated the most accurate links, while the Siamese-BERT architecture produced comparable results with significantly less execution time. By learning and transferring knowledge, all three models in the framework outperformed classical IR trace models.

In the paper [21] titled "Analyzing close relations between target artifacts for improving IR-based requirement traceability recovery", the authors proposed a method for trace link recovery by combining the IR method and the close relations between the target artifacts. This approach was referred to as IR_CRT. Aside from textual similarity, the close semantic relations between the target artifacts were considered. Experiments on five public datasets indicated that the precision on these datasets improved by 15.6% on average, showing that their method outperformed the baseline when working under the same conditions.

In the paper [22] titled "Ontology-based Trace Retrieval", the authors solved the term mismatch problem in automated requirement trace retrieval by incorporating information from the general and domain-specific ontologies into the tracing process. They used ontologies to identify relationships that would not be recognized by standard IR techniques. They then experimentally evaluated their approach against the standard vector space model (VSM). Their results showed that a domain ontology combined with a generalized ontology returns the greatest improvements in trace accuracy.

In the paper [23] titled "Towards an Intelligent Domain-Specific Traceability Solution", the authors solved the term mismatch problem in automated requirement trace retrieval by presenting the domain-contextualized intelligent traceability (DoCIT) solution. This approach mimicked some of the higher-level reasoning that a human trace analyst performs. The authors focused their efforts on the complex domain of communication and control in a transportation system and found that their approach can significantly improve the quality of the generated trace links. They illustrated and evaluated DoCIT with examples and experiments from the control and communication sector of a transportation domain.

In the paper [24] titled "Combining Machine Learning and Logical Reasoning to Improve Requirements Traceability Recovery", the authors proposed a novel traceability link recovery approach that measures the similarity between requirements and the source code by exploring their features.

They combined machine learning and logical reasoning models and conducted a series of experiments on four datasets to evaluate the performance of their method against existing approaches. Their experiments showed that their approach is substantially better than other methods.

III. Proposed Approach

In this section, we divided the task of the proposed approach using clustering to improve the recovery of links between source code and requirements, into two main phases. Fig. 1 illustrates the two phases of the proposed trace retrieval approach.

B. Phase 1: Choosing Suitable Datasets

Research on automated requirement traceability relies on the availability of different dataset types. In general, obtaining datasets has been one of the reported barriers for researchers in the software engineering domain [25]. This phase introduces the three datasets used in this study to evaluate automated trace retrieval between the source code and requirements. Table I defines the characteristics of the datasets used in the proposed approach. These datasets are available from http://www.CoEST.org.

TABLE I. Dataset Characteristics [25],[26],[27], [28][29][30]

Dataset description	Freq	Traceability details	Trace space	CC	UC	Trace links
eTour: "an electronic tourist guide developed by students"	10	Use cases to code	6728	116	58	366
SMOS: "an application that is used to monitor high school students (e.g., absence, grades)"	7	Use cases to classes	6700	100	67	1044
eANCI: "system providing support to manage Italian municipalities"	3	Use cases to classes	7645	55	140	567

The above-mentioned datasets were chosen because they are available for all researchers and compatible with TraceLab's environment.

With these datasets, we ran into some issues related to the two artifacts (requirements and source code). For example, all three datasets are imbalanced datasets. They included useless words, with some words starting with extra letters and ending with extra letters without meaning. A few single characters can be found between the lines of the files that did not have a clear meaning. According to the Answer set of the dataset, some source codes did not have any requirements, which affected the links between the two artifacts and gave wrong similarities. Some words and sentences were written in different languages. Some of them were in English, while others were in Italian.

C. Phase 2: Choosing and Applying Clustering Models

Clustering is an unsupervised machine learning task used to gather data into many collections or clusters according to the similarities of the data point features and characteristics [31]. In this phase, we followed seven main steps to apply clustering models to the three datasets:

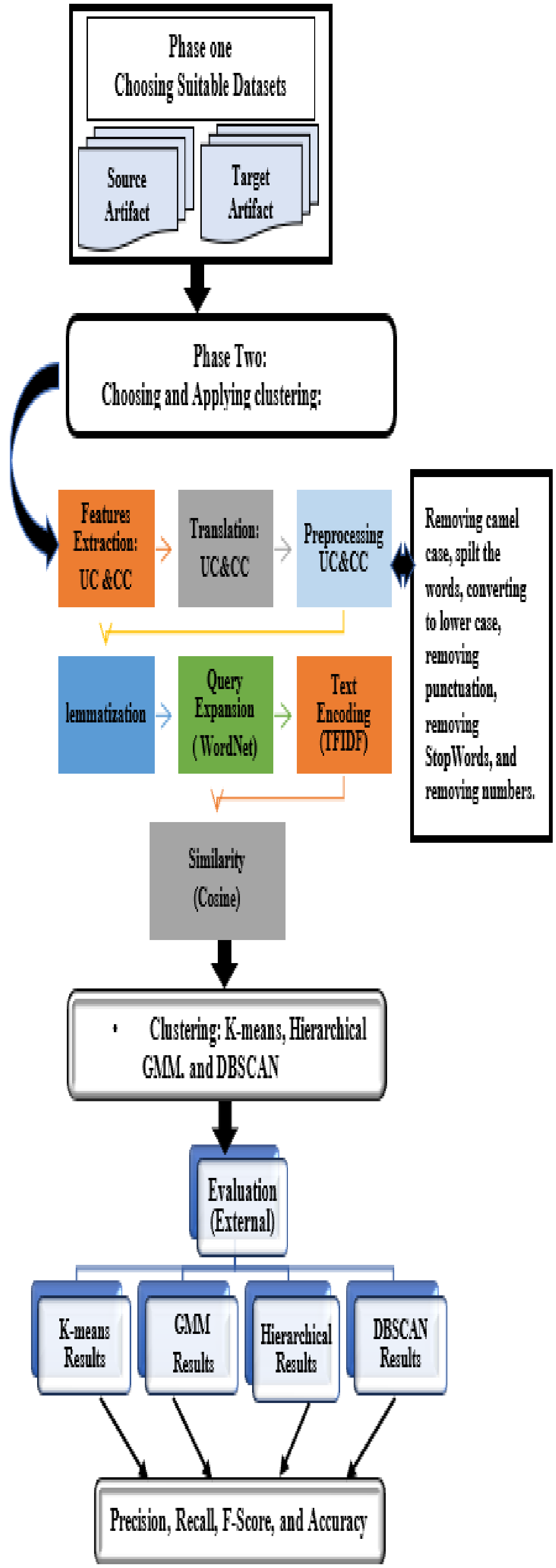


Fig. 1. Proposed trace retrieval approach

Step 1, *Feature extraction*: The requirements (source artifact) and the source code (target artifact) were written in natural and programming languages, respectively. Hence, they included different types of information that were not all useful for identifying the traceability links between the two artifacts. For that, we extracted the most important source code feature, namely (*class_name, class_attribute, class_comments, method_comment,method_name,method_parameter,method_return*). We also extracted the two main features related to the requirements (i.e., Title and Description) as shown in Fig. 3.

Step 2, *Translation*: The content of the two artifacts (i.e., UC and CC files) included Italian words and sentences in some files and some parts of the file lines. Therefore, the content of the two artifacts was translated into English. To this end, we leveraged the advanced translation engine to translate documents into English if they were originally unspecified in English. We used one of the tools used for translation into the English language namely Google Translate [32], which enables a method of translating text from one language to another.

Step 3, *Preprocessing*: We followed a classic process to pre-process the contents of the software artifacts, UC and CC. The preprocessing comprised the following steps: removing the camel case, splitting the words, converting to a lower case, and removing punctuations, Stop Words, and numbers. We performed lemmatization for all words to facilitate natural language analysis, such as transforming the verb, noun, adjectives, and adverbs into their bases and examining the study of word texts with the intention of finding something that adds meaning to the word text [33] and removing any single character from both artifacts.

Step 4, *Query expansion using WordNet*: WordNet is an online lexical system (database) for the English language that provides diverse and wide-ranging semantic information [34]. We used query expansion to add related words to a query to increase the number of returned documents and improve the recall accordingly. In most cases, all words in each query should first be extracted. For each word, the synonyms were automatically selected.

Step 5, *Text encoding*: Many methods can be used to convert text into numerical vectors, such as TF-IDF encoding, Dec2Vec, Word2Vec, and Bag of Words (BOW) [35]. We experimented Dec2Vec and Word2Vec, but they did not improve experiments results, so we used the TFIDF to have better results.

Step 6, *Computing similarity*: Using the cosine similarity [36], we computed the similarity between two vectors after converting the corresponding contents of each feature into a vector by using the TFIDF algorithm. The cosine similarity algorithm measures the similarity between two vectors, which can represent paragraphs, sentences, words, or the entire document.

Step 7: In this step, four types of clustering models were used, namely K-means++, GMM (hard clustering type), Hierarchical, and DBSCAN. After the previous steps, we first prepared the dataset. Fig. 2 illustrates the total columns and rows of the dataset which are 14 columns. Second, we chose the clustering type. Third, evaluating the results, we performed external evaluation criteria and computed the confusion matrix (i.e., precision, recall, F-score, and accuracy measures. Table II shows the meaning of the confusion matrix symbols that are used for precision, recall, F-Score, and accuracy. The mathematical formulas of accuracy, precision, recall, and F-score are as follows:

Accuracy = $(TP + TN) / (TP + FP + FN + TN)$.

Precision = $TP / (TP + FP)$

Recall = $TP / (TP + FN)$

F-score = 2*(precision*recall)/ (precision + recall).

TABLE II. MEANING OF THE CONFUSION MATRIX SYMBOLS [18]

Symbol	Description
TP	The number of data pairs found in the same cluster, both in C and in P.
FP	The number of data pairs found in the same cluster in C but in different clusters in P.
FN	The number of data pairs found in different clusters in C but in the same cluster in P.
TN	The number of data pairs found in different clusters, both in C and in P.

C means the actual cluster and P means the prediction cluster.

F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
0	0	0	0	0	0	0	0	0	0.1026	0.1003	0	0.0231	0
0	0	0	0	0	0	0	0	0.0277	0.0795	0.0242	0	0.0606	0.0247
0	0	0	0	0	0	0	0	0	0.0216	0.0135	0.0193	0.0273	0.0214
0	0	0	0	0	0	0	0	0	0	0.0305	0	0	0.0172
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.1127	0	0.0115	0	0	0	0	0.1301	0	0.0095	0	0
0	0	0	0	0	0	0	0	0.0284	0	0	0.0371	0	0.0279
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.3077	0	0.2007	0	0.0672	0	0
0	0	0	0	0	0	0	0.1521	0.0122	0	0	0.0286	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0.0036	0	0
0	0.0795	0.1109	0	0.057	0.0705	0.0255	0	0.0787	0.0916	0	0.0471	0.0814	0.0211
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0.0247	0.0835	0.189	0.044	0.1305	0.0109	0	0.0519	0.2206	0.1859	0.0508	0.1178	0.0126
0.1465	0.1176	0.0473	0.0685	0.1205	0.1713	0	0.2479	0.1165	0.0583	0.0954	0.0995	0.1554	0.037
0	0.0172	0	0	0.0115	0	0	0.0953	0.0142	0.0485	0.0255	0.0202	0.0089	0
0	0	0	0	0	0	0	0.0716	0.0634	0	0.0128	0	0	0
0.1279	0.0774	0.0507	0.0503	0.0774	0.1032	0	0.2165	0.0894	0.0419	0.0415	0.0894	0.0852	0.0288
0	0	0	0	0	0	0	0	0.0276	0.1323	0.0322	0.028	0	0
0	0	0	0	0	0	0	0	0.0138	0.0354	0	0.0187	0	0

Fig. 2. The 14 columns of the dataset

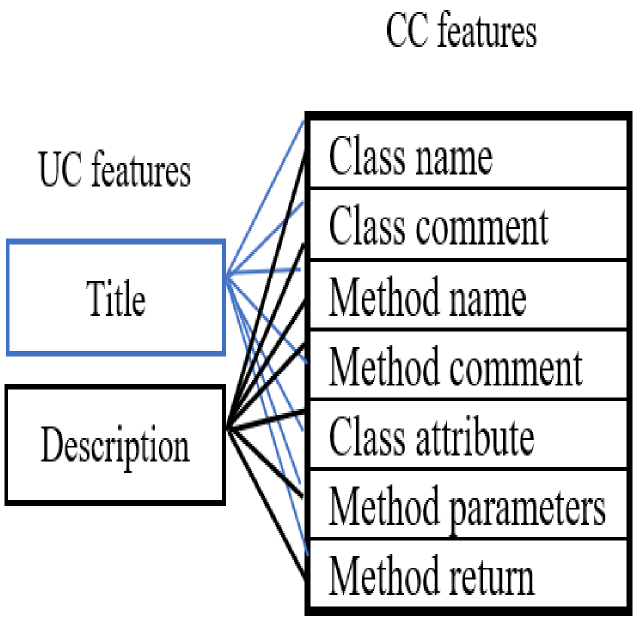


Fig. 3. Feature extraction

IV. RESULTS

In this section, we present the results of applying the Proposed Trace Retrieval Approach described in Section 3, which consists of two main phases as shown in Fig 1.

As mentioned in Section III, we applied four clustering models (i.e., K-means++, GMM, Hierarchical, and DBSCAN) to the three datasets described in Phase 1 of the proposed trace retrieval approach.

The results of the new proposed trace retrieval approach are presented here which include four measures: Precision, recall, F-score, and accuracy as shown in (Table III).

Comparing the proposed approach with two experiments, namely TraceLab and IR as shown in Tables IV (A) to (C). Tables V (A) to (C) show the comparison results of the proposed method and the two studies (i.e., studies [19] and [24]).

First, Table III and Fig. 4 present the results of K-means++ for three datasets: eTour, SMOS, and eANCI achieved high values compared to GMM, hierarchical DBSCAN results. The proposed trace retrieval approach is effective in improving precision, recall, F-score, and accuracy. The extent of improvement differed from one dataset to another depending on the dataset type. We used accuracy as an additional measure, whereas TraceLab and Information retrieval R and two studies [19] and [24] did not. The proposed approach achieved highest results using accuracy with the three datasets (i.e., eTour = 0.91, SMOS = 0.66, and eANSI = 0.60). Three parameters used in the experiments: max_feature with a value of 400 for the eTour dataset, SMOS, featurewiz when (corr_limit values) equal 0.5, and corr_limit values equal 0.5 for eANSI.

TABLE III. SUMMARY OF THE RESULTS OF THE PROPOSED TRACE RETRIEVAL APPROACH USING TWO CLUSTERING ALGORITHMS FOR THREE DATASETS

eTour dataset:				
Model Name	**Precision**	**Recall**	**FScore**	**Accuracy**
K-means++	**0.93**	**0.97**	**0.94**	**0.91**
SMOS dataset				
K-means++	**0.73**	**0.76**	**0.74**	**0.66**
eANCI dataset				
K-means++	**0.64**	**0.77**	**0.70**	**0.60**

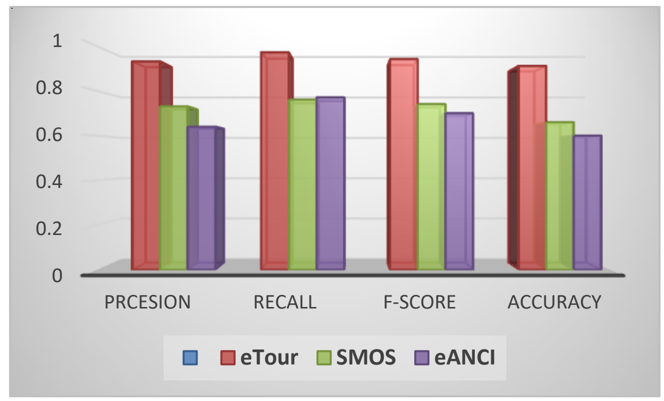
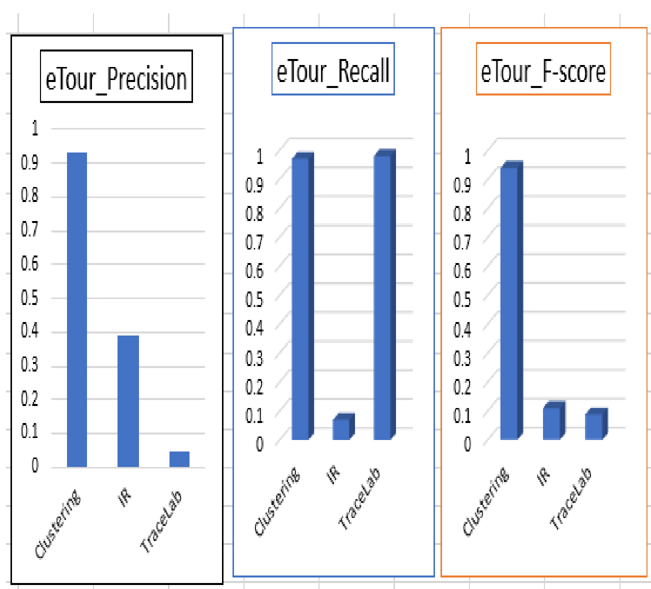
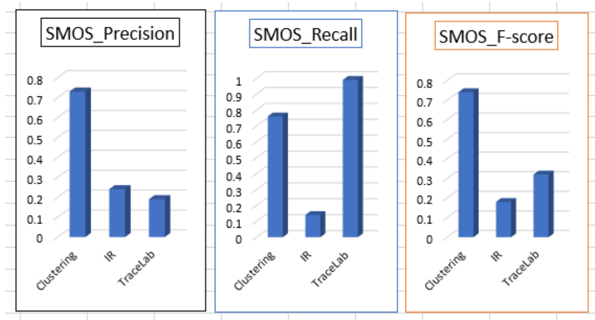


Fig. 4. Summary of the results of the proposed trace retrieval approach using two clustering algorithms for three datasets

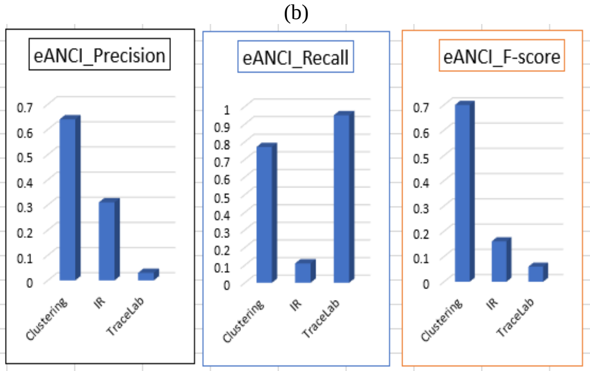
Second, presenting a comparison between the proposed trace retrieval approach and two experiments, namely ***TraceLab and IR***. Table IV (A) to (C) and Fig. 5(a) to (c) present a comparison of the proposed approach and the two experiments. The results showed that the former returns the greatest improvements in the trace measures.



(a)



(b)



(c)

Fig. 5. (a). eTour result comparison: clustering, IR, and TraceLab (b) SMOS result comparison: clustering, IR, and TraceLab (c) eANCI result comparison: clustering, IR, and TraceLab

1) Experiment one (TraceLab): Study [37] designed to help and support the reproducibility of experiments in software engineering and maintenance. In TraceLab's visual modeling environment [38], IR algorithms are implemented as experiments using a library of reusable and user-defined components [39]. An already existing experiment, called Basic IR [40], which is a standard support vector model experiment, was conducted to implement the VSM/TF*IDF algorithm.

2) Experiment two: four steps were applied. *The first step* was preprocessing, which included transforming the upper case to the lower case, removing the punctuations and digits, converting plurals to singulars, and transforming verbs into infinitives. *The second* step was getting the weight of each term into artifacts using TFIDF. *The third step* was computing the similarity between the source and target artifacts using the cosine similarity. *The fourth* and final step was computing the precision, recall, and F-score.

TABLE IV. (A). ETOUR COMPARED WITH IR AND TRACELAB

Measure	Clustering	IR	TraceLab
Precision	**0.93**	0.39	0.046

Measure	Clustering	IR	TraceLab
Recall	**0.97**	0.07	0.98

Measure	Clustering	IR	TraceLab
F-score	**0.94**	0.11	0.088

(B). SMOS COMPARED WITH IR AND TRACELAB

Measure	Clustering	IR	TraceLab
Precision	**0.73**	0.24	0.19

Measure	Clustering	IR	TraceLab
Recall	0.76	0.14	0.99

Measure	Clustering	IR	TraceLab
F-score	**0.74**	0.18	0.32

(C). EANCI COMPARED WITH IR AND TRACELAB

Measure	Clustering	IR	TraceLab
Precision	**0.64**	0.31	0.03

Measure	Clustering	IR	TraceLab
Recall	**0.77**	0.11	0.95

Measure	Clustering	IR	TraceLab
F-score	**0.70**	0.16	0.06

Tables IV (A) to (C) and Fig. 5(a) to (c) show that the proposed trace retrieval approach achieved high results with three datasets in the precision measure, which was the concern in this work. TraceLab achieved high results in recall because the precision and F-score measures were very low, and the false negative (FN) of the links was more than the true positive with the three datasets. The proposed approach achieved high results with the three datasets in F-score compared to the IR and TraceLab.

Third, compared the results of the proposed approach and those of other studies. Tables V (A) to (C) and Fig. 6(a) to (c) present the results obtained from applying clustering models into the three datasets. The performance of the proposed approach was compared to those of studies 19 and 24, which addressed the same problem and used the same datasets. Four measures were used: precision, recall, and F-score as the comparison metrics. The two studies used precision, recall, and F-score, but not accuracy.

Tables V (A) to (C) and Fig. 6(a) to (c) depict that the proposed trace retrieval approach achieved high results with precision, recall, and F-Score in the eTour dataset than in the two studies. In the SMOS dataset, the proposed approach trace retrieval achieved high results using precision, recall, F-Score than in the two studies. Meanwhile, in the eANCI dataset, the proposed approach obtained low results using precision than in the two studies and highest results in recall and F-Score.

TABLE V. (A). ETOUR COMPARED WITH THE TWO STUDIES

Measure	Clustering	Study [19]	Study [24]
Precsion	**0.93**	0.68	0.66

Measure	Clustering	Study [19]	Study [24]
Recall	**0.97**	0.34	0.59

Measure	Clustering	Study [19]	Study [24]
F-score	**0.94**	0.46	0.61

(B). SMOS COMPARED WITH THE TWO STUDIES

Measure	Clustering	Study [19]	Study [24]
Precsion	**0.73**	0.57	0.75

Measure	Clustering	Study [19]	Study [24]
Recall	**0.76**	0.29	0.33

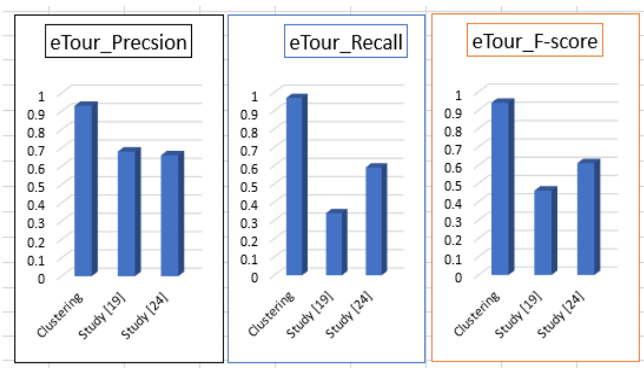
Measure	Clustering	Study [19]	Study [24]
F-score	**0.74**	0.39	0.51

(C). EANCI COMPARED WITH THE TWO STUDIES

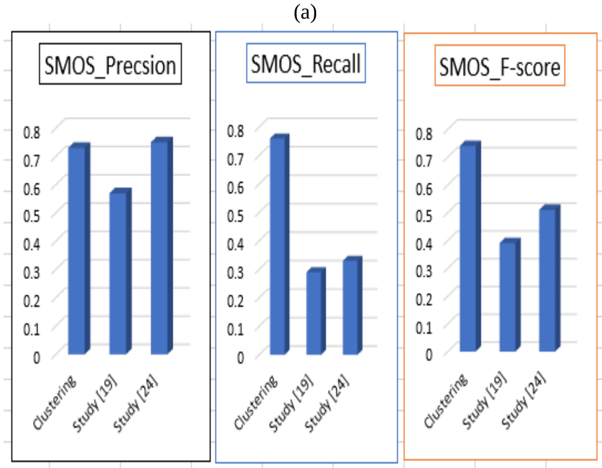
Measure	Clustering	Study [19]	Study [24]
Precsion	**0.64**	0.73	0.62

Measure	Clustering	Study [19]	Study [24]
Recall	**0.77**	0.44	0.54

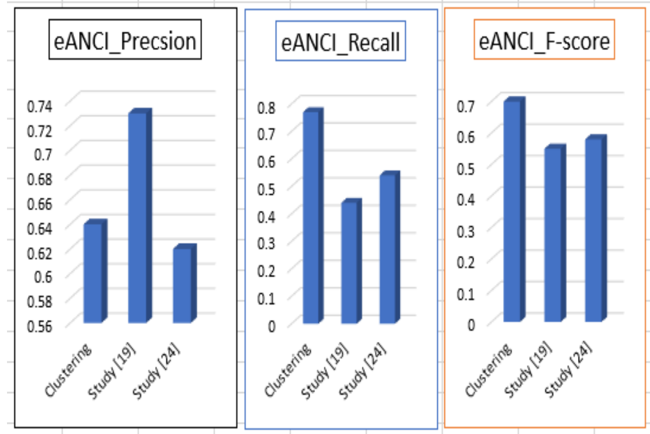
Measure	Clustering	Study [19]	Study [24]
F-score	0.70	0.55	0.58



(a)



(b)



(c)

Fig. 6. (A). eTour result comparison: clustering and studies [19] and [24] (B). SMOS result comparison: clustering and studies [19] and [24] (C). eANCI result comparison: clustering and studies [19] and [24]

Finally, the authors will present all the experimental results obtained from K-means++, hierarchical, and DBSCAN as shown in (Tables VI-XVII). Those tables include different results using various parameters used during running the experiments such as (max_features). The max_features set the maximum number of features to be used by specifying the value between (100 and 900). Also, featureswiz package was used. The package is available at https://pypi.org/project/featurewiz/. featurewiz package is used for selecting the most important features using different parameters like ***corr_limit*** parameter using different values (0.1 – 0.9) as follows:

*features, train = featurewiz (data, target, **corr_limit=0.9**, verbose=2, sep=",", header=0, test_data=" ", feature_engg=" ", category_encoders="").*

First: eTour dataset, (Tables VI and VII) presents the results of the experiments obtained from K-means++ and hierarchical.

Table VI shows the result of the proposed trace retrieval approach using K-means++ based on max_feature values (100,300,400,800). The Table indicated that the precision achieved the highest value with all max_feature values. Recall achieved the highest results when is max_feature values=400 and 800 and F-Score with max_feature values=300 and 400.Finally, the accuracy achieved highest results when is max_feature values=300,400, and 800.

TABLE VI. ETOUR_K-MEANS ++ (MAX_FEATURE)

Measure's name	max_feature values			
	800	400	300	100
Precision	0.93	**0.93**	0.93	0.93
Recall	0.97	0.96	0.96	0.95
F-Score	0.93	0.94	0.94	0.93
Accuracy	0.91	0.91	0.91	0.90

According to Table VII, the proposed trace retrieval approach using hierarchical based on the max_feature values achieved the highest results with precision, recall, F-Score, and accuracy when is max_feature value =950.

TABLE VII. eTOUR_ HIERARCHICAL (MAX_FEATURE)

Measure's name	Max_feature vlues			
	800	**400**	**300**	**950**
Precision	0.91	**0.91**	0.91	**0.93**
Recall	0.56	0.56	0.56	**0.95**
F-score	0.69	0.69	**0.69**	**0.93**
Accuracy	0.56	0.57	0.56	**0.90**

Second: SMOS dataset, (Tables VIII–XII) presents the results of the experiments obtained from K-means++ and hierarchical.

Table VIII indicates the results of the proposed trace retrieval approach using K-means++ based on max_feature values (100 - 500). Precision achieved the highest results when are max_feature values-=400. Both Recall and accuracy achieved the highest results when is max_feature values=500, and F-score achieved the highest results when are max_feature values=300

TABLE VIII. SMOS_ K-MEANS ++MAX_FEATURE (MINMAX_SCALARE)

Measure's name	max_feature values				
	500	**400**	**300**	**200**	**100**
Precision	**0.71**	**0.72**	0.71	0.71	**0.71**
Recall	**0.80**	**0.74**	**0.79**	**0.73**	0.69
F-score	0.75	0.73	**0.77**	**0.70**	0.68
Accuracy	**0.66**	0.65	0.65	0.64	0.62

The authors inferred that the corr_limit values enhance precision, recall, F-Score, and accuracy with different values as shown in Table IX. Precision achieves the highest results when the corr_limit values =0.7 and 0.5. Recall and accuracy obtain the highest results when the corr_limit values = 0.7. F-Score obtains the highest result when corr_limit values=0.7 and 0.5.

TABLE IX. SMOS_KMEANS++ USING FEATURE_WIZ (MIN, AXSCALER)

Measure's name	Corr_limit values			
	0.7	**0.5**	**0.4**	**0.3**
Precision	**0.72**	**0.72**	**0.71**	0.71
Recall	0.71	0.70	0.67	**0.67**
F-score	0.71	0.71	0.69	**0.69**
Accuracy	**0.64**	**0.63**	**0.61**	**0.61**

As shown in Table X, the proposed trace retrieval approach using K-means++ based on standard scalar (corr_limit values). Precision and accuracy achieve the highest results when corr_limit value=0.5. Also, the recall and F-Score achieved the highest results when is corr_limit value=0.3.

TABLE X. SMOS_KMEANS ++ USING (STANDARD SCALAR)

Measure's name	Standard scalar (corr_limit values)			
	0.5	**0.4**	**0.3**	**0.6**
Precision	**0.73**	**0.70**	**0.66**	0.71
Recall	**0.76**	0.75	**0.96**	0.77
F-score	**0.74**	0.72	**0.78**	0.74
Accuracy	**0.66**	**0.63**	0.65	0.64

Table XI lists the results of the proposed trace retrieval approach using hierarchical based on max_feature values (100,200, 300, and 500). The precision achieved the highest value with all values of the max_feature. Recall, F-Score, and accuracy obtains the highest value when is max_feature values =500.

TABLE XI. SMOS_ HIERARCHICAL MAX_FEATURE(MINMAX_SCALARE)

Measure's name	max_feature values			
	500	**300**	**200**	**100**
Precision	**0.71**	**0.71**	**0.71**	**0.71**
Recall	**0.80**	0.79	0.73	0.77
F-score	0.76	0.75	0.72	0.74
Accuracy	**0..66**	0.65	0.64	0.64

Table XII presents the proposed trace retrieval approach using hierarchical based on featurewiz using (corr_limit values). Precision achieved highest results when is corr_limit value=0.5. The recall, F-Score, and results enhance when corr_limit values equal 0.6. Accuracy when corr_limit equal 0.5 and 0.6

TABLE XII. SMOS _ HIERARCHICAL USING SELECT FEATURES (FEATUREWIZ)

Measure's name	Standard scalar (corr_limit values)			
	0.5	**0.6**	**0.4**	**0.8**
Precision	**0.73**	0.67	0.71	**0.71**
Recall	0.75	**0.94**	**0.68**	0.76
F-score	0.74	**0.78**	0.69	0.73
Accuracy	**0.66**	**0.66**	0.62	**0.65**

Third: eANCI dataset, (Tables XIII-XVII) presents the results of the experiments obtained from DBSCAN, K-means++, and hierarchical.

The proposed trace retrieval approach using DBSCAN achieved the highest value in precision when is corr_limit values-=0.5 and 0.7. Recall F-Score,and accuracy achieved the highest value when is corr_limit values=0.7 as shown in Table XIII.

TABLE XIII. EANCI_DBSCAN USING FEATUREWIZ

Measure's name	Standard scalar (corr_limit values		
	0.5	**0.6**	**0.7**
Precision	**0.64**	0.61	**0.64**
Recall	0.71	0.53	**0.77**
F-score	0.67	0.57	**0.70**
Accuracy	0.57	0.51	**0.60**

As shown in Table XIV, the proposed trace retrieval approach using K-means++ based on standard scalar achieves highest precision with all values of the Max_feature. Recall-Score, and accuracy achieves the highest results when are Max_feature =200.

TABLE XIV. EANCI_KMEANS ++ USING MAX_FEATURE

Measure's name	Max_feature		
	100	**200**	**300**
Precision	**0.63**	**0.63**	**0.63**
Recall	0.82	**0.84**	0.84
F-Score	**0.71**	**0.72**	**0.72**
Accuracy	0.59	0.60	**0.72**

From Table XV, the authors inferred that the proposed trace retrieval approach using K-means++ based on standard scalar (corr_limit values) achieves the highest results with precision when is corr_limit value=0.5 and 0.7. Recall, F-Score, and accuracy achieved the highest results when is the corr_limit value=0.7.

TABLE XV. EANCI_ KNEAMS USING FEATURERWIZ

Measure's name	Standard scalar (corr_limit values)		
	0.5	**0.6**	**0.7**
Precision	**0.64**	0.61	0.64
Recall	**0.71**	0.53	**0.77**
F-score	**0.67**	0.57	0.70
Accuracy	**0.57**	**0.51**	0.60

According to Table XVI, the proposed trace retrieval approach using hierarchical achieves the highest results for precision, Recall, F-Score, and accuracy when is MinMax scalar = 100.

TABLE XVI. EANCI_ HIERARCHICAL USING MAX_FEATURE

Measure's name	Standard scalar = 400	Standard scalar = 100
Precision	0.62	**0.63**
Recall	**0.68**	0.77
F-score	0.65	**0.69**
Accuracy	0.56	**0.59**

As shown in Table XVII, the proposed trace retrieval approach using a GMM based on a standard scalar achieves the highest value in precision, Recall, F-Score, and accuracy when is corr_limit values = 0.5.

TABLE XVII. EANCI _GMM_ USING FEATUREWIZ

Measure's name	Standard scalar (corr_limit values)	
	0.5	**0.6**
Precision	**0.63**	**0.61**
Recall	**0.65**	**0.52**
F-score	**0.63**	0.56
Accuracy	**0.56**	**0.50**

Regardless of the results, the suitable intelligent solution for addressing the term mismatch problem between requirements and source code for a few labels or unlabelled data is UNSUPERVISED machine learning using clustering, which is the only way to solve the problem of labels and enhance low precision.

V. EVALUATION

This section presents the evaluation method of the retrieved links evaluated with respect to two criteria. The first criterion is the performance of the three measures (i.e., precision, recall, and F-score), when evaluating the traceability links between the source and target artifacts. The second criterion is the comparison of the proposed trace retrieval approach with two experiments (IR and TraceLab) as shown in Tables IV (A) to (C) and two studies ([19]and [24]) as shown in Tables V (A) to (C). The proposed trace retrieval approach achieved high results compared to the two experiments and the two studies.

VI. CONCLUSION AND FUTURE WORK

This study presented new clustering-based approach that addresses the term mismatch problem to obtain the greatest improvements in precision. The study followed the proposed research direction toward realizing automated trace retrieval by developing intelligent tracing solutions. Then the authors applied the intelligent solution that is based on unsupervised learning using clustering. After that, evaluate the proposed approach results with respect to two criteria: performance of the confusion matrix (i.e., precision, recall, F-score) and comparison of the proposed trace retrieval approach with two experiments and two studies. Clustering yields high results in precision with other measures (i.e., recall, F-score, and accuracy), which is a big concern associated with trace retrieval precision.

In the future work, the authors will look for another intelligent solution that can be applied to the same datasets (i.e., eTour, SMOS, and eANCI) for improving trace retrieval precision.

REFERENCES

[1] A. Kannenberg, & H. Saiedian, why software requirements traceability remains a challenge. CrossTalk The Journal of Defense Software Engineering, (2009). 22(5), 14-19).

[2] N.Al-walidi, A. Khamis, & N. Ramadan, "Systematic Literature Review of Recommender Systems for Requirements Engineering", The International Journal of Computer Applications (IJCA) (2020).

[3] J. Cleland-Huang, O.Gotel, C. Huffman Hayes, J. P Mäder,& A. Zisman, Software traceability: trends and future directions. In Future of software engineering proceedings (2014). (pp. 55-69).

[4] D.Hanspeter, A.Janes, A.Sillitti, & G. Succi, Improving the identification of traceability links between source code and requirements. In DMS (2012). (pp. 95-100).

[5] X.Zou, R. Settimi, & J. Cleland-Huang, improving automated requirements trace retrieval: a study of term-based enhancement methods. Empirical Software Engineering, (2010) 15(2), 119-146.

[6] H. Kim.Unsupervised learning. In Artificial Intelligence for 6G (2022). (pp. 35-86). Springer, Cham.

[7] A. Ayed, B. M. B.Halima, & AM.limi, Survey on clustering methods: Towards fuzzy clustering for big data. In 2014 6th International conference of soft computing and pattern recognition (SoCPaR) (2014, August) (pp. 331-336). IEEE.

[8] R. Petegrosso, Z. Li, & R. Kuang, Machine learning and statistical methods for clustering single-cell RNA-sequencing data. Briefings in bioinformatics, (2020)21(4), 1209-1223.

[9] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, & P. A Siano, comparative study of clustering techniques for electrical load pattern segmentation. Renewable and Sustainable Energy Reviews, (2020). 120, 109628.

[10] S.Babichev, V.Lytvynenko, &V. Osypenko, Implementation of the objective clustering inductive technology based on DBSCAN clustering algorithm. In 2017 12th international scientific and technical conference on computer sciences and information technologies (2017, September). (csit) (Vol. 1, pp. 479-484). IEEE.

[11] X. Zou, R.Settimi, &J. Cleland-Huang, Improving automated requirements trace retrieval: a study of term-based enhancement methods. Empirical Software Engineering, (2010). 15(2), 119-146.

[12] N. Ali, H. Cai, H., Hamou-Lhadj, A. J, & Hassine,. Exploiting Parts-of-Speech for effective automated requirements traceability. Information and Software Technology, (2019). 106, 126-141.

[13] D. Cuddeback, A.Dekhtyar, & J.Hayes, Automated requirements traceability: The study of human analysts. In 2010 18th IEEE International Requirements Engineering Conference (2010, October). (pp. 231-240). IEEE.

[14] C. Mills, J. Escobar-Avila, & Haiduc, S. Automatic traceability maintenance via machine learning classification. In 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME) (2018, September). (pp. 369-380). IEEE.

[15] R. Lapeña,J . Font,C. Cetina, &O. Pastor, (2018, June). Exploring new directions in traceability link recovery in models: The process models' case. In International conference on advanced information systems engineering (pp. 359-373). Springer, Cham.

[16] W.A. Zogaan, (2019). Towards an Intelligent System for Software Traceability Datasets Generation. Rochester Institute of Technology.

[17] A.I. Montoya-Munoz, &O. M. C. Rendon, (2020). An approach based on fog computing for providing reliability in iot data collection: A case study in a colombian coffee smart farm. Applied Sciences, 10(24), 8904.

[18] J. O. Palacio-Niño, & F.Berzal, (2019). Evaluation metrics for unsupervised learning algorithms. arXiv preprint arXiv:1905.05667.

[19] T.Du, B. Shen, G. H. Huang, Z. Q. Y. S. Yu, , & Wu, D. X. Automatic traceability link recovery via active learning. Frontiers of Information Technology & Electronic Engineering, (2020) 21(8), 1217-1225.

[20] J.Lin, Y. Liu, , Q. Zeng,M. Jiang, &J. Cleland-Huang, Traceability transformed: Generating more accurate links with pre-trained bert models. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE) (2021, May). (pp. 324-335). IEEE.

[21] H. Wang, Shen,Z . Huang,Y.Yu, &K. Chen, Analyzing close relations between target artifacts for improving IR-based requirement traceability recovery. Frontiers of Information Technology & Electronic Engineering, (2021). 22(7), 957-968.

[22] Y.Li, &J. Cleland-Huang, Ontology-based trace retrieval. In 2013 7th International Workshop on Traceability in Emerging Forms of Software Engineering (TEFSE) (2013, May). (pp. 30-36). IEEE.

[23] J. Guo,N . Monaikul, C.Plepel, &J. Cleland-Huang, Towards an intelligent domain-specific traceability solution. In Proceedings of the 29th ACM/IEEE international conference on Automated software engineering (2014, September). (pp. 755-766).

[24] T.Li, S. Wang, D. Lillis, &Z. Yang, combining machine learning and logical reasoning to improve requirements traceability recovery. Applied Sciences, (2020). 10(20), 7253.

[25] P. Sharma, (2017). Datasets Used in Fifteen Years of Automated Requirements Traceability Research. Rochester Institute of Technology.

[26] F. Faiz, R.Easmin, &A.U.Gias, (2016). Achieving Better Requirements to Code Traceability: Which Refactoring Should Be Done First? 2016 10th International Conference on the Quality of Information and Communications Technology (QUATIC). doi:10.1109/quatic.2016.012

[27] CoEST: Center of excellence for software traceability, http://www.CoEST.org.

[28] C. Mills,J.Escobar-Avila, Bhattacharya, A.,Kondyukov, G.S. Chakraborty, &S. Haiduc, (2019, September). Tracing with less data: active learning for classification-based traceability link recovery. In 2019 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 103-113). IEEE.

[29] R. Oliveto, M.Gethers,D. Poshyvanyk, &A. De Lucia, On the Equivalence of Information Retrieval Methods for Automated Traceability Link Recovery. 2010 IEEE 18th International Conference on Program Comprehension. (2010). doi:10.1109/icpc.2010.20.

[30] D. Falessi,M . Di Penta,G. Canfora,&G. Cantone, (2017). Estimating the number of remaining links in traceability recovery. Empirical Software Engineering, 22(3), 996-1027.

[31] A. E.Ezugwu,A.M. Ikotun,O.O. Oyelade,L. Abualigah,L.O. Agushaka, , C.I.Eke, &A.A. Akinyelu, (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.

[32] H. Bahri, & T. S. T. Mahadi, (2016). Google translate as a supplementary tool for learning Malay: A case study at Universiti Sains Malaysia. Advances in Language and Literary Studies, 7(3), 161-167.

[33] D. Khyani, B. S.Siddhartha, N. M. Niveditha, , & B. M. Divya, (2021). An Interpretation of Lemmatization and Stemming in Natural Language Processing. Journal of University of Shanghai for Science and Technolog

[34] A. Abubakar, I.A. Roko, A.B. Muhammad, &I. Saidu, Hausa WordNet: An Electronic Lexical Resource. Saudi Journal of Engineering and Technology, (2019). 4(8), 279-285.

[35] M. Khatun, Evaluating Word Embedding Models for Traceability (Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College) (2021).

[36] N. Alsuraiby, &S. Albarrak, Cosine similarity-based algorithm for social networking recommendation. International Journal of Electrical & Computer Engineering (2022). (2088-8708), 12(2).

[37] B. Dit,E . Moritz, M. Linares-Vásquez,D. Poshyvanyk,&J. Cleland-Huang, Supporting and accelerating reproducible empirical research in software evolution and maintenance using tracelab component library. Empirical Software Engineering, (2015). 20(5), 1198-1236.

[38] D.Farrar, & J. H. Hayes, A comparison of stemming techniques in tracing. In 2019 IEEE/ACM 10th International Symposium on Software and Systems Traceability (SST) (2019, May). (pp. 37-44). IEEE.

[39] J.Payne, & J. H.Hayes, (2019, May). University of kentucky tracelab component similarity matrix voting merge. In 2019 IEEE/ACM 10th International Symposium on Software and Systems Traceability (SST) (pp. 17-20). IEEE.

[40] http://selab.netlab.uky.edu/homepage/pages/TraceLabCatalogue/TracelabCatalogue/componentspage.html(2018).

Application of Multi-Scale Convolution Neural Network Optimization Image Defogging Algorithm in Image Processing

Weihan Zhu
School of Electrical and Information Engineering-Jiangsu University, Jingjiang College, Zhenjiang, 212000, China

***Abstract*—To improve the ability to detect and identify smog images in complex road traffic scenes, smog images need to be defogged, and an optimized image defogging algorithm on the basis of multi-scale convolutional neural network (MCNN) is proposed. The physical model of road traffic scene smog scattering is constructed, and the image is divided sky area, road surface area and road sky boundary area. The road sky boundary line is the boundary line between road surface and sky area. The dark channel of traffic scene smog image is established by Canny edge detection and MCNN optimization, and the smog image is subjected to detail compensation and gray enhancement processing through prior knowledge. After substituting the atmospheric light value and transmittance map into the atmospheric scattering model (ACM), the MCNN learning model is combined to realize the filtering processing and defogging optimization of smog images in complex road traffic scenes. The color saturation, defogging degree, peak signal-to-noise ratio (PSNR), texture effect as well as other aspects of the image are taken as test indexes for the simulation experiment. The simulation results show that the color saturation, defogging degree and image definition of the defogged haze images in complex road traffic scenes are higher by using this method, which improves the output PSNR of the defogged haze images in complex road traffic scenes, and has a good application value in image defogging.**

Keywords*—Multi-scale convolution neural network; complex road traffic scene; Image defogging; dark passage; Gray enhancement

I. Introduction

Nowadays, the society has entered the information age, and the way that human beings know and transform the world depends not only on language and characters, but also on images, which gradually become the main medium for transmitting real-world information. According to scientific statistics, in the process of people learning and recognizing social changes, most of the effective information comes from images, which means that image research has a significant role in production and life. In actual scenes, image semantics is often changed due to the introduction of some noises, and some directions related to computer vision are just born for solving this kind of image tasks, such as common image denoising and image enhancement, etc. This paper mainly focuses on the research of the related methods of image defogging. Recently, daily transportation, aerospace and road monitoring are all seriously affected by smog weather due to rapid development of domestic industrial production,. Fog refers to tiny water droplets suspended in the air, which are formed by condensation of water vapor in the ground layer; Haze, which consists of dust, sulfuric acid, organic hydrocarbons and other particles in the air, leads to atmosphere turbid. In foggy scenes, natural light will be absorbed or scattered by these suspended water droplets and particles in the process of transmission, which will cause the outline of the shot scene to be blurred, the image clarity and color saturation to decrease, thus affecting the overall visual effect of images. Especially when the smog concentration is high, the information of the real image will be seriously lost. Hinder the subsequent application of images. For example, in daily traffic

The main goal of image defogging algorithm is to reduce the noise caused by haze impurities to the image, increase the image's clarity and color saturation, and repair the image details. Traditional image defogging methods are mainly divided into two types. The first type is the defogging algorithm on the basis of non-physical model, which only depends on enhancing the image to enhance the image visual effect. This kind of method does not involve the principle of fog formation and the process of image imaging, but only enhances some information in the foggy image according to the actual needs. The second kind is the defogging algorithm on the basis of physical model. This method, on the basis of the ACM, estimates the undetermined parameters in the model by statistical methods to recover the fog-free image. Recently, the deep learning theory has been extensively applied in image defogging. Some scholars use a large number of foggy image data sets to train and learn, so that they can achieve defogging more efficiently.

The early defogging algorithm did not consider the objective physical model of image degradation in foggy environment, but was guided by the subjective feeling of human vision. By enhancing the contrast of foggy images, the visual feeling of images was natural and clear, and the details of images were richer. This kind of method can improve the subjective quality of images to some extent[1]. However, there is no fundamental analysis of the factors affecting image degradation, so some prominent information may cause some losses. This kind of algorithm mainly includes: histogram equalization algorithm. And wavelet homomorphic filtering algorithm, Re tinex algorithm, etc. Among them, literature [2] puts forward an image defogging algorithm for complex road traffic scenes, which adopts the smooth transition of abrupt points in the image. Because of the unknown a priori, it is difficult to build a statistical model. Literature puts forward a simple defogging model on the basis of physical laws. Through constructing the scene's physical texture features, the atomization points are virtually reconstructed. The method is simple and easy, but the accuracy is not high. Literature analyzes the defogging algorithm of atomized images collected on the water surface, and mainly adopts the defogging algorithm of complex road traffic scenes on the basis of constrained evolutionary time-frequency weighted filtering. The algorithm is limited by the scattering interference of edge fog points, so the defogging effect of complex road traffic scenes and other images is not good.

To solve the above problems, this paper puts forward a haze image defogging algorithm on the basis of MCNN. Firstly, the paper builds a physical model of haze scattering in road traffic scenes, and builds an ACM according to the sky brightness estimation and road sky line segmentation method. By using MCNN method, the largest line segment is found out based on the calculated length of each straight line segment, so as to recover the image scene and defog the image in complex road traffic scenes. Finally, the performance of this algorithm is checked by simulation experiment, indicating that his algorithm has superior performance.

II. Related Works

The work of removing smoke from traffic scene images belongs to image denoising, and people have carried out a lot of research in this field before. With the development and maturity of artificial intelligence technology, it is increasingly used in image denoising. Jian J et al. found that the image denoising algorithm based on sparse expression has good denoising effect. Therefore, a convolutional neural network algorithm based on sparse expression is designed and applied to image denoising. The experimental results show that the algorithm significantly improves the image quality and the denoising effect is significantly better than the algorithm model without sparse expression [3]. Jh A et al.'s desert images tend to contain noise caused by large amounts of sand and dust in the air. In order to improve the quality of images taken in desert areas, an improved neural network algorithm was designed. The test results show that the algorithm has better image denoising effect than the current algorithm [4]. Liu Y's research team found that the quality of polarized images is easily affected by the noise in the images obtained by the polarization camera. A denoising method using pulse coupled neural network to optimize polarization images is proposed. The calculation results show that compared with the most advanced image denoising algorithm, the proposed optimized image denoising method effectively suppresses the noise in polarized images [5]. Golilarz NA et al. designed a neural network algorithm incorporating adaptive Gaussian distribution to solve the noise problem in remote sensing satellite images. The test results show that, compared with the common denoising algorithm, the denoised image signal-to-noise ratio of this algorithm is significantly higher [6].

To sum up, in order to denoise the image to improve the image quality, people have proposed many improved image appearance processing methods, but few studies have involved removing the influence of fog caused by weather in the image, which is crucial to improve the image recognition quality of the transportation industry, which is the main purpose of this study.

III. Construction of Atmospheric Scattering Physical Model and Image Pre-Processing of Image Environment on the Basis Of GIS

A. Physical Model of Haze Scattering in Road Traffic Scenes

Firstly, a physical model of smog scattering in road traffic scenes is established[7]. When light travels, some of it will deflect due to contact with suspended particles in the atmosphere[8-9]. Besides, the deflection intensity is closely related to the suspended particle's size, type and distribution. Generally, the camera equipment receives two types of light sources: one is the reflected light of the target object. During reflected light propagation from the target object to the imaging equipment, the light received by the equipment is attenuated due to the absorption and scattering of impurities, resulting in the decrease of imaging brightness and contrast; Second, the ambient atmospheric light, the main sources of which are direct sunlight, atmospheric scattered light and ground reflected light, etc. Atmospheric light will also experience the process of being absorbed or scattered by particles, but when it propagates to the imaging equipment, the original image may be blurred because the intensity of this split light is larger than the target's reflected light. Therefore, the light intensity which is received by the imaging sensor is mainly from the superposition of the above two remaining light intensities after being influenced by impurities. They can be respectively expressed by the incident light attenuation model (ILAM) and the atmospheric light imaging model. Seen from image processing, the contrast of degraded images in foggy days is enhanced. Aiming at the characteristics of clear boundary line between highway and sky and large sky area, the image is segmented first, and the pre-processing of image defogging is realized [10]. The schematic diagram of the physical model of water surface atmospheric scattering is shown in Fig. 1.

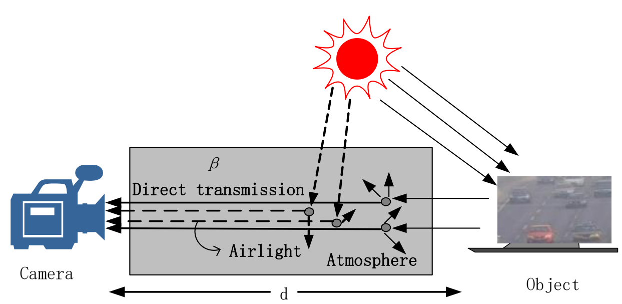


Fig. 1. Physical model of atmospheric scattering

Fig. 1 shows Narasimhan ACM. On the basis of this, a monochromatic ACM under fog and haze weather conditions is obtained. Depending on the physical model, the weather degradation image restoration method, the human eye observes the object depending on the reflection phenomenon formed when the light meets the target scene[11-12]. During the reflection process, the light will collide with the suspended impurities in the atmosphere, resulting in the deflection of some light propagation directions and attenuation of the incident light actually entering our eyes or imaging equipment[13]. At the same time, the scattering degree of the incident light is relevant to the distance between the measured object and the imaging equipment. The larger the distance, the stronger the scattering ability, and the less the light that finally enters the camera[14]. The atmospheric scattering process and image degradation process during the generation of the incident ACM are described by the model state equation:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

Wherein, A indicates the environmental light effect component of smog image in complex road traffic scene, $t(x)$ indicates the atomization transmittance of smog image in complex road traffic scene, and $J(x)t(x)$ indicates the attenuation factor of smog image atomization process in complex road traffic scene. According to the above equation, it is assumed that A represents the sky brightness, $\rho(x)$ represents the scene albedo of the smog image of the complex road traffic scene at the spatial coordinate $d(x)$, and S x represents the spatial coordinate, so that the pixel of the smog image of the complex road traffic scene captured by $I(x)$ is expressed as:

$$I(x) = A\rho(x)e^{-\beta d(x)} + A(1 - e^{-\beta d(x)}) \quad (2)$$

Wherein, β represents the scattering degree of incident light. In the complex road network environment, the transmission wavelength coefficient of smog image imaging in the atmosphere of complex road traffic scenes is:

$$L = J(w,e) - \sum_{i=1}^{N} a_i \left\{ w^T \varphi(x_i) + b + e_i - y_i \right\} \quad (3)$$

Wherein, $J(w,e)$ represents the distance between the measured object and the imaging equipment, a_i represents the unit cross-sectional thickness through which the incident light vertically passes, b represents the wavelength of the incident light, and e_i is the scene depth. When the reflected light of the target scene propagates to the imaging equipment, the impurities suspended in the natural environment will scatter the reflected light, resulting in the attenuated incident light, direct sunlight and scattered light of other objects entering the image acquisition equipment simultaneously. This process is named atmospheric light imaging. Therefore, contrary to the ILAM, the acquisition equipment in the atmospheric light imaging model can not only receive the target object's reflected light, but also have more types of atmospheric light. With the fluctuation of light and the change of probability density distribution model, the objective optimization function of atomization image processing and defogging is expressed as:

$$\left.\begin{aligned} J_1(w,e) &= \frac{\mu}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{N} e_i^2 \\ s.t.\, y_i &= w^T \varphi(x_i) + b + e_i, \quad i = 1,\cdots,N \end{aligned}\right\} \quad (4)$$

In the above formula, φ () is the wavelength kernel space mapping function of the local light spot information of the haze image in the complex road traffic scene, and $w^T \varphi(x_i)$ is the atmospheric penetration factor of the haze image in the complex road traffic scene. The partial derivative of the above formula is obtained as follows:

$$\left.\begin{aligned} \frac{\partial L}{\partial w} &= 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} &= 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} &= 0 \rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} &= 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i \end{aligned}\right\} \quad (5)$$

In the above formula, $i = 1,\cdots,N$, streamline amplification processing is performed on the smog image of complex road traffic scene in the Z-axis direction of each vertex of the entity constructed by the atomized environment model, and the LOD with the highest resolution is used to form the entity environment object model of the smog image of complex road traffic scene. During the propagation of the reflected light of the smog image of complex road traffic scene, because of the scattering effect of atmospheric particles, some reflected light on the surface of the object is lost due to the scattering of the particles of the smog image of complex road traffic scene, and the fog point scattering loss of the smog image of complex road traffic scene is obtained as follows:

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A \quad (6)$$

Wherein, A represents the opposite of the ILAM, which is collected in the atmospheric light imaging model, $I(x)$ represents the reflected wavelength of the received target object, and $\max(t(x), t_0)$ is the wavelength of the reflected light scattered by the suspended impurities in the natural environment. Therefore, through the ACM, a scene picture is formed, and the reflected light intensity will decrease exponentially with the propagation distance. On this basis, this paper adopts the defogging method on the basis of the physical model, which essentially uses the ACM to solve the scene albedo and realize the image defogging.

B. *Mathematical Model Construction and Preprocessing of Image*

On the basis of the prior knowledge, the smog image is subjected to detail compensation and gray enhancement. After substituting the atmospheric light value and transmittance map into the ACM, the MCNN learning model is combined to construct the defogging processing model of the smog image in the complex road traffic scene, and the mathematical model of the image is obtained. Collecting image information by block processing method, and selecting initial block size as a fixed value: 15*15; determining the image processing template by the image size m*n. And solving the dark primary color of the scattered atomized image in the 3*3 template of the haze image of complex road traffic scene, where $t(x)=e^{-\beta d(x)}$ indicates the propagation function or transmittance of the medium at the image position $0<t(x)<1$, so as to obtain the new features of the haze image of complex road traffic scene. The solution process is described as follows: Assume that the signal form of the haze image information of a complex road traffic scene is:

$$s(t)=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}a_{mn}g_{mn}(t) \tag{7}$$

Wherein, a_{mn} is the extra luminous flux of the imaging acquisition equipment, $g_{mn}(t)$ is haze images of miscellaneous road traffic scenes , the included angle projected in the imaging acquisition equipment is fixed, and the atmospheric light source is the cross section of the vertebral body, and the largest pixel value of the imaging acquisition is determined as the estimated value of sky brightness, which is expressed as:

$$J^{dark}(x)=\min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(J^c(y))) \tag{8}$$

Wherein, J^c represents a channel of smog image information of complex road traffic scene. A dark original smog image distribution area is set in the 3*3 template with X as the center. The edge of the smog image of complex road traffic scene is detected by Canny operator, and the smog image of complex road traffic scene is counted in blocks to find out the candidate sky area, and the intensity of the smog image of complex road traffic scene is close to 0. Based on the above processing, the haze image mathematical model in complex road traffic scene is obtained, and the atmospheric dissipation function is further defined, and the haze image in complex road traffic scene is defogged. The atmospheric dissipation function represents the extra part of ambient light to the image. In this way, the atmospheric dissipation function is simplified to $U(x)=1-e^{-\beta d(x)}$, which apparently has the characteristic resolution of the haze image in complex road traffic scene. Therefore, the captured image pixels is simplified as follows:

$$I(x)=J(x)t(x)+AU(x) \tag{9}$$

Wherein, $J(x)$ is the regular quantity of incident light attenuation of a point light source, A is the illumination amplitude and $U(x)$ is the radiance of the light source. For a color image, J^c is defined to represent a certain color component J, namely one of the three channels of RGB. The global non-significant mutation information extraction method is adopted to increase the sampling accuracy. The information sampling plane of smog image in complex road traffic scene is composed of single-layer square grid in color space. In the above-mentioned modeling coordinate system of background information of smog image in complex road traffic scene, the formula for collecting atomization information of smog image in complex road traffic scene is as follows:

$$y=\bar{y}+R_t d \quad , \quad z=\bar{z}+R_h d \tag{10}$$

Among it, R_t indicates the transmittance estimation value of the haze image in the complex road traffic scene, R_h is the estimation of the edge change pixels of the haze image Grid structure in the complex road traffic scene is the change rate of the abrupt information texture after the image is disturbed by haze. Through the above processing, the ACM is constructed according to the sky brightness estimation and highway sky line segmentation method, and the haze scattering image of road traffic scene is defogged.

IV. Scale Convolution Neural Network Method and Improved Implementation of Defogging Algorithm

A. MCNN Learning

Narasimhan and Nayar explained the elements contained in foggy images and the imaging process by establishing a mathematical model. The model considered that the reasons for the image quality degradation in foggy days mainly included two aspects: first, the energy attenuation caused by the influence of suspended particles in the atmosphere on the absorption and scattering of the target's reflected light, and finally reduced the imaging brightness; Second, atmospheric light imaging, where ambient light like sunlight is scattered by atmospheric media to form background light, and its intensity is too high, which affects the imaging clarity and causes the image color to be unnatural. Therefore, the MCNN learning method is applied to learn the image defogging. Fig. 2 shows the MCNN learning model.

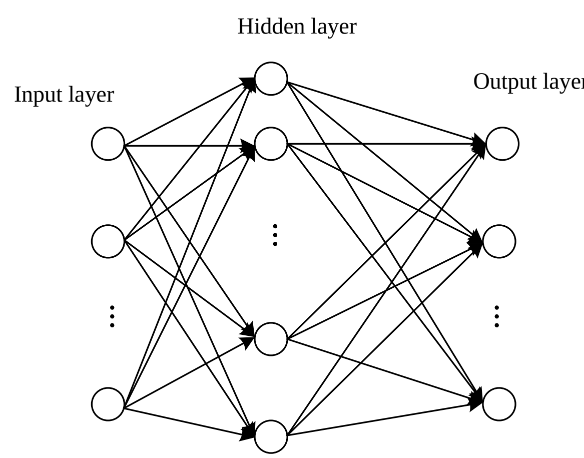


Fig. 2. MCNN

In this study, a smog image defogging algorithm on the basis of MCNN for complex road traffic scenes is adopted. The improved idea and key technology realization of the algorithm are described as follows. Firstly, the texture information features of smog images in complex road traffic scenes are calculated. Assuming that the transmittance of smog in complex road traffic scenes is the same in a local area, the 3*3 median filter is used to preprocess the smog in complicated scenes of road traffic and filter out the noise in the smog in complex road traffic scenes, and the texture information of smog in complex road traffic scenes is obtained as follows:

$$pk = \left\langle x_0, (x_i)_{0\le i\le \tau}, (x_i^{'})_{0\le i\le l-1}, (\Pi_i)_{0\le i\le l-1} \right\rangle \quad (11)$$

Wherein, x_0 is the transmittance of the window, x_i is the selected brightness in the dark channel diagram of the foggy image, Π_i is the pixel point at the same position, and the error vector of random logarithmic fluctuation is used to obtain integer vectors of haze fluctuation and cutoff as follows:

$$\boldsymbol{b} = (b_i)_{1\le i\le \tau} \in \left(-2^{\alpha}, 2^{\alpha}\right)^{\tau} \quad (12)$$

Wherein, b_i is the atmospheric light value, and the image S is decomposed into an orthogonal projection sequence of smog in complex road traffic scenes by one-dimensional wavelet decomposition method. The learning function of neural network of smog in complex road traffic scenes is:

$$c = \left[\sum_{1\le i,j\le \mu} m_{i,j} \cdot x_{i,0}^{'} \cdot x_{j,1}^{'} + \sum_{1\le i,j\le \mu} b_{i,j}^{'} \cdot \Pi_{i,0} \cdot \Pi_{j,1} + \sum_{1\le i,j\le \beta} b_{i,j} \cdot x_{i,0} \cdot x_{j,1} \right]_{x_0} \quad (13)$$

Wherein, $b_{i,j}^{'}$ is the transmitted light intensity of smog in complex road traffic scene, and for the point set in the edge center area, in a local sub-block area in any channel of the original road fog image, it satisfies:

$$\min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(I^c(y))) = \tilde{t}(x) \min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(J^c(y))) + (1-\tilde{t}(x))A^c \quad (14)$$

: Wherein, A^c represents the sky brightness of a certain color channel component, which is known from the previous sky brightness estimation algorithm as A^c, and $A > 0$. The texture features of the haze image of the complex road traffic scene in the haze obtained above are uniformly traversed and addressed, and the haze on the water surface is segmented by the highway sky line to obtain two parts, the regional difference $Dif(C_1, C_2)$ of which is larger than the internal difference $Int(C_1)$ or $Int(C_2)$ of no less than one part. In the area near the highway sky boundary line, the target search area can be reduced through detecting the highway sky line, which can reduce the subsequent target detection and recognition. Here, the estimated value of transmittance is expressed by. Because of $A > 0$, equation (14) is equivalent to:

$$\min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{I^c(y)}{A})) = \tilde{t}(x) \min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{J^c(y)}{A})) + (1-\tilde{t}(x)) \quad (15)$$

In the sky area of the complex road traffic scene smog image, the intensity value of the complex road traffic scene smog image with the smallest channel tends to the sky brightness A, that is, $\min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{I^c(y)}{A})) \to 1$, then $\tilde{t}(x) \to 0$ at this time, because of $A > 0$, so

$$\min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{J^c(y)}{A})) = 0 \quad (16)$$

Substituting the formula (16) into the formula (15) gives:

$$\tilde{t}(x) = 1 - \min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{I^c(y)}{A})) \quad (17)$$

$$\tilde{U}(x) = 1 - \tilde{t}(x) = \min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{I^c(y)}{A})) \quad (18)$$

Edge detection is used to get the smog image of complex road traffic scene, and Hough transform is used to detect the straight line, and the straight line segment is extracted. Here, the expression is the estimated value of atmospheric dissipation function.

A. Haze Image Defogging Processing in Complex Road Traffic Scenes

According to the above algorithm, the prior defogging algorithm flow of dark channel can be obtained. Firstly, the foggy image is used as input, and the dark channel map is obtained in accordance with the ACM deformation. Then, the atmospheric light value is estimated, thus obtaining the transmittance map; at last, on the premise that I(x,y) is known, the clear image is inverted by t(x) and A. The defogging algorithm can restore the image color and visibility, and at the same time, it can estimate the distance of the object by using the fog concentration, which have important applications in computer vision. However, according to the experimental statistics, if the image to be searched contains a large sky scene, the restored fog-free image will have obvious excessive areas. Realize the restoration of the image scene, and realize the image defogging processing in the complex road traffic scene. The estimated atmospheric dissipation function is:

$$U(x) = 1 - \tilde{t}'(x) = \omega\tilde{U}(x) = \omega \min_{c\in\{r,g,b\}}(\min_{y\in\Omega(x)}(\frac{I^c(y)}{A})) \quad (19)$$

In some scenes, it is difficult to extract features manually. Convolutional neural network is used to learn image features automatically. The influence weight in the field of image vision is gradually increasing. Following the example of human visual system, CNN obtains local features of data through local receptive fields, and gradually expands receptive fields by convolution kernel stacking. The restored image can be expressed as follows:

$$\begin{aligned} J(x) &= [I(x) - AU(x)]/(1-U(x)) \\ &= [I(x) - A\omega\tilde{U}(x)]/(1-\omega\tilde{U}(x)) \end{aligned} \quad (20)$$

Wherein, $I(x)$ is the convolution layer parameter, A is the input feature map size, the output feature map size size, $U(x)$ is the convolution kernel size, $U(x)$ is the abstract feature and context information. The convolution layer plays an important role in feature extraction of the input data information, and it contains multiple filters. The convolution kernel's each element has its corresponding weight coefficient and bias constant, which is similar to the forward propagation neuron. Each neuron in the convolution layer is associated with multiple neurons, and the associated area is defined as "receptive field". To restore the haze image of complex road traffic scenes more naturally, a constant ω is introduced here to improve the estimation of transmittance, where $0 < \omega \leq 1$, and

$$\tilde{t}'(x) = 1 - \omega \min_{c\in\{r,g,b\}} (\min_{y\in\Omega(x)} (\frac{I^c(y)}{A})) = 1 - \omega\tilde{U}(x) \tag{21}$$

Wherein, $\tilde{t}'(x)$ is the estimated value of the improved transmittance, and the value of ω relies on the fog concentration and the size of the sky area in the specific complex road traffic scene. For the atomized image, in the convolution layer, the step parameter S transformation can affect the size of the feature map. It can be seen that when S increases, the network parameters of the complex road traffic scene fog will decrease. However, setting the S value is complicated, and the pool layer can be used to achieve the same function, that is, the pool layer can also change the image size, so it is also called the sampling layer. It can not only reduce the data size, but also detect more abstract features and contextual information, thus enriching semantic information. The pool layer is generally used behind the convolution layer, and two ways of maximum pool and average pool are often adopted. Maximum pooling refers to selecting the largest element value from locally related elements, while average pooling refers to calculating the average value from these elements and returning it. The adjacent gray levels of image pixels obtained by constraining evolutionary conditions are:

$$P(t, f) = \int_{-\infty}^{\infty} s^*(u - \frac{\tau}{2})\alpha(\tau, v)e^{-j2\pi(vt+f\tau-vu)}dudvd\tau \tag{22}$$

In the above formula, when there is large sky area and thick fog, the light transmittance of smog in complex road traffic scenes is generally 0.75. When the fog is thick, the ω value is larger, and the value is smaller when the proportion of smog image features in complex road traffic scenes is larger in the sky area. Because the fog concentration of the smog in the complex road traffic scene in the same background has little change, we can only get the fog distribution map of one of the video frames in the same background, and use the MCNN to directly subtract the fog noise of other frames from this fog distribution map to achieve the effect of video clarity. To sum up, the algorithm implementation process is shown in Fig. 3.

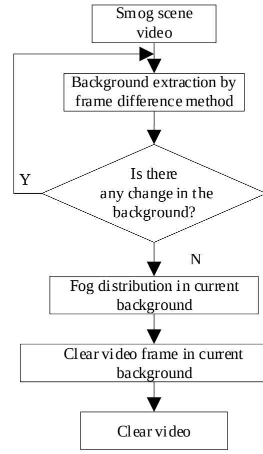


Fig. 3. Flow chart of defogging and clearing processing of road surface scattering image

V. SIMULATION EXPERIMENT AND RESULT ANALYSIS

To test the algorithm's defogging and clearing performance after collecting smog images of complex road traffic scenes under smog, a simulation experiment was carried out. The hardware environment of the simulation is PC, the operating system is Windows 7, and the system memory is 4GB. The algorithm simulation is realized by Matlab 2012. To prove the image defogging effectiveness, it is evaluated from objective and subjective perspectives. According to the subjective visual perception of the observer, evaluate the color saturation and clarity of the restored image. Evaluation indexes are used to quantitatively analyze the restoration effect of fog-free images. Subjective evaluation refers to the subjective evaluation of the defogged image by observers on the basis of its color saturation, defogged degree and texture effect. Generally, the scoring criteria for comparison between fog-free images and restored images include: whether the color is distorted, whether the image is clear, whether there is exposure problem, and whether the information of the image is lost. Image subjective evaluation can not only reflect subjective feelings, but also the evaluation results are authentic and effective, which has certain reference significance. Although the subjective evaluation is intuitive, there are still many problems, which require repeated observation of fog-free images, and it is time-consuming. Firstly, the images in the road traffic scene under the fog scattering environment are collected. Figure shows the collected original image information, which is used as a sample for image defogging. In Fig. 4, when the road empty background image is viewed from a long distance, the image is divided into sky area, road surface area and road sky boundary area. The road sky boundary line is the boundary line between road surface and sky area, and the Canny edge detection and Hough line detection are combined to obtain the image detail enhancement image in Fig. 5.

Fig. 4. Original image sample

Fig. 5. Image feature enhancement extraction diagram

Using the prior information of the sky area position, a connected component above the image is selected as the sky area to segment the road sky, and the final road sky boundary line graph is obtained, so as to pre-process image defogging. Through the above processing, on the basis of this, the fog processing of road fog image is carried out. To compare the performance of the algorithm, this algorithm is compared with the traditional Retinex algorithm and MCNN algorithm.

The analysis results in Fig. 6 show that most of the fog in the image is removed by using this algorithm, and the image is realistic with good visual effect. The algorithm in this paper has obvious defogging effect on smog, especially suitable for the processing of images with obvious road-sky boundary lines. Test the convergence of the image and get the convergence curve as shown in Fig. 7.

(a) Retinex

(b) Time-frequency weighting algorithm

(c) Algorithm in this paper

Fig. 6. Comparison of simulation results of image defogging algorithm

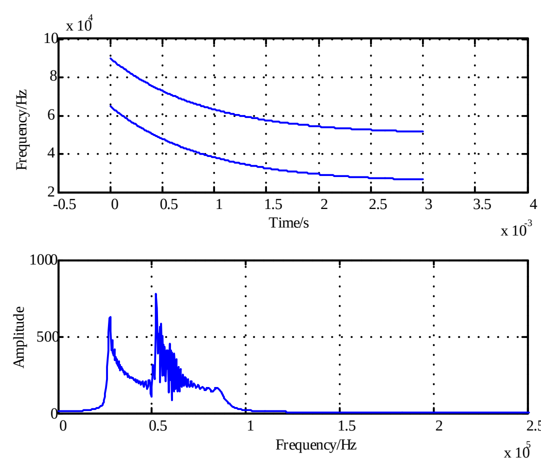


Fig. 7. Convergence test of image defogging

The output SNR is tested. The comparison results are shown in Fig. 8 and Table I. The results in Fig. 8 and Table I show that this method has higher color saturation, higher defogging degree and better image definition, which improves the peak SNR of the defogged haze images in complex road traffic scenes, and has a good application value in image defogging.

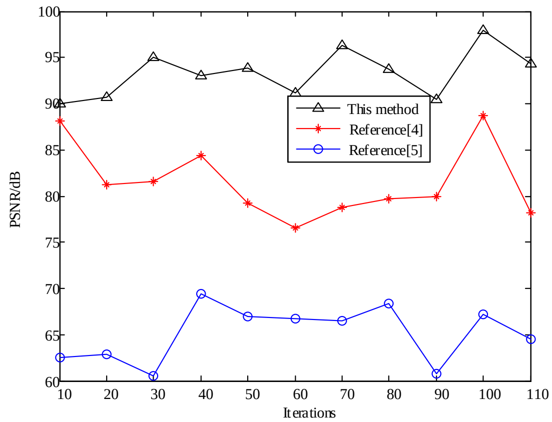


Fig. 8. Signal-to-noise ratio of image defogging output

TABLE I. OUTPUT PSNR OF IMAGE DEFOGGING (UNIT: DB)

Test frame	This method	Retinex	Time-frequency weighting algorithm
10	90.05	88.09	62.54
20	90.76	81.27	62.87
30	95.06	81.62	60.54
40	93.06	84.35	69.45
50	93.85	79.19	66.97
60	91.17	76.56	66.70
70	96.36	78.77	66.44
80	93.71	79.68	68.32
90	90.47	79.92	60.76
100	97.90	88.77	67.18
110	94.31	78.22	64.55

VI. DISCUSSION

The existence of smoke in complex traffic scenes will further reduce the image clarity, which will affect the efficiency of the image recognition system in traffic scenes. In this study, a smoke denoising algorithm for traffic scene images is proposed, which combines multi-scale idea and Canny edge detection. The test results show that, from a qualitative point of view, the part of the image that is occluded by smoke using the improved algorithm is clearer by naked eyes, and the definition and color saturation of the entire image are significantly higher than those of the image denoised by the contrast algorithm. The research results of WANGFS also show that the denoising ability of the denoising algorithm has been improved to a certain extent by using the fusion multi-scale idea [15].

From a quantitative perspective, as the number of training iterations of each algorithm increases, the peak signal-to-noise ratio of the denoised image of each algorithm gradually increases, but on the whole, the peak signal-to-noise ratio of the image processed by the improved algorithm designed in this study is the highest. When the number of iterations exceeds 100, the peak signal-to-noise ratio of the test set image after denoising based on the improved multi-scale convolutional neural network algorithm, Retinex algorithm, and MCNN algorithm is about 94dB, 81dB, and 65dB respectively. This is mainly because Canny edge detection can retain more feature information of the original image. Suyu WANG's research results show that the improved algorithm using Canny edge detection has a lower peak signal to noise ratio, which is mainly because the algorithm has too few computational levels to fully extract image features [16].

VII. CONCLUSION

In the haze scene, the images taken by the imaging equipment will have some problems. Therefore, the research of fog removal technology is of great significance. An image defogging optimization algorithm based on MCNN is proposed. In the real fog image experiment, qualitative observation with the naked eye shows that the defogging image processed by the improved MCNN algorithm can better process the details on the basis of maintaining the integrity of the image information, and the defogging effect is better than the comparison algorithm. In addition, the contrast and brightness of the real fog image can be closer to the natural and clear image. From a quantitative point of view, the improved algorithm designed in this study has the highest peak signal to noise ratio of the image after processing. When the number of iterations exceeds 100, the peak signal-to-noise ratio of the test set image after denoising based on the improved multi-scale convolutional neural network algorithm, Retinex algorithm, and MCNN algorithm is about 94dB, 81dB, and 65dB, respectively. The above research results show that the image smoke removal algorithm based on the improved MCNN algorithm proposed in this study for the application of traffic scenes has certain application potential. However, due to the limitations of research conditions, this study failed to select more data sets for experiments, which is also the part that needs to be improved in subsequent experiments.

REFERENCES

[1] Kai LI, Jie LI. Structure-fuzzy multi-class support vector machine algorithm on the basis of pinball loss[J]. Journal of Computer Applications, 2021, 41(11): 3104-3112.

[2] Xiaoqing ZHANG, Chenxi WANG, Yan LYU, Yaojin LIN. Hierarchical classification online streaming feature selection algorithm on the basis of ReliefF algorithm[J]. Journal of Computer Applications, 2022, 42(3): 688-694.

[3] Jian J , Ren F , Ji H F , et al. Generalised non-locally centralised image de-noising using sparse dictionary[J]. Iet Image Processing, 2018, 12(7):1072-1078.

[4] Jh A , Jh B , Jw A , et al. Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm[J]. Renewable Energy, 2020, 162(12):1208-1226.

[5] Liu Y , Liu Y , Hong Y , et al. An optimized pulse coupled neural network image de-noising method for a field-programmable gate array based polarization camera[J]. Review of Scientific Instruments, 2021, 92(11):113703-113706.

[6] Golilarz N A , Gao H , Pirasteh S , et al. Satellite Multispectral and Hyperspectral Image De-Noising with Enhanced Adaptive Generalized Gaussian Distribution Threshold in the Wavelet Domain[J]. Remote Sensing, 2020, 13(1):101-108.

[7] GAN J Y,ZHAI Y K,XIANG L,et al. Spatial-temporal texture cascaded feature method for face liveness detection[J]. Pattern Recognition and Artificial Intelligence,2019,32(2):117-123.

[8] GUO Qingrong, JIA Zhenhong, YANG Jie, Nikola KASABOV. Synthetic aperture radar image enhancement method on the basis of combination of non-subsampled shearlet transform and fuzzy contrast[J]. Journal of Computer Applications, 2018, 38(9): 2701-2705.

[9] LI Ping, XU Xin, DONG Hao, DENG Xu. Polarimetric SAR image feature selection and multi-layer SVM classification using divisibility index[J]. Journal of Computer Applications, 2018, 38(1): 132-136.

[10] SODHRO A H,OBAIDAT M S,ABBASI Q H,et al. Quality of service optimization in an IoT-driven intelligent transportation system[J]. IEEE Wireless Communications,2019,26(6):10-17.

[11] LENG J,LIU Y. An enhanced SSD with feature fusion and visual reasoning for object detection[J]. Neural Computing and Applications,2019,31:6549-6558.

[12] DONG Pengxin, DONG Anguo, LI Chuting, LIANG Miaomiao. Hyperspectral Image Classification on the basis of Fully Convolutional Network and Auto-Encoder[J]. Computer Engineering and Applications, 2022, 58(5): 256-263.

[13] XU Li, LU Guiming, QIU Zhenguang. Adaptive Retinex Algorithm on the basis of Detail Selection Used in Underwater Image Enhancement[J]. Computer Engineering and Applications, 2022, 58(11): 224-233.

[14] SHI Wenxu, BAO Jiahui, YAO Yu. Remote sensing image target detection and identification on the basis of deep learning. Journal of Computer Applications, 2020, 40(12): 3558-3562.

[15] WANG F S, WANG J,LI B,et al. Deep attribute learning based traffic sign detection[J]. Journal of Jilin University (Engineering and Technology Edition),2018,48(1):319-329.

[16] Suyu WANG, Jing YANG, Yue LI. Image super-resolution restoration algorithm on the basis of information distillation network with dual attention mechanism[J]. Journal of Computer Applications, 2022, 42(1): 239-244.

Data Clutter Reduction in Sampling Technique

Nur Nina Manarina Jamalludin[1], Zainura Idrus[2], Zanariah Idrus[3], Ahmad Afif Ahmarofi[4], Jahaya Abdul Hamid[5], Nurul Husna Mahadzir[6]

College of Computing, Informatics and Media, Universiti Teknologi Mara (UiTM), Selangor, Malaysia[1, 2]
Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA Kedah, Kedah, Malaysia.[3, 4, .6]
Kolej Matrikulasi Kedah,Kementerian Pendidikan Malaysia, Changlun, Kedah.[5]

Abstract*—Visualization is a process of converting data into its visual form as such data patterns can be extracted from the data. Data patterns are knowledge hidden behind the data. However, when data is big, it tends to overlap and clutter on visualization which distorts the data patterns. Data is overly crowded on visualization thus, it has become a challenge to extract knowledge patterns. Besides, big data is costly to visualize because it requires expensive hardware facilities due to its size. Moreover, it is timely to plot the data since it takes time for data to render on visualizations. Due to those reasons, there is a need to reduce the size of big datasets and at the same time maintain the data patterns. There are many methods of data reduction, which are preprocessing operations, dimension reduction, compression, network theory, redundancy elimination, data mining, machine learning, data filtering and sampling techniques. However, the commonly used data reduction technique is sampling technique that derives samples from data populations. Thus, sampling technique is chosen as a study for data reduction in this paper. However, the studies are scattered and are not discussed in a single paper. Consequently, the objective of this paper is to collect them in a single paper for further analysis in order to understand them in great detail. To achieve the objective, three interdisciplinary databases which are ACM Digital Library, IEEE Explore and Science Direct have been selected. From the database, a total of 48 studies have been extracted and they are from the years 2017 to 2021. Other than sampling techniques, this paper also seeks information on big data, data visualization, data clutter, and data reduction.

Keywords*—*Sampling technique; probability sampling; non-probability sampling; data clutter; big data; data visualization; data reduction

I. Introduction

Data visualization is a technique to convert data to a visual form to extract knowledge hidden behind the data through data patterns. According to [1], data visualization involves a combination of people with distinct visualization-related skills. Data visualization is also a technology to explore data interactively. Through data exploration, various data patterns can be revealed.

Big data plays a bigger role in our latest technologies today. Communities are particularly depending on the data to gain more information for decision making [2]. The advantage of data visualization is that it supports analysis, identifies issues and tackles problems faster through data patterns [3].

Big data technology is designed to process an enormous dataset for process optimization and decision making [4, 5]. However, an enormous dataset, both structured and unstructured are complex as they deal with an extensive amount of data. Thus, consistently, they are inadequate to operate with conventional processing techniques and algorithms [2, 6]. This is true when data is from various sources with various forms and format and yet they need to be integrated prior to processing.

Other challenges when dealing with big data are the effectiveness and efficiency in understanding, storing, managing, and developing data visualization [7]. Plotting these big data to form visualizations, require high end and expensive hardware and software facilities.

Nevertheless, when data have been successfully plotted and converted to visualization forms, it is common for the data to overlap on top of each other which could lead to data clutter issues. Data clutter can be defined as data that overlapping on top of each other which can lead to a massive number of false detections over the search space that relies on pixel patterns [8]. Fig. 1 below show the example of data clutter.

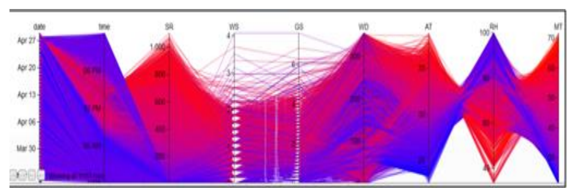


Fig. 1. Example of data clutter (Source: [51])

Another issue of concern regarding data clutter is plotting efficiency. It takes a longer time to plot data into its visual and is defined as computational overhead which is costly [9].

Data clutter also leads to unrecognized data patterns when in fact extracting the patterns is the main objective of data visualization [10] for strategic planning and decision making. In other words, the whole point of data visualization is to extract data patterns in order to uncover the gems of knowledge hidden behind the data.

Thus, there is a need to overcome data clutter and one of the techniques is through data reduction. Data reduction is one of the methods to shrunk computational overhead [11]. Although the dataset is reduced, the original information in the dataset should be preserved without scarifying any data patterns. However, data reduction could somehow remove some information from an original dataset that can lead to an unknown output of the dataset [12]. Nevertheless, data reduction can solve the difficulties that both data and visualization scientists suffer [13].

By reducing data in a dataset, data visualization can be more comprehensible with clearer data patterns. With proper data processing, data reduction can generate accurate visualization without changing data patterns [14].

The sampling technique is about choosing a subgroup from a large dataset and at the same time maintaining its properties or attributes [15, 16]. The sampling technique is a crucial way to analyze a massive dataset where its size is reduced for effective use of equipment and space [17].

Moreover, sampling techniques can lower data error that happens due to human factors when dealing with a large dataset [18]. The sampling technique is an outstanding technique to handle large datasets and when the resources are restricted. It generates results rapidly and accurately as data are smaller in size [19]. Normally, data reduction is implemented along with a machine-learning technique to analyze the size of the dataset to find an accurate output [14].

During analysis, it is common to visualize and analyze smaller datasets at a time as it is easier to identify data patterns [7, 20]. On the other hand, a larger dataset is explored at an early stage of data analysis to get a bird eye view and to identify interesting patterns for further analysis [13, 21]. Besides that, a larger dataset is used to understand the structure and the flow of the dataset [13]. However, it is burdensome for the researchers and the algorithms if a large dataset is used to discover data patterns.

Various techniques have been used for data reduction such as pre-processing operations, dimension reduction, compression, network theory, redundancy elimination, data mining, machine learning, data filtering and sampling technique However, in this study, the sampling technique is chosen as a data reduction technique. It is chosen because it is a well-known technique and consistently gives a positive result in scaling down the number of massive datasets [11].

However, there are various types of sampling techniques. Each is with its own strength and weakness. However, the discussion of these sampling techniques is scattered and not in a single paper. It is inconvenient to gather information from different papers. Thus, this review paper combines various sampling techniques into a single paper to ease their comparison and contrast for further analysis. Besides, the analysis could be used in choosing the right sampling technique for optimum outcome.

To be specific, the objective of this review paper is to understand the various sampling techniques for data reduction. This study is intended to find the answers to three main research questions (RQ) which are: RQ1: What are the various types of sampling techniques to reduce data clutter? RQ2: What are the distinct behaviors of these sampling techniques? Finally, RQ3: What is the outcome of these techniques in various dataset and applications?

To achieve the objectives, this paper starts with an introduction, followed by research method. Then, the paper continues with the synthesis of results, followed by discussion. Finally, the last section summarizes and presents conclusions.

II. Research Method

The aim of this paper is to explore the various types of sampling techniques that can assist in encountering the problem of data clutter and provide a snapshot to a direct future design and research.

Since there are huge research papers available in various areas of research, they need to be filtered in order to focus only on related papers that give most input. Thus, this section focuses on the three steps on how the filtering process is designed.

The first step is to identify research questions, followed by a description for each of the questions. This phase is vital as it gives direction to the filtering process. Next, is to identify related keywords based on the research questions. These keywords become the basis for research paper filtering. Three online databases have been identified that are suitable for cluster analysis. They are ACM Digital Library, IEEE Explore and Science Direct. Lastly, inclusion and exclusion criteria for the research paper are identified. These criteria are used to further filter research papers based on the keywords selected. It is to ensure that the research papers are qualified and within the scope of the research.

A. Research Questions

The first step of the research paper filtering process is to formulate research questions. Thus, to achieve this paper's objectives, it has to answers three research questions as depicted in Table I.

TABLE I. THE RESEARCH QUESTIONS ADDRESSED

Research Questions	Motivation
RQ1: What are the various types of sampling techniques to reduce data clutter?	To identify various sampling techniques that have been used to reduce data.
RQ2: What are the distinct behaviors of these sampling techniques?	Identify the distinct behaviors or attributes of these sampling techniques and how they are different from each other.
RQ3: What are the outcomes of these techniques in various dataset and applications?	Find out how these sampling techniques are being applied in previous research to view it from a bigger perspective.

B. Search Strategy

This paper covers English language articles that have been published from 2017 to 2021. The primary collection method is through online databases. Three popular online research databases have been selected which are ACM Digital Library, IEEE Explore, and Science Direct.

The next step is to identify keywords. There are several sets of keywords that have been built based on the research questions.

Next, the keywords are embedded with Boolean operators which are 'OR' and 'AND' operators. Table II shows the research questions and their related keywords.

TABLE II. KEYWORDS FOR RESEARCH QUESTION

Research Question	Keywords
What are the various types of sampling techniques to reduce data clutter?	"Types" OR "Sampling techniques" AND "Reduce" AND "Data clutter" OR "Data overlap"
What are the distinct behaviors of these sampling techniques?	"Distinct" OR "Behaviors" AND "Sampling techniques"
What are the outcomes of these techniques in various dataset and applications?	"Sampling Techniques" AND "Application"

After the keywords have been built, they are then used to filter the three online databases. The first search string retrieves 85,729 search results on ACM Digital Library, IEEE Explore returns 97,668 search results and Science Direct fetches 2,218,712 results. The second search string receives 138,341 from ACM Digital Library, IEEE Explore returns 241,709 results and Science Direct fetches 2,648,324. The last search string receives 144,579 from ACM Digital Library, IEEE Explore fetches 680,554 results and Science Direct receives 11,397 results. The total for all keywords in three different online databases is 6,267,013 articles.

C. Inclusion and Exclusion Criteria

Once the databases have been filtered, the next stage is to further filter the articles through inclusion and exclusion criteria. This step is to ensure the articles fall within the paper scope. Table III shows the inclusion and exclusion criteria for this research.

TABLE III. CRITERIA OF INCLUSION AND EXCLUSION

Inclusion Criteria	Exclusion Criteria
The research papers published in the English language are included	Papers written other than English language are not included
Primary studies like original research papers are selected	Papers that fail to answer the research questions are excluded
Research papers, book chapters that are relevant to the main topic are selected	Elimination of duplicated papers
Research papers ranging from 2016 to 2021 are included in the studies	Research paper less than two pages are removed

The inclusion and exclusion criteria are applied to the 6,267,013 articles filtered earlier. Then, the references from previous systematic reviews are also added, in order to collect as many papers as possible. Finally, a total of 52 papers have been identified suitable for the review. From the 52 articles, they are then grouped into their related sampling techniques. Fig. 2 shows the tabulation of the papers and their topics on sampling techniques.

The total number of papers on simple random sampling is nine while six papers focusing on systematic random sampling. Stratified random sampling is found in seven papers. Moreover, multi-stage sampling has been discussed in three papers while convenient sampling appears in four papers as well as seven papers on snowball sampling. Lastly, quota and judgmental sampling are each discussed in other two papers.

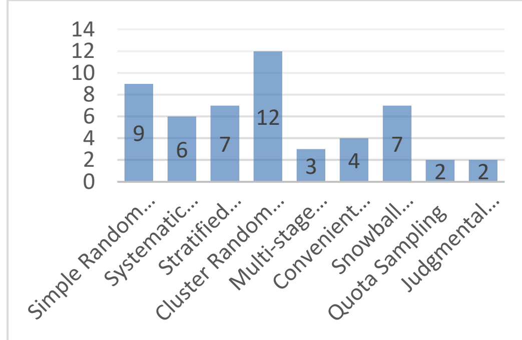


Fig. 2. Histogram for the articles used for each sampling technique

III. SYNTHESIS OF RESULTS

This section answers the research questions as depicted on Table I, by synthesizing and analyzing the knowledge collected from the filtered research papers.

A. RQ1: What are the Various Types of Sampling Techniques to Reduce Data Clutter?

Sampling technique can be divided into two types which are probability sampling and non-probability sampling. Probability sampling is also known as random sampling or representative sampling.

Unlike probability sampling, non-probability sampling does not have a random selection of the sample.

1) Probability sampling: Probability sampling chooses the sample data randomly. The data population must be defined precisely to make the sample selection process easier. The advantage of probability sampling is that it can reduce the chances of systematic error during the sampling process [22]. A systematic error is caused by incorrect measurement of data. Moreover, the sampling technique under-probability sampling can reduce bias that is a common problem in sampling technique [22]. It can select sample data fairly without bias. Moreover, sampling techniques in probability sampling can generate a better sample as compared to non-probability sampling [22].

However, probability sampling demands a lot of training. In addition, to ensure the sample data collected is fair and generates the desired outcome, some calculations need to be applied to the technique. Other than that, probability sampling is timely to generate samples because of the calculation and the implementation of the technique happens layer by layer [23]. There are a few techniques that fall under probability sampling which are simple random sampling, systematic random sampling, stratified random sampling, cluster sampling, and multi-stage sampling.

2) Simple random sampling: One of the sampling techniques that fall under probability sampling is simple random sampling. Simple random sampling is a sampling where each data in the population has a similar chance to be selected as the sample. Each data in the population can participate to be selected as a sample without exception. Usually, the researcher uses computer-generated random numbers to select the sample. Simple random sampling is suitable to be used when the entire population is available and the researcher has a list of all subjects from the population [23]. Fig. 3 below shows how the simple random sampling selects samples from the population.

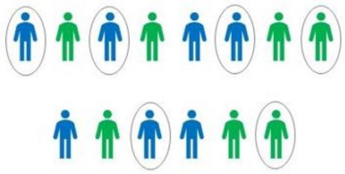


Fig. 3. Selection of sample using simple random sampling

However, using simple random sampling techniques may reduce the possibilities of bias. It is because the dataset is selected randomly, hence there is no bias. The sample from the population is a righteous representative of the whole population [22]. For example, if there exist a few categories in the dataset, all the categories will reform to represent the whole categories. Simple random sampling lowers the vulnerability suitable to the finite size of the sample [23].

Simple random sampling might be pricey and time-consuming if the population involved a broadly spread geographical location. It needs a lot of attempts when a dataset used is large [22]. However, if there is a minority in the population, they will be diminished.

3) Systematic random sampling: The next technique of probability sampling is systematic random sampling which makes use of a specific formula to select the sample. The formula is also known as regular interval [24]. The process of data selection is initiated by selecting a random data and then the selection continues at regular intervals [23].

A population can be described based on any characteristics that are suitable for the studies. The characteristics can be age, gender, race, location, and others, as long as there are different characteristics.

Fig. 4 below shows how systematic random sampling selects the sample from the population.

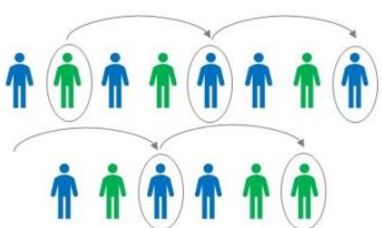


Fig. 4. Selection of sample using systematic random sampling

Systematic sampling is to ensure that the whole population participates in the sampling selection and there is no exception [22]. Thus, there is certainly a sample from each data category since all the categories are involved. Similar to simple random sampling, systematic sampling chooses the sample randomly and the location of the sample is not important as long as all the elements are included [22]. Systematic sampling is less expensive [24].

Due to the calculation that is needed for this technique, it might be timely and requires lots of effort especially if the population involved is scattered in the widely spread geographical area and it is also a challenge to access the population [22].

4) Stratified random sampling: One of the sampling techniques that uses strata or subgroups of the population is stratified random sampling. Stratified random sampling uses a subgroup to give an equal possibility to select data randomly from the strata [21]. The population is grouped into similar characteristics and the sample is selected randomly from the subgroup [22]. Stratified sampling is derived from the simple random sampling; hence the sample frame is needed. Fig. 5 below shows how stratified random sampling is selected from the population.

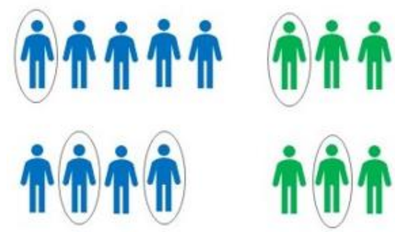


Fig. 5. Selection of sample using stratified random sampling

The advantage of stratified random sampling is the researcher can collect samples from each of the strata and the sample size will be different from each stratum [22]. Stratified random sampling also collects samples from the minority population, hence there is no exception between the majority population and minority population [22]. Thus, stratified sampling can represent the actual data population [25].

However, stratified random sampling is costly, timely and requires a lot of effort due to the subgroup process and randomly selected samples [20, 25]. If the population is not sub grouped into the same characteristic, the entire research may be useless [20]. Hence, before proceeding to select the samples, the researcher must make sure that the population has been sub grouped correctly based on the characteristics.

5) Cluster sampling: Cluster sampling is a sampling technique where the samples of the population are from a geographical area that is spread and possible to be accessed simultaneously [21]. The researcher splits the population into clusters based on the geographical area and later extracts the sample from the cluster. Cluster sampling is similar to simple random sampling as the samples are randomly selected from the cluster. Fig. 6 shows how samples are selected from the population using cluster sampling.

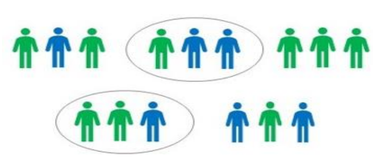


Fig. 6. Selection of sample using stratified random sampling

If the population is widely spread over a geographical area, cluster sampling can lessen the cost, time, and efforts compared to other sampling techniques [26]. Cluster sampling requires less effort and time because the population can be visited once. Moreover, cluster sampling does not need to define the number of clusters prior to processing [27]. The number of clusters is defined after the whole population has been clustered. According to [28], cluster sampling does not require additional information while the algorithm is applied to the population. By using hybrid models, it can enhance the variances for the data [29].

However, cluster sampling might not perform the true diversity of the population [30]. Moreover, biases and systematic errors could happen sometimes. In addition, after the cluster sampling has been applied to the population, it might be challenging to do a correction to the sample [27, 28].

6) Multi-stage sampling: The last sampling technique of probability sampling is multi-stage sampling. Multi-stage sampling is similar to cluster sampling where the population is sub grouped into the same cluster and a sample from the cluster is selected in the next process [21].

The advantages of multi-stage sampling are time and cost efficiency. This is because the population is sub grouped into the same category and samples from the subgroup are selected at a later stage [22] which reduced the process flow. The disadvantage of multi-stage sampling is the sample does not represent the population if the selected clusters do not capture the characteristics of the population [22].

7) Non-probability sampling: Another type of sampling technique is non-probability sampling where selection is not random. The dataset for this type of sampling does not need to be precisely defined. Unlike probability sampling, non-probability sampling can be used either for specific and general categories.

The advantage of using the non-probability type of sampling is that they require less effort and less time to generate the sample [22]. However, their disadvantage is that they are easily exposed to systematic errors and bias issues [20]. Thus, the samples at times might not be an accurate representation of the population [20].

There are a few sampling techniques that fall under non-probability sampling, which are convenient sampling, snowball sampling, quota sampling, and judgment sampling.

8) Convenient sampling: Convenient sampling is the process of selecting sample subjects based on their availability and accessibility [24]. It is common for researchers to conduct interviews from the available pool of respondents as shown in Fig. 7. It is not the researchers' concern if the selected data fail to represent the population.

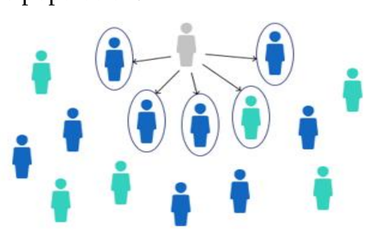


Fig. 7. Selection of sample using convenient sampling

With this kind of method, data could be collected through online channels where respondents are those who are willing to spend time in the data collection process [31]. Thus, it takes less effort. However, convenient sampling is easily facing problems of biasness and systematic error.

9) Snowball sampling: Another non-probability sampling technique is snowball sampling. Snowball sampling is where the researcher makes the first arrangement with a small group of people that are relevant to the subject and uses them as the criterion to contact other people [21]. In other words, data are collected form a small group of respondents and through them, more responded are identified as illustrated in Fig. 8.

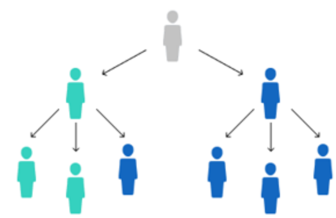


Fig. 8. Selection of sampling using snowball sampling

The technique is suitable when the population is not located in a particular area [24]. It is also suitable for scarce and very small population [22]. Thus, there may occur biases and systematic errors due to non-random network connection [20].

10) Quota sampling: Quota sampling is also categorized as non-probability sampling. Quota sampling is commonly used if the elements of the population is not matched with another characteristic of the criteria that has been defined. The population is sub grouped into its same elements and the quota is set for each subgroup.

The advantage of quota sampling is every single subject in the population has its own subgroup. Compared to stratified sampling, quota sampling is less time consuming and inexpensive [22]. However, quota sampling might not be the best method to represent the whole population. Hence, it cannot counter issues where generalizability needs to be made [20].

11) Judgmental sampling: Judgmental sampling is the sampling where the subjects of the population are selected by the researcher [20]. The process starts with the researcher generally evaluating the population's characteristics. From there, samples are selected with the aim that they represent the whole population as illustrated in Fig. 9.

There are not many researchers who use this sampling technique because the researcher's judgment might be biased [22].

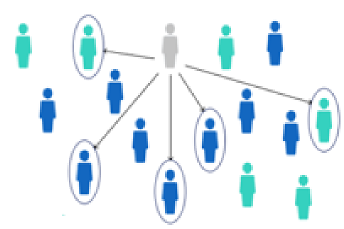


Fig. 9. Selection of sample using judgmental sampling

In short there are various sampling techniques, and they can be grouped into probability and non-probability. Each has its own behaviors and characteristics. Thus, the next section answers the RQ2 where the sampling techniques differences are analyzed.

B. *RQ2: What are the Distinct behaviours of these Tecniques?*

To answer research question 2 (RQ2) about the distinct behaviors of the sampling techniques, this section compares and contrasts these sampling techniques. Table IV shows their comparison.

TABLE IV. COMPARISON OF SAMPLING TECHNIQUES

Sampling Techniques \ Behaviors	Use sample	Costly	Bias problem	Timely	Sample represent population	Lots of efforts	Systematic error	Size of population
Simple Random Sampling	Yes	Depends	No	Depends	Yes	No	-	Large
Systematic Random Sampling	Yes	No	-	Yes	Yes	Yes	-	Large
Stratified Random Sapling	Yes	Yes	-	Yes	Yes	Yes	-	Large
Cluster Sampling	Yes	No	Yes	No	Yes	No	Yes	Large
Multi-stage Sampling	Yes	-	-	-	-	-	-	Large
Convenient Sampling	No	No	-	-	-	No	-	From small to large
Snowball Sampling	No	No	Yes	Yes	-	Yes	Yes	From small to large
Quota Sampling	Yes	No	-	No	No	No	-	Small
Judgmental Sampling	No	No	-	No	-	Yes	-	From small to large

The comparison involves a group of probability and non-probability sampling techniques which have been introduced in the earlier sections. The comparison involves a few attributes which are sample, cost, bias problem, time-consuming, population representation, efforts used, systematic error and size of population.

In terms of costing, stratified technique is more costly as compared to others. On the other hand, costing for simple random sampling depends on the dataset size. The bigger the size the more cost involved. Sampling techniques that are less costly are systematic random sampling, cluster sampling, convenient sampling, snowball sampling, quota sampling and judgmental sampling.

Next attribute that is important in the area or sampling is bias. Sampling techniques that are prone to bias are cluster sampling and snowball sampling. Meanwhile, simple random sampling is the least with the issue of bias.

Timely is another attribute for comparison. Sampling techniques that are timely to be executed are systematic random sampling, stratified random sampling and snowball sampling. On the other hand, sampling techniques that are not commonly related to timely issues are cluster sampling, quota sampling and judgmental sampling. While simple random sampling depends on the size of population or dataset. The bigger size, the longer sampling process.

In terms of accurately representing population, the sampling techniques are simple random sampling, systematic random sampling, stratified random sampling and cluster sampling. Meanwhile, quota sampling is not included in this group of accurately representing the population.

Besides, effort is another attribute of sampling techniques that differ from each other. Systematic random sampling, stratified random sampling, snowball sampling and judgmental sampling require more effort for implementation. On the contemporary level, simple random sampling, cluster sampling, convenient sampling and quota sampling are categorized as less effort when it comes to implementation.

Systematic error is another vital attribute of sampling technique that needs to be compared. Cluster sampling and snowball sampling are prone to systematic error. While for other sampling techniques, it could not be identified whether they are prone to systematic error.

Lastly, population size varies with sampling techniques. Different techniques are suitable for different population sizes. Techniques that use large populations are simple random sampling, systematic random sampling, stratified random sampling, cluster sampling and multi-stage sampling. On the other hand, quota sampling is suitable for a small population. However, convenient sampling, snowball sampling and judgmental sampling need two steps. Thus, the first step is to make use of the small population and the population will grow with time.

C. *RQ3: What is the Outcome of these Techniques in Various Dataset and Application*

This section is to identify the various applications of the sampling techniques based on the previous research.

The discussion starts with probability sampling and continues with a non-sampling group of techniques.

1) Probability sampling: This section discusses the application of various probability sampling in various areas.

2) Simple random sampling: Many previous studies used simple random sampling as the technique in their studies. According to [32], to find the random information that is adjacent to the expected distribution, the anticipated distribution and the sampling distribution are extricated frequently by using simple random sampling. The purpose of extracting frequently is to make sure that the population is placed randomly and not follow any arrangement. By combining the simple random sampling technique with Arithmetic Mean which is then known as Random Sampling-Arithmetic Mean (RS-AM), it could reduce the data conflict and increase the sampling accuracy [32]. The benefit of using RS-AM is that an efficient calculation method could be achieved if the distance function is used effectively [32].

Based on [33], block pool is the output of I-sampling. It is also known as the block-based sampling method. It is used with data that has almost similar probability distribution.

The sampling process start with splitting a large dataset into non-overlap data. Next, block pools are created. Finally, the data blocks are randomly be selected from the block pool. This is to ensure data are selected fairly and without bias.

In addition, by using distinct data sizes to randomly extract the documents especially in medical, the chances of certain words appearing in the documents and sentences are high. Identifying the distance between target words in different dimensions is used to categorize the documents and to differentiate the documents. Each alphabet has its own size and distance between each stroke. For the word centroids distance training, random sampling is applied to verify the accuracy of the output and the variety of the selected documents [30, 48].

Simple random sampling can also be used to allocate resources to the network and all subsystems in the Markov chain [31, 34]. It has been proven that simple random sampling can be used in the network and not limited to dataset. Due to the power supply and network restraint, it is suitable to make a schedule especially a wireless network system that incorporates the immense number of nodes. Scheduling is used to ensure the stability of performance specification in the network systems. Scheduling performs independent control loops based on the requirement and overall network supplies.

3) Systematic random sampling: Systematic sampling technique is used in an alternative method to estimate the Banzhaf-Owen value in a large class of TU games with scientific [32, 34]. The systematic sampling technique is suitable with the arrangement of a priori unions. Systematic sampling is used as a technique in this study because this technique can handle memory size issues and can reduce massive situations.

In addition, Stochastic bilevel programming is a program that has randomness in the problem, hence it will face problems with computationally expensive and challenging [33, 34]. The randomness that happens in bilevel programming is the randomness property and hierarchical nature of optimization. Systematic sampling is chosen as the algorithm because this technique will determine a representative from the leader's opinion and hybrid particle swarm optimization procedure.

Another application of systematic sampling is to measure the correlation of two objects. The correlation is defined as the max-min distance between two points set and could identify the parallel between two points [34, 48]. A Measure two-point set is known as Hausdorff distance or HD. HD is difficult to figure out because of a very massive scale point set and at the same time to assure the certainty of HD. Systematic sampling is chosen as the algorithm because it spends less time to choose the sample and achieve maximum dispersion of the sample and the starting point of this sampling is selected random samples [34, 48].

4) Stratified random sampling: Stratified sampling is a widely used tool for variance reduction used for failure probability estimation. This study merges the stratified sampling with importance sampling and stratified importance sampling. Stratified sampling is used to reduce the samples to approximately a failure probability with the same coefficient of variation [34, 35].

The common use of stratified sampling is to subgroup the population [34, 36]. Then samples are selected from the subgroup. The sample size is to resolve the strata by acknowledging the compatibility between sampling design and load research objectives. The population from the study consists of electricity tariffs, contract power, geographical area, and region type. Stratified random sampling subgroups the population into the same categories.

In the study conducted by [37, 48], stratified sampling is used to preprocess the wastewater condition dataset that has a sophisticated nonlinear relationship, performance, and MLR models that are not good. The researcher partitions the dataset into different subgroups and selects data points from distinct strata for different purposes. Stratified random sampling trains and tests the dataset that contains the same proposition of each class label [37].

5) Cluster sampling: A study conducted by [38] stated that the cluster sampling technique is applied to get a sample from Java software that consists of a similar system and to display the differences between the clusters. The software is grouping into the cluster using the CrossSim algorithm to observe the similarities. After finding out the similarities, the software is clustered using cluster sampling. Lastly, to extract the description of the project and group the systems based on the extraction, Python implementation of the Latent Dirichlet Allocation (LDA) is used.

Moreover, the clustering technique can be used as an alternative technique for subdividing the input space especially involving the high-dimension input spaces and to model MFs into partitions. This study uses the cluster sampling approach initialized ANFIS and MF which can take the entire advantage of intrinsic data distribution. The cluster sampling is used to develop ANFIS nearest-neighbourhood and to allow the online generation of advanced rules by excluding the nearest-neighbourhood that is not effective anymore [20, 39].

On the other hand, conditional cluster sampling is used as the approach for the current pandemic around the world that is COVID-19. This study is conducted by [40] where conditional cluster sampling is used to test patients in pools rather than individual testing by using a numerical method, statistical data, and machine learning. After the output has been generated for each pool, the researcher makes a decision either to continue the testing or abort it. Conditional cluster sampling is applied to cluster the population depending on the patients' condition. The major reason behind this study is the possibility of COVID-19 critical patients is higher than other patients' diseases.

Lastly [41] stated that the cluster sampling is not efficient as SRS, but it is cost-efficient in the statistical information that is to come across a wide geographical area. Cluster sampling is the most economical when a group of the population element establishes a sampling unit than a single element [41].

6) Multi-stage sampling: There are a few studies that used multi-stage sampling as the approach. One of the studies is conducted by [42], where the multistage sampling is the extension to the acceptance sampling technique where the inspection happened several times and only accepted if it travelled as such it covered all stages. This approach is called the multi-stage acceptance technique when the starting test is pursued by the next inspections.

7) Non-probability sampling: This section discussed the non-probability type of sampling technique.

8) Snowball sampling: The study conducted by [43] stated that snowball sampling is used when it is difficult to approach subjects with distinct characteristics. The study used qualitative research which coordinates the approach of describing people's experiences and internal feelings. This research collects data with a different approach such as interviews, observations, focus groups, narratives, notes, reports, and a review of archives. Snowball sampling assembles the information to approach specific groups of people.

Moreover, [44] using snowball sampling to create samples from simulated networks and cut down the distance from network statistics across network sizes. In this study, snowball sampling generates samples with an identical number of waves and seeds as the samples taken from simulated population networks.

Besides, snowball sampling is used to identify the effectiveness of assumptions about the existence of effects such as network closure and attribute homophiles [45]. This study uses snowball sampling to generate specimens of nodes in a network by applying the network structure itself that can be represented as follows.

Lastly, snowball sampling starts with the entity that has preference characteristics and uses that individual's connection to attract other people with the same characteristics [46]. This study uses a snowball sampling to gain information about the mothers with children that have developmental disabilities. Mothers are requested to pass the information to other mothers that might have a child that suffered from developmental disabilities. However, previous studies for judgmental sampling, quota sampling, and convenient sampling are quite difficult to find.

IV. Conclusion

There are two main types of sampling techniques which are probability sampling and non-probability sampling. The sampling techniques that fall under probability sampling are simple random sampling, systematic random sampling, stratified random sampling, cluster sampling, and multi-stage sampling.

The population involved in the probability sampling can be large as the sampling techniques take the samples directly from the population. Hence, the large population should ensure that the output generated is correct and accurate. Each sampling technique in probability sampling has its methods or calculation to take samples from the population.

The sampling techniques for the non-probability sampling are convenience sampling, snowball sampling, quota sampling, and judgment sampling. The majority of the sampling techniques in the non-probability sampling are suitable for small populations. Thus, most researchers use this sampling technique in their research because of the population size.

On the other hand, non-probability sampling is not quite popular among researchers because of the small population. This can lead to inaccurate output. Therefore, in this paper, there are sampling techniques in non-probability sampling that do not have previous studies. Besides, non-probability sampling majorly uses primary data which could become a burden to researchers.

Probability sampling is usually used by the researcher because the data can be primary and secondary. Hence, the researcher can choose how to gather the data (primary or secondary). Moreover, probability sampling uses a large population that can lead to accurate results because of the number of the population involved in the research.

To make researchers to understand the pattern well, researcher can visualize the graph. One of the well-known visualization graphs is parallel coordinates as it can be used for large and multi-dimensional data set visualization [46, 47]. One of benefits of using parallel coordinates graph is the ability to identify the relationship of multivariate data [20].

References

[1] Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. "Data Changes Everything : Challenges and Opportunities in Data Visualization Design Handoff.", *IEEE Transactions On Visualization And Computer Graphics*, *26*(1), 12–22, 2020.

[2] Taylor-sakyi, K. "*Big Data : Understanding Big Data*.", January 2016.

[3] Arockia Panimalar.S, Komal M.Khule, Karthika.S, N. K. "Data Visualization Tools and Techniques For Datasets In Big Data., *International Research Journal of Engineering and Technology(IRJET)*, *4*(8). Retrieved from https://irjet.net/archives/V4/i8/IRJET-V4I8296.pdf, 2017

[4] Babu, A. G. L., Reddy, S. G., Agarwal, & Swathi. " An effective approach for Visualizing Big Data.", *International Journal of Innovations in Engineering and Technology*, *7*(2), 77–81,2016.

[5] Samuel, S. A., & Anthonia, A. "An overview of big data visualization techniques in data mining.", *International Journal of Computer Science and Information Technology Research*, *4*(3), 105–113. Retrieved from https://www.researchgate.net/publication/305905594, 2016

[6] Bryner, D., Huffer, F., Rosenthal, M., Tucker, J. D., & Srivastava, A. "Estimation of linear target-layer trajectories using cluttered point cloud data.", *Computational Statistics and Data Analysis*, *102*, 1–22. https://doi.org/10.1016/j.csda.2016.04.002, 2016

[7] Idrus, Z., Zainuddin, H., Ja'afar, A.D.M.: Visual analytics: designing flexible filtering in parallel coordinate graph. J. Fundam. Appl. Sci. **9**, 23–32. https://doi.org/10.4314/jfas.v9i5s.3, 2019

[8] Li, T., De la Prieta Pintado, F., Corchado, J. M., & Bajo, J. "Multi-source homogeneous data clustering for multi-target detection from cluttered background with misdetection.", *Applied Soft Computing Journal*, *60*, 436–446. https://doi.org/10.1016/j.asoc.2017.07.012, 2017

[9] Yongjoo Park, Michael Cafarella, B. M. "*Visualization-Aware Sampling for Very Large Databases*.", *2*, 5,2015.

[10] Bum Chul Kwon and Janu Verma, Peter J. Haas, C. D. "Visualization Viewpoints.", *IEEE Computer Society*, (February), 100–108, 2017.

[11] Li, S. "*Data Reduction Techniques for Scientific Visualization and Data Analysis"*, Vol. 36, 2017.

[12] Rojas, J. A. R. "Sampling Techniques to Improve Big Data Exploration.", *IEEE Symposium on Large Data Analysis and Visualization*, 26–35, 2017.

[13] Bhardwaj, P. "Types of Sampling in Research.", *2019 Journal of the Practice of Cardiovascular Sciences*, 157–163. https://doi.org/10.4103/jpcs.jpcs, 2019

[14] Iqbal Jeelani M, F. D. and M. G. "A Review on the Recent Development on the Cluster Sampling.", *Biostatics and Biometrics Open Access Journal*, *5*(5), 146–150. https://doi.org/10.19080/BBOAJ.2018.05.555673, 2018

[15] Bruno, P., Calimeri, F., Kitanidis, A. S., & De Momi, E. "Data reduction and data visualization for automatic diagnosis using gene expression and clinical data.", *Artificial Intelligence in Medicine*, *107*(November 2019), 101884. https://doi.org/10.1016/j.artmed.2020.101884, 2020

[16] Hepworth, K. "Big Data Visualization : Promises & Pitfalls.", *Communication Design Quarterly*, *4*, 7–19, 2017.

[17] Alvi, M. H. "*A Manual for Selecting Sampling Techniques in Research Mohsin Hassan Alvi"*, 2016.

[18] Rahi, S. "Research Design and Methods : A Systematic Review of Research Paradigms, Sampling Issues and Instruments Development.", *International Journal of Economics & Management Sciences*, *6*(2), 1–5. https://doi.org/10.4172/2162-6359.1000403, 2017

[19] Imad Rida, S. A. "AN ENSEMBLE LEARNING METHOD BASED ON RANDOM SUBSPACE SAMPLING FOR PALMPRINT IDENTIFICATION.", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2047-2051, 2018.

[20] Qing Xu, G. S. "Generalizing systematic adaptive cluster sampling for forest ecosystem inventory.", *Forest Ecology and Management*, 1-10, 2021.

[21] Elfil, M., & Negida, A. "Sampling methods in Clinical Research ; an Educational Review.", 5(1), 3–5, 2017.

[22] Bashar I. Ahmad, A. T. "Spectral Analysis of Stratified Sampling: A Means to Perform Efficient Multiband Spectrum Sensing.", *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, 178-187, 2012.

[23] Saumya Singh, S. S. "Review of Clustering Techniques in Control System.", International Conference on Smart Sustainable Intelligent Computing and Application under ICITETM 2020, 272-280, 2020.

[24] Sumanta Das, J. C. "UAV-Thermal imaging and agglomerative hierarchical clustering.", *ISPRS Journal of Photogrammetry and Remote Sensing*, 221-237, 2021.

[25] Aly, H. H. "A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid.", *Electric Power Systems Research*, 1-13, 2021.

[26] Ly, T., Cockburn, M., & Langholz, B. "Cost-efficient case-control cluster sampling designs for population-based epidemiological studies.", *Spatial and Spatio-Temporal Epidemiology*, *26*, 95–105. https://doi.org/10.1016/j.sste.2018.05.002, 2018

[27] Yu Han, Q. S.-Y.-L. "A convenient sampling and noninvasive dried spot method of uric acid in human saliva: Comparison of serum uric acid value and salivary uric acid in healthy volunteers and hyperuricemia patients.", *Journal of Chromatography B*, 1-8, 2021.

[28] Tian, S., Zhang, J., Chen, L., Liu, H., & Wang, Y. "Random Sampling-Arithmetic Mean: A Simple Method of Meteorological Data Quality Control Based on Random Observation Thought.", *IEEE Access*, *8*, 226999–227013. https://doi.org/10.1109/ACCESS.2020.3045434, 2020

[29] He, Y., Huang, J. Z., Long, H., Wang, Q., & Wei, C. "I-Sampling: A New Block-Based Sampling Method for Large-Scale Dataset.", *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, 360–367. https://doi.org/10.1109/BigDataCongress.2017.53, 2017

[30] Xie, H., Bin Ahmadon, M. A., Yamaguchi, S., & Toyoshima, I. "Random Sampling and Inductive Ability Evaluation of Word Embedding in Medical Literature.", *2019 IEEE International Conference on Consumer Electronics, ICCE 2019*. https://doi.org/10.1109/ICCE.2019.8662022, 2019

[31] Lu, Z., Zhuang, Y., & Yuan, L. "Random sampling and performance analysis for networked systems.", *Proceedings off the 36th Chinese Control Conference*, 7847–7851. https://doi.org/10.23919/ChiCC.2017.8028597, 2017

[32] Saavedra-nieves, A. "Assessing systematic sampling in estimating the Banzhaf – Owen value.", *Operations Research Letters*, *48*(6), 725–731. https://doi.org/10.1016/j.orl.2020.08.015, 2020

[33] Goshu, N. N., & Kassa, S. M. "Computers and Operations Research A systematic sampling evolutionary (SSE) method for stochastic bilevel programming problems.", *Computers and Operations Research*, *120*, 104942. https://doi.org/10.1016/j.cor.2020.104942, 2020

[34] Ryu, J., & Kamata, S. "An efficient computational algorithm for Hausdorff distance based on points-ruling-out and systematic random sampling.", *Pattern Recognition*, *114*, 107857. https://doi.org/10.1016/j.patcog.2021.107857, 2021

[35] Xiao, S., Oladyshkin, S., & Nowak, W. "Reliability analysis with stratified importance sampling based on adaptive Kriging.", *Reliability Engineering and System Safety*, *197*(January), 106852. https://doi.org/10.1016/j.ress.2020.106852, 2020

[36] Raeisi-Gahrooei, Y., Khodabakhshian, A., & Hooshmand, R. A. "A new stratified random sample customer selection for load research study in distribution networks.", *International Journal of Electrical Power and Energy Systems*, *97*(July 2017), 363–371. https://doi.org/10.1016/j.ijepes.2017.11.029, 2018

[37] Fu, Z., Cheng, J., Yang, M., Batista, J., & Jiang, Y. "Wastewater discharge quality prediction using stratified sampling and wavelet de-noising ANFIS model.", *Computers and Electrical Engineering*, *85*. https://doi.org/10.1016/j.compeleceng.2020.106701, 2020

[38] Capiluppi, A., Di Ruscio, D., Di Rocco, J., Nguyen, P. T., & Ajienka, N. "Detecting Java software similarities by using different clustering techniques.", *Information and Software Technology*, *122*(September 2019), 106279. https://doi.org/10.1016/j.infsof.2020.106279, 2020

[39] Leonori, S., Martino, A., Luzi, M., Frattale Mascioli, F. M., & Rizzi, A. "A generalized framework for ANFIS synthesis procedures by clustering techniques.", *Applied Soft Computing Journal*, *96*, 106622. https://doi.org/10.1016/j.asoc.2020.106622, 2020

[40] Zoha, N., Ghosh, S. K., Arif-Ul-Islam, M., & Ghosh, T. "A numerical approach to maximize the number of testing of COVID-19 using conditional cluster sampling method.", *Informatics in Medicine Unlocked*, *23*, 100532. https://doi.org/10.1016/j.imu.2021.100532, 2021

[41] Ly, T., Cockburn, M., & Langholz, B. "Cost-efficient case-control cluster sampling designs for population-based epidemiological studies.", *Spatial and Spatio-Temporal Epidemiology*, *26*, 95–105. https://doi.org/10.1016/j.sste.2018.05.002, 2018

[42] Sommer, A., & Steland, A. "Journal of Statistical Planning and Inference Multistage acceptance sampling under nonparametric

dependent sampling designs.", *Journal of Statistical Planning and Inference*, *199*, 89–113. https://doi.org/10.1016/j.jspi.2018.05.006, 2019

[43] Naderifar, M., Goli, H., & Ghaljaie, F. "Snowball Sampling : A Purposeful Method of Sampling in Qualitative.", *Strides in Development of Medical Education*, *14*(3). https://doi.org/10.5812/sdme.67670.Research, 2017

[44] Rolls, D. A., & Robins, G. "Minimum distance estimators of population size from snowball samples using conditional estimation and scaling of exponential random graph models.", *Computational Statistics and Data Analysis*, *116*, 32–48. https://doi.org/10.1016/j.csda.2017.07.004, 2017

[45] Stivala, A. D., Koskinen, J. H., Rolls, D. A., Wang, P., & Robins, G. L. "Snowball sampling for estimating exponential random graph models for large networks.", *Social Networks*, *47*, 167–188. https://doi.org/10.1016/j.socnet.2015.11.003, 2016

[46] Lee, J., & Spratling, R. "Recruiting Mothers of Children With Developmental Disabilities : Adaptations of the Snowball Sampling Technique Using Social Media.", *Journal of Pediatric Health Care*, *33*(1), 107–110. https://doi.org/10.1016/j.pedhc.2018.09.011, 2019

[47] N. N. S. 'Asri, Z. Idrus, H. Zainuddin and Z. Idrus, "Parallel Coordinates Graph in Bundling Technique," 2019 1st International Informatics and Software Engineering Conference (UBMYK), pp. 1-6, https://doi.org10.1109/UBMYK48245.2019.8965511, 2019

[48] Idrus, Zainura & Zainuddin, H. & Ja'afar, A.D.M. Visual analytics: designing flexible filtering in parallel coordinate graph. Journal of Fundamental and Applied Sciences. 9. 23. https://doi.org/10.4314/jfas.v9i5s.3, 2018

Research on the Optimization Problem of Agricultural Product Logistics based on Genetic Algorithm under the Background of Sharing Economy

Na Wang
College of Humanities and Management
Xi'an Traffic Engineering Institute, Xi'an, 710300, China

***Abstract*—China's national development and reform commission issued the "logistics industry adjustment and revitalization plan" in 2009 to support the development of agricultural product logistics and distribution centers. China's agricultural product logistics and distribution have entered a stage of rapid development. With the rise of the sharing economy, logistics has become a bottleneck restricting the further development of agricultural product distribution. In order to realize the effective cooperation among the main body of agricultural product logistics distribution, improve the distribution efficiency and reduce the distribution cost, a logistics distribution optimization model based on the two-layer planning idea and genetic algorithm is proposed. A two-level programming model is constructed by combining qualitative and quantitative methods, theory and examples, and insertion and deletion operators are introduced to optimize the genetic algorithm. The research results show that the optimized genetic algorithm has a 54.55% increase in convergence speed, 1.08% in performance, and a 54.231% reduction in path length compared to the benchmark algorithm. It effectively improves the efficiency of path planning and saves the planning cost, and the final target value is reduced by 48.19%.**

***Keywords*—Sharing economy; two-level programming; genetic algorithm; path optimization**

I. Introduction

With the rise of the sharing economy, urban residents in my country have higher requirements for the distribution of perishable agricultural products, and relevant online platforms need to have more professional logistics and distribution planning [1]. Due to the late start of the perishable agricultural product distribution network in China, the lack of infrastructure and equipment, and the lack of relevant technical and legal support are discussed in [2]. In addition to strengthening the construction of perishable agricultural product distribution infrastructure and improving relevant laws, it is also necessary to establish an efficient logistics network system for the distribution of perishable agricultural products [3]. In China, the distribution network of perishable agricultural products usually adopts the network model of a single economy. The production base is directly connected with the sales terminal, and the merchant arranges the delivery order according to the user's order delivery time. The lack of overall planning in each link of this model will not only result in waste of transport capacity, but once traffic congestion and other problems occur midway, the goods may not be delivered in time. In order to build a more reasonable logistics network and solve the problems existing in the layout of the existing logistics network, a genetic algorithm (GA) optimization logistics distribution model for two-level planning is proposed. The upper model is to determine the optimal location of logistics distribution nodes, and the lower model is to determine the best path for logistics distribution. For the optimization of the genetic algorithm, insertion operator and deletion operator are introduced to increase the coherence constraint of path planning and reduce the path mutation before and after the current planning. The combination of node location selection and distribution path optimization in the logistics distribution network can fundamentally realize the coordination and cooperation of various subjects, so as to improve logistics distribution efficiency, reduce distribution costs, and meet consumer demand for quality, time and price.

II. Related Work

The State Information Center of China mentioned in the "China Sharing Economy Development Report (2019)" released on February 28, 2019 that the transaction volume of China's sharing economy market exceeded three trillion yuan, a year-on-year increase of 41.6%, and the number of relevant participants was close to 800 million. The "sharing economy" has become a new economic growth point with huge potential in China [4]. The transaction scale of fresh agricultural products is about 162 billion yuan, maintaining a steady growth of 29.2% [5].

In order to solve the problems of high cost and low resource utilization in the process of logistics distribution under the sharing economy, many domestic and foreign scholars have carried out related research. Scholars such as Lv proposed a linear multi-objective bi-level programming method, the core of which is to replace the lower-level problem with optimal conditions, and use the complementary constraint as the penalty term for the upper-level objective. And introduce the concept of problem equilibrium point and analyze its characteristics, and propose an equilibrium point algorithm based on the penalty method [6]. Scholars such as Zeng proposed a two-level programming model with equilibrium constraints to optimize the planning and design of renewable energy electric vehicle charging stations. The lower-level problem is used to confirm the user's charging strategy [7].

Scholars such as Moon K proposed a two-level programming model to find a single minimum genetic variation by studying the logical reasoning of Boolean networks, and developed a branching and constraint algorithm, which can effectively find all the minimum mutations. The effectiveness of this model is validated through computational studies on a variety of Boolean networks [8]. Song et al. constructed an energy optimal scheduling model based on uncertain two-level programming. The upper model takes the transition matrix of the energy hub as the upper decision maker, and the minimum operating cost in the form of confidence as the objective function; the lower model uses the power subnet, the optimal operation scheme of the thermal energy sub-network and the gas sub-network is the lower-level decision-maker, aiming at the operation economy of each sub-network and taking its operation as a necessary constraint [9]. Aboelnaga Y and other scholars proposed an improved genetic algorithm and chaotic search to solve the two-layer programming problem, which improved the performance and convergence speed of the algorithm, and successfully got rid of the local optimum, allowing the algorithm to solve the global optimum [10].

Scholars such as Zhu believe that a genetic algorithm model based on the attention mechanism is proposed, and the attention mechanism is used to assign different weights to each feature, so that the model can focus more attention on key features. And through the genetic algorithm to optimize the structure of the model and the parameter selection of the data, the global search ability of the model is improved [11]. Wang et al. proposed a multi-objective trajectory planning method based on an improved elite non-dominant sorting genetic algorithm. The trajectory function is composed of a quintic polynomial and a cubic Bezier curve, and then three genetic operators are introduced: sorting group selection, direction-based crossover and mutation with adaptive precision control. The optimal solution of the algorithm is determined through fuzzy comprehensive evaluation to obtain the optimal trajectory [12]. Huang et al. (2022) proposed an adaptive optics technique based on genetic algorithm to detect the twisted wavefront of a laser beam, and then perform aberration correction, which has optimized the performance of two-photon fluorescence microscopy. With the spatial light modulator acting as a wavefront controller, the corrected phase is obtained through a signal feedback loop and a natural selection process [13]. Scholars such as Zemliak (2022) introduced the generalized optimization idea of circuit into the optimization of genetic algorithm, changing the control vector that determines the method of calculating the fitting function makes it possible to bypass the local minimum and find the global minimum, and the accuracy is high, and the central processing unit Time is also greatly reduced [14]. Scholars such as Al-Obaidi et al. (2021) incorporated the developed and validated process model into an optimization framework based on species conservation genetic algorithms to optimize the design and operating parameters of the process. And a multi-objective function is proposed to optimize the membrane design parameters, xylenol repulsion and required energy consumption [15]. Scholars at home and abroad have used two-level programming and genetic algorithm in the field of path planning, and achieved certain results. However, in terms of genetic algorithm optimization, only the smoothness of a single planned path is considered, but the coherence of multiple planned paths is not considered, and it is impossible to guarantee that the planned path has no mutation [16]. Aiming at the incoherence problem of common genetic algorithms in path planning, this paper proposes an optimization method that introduces insertion operator and deletion operator, and adds coherence constraints in the fitness function, in order to reduce the probability of path mutation and ensure the path Coherence and smoothness of planning.

III. Genetic Algorithm Optimization and Two-Layer Model Construction

A. *Eight-neighborhood Path Selection Genetic Algorithm with Insertion Operator and Deletion Operator*

The GA algorithm is an optimization algorithm based on a biological optimization mechanism. It has strong global search ability and scalability, but has problems such as weak local search ability and high randomness [17]. In order to reduce the path length of the transport vehicle, and improve the stability of the vehicle during driving. The research introduces genetic operators such as deletion and insertion on the basis of standard GA algorithm, and correlates with the last path planning result. The main idea is to compare the angle difference between candidate paths. The larger the difference, the lower the weight given to the path and the smaller the probability of being selected. Assume that each path is a chromosome, and each individual in the population corresponds to only one chromosome, and the genes on the chromosome are path grounding singles. Compared with the coordinates, the grid number of the grid is simpler in form, which is convenient for the operation of the genetic operator, so the sequence number of the grid is used to encode the chromosome. Let any $P=[p_1,p_2,\cdots,p_n]$ path be a node on the path. $p_i(i=1,2,\cdots,n)$ As shown in Fig. 1.

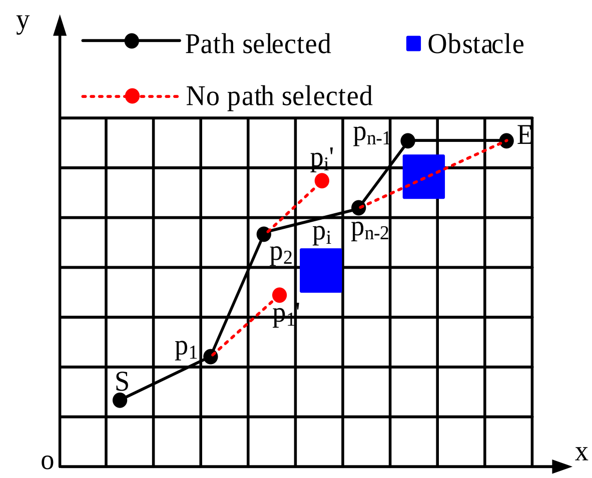


Fig. 1. Path coding of improved genetic algorithm.

In Fig. 1, S is the start point of the path and E is the end point of the path. On the node, the p_1 path is chosen $p_1 \to p_2$ instead $p_1 \to p_1'$ of, because of $\angle Sp_1p_2 < \angle Sp_1p_1'$. The quality of the initial population of the GA algorithm determines the quality of the final output of the GA algorithm, and the diversity and randomness of the initial population are important factors affecting the quality of the initial population [18]. The research uses two methods of directed search and random search to generate candidate paths, and the weights of the two methods are equal. When a path node is near the end, no new path nodes are generated and the current path is preserved. When there is an obstacle in front of the generated path node, the path is re-planned. As shown in formula (1) and formula (2).

$$x_{i+1} = (x_G - x_i)\cdot rand + x_i$$
$$y_{i+1} = (y_G - x_i)\cdot rand + y_i \qquad (1)$$

$$x_{i+1} = N\cdot rand$$
$$y_{i+1} = (y_G - y_i)\cdot rand + y_i \qquad (2)$$

In formula (1) and formula (2), (x_{i+1}, y_{i+1}) is the Cartesian coordinate of the candidate path node, is p_{i+1} the Cartesian coordinate (x_i, y_i) of the current path node p_i, (x_G, y_G) is the coordinate of the coordinate end point, and $rand$ is $0:1$ a random number between. Genetic manipulation refers to the use of a series of genetic operators to perform operations [19]. Due to the randomness of path and path node generation, there is a probability of adjacent node breakpoints in the path, which is not conducive to genetic operations. And there are redundant path nodes in the randomly generated path, which is not the optimal path.

Research on adding an insert operator to connect breakpoint path nodes, and a new delete operator to delete redundant path nodes. In the GA algorithm, the roulette algorithm is a commonly used selection operation, but the roulette algorithm may lead to a large selection error due to its strong randomness, resulting in the frequent occurrence of individuals with large offspring fitness and falling into a local optimum [20]. The study introduces the deterministic sampling selection method to replace the roulette algorithm, and first calculates the expected survival number of the next generation of individuals, as shown in formula (3).

$$N_i = \frac{M_p f_i}{\sum_{i=1}^{M_p} f_i} \qquad (3)$$

In formula (3), M_p is the number of individuals in the population, N_i where i is the expected survival number, f_i of the i th individual, and is the fitness value of the th individual. The integer part is taken N_i as the survival number of individuals in the next generation, the fractional part is sorted in descending order, and the top $M_p - \sum_{i=1}^{M_p}[N_i]$ individual is selected to join the next generation population, which $[N_i]$ is rounded.

In order to avoid the breakpoint path generated by the crossover operation, a single-point crossover method is used for crossover. When there are redundant path nodes in the path, a crossover is randomly selected, and when the path has no redundant path nodes, no crossover is performed. Compared with roulette algorithm, deterministic sampling selection method adopts single point crossing in the crossing process. The algorithm reduces the frequency of individuals with large fitness, effectively avoids falling into local optimum, and improves the local search ability of GA algorithm. In the GA algorithm, the most commonly used mutation method is the random mutation method. In practical problems, however, random variation may lead to poor or even impassable logistics paths. The study adopts the eight-neighbor random non-obstacle node method, as shown in Fig. 2.

In Fig. 2, the path nodes with higher fitness are selected in the eight neighborhoods near the mutation point. When the obtained path is better than the original path, it is replaced with the mutated path, otherwise the original path remains unchanged.

It should be noted that when there is an obstacle between the path node with higher fitness and the target node, another path node with higher fitness needs to be selected. Since the mutation operation will not produce a worse path, the research only performs mutation operation on the optimal path in each generation, which can greatly improve the efficiency and performance of the algorithm. When there is a breakpoint path, the insertion operator fills it with free grids to make it a feasible continuous path, as shown in formula (4).

$$\mathrm{V} = \max\left\{abs(x_{i+1} - x_i), abs(y_{i+1} - y_i)\right\} \qquad (4)$$

In formula (4), at that time, it is determined that the two path nodes are continuous, otherwise it is determined to be discontinuous. $\mathrm{V}=1$ When the path is discontinuous, the average method is used to fill the discontinuous path, as shown in formula (5).

$$x'_{i+1} = \frac{(x_{i+1} + x_i)}{2}$$
$$y'_{i+1} = \frac{(y_{i+1} + y_i)}{2}$$
$$n_i = x'_{i+1} + Ny'_{i+1} \qquad (5)$$

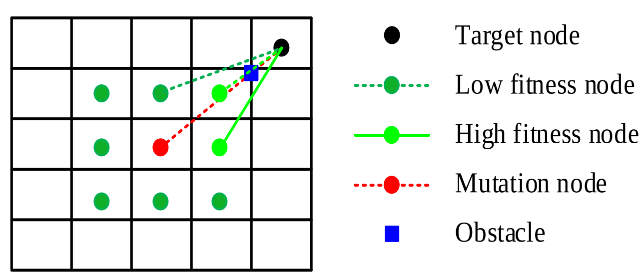


Fig. 2. Eight-neighborhood random non-obstacle node method.

In formula (5), it (x'_{i+1}, y'_{i+1}) is the coordinates of the free grid, and when it n_i is an obstacle grid, it is n_i filled with the nearest grid grid. In order to reduce the length of chromosomes and improve the efficiency of the algorithm, a deletion operator is introduced. Retrieve all path nodes from the starting point. If the weak current path node and the end point are connected without obstacles, it means that the path nodes in between are redundant nodes. These redundant path nodes should be deleted and the path should be re-planned, as shown in Fig. 3.

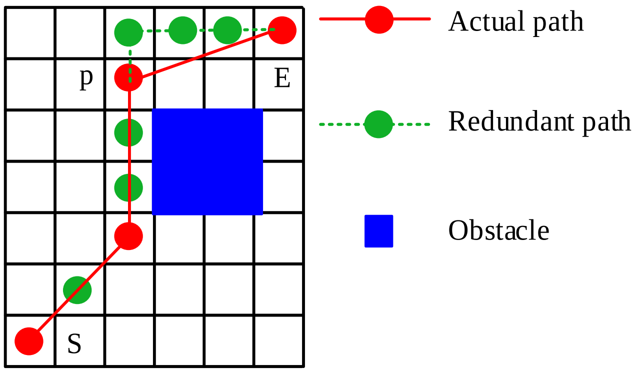


Fig. 3. Paths before and after deleting redundant nodes.

In Fig. 3, redundant path nodes are meaningless and only add computational burden to the algorithm. In particular p, the redundant path nodes between the path node and the end point not only reduce the operating efficiency of the algorithm, but also directly cause the algorithm to fail to select the optimal path. E The elite retention strategy is a commonly used strategy in GA algorithms to ensure that individuals with the highest fitness can be retained after each iteration. Assume that the individual with the highest fitness in the current candidate path is a, compare the fitness of the candidate path a with the fitness of the optimal path so far A, and then use it $a > A$ a instead A. The fitness function is a performance index to evaluate the fitness of an individual, which directly affects whether the final output of the GA algorithm is the optimal solution. The purpose of the research is to find the most suitable logistics distribution path and distribution node, so it is necessary to optimize the length and coherence of the path at the same time. The fitness function is shown in formula (6).

$$f(x) = Inf - w_1 \cdot f_1(p) - w_2 \cdot f_2(p) \tag{6}$$

In formula (6), Inf is a large enough real number, w is $f(p)$ the weight, $f_1(p)$ is the path length function, $f_2(p)$ is the path coherence function. As shown in formula (7) and formula (8).

$$f_1(p) = \sum_{i=0}^{L-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{7}$$

$$f_2(p) = |\theta_k - \theta_{k-1}| \tag{8}$$

In Eq. (7), L is the number of all path nodes, in Equation (8), θ_k is the angle between the direction of the transportation vehicle at the current moment and the planned path, and θ_{k-1} is the angle between the direction of the transportation vehicle and the planned path at the previous moment. The research abstracts the transportation reserve as a mass point, which can θ also be regarded as the expected turning angle of the vehicle, as shown in Eq. (9).

$$\theta = arc\tan\left(\frac{x_1 - x_S}{y_1 - y_S}\right) \tag{9}$$

In formula (9), (x_1, y_1) is the first path node after the operator is deleted, and (x_S, y_S) is the starting point coordinate.

B. Construction of Upper Model and Lower Model

Agricultural products include products with simple storage conditions and long storage time, as well as perishable products. Under the sharing economy, the logistics and distribution of agricultural products in cities are mostly perishable products such as fresh and aquatic products, so such products are mainly considered when constructing a two-tier planning model. There are many kinds of perishable agricultural products, and it is difficult to unify the storage conditions of each logistics distribution node, so it is impossible to find suitable variables to calculate the real loss during transportation. All perishable products, even with the most advanced preservation techniques, have a limited shelf life and can be roughly divided into three stages, as shown in Fig. 4.

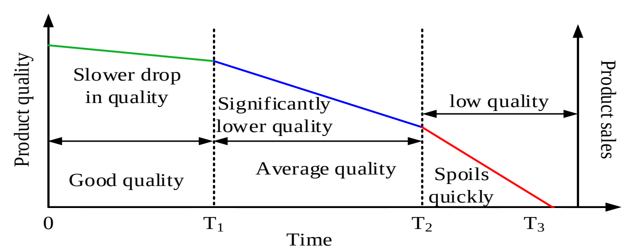


Fig. 4. Quality-time change of perishable agricultural products.

In Fig. 4, the perishable agricultural products are divided into three stages. The first stage is when the initial time 0 arrives T_1, the product is in a state of good quality, and the quality of the product declines slowly; in the second stage, the quality of the product is relatively obvious. In this process, the speed of quality decline is accelerated; in the third stage, the quality of the product deteriorates, and the speed of quality decline increases sharply, and it is T_3 no longer suitable for consumption at any time. In the entire logistics network, the investment and construction of distribution nodes is the focus of optimization, and reasonable distribution node construction can greatly improve the ability to collect and disperse goods. Build the upper-level model for decision-making of distribution nodes, and the cost is shown in formula (10).

$$H=\min\sum_{r\in G}C_{pr}d_{pr}Z_r+\sum_{r\in G}F_rZ_r+\theta_1\sum_{r\in G}\left(t_{pr}Z_r\right)^2$$
$$\sum_{r\in G}Z_r\geq 1$$
$$\sum_{r\in G}Q_rZ_r\geq\sum_{j\in H}q_j$$
$$Z_r=\begin{cases}1,\text{Create a distribution node at } r\\ 0,\text{else}\end{cases}\tag{10}$$

In formula (10), $\sum_{r\in G}C_{pr}d_{pr}Z_r$, $\sum_{r\in G}F_rZ_r$ and $\theta_1\sum_{r\in G}\left(t_{pr}Z_r\right)^2$ represent the transportation cost of perishable agricultural products from the supplier p to the distribution node r, the construction and operation cost of the distribution node, and the consumption cost of perishable agricultural products in the distribution process, respectively. $\sum_{r\in G}Z_r\geq 1$ Indicates that at least one distribution node must be established, $\sum_{r\in G}Q_rZ_r\geq\sum_{j\in H}q_j$ indicating that the carrying capacity of the distribution node is greater than the demand of the terminal retailer. In actual distribution, the collection and distribution of goods need to be flexibly changed according to the specific requirements of customers. Considering the storage capacity, transportation capacity and time requirements of customers, the lower-level model of distribution path decision-making is constructed, as shown in formula (11).

$$U=\min\sum_{i\in S}\sum_{j\in H}\sum_{k\in V}C_{ij}X_{ijk}d_{ij}+\sum_{k\in V}C_kX_k+\sum_{j\in H}R_jE_j+\theta\sum_{k\in V}\left(\sum_{i\in S}\sum_{j\in H}X_{ijk}t_{ij}\right)^2$$
$$\sum_{k\in V}\sum_{j\in H}C_{ij}X_{ijk}=1;\ j\in H$$
$$\sum_{j\in H}\sum_{i\in S}q_{ij}X_{ijk}\leq Q_k;k\in V$$
$$\sum_{i\in S}X_{ijk}-\sum_{i\in S}X_{pjk}=1;k\in V,p\in S$$
$$\sum_{k\in V}X_{rmk}+Z_r+Z_m\leq 2;r\in G,m\in G$$
$$\sum_{i\in S}\sum_{j\in H}X_{ijk}\leq 1,k\in V$$
$$E_{ij}\leq T_j\leq L_{ij}$$
$$X_k=\begin{cases}1,\text{vehicle } k \text{ is used}\\ 0,\text{else}\end{cases}$$
$$X_{rjk}=\begin{cases}1,\text{the } k\text{-th vehicle delivers from } r \text{ to } j\\ 0,\text{else}\end{cases},k\in V,j\in H$$
$$Y_{rj}=\begin{cases}1,j \text{ deliver from } r\\ 0,\text{else}\end{cases},r\in G,j\in H$$
$$R_j=\begin{cases}1,\text{time window met}\\ 0,\text{else}\end{cases},j\in H\tag{11}$$

In Eq. (11), $\sum_{i\in S}\sum_{j\in H}\sum_{k\in V}C_{ij}X_{ijk}d_{ij}$, $\sum_{k\in V}C_kX_k$, $\sum_{j\in H}R_jE_j$ and $\theta\sum_{k\in V}\left(\sum_{i\in S}\sum_{j\in H}X_{ijk}t_{ij}\right)^2$ points are represented by the distribution cost of the distribution node to the retailer, the use cost of the vehicle, the penalty cost of overtime and the loss cost on the way. $\sum_{k\in V}\sum_{j\in H}C_{ij}X_{ijk}=1;\ j\in H$ Indicates that each retailer has only one vehicle for distribution, $\sum_{j\in H}\sum_{i\in S}q_{ij}X_{ijk}\leq Q_k;k\in V$ indicates that the total amount of goods in each distribution route does not exceed the maximum capacity of each vehicle, $\sum_{i\in S}X_{ijk}-\sum_{i\in S}X_{pjk}=1;k\in V,p\in S$ indicates that each distribution process is continuous, and $\sum_{k\in V}X_{rmk}+Z_r+Z_m\leq 2;r\in G,m\in G$ indicates that there is no distribution relationship between the two distribution nodes., which $\sum_{i\in S}\sum_{j\in H}X_{ijk}\leq 1,k\in V$ means that each vehicle belongs to at most one distribution node, which $E_{ij}\leq T_j\leq L_{ij}$ is the time constraint of the retailer. It can be seen that, whether it is the upper-level distribution node decision-making model or the lower-level distribution path arrangement model, the cost of the entire logistics activity is required to be the lowest. By changing the location class of distribution nodes, the choice of distribution path is affected, and the continuous optimization of the distribution path in turn affects the choice of the location of distribution nodes. This process is a process of mutual influence. The process of solving the model is shown in Fig. 5.

The process of solving the two-layer model is shown in formula (12).

$$p_{ij}^k(t)=\begin{cases}\dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum_{s\in allowed_k}\tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)},j\in allowed_k\\ 0,\text{else}\end{cases}\tag{12}$$

In Eq. (12), the α sum β score is denoted by the weight of the distance-influenced heuristic factor. α The larger the value, the more inclined the shorter distance is when planning the path, the β larger the more inclined the path with fewer obstacles. $allowed_k$ Indicates k the path nodes that can be selected in the next planning of the path, τ_{ij} and η_{ij} is the amount of information obtained by the two weights when planning the path. The update of information is shown in formula (13).

$$\tau_{ij}(t+n)=\rho\tau_{ij}(t)+\Delta\tau_{ij}$$
$$\Delta\tau_{ij}=\sum_{k=1}^{n}\tau_{ij}^k\tag{13}$$

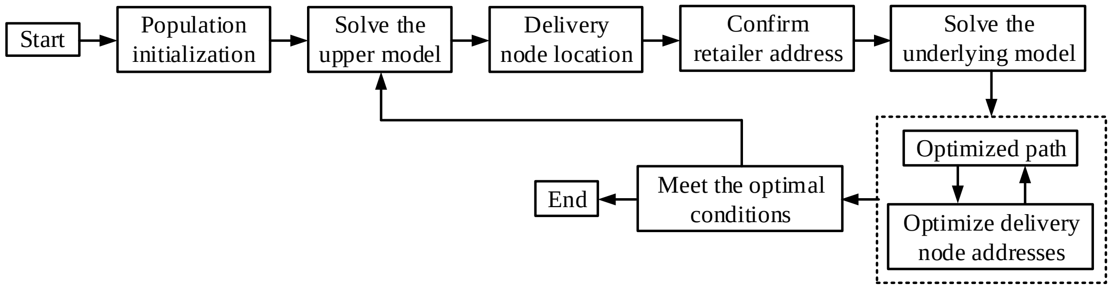


Fig. 5. Genetic algorithm to solve the two-layer model.

In formula (13), it $\rho(0<\rho<1)$ is the residual coefficient of information, and η_{ij} the update is the same as formula (13), so it is omitted. This process is repeated continuously, and the iteration is stopped when the end condition is satisfied. If it is not satisfied, a new round of optimization is started after jumping to the initialized population.

IV. Path Planning Model Training and Examples

A. Path Planning using Different Genetic Algorithms in Simple and Complex Environments

Two training scenarios are constructed. The first environment is relatively simple, and the second environment is relatively complex. Assuming that the length of each fence grid is 1, the evaluation indicators are the path length and the number of corners. The comparison algorithms are the A-Star algorithm and the GA algorithm commonly used in path planning, and the results are shown in Fig. 6.

In the simple scenario of Fig. 6(a), all three algorithms successfully plan paths. Among them, the path length of the ImproveGA algorithm is about 13.544 and the number of corners is one. The path length of the StandardGA algorithm is 12.857, and the number of corners is three. The path length of Astar algorithm is 22, and the number of corners is 6. In the complex scene of Fig. 6(b), the path length of the ImproveGA algorithm is about 17.890, and the number of corners is 4. The StandardGA algorithm falls into a local optimal solution and fails to plan a path. The path length of Astar algorithm is 22, and the number of corners is 7. In a simple scene, the path length of the StandardGA algorithm is 5.72% shorter than that of the ImproveGA algorithm, but the number of corners is three times. In a complex scene, the path cannot be planned because it falls into a local optimal solution. The ImproveGA algorithm can successfully plan paths in both simple and complex environments. Compared with the Astar algorithm, the paths in the two environments are 38.44% and 18.68% shorter, respectively, and the number of corners is much lower than that of the Astar algorithm. In order to further study the performance of the algorithm, the traveling salesman problem dataset maintained by Heidelberg University is used to train the algorithm, and the results are shown in Table I.

In Table I, the path length of the ImproveGA algorithm is optimized by 54.231% and 25.554% compared with the other two algorithms, the maximum yaw angle is optimized by 39.939% and 18.257%, and the sum of the absolute value of the turning angle is optimized by 46.713% and 25.779%. Combining Fig. 6 and Table I, it can be concluded that compared with the Astar algorithm, the GA algorithm has certain advantages, and it can plan the path more flexibly, reduce the turning angle, reduce the path length and the calculation amount of the algorithm, but the StandardGA algorithm is easy to fall into in the face of complexity. The ImproveGA algorithm can not only deal with complex environments, but also has better performance indicators than the StandardGA algorithm.

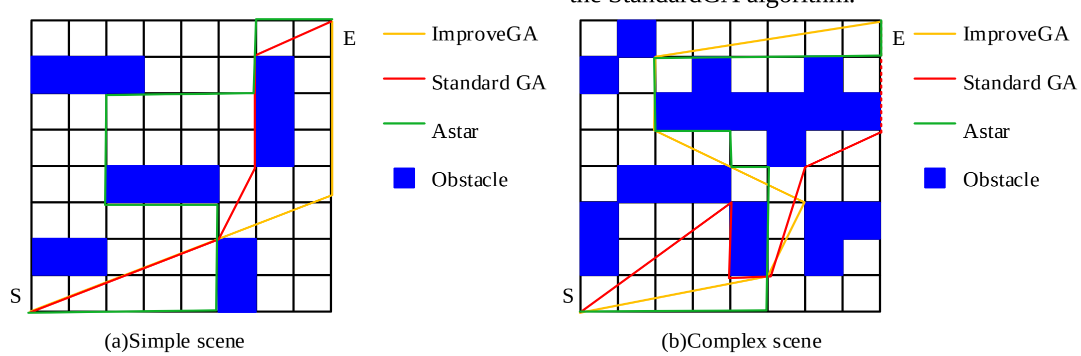


Fig. 6. Simulation results of path planning scenarios.

TABLE I. Path Planning Training Results

Algorithm	Path length	Maximum yaw angle change	The sum of the absolute values of the corners
Astar	27.18km	32.8°	371.2°
StandardGA	16.71km	24.1°	266.5°
ImproveGA	12.44km	19.7°	197.8°

B. *Practical Application of Two-level Planning Model to Logistics Network Optimization*

In order to test the optimization effect of the two-level planning model on the logistics network, an example containing five potential distribution nodes and 20 demand points is used for testing. An enterprise selects two out of five alternative distribution nodes to serve 20 demand points, numbers the five alternative distribution nodes as A, B, C, D, E, and the 20 demand points as 1, 2, 3,..., 20. There are a total of five delivery vehicles, all of the same model, and the demand at each demand node is shown in Table II.

For the convenience of calculation, the unit transportation cost is set at 10 yuan per ton per kilometer, and the cost of using vehicles is 50 yuan per vehicle. The average driving speed of the vehicle is 20 kilometers per hour, the load capacity of the vehicle is 20 tons, and the consumption cost of perishable agricultural products is 1 yuan per hour. Since the relative positions of delivery nodes and demand nodes are fixed, unit price and vehicle cost will not affect the results no matter how they are chosen. The total demand of all demand nodes is 370, and the coordinates and capacities of potential distribution nodes are shown in Table III.

In Table III, the total supply of any two nodes is required to be greater than or equal to 370, so the only combinations considered are point A, any point, and CD. The research uses Matlab to solve the model. Since the problem is not complicated, the fixed value method is used instead of the adaptive probability method in parameter setting. In order to better find the global solution, the initial population size is set to 100. In order to prevent falling into local optimum, the crossover probability is set to 0.5, and the mutation probability is set to 0.05. Due to the large initial population size and low mutation probability, the number of iterations is set to 500 to ensure that the optimal solution can be found. Using fitness as the evaluation index, the change trajectory of the optimal target value is shown in Fig. 7.

TABLE II. C COORDINATES AND DEMAND QUANTITY OF DEMAND NODES

Numbering	X coordinate	Y coordinate	Time window upper limit	Lower time window	Service hours	Demand
1	Twenty-three	28	0	110	10	20
2	52	34	0	120	10	30
3	27	6	0	120	10	15
4	95	38	0	100	10	25
5	64	Twenty-four	0	115	10	10
6	Twenty-three	15	0	120	10	10
7	6	16	0	140	10	15
8	41	10	0	160	10	20
9	38	26	0	105	10	30
10	91	32	0	120	10	10
11	82	14	0	110	10	10
12	64	Twenty-two	0	100	10	5
13	57	71	0	150	10	10
14	56	82	0	100	10	20
15	19	66	0	180	10	10
16	Twenty-two	9	0	105	10	25
17	13	18	0	115	10	25
18	41	35	0	110	10	15
19	69	32	0	115	10	30
20	32	91	0	155	10	35

TABLE III. C COORDINATES AND SUPPLY VOLUME OF EACH DISTRIBUTION NODE

Numbering	X coordinate	Y coordinate	Capacity	Construction Cost
A	3	38	22 0	100
B	54	32	1 70	50
C	41	7	1 90	70
D	50	80	1 80	90
E	25	71	1 70	60

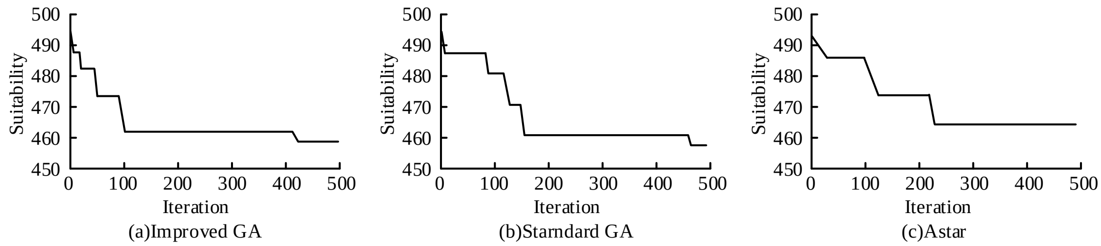


Fig. 7. Trajectory of optimal target value change.

In Fig. 7(a), the two-level programming model using the ImproveGA optimization algorithm tends to converge after about 100 iterations. When the number of iterations is less than 100, the fitness value decreases rapidly. After convergence, the fitness value is 462. The value converges to 459. In Fig. 7(b), the two-level programming model using the StamdardGA algorithm tends to converge at about 150 iterations, but the adaptation value will be slower before the 80 iterations. After convergence, the adaptation value is around 461, and the optimal target The value converges to 458. In Fig. 7(c), the two-level programming model using the Aster algorithm tends to converge after about 220 iterations, the adaptive value after convergence is about 464, and the optimal target value converges to 464. Compared with the Asters algorithm, the ImproveGA algorithm improves the performance by 1.08% and the convergence speed by 54.55%. It can be seen that although the other two algorithms are not much worse than ImproveGA in terms of optimal target value when used in bi-level programming, there is a large gap in stability and optimization speed. It cannot meet the needs of timeliness in the distribution route planning of perishable agricultural products. The optimal distribution node location and distribution path are shown in Fig. 8.

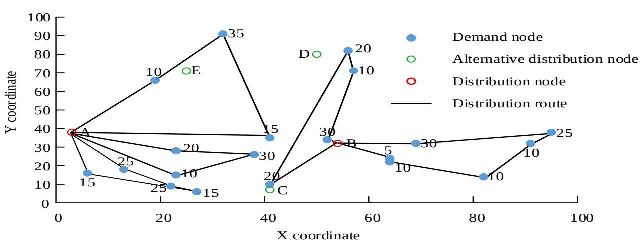


Fig. 8. Optimal distribution node and distribution path.

Combining Table II, Table III and Fig. 8, it can be seen that if only the optimization of distribution nodes is considered, the selected distribution nodes are also points A and B. Since the distribution is not carried out in a roving manner, the total logistics cost is as high as 1072. However, if only the delivery route is considered and the delivery nodes are not considered, the final target value is 884. The bi-level programming model considers both the distribution node and the distribution route, and the final target value is only 458, which is 57.28% and 48.19% lower than the single problem.

V. Conclusion

On the basis of consulting relevant literature and referring to domestic and foreign research results, the distribution network of perishable agricultural products is optimized. The main optimization direction is to establish a two-level planning model to optimize the distribution node location and route planning of the logistics network. And improve the genetic algorithm in the two-level programming model to improve its path planning ability. Compared with the standard GA, the improved GA introduces insertion and deletion operators in the crossover stage. The improved genetic algorithm, Astar algorithm and standard genetic algorithm are tested for path planning. The performance of the improved genetic algorithm is increased by 18.68% and 38.44% respectively. The maximum yaw angle decreased by 18.257% and 39.939%, respectively, and the sum of absolute angles decreased by 25.554% and 46.713%, respectively. Then, the constructed bi-level programming model is tested, and the improved genetic algorithm can plan the optimal configuration faster than the other two comparison algorithms. Speed increased by 54.55%. Finally, comparing the bi-level programming model with the single-planning model, it is found that the cost of the bi-level planning is reduced by 57.28% compared with only considering the distribution node optimization. Compared with only considering the distribution route optimization, the cost of bi-level planning is reduced by 48.19%. This model can effectively save planning costs and improve delivery efficiency. The limitation of this study is that it simplifies the calculation of consumption cost, which is not enough to fully simulate the actual situation. Therefore, future research will consider adding the loss cost in the calculation to make the experiment more realistic.

References

[1] A. Yj, B. A. Bei, L. B. Yang. "Integrated multi-item packaging and vehicle routing with split delivery problem for fresh agri-product emergency supply at large-scale epidemic disease context – ScienceDirect," Journal of Traffic and Transportation Engineering (English Edition), vol. 8(2): 196-208, pp. 2021.

[2] C. Liu, W. Chen, Q. Zhou, et al. "Modelling dynamic freshness-keeping effort over a finite time horizon in a two-echelon online fresh product supply chain," European Journal of Operational Research, vol. 293(2), pp. 511-528, 2020.

[3] H. Haller, A. S. Fagerholm, P. Carlsson, et al. "Towards a resilient and resource-efficient local food system based on industrial symbiosis in hrnsand: A swedish case study," Sustainability, vol. 14(4), pp. 2197-2214, 2022.

[4] C. Liu, R. Chan, M. Wang, et al. "Mapping the sharing economy in China," Sustainability, vol. 12(16), pp. 6333-6352, 2020.

[5] A. J. Ren. "Model innovation and operation optimization of smart supply chain of fresh agricultural products under the background of new technology," Modern Management, vol. 11(3), pp. 227-233, 2021.

[6] Y. Lv, T. Hu, J. Jiang. "Penalty method-based equilibrium point approach for solving the linear bilevel multiobjective programming problem," Discrete, Continuous Dynamical Systems - S, vol. 13(6), pp. 1743-1755, 2020.

[7] B. Zeng, H. Dong, F. Xu, et al. "Bilevel programming approach for optimal planning design of EV charging station," IEEE Transactions on Industry Applications, vol. 56(3), pp. 2314-2323, 2020.

[8] K. Moon, K. Lee, S. Chopra, et al. "Bilevel integer programming on a Boolean network for discovering critical genetic alterations in cancer development and therapy," vol. 300(2), pp. 743-754, 2022.

[9] X. Song, H. Lin, G. De, H. Li, X. Fu, Z. Tan. "An energy optimal dispatching model of an integrated energy system based on uncertain bilevel programming," Energies, vol. 13(2), pp. 477-503, 2020.

[10] Y. Aboelnaga, S Nasr. "Modified evolutionary algorithm and chaotic search for bilevel programming problems," Symmetry, vol. 12(5), pp. 767-796, 2020.

[11] T. Zhu, Y. Li, Z. Li, et al. "Inter-hour forecast of solar radiation based on long short-term memory with attention mechanism and genetic algorithm," Energies, vol. 15(3), pp. 1-14, 2022.

[12] Z. Wang, Y. Li, K. Shuai, et al. "Multi-objective trajectory planning method based on the improved elitist non-dominated sorting genetic algorithm," Chinese Journal of Mechanical Engineering, vol. 35(1), pp. 70-84, 2022.

[13] Y. Huang, Y. Guo, J. Li, et al. "Aberration correction to optimize the performance of two-photon fluorescence microscopy using the genetic algorithm," Microscopy and Microanalysis, vol. 28(2), pp. 383-389, 2022.

[14] A. Zemliak. "A modified genetic algorithm for system optimization," COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, vol. 41(1), pp. 499-516, 2022.

[15] M. A. Al-Obaidi, A. Ruiz-García, G. Hassan, et al. "Model based simulation and genetic algorithm based optimisation of spiral wound membrane RO process for improved dimethylphenol rejection from wastewater," Membranes, vol. 11(8), pp. 595-614, 2021.

[16] G. Singh, V. K. Banga. "Combinations of novel hybrid optimization algorithms-based trajectory planning analysis for an industrial robotic manipulators," Journal of Field Robotics, vol. 39(5), pp. 650-674, 2022.

[17] M. Chattoraj, U. R. Vinayakamurthy. "A self adaptive new crossover operator to improve the efficiency of the genetic algorithm to find the shortest path," Indonesian Journal of Electrical Engineering and Computer Science, vol. 23(2), pp. 1011-1017, 2021.

[18] Y. M. Wu, Z. Li, C. Sun, et al. "Measurement and control of system resilience recovery by path planning based on improved genetic algorithm," Measurement and Control, vol. 54(7-8), pp. 1157-1173, 2021.

[19] R. Havangi. "Robust square-root cubature FastSLAM with genetic operators," Robotica, vol. 39(4), pp. 1-21, 2020.

[20] N. P. Marri, NR Rajalakshmi. "MOEAGAC:an energy aware model with genetic algorithm for efficient scheduling in cloud computing," International Journal of Intelligent Computing and Cybernetics, vol. 15(2), pp. 318-329, 2022.

Application Research of Trademark Recognition Technology based on SIFT Feature Recognition Algorithm in Advertising Design

Weina Zhang
Department of Publishing and Communication, Shanghai Publishing and Printing College, Shanghai, 200093, China

***Abstract*—Now-a-days, due to the sharp increase in the number of advertising designs, creative duplication is easy to occur in advertising design. If this situation is not discovered in time, it may cause legal disputes and cause damage to the reputation and property of the enterprise. In view of the above situation, this paper proposes a trademark recognition technology based on SIFT feature recognition algorithm to avoid duplication of advertisement design and cause copyright disputes. Aiming at the defect that the dimension of the image feature vector extracted by SIFT algorithm is too high, the principal component analysis method is used to reduce its dimension. For the problem of unsatisfactory image recognition rate accuracy of SIFT algorithm, a support vector machine is used to classify the extracted feature vector, so as to improve the image recognition rate. Based on the above content, build a trademark recognition model. The research results show that the recognition accuracy of the model reaches 98.82%, 0.66% and 0.58% higher than that of Model 1 and Model 2; AUC value of model 3 is 0.962, 0.039 higher than model 2 and 0.107 higher than model 1. The above results show that the proposed trademark recognition model can better identify similar advertising designs, thereby avoiding design duplication and legal disputes.**

Keywords*—SIFT algorithm; trademark recognition; advertising design; support vector machine; principal component analysis

I. Introduction

Good advertising design can keep a product in the memory of consumers for a long time, thereby increasing the popularity and sales of the product. Therefore, advertising design is of great significance to the development of enterprises. Trademark design is an important part of advertising design. Due to the increase in the number of enterprises and the huge number of trademark designs, duplication of ideas is an unavoidable problem. Reputation and property have caused certain losses [1-2]. Therefore, trademark recognition technology based on computer vision technology has gradually emerged and has received extensive attention from all walks of life. It has important applications in product monitoring, trademark protection and commercial information mining. The application of trademark recognition technology in advertising design is as follows: through the identification of trademark design features, trademark images with similar characteristics are matched in the data set, so as to determine whether the design results are repeated with other trademark designs [3]. Scale invariant feature transform (SIFT) algorithm has strong local feature extraction ability and matching ability, and is less affected by light and noise, so it has an important application in trademark recognition. However, the feature dimension extracted by the SIFT algorithm is high, which leads to long image recognition time and unsatisfactory recognition accuracy, so it needs to be improved [4-5]. The principal components analysis (PCA) is used to reduce the dimensions of the feature vectors extracted by SIFT algorithm, and the support vector machine (SVM) is used to classify the feature vectors. Through the above operations, the trademark recognition model based on SIFT-PCA-SVM is built, so as to improve the efficiency and accuracy of image recognition and matching. The results show that the precision and efficiency of the research model are significantly higher than the existing models. The research builds a trademark recognition model based on SIFT-PCA-SVM to achieve efficient trademark recognition and avoid the repetition of advertising design ideas to the greatest extent. There are three main innovations in the research. The first point is to build a trademark recognition model based on SIFT algorithm to achieve intelligent detection of plagiarism in advertising design. The second point is to use PCA to reduce the dimension of feature vectors and improve the processing efficiency and accuracy of SIFT algorithm. The third point is to introduce SVM to classify the output feature vectors, so as to improve the accuracy of trademark recognition.

II. Related Works

In today's highly developed economy and technology, the number of enterprises is increasing, and various products are emerging on the market one after another. In an increasingly fierce market environment, excellent advertising can make products more famous, thereby increasing product sales and improving corporate efficiency. Therefore, the importance of advertising design is becoming more and more prominent. In order to improve the effect of advertising design, many scholars take the advertising design of various products as an example, and put forward guiding suggestions for advertising design. Choi SJ et al. assessed the correlation between ad spend and data breaches using a gamma distribution using data from non-federal acute care inpatient hospitals as an example [6]. Kininmonth S believes that the frontier media market will have a certain impact on the advertising industry, and discusses the advertising market prospects under the trend of digital advertising [7]. Shieh CH et al. invited 363 consumers to participate in a field experiment in a shopping mall to explore the correlation between consumers' purchase intention and location-based advertising [8]. Wright MJ et al. used the logistic regression analysis method to discuss and analyze the marketing effect of digital advertising in the case of consumer distraction, and believed that the marketing effect of emotional advertising is better than rational advertising in special circumstances [9]. Lacoste-Badie S et al. took 50 French children as the research object, evaluated the children's attention to the health problems in snack advertisements, and suggested that the French authorities should take measures to strengthen the health warnings in advertisements based on the results [10]. Taking a coffee shop brand as an example, Busser JA et al. used structural equation model to study the impact of consumer-generated advertising on consumers' brand trust and brand loyalty. Brand loyalty has a positive impact [11]. Levin E et al. discussed the relationship between the service quality of advertising companies, advertising companies and customer relationship management from the social and economic aspects [12]. M Méndez-Suárez et al. used the partial least squares equation model to study the impact of Internet advertising and digital advertising on brand awareness, as well as the impact of brand awareness on product sales [13]. Basch CH et al. analyzed the prevalence of sugar-sweetened beverage advertisements on LinkNYC kiosks in New York City to evaluate and recommend the work of the New York Department of Health [14]. Sahin S et al. analyzed the influence of green advertising image on consumer psychology, and the results show that green advertising image can promote consumers' purchasing behavior [15].

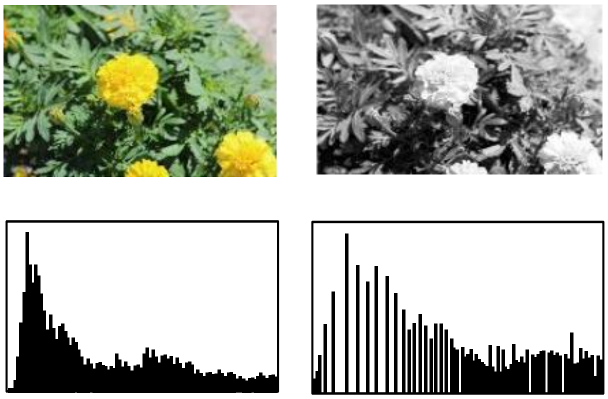
The SIFT algorithm is a common algorithm in image processing, which can detect and describe the key features of the image, and is a local feature descriptor. The SIFT algorithm has the characteristics of strong discrimination, more feature vector extraction, and less influence by noise and illumination. It has a key application in digital image feature matching. A large number of scholars have conducted in-depth discussion and analysis on the application and improvement of SIFT. Combining tensor theory and based on spatial and spectral information, LI et al. proposed a tensor gradient SIFT and applied it to feature extraction and matching of hyperspectral images [16]. Li J et al. used PCA and SIFT to predict protein-protein interactions in organisms through protein sequences, and verified by five-fold cross-validation method, the accuracy of this method exceeded 97% [17]. Combining SIFT and patch matching, Henan et al. proposed a 3D view image quality improvement method. The simulation results show that the method has strong feasibility and effectiveness [18]. Raveendra K et al. optimized the performance of SIFT through the global optimization feature of ant colony optimization (ACO), and constructed a hybrid model for document retrieval in the dataset, and the accuracy of the model exceeded 95% [19]. Hwang SW uses artificial neural networks (ANN) and SIFT algorithms to build a model to automatically classify wood knots based on macroscopic images [20]. Wulandari I et al. introduced the SIFT algorithm in the citra fusion technology and used it in treatment and medical treatment. The experimental results show that the method has good application effect in the medical field [21]. Taking the official currency of Singapore as an example, Prasasti AL [22] proposed a foreign currency identification method based on the improved SIFT algorithm based on ANDROID. Dalai R et al. introduced the SIFT algorithm into the Mask-R-Convolution Neural Network (Mask-RCNN) to extract and match the feature points of the image, thereby realizing image processing and fast detection of objects in the image [23]. Karim A et al. proposed an Arabic handwriting recognition model combining SIFT and SVM. The experimental results show that the recognition accuracy of this model reaches 99.08% [24]. Misra I et al. used the pattern-guided SIFT algorithm for feature detection, thereby improving the accuracy of remote sensing image registration and facilitating subsequent Earth observation data analysis. The experimental results show that the RMSE of this method is 0.12 [25].

As can be seen from the above content, there are many research results on advertising design and SIFT algorithm, but few people apply SIFT algorithm to advertising design. Through the feature extraction and feature matching capabilities of the SIFT algorithm, it can help advertising designers avoid creative duplication, thereby improving the effect of advertising design and increasing corporate profits. To this end, the research builds a trademark recognition model based on the SIFT algorithm, and discusses the practical application effect of the model in advertising design.

III. Construction of Trademark Recognition Model Based on Improved SIFT

A. *Image Preprocessing and SIFT-Based Trademark Recognition*

For the survival and development of enterprises, the use of advertisements to enhance the popularity and sales of products has become a major concern of major enterprises. Therefore, the number of advertisements is increasing, flooding all aspects of people's lives. In advertising design work, especially in trademark design, creative duplication is an unavoidable problem. When the duplication of trademark design is not discovered in time, copyright and legal disputes are likely to occur, causing damage to the reputation and property of enterprises, products and advertising designers. Therefore, it is very necessary to realize trademark recognition through image recognition technology, so as to avoid repeated plagiarism of trademark design. When performing trademark image recognition, it is necessary to extract the feature values of the image to achieve image feature matching. However, the original image has many problems, such as noise and different sizes [26]. Therefore, image preprocessing is required. First, the digital image histogram is used to process the gray value of the image, as shown in Fig. 1.



(a) Before pretreatment (b) After pretreatment

Fig. 1. Gray value processing of image by histogram

Then, a Gaussian filter is used to filter the image, and the Gaussian function is shown in formula (1).

$$g(i,j)=e^{-\frac{(i^2+j^2)}{2\sigma^2}} \tag{1}$$

In formula (3), $[i,j]$ is a certain pixel point, (i^2+j^2) is the radius of the Gaussian function, and σ is the standard deviation. Finally, the redundant data of the image is removed by the method of data compression, and the image compression is realized. Data compression refers to reducing the amount of data processed with the same amount of information. Data redundancy can be expressed as formula (2).

$$R=1-1/C \tag{2}$$

Formula (2), R,C the number of bits of data redundancy and the compression rate are respectively expressed. The definition of the compression ratio is shown in formula (3).

$$C=b/b' \tag{3}$$

Formula (3), , b and b' respectively represent the number of bits in the information to be processed before compression and the number of bits in the information to be processed after compression. The SIFT algorithm has strong anti-interference ability, and has excellent feature extraction and feature matching capabilities. The SIFT algorithm is used to build a trademark recognition model. First, the scale space is constructed with the help of Gaussian blur. For an image whose scale space is set to $L(x,y,\delta)$, its size can be expressed as Equation (4).

$$L(x,y,\delta)=G(x,y,\delta)*I(x,y) \tag{4}$$

In formula (4), (x,y) is I the coordinate of the pixel point of the image, δ which is the scale space factor, and its value determines the smoothness of the image. The δ smaller the value, the more obvious the detailed features of the image, which $*$ means convolution, which G means the Gaussian function, such as formula (5) shown.

$$G=\frac{1}{2\pi\delta^2}e^{-\frac{\left(x-\frac{m}{2}\right)^2+\left(y+\frac{n}{2}\right)^2}{2\delta^2}} \tag{5}$$

In formula (5), it m,n represents the dimension of the Gaussian template. SIFT can use Gaussian blur to search for key feature points in each scale space. The calculation of the Gaussian blur template is mainly based on the normal distribution, as shown in formula (6).

$$G(r)=\frac{1}{\sqrt{2\pi\delta^2}^{N}}e^{-\frac{r^2}{2\delta^2}} \tag{6}$$

In formula (6), δ is the standard deviation of the normal distribution, the smaller the value, the clearer the image; r it represents the distance between the elements in the fuzzy template and the center of the template; N it represents the dimension of the space where the normal distribution is located. Gaussian blurring is performed on the image at different scales and downsampling is performed to construct a Gaussian image pyramid, as shown in Fig. 2.

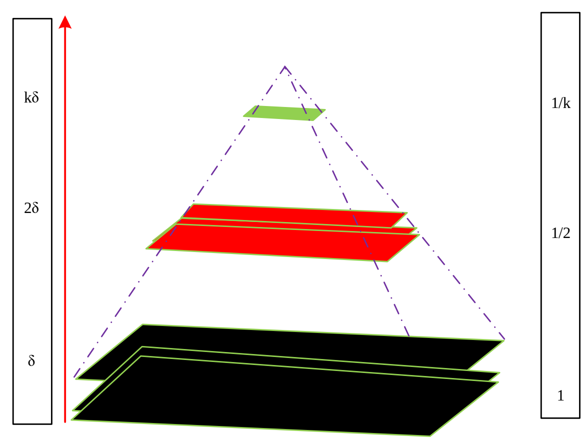


Fig. 2. Gaussian image pyramid

The difference between each layer of the Gaussian pyramid is recorded as Difference of Gaussian (DOG), and the key feature points can be obtained by using the difference Gaussian to describe the image features. The extreme points in the DOG space are sampled and detected to locate key feature points. DOG is greatly affected by edges and noise. In order to obtain a more stable image feature description, it is necessary to remove low-contrast and unstable points, which are realized by using Harris Corner detector and two-dimensional Hessian matrix respectively. After obtaining the stable extreme feature points, in order to describe the rotation invariance of the feature points, the pixel gradient and direction of the feature point neighborhood are obtained by calculating the image gradient, as shown in Fig. 3.

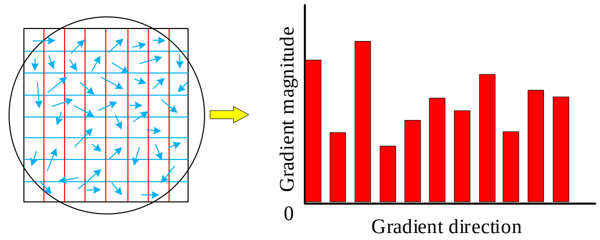


Fig. 3. Description of feature point direction

Based on the above content, information such as the position, scale, and direction of the feature points can be obtained, so as to ensure the uniqueness of the feature points. After using the SIFT algorithm to obtain the eigenvalues of an image, the eigenvalues of the image are matched with the eigenvalues of other images to achieve image recognition. However, the eigenvalues between images are not in a one-to-one correspondence, so it is necessary to quickly match the eigenvalues between images through data indexing [27]. The research uses KD-Tree algorithm crop eigenvalue matching algorithm. For two images A and B, first obtain a set of eigenvalues from image A, denoted as $F_a = \left\{ y_1^a, y_2^a, \ldots, y_N^a \right\}$, and denote the set of eigenvalues in image B $F_b = \left\{ y_1^b, y_2^b, \ldots, y_N^b \right\}$ as.

$$d\left(F_a, F_b\right) = \sqrt{\sum_{i=1}^{n} \left(y_i^a - y_i^b\right)^2} \tag{7}$$

After obtaining the ratio between the closest distance and the second closest distance, compare whether the ratio satisfies the set threshold, so as to determine whether it is a similar point. Repeat the above operation to match all similar feature points. Through the above content, we can realize trademark recognition and judge whether the trademark design is repeated on a large scale.

B. Optimization path of Trademark Recognition Model based on SIFT Algorithm

When the trademark recognition model based on SIFT algorithm is used for trademark recognition, its recognition effect is more dependent on the quality of the extracted feature values. However, the feature dimension extracted by the SIFT algorithm is high, which leads to long image recognition time and unsatisfactory recognition accuracy. Aiming at the defect that the dimension of the image feature vector extracted by the SIFT algorithm is too high, the PCA algorithm is used to reduce the dimension of the feature vector. Since the image feature vector extracted by the SIFT algorithm is basically 128-dimensional, the dimension is relatively high, and it is very easy to cause a dimensional disaster, resulting in an exponential increase in the amount of data calculation, which seriously affects the accuracy and efficiency of trademark recognition. The PCA algorithm can simplify the 128-dimensional information, thereby reducing the dimension of the eigenvalue descriptor. The dimensionality reduction process of PCA for eigenvalues is: obtaining a 128-dimensional matrix vector C_X, and each column in the matrix represents an eigenvector. Calculate the covariance of all eigenvectors Σ_X, and obtain the eigenvalues of the covariance matrix by solving Equation (8).

$$\det\left(\lambda_i I - \Sigma_X\right) = 0 \tag{8}$$

In formula (8), λ_i represents the i th eigenvector. According to the solution of formula (8), the corresponding eigenvectors are obtained w_i. Considering the acquired feature vector as a column vector W, the new transformed feature is obtained by formula (9).

$$C_Y = C_X W^T \tag{9}$$

In formula (9), W^T is the column vector matrix. Select eigenvectors with larger values and set eigenvectors with smaller values to 0, thereby reducing the eigenvector dimension. The identification and matching of trademark images can be regarded as a binary classification problem. In view of the problem of too many feature points in feature matching, which leads to low efficiency and accuracy, SVM is used to match feature values [28]. SVM is essentially a linear classifier, which has excellent performance for nonlinear identification problems in high-dimensional space. The mathematical model of SVM can be expressed as Equation (10).

$$f(x) = sign(wx + b) \tag{10}$$

In formula (10), w the normal vector of the b classification surface is represented, and the constant term of the classification surface is represented. When formula (11) is satisfied, it means that the current hyperplane is the optimal hyperplane.

$$wx + b = 0 \tag{11}$$

After obtaining the optimal hyperplane by formula (11), in order to make the model more optimized, it is necessary to find the farthest distance from the sample points of different sample classifications in the data set to the hyperplane, so as to find the maximum interval hyperplane. The specific model is shown in the Fig. 4.

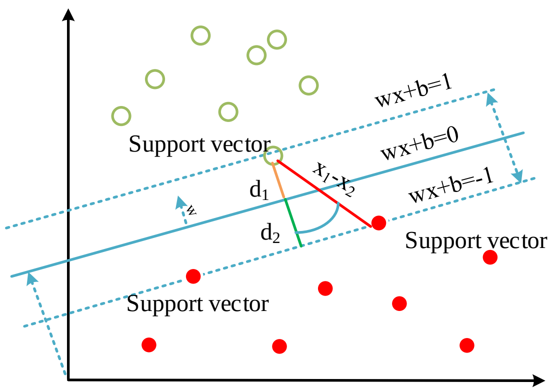


Fig. 4. Schematic diagram after optimization of support vector machine

In Fig. 4, the maximum margin hyperplane solution can be transformed into the function minima problem in Eq. (12).

$$\varphi(w) = \frac{1}{2}\left\|w^2\right\| \tag{12}$$

By formula (12), some linear problems can be solved. However, for nonlinear problems, SVM needs to use the conversion function to upgrade the original space to convert the original problem into a linear problem in a high-dimensional space, so as to facilitate the calculation and solution of the optimal classification hyperplane in the high-dimensional space, and finally to the non-linear problem. Solve linear problems. In SVM, the process of converting a nonlinear problem into a linear problem using a transfer function is shown in Fig. 5.

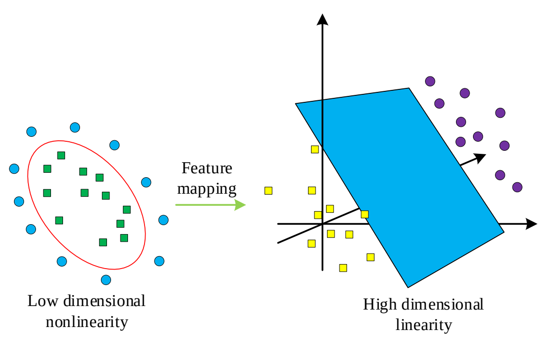


Fig. 5. The process of transforming a nonlinear problem into a linear problem

In the research, the Sigmoid kernel function is used as the transformation function. The expression of the sigmoid kernel function is shown in formula (13).

$$K(x_1, x_2) = \tanh\left(\eta\langle x_1, x_2\rangle + \theta\right) \tag{13}$$

In Formula (13), η is the learning rate and θ is the vector feature. Nonlinear mapping $\phi : R^d \to H$ can map the sample data in the input space to the high-dimensional feature space H, the kernel function K can be used to obtain formula (14).

$$K(x_1, x_2) = \phi(x_i)\cdot\phi(x_j) \tag{14}$$

Through the formula (14), the inner product operation is performed in the high-dimensional space to obtain a nonlinear classification decision function. SVM mainly solves the two-class problem, and has poor performance for solving the multi-class problem. The classification of trademark image feature values is a multi-classification problem, so it is necessary to build a multi-classifier according to SVM. There are generally two ways to construct the SVM multi-classifier. The first is to modify the objective function of the multi-classification problem, so as to solve the multi-classification problem at one time. However, this method requires a huge amount of calculation, takes a long time, and the establishment of the objective function is very difficult. Therefore, another method is used to construct SVM multi-classifiers in this study, that is, construct n binary classifiers, and perform the analysis on these binary classifiers. Combinatorial solutions to solve multi-class problems. When the feature values of the image have N classes, the number of binary classifiers of SVM is shown in formula (15).

$$n = \frac{N(N-1)}{2} \tag{15}$$

By combining the binary classifiers, the voting method is used to obtain the category of the feature value of the trademark image. Based on the above content, a trademark recognition model based on SIFT-PCA-SVM is constructed, so as to realize the recognition and matching of trademark images, prevent legal disputes caused by plagiarism and infringement, and avoid economic and reputation losses of enterprises and advertising designers.

IV. Performance Analysis of Trademark Recognition Model Based on SIFT-PCA-SVM

Aiming at the problem that the current trademark design is prone to creative duplication and legal disputes, this paper studies the construction of a trademark recognition model based on SIFT-PCA-SVM. In order to verify the performance of the model, the research uses the FlickrLogos dataset for experiments. The FlickrLogos dataset contains images related to 32 kinds of trademarks. The study selects 3000 images, of which 2500 images are used as the training sample set to train the model, and the other 500 images are used as the test sample set to test the performance of the model. The test environment is as follows. The operating system uses Windows, OpenCV3.0 alpha as the development tool, and Python as the programming tool. The processor is Intel (R) Core (TM), and the memory is 6GB.The data of training sample set and test sample set are preprocessed by the method proposed in the study. Build a trademark recognition model based on SIFT algorithm (Model 1), a trademark recognition model based on SIFT-SVM (Model 2) and a trademark recognition model based on SIFT-PCA-SVM (Model 3), respectively, and compare the training of the above three models Effects and test effects to validate the performance of the proposed trademark identification method in the study. The three models are trained using the training sample set. During the training process, the performance of the three models is shown in Fig. 6.

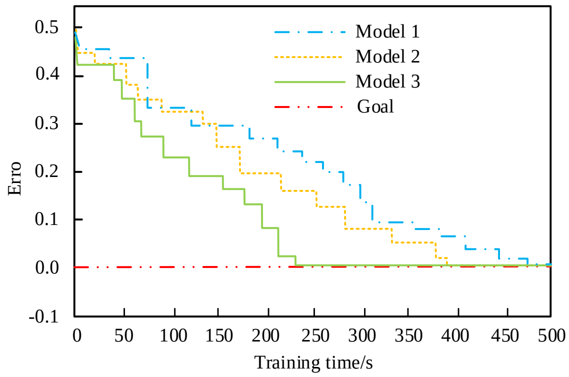


Fig. 6. The performance of the three models in the training process

As can be seen in Fig. 6, the training performance of Model 3 is significantly better than the other two models. Model 3 achieves the target accuracy when the training time is 241 seconds; Model 2 achieves the target accuracy when the training time is 393 seconds, which is 152 seconds more than the training time of Model 3; Model 1 achieves the target accuracy when the training time is 476 seconds, more than Model 3 takes 235 seconds more to train. From the above results, it can be shown that the training performance of Model 3 is better, which can greatly save training time and improve the efficiency of trademark recognition. Aiming at the defect that the dimension of image feature vector extracted by SIFT algorithm is too high, the PCA algorithm is used to reduce the dimension of its feature vector , so as to reduce the number of invalid feature points extracted and improve the feature extraction efficiency and feature matching accuracy. In order to verify the improvement effect of the PCA algorithm on the SIFT algorithm, the study randomly selects three trademark images from the FlickrLogos data set, denoted as image A, image B and image C respectively. Three images are used to test the feature point extraction effect of SIFT algorithm and PCA-SIFT algorithm. The number of feature points extracted by the SIFT algorithm and the PCA-SIFT algorithm for the three images and the matching accuracy of the feature points are shown in Fig. 7.

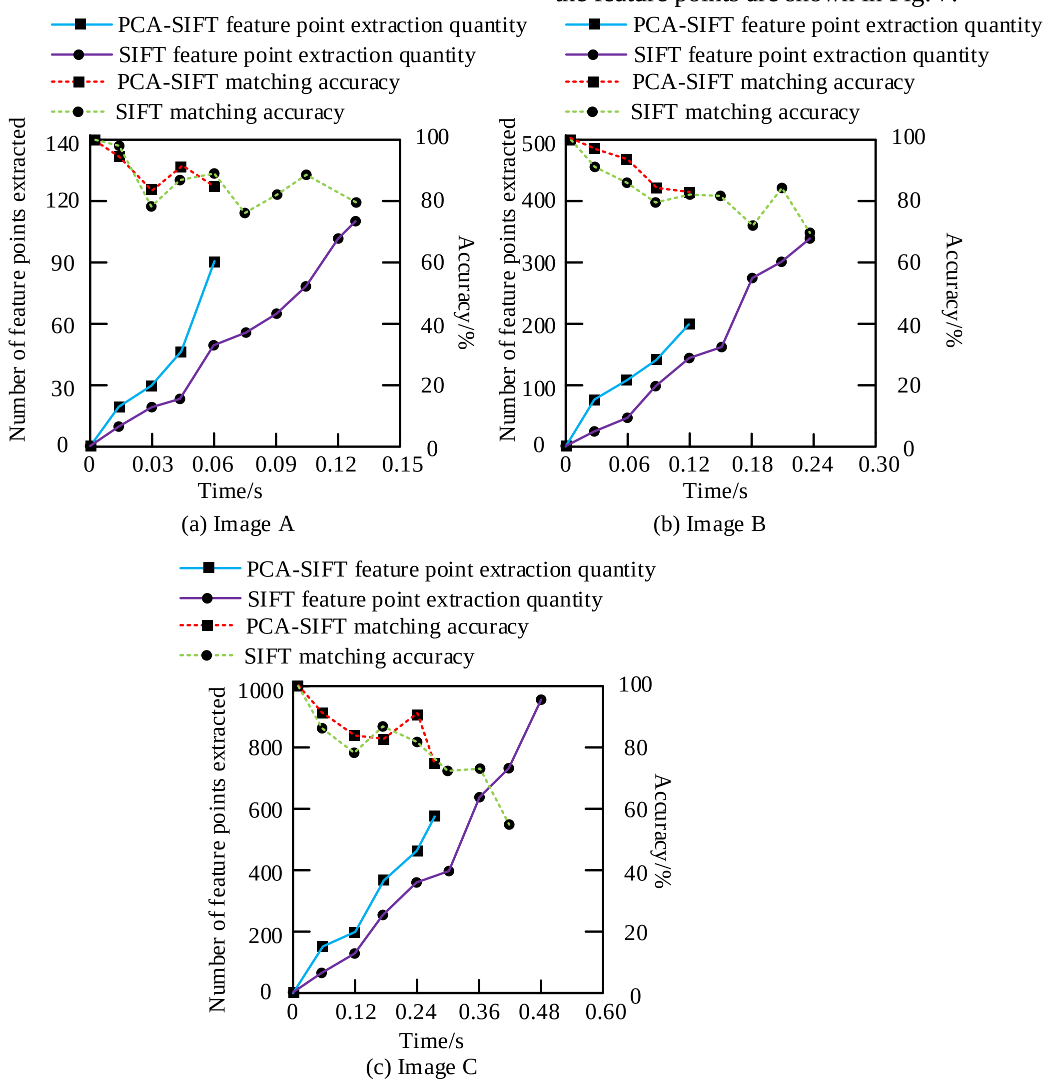


Fig. 7. Improvement effect of PCA algorithm on SIFT algorithm

In Fig. 7, it can be seen that when processing the same image, the number of image feature points and feature point extraction time of the PCA-SIFT algorithm are significantly lower than those of the SIFT algorithm, and the feature point matching accuracy is significantly higher than that of the SIFT algorithm. For the three images, the average extraction time of feature points of PCA-SIFT algorithm is 0.16s, the average number of feature points extracted is 284, and the average matching accuracy of final feature points is 80.62%; the average extraction time of feature points of SIFT algorithm is 0.31s, Compared with the PCA-SIFT algorithm, the average number of feature points extracted is 485.2, which is 201.2 more than that of the PCA-SIFT algorithm. The average matching accuracy of the final feature points is 78.41%, which is 2.21% lower than that of the PCA-SIFT algorithm. The identification and matching of trademark images can be regarded as a binary classification problem. In view of the problem of too many feature points in feature matching, which leads to low efficiency and accuracy, SVM is used to match feature values. In order to verify the recognition accuracy of the proposed trademark recognition model, a test sample set was used to test Model 1, Model 2 and Model 3 to compare the trademark recognition accuracy of the three models, as shown in Fig. 8.

In Fig. 8, the trademark recognition accuracy of Model 3 is significantly better than Model 1 and Model 2, and the number of iterations required to achieve the best accuracy is also significantly less than Model 1 and Model 2. The trademark recognition accuracy of Model 3 is 98.82 %, which is 0.66% higher than Model 1's 98.16% and 0.58% higher than Model 2's 98.24%. In the FlickrLogos dataset, images of four trademarks including Adidas, Aldi, Apple, and Becks were selected for testing. For the above four trademarks, the recognition time and recognition accuracy of the three models are shown in Table I.

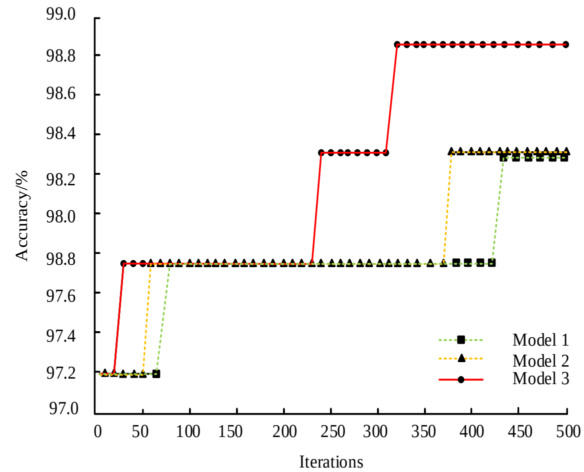


Fig. 8. Trademark recognition accuracy of three models

TABLE I. RECOGNITION TIME AND ACCURACY OF THREE MODELS

Model	Trademark	Time/s	Accuracy/%
Model 1	Adidas	68.6	89.32
Model 2		32.4	93.65
Model 3		35.3	98.42
Model 1	Aldi	76.6	86.46
Model 2		44.3	92.03
Model 3		45.6	98.14
Model 1	Apple	55.2	90.06
Model 2		24.0	96.33
Model 3		26.8	99.25
Model 1	Becks	108.4	88.44
Model 2		65.3	91.09
Model 3		70.2	98.62
Model 1	Average	77.2	88.57
Model 2		41.5	93.28
Model 3		44.5	98.61

In Table I, the trademark recognition accuracy of model 3 is significantly higher than that of model 1 and model 2, and the recognition time of model 3 is much shorter than that of model 1, slightly higher than that of model 2. Among them, in the identification of Adidas trademark, the identification time of model 3 is 35.3s, and the identification accuracy is 98.42%; The recognition time of model 2 is 32.4s, 2.9s less than model 3, and the recognition accuracy is 93.65%, 4.77% less than model 3; The recognition time of model 1 is 68.6s, 33.3s more than model 3, and the recognition accuracy is 89.32%, 9.10% less than model 3. This shows that the accuracy and efficiency of Model 3 are much higher than those of Model 1 and Model 2. In order to further evaluate the performance of the three models, the area under the ROC curve is introduced to evaluate the three models. Generally speaking, the larger the

AUC value is, the better the performance of the model is. Test the two prediction models with the test data sample set, and the AUC curves of the three models are shown in Fig. 9.

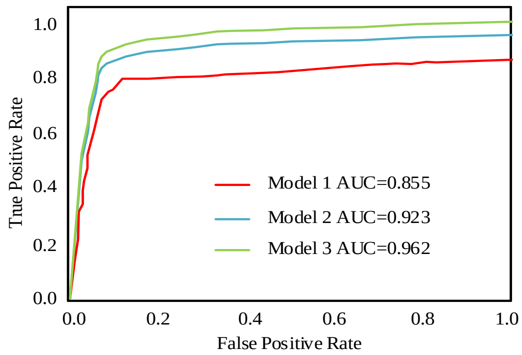


Fig. 9. AUC curve of three models

As can be seen in Fig. 9, the AUC value of Model 3 is 0.962, which is 0.039 higher than Model 2's 0.923 and 0.107 higher than Model 1's 0.855. To sum up, the proposed SIFT-PCA-SVM-based trademark recognition model can identify trademarks with high efficiency and high precision, so as to avoid legal disputes caused by infringement and plagiarism, and has positive significance for advertising design.

V. Conclusion

Aiming at the problem that the current trademark design pattern is easy to repeat and causes legal disputes such as plagiarism and infringement, an SIFT-PCA-SVM algorithm is proposed to build a trademark recognition model. The experimental results show that model 3 achieves the target accuracy when the training time is 241s, 152 seconds less than model 2, and 235 seconds less than model 1. The trademark recognition accuracy of model 3 is 98.82%, 0.66% higher than 98.16% of model 1, and 0.58% higher than 98.24% of model 2. The area under ROC curve is introduced to evaluate the three models. AUC value of model 3 is 0.962, 0.039 higher than 0.923 of model 2, and 0.107 higher than 0.855 of model 1. In conclusion, the trademark identification model based on SIFT-PCA-SVM proposed in the study can identify trademarks efficiently and accurately, so as to avoid legal disputes caused by infringement, plagiarism and other issues, which is of positive significance to advertising design. The trademark recognition model proposed in the study is mainly used to identify the problem of pattern type advertisement design plagiarism. The recognition effect of video and audio advertisement design plagiarism is poor, which is also the direction that needs to be improved in the next step of the study.

References

[1] S. Wang, X. Lehto, "The Interplay of Travelers' Psychological Distance, Language Abstraction, and Message Appeal Type in Social Media Advertising," Journal of Travel Research, Vol. 59, No. 8, pp. 1430-1446, 2020.

[2] S. Bellman, J. Murphy, S.V. Arismendez, D. Varan. "How TV sponsorship can help television spot advertising," European Journal of Marketing, Vol. 53, No. 1, pp. 121-136, 2019.

[3] W. Yousaf, A. Umar, S.H. Shirazi, S.H. Shirazi, Z. Khan, M. Zaka. "Patch-CNN: Deep learning for logo detection and brand recognition," Journal of Intelligent and Fuzzy Systems, Vol. 40, No. 2, pp. 1- 14, 2021.

[4] Q. Yu, S. Zhou, Y. Jiang, P. Wu, Y. Xu. "High-Performance SAR Image Matching Using Improved SIFT Framework Based on Rolling Guidance Filter and ROEWA-Powered Feature," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 12, No. 3, pp. 920-933, 2019.

[5] Z. Zhou, Q. Wu, S. Wan, W. Sun, X. Sun. "Integrating SIFT and CNN Feature Matching for Partial-Duplicate Image Detection," IEEE Transactions on Emerging Topics in Computational Intelligence, Vol. 4, No. 5, pp. 593-604, 2020.

[6] S.J. Choi, M.E. Johnson, "Understanding the relationship between data breaches and hospital advertising expenditures," The American Journal of Managed Care, Vol. 25, No. 1, pp. e14-e20, 2019.

[7] S. Kininmonth "Anything That Can Be Traded, Will Be Traded": The Contests to Automate and Financialize Advertising Futures Markets: Television & New Media, Vol. 23, No. 6, pp. 629-645, 2022.

[8] C.H. Shieh, Y. Xu, I.L. Ling, "How location-based advertising elicits in-store purchase," Journal of Services Marketing, Vol. 33, No. 4, pp. 380-395, 2019.

[9] M.J. Wright, M. Avis, I. Santoso, G. Trinh, "Mind the attention gap: how does digital advertising impact choice under low attention," European Journal of Marketing, Vol. 56, No. 2, pp. 442-466, 2022.

[10] Lacoste-Badie S, Minvielle M, Droulers O. "Attention to food health warnings in children's advertising: a French perspective." Public Health, Vol. 173, pp. 69-74, 2019.

[11] J.A. Busser, L.V. Shulga, "Involvement in consumer-generated advertising: Effects of organizational transparency and brand authenticity on loyalty and trust." International Journal of Contemporary Hospitality Management, Vol. 31, No. 4, pp. 1763-1784, 2019.

[12] E. Levin, T.N. Quach, P. Thaichon, "Enhancing client-agency relationship quality in the advertising industry - an application of project management," The Journal of Business & Industrial Marketing, Vol. 34, No. 2, pp. 463-473, 2019.

[13] M. Méndez-Suárez, A. Monfort, "The amplifying effect of branded queries on advertising in multi-channel retailing." Journal of Business Research, Vol. 112, pp. 254-260, 2020

[14] C.H. Basch, M. Leblanc, D. Ethan, C.E. Basch. "Sugar sweetened beverages on emerging outdoor advertising in New York City," Public Health, Vol. 167, No. 4, pp. 38-40, 2019.

[15] S. Sahin, S. Baloglu, E. Topcuoglu. "The Influence of Green Message Types on Advertising Effectiveness for Luxury and Budget Hotel Segments," Cornell Hospitality Quarterly, Vol. 61, NO. 4, pp. 443-460, 2020.

[16] LI, F. Yanshan, X. Leidong, X. Wei, "TGSIFT: Robust SIFT Descriptor Based on Tensor Gradient for Hyperspectral Images. Chinese Journal of Electronics, Vol. 29, NO. 05, pp. 128-137. 2020

[17] J. Li, X. Shi, Z.H. You, Z.Chen, M. Fang. "Using Weighted Extreme Learning Machine Combined With Scale-Invariant Feature Transform to Predict Protein-Protein Interactions From Protein Evolutionary Information," IEEE/ACM Transactions on Computational Biology and Bioinformatics, PP(99):1-1, 2020.

[18] Henan, Shigang, Wang, Yan, Zhao, Jian, Wei, Meilan, Piao. "3D view image reconstruction in integral computational imaging using scale invariant feature transform and patch matching," Optics express, Vol. 27, No. 17, pp. 24207- 24222, 2019.

[19] K. Raveendra, R. Vinothkanna, "Hybrid ant colony optimization model for image retrieval using scale-invariant feature transform local descriptor," Computers & Electrical Engineering, Vol. 74, pp. 281-291, 2019.

[20] S.W. Hwang, T. Lee, H. Kim, H. Chuang, J.G. Choi, H. Yeo, "Classification of wood knots using artificial neural networks with texture and local feature-based image descriptors," Holzforschung, Vol. 76, No. 1, pp. 1-13, 2022.

[21] I. Wulandari, "Fusi Citra Dengan Scale Invariant Feature Transform (Sift) Sebagai Registrasi Citra," Jurnal Ilmiah Informatika Komputer, Vol. 25, No. 2, pp. 137-146, 2020.

[22] A.L. Prasasti, "Design Of Foreign Currency Recognition Application Using Scale Invariant Feature Transform (Sift) Method Based On Android (Case Study: Singapore Dollar)," Journal of Engineering and Applied Science, Vol. 14, No. 19, pp. 6991-6997, 2019.

[23] R. Dalai, K.K. Senapati, "A MASK-RCNN Based Approach Using Scale Invariant Feature Transform Key points for Object Detection from Uniform Background Scene," Advances In Image and Video Processing, Vol. 7, No. 5, pp. 01-08, 2019.

[24] A. Karim, S. Mahdi, A.A.Mohammed, "Arabic Handwriting Word Recognition Based on Scale Invariant Feature Transform and Support Vector Machine," Iraqi Journal of Science, Vol. 60, No. 2, pp. 381-387, 2019.

[25] I. Misra, M.K. Rohil, S.M. Moorthi, D. Dhar, "FIRM: Framework for Image Registration Using Multistage Feature Detection and Mode-Guided Motion Smoothness Keypoint Optimization," IEEE Transactions on Geoscience and Remote Sensing, PP(99):1-12, 2021.

[26] I. Aranguren, A. Valdivia, M. Pérez-Cisneros, et al. "Digital image thresholding by using a lateral inhibition 2D histogram and a Mutated Electromagnetic Field Optimization," Multimedia Tools and Applications, Vol. 81, No. 7, pp. 10023-10049, 2022.

[27] H. Hu, Z.Zhang, Z. Xie, et al. "Local relation networks for image recognition," Proceedings of the IEEE/CVF International Conference on Computer Vision.: pp. 3464-3473, 2019

[28] M. Hosseinzadeh, A. M. Rahmani, B.Vo, et al. "Improving security using SVM-based anomaly detection: issues and challenge." Soft Computing, Vol. 25, No. 4, pp. 3195-3223, 2021.

Utilizing Deep Learning in Arabic Text Classification Sentiment Analysis of Twitter

Nehad M. Ibrahim[1], Wael M.S. Yafooz[2], Abdel-Hamid M. Emara[3], Ahmed Abdel-Wahab[4]

Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia[1]

Department of Computer Science-College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia[2]

Department of Computers and Systems Engineering-Faculty of Engineering, Al-Azhar University, Cairo 11884, Egypt[3]

Arab Open University, Riyadh, Saudi Arabia[4]

***Abstract*—The number of social media users has increased. These users share and reshare their ideas in posts and this information can be mined and used by decision-makers in different domains, who analyse and study user opinions on social media networks to improve the quality of products or study specific phenomena. During the COVID-19 pandemic, social media was used to make decisions to limit the spread of the disease using sentiment analysis. Substantial research on this topic has been done; however, there are limited Arabic textual resources on social media. This has resulted in fewer quality sentiment analyses on Arabic texts. This study proposes a model for Arabic sentiment analysis using a Twitter dataset and deep learning models with Arabic word embedding. It uses the supervised deep learning algorithms on the proposed dataset. The dataset contains 51,000 tweets, of which 8,820 are classified as positive, 37,360 neutral, and 8,820 as negative. After cleaning it will contain 31,413. The experiment has been carried out by applying the deep learning models, Convolutional Neural Network and Long Short-Term Memory while comparing the results of different machine learning techniques such as Naive Bayes and Support Vector Machine. The accuracy of the AraBERT model is 0.92% when applying the test on 3,505 tweets.**

***Keywords*—Arabic sentiment analysis; machine learning; convolutional neural networks; word embedding; Arabic word2Vec; long short-term method; AraBERT**

I. Introduction

Recently, sentiment analysis has been prioritized by researchers because it plays an important role in many domains. It is primarily used to study user feedback (user opinion) on a specific event, product or social phenomenon. Many studies have proposed models, approaches or novel databases to predicate and detect user opinions. These methods use machine learning classifiers, deep learning models and natural language techniques as pre-processing methods. Most of the sentiment analysis research focuses on languages other than Arabic. Recent Natural Language Processing research is now increasingly focused on using deep neural learning [1]. Some research initiatives are being launched in a competition funded by the King Abdullah University of Science and Technology (KAUST). They focus on the Arabic language and some individual research efforts.

Generally, in other languages, specifically English, the universal language has proven to be significant due to the vast amount of data contributed by users on social networks (Facebook, Twitter, etc.). In machine learning, a classification known as supervised learning is used in sentiment analysis. There are several methods used in sentiment analysis which can be categorized into binary classification, multi-classification, polarity, multilingual and aspect-based sentiment analysis. In binary classification, the classes can be represented only as positive and negative. In multi-class, there are more than two classes. Additionally, there are classifiers used in binary classification such as DT and TH, while KNN and LR are used in multi-classification. Polarity in sentiment analysis is based on a dictionary that assigns a score to each word. Multilingual sentiment analysis requires many pre-processing steps to be performed in option detection and aspect-based. It is focused on one aspect, concept or word.

To the best of our scholarly knowledge, less attention has been given to Arabic sentiment analysis and there are fewer public Arabic datasets [2]. Therefore, this paper proposes a model for Arabic sentiment analysis based on the proposed dataset. This work uses supervised deep learning algorithms. The original dataset before the cleaning process contains 51,000 tweets classified as 8,820 positive, 37,360 neutral and 8,820 negative. After cleaning, it contains 31,413 tweets classified as 4,855 positive, 21,842 neutral and 4,716 negative. This work introduces and applies deep learning methods on Arabic sentiment analysis text multi-classes with parameter optimization, and improves the process in the text pre-processing area. We apply the deep learning methods Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) to compare the results of different supervised machine learning techniques such as Naive Bayes (NB) and Support Vector Machine (SVM). The accuracy of the best CNN model is 95.8% and the accuracy of LSTM is 96.6%, which are better than the SVM and NB results, which are 82.5% and 69.4%, respectively. We used BERT pre-trained specifically in Arabic to achieve the same success that BERT achieved in English [3]. Based on a review of the literature and the high accuracy achieved in the deep learning models, the main contributions of this paper can be summarized as follows.

- Develop a model for Arabic sentiment analysis using machine learning and deep learning models.
- Explore the most recent approaches to Arabic sentiment analysis.
- Propose a novel dataset called ASAD that is publicly available.
- Perform a comparative analysis of the results of MLC and DLM.

The remainder of this paper is organized as follows: Section II presents an overview of related studies on sentiment analysis. Our research methods and materials are explained in Section III. Section IV presents the results. The conclusion is in Section V.

II. Related Studies

Scholars have not given enough attention to the Arabic Sentiment Analysis Dataset (ASAD). It [4] provides a comprehensive overview of a new Twitter-based benchmark dataset for Arabic sentiment analysis. ASAD is a massive, high-quality annotated dataset (including 95,000 tweets) with three-class sentiment labels, compared to other publicly released Arabic datasets (positive, negative and neutral) in [5], researched Twitter's sentiment analysis. Three machine learning algorithms are used, Logistic Regression, Aid Vector Classification, and NB, with two sets of characteristics. The word frequency approach, word embeddings, and machine learning classifiers can correctly identify rumour-related tweets with 84% accuracy, which classifies tweets into four categories: academic, media, government and health [6].

In [7], two new Arabic text categorization datasets are introduced. The first consists of Twitter, Facebook and YouTube posts from well-known Arabic news channels, and the second consists of tweets from popular Arabic accounts. In modern standard Arabic, the papers in the former are almost entirely written (MSA), while the tweets in the latter contain both MSA and dialectal Arabic.

In [8], Word2Vec models were collected from 10 newspapers in different Arabic countries from a broad Arabic corpus. The reports increased the accuracy of sentiment classification after applying various machine learning algorithms and convolution neural networks with different text feature choices (91%–95%).

In [9], an in-house-built dataset shows tweets and comments where three classifiers were applied, including NB, SVM and K-Nearest Neighbour, in particular. The findings show that SVM provides the highest accuracy, while KNN (K=10) provides the highest recall.

In [10], four classifiers were trained to incorporate a dataset consisting of 4,712 tweets to conduct a comparative study on the output of the classifiers, namely NB, SVM, Multinomial Logistic Regression and K-Nearest Neighbour. When running against the tweet's dataset, these algorithms revealed that SVM gives the highest F1 score (72.0), while KNN (K=2) achieved the best accuracy, equivalent to 92.0.0.

In [11] the processes of gathering Twitter data and filtering, pre-processing and annotating the Arabic text to create a large dataset of sentiment analysis in Arabic are summarized. In addition to deep learning and CNN, machine learning algorithms (NB, SVM, and Logistic Regression) were used on the health dataset in the sentiment analysis experiments. The keywords are Machine Learning, Deep Neural Networks, Arabic Language and Emotion Analysis.

Several versions of RNN and CNN classifiers using GloVe-based word embedding were introduced. All classifiers performed well, while the classifiers between 90% and 91% had the highest accuracy. Experimental findings indicate that BRAD 2.0 is rich and stable [12]. To encourage more study in the field of Arabic computational linguistics, the benchmark dataset was made available as the key contribution.

In [13] a new mix model from CNN and LSTM is proposed using vector representations of sentences and the SoftMax regression classifier to identify the sentiment tendencies in the text. In [14] a method for evolving the CNN and creating an Arabic sentiment classification system is proposed using the differential evolution (DE) algorithm. In [15] a novel architecture for Arabic word classification and understanding is proposed. This is based on CNNs and recurrent neural networks that address the difficulty of handling unstructured social media texts in low data availability. Therefore [16] attempts to identify expressions related to feelings, such as happiness, rage, anxiety and sadness. In addition, it presents the emotion classification in Arabic tweets by using the CNNs and compares them with the machine learning methods SVM, NB and Multi-Layer Perceptron (MLP).

In [17] an Arabic sentiment analysis corpus culled from Twitter is presented, consisting of 36,000 tweets categorized as positive or negative, plus 8,000 tweets manually annotated and used to assess the corpus intrinsically by comparing it to human classification and pre-trained sentiment analysis models, with an accuracy of 86%. In [18] a survey is proposed that focuses on 90 recent research papers (74% were published after 2015). In [19] supervised and unsupervised transformation methods, such as principal component analysis (PCA) and latent Dirichlet allocation (LDA), are presented. They are tested on five Arabic opinion text datasets of various domains and sizes (1.6–94,000 reviews). In the two-class classification problem, accuracy values range from 95.5%–99.8%, and for the three-class classification problem, accuracy values ranged from 92%–97.3%.

In [20] a new study is presented to develop a new model to predict an individual's awareness of the precautionary procedures. Tweets related to COVID-19 were collected from the five main regions in Saudi Arabia, and the accuracy level achieved was 85%. A systematic comparative overview of the most appropriate methods for analysing Arabic sentiment is presented in [21]. It carries out a thorough comparison of various machine learning methods for Arabic sentiment analysis, such as NB, SVM, CNN, LSTM and several recently developed language models. The model achieves F-scores of 0.69, 0.76 and 0.92. A method for extracting knowledge from Arabic text on social media in four stages – data collection, cleaning, enrichment and availability – is shown in [22]. It offers an integrated solution for the challenges of pre-processing Arabic text on social media. This was undertaken to investigate the performance metrics as given in [23, 24, 25, 26, 27, 28, 29] and validates the proposed model for small-and large-scale datasets. Disambiguation using the deep learning techniques with the Arabic corpus is presented in [30]. An Arabic model for text clustering using word embedding and Arabic word net is presented in [31].

III. METHODOLOGY

This section describes the research methodology that was used to conduct this research. It consisted of six main interrelated phases: text retrieval; text pre-processing; tokenization and feature extraction; application of the deep learning model; model performance evaluation as shown in Fig. 1; and use of the transfer learning by applying the AraBERT Model.

The analysis concentrated on three cases of tweets, positive, negative and neutral. To perform the sentiment analysis experiment, the large number of collected tweets was necessary. A lot of noisy data was used in the total number of tweets (51,000 tweets). The model architecture is shown in Fig. 1. The first phase in this work is text retrieval; the second is text cleaning; the third is tokenization; the fourth is embedding the text using the Word2Vec corpus in [32]; the fifth is to apply deep learning models; and the last phase is to evaluate the results.

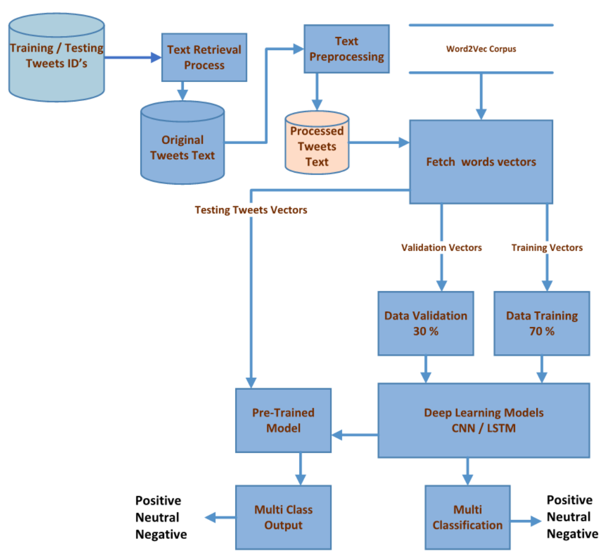


Fig. 1. Model architecture

A. Tweet Text Retrieval

This section describes the first phase, text retrieval. Data retrieval is performed using Tweepy API, and because the text characters are in Arabic, we implemented an Arabic text retrieval module using the Tweepy Twitter API library in Python.

B. Text Pre-Processing

Improving the accuracy of the text classification required enhancing the text features [22]. We added some feature selection improvements, such as noise removal. Removing the noisy characters from the text enhanced the word representation. The following steps were run on the text to remove the noisy data.

- Remove the advertisement tweets.
- Remove the retweeted tweets, which started with the segment 'RT'.
- Remove the duplicate tweets, which were retrieved more than once.

1) Normalization: Normalization is a pre-processing method of text data cleaning, to format a sequence of texts into a standard uniform [33]. It is difficult to analyse Arabic texts because of the nuances of the language, both in terms of infrastructure and conformation. Arabic has an abundance of diverse inflexions, dialects and spellings that change the meanings of the words. Using special labels, called configuration, rather than vowels, they differ according to the shape of the word. This method is necessary and useful in word processing to minimise uncommon terms and increase classification accuracy. The following steps were applied to the stored tweets in the dataset on two levels. The first level retained the emoji and the second level removed the emoji from the data collection phase.

- Remove digits, non-Arabic letters, single letters, punctuations, diacritics and special characters ($, %, &, #, .).
- Separate all words by spaces.
- Remove (ها, وا, ان, ون, ين , ات) from the end of the word.
- Remove (لل, ولل, وتال, وال,كال, وكال, تال, ال) from the beginning of the word except اللهم, الله, إله لله.
- Normalize some characters with a single character, such as (آ, إ , أ) with (ا), (ئ,ي) with (ى) and (ة) with (ه).
- Replace repeated characters (e.g.,يااااااااااااااارب) with a single character (e.g., يارب).

2) Stop words removal: Stop words are words that do not affect the meaning of the sentence, and eliminating the stop words from the text helps to identify the most important words as the following [34].

- Relative pronouns (الأسماء الموصولة)
- Referral names/determiners (أسماء الإحالة / المحددات)
- Transformers (verbs, letters) ((الأفعال والحروف) المحولات)
- Verbal pronouns (الضمائر اللفظية)
- Adverbs (الاحوال)
- Interrogative pronouns (ضمير الاستفهام)
- Conditional pronouns (الضمير مشروط)
- Prepositions (حرف جر)
- Pronouns (الضمائر)

a) Pre-processing: We performed pre-processing operations on the data, such as removing the stop words, special characters, such as '@' and '#', URLs, non-Arabic characters and punctuation to create a clear analysis of the text. In addition, we used NLTK (https://www.nltk.org) word tokenization on the text data (refer Table I)

TABLE I. SAMPLES OF TWEETS BEFORE AND AFTER THE NOISE CLEANING PROCESS

#	Label	Text before cleaning	Text normalized with emoji retained	Text normalized with emoji removed	English meaning
1	Neutral	'@elm عندي مشكله لما تبي اجدد إقامة عامل حاولت أتواصل بس يفصل معي الخط من قبل مقيم'	عندي مشكله لما تبي اجدد إقامة عامل حاولت أتواصل بس يفصل معي الخط من قبل مقيم	عندي مشكله لما تبي اجد إقامة عامل حاولت أتواصل بس يفصل معي الخط من قبل مقيم	I have a problem when I want to find the residency of a worker, I tried to contact me, but the line was separated by a resident
2	Positive	'@nas_alharbi8 والله حسب الأرقام سيكون مخيب للآمال ولكن الأهلي قدها برجاله في الملعب 🖤 '	والله حسب الأرقام سيكون مخيب للآمال ولكن الأهلي قدها برجاله في الملعب 🖤	والله حسب الأرقام سيكون مخيب لآمال ولكن الأهلي قدها برجاله في الملعب	By God, according to the numbers, it will be disappointing, but Al-Ahly has led it with its men on the field
3	Neutral	"' الزعل بيغير ملامحك ، بيغير نظرة العين ، بيغير شكلك في الصور ، الزعل ممكن يطفيك تماما."'	الزعل بيغير ملامحك بيغير نظرة العين بيغير شكلك في الصور الزعل ممكن يطفيك تماما	الزعل بيغير ملامحك بيغير نظرة العين بيغير شكلك في الصور الزعل مكن يطفيك تماما	Upset changes your features, changes the look of the eyes, changes your appearance in the photos, upset you can completely eliminate you
4	Positive	'ثقي قلباً وقالباً معاكِ وفخورين فيكِ وين مارحتي وربي يوفقك ويسهلك☹♥□ طريقك https://t.co/N8vVjcwD08'	' ثقي قلباً وقالباً معاكِ وفخورين فيكِ وين مارحتي وربي يوفقك ويسهلك طريقك☹♥ □	ثقي قلبا وقالبا معاك وفخورين فيك وين مارحتي وربي يوفقك ويسهلك طريقك	Trust your heart and soul with you and be proud of you. Where have you gone, and my Lord will help you and make your way
5	Negative	' هذا المخلوق ينتمي إلى المجتمع الشيطاني ما في منه اي خير . https://t.co/IYoS64SKJZ'	' هذا المخلوق ينتمي إلى المجتمع الشيطاني ما في منه اي خير	هذا المخلوق ينتمي إلى المجتمع الشيطاني ما في منه اي خير	This creature belongs to the satanic community, there is no good in it
6	Negative	' ادور لعيني النوم لو كانه سلف .. من بارحة الاولى و انا اودي و اجيب .'	ادور لعيني النوم لو كانه سلف من بارحة الاولى و انا اولادي و اجيب	ادور لعيني النوم لو كانه سلف من بارحة الاولى و انا اولادي و اجيب	I turn to my eyes to sleep if it was an ancestor from yesterday and I am my children and I answer

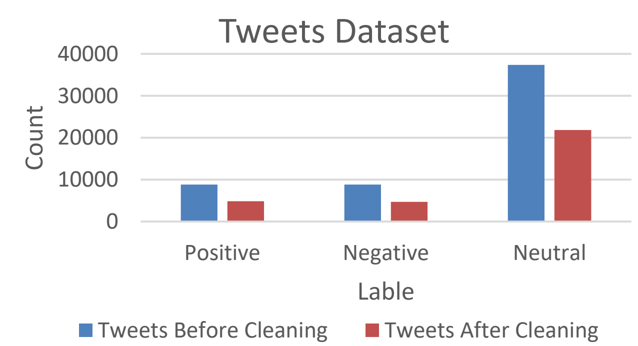


Fig. 2. Dataset statistics

Fig. 2 shows the dataset before and after cleaning. As the dataset published by [34] contains IDs and annotations only, the first step must be to retrieve the tweeter text using the authorized API object from Tweepy, using the API to retrieve all IDs from the file one by one.

3) Tokenization: Tokenization is the method used to break down text into individual tokens separated by white space. Tokenization removes all special characters, determines phrase boundaries, and processes abbreviations and numbers [35]. Due to the morphological complexity of the language, the number of tokens in Arabic can exceed four. Since Arabic words often contain many affixes and clitics, the tokenization process was preceded by a segmentation process to eliminate suffixes.

C. Word Vectors Lookup

Word embedding [36] is a language modelling and feature learning technique, where each word is mapped to a vector of real values in such a way that words have a similar representation with similar meanings. Using neural networks, value learning can be achieved. Word2Vec, which has models such as skip-gram and continuous bag of words (CBOW), is a widely used word embedding method. The likelihood of words occurring in proximity to one another is dependent on both models. Skip-gram allows a word to begin with and to anticipate the words that are likely to accompany it. By predicting a word that is likely to occur based on particular background terms, CBOW reverses that. The CBOW model used to learn domain-specific word embeddings from large amounts of Arabic text was collected from the free online encyclopaedia Wikipedia (2,000,000 words vectors of Word2Vec) to create the corpus defined in [31].

The Word2Vec corpus was used to look up all words' vectors from the corpus. In [11] it is suggested that pre-trained word embeddings trained on very large text corpora, such as the free Word2Vec vectors trained on 100 billion Google news tokens, can provide universal features for use in natural language processing.

D. Deep Learning Models

This section presents the two types of deep learning models used in the experiments. Using deep learning in text classification is powerful for feature extraction. In this work, we improve the CNN model in [13] as shown in Fig. 3, and the LSTM model as shown in Fig. 4 with parameter optimization to improve the classification accuracy and compare it with the different ML methods (KNN, NB and SVM). Additionally, we propose the new CNN model as shown in Table III.

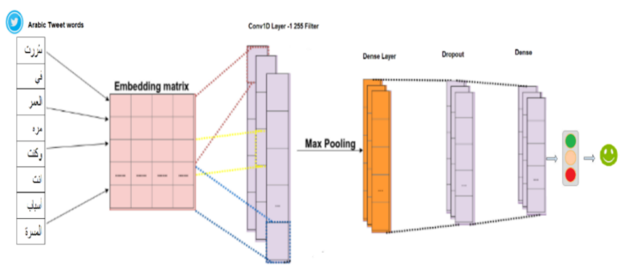


Fig. 3. CNN architecture

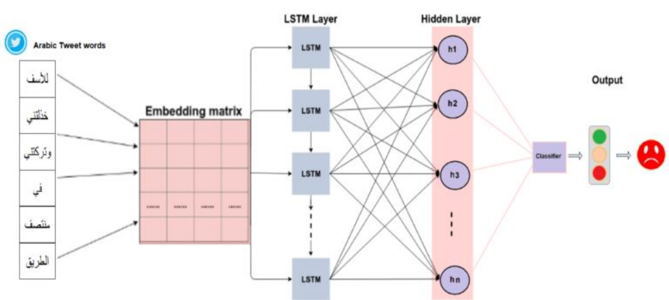


Fig. 4. LSTM architecture

IV. Results

This section presents the results of the conducted experiments. These three experiments, namely, the experiment based on the CNN model, the experiment based on the LSTM model, and the experiment based on the feature methods. These experiments were carried out based on different experiment settings.

A. Experiment Settings

This section describes the experiment settings of all three experiments. The experiment settings and hyperparameter tuning were performed to improve the accuracy of the performance model. In the first experiment, each CNN layer has various parameters such as the number of filters, kernel size, strides, padding, dropout rate, batch size and activation function. All CNN parameters are tuned and optimized to achieve high accuracy. In the second experiment, LTSM is used with the architecture. The embedding layer is used in the LSTM layers as in the CNN model. In the third experiment, the CNN and LSTM models with N-gram ranges were applied to achieve highly accurate model performance. The N-gram ranges method is applied with embedding vectors CBOW and skip-gram. All the parameters and details of the experiment are shown in Table II.

TABLE II. EXPERIMENTS SETTINGS

Parameter	Experiment 1	Experiment 2	Experiment 3
No. of layers	10	22	5
Input shape	300	300	300
Filters	64	64	64
Embedding vectors	CBOW	CBOW	CBOW
Max sequence length	300	300	300
Activation function	SoftMax	SoftMax	SoftMax
Optimization	Adam (learning rate = 0.001)	Adam (learning rate = 0.001)	Adam (learning rate = 0.001)

In these experiments, we applied the SoftMax activation function as shown in formula 1. The key benefit of using SoftMax is the performance probabilities range. The range will vary from 0 to 1, and the sum of all the odds will be equal to 1. If the SoftMax function is used for the multi-classification model it returns the probabilities of each class and the target class would have a high probability.

$$f(S)_i = \frac{e^{S_i}}{\sum_j^C e^{S_j}} \quad (1)$$

While for the loss function in all experiments a 'Categorical Cross-Entropy loss' has been utilized to train a CNN to output a probability over the C classes for each tweet, the target is three classes in the SoftMax activation function at the last layer. The mathematical formula of SoftMax activation and Cross-Entropy loss is shown in formula 2.

$$CE = -\log\left(\frac{e^{S_p}}{\sum_j^C e^{S_j}}\right) \quad (2)$$

where S_p is the CNN score for the positive class.

Before training the model, the dataset is divided into 25,130 training samples and validated on 6,283 samples.

TABLE III. LSTM ARCHITECTURE LAYERS

Layer (type)	Output Shape	Parameters #
input (Input Layer)	(None, 300)	0
embedding (Embedding)	(None, 300, 300)	17421900
LSTM	(None, 18)	22968
Batch Normalization	(None, 18)	72
Dense	(None, 3)	57
Total Parameters: 17,444,997 **Trainable Parameters: 17,444,961** **Non-trainable Parameters: 36**		

In the third experiment, two experiments applied CNN and LSTM models to get a highly accurate model performance. The first experiment used the N-gram ranges method along with embedding vectors using CBOW. The second experiment used the N-gram ranges method to extract the required features, but with embedding vectors using skip-gram.

B. *Experiments Results*

This section explains the results of the experiments. Several experiments were conducted, and these can be categorized into three main experiments, namely, the experiment based on the CNN model, the experiment based on the LSTM model, and the experiment based on the feature methods.

1) Experiments based on CNN model: In this experiment, the CNN model applied two different architectures as shown in Table III, Table IV and Table V. Fig. 5 shows a scatter plot diagram for tweets labelled against the predicted labels. In addition, this figure explains the correlation with different epochs compared to the pre-processing approaches, and whether the emoji were retained or removed.

TABLE IV. CNN-1 ARCHITECTURE LAYERS

Layer (type)	Output Shape	Parameters #
input (Input Layer)	(None, 300)	0
embedding (Embedding)	(None, 300, 300)	17421900
CNN	(None, 300, 64)	96064
Global max pooling	(None, 64)	0
dropout_1 (Dropout)	(None, 64)	0
batch_normalization_1	(None, 64)	256
dense_1 (Dense)	(None, 3)	195
dropout_2 (Dropout)	(None, 3)	0
Batch Normalization	(None, 3)	12
Dense	(None, 3)	12
Total Parameters: 17,518,439 **Trainable Parameters: 17,518,305** **Non-trainable Parameters: 134**		

TABLE V. CNN-2 ARCHITECTURE LAYERS

Layer (type)	Output Shape	Parameters #
input (Input Layer)	(None, 300)	0
embedding (Embedding)	(None, 300, 300)	17421900
CNN-1	(None, 300, 64)	96064
CNN-2	(None, 300, 64)	20544
max pooling	(None, 150,64)	0
CNN-3	(None, 150, 64)	20544
CNN-4	(None, 150, 64)	20544
max pooling	(None, 75,64)	0
CNN-5	(None, 75, 64)	20544
CNN-6	(None, 75, 64)	20544
CNN-7	(None, 75, 64)	20544
max pooling	(None, 75,64)	0
CNN-8	(None, 37, 64)	20544
CNN-9	(None, 37, 64)	20544
CNN-10	(None, 37, 64)	20544
Global max pooling	(None, 75,64)	0
CNN-11	(None, 18, 64)	20544
CNN-12	(None, 18, 64)	20544
CNN-13	(None, 18, 64)	20544
Global max pooling	(None, 64)	0
Batch Normalization	(None, 64)	256
Dense	(None, 3)	195
Total params: 17,764,943 **Trainable params: 17,764,815** **Non-trainable params: 128**		

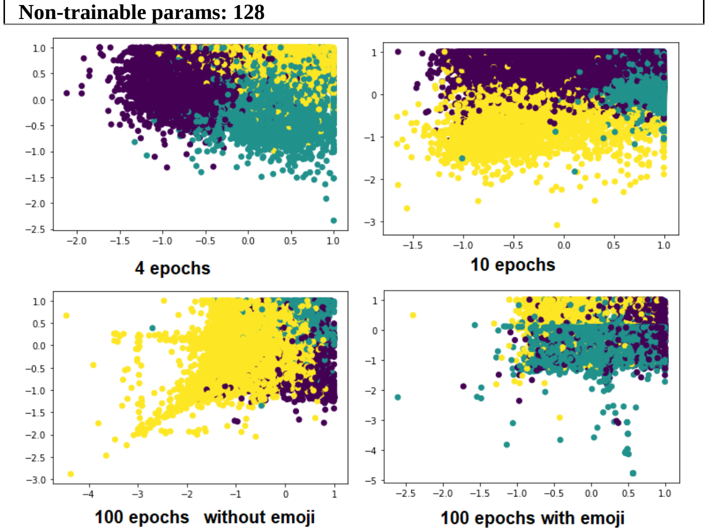


Fig. 5. Scatter plot diagram for tweets labels

2) Experiment based on the LSTM model: In this experiment, the LSTM model has been utilized. LSTM is used to enhance the memorisation of important information. In the text classification, LSTM is used in multiple word strings to identify the class to which it belongs. In this experiment, the dataset has been divided into three groups with different dataset sizes of 3,000 tweets, 15,000 tweets and 31,000 tweets, respectively. The results of this experiment are shown in Evaluation Results of Our Models' Accuracy with Different Dataset Sizes

Model	Accuracy-without emoji	Accuracy-with keeping emoji
SVM	81.79 %	82.5 %
NB	69 %	69.4 %
CNN-1	95 %	95.8 %
CNN-2	70 %	82.7 %
LSTM	96.6 %	95.5 %

TABLE VI. EVALUATION RESULTS OF OUR MODELS' ACCURACY WITH 31,000 TWEETS, WITH EMOJI RETAINED AND REMOVED

3) VI.

Model	Accuracy-without emoji	Accuracy-with keeping emoji
SVM	81.79 %	82.5 %
NB	69 %	69.4 %
CNN-1	95 %	95.8 %
CNN-2	70 %	82.7 %
LSTM	96.6 %	95.5 %

Table VII shows the results of our models' accuracy with the maximum tweets (31,000) and compares the classification accuracy when emoji are retained or removed from the content of the tweets.

TABLE VII. EVALUATION RESULTS OF OUR MODELS' ACCURACY WITH DIFFERENT DATASET SIZES

Model	Accuracy-without emoji	Accuracy-with keeping emoji
SVM	81.79 %	82.5 %
NB	69 %	69.4 %
CNN-1	95 %	95.8 %
CNN-2	70 %	82.7 %
LSTM	96.6 %	95.5 %

TABLE VIII. EVALUATION RESULTS OF OUR MODELS' ACCURACY WITH 31,000 TWEETS, WITH EMOJI RETAINED AND REMOVED

Dataset Size	SVM	NB	CNN-1(9L)	CNN-2(21L)	LSTM
Dataset-1(3000 Tweets)	81 %	23 %	93 %	89 %	90 %
Dataset-1 (15000 Tweets)	78.9 %	63 %	96 %	74.8	90 %
Dataset-1 (31000 Tweets)	82.5 %	69.4%	95.8%	83%	96.6 %

Training and testing accuracy for CNN and LSTM models is shown in Fig. 6, which also shows the accuracy of retaining or removing emoji. This comparison was implemented with different epochs, up to 100 epochs.

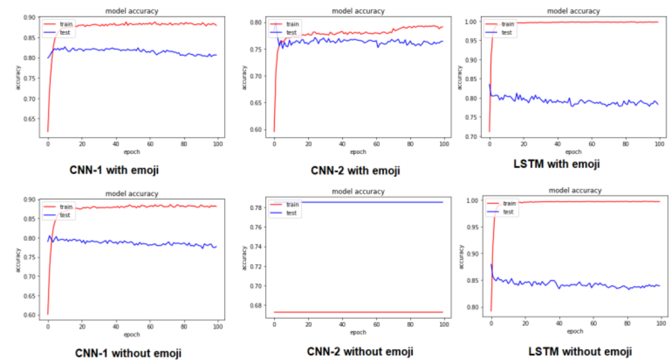


Fig. 6. Deep learning accuracy comparing retaining and removing emoji

4) Experiments based on features methods: In this experiment, two methods have been utilized to examine the accuracy of the model performance. In the first experiment, the two methods used are N-gram and CBOW. The experiments' results are shown in Table VIII.

TABLE IX. CBOW WITH N-GRAM RESULTS COMPARISON

Classifier	Without N-gram	N-gram range 1-2	N-gram range 1-3
SVM	82.22	82.64	83.2
LSTM	96.6	96.7	96.9
CNN-1	95.8	95.82	95.83
CNN-2	83	83.3	84
NB	69.4	69.7	70

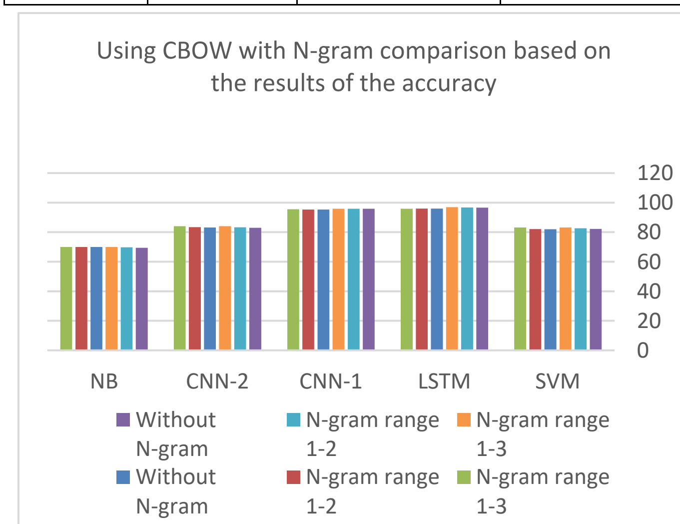


Fig. 7. CBOW and N-gram results comparison

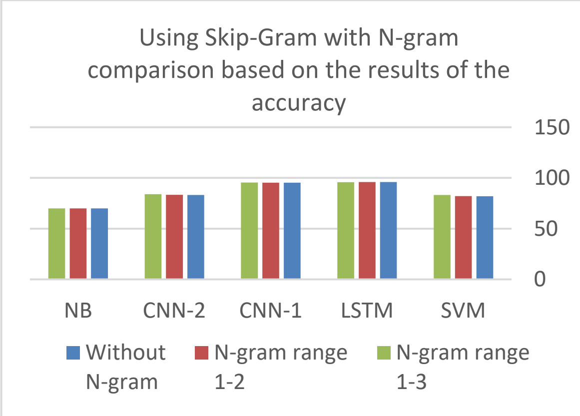


Fig. 8. Skip-gram and N-gram results comparison

In the second experiment, skip-gram and N-gram have been utilized.

TABLE X. SKIP-GRAM AND N-GRAM RESULTS COMPARISON

Classifier	Without N-gram	N-gram range 1-2	N-gram range 1-3
SVM	82	82.06	83.2
LSTM	**96**	**96**	**95.8**
CNN-1	95.3	95.3	95.5
CNN-2	83.2	83.4	84
NB	70	70	70

The accuracy of CNN-1 with emoji retained is 95.8%, and with emoji removed is 95%. The accuracy of CNN-2 with emoji retained is 82.7%, and with emoji removed is 70%. The CNN-1 model is more accurate than CNN-2. The accuracy of LSTM with emoji retained is 95.5% and 96.6% with emoji removed. When applied to SVM and NB the results are 82.5% and 69.4%, respectively. Fig. 8 shows the comparison between the deep learning accuracy with emoji retained and removed. By applying the same models with N-gram and with skip-gram, the results in Table VIII and Table IX, and represented in Fig. 7 and Fig. 8, show the LSTM model is better in both CBOW and skip-gram.

5) Experiment based on the AraBERT model: In this experiment the AraBERT for Sequence Classification is applied (Transformer-based Model for Arabic Language Understanding) with the following parameters:

TABLE XI. ARABERT PARAMETERS

Parameter	Value
Attention probs dropout prob	0.1
Hidden act	gelu
Hidden size	768
Initializer range	0.02
Intermediate size	3072
Layer norm eps	1e-12
Max position embeddings	512
Num attention heads	12
Num hidden layers	12
Transformers version	4.17.0
Vocab size	64000

After training the dataset on the AraBERT model using the parameter list as shown in Table X. The training accuracy (Fig. 9) of the AraBERT model is 0.92% when the test is applied on 3,505 tweets.

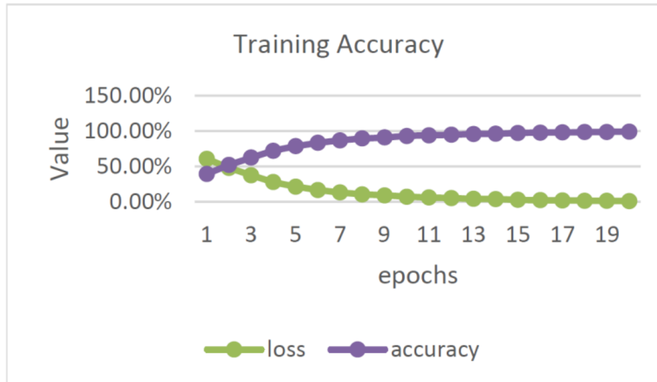


Fig. 9. Results of the training for AraBERT Model

V. CONCLUSION AND FUTURE WORKS

Results of this work show that there is improvement in CNN model accuracy by retaining emoji in text content, and LSTM is more accurate when emoji are removed. These results are summarized in Table II. The accuracy of CNN-1 with emoji retained is 95.8%, and with emoji removed is 95%. The accuracy of CNN-2 with emoji retained is 82.7%, whereas its accuracy with emoji removed is 70%. CNN-1 outperforms CNN-2 in terms of accuracy. The accuracy of LSTM when the emoji are retained is 95.5 %, and it is 96.6% when the emoji are removed. When we use SVM and NB, the outcomes are 82.5% and 69.4 %, respectively. The accuracy of the AraBERT model is 0.92%. In this work, we have shown that the LSTM architecture is the most suitable for the analysis of Arabic tweets. In the future, we can build a new system to analyse Arabic texts using the modern model GPT-3 from Open AI and apply the sentiment analysis on this dataset.

ACKNOWLEDGMENT

The authors would like to thank the Arab Open University for supporting this research paper.

REFERENCES

[1] W. Antoun, F. Baly, and H. Hajj, “AraBERT: Transformer-based Model for Arabic Language Understanding,” arXiv:2003.00104 [cs], Mar. 2021, Accessed: Dec. 01, 2022. [Online]. Available: http://arxiv.org/abs/2003.00104.

[2] Alhejaili, R., Alhazmi, E. S., Alsaeedi, A., & Yafooz, W. M. (2021, September). Sentiment Analysis of The Covid-19 Vaccine For Arabic Tweets Using Machine Learning. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE.

[3] B. Alharbi et al., "ASAD: A Twitter-based Benchmark Arabic Sentiment Analysis Dataset," arXiv:2011.00578 [cs], Mar. 2021, Accessed: Dec. 01, 2022. [Online]. Available: https://arxiv.org/abs/2011.00578.

[4] H. Saif, Y. He, and H. Alani, "Semantic Sentiment Analysis of Twitter," in International semantic web conference, 2012, pp. 508–524.

[5] L. Alsudias and P. Rayson, "COVID-19 and Arabic Twitter : How can Arab World Governments and Public Health Organizations Learn from Social Media?," in ACL 2020 Workshop, 2020, pp. 1–9.

[6] S. A. Chowdhury, A. Abdelali, K. Darwish, J. Soon-Gyo, J. Salminen, and J. B. Jansen, "Improving Arabic text categorization using transformer training diversification," in In Proceedings of the fifth arabic natural language processing workshop, Dec. 2020, pp. 226–236.

[7] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "Improving sentiment analysis in Arabic using word representation," in In 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), Mar. 2018, pp. 13–18.

[8] R. M. Duwairi and I. Qarqaz, "Arabic sentiment analysis using supervised classification," in In 2014 International Conference on Future Internet of Things and Cloud, Aug. 2014, pp. 579–583

[9] R. Ismail, M. Omar, M. Tabir, N. Mahadi, and I. Amin, "Sentiment analysis for Arabic dialect using supervised learning," in In 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Aug. 2018, pp. 1–6.

[10] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "Arabic language sentiment analysis on health services," 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), Apr. 2017, doi: 10.1109/asar.2017.8067771.

[11] A. Elnagar, L. Lulu, and O. Einea, "An Annotated Huge Dataset for Standard and Colloquial Arabic Reviews for Subjective Sentiment Analysis," Procedia Computer Science, vol. 142, pp. 182–189, 2018, doi: 10.1016/j.procs.2018.10.474.

[12] Y. Zhang, J. Zheng, Y. Jiang, G. Huang, and R. Chen, "A Text Sentiment Classification Modeling Method Based on Coordinated CNN-LSTM-Attention Model," Chinese Journal of Electronics, vol. 28, no. 1, pp. 120–126, Jan. 2019, doi: 10.1049/cje.2018.11.004.

[13] A. Dahou, M. A. Elaziz, J. Zhou, and S. Xiong, "Arabic Sentiment Classification Using Convolutional Neural Network and Differential Evolution Algorithm," Computational Intelligence and Neuroscience, vol. 2019, pp. 1–16, Feb. 2019, doi: 10.1155/2019/2537689.

[14] M. Beseiso and H. Elmousalami, "Subword Attentive Model for Arabic Sentiment Analysis," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 19, no. 2, pp. 1–17, Mar. 2020, doi: 10.1145/3360016.

[15] M. Baali and N. Ghneim, "Emotion analysis of Arabic tweets using deep learning approach," Journal of Big Data, vol. 6, no. 1, Oct. 2019, doi: 10.1186/s40537-019-0252-x.

[16] K. Abu Kwaik, S. Chatzikyriakidis, S. Dobnik, M. Saad, and R. Johansson, "An Arabic Tweets Sentiment Analysis Dataset (ATSAD) using Distant Supervision and Self Training," ACLWeb, May 01, 2020. https://www.aclweb.org/anthology/2020.osact-1.1 (accessed Dec. 01, 2022).

[17] I. Guellil, H. Saâdane, F. Azouaou, B. Gueni, and D. Nouvel, "Arabic natural language processing: An overview," Journal of King Saud University - Computer and Information Sciences, Feb. 2019, doi: 10.1016/j.jksuci.2019.02.006.

[18] R. M. K. Saeed, S. Rady, and T. F. Gharib, "Optimizing Sentiment Classification for Arabic Opinion Texts," Cognitive Computation, vol. 13, no. 1, pp. 164–178, Jan. 2021, doi: 10.1007/s12559-020-09771-z.

[19] S. S. Aljameel et al., "A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia," International Journal of Environmental Research and Public Health, vol. 18, no. 1, p. 218, Dec. 2020, doi: 10.3390/ijerph18010218.

[20] I. Abu Farha and W. Magdy, "A comparative study of effective approaches for Arabic sentiment analysis," Information Processing & Management, vol. 58, no. 2, p. 102438, Mar. 2021, doi: 10.1016/j.ipm.2020.102438.

[21] M. O. Hegazi, Y. Al-Dossari, A. Al-Yahy, A. Al-Sumari, and A. Hilal, "Preprocessing Arabic text on social media," Heliyon, vol. 7, no. 2, p. e06191, Feb. 2021, doi: 10.1016/j.heliyon.2021.e06191.

[22] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning Word Vectors for 157 Languages," arXiv preprint arXiv:1802.06893, Aug. 2018.

[23] A. Brahmi and A. E. Abdelkader, "Arabic texts analysis for topic modeling evaluation," Inf. Retr. Boston., vol. 15, no. 1, pp. 33–53, 2012, doi: 10.1007/s10791-011-9171-y.

[24] Z. Kaoudja, B. Khaldi, and M. L. Kherfi, "Arabic artistic script style identification using texture descriptors," CCSSP 2020 - 1st Int. Conf. Commun. Control Syst. Signal Process., pp. 113–118, 2020, doi: 10.1109/CCSSP49278.2020.9151569.

[25] L. Srinivasan and C. Nalini, "An improved framework for authorship identification in online messages," Cluster Comput., vol. 22, no. s5, pp. 12101–12110, 2019, doi: 10.1007/s10586-017-1563-3.

[26] M. Martinc and S. Pollak, "Combining n -grams and deep convolutional features for language variety classification," vol. 2013, pp. 607–632, 2019, doi: 10.1017/S1351324919000299.

[27] T. K. Mustafa, A. A. Abdul Razzaq, and E. A. Al-Zubaidi, "Authorship Arabic Text Detection According to Style of Writing by Using (SABA) Method," Asian J. Appl. Sci., vol. 5, no. 2, pp. 483–490, 2017, doi: 10.24203/ajas.v5i2.4750.

[28] M. H. Altakrori, F. Iqbal, B. C. M. Fung, S. H. H. Ding, and A. Tubaishat, "Arabic authorship attribution: An extensive study on twitter posts," ACM Trans. Asian Low-Resource Lang. Inf. Process., vol. 18, no. 1, 2018, doi: 10.1145/3236391.

[29] Alhujaili, R. F., & Yafooz, W. M. (2022, May). Sentiment Analysis for YouTube Educational Videos Using Machine and Deep Learning Approaches. In 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI) (pp. 238-244). IEEE.

[30] N. M. A. Rahman, S. A. Nouh, and R. H. A. Alez, "A Language Model for Arabic Texts Disambiguation using Deep Learning," vol. 6, no. 2, pp. 1–16, 2019.

[31] N. M. Abdel and R. Ibrahim, "A New Model for Arabic Text Clustering by Word Embedding and Arabic Word Net," Saudi J. Eng. Technol., vol. 6272, pp. 401–406, 2019, doi: 10.36348/SJEAT.2019.v04i10.001.

[32] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "ASA: A framework for Arabic sentiment analysis," Journal of Information Science, p. 016555151984951, May 2019, doi: 10.1177/0165551519849516.

[33] O. Oueslati, E. Cambria, M. B. HajHmida, and H. Ounelli, "A review of sentiment analysis research in Arabic language," Future Generation Computer Systems, vol. 112, pp. 408–430, Nov. 2020, doi: 10.1016/j.future.2020.05.034.

[34] M. A. Hedderich, L. Lange, H. Adel, J. Strötgen, and D. Klakow, "A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios," arXiv:2010.12309 [cs], Apr. 2021, Accessed: Dec. 01, 2022. [Online]. Available: http://arxiv.org/abs/2010.12309

[35] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Advances in neural information processing systems, 2013, vol. 26.

[36] N. M. Ibrahim, "Text Mining using Deep Learning Article Review," vol. 9, no. 9, pp. 1916–1933, 2018.

Synthesis of Comments to Social Media Posts for Business Applications

Peter Adebowale Olujimi[1], Abejide Ade-Ibijola[2]
Department of Applied Information Systems-College of Business and Economics,
University of Johannesburg, Johannesburg, South Africa[1]
Research Group on Data-Artificial Intelligence-and-Innovations
for Digital Transformation-Johannesburg Business School,
University of Johannesburg, Johannesburg, South Africa[2]

***Abstract*—Responding to enormous comments on social media platforms is one major challenge facing businesses in recent times, especially when dealing with irate consumers. Customers have increasingly adopted social networks as a platform for expressing their concerns and posting comments on business pages, posing a great challenge for customer support agents and digital marketers alike. Analyzing and responding manually to these enormous comments is a time-consuming task, necessitating the adoption of Artificial Intelligence (AI) tool that can complete the task swiftly — *automatic comprehension of social media posts for comment generation*. In this paper, we present algorithms and a tool for the automatic comprehension of customer tweets and generation of responses to these tweets. This was done in two-fold: using existing Natural Language Processing (NLP) libraries to preprocess and tokenize these tweets, and secondly, using rule-based algorithms to find *a matching response* to each customer, based on the array of extracted tokens from the customer's tweet. This was built into a tool called `Comment-Synthesizer`. This tool takes unfiltered tweets as input, preprocesses the tweets, and matches the tweet with predefined responses using a rule-based algorithm with a success rate of 76%. This tool, if implemented in a desktop automation application, can be used to respond automatically to a large volume of customers' social media comments/posts.**

Keywords*—Natural language comprehension; social media; natural language processing; customer engagements; artificial intelligence; comment generation

I. Introduction

In recent years, social media has become an indispensable resource for companies to incorporate into their marketing plans [1]. Social media is described as a collection of Web applications that allow people to produce and generate content through the innovative use of Internet technology [2]. The widespread use of social networks by customers influences businesses that must join to connect with customers [3]. Businesses are embracing social media for an expanding number of marketing tasks, including image, market analysis, demand planning, provision of services, and commerce. Additionally, businesses have found that social networking websites, in particular, are crucial for drawing customers to their products and brands [4]. As a result of the recent explosion of interest in social networks like Facebook, Twitter, and LinkedIn, tons of businesses are posting information to influence consumers about their products/services [5]. The number of posts and comments received every day gives a surprising indication of the enormous influence of social networks, with Facebook accounting for one billion active users, YouTube with over four billion views, and nearly 140 million active users on Twitter [6], [7], [8]. As a means of communication, the usage of social networks is becoming significantly important, particularly because it enables firms to cultivate relationships with their clients [9]. Social networks can be used to distribute well-known e-Commerce solutions, but businesses that do not engage with their customers lose their potential [10].

In the early days of social media, PR firms would keep an eye on customer comments on a company's website to find and handle irate clients, but with the proliferation and volume use of social media platforms, this is no longer nearly adequate [6]. When a business adopts a Social Media strategy, the volume of postings increases quickly, requiring considerable human work for manual analysis [11]. Consequently, many businesses designate a human expert to constantly update their pages with relevant content, monitor comments, and respond to them on time. However, responding manually to these enormous comments is a time-consuming task, necessitating the adoption of Artificial Intelligence (AI) tool that can complete the task swiftly — *automatic comprehension of social media posts for comment generation*.

A long-standing aim of AI is the creation of programs that can understand and generate human language [12]. Natural Language Processing (NLP), an aspect of AI that deals with the interaction between machines and human language and has been around for more than four decades [13], is one of such programs. Machines are programmed to successfully process large natural language data. NLP makes it feasible for Intelligent machines to analyse the complexities of human expression in order to obtain information from numerous types of text, such as blogging, evaluation of products, as well as the numerous daily posts on Twitter, Facebook, and news feeds [14]. The digital representation and understanding of human language have recently attracted significant interest in this domain. Its applications have spread to numerous domains such as email spam surveillance, translation software, and retrieval of information, summarization, medicine, and question answering [15].

In this paper a set of rules was used to develop an algorithm that automatically comprehends social media comments and responds to them in real-time. This involves performing lexical analysis on the raw tweets with several preprocessing techniques, such as eliminating Twitter handles, URLs, whitespaces, etc. In addition, we performed sentiment analysis to identify the underlying emotions in the dataset, and then we defined rule-based algorithms to find *a matching response* to each customer, based on the array of extracted tokens from the customer's tweet. Fig. 1 depicts the processes.

The following is a summary of the significant contributions made in this paper.

1) leverage the Python Natural Language Toolkit (NLTK) for the analysis of the dataset. A tool that supports NLP data structures and algorithms in Python,
2) design a Natural Language Generation (NLG) algorithm system that uses a rule-based approach to automatically generate social media comments,
3) create `Comment-Synthesizer`, a software tool that deploys the techniques described in (3) above, and
4) present deployment of the tool, as well as results demonstrating the social media comments generated.

The remaining sections of this paper are organized as follows. Section II provides background and discusses the related work. Section III elaborates on the design concept used in this work. In Section IV, we demonstrate the implementation of the `Comment-Synthesizer` tool and the results obtained from the synthesized comments. In Section V, the results are evaluated, while Section VI provides a summary of the work and the further study.

II. Background and Related Work

This section examines works related to the problem domain and presents the various terms and techniques for solving the problem. It provides a definition of the terms and the motivation behind this work.

A. Social Media Marketing in the Digital Age

Social media is a dynamic and fast-moving domain across the globe. In the digital age, leveraging social media to advertise to a targeted demographic of potential customers is an effective strategy [16]. This allows a business to remain connected with customers and respond to their queries without delay. Every day, people around the world use the Web, social networks, smartphone applications, and other electronic communication systems [17]. Since the emergence of social networks, many businesses have realized that they can add value to a firm's entire business plan, particularly in terms of improving brand perception.

Businesses are leveraging alternative marketing strategies that are more cost-effective and beneficial for consumer engagement than conventional ways, as they recognize the potential of the Internet as a crucial component of their platform. They are now supporting open communication practices within the firm, such as tweeting, texting, and other methods, to promote branding outside the company [18]. Its no secret that platforms like Facebook, Instagram, and Twitter are essential tools for interacting with customers and spreading the word about a firm's products and services. Given how frequently consumers use social media for several hours each day, marketers have adopted it as a powerful marketing tool [19]. According to the statistics, the average internet user spends approximately 144 minutes every day exploring social networks. As a result, large corporations and their strategists employ social media to create brand images [20].

B. Social Media Automation

Digital marketing has metamorphosed significantly over the past two decades. Businesses are now expected to improve their strategies to earn customers' loyalty. The social web statistics indicate that automation of digital marketing applications, especially social media, is the key as this enables firms to make effective use of their time by concentrating on the more important side of the business. Social media automation can be described as an approach to remodelling business social media campaigns by utilising automation tools, such as Hootsuite, Buffer, Social Pilot, and the like [21]). Automating social media posts reduces the time spent on maintaining and growing brand accounts. Therefore, time and resources could be allocated to other areas of the marketing budget and to meet strategic goals [22]. A significant advantage of automation is the propensity to organize the company's entire social media strategy in advance. Content or posts can be scheduled on Twitter, Facebook, Instagram, and other social networks, which are then automatically published [23].

C. Challenges of Digital Marketing

One of the most direct channels of engagement between a business and a target audience is through social media marketing — bidirectional communication. As consumers spend more time than ever online, the ability to effectively engage with them, respond to their comments, and solve their problems in real time is crucial [24]. Marketing experts have less control over the message on social networks due to its conversational nature than they do with more conventional means of marketing communication [3]. There is more exposure to digitization and social media from several consumers. Social media power is a fundamental challenge for any digital marketer. The management of customer queries is one of those challenges.

A key goal of marketers is customer engagement, a study in 2012 found that customer engagement was reinforced by 78 per cent of marketing reports when they use social media to promote their products [25]. Customer expectations must be managed to avoid harmful or negative posts. The use of negative post-reactions in social media marketing is one facet of digital marketing that can be particularly destructive to marketing campaigns. There is not much a marketer can do to prevent unfavorable comments, images, or videos from being posted online by dissatisfied consumers or competitors in the same business [26].

D. Artificial Intelligence and Automation

Artificial Intelligence is the ability to make machines execute intelligent tasks like humans. It uses intelligence to perform automated tasks. A more in-depth definition of AI describes it as the flexibility in a machine's ability to learn from experience, incorporate that knowledge into its operation, and use that performance to achieve predetermined ends [27]. As automation becomes more complex, the demand for AI is growing due to its ability to solve complex problems even with little human expertise and resources in a short timeframe [28]. With abstraction to phase out human efforts in routine tasks,

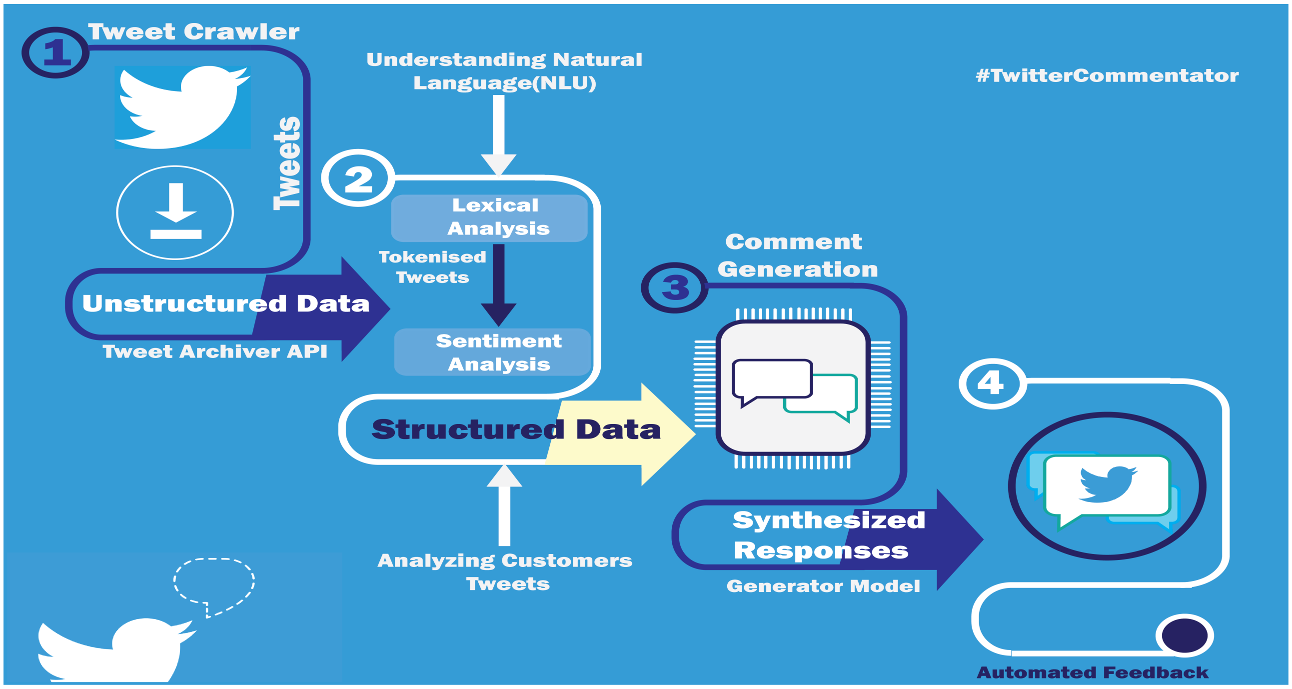


Fig. 1. Processes of social media comment generation

advancing machine learning (ML) will continuously activate paradigm shifts in multiple sectors of the technology industry. Among the components of AI is NLP, which is concerned with the interaction of machines and humans through natural language. As AI is penetrating the business arena, it has led to a fundamental change in the concept and application [29].

The application of AI is used significantly in different business domains, including healthcare, finance, manufacturing, law, education, and more. Doctors can now diagnose diseases faster than before with the assistance of ML. The application of AI in social media marketing results in spending less time on routine administrative tasks and providing customers with satisfaction. The adoption of AI can be cost effective and complement customer engagement [12]. Applications such as chatbot assist patients with invoicing and assist clients with appointment scheduling. In pedagogy, AI can provide auto-grading, assist students with learning by catering to their individual needs and ensuring that they stay on course. Artificial intelligence in the legal field has made it easier for lawyers to precisely and efficiently evaluate thousands of large legal documents, which is generally a difficult undertaking. Industrial robots have made manufacturing considerably easier and more efficient than it was only a few years ago [28].

E. Motivation

There is an increase in the number of comments available on a business social media platform every single day, especially on Twitter. It is projected that there will be 3.43 billion monthly active users on social media platforms worldwide in 2023. This figure represents almost one-third of the total population of the world [20]. Thus, it becomes a problem for a human expert to deal with a large volume of questions in different real-life data handling applications. Social media comments are important data in social media marketing, as those comments can make or break a business. However, all posts must be evaluated to effectively utilize the data acquired in the social media posts [11]. Against this background, there is the necessity for the creation of an algorithm that can complete the task swiftly.

F. The Gap

Unprecedented social media marketing growth has led to a constant influx of customers posting a massive volume of comments on business social media platforms that may be too much for a human expert to handle. When customers do not receive a prompt response to their inquiries, there is an underlying unhappiness that exists. Therefore, it is necessary to create a novel strategy to solve this issue without involving humans in easy-to-automate tasks.

G. Why Use Rule-Based Techniques?

The rule-based system has advanced several NLP systems. Rule-based logic is at the centre of most automated processes. A rule-based system encrypts expert human knowledge in a rather narrow domain using an automated system [30]. The foundation of rule-based systems is solid language comprehension and the syntactic insights gained with an NLP system can be applied to a similar task in another system [31]. One of the simple methods to represent the human intellect in AI is to transform it into plain linguistic phrases utilizing the format IF-THEN rules [32]. The knowledge expert system is an example of a rule-based technique that uses predefined rules to make deductions and draw conclusions to perform automated tasks.

A rule-based approach consists primarily of a collection of rules, a statement of facts, and a termination criterion [30]. Despite the fact that a rule-based method is neither AI nor ML, it may be used to power certain components of AI and

ML, as it essentially follows the rules specified by humans. Although an AI can determine the action to take itself, the rule-based approach does exactly as humans instructed. ML, on the other hand, is concerned with the possibility of computer programs accessing data and using it to learn for themselves. The objective is to provide machines with the ability to learn independently of human intervention and modify their behavior as necessary. Therefore, it is essential to use an approach that is most suitable for a specific field of study as this will influence the specified result. ML components in this case lack enough data to form a proper generalization. The benefit of the rule-based system is that it offers a detailed analysis of social media comments using morphological and syntactic information, while ML approaches face some challenges with specific NLP tasks, such as detecting well-formed and poorly formed inputs [31].

H. NLP Tools and Techniques

NLP deals with computer-natural language communication. Algorithmic systems can integrate language understanding and language generation through the use of NLP metrics [15]. NLP techniques enable enormous collections of social media posts to be automatically analysed and useful information extracted. Table I illustrates some examples of NLP tools and techniques.

TABLE I. SOME EXAMPLES OF NLP TOOLS AND TECHNIQUES

Tools	Techniques
AllenNLP: Premised on PyTorch tools and frameworks to use deep learning techniques for applications in NLP research and industry [33].	**Text Classification:** NLP techniques to classify vast volumes of unstructured text [34].
Keras: A library for neural networks that is highly modular, which is built on top of TensorFlow [35].	**Summarisation:** Using an NLP system to synthesize complex text by generating a concise summary [36].
NLP Architect: A Python module for evaluating deep learning architectures and techniques for optimizing NLP and NLU neural networks [37].	**Sentiment Analysis:** The analysis of data to assess its positivity, neutrality, or negativity [38].
SpaCy: Open-source NLP library that specializes in syntactic analysis, word-to-vector transformations, and conversational UIs [35].	**Keyword Extraction:** Automatically extracting the most useful textual data employing AI and machine learning algorithms [39].
NLTK: The Natural Language Toolkit is an accessible NLP resource for tokenising, stemming, labeling parts of speech, parsing, and logic [40].	**Named Entity Recognition:** Tags and extracts named identities from the using NLP algorithms for further analysis [41].

I. Related Work

According to our understanding and considerable research, no effort has been attempted to automatically comprehend and generate comments for social media posts using rule-based technique. Nevertheless, there are numerous similar works in NLP, particularly in the area of tweet message analysis on social media sites like Twitter, and various works in automatic comprehension and summarisation, language generation, and synthesis of things. Several of these interrelated concepts will be covered in this section.

1) *Twitter Immediate Response Using NLP Techniques:* Ahmad et al [42] proposed implementing NLP algorithms to analyze and evaluate various preprocessing, data augmentation, classification, and fusion approaches to provide instant response on water quality from social media posts.

2) *Social Media Posts as Programming Feedback:* Kabaso and Ade-Ibijola [43] presented a Context-Free-Grammar for automatically generating educational feedback on Twitter for novices in programming.

3) *Comment Generation Based on User Profile:* Zeng et al [44] proposed that each user's personalized comments be generated automatically based on the data in their profile. The work employed a large-scale and high-quality Chinese dataset as the foundation for creating a link between user profiles and the comments made on social media platforms.

4) *Social Media Profile Synthesis:* Ade-Ibijola [45] presented novel approaches to create imaginary social media profiles on an automatic basis utilizing probabilistic CFG.

5) *Financial Chat Analyzer (FINCHAN):* IMs with FX options obtained from Bloomberg Terminals were analyzed by [13] using a formal grammar-based technique. The work presented a novel algorithm to produce descriptive overviews of these financial chats by cleaning and retrieving terms from the conversations. The conversation token syntax was then specified using a CFG.

6) *Analysis of Tweets for Identifying Trends:* Novel algorithms were developed by [46] to identify trends in tweets related to natural disasters, such as earthquakes.

J. Definition of Terms

Definition 1: Symbol, Alphabet, and String [47]: A single element is regarded as a symbol. Any finite collection of symbols is represented by an alphabet Σ, while a string consists entirely of zero or more symbol combinations.

Definition 2: Tokenization: The process of dividing raw text into small parts is called tokenization. Raw text is divided into words and sentences and is referred to as tokens. These tokens are helpful in the creation of NLP contexts and models. Tokenizing complete sentences or individual words is possible using this method. The term "word tokenization" refers to the act of splitting a string of words into constituent lexemes. Sentence tokenization is the term used for the process of tokenizing a sentence [48].

Definition 3: Regular Languages and regular expressions [47], [13]: Languages that can be accepted by finite automata are regular languages that can be expressed by formulas known as regular expressions that use the operations of union, concatenation, and the Kleene star. An algebraic notation to describe regular language is called regular expressions (RE). The following terms apply to a regular language L over some alphabet Σ:

1. Ø denotes an empty set in regular language.

2. λ is a collection of all conceivable null strings in regular language.

3. For each $a \epsilon \Sigma$, a denotes the regular language $\{a\}$.

4. Given L_1 and L_2 are R, then the following is true: L_1 and L_2 is R that combines both languages. The combined language, identified as L_1, L_2, is considered a regular language. L_1^* is another common language that symbolizes the Kleene closure feature and is also referred to as the Kleene star.

III. Design

The rules for generating comments for social media posts are described in this section. Our test case focuses on Twitter. Twitter is a social networking and microblogging platform that was launched in 2006 by a small group of people, including Jack Dorsey [49]. Every second, over 6,000 tweets are sent, equating to about 350,000 messages sent in a single minute, 500 million tweets produced in a single day, and 200 billion tweets sent in a single year [50]. To increase user and consumer participation, marketers employ Twitter and other social networking platforms in their campaigns [51]. A tweet has a structure that allows content to be shared among its audience with a 280-character limit [52]. Twitter can facilitate brand engagement with consumers by monitoring and responding to online consumer opinions [53]. In this paper, the structure of a tweet is described using the following attributes:

1) **Tweet Text:** it is a short message containing 280 characters or less that a Twitter user shares with a specific audience. For example, text, pictures, graphics and videos,
2) **Screen Name:** a Twitter handle or username that is assigned when a user creates their first Twitter account. This is followed by the "@" symbol, which your followers use when responding, mentioning, and sending direct messages (DM). For instance, @steveleo,
3) **Full Name:** it is a unique identifier that is different from a Twitter username. For instance, Jay Paul,
4) **Date:** the time and date when a tweet was published on Twitter,
5) **Tweet ID:** Twitter's unique identifier for each user's tweet,
6) **Location:** the location of the Twitter user,
7) **Retweet:** sharing a tweet from another account with your followers,
8) **Type of App:** the device that a user uses to send Tweets,
9) **Hashtag:** a topic or narrative preceded by the sign (#) symbol. Users can use hashtags to explore topics of interest, and
10) **Url:** the address to a user's profile on Twitter.

A. Tweet Extraction

To begin with, we first crawl tweets of a keyword associated with the hashtag "loadshedding". In South Africa, the country's electricity company, Eskom, frequently uses the term "load shedding" to prevent blackouts. As the primary source of electricity in South Africa, Eskom is tasked with ensuring the reliability of the country's power supply to best support the nation's economy and society. Their main responsibilities in the country of South Africa include the generation, transmission, and distribution of electricity. Loadshedding is a well-planned response to unplanned events that is carried out with the objective of preventing the complete collapse of the energy grid [54]. The primary goal of load shedding is to ensure that the grid does not fail completely. Consumers are frequently notified of imminent load shedding, which typically takes place in a sequence of stages: 1, 2, 3, 4 etc. The tweets were extracted using Tweet Archiver API that spans a period of 12 weeks, which are then stored in Google sheets for further preprocessing.

B. Tweet Preprocessing

It is crucial to conduct a lexical analysis to parse the tweet text. Tweets are unstructured data that frequently contains graphics and abbreviations and must be converted into a format that machines can understand and analyse. Tokenization is the initial stage in any NLP pipeline. Algorithm 1 shows the tweet handling method using regex to search for the instance of tweet data in the text. A tokenizer as presented in Algorithm 2 was designed to separate the dataset into distinct pieces of information. The dataset was first preprocessed using the NLTK tool, which involves the following cleaning process: removing Twitter handles, whitespace, emojis, URLs from the tweet, hashtag and punctuation removal, converting text to lowercase, stop word removal, tokenizing the string with regular expressions into a list of words, and reducing words to their stems, i.e., buying, buys become buy. During the tokenization process, some stop words, such as "the" and "a", will be eliminated because these words do not contribute significantly to the information obtained. An illustration of a tweet preprocessing technique can be seen in Fig. 2.

Algorithm 1 : Tweet Handling

```
Function text_handle_remove(input_text):
    using regex to search for the instance of tweet data in
  the text
    regex.sub(r'https?://[^\s\n\r]+', ”, tweet)
    regex.sub(r'@[^\s\n\r]+', ”, tweet)
    re.sub(r'#', ”, tweet)
    return new_tweet_text
    end
End Function
```

C. Sentiment Analysis

The comprehension of massive amounts of unstructured data (tweets) can be facilitated for businesses using sentiment analysis. We are primarily interested in gaining a sense of the consumer perspective with respect to the dataset and the polarity of that perspective, that is, whether it is positive, negative, or neutral. The dataset was augmented with labels to facilitate the computation of opinion scores and their classification. The results shows a slightly high positive score (35.6%) compared to the negative score (35.2%), as shown in Fig. 3. It is observable that loadshedding has a negative impact on the consumers, and they viewed it as incompetence and maladministration. In addition, we utilised a word cloud as a

	Tweet Text	Tweet_No_url_cleaned	Tokenized_tweet
1	Please Fix #Eskom 😭 RT @elonmusk: Hate to say ...	Please Fix Eskom RT Hate to say it but we ne...	[please, fix, eskom, rt, hate, say, need, incr...
2	South Africa's state-owned power utility #Esko...	South Africas stateowned power utility Eskom h...	[south, africa, stateowned, power, utility, es...
3	If indeed we are serious about the township ec...	If indeed we are serious about the township ec...	[indeed, serious, township, economy, please, e...
4	To avoid impacting South Africa's electricity ...	To avoid impacting South Africas electricity s...	[avoid, impact, south, africa, electricity, su...
5	VIDEO | The Koeberg Power Station's Unit 2 has...	VIDEO The Koeberg Power Stations Unit 2 has b...	[video, koeberg, power, station, unit, 2, offl...
...	...	...	...
4037	The National Energy Regulator of South Africa ...	The National Energy Regulator of South Africa ...	[national, energy, regulator, south, africa, n...
4038	#Eskom South Africans quality for 100units of ...	Eskom South Africans quality for 100units of f...	[eskom, south, african, quality, 100units, fre...
4039	There you have it #Eskom RT @Eskom_SA: #EskomG...	There you have it Eskom RT EskomGauteng Media...	[eskom, rt, eskomgauteng, mediastatement, elec...
4040	[WATCH] A heavy police presence in #Diepkloof ...	WATCH A heavy police presence in Diepkloof for...	[watch, heavy, police, presence, diepkloof, fo...
4041	Don't you love how #Eskom is being sabotaged a...	Dont you love how Eskom is being sabotaged and...	[dont, love, eskom, sabotage, publicly, loot, ...

Fig. 2. Raw tweets before and after tokenization

Algorithm 2 : Tweet Tokenization

```
Function tokenize_text(input_text):
    tokenized_tweet=tweet_tokenizer.tokenize(tweet)
    stemmer = PorterStemmer()
    intialise english stopwords
    clean_tweet list
    for word in tokenized_tweet:
        if (word not in stopwords_english and
                word not in string.punctuation):
        wordnet_lemmatizer.lemmatize(word,
pos='verb')
        wordnet_lemmatizer.lemmatize(lemmatizer_word,
pos='noun')
        add lemmatized_word to clean_tweet list
    return tweets_clean
    end
End Function
```

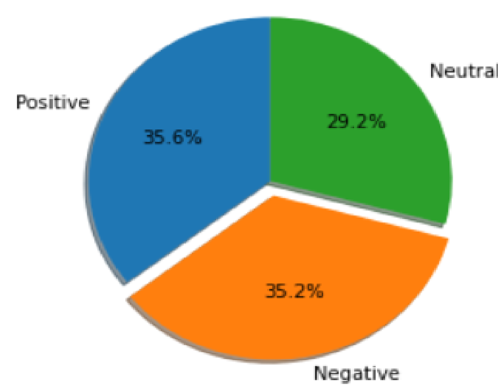


Fig. 3. Distribution of sentiments in the tweet

means of disseminating crucial information and developing an appealing visualisation strategy for the purpose of highlighting important textual data points, which helped monotonous data stand out as seen in Fig. 4.

IV. Implementation and Results

This section presents a new tool for implementing the techniques described in this article and the results obtained. The name of this tool is called `Comment-Synthesizer`.

A. *Implementation details*

We discuss the components of our technique for synthesizing comments and generating automated responses. First, we generated a list of intents and responses that spans the entire category of customer posts in the dataset — preprocessed Tweets. Based on the analysis of the patterns observed in the dataset, the intent of the users was classified into 19 groups, each class representing a distinct user response. We tagged each post response based on the human expert's estimation using the process below:

1) Create a list of intents and responses.
2) Use WordNet to find similar words with a tag of intent.
3) Combine words that are similar in meaning.
4) Use Regular Expressions (RegEx) to find similar patterns.

The tag in each intent is determined by the categorization of the dataset. For example, a question concerning power outages was tagged "outages" while inquiries from customers regarding

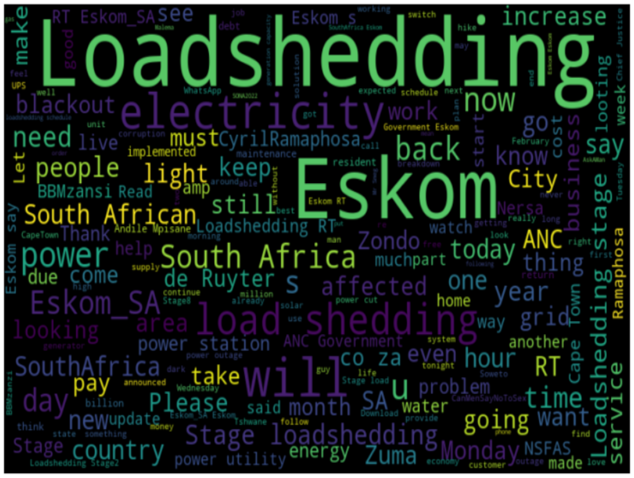


Fig. 4. Visualization of the frequent words in the tweet

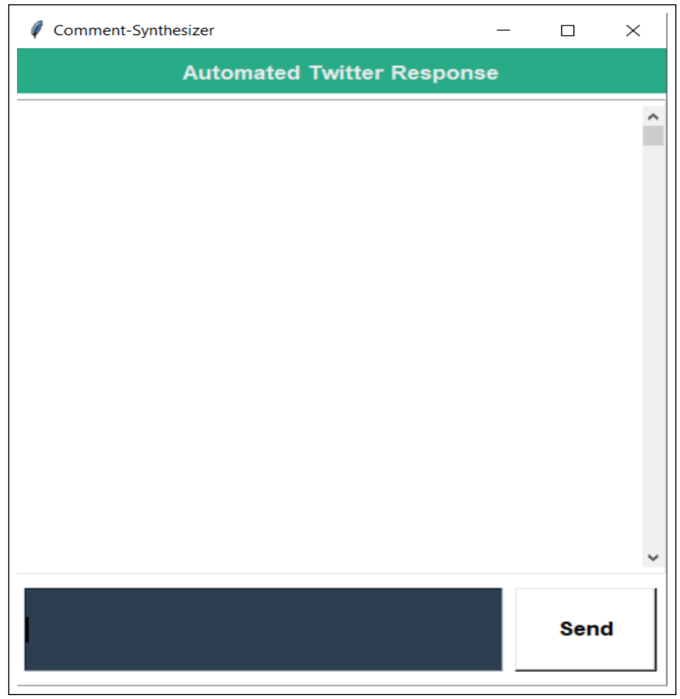


Fig. 5. The `Comment-Synthesizer` user interface

unavailable lights were tagged "lights out", etc. In addition to the tags that were entered manually during the process, the WordNet library was utilized to locate words that were related to those tags. A RegEx was used for word combinations to determine the right number of data points within the dataset. This involves querying similar tag information and retrieving the assigned tag when relevant data are discovered during the process. The full dataset and algorithm implementation were subsequently embedded into a desktop interface. Table II illustrates the function of RegEx to find matching patterns within the dataset while some implementation aspects of the tool are indicated by the steps of Algorithm 3.

TABLE II. REGEX FOR TWEET PATTERN MATCHING

Token	Symbol
Newline	\n
Alphabet	[a-zA-Z]
Number	[0-9]
Carriage return	\r
Tab	\t
Number and alphabet	[ˆa-zA-Z0-9 \n\.]
BackSpace character	\b

B. Desktop Application

The tool was implemented and integrated into a desktop application using the Python Tkinter library. This application shows a simple interface that allows users to type in their tweet and receive an automated response based on that tweet. We integrated the algorithm into a Python script using Tkinter to develop the interface and then made it feasible to employ interface formatting to create a user experience that will represent a functional chatbot application as displayed in Fig. 5.

C. Results

We have tested `Comment-Synthesizer` on tweets that were extracted from Twitter using the Tweet Archiver API between December 2021 and March 2022. The dataset was extracted using keywords associated with #Loadshedding, #EskomLoadshedding, and #Eskom_SA. In total, the dataset comprises 4,041 Tweets, which are then stored in a Google Sheet for further preprocessing. After preprocessing approaches, such as removing Twitter handles, emojis, URLs, punctuation, and whitespace, we divided the dataset into training and test datasets. There are 2,733 tweets in the training set and 1,308 tweets in the test set. The tool worked successfully and produced Twitter responses that were, for the most part, accurate. Fig. 6 and Fig. 7 shows the program interface during testing execution.

V. EVALUATION OF THE TOOL

The performance of a model can be explained using evaluation metrics. The ability of assessment metrics to differentiate between different model outputs is an essential characteristic of these metrics. The development of an algorithm and a tool that achieves a high level of accuracy when applied to data that are not part of the sample dataset is the primary objective of this research. As a result, it is essential to perform an accuracy check on this model before attempting to compute anticipated values. The evaluation of `Comment-Synthesizer` is based on the following metrics: recall, precision, and accuracy while predicting syntactically correct or incorrect responses in the dataset. In the evaluation, the confusion matrix classification was used, and the tool performed well and generated accurate tweet responses for the most part — about 76% accuracy. The classification of the tool performance is as shown below:

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{449}{449+163} = 0.73$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{449}{449+153} = 0.75$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{449+543}{449+543+153+163} = 0.76$$

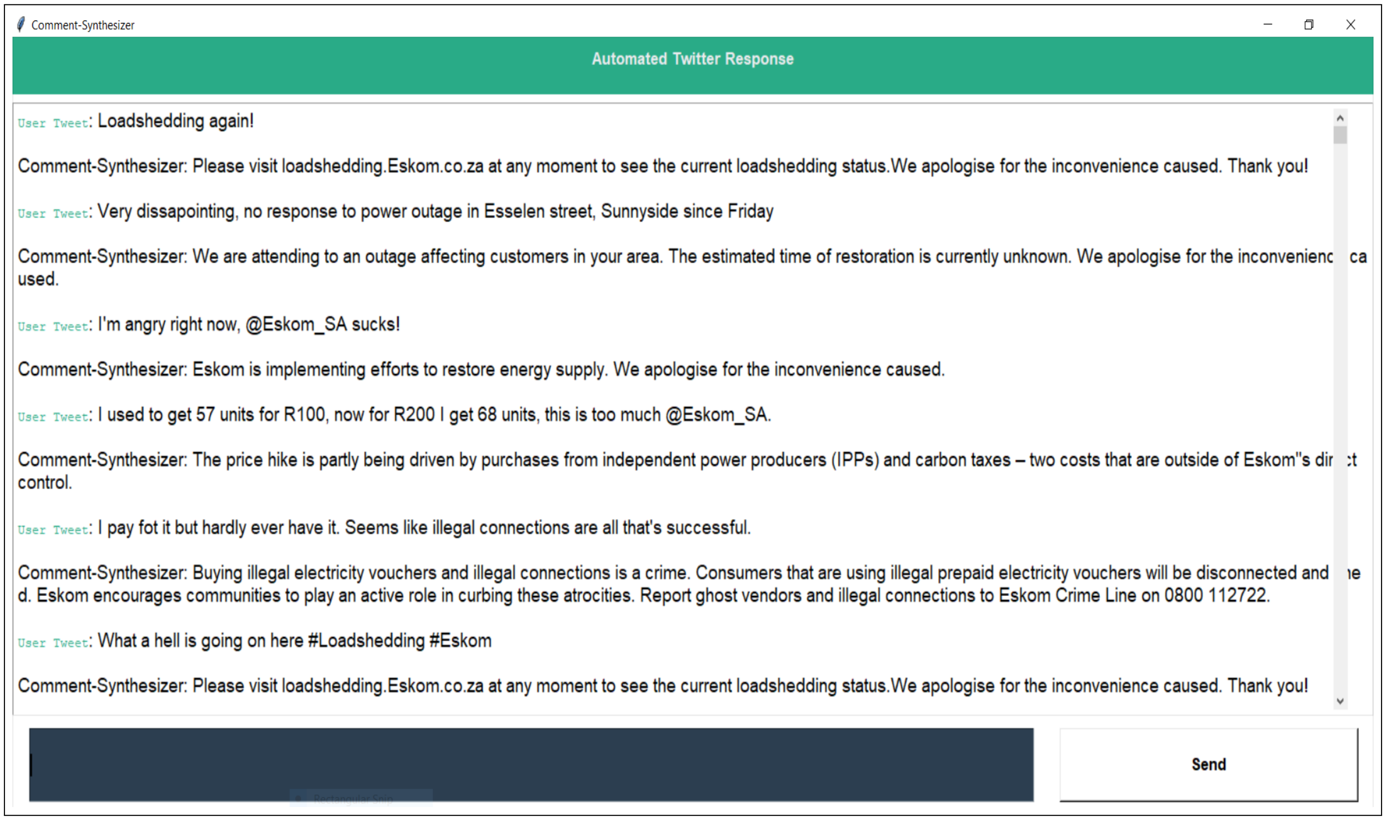


Fig. 6. The `Comment-Synthesizer` tool displays the users' tweet and the responses generated automatically (1/2)

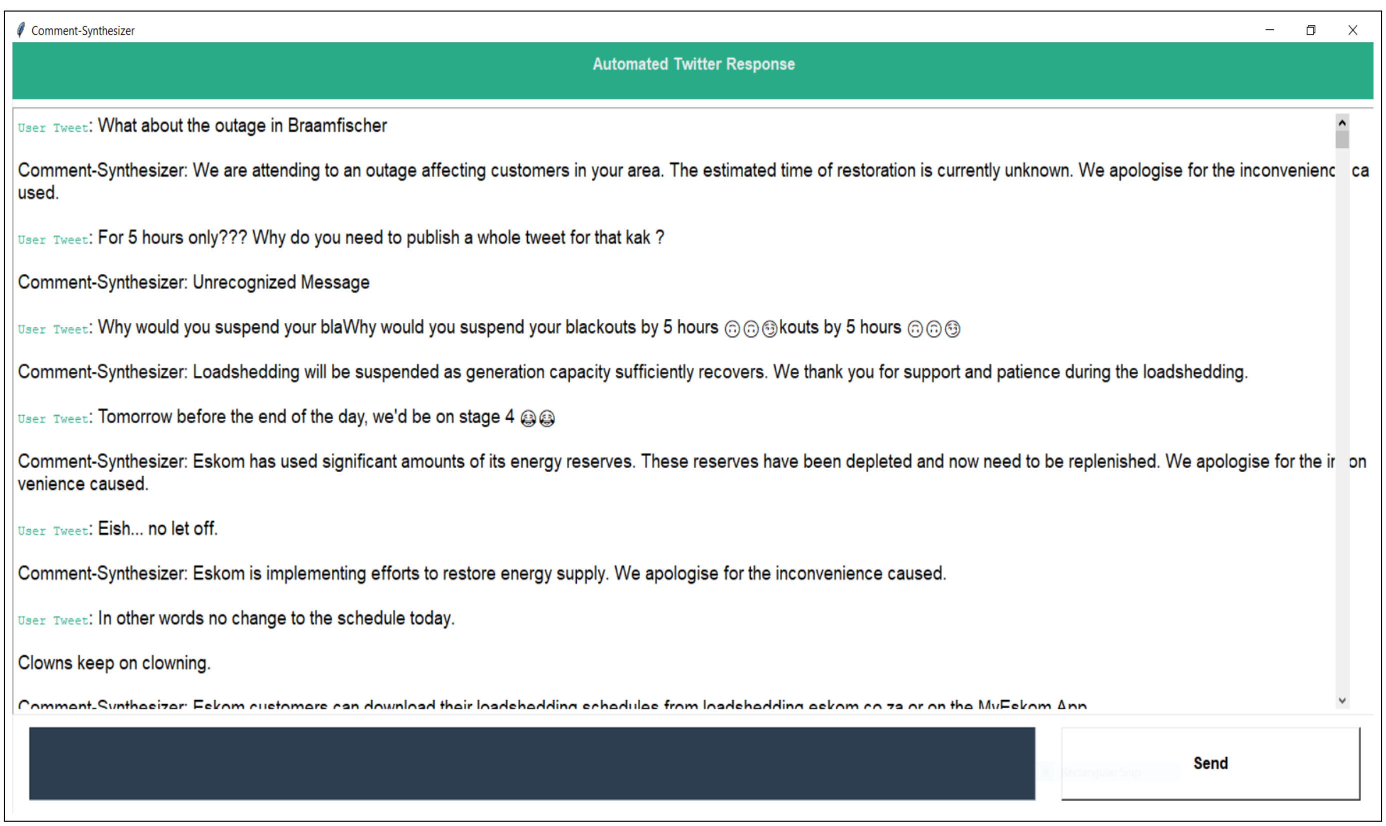


Fig. 7. The `Comment-Synthesizer` tool displays the users' tweet and the responses generated automatically (2/2)

Algorithm 3 : Comment-Synthesizer Algorithm

```
Function get_tweet_synonyms(tag)
    define the list of intents, response and tag
    preprocess tweet
    for each_tag in tag
        get synonyms of tag from Wordnet.synsets
            for each_word in synonyms
                lemmatize each_word
        add user customized synonyms
    return processed_synonym
    end
End Function
Function create_pool_of_intent(tag)
    for each_synonmys in tags_of_intent:
        add searchable regex function(".*\\b")
        Add regex compile action()
    return new_synonyms
    end
End Function
Function algorithm_implementation(user_tweet)
    for each_word in user_tweet
        convert each_word to lower_case
        remove tweet_handle, hash_tag in each_word
    produce user_tweet
    for intent in dictionary_of_intent:
        regex search for user_tweet_pattern
    if user_tweet_pattern is found:
        produce the appropriate tag and response
    else
        produce unrelated message
    return appropriate response to the tag found
    end
End Function
```

where TRUE POSITIVES are represented by the letters TP, TRUE NEGATIVES by the letters TN, FALSE POSITIVES by the letters FP, and FALSE NEGATIVES by the letters FN.

VI. Conclusion and Future Work

The design of a software program called `Comment-Synthesizer` is described, which is designed to automatically comprehend and generate comments on social media posts. This entails performing lexical analysis on raw tweets using various preprocessing techniques, such as removing Twitter handles, URLs, whitespaces, etc. In addition, we performed sentiment analysis to identify the underlying emotions in the dataset, and then we defined rule-based algorithms to find *a matching response* to each customer, based on the array of extracted tokens from the customer's tweet. The Python library Tkinter was used to implement and build the tool into a desktop application. We demonstrate the deployment of the tool with the results of the social media comments it generated. During the testing, the `Comment-Synthesizer` algorithm predicted with 76 percent accuracy for the most part. This tool, if implemented in a desktop automation application, can be used to respond automatically to a large volume of customers' social media comments/posts, thereby improving customer experience. In the future, we will increase the classification of datasets to enhance the functionality of the `Comment-Synthesizer` tool.

Acknowledgment

This research was partially supported by the Global Excellence Stature (GES) awards and National Research Fund (NRF) with grant number - 119041.

References

[1] C. Castronovo and L. Huang, "Social media in an alternative marketing communication model," *Journal of marketing development and competitiveness*, vol. 6, no. 1, pp. 117–134, 2012.

[2] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[3] J. B. Thomas, C. O. Peters, E. G. Howell, and K. Robbins, "Social media and negative word of mouth: strategies for handing unexpecting comments," *Atlantic Marketing Journal*, vol. 1, no. 2, p. 7, 2012.

[4] M. Monika and S. Sharma, "Exploring the influence of social media promotions on brand engagement, trust, perceived value and perceived risk of online shopping behavior," 2019.

[5] K. Bontcheva, G. Gorrell, and B. Wessels, "Social media and information overload: Survey results," *arXiv preprint arXiv:1306.0813*, 2013.

[6] W. Fan and M. D. Gordon, "The power of social media analytics," *Communications of the ACM*, vol. 57, no. 6, pp. 74–81, 2014. [Online]. Available: doi.org/10.1145/2602574.

[7] L. Yu, S. Asur, and B. A. Huberman, "What trends in chinese social media," *arXiv preprint arXiv:1107.3522*, 2011.

[8] X. Jin, C. X. Lin, J. Luo, and J. Han, "Socialspamguard: A data mining-based spam detection system for social media networks," *Proceedings of the VLDB Endowment*, vol. 4, no. 12, pp. 1458–1461, 2011.

[9] L. Zailskaitė-Jakštė, A. Ostreika, A. Jakštas, E. Stanevičienė, and R. Damaševičius, "Brand communication in social media: The use of image colours in popular posts," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2017, pp. 1373–1378.

[10] M. J. Culnan, P. J. McHugh, and J. I. Zubillaga, "How large us companies can use twitter and other social media to gain business value." *MIS Quarterly Executive*, vol. 9, no. 4, 2010.

[11] J. M. Schwaiger, M. Lang, C. Ritter, and F. Johannsen, "Assessing the accuracy of sentiment analysis of social media posts at small and medium-sized enterprises in southern germany," 2016.

[12] R. Akerkar, *Artificial intelligence for business*. Springer, 2019. [Online]. Available: doi.org/10.1007/978-3-319-97436-1.

[13] A. Ade-Ibijola, "Finchan: A grammar-based tool for automatic comprehension of financial instant messages," in *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, 2016, pp. 1–10. [Online]. Available: https://doi.org/10.1145/2987491.2987518.

[14] J. Kietzmann, J. Paschen, and E. Treen, "Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey," *Journal of Advertising Research*, vol. 58, no. 3, pp. 263–267, 2018.

[15] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia Tools and Applications*, pp. 1–32, 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-13428-4.

[16] S. Kabaso and A. Ade-Ibijola, "Sell-bot: An intelligent tool for advertisement synthesis on social media," in *The Disruptive Fourth Industrial Revolution*. Springer, 2020, pp. 155–178.

[17] A. T. Stephen, "The role of digital and social media marketing in consumer behavior," *Current opinión in Psychology*, vol. 10, pp. 17–21, 2016. [Online]. Available: https://doi.org/10.1016/j.copsyc.2015.10.016.

[18] A. Adamson, "Companies should encourage social networking among employees," *Social Media, Forbes. com, available at http://www. forbes. com/2009/06/02/charles-schwab-spyfacebook-leadership-cmo-net*, 2009.

[19] G. Appel, L. Grewal, R. Hadi, and A. T. Stephen, "The future of social media in marketing," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 79–95, 2020. [Online]. Available: https://doi.org/10.1007/s11747-019-00695-1.

[20] P. H. Tankovska, "Social media - statistics & facts," 2021.

[21] H. Pham, "Top 12 Social Media Automation Tools You Should Know About," 2021.

[22] C. Samanta, "Social Media Automation – A Social Media Marketing Trend," 2020.

[23] A. Sonnenberg, "What Is Social Media Automation? What Isn't Social Media Automation? Why Do You Need Social Media Automation?" 2021.

[24] L. Hailey, "The importance of social media marketing today," *Retrieved June*, vol. 26, no. 2012, pp. 233–240, 2010.

[25] C. Nguyen, N. Nguyen, and A. Duong, "The relationships of social media marketing, consumer engagement and purchase intention," *Test Eng. Manag*, vol. 2, pp. 24 653–24 666, 2020.

[26] C. M. Cheung, M. K. Lee, and D. R. Thadani, "The impact of positive electronic word-of-mouth on consumer online purchasing decision," in *World Summit on Knowledge Society*. Springer, 2009, pp. 501–510.

[27] A. Kaplan and M. Haenlein, "Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, vol. 62, no. 1, pp. 15–25, 2019.

[28] S. S. Shekhar, "Artificial intelligence in automation," *Artificial Intelligence*, vol. 3085, no. 06, pp. 14–17, 2019.

[29] M. S. S. E. Namaki, "How Companies are Applying AI to the Business Strategy Formulation," *Scholedge International Journal of Business Policy & Governance ISSN 2394-3351*, vol. 5, no. 8, p. 77, 2018.

[30] C. Grosan and A. Abraham, *Intelligent systems*. Springer, 2011. [Online]. Available: https://doi.org/10.1007/978-3-642-21004-4.

[31] K. Shaalan, "Rule-based Approach in Arabic Natural Language Processing," *the International Journal on Information and Communication Technologies (IJICT)*, vol. 3, no. 3, p. 11–19, 2010.

[32] S. Vashishtha and S. Susan, "Fuzzy rule based unsupervised sentiment analysis from social media posts," *Expert Systems with Applications*, vol. 138, no. July, 2019. [Online]. Available: doi.org/10.1016/j.eswa.2019.112834.

[33] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. Liu, M. Peters, M. Schmitz, and L. Zettlemoyer, "Allennlp: A deep semantic natural language processing platform," *arXiv preprint arXiv:1803.07640*, 2018.

[34] M. Thangaraj and M. Sivakami, "Text classification techniques: A literature review," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 13, p. 117, 2018. [Online]. Available: https://doi.org/10.28945/4066.

[35] B. Srinivasa-Desikan, *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018.

[36] S. Dohare, H. Karnick, and V. Gupta, "Text summarization using abstract meaning representation," *arXiv preprint arXiv:1706.01678*, 2017.

[37] J. Mamou, O. Pereg, M. Wasserblat, A. Eirew, Y. Green, S. Guskin, P. Izsak, and D. Korat, "Term set expansion based nlp architect by intel ai lab," *arXiv preprint arXiv:1808.08953*, 2018.

[38] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Information Sciences*, vol. 311, pp. 18–38, 2015.

[39] L. Marujo, W. Ling, I. Trancoso, C. Dyer, A. W. Black, A. Gershman, D. M. de Matos, J. P. Neto, and J. G. Carbonell, "Automatic keyword extraction on twitter," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2015, pp. 637–643.

[40] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

[41] C. Li, J. Weng, Q. He, Y. Yao, A. Datta, A. Sun, and B.-S. Lee, "Twiner: named entity recognition in targeted twitter stream," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 721–730.

[42] K. Ahmad, M. A. Ayub, K. Ahmad, J. Khan, N. Ahmad, and A. Al-Fuqaha, "Merit-based fusion of nlp techniques for instant feedback on water quality from twitter text," *arXiv preprint arXiv:2202.04462*, 2022.

[43] S. Kabaso and A. Ade-Ibijola, "Synthesis of social media messages and tweets as feedback medium in introductory programming," in *Annual Conference of the Southern African Computer Lecturers' Association*. Springer, 2019, pp. 3–16.

[44] W. Zeng, A. Abuduweili, L. Li, and P. Yang, "Automatic generation of personalized comment based on user profile," *arXiv preprint arXiv:1907.10371*, 2019.

[45] A. Ade-Ibijola, "Synthesis of social media profiles using a probabilistic context-free grammar," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*. IEEE, 2017, pp. 104–109. [Online]. Available: doi.org/10.1109/robomech.2017.8261131.

[46] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919–931, 2013.

[47] J. C. Martin, *Introduction to Languages and the Theory of Computation*. McGraw-Hill Higher Education, 2011.

[48] D. Ramachandran and R. Parvathi, "Analysis of twitter specific preprocessing technique for tweets," *Procedia Computer Science*, vol. 165, pp. 245–251, 2019.

[49] K. T. Taka, M. Paora, and R. Stacey, "An analysis of te reo mäori tweets," pp. 1–17, 2015.

[50] LiveStats. (2022) Twitter usage statistics. [Online]. Available: https://www.internetlivestats.com/twitter-statistics.

[51] L. D. Hollebeek, M. S. Glynn, and R. J. Brodie, "Consumer brand engagement in social media: Conceptualization, scale development and validation," *Journal of Interactive Marketing*, vol. 28, pp. 149–165, 2014.

[52] Twitter. (2022) What to tweet. [Online]. Available: https://business.twitter.com/en/basics/what-to-tweet.html.

[53] A. I. Canhoto and M. Clark, "Customer service 140 characters at a time: The users' perspective," *Journal of Marketing Management*, vol. 29, pp. 522–544, 2013.

[54] Eskom. (2022) What is load shedding? [Online]. Available: https://loadshedding.eskom.co.za/LoadShedding/Description.

PDE: A Real-Time Object Detection and Enhancing Model under Low Visibility Conditions

Zhiying Li[1], Shuyuan Lin[2]*, Zhongming Liang[3],
Yongjia Lei[4], Zefan Wang[5], Hao Chen[6]
Jinan University, Guangzhou, China[1,2,3,4,5]
The Hong Kong Polytechnic University, Hong Kong, China[6]

***Abstract*—Deep object detection models are important tools that can accurately detect objects and frame them for the user in real time. However, in low visibility conditions, such as fog or low light conditions, the captured images are underexposed and blurred, which negatively affects the recognition accuracy and is not well visible to humans. In addition, the image enhancement model is complex and time-consuming. Using the image enhancement model before the object recognition model cannot meet the real-time requirements. Therefore, we propose the Parallel Detection and Enhancement model (PDE), which detects objects and enhances poorly visible images in parallel and in real time. Specifically, we introduce the specially designed tiny prediction head along with coordinated attention and multi-stage concatenation modules to better detect underexposed and blurred objects. For the parallel image enhancement model, we adaptively develop improved weighting evaluation models for each "3D Lookup Table" module. As a result, PDE achieves better detection accuracy for poorly visible objects and more user-friendly reference in real time. Experimental results show that PDE has significantly better object recognition performance than the state-of-the-art on real foggy (8.9%) and low-light (20.6%) datasets.**

Keywords*—Low-visibility condition; image enhance; object detection

I. Introduction

Deep learning is used for many tasks, such as model fitting [1]–[3], object detection [4], [5], and so on. Recently, deep object detection models [6]–[8] have been widely used in daily life. These models provide accurate and 7×24 consistent object recognition, which facilitates people's work and helps them detect inconspicuous objects. However, in most places in the world, it is dark 42% of the day and there are one to three rainy or foggy days per week. These natural phenomena inevitably affect the performance of deep object detection models [9]–[11]. In addition, object detection models must filter out objects for user reference in real time (processing more than 30 frames per second). In this context, object detection in low visibility conditions has attracted much attention in both academia and industry to enable accurate and view-friendly object detection at all times of the day and under all climatic condition [12]–[14].

To mitigate the negative effects in low visibility conditions, current research can be divided into two classes: 1) Two-stage models: Two-stage models use image enhancement models to first enhance the images with poor visibility and then train the object detection model on the enhanced images [15], [16]. For example, [17] used "GridDehaze" to denoise foggy images, and [18] introduced a brightening step to lighten the images before object detection. 2) Joint Learning Models: Joint learning models jointly train an image brightening model and an object detection model to deal with poorly visible images [19], [20]. More specifically, two subnetworks [21] with common feature extraction layers are used to simultaneously detect the objects and brighten the images.

As shown in Fig. 1, current object detection models cannot detect and display objects well due to the following three challenges: 1) Out-of-focus and low-contrast objects. In low visibility conditions, objects are out of focus and low contrast, so their detectable areas are smaller and blurrier than in high quality images. This negatively affects the accuracy of conventional object detection methods originally developed for high-quality images. 2) Unfavorable representation in poor visibility conditions. Images captured in poor visibility conditions are not well visible to the user. It is necessary to enhance the images and frame the objects clearly for the user. 3) Real-time processing requirements. To meet real-time requirements, the entire process of image enhancement and object detection should be performed at more than 30 frames per second. In the two-stage models, the image enhancement step and the object recognition step are processed serially, which further limits the processing time of each step and results in unsatisfactory performance of both steps. In the joint learning models, the image enhancement model and the object detection model have completely different optimization objectives. The joint optimization of these two models may result in a wobble phenomenon, leading to better performance in image enhancement or recognition, but negatively affecting the other objective.

To this end, we study the problem of how to achieve better enhancement and detection while meeting real-time requirements. Specifically, we investigate the following three research questions. 1) How can the enhancement and detection steps be decoupled to improve both together? 2) How can objects with blurred edges be accurately detected? 3) How can images with poor visibility be enhanced to better present recognition results to users? By exploring the above questions, our work makes the following three contributions.

- We propose a novel parallel framework called Parallel Detecting and Enhancing models (PDE) that can solve the wobble problem while improving detection and enhancement performance.
- PDE introduces a tailored model for detecting objects

* Corresponding authors.

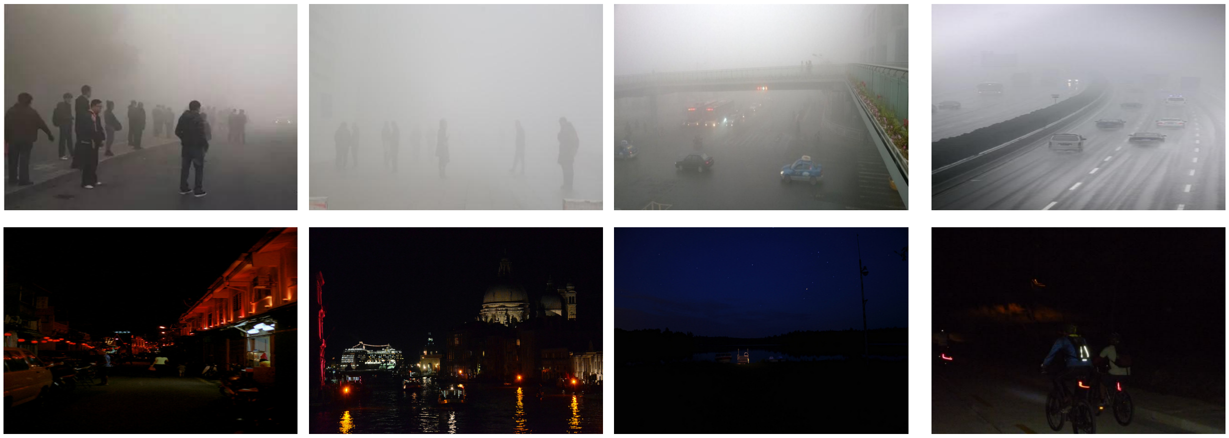


Fig. 1. Intuitive cases explaining the problems of object detection in low visibility conditions. The first row shows the images taken on foggy days. The second row shows the images taken in low-light conditions.

in low-visibility images by introducing a tiny prediction head to detect objects with smaller detection areas. In addition, PDE uses coordinated attention and multi-stage concatenation to further improve detection performance in low visibility.

- PDE incorporates a specially designed image enhancement model by developing an adaptively enhanced weighting model for each "3D Lookup Table" module to achieve better enhancement performance.

To prove the efficiency of our approach in object detection, we evaluate the proposed model on synthetic and real low visibility datasets. Experimental results show that PDE has significantly better object detection performance on two real datasets with fog (8.9%) and low light (20.6%). In addition, case studies show that PDE provides more accurate object detection and clearer rendering than other low visibility models.

The remainder of this article is organized as follows: Section II provides related work, including object detection, image enhancement, and multi-tasking in low visibility. Section III describes the proposed method used for low visibility images. Section IV presents the comprehensive experimental results of our method compared with other methods. Section V concludes our work.

II. Related Work

A. Object Detection

CNN-based target detectors can be divided into two types according to the steps of image processing: 1) Single-stage detectors: YOLOv4 [22], YOLOv5 [23], FCOS [24], and EfficientDet [25]. 2) Two-stage detectors: R-CNN [26], R-FCN [27], Mask R-CNN [28], Fast R-CNN [29], etc. From the point of view of composition, they both consist of two parts. One part is the CNN-based basic framework, which is used to extract image features. The other part is the prediction head, which is responsible for classification and localization. In addition, existing object detectors add some layers between the basic framework and the head, which are called the neck of the detector. The three structures are described in detail below.

Backbone. The backbone often uses VGG [30], ResNet [31], EfficientNet [32], CSPDarknet53 [33], Swin-Transformer [34], etc., rather than networks designed by ourselves, since these networks have been shown to have strong feature extraction capability in computer vision tasks. However, the backbone network can be fine-tuned to make it more suitable for specific tasks.

Neck. The neck was designed to make more efficient use of features extracted from the backbone network. Its main task is to further process and use the features extracted from the backbone in different stages. The neck usually consists of several top-down and several bottom-up paths. The neck is an important component of the object recognition network and connects the backbone to the head. Commonly used linking modules for the neck include FPN [35], NAS-FPN [36], PANet [37], BiFPN [25], ASFF [38], etc. The common point of these modules is the iterative use of various upsampling, downsampling, dot-sum or dot-product methods to develop aggregation strategies.

Head. In the detection task, the backbone cannot perform the localization task. Therefore, the head network is responsible for detecting the location and class of objects based on the feature maps extracted from the backbone. Head networks are generally divided into two categories: single-stage object detectors and two-stage object detectors. The most representative two-stage object detector is the R-CNN [26], [39] series. Compared to the two-stage detector, the single-stage object detector predicts both the bounding box and the object class simultaneously. The most representative single-stage object detectors are YOLO [22], [40], SSD [41], and RetinaNet [42] series.

B. Image Enhancement

Image adjustment determines a threshold based on the gray level range of the image. If it is below the threshold, automatic color gradation enhancement is applied. On the other hand, if it is above the threshold, enhancement methods based on histogram equalization and inverse equalization are performed. The adaptive image enhancement method can enhance not only low-contrast images, but also partially dark and partially

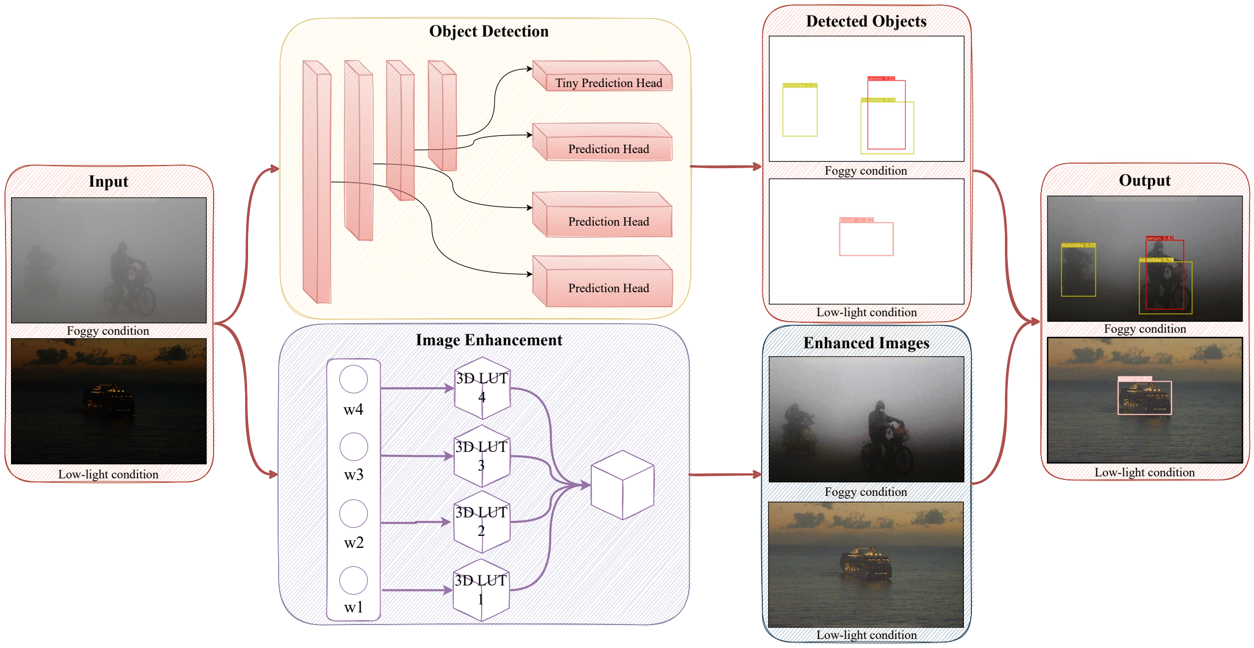


Fig. 2. An overview of the working pipeline with PDE. PDE can detect objects and enhance the low-visibility images in parallel and in real time. The label and coordinate values are obtained by the object detection model. Then we overlay the detected images with the enhanced image to get the final output.

light images with high robustness, so that the enhanced images have a better visual effect. Image adjustment is a widely used technique in image enhancement. Some classical methods [43], [44] use adaptive filters to control the contributions of the various enhancement operations so that contrast enhancement occurs in regions of high detail. [45] proposed a Deep Learning model that trains data on unpaired images. A Deep Reinforcement Learning approach is also used to decide what action to take given the current state of the images. [46] effectively transforms the color and hue of the source image by using a small CNN to learn image-adaptive 3D lookup tables.

C. Multi-Task in Low-Visibility Conditions

Existing models for object detection in low visibility include several tasks, such as image denoising and object detection. Depending on the order in which the different tasks are performed, they can be divided into two classes: two-stage models and joint learning models. Two-stage models [5], [15], [17] use classical visualization enhancement methods to process images before detection. For example, [47] proposed an AOD-Net for foggy conditions that denoises images before detection. However, the object detection models have strict requirements for deriving the time. When the image enhancement model and the object detection model are connected in series, the recognition time of both models is constrained, resulting in suboptimal performance in enhancement and detection. Joint learning models [19], [21] have performed image enhancement and object detection using a joint structure to better recognize images with low visibility. However, it is difficult to adjust the parameters to balance the completely different optimization goals of image enhancement and object detection. For this reason, [48] proposed an unsupervised adaptive system for object detection in rain and fog. After that, many works [49]–[51] emerged to improve the detection performance by using range adaptation. [52] proposed a robust learning method to resist interference from poor visibility and reduce the information loss caused by range adaptation. [20] developed a joint learning model (IA-YOLO) that combines image matching enhancement and object detection to meet the requirements of real-time recognition. In this work, we use the classical single-stage model YOLOv5 [53] as a basis and improve its performance under low visibility conditions.

III. Proposed Method

The PDE synchronizes the input image with the object detection module and the image enhancement module, as shown in Fig. 2. First, the object detection module detects the image to obtain the coordinates and classification information of the target. Second, the image enhancement module reduces the weather noise and increases the brightness of the input image to obtain a more user-friendly reference. Finally, the coordinates and classification information of the target are written into the enhanced image to obtain the result. In this section, we introduce the object detection and image enhancement modules.

A. Detection Network Module

Images captured in low visibility conditions contain interference from environmental information that makes object detection difficult. To overcome this challenge, PDE introduces a customized model for low-visibility object detection by importing a tiny prediction head and further employing coordinated attention and multi-stage concatenation to improve the performance of low-visibility detection. As shown in Fig. 3, the object detection model is a newly developed implementation of YOLOv5. These tricks help deep neural networks accurately locate and identify objects by reducing the detrimental effects

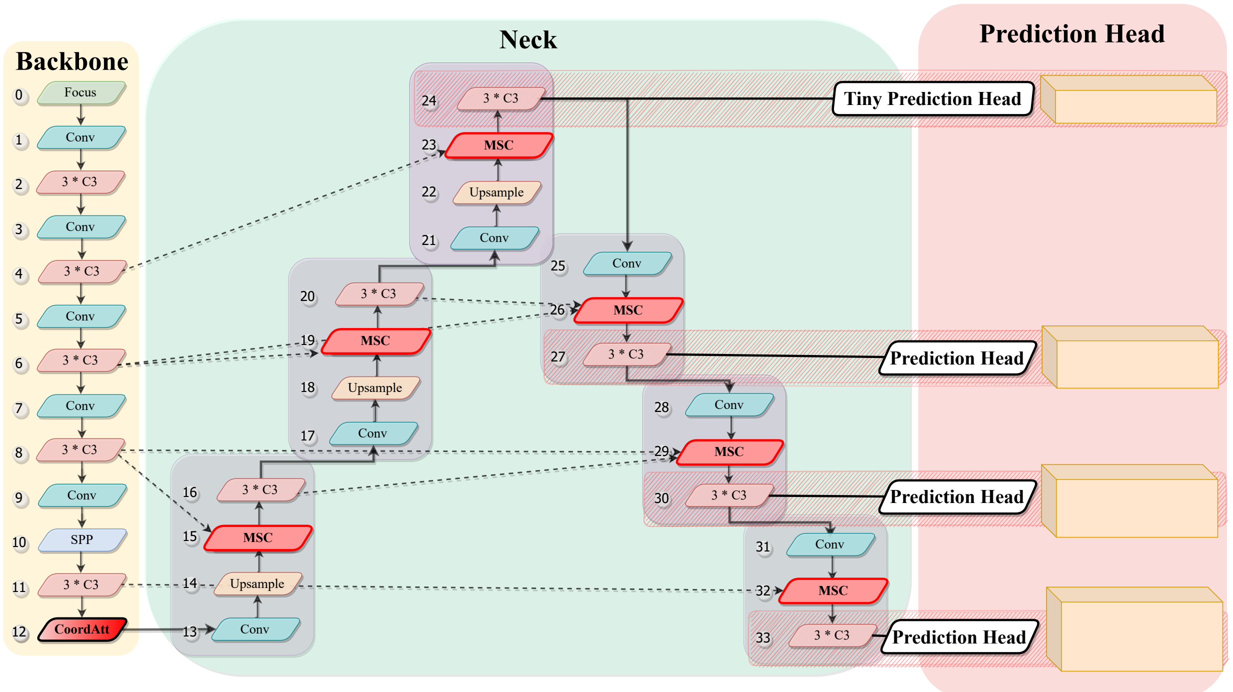


Fig. 3. The object detection module. 1) A coordinate attention block (CoordAtt) is located at the bottom of the backbone network. 2) The multi-stage concatenation module (MSC) replaces the original concatenation module. 3) The number of prediction heads has been increased from 3 to 4.

of detectable zones that are smaller and blurrier than in high-quality images. In addition, the model combines feature maps generated by shallow and deep neural networks so that semantic information and location features can be fully utilized. Moreover, an explicit supervised learning task is formed by setting a learning weight parameter. In this way, excellent learning results are obtained by perfectly distinguishing the importance of feature maps from shallow and deep neural networks. Therefore, the representative features can be accurately retrieved and appropriately represented, which improves the overall detection performance of the model. To get a clearer picture of the core of the object recognition module, we will describe the above methods and loss functions in detail below.

Coordinate Attention. Attentional mechanisms have been shown to be effective in many visual tasks. The core of the attentional mechanism is to enhance the model's ability to extract and represent important features, similar to the way humans selectively focus on important parts rather than the totality of information. However, most attention mechanisms only consider the information between channels and not the information about spatial location. This ignores the part of the information that is hidden in space and fails to extract the optimal representation of the features. Moreover, the convolution operation can only extract local relations, but not relations over long distances. To this end, we use coordinate attention [54] to capture spatial relationships over long distances with precise location data by embedding the location information into the channel attention. Specifically, each input X is decomposed into w and h dimensions, and the decomposed tensors are processed by global pooling to generate X_w and X_h, respectively. Feeding into a convolutional block with the concatenated tensors X_w and X_h generates an encoded Y that summarizes the extracted features of X_w and X_h. The set of operations can be formulated as follows:

$$Y = f(G(\{X_w, X_h\} \circ W)) \tag{1}$$

where $\{X_w, X_h\}$ means concatenating X_w and X_h, $\circ$ means convolution operation, G and f denote normalization and activation function, respectively, and W is the convolution filter. Furthermore, we split Y again to obtain $\hat{X}_w$ and $\hat{X}_h$:

$$\hat{X}_w, \hat{X}_h = Split(Y) \tag{2}$$

Also, $\hat{X}_w$ and $\hat{X}_h$ are convoluted and activated to get the final output:

$$Y_{out} = X \times \sigma(F(\hat{X}_w)) \times \sigma(F(\hat{X}_h)) \tag{3}$$

where F denotes convolution and σ denotes the sigmoid function.

By combining attention along the horizontal and vertical directions of the input sensor, each element of the attention maps can reflect in two directions whether the object of interest is present in the corresponding row and column. In this way, coordinated attention can more accurately determine the exact location of the object so that the entire model can better identify objects.

A Tiny Prediction Head. Feature sensitivity is one of the most important properties of the model for extracting key information from noisy images. The best way to achieve this is to add observations from different viewpoints and combine them to make better use of the fine-grained features from different viewpoints and achieve better feature representation.

However, we find that there are three different scales of prediction heads in the original YOLOv5 model, namely 256×256, 384×384 and 512×512, and the number of anchor images is 9. Although YOLOv5 has been observed from three perspectives, the recognition results are still not satisfactory for noisy images, as shown in Fig. 4. To this end, we add a tiny prediction head with a scale of 128×128 to extract features from a more microscopic perspective. We also increase the number of anchor frames from 9 to 12. Although this is only an incremental change to the detection head, the structure of the multi-detection head contributes significantly to improving the model performance, as shown in the Table IV of experimental results. We also found that the number of detection heads is not as large as possible and that the structure with four detection heads is the most stable and effective. Given the input X, the individual steps are as follows:

$$Z = \sum_{i=1}^{N} Anchor_i \oplus X_i \tag{4}$$

where N denotes the number of prediction heads. $Anchor$ denotes an anchor frame set based on prior knowledge. $\oplus$ denotes the matching of $Anchor$ with X to obtain a set of prediction boxes. Then, the result after matching is concatenated with the original input X. Finally, the convolution operation is used to continuously traverse the entire region to obtain the final feature map. The details are as follows:

$$Out = Conv(Concat(Z, X)) \tag{5}$$

where Out denotes the final feature map.

Multi-Stage Concatenation. The receptive field, the most important component in CNN-based models, is used to extract abstract features layer by layer. In deep layers, the receptive field is relatively large to extract features, and conversely, it is smaller in shallow layers. In general, feature maps captured by larger receptive fields have stronger semantic representation but weaker spatial representation. In contrast, feature maps captured by smaller receptive fields have weaker semantic representation but stronger spatial representation. To this end, we combine the original concatenation module with the BiFPN algorithm [25] to fully exploit the properties of feature maps from deep and shallow layers. The importance of features from different layers is defined by a learnable weighting parameter W. The parameter is defined as :

$$W_i = \frac{X_i}{\sum X_i + \varepsilon} \tag{6}$$

where X_i denotes the input of each layer, and W_i denotes the weight parameter of each input layer. ε is set to 0.0001.

According to Eq. 6 the formation of the result can be formulated as follows:

$$Y = Conv(ReLu(\sum_{i=0}^{N-1} W_i \times X_i)) \tag{7}$$

where Y denotes the result and $Conv$ denotes the convolution operation; $Relu$ denotes the activation function and N denotes the number of input layers. Both semantic features and spatial features can be transferred to different depth layers by feature fusion and mapping. This method improves the model's ability to extract and express features, and thus improves recognition performance.

Loss Function. The loss function of YOLOv5 divides the objective function into three subfunctions, namely object, classification, and regression. However, we found that the original loss function uses the basic Intersection-over-Union (IoU) loss, which limits the ability to measure the overall performance of the model. The total loss is calculated as follows:

$$Loss = w_{obj} \times loss_{obj} + w_{cls} \times loss_{cls} + w_{reg} \times loss_{reg} \tag{8}$$

where $loss_{obj}$, $loss_{cls}$, and $loss_{reg}$ denote the object objective function, classification objective function, and regression objective function, respectively. w_{obj}, w_{cls}, and w_{reg} denote their weighting values set a priori to 0.3, 0.05, and 0.7, respectively.

To better measure the difference between confidence in the predicted object and the true value, we construct this objective function $loss_{obj}$ based on cross-entropy loss. For a given predicted value x and a true value y, the equation is as follows:

$$loss_{obj} = -\frac{1}{n}\sum_{i=1}^{n}(y_i \times \ln i + (1 - y_i) \times \ln(1 - x_i)) \tag{9}$$

Although the predictions are multiclassification, there is only one positive sample, so we use the loss of cross entropy. After we use the cross entropy as the loss function, the gradient of the backpropagation is no longer associated with the derivative of the sigmoid function. This avoids the disappearance of the gradient to some extent. $loss_{obj}$ is the same as $loss_{cls}$.

In the regression task, the most direct indicator to determine the distance between the predicted box and the ground truth is the intersection over union (IoU), and $IoU = \frac{|A \cap B|}{|A \cup B|}$, however, does not accurately reflect the intersection of the two boxes and cannot be trained further due to disjunction. Therefore, we use the complete-IoU [55] to construct the objective function of the regression task. We consider the similarity of the aspect ratio between the ground truth and the predicted box.

$$loss_{reg} = 1 - IoU + \frac{\beta^2 \times (b_p, b_g)}{c^2} + \alpha \times \nu \tag{10}$$

$$\alpha = \frac{\nu}{\nu - IoU + (1 + \varepsilon)} \tag{11}$$

$$\nu = \frac{4}{\pi_2} \times (\frac{w_g}{h_g} - \frac{w_p}{h_p})^2 \tag{12}$$

where b_p and b_g denote the centers of the prediction box and the ground truth, respectively. β denotes the Euclidean distance between the two centers. c denotes the distance of the diagonals of the smallest region containing both the prediction box and the ground truth. α denotes the weighting parameter. ν is used to measure the similarity between the aspect ratio of the prediction box and the ground truth.

B. Image Enhancement Module

The 3D lookup table is an algorithm that reconstructs the hue of an image by creating a color map. The essence of the 3D lookup table is an mapping relation: $(R, G, B) = f(r, g, b)$, where f represents the mapping function. Moreover, it is an intuitive idea to learn a classifier to classify the scene. Suppose M 3D lookup tables, denoted by $\{\mu_n\}_{n=1,...,M}$, are learned. The classifier outputs N probabilities $\{p_n\}_{n=1,...,N}$ for classifying the scene. The process of 3D lookup table selection can be described as follows:

$$q = \mu_i(x), \quad s.t. \quad i = arg \ \max_n p_n \tag{13}$$

where x denotes an input image and q the output. Another common method for improving image quality is to manually adjust the parameters of a 3D lookup table. However, manually adjusting the parameters is extremely time consuming when processing large images. The parameters need to be adjusted based on scenarios that have different negative effects. Therefore, the applicability of the method is hindered by a lack of flexibility and practicality.

H. Zeng developed an end-to-end adaptive image enhancement method [46] based on 3D lookup tables and a convolutional network. The model learns how to improve image quality based on paired data, namely the affected images and the images optimized by experts. To this end, we propose a specific model for image enhancement by developing adaptive models for each "3D Lookup Table" to improve the weighting evaluation. Moreover, in this work, we train the image enhancement model by combining preprocessed images, i.e., foggy images and low-light images, with clear original images.

The image enhancement model introduces 4 basic 3D lookup tables along with a CNN-based model g that predicts weights for the output of each 3D lookup table. For an input image x, the final enhancement result is as follows:

$$q = \sum_{n=1}^{4} w_n \mu_n(x) \tag{14}$$

where $\{w_n\}_{n=1,...,4} = g(x)$ are the content-dependent weights output by the CNN-based model. Specifically, we use different 3D lookup tables to enhance different images. Moreover, the color space of the image is transformed using 3D lookup tables, while the CNN weight predictor extracts information about the image content, including hue, brightness, contrast, etc. The weights obtained by the CNN predictors are assigned to the corresponding 3D lookup tables. Therefore, our model adaptively improves the image quality according to the image content and scene in low visibility conditions.

IV. Experiments

We evaluate the effectiveness of PDE in fog and low-light conditions. We report the object detection metric mAP (average of all 10 IoU thresholds in the range of [0.5: 0.95]) and the image enhancement metrics PSNR (Peak Signal to Noise Ratio) and SSIM (Structure Similarity). We will present this section under the following aspects.

A. Experimental Details

Datasets. For the two tasks that PDE faces, i.e., target detection and image enhancement, we need to take different approaches to create datasets for the corresponding tasks so that we can effectively evaluate the performance of the model.

In object detection, we first evaluate the detection performance of the model under three conditions, including normal, foggy, and low light. We use the VOC dataset [56], [57] as a benchmark and the RTTS dataset [58] and the ExDark dataset [59] as test sets. To make better use of these datasets, we filtered out the common categories of the datasets. The VOC dataset shares five categories with the RTTS dataset, namely pedestrians, cars, buses, bicycles, and motorcycles. Similarly, the VOC dataset shares 10 categories with the ExDark dataset, namely, bicycles, boats, bottles, buses, cars, cats, chairs, dogs, motorcycles, and people. The VOC_5c training dataset and the VOC_5c test dataset, namely VOC_5c_train and VOC_5c_test, are created after screening and consist of 8111 and 2734 images, respectively.

Although we already have a dataset for normal conditions, we lack sufficient images of foggy conditions and low light conditions. Therefore, we use a weather simulation algorithm to simulate images under low visibility conditions. According [60], for the original input image $O(x)$, the foggy image $F(x)$ can be calculated as follows:

$$F(x) = O(x) \times g(x) + L \times (1 - g(x)) \tag{15}$$

where L denotes global atmospheric light, and $g(x)$ denotes the medium transmission map, which is defined as:

$$g(x) = e^{-\beta} \times s(x) \tag{16}$$

where β denotes the scattering coefficient of the atmosphere, and $s(x)$ denotes the scene depth which is calculated by

$$s(x) = -0.04 \times \rho + \sqrt{max(row, col)} \tag{17}$$

where ρ denotes the Euclidean distance from the current coordinate to the pixel coordinate of the image center, row and col represent the number of rows and columns of the images. Combining the Eq. 15, 16 and 17, we obtain the following equation for the generation of fog images:

$$F(x) = O(x) \times e^{-\beta} \times s(x) + L \times (1 - e^{-\beta} \times s(x)) \tag{18}$$

In this work, L is set to 0.5 and β is calculated using the formula $\beta = 0.05 + 0.01 \times Num$. Num is set to a random integer between 0 and 9. In this way, for each input image, we get up to 10 foggy images with different effects of fog concentration.

Similarly, we simulate low lighting conditions to create the low lighting conditions dataset. For a given input image, each pixel x in the image is transformed as follows:

$$f(x) = x^{\gamma} \tag{19}$$

where γ is determined randomly from a uniform distribution with a range of values of [1.5, 5].

TABLE I. An Overview of All Data Sets used in this Experiment

Dataset	Number
VOC_5c_train	8111
VOC_5c_test	2734
VOC_10c_train	12334
VOC_10c_test	3760
VOC_fog_train	8111
VOC_fog_test	2734
VOC_low-light_train	12334
VOC_low-light_test	3760
RTTS	4322
ExDark	2563
LUTs_fog_train	20000
LUTs_fog_test	2000
LUTs_low-light_train	20000
LUTs_low-light_test	2000

To achieve ideal recognition performance under normal and low visibility conditions, we use a hybrid data training scheme for PDE. Each image in the normal datasets has a 2/3 probability of being randomly tagged with some kind of fog or converted to a low-visibility image before being input to the model for training. The hybrid data contains images from both normal and low visibility situations. The model becomes more robust when it learns with normal and low visibility images simultaneously, resulting in high performance.

Second, training an image enhancement model for image enhancement tasks requires a large amount of data to achieve an excellent result. Therefore, we extend the data again based on the simulated images in foggy and low-light conditions in the object recognition task. For the foggy conditions, we first add three random fog patches to each image in the VOC_5c_train dataset. Second, we randomly select 20,000 images from this dataset to form the fog training dataset, i.e., LUTs_fog_train. Similarly, we first randomly add fog to each image in the VOC_5c_test dataset. Second, we randomly select 20,000 images from this dataset to form the test dataset under foggy conditions, i.e., LUTs_fog_test.

In low light conditions, we perform the same steps to obtain the dataset, i.e. LUTs_low-light_train and LUTs_low-light_test.

We count the number of all records for this experiment, as shown in Table I. VOC_5c_train, VOC_5c_test, VOC_10c_train, VOC_10c_test, VOC_fog_train, VOC_fog_test, VOC_low-light_train, and VOC_low-light_test denote training and test sets, respectively, for object detection under normal, foggy, and low-light conditions. RTTS and ExDark are real-world datasets consisting of images taken under foggy and low-light conditions, respectively. LUTs_fog_train, LUTs_fog_test, LUTs_low-light_train, and LUTs_low-light_test denote training and test datasets for image enhancement in foggy and low-light conditions, respectively.

Baselines. This work focuses on improving the accuracy of object detection in low visibility conditions, complemented by image enhancement techniques to obtain more user-friendly references. Therefore, we perform comparison experiments and ablation experiments mainly for the object recognition module, while for the image enhancement module, we only present its experimental results without detailed comparison with other excellent methods.

To evaluate the universality and effectiveness of PDE in fog and low-light conditions, we choose YOLOv5 as our baseline model. In addition, we compare our model with other excellent models for detecting objects in low visibility. We choose the real-time target detection model YOLOv3 as our comparison model. We also choose GridDehaze [35], MSBDN [5], and ZeroDCE [15], the most widely used CNN-based image enhancement methods, to process images before detection and then combine them with the object detection model YOLOv3 [43]. GridDehaze and MSBDN are both image enhancement models for removing fog by developing novel network modules to learn more effective feature representations for image unveiling. ZeroDCE achieves effective image enhancement by implementing intuitive and simple nonlinear curve mapping to adapt to different lighting conditions. For the domain adaptation approach, we choose DAYOLO [19], which combines multiple adaptation paths and corresponding domain classifiers with the YOLO object detector to produce domain-invariant features. For the multi-task learning algorithm, we choose DSNet [22], which can learn denoising and detection together. We also choose IA-YOLO [34], which can adaptively enhance each image to improve detection performance.

B. Experiments Results

To fully demonstrate detection performance, for each model we evaluate the model's ability to recognize objects under different conditions, namely normal, foggy, and low light. The improvements are calculated by comparing PDE with the best baseline (underlined). From Table II and Table III, it can be seen that PDE significantly outperforms the other SOTA models at low visibility in the detection scene in all data sets and at all settings. In particular, for the mAP metric, PDE outperforms the baseline model by 8.9% (RTTS) and 19.7% (VOC_fog_test) in foggy conditions. In low-light conditions, PDE outperforms the baseline model by 20.6% (ExDark) and 15.8% (VOC_low-light_test).

These results demonstrate the consistent superiority of our PDE in detection performance under poor visibility conditions. Moreover, the PDE also performs better than the corresponding best baselines in a normal scene. This phenomenon proves the strong scalability of PDE.

Image enhancement is a secondary task that helps improve the display for the user, as this work focuses on target detection. Therefore, we did not perform comparison experiments for the image enhancement task. We evaluate the model's ability to enhance images in fog and low-light conditions. For the PSNR metric, PDE achieves a score of 23.64 (fog test set) and 23.97 (low-light test set). For the SSIM metric, PDE achieves a value of 0.838 (fog test set) and 0.827 (low-light test set). In the following subsection IV-D, we conduct a case study to illustrate the excellent results of the image enhancement task.

C. Ablation Study

To test the effectiveness of the object detection model in PDE, we compare the detection performance of our model

TABLE II. COMPARISON OF DETECTION PERFORMANCE WITH BASELINES IN TWO SCENARIOS, INCLUDING NORMAL AND FOGGY CONDITIONS. THE IMPROVEMENTS ARE COMPUTED BY COMPARING OUR MODEL WITH THE CORRESPONDING BEST BASELINES (UNDERLINED)

Model	Train data	VOC_5c_test	VOC_fog_test	RTTS
MSBDN [5]	VOC_5c_train	-	57.4	30.2
GridDehaze [35]	VOC_5c_train	-	58.2	31.4
DAYOLO [19]	Hybrid data	56.5	55.1	29.9
DSNet [22]	Hybrid data	53.3	67.4	28.9
IA-YOLO [34]	Hybrid data	73.2	72.0	37.0
YOLOv3 [43]	VOC_5c_train	70.1	31.1	28.8
YOLOv3 [43]	Hybrid data	64.1	63.4	30.8
YOLOv5 [23]	VOC_5c_train	86.2	68.5	45.1
YOLOv5 [23]	Hybrid data	<u>85.6</u>	<u>71.4</u>	<u>50.5</u>
PDE	Hybrid data	**86.7**(1.3%↑)	**85.5**(19.7%↑)	**55.0**(8.9%↑)

TABLE III. COMPARISON OF RECOGNITION PERFORMANCE WITH BASELINES IN TWO SCENARIOS, INCLUDING NORMAL AND LOW LIGHT CONDITIONS. THE IMPROVEMENTS ARE COMPUTED BY COMPARING OUR MODEL WITH THE CORRESPONDING BEST BASELINES (UNDERLINED)

Model	Train data	VOC_10c_test	VOC_low-light_test	ExDark
ZeroDCE [15]	VOC_10c_train	-	33.6	34.4
DAYOLO [19]	Hybrid data	41.7	21.5	18.2
DSNet [22]	Hybrid data	64.1	43.8	37.0
IA-YOLO [34]	Hybrid data	70.0	59.4	40.4
YOLOv3 [43]	VOC_10c_train	69.1	45.9	36.4
YOLOv3 [43]	Hybrid data	65.3	52.3	37.0
YOLOv5 [23]	VOC_10c_train	78.2	60.8	43.2
YOLOv5 [23]	Hybrid data	<u>77.1</u>	<u>64.5</u>	<u>45.0</u>
PDE	Hybrid data	**79.5**(3.1%↑)	**74.7**(15.8%↑)	**54.3**(20.6%↑)

TABLE IV. ABLATION ANALYSIS OF MODULES OF OUR MODEL IN REAL DATA SETS UNDER LOW VISUAL CONDITIONS. CA DENOTES THE COORDINATE ATTENTION MODULE. MSC DENOTES THE CONCATENATION MODULE COMBINED WITH MULTI-STAGE FEATURE FUSION. MH DENOTES THE MODULE WITH MULTIPLE PREDICTION HEADS. THE IMPROVEMENTS ARE COMPUTED BY COMPARING THE VARIANTS WITH YOLOV5 (UNDERLINED)

Model	Method			RTTS	ExDark
	CA	MSC	MH	mAP	mAP
YOLOv5	✗	✗	✗	<u>50.5</u>	<u>45.0</u>
PDE w/o MH	✓	✓	✗	52.5 (3.9%↑)	49.9 (10.9%↑)
PDE w/o MSC	✓	✗	✓	52.9 (4.7%↑)	53.6 (19.1%↑)
PDE w/o CA	✗	✓	✓	**55.0** (8.9%↑)	53.8 (19.5%↑)
PDE	✓	✓	✓	**55.0** (8.9%↑)	**54.3** (20.6%↑)

with its variants on two real datasets (RTTS and ExDark) in Table IV. In the following experiments, we use the data as our training dataset. "PDE w/o MH" means we omit the tiny prediction head in PDE. "PDE w/o MSC" means we omit the multi-stage concatenation module in PDE. "PDE w/o CA" means we omit the coordinate attention module in PDE.

As shown in Table IV, "PDE w/o MH" is 3.9% and 10.9% higher than YOLOv5 in RTTS and ExDark, respectively. However, "PDE w/o MSC" is 4.7% and 19.1% higher than YOLOv5, respectively, whereas "PDE w/o CA" is 8.9% and 19.5% higher than YOLOv5, respectively. Although the coordinate attention module can improve the performance of the model, the effect is not very large when the Table IV is analyzed. On the contrary, the tiny prediction head and multi-stage concatenation module significantly improve the performance of the model. In particular, the growth rate obtained with "PDE w/o MSC" reaches the maximum in ExDark, which proves that the module uses the features extracted from the backbone network very effectively under low light conditions. Moreover, the growth rate of "PDE w/o CA" reaches the maximum in RTTS, where the multi-stage concatenation module fully utilizes the effective features in the images combined with the tiny prediction head to perform target detection.

According to Table IV, PDE consistently outperforms the other variants, underscoring the need for and effectiveness of these methods, as noted in III-A.

D. Case Study

In Fig. 4, we visualized the detection result on two real datasets (RTTS and ExDark). In particular, we compare the detection results of the base models YOLOv5 and PDE. As you can see in Fig. 4, PDE can achieve better object detection accuracy and user representation in low- visibility. Moreover, inference time is an important metric to evaluate

Fig. 4. Visualization. Detection results of YOLOv5 (middle row) and PDE (bottom row) on RTTS (columns 1, 2) and ExDark (columns 3, 4). PDE achieves better object detection accuracy and better visualization for the user in low visibility images.

the practicality of models. Therefore, we conduct extensive test experiments to evaluate PDE by processing 480×480 images on a single GTX 2080Ti GPU. The experiment showed that PDE can process more than 30 frames per second. Therefore, PDE can achieve better user representation and detection while meeting real-time requirements.

V. Conclusion

In this paper, we note that the existing low visibility models suffer from the wobble phenomenon caused by the absence of better detection and image enhancement performance. We propose the parallel detection and enhancement model (PDE) to ensure that image enhancement and object detection perform their tasks. For object detection, PDE introduces a tailored model for low-visibility object detection by introducing a tiny prediction head, combined with coordinate attention and multi-stage concatenation modules. For image enhancement, PDE proposes a dedicated image enhancement model by developing an adaptively enhanced weighting model for each "3D Lookup Table" module. By decoupling these two concepts, PDE can improve the overall performance. Extensive experiments show that PDE achieves better accuracy in detecting low-visibility objects and more user-friendly reference in real time in all situations.

Acknowledgment

This work was supported by the National Natural Science Foundation of P. R. China (Nos. 62002068).

References

[1] Lin S, Xiao G, Yan Y, Suter D, Wang H. "Hypergraph optimization for multi-structural geometric model fitting," In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 8730-8737.

[2] Lin S, Luo H, Yan Y, Xiao G, Wang H. "Co-clustering on Bipartite Graphs for Robust Model Fitting," *IEEE Transactions on Image Pro cessing*, 2022, 31: 6605-6620.

[3] Lin S, Wang X, Xiao G, Yan Y, Wang H. "Hierarchical representation via message propagation for robust model fitting," *IEEE Transactions on Industrial Electronics*, 2020, 68(9):8582-8592.

[4] Z. Xia, S. Song, L. E. Li, and G. Huang, "3d object detection with pointformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7463–7472.

[5] Yang H, Lin S, Cheng L, Lu Y, Wang H. "SCINet: Semantic Cue Infusion Network for Lane Detection," *IEEE International Conference on Image Processing*, 2022, 1811-1815.

[6] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao, and Z. Han, "Effective fusion factor in FPN for tiny object detection," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 1160–1168.

[7] P. Sun et al., "Sparse r-cnn: End-to-end object detection with learnable proposals," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14454–14463.

[8] T. Yin, X. Zhou, and P. Krahenbuhl, "Center-based 3d object detection and tracking," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 11784–11793.

[9] A. Pfeuffer, M. Schön, C. Ditzel, and K. Dietmayer, "The ADUULM-Dataset-a Semantic Segmentation Dataset for Sensor Fusion.," 2020.

[10] T. Song, Y. Kim, C. Oh, and K. Sohn, "Deep Network for Simultaneous Stereo Matching and Dehazing.," in *BMVC*, 2018, p. 5.

[11] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640.

[12] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[13] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-Light Image/Video Enhancement Using CNNs.," in *BMVC*, 2018, vol. 220, no. 1, p. 4.

[14] Q. Zhu, J. Mai, and L. Shao, "Single image dehazing using color attenuation prior.," 2014.

[15] H. Dong et al., "Multi-scale boosted dehazing network with dense

feature fusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2157–2167.

[16] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[17] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7314–7323.

[18] W. Ren et al., "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.

[19] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.

[20] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions," *arXiv preprint* arXiv:2112.08088, 2021.

[21] S.-C. Huang, T.-H. Le, and D.-W. Jaw, "DSNet: Joint semantic learning for object detection in inclement weather conditions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 8, pp. 2623–2633, 2020.

[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint* arXiv:2004. pp. 10934, 2020.

[23] G. Jocher et al., "ultralytics/yolov5: v5. 0-YOLOv5-P6 1280 models AWS Supervise. ly and YouTube integrations," *Zenodo*, vol. 11, 2021.

[24] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9627–9636.

[25] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10781–10790.

[26] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[27] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, vol. 29, pp. 379-387, 2016.

[28] 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.

[29] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint* arXiv:1409.1556, 2014.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[32] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019, pp. 6105–6114.

[33] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.

[34] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.

[35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[36] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Nas-fpn: Learning scalable feature pyramid architecture for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7036–7045.

[37] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.

[38] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," *arXiv preprint* arXiv:1911.09516, 2019.

[39] T. Cheng, X. Wang, L. Huang, and W. Liu, "Boundary-preserving mask r-cnn," in *European conference on computer vision*, 2020, pp. 660–676.

[40] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint* arXiv:1804.02767, 2018.

[41] W. Liu et al., "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016, pp. 21–37.

[42] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[43] A. Polesel, G. Ramponi, and V. J. Mathews, "Image enhancement via adaptive unsharp masking," *IEEE transactions on image processing*, vol. 9, no. 3, pp. 505–510, 2000.

[44] W. Wang, Z. Chen, X. Yuan, and F. Guan, "An adaptive weak light image enhancement method," in *Twelfth International Conference on Signal Processing Systems*, 2021, vol. 11719, p. 1171902.

[45] Y. Hu, H. He, C. Xu, B. Wang, and S. Lin, "Exposure: A white-box photo post-processing framework," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 2, pp. 1–17, 2018.

[46] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, "Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[47] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "An all-in-one network for dehazing and beyond," *arXiv preprint* arXiv:1707.06543, 2017.

[48] V. A. Sindagi, P. Oza, R. Yasarla, and V. M. Patel, "Prior-based domain adaptive object detection for hazy and rainy conditions," in *European Conference on Computer Vision*, 2020, pp. 763–780.

[49] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3339–3348.

[50] M. Hnewa and H. Radha, "Multiscale domain adaptive yolo for cross-domain object detection," in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 3323–3327.

[51] S. Zhang, H. Tuo, J. Hu, and Z. Jing, "Domain Adaptive YOLO for One-Stage Cross-Domain Detection," in *Asian Conference on Machine Learning*, 2021, pp. 785–797.

[52] M. Khodabandeh, A. Vahdat, M. Ranjbar, and W. G. Macready, "A robust learning approach to domain adaptive object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 480–490.

[53] Y. Fang, X. Guo, K. Chen, Z. Zhou, and Q. Ye, "Accurate and Automated Detection of Surface Knots on Sawn Timbers Using YOLO-V5 Model.," *BioResources*, vol. 16, no. 3, 2021.

[54] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13713–13722.

[55] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 07, pp. 12993–13000.

[56] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[57] M. Everingham and J. Winn, "The pascal visual object classes challenge 2012 (voc2012) development kit," *Pattern Analysis, Statistical Modelling and Computational Learning*, Tech. Rep, vol. 8, no. 5, 2011.

[58] B. Li et al., "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[59] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2019.

[60] H. Israël and F. Kasten, "Koschmieders theorie der horizontalen sichtweite," in *Die Sichtweite im Nebel und die Möglichkeiten ihrer künstlichen Beeinflussung, Springer*, 1959, pp. 7–10.

Design and Implementation of an Unreal Engine 4-Based Smart Traffic Control System for Smart City Applications

Md. Imtiaz Hossain Subree[1], Md. Rakib Hasan[2], Maksuda Haider Sayma[3]
Dept. of ICT, Comilla University, Cumilla - 3506, Bangladesh

***Abstract*—Traffic congestion is a serious problem nowadays, especially in Dhaka city. With the increasing population and automation, it has become one of the most critical issues in our country. There can be a lot of causes of congestion in traffic, such as insufficient capacity, large red signal delay, unrestrained demand, etc, which causes extra time delay, extra fuel consumption, a speed reduction of vehicle, and financial loss. The traffic control system is one of the most important factors affecting traffic flow. Poor traffic management around these hotspots could result in prolonged traffic jams. Small critical locations that are frequent hotspots for congestion are a common byproduct of poorly constructed road networks in many developing countries. In this research, we first offer a straightforward automated image processing method for analyzing CCTV camera image feeds to determine the level of traffic congestion. Our system's design seeks to use real-time photos from the cams at traffic intersections to calculate traffic density using image processing, and to adjust the traffic signal on the data obtained based on the current traffic congestion on the road. We suggest tailoring our system to erratic traffic feeds with poor visual quality. Using live Surveillance camera feeds from multiple traffic signals in Dhaka city, we demonstrate evidence of this bottleneck breakdown tendency persisting over prolonged time frames across multiple locations. To partially address this problem, we offer a local adapting algorithm that coordinates signal timing behavior in a restricted area and can locally minimize congestion collapse by maintaining time-variant traffic surges. Using simulation-based research on basic network topologies, we show how our local decongestion protocol may boost the capacity of the road and avoid traffic collapse in limited scenarios.**

Keywords*—*Traffic control system; traffic congestion; CCTV; image processing; simulation

I. INTRODUCTION

Within the cutting-edge generation, the exponential growth of traffic congestion is among the most vital problems facing our civilization today. Due to the rise in automobiles in urban regions, many road networks are experiencing issues with the decline in road capacity and the accompanying stage of the carrier.

Traffic congestion eats up 5 million working hours each day in Dhaka city, which costs the country 11.4B USD a year. Vehicular speed has decreased from 21kmph to 6.5kmph in the last 12 years, which causes 40% of fuel wastage. 40% of the fuel is being wasted. There are 70 signalized intersections in Dhaka city. In our country, generally, 2 types of systems are used for controlling traffic at intersections.

1) Manual Traffic Control System: In the manual traffic control system, to manage traffic, a certain region is assigned to traffic police. To regulate traffic, traffic police utilize signboards, whistles, lasers, and other tools.
2) Automatic Traffic Control System: In an automatic system, Traffic Lights are used to control the traffic. A timer is used to control the signals of the traffic lights. The lights get off and on based on the timer value. [1]

The conventional methods have distinct drawbacks. There is a significant labor requirement for the manual controlling system. We are unable to have traffic police manually regulate traffic in all sections of a town or city due to the inadequate number of traffic officers. Conventional traffic control employs a traffic signal with a timer for each phase that is static and cannot change based on the volume of traffic on that particular road in real time. It's always fixed and do not alter or update in real-time in accordance with traffic on the roads [1].

The main problem with the conventional automatic traffic control system is, the traffic lights use a timer for every segment, which is rigid and does update or change based on the real-time traffic on the road. During morning and evening, the majority of the traffic movement direction is always opposite. Taking into account this theory, one particular direction with heavier traffic is given the same amount of green signal time as the other direction, which has lesser traffic than the other or in some cases, no traffic at all, which is inconvenient.

So, the main problems identified by us were

- Manual Traffic control system is inconvenient and faulty.
- Automatic Traffic control system uses a fixed timer which is rigid and does not update with the real-time traffic on the road.

From this existing problems, the following research questions arise:

- Is there a better way to replace the existing system?
- Can we upgrade the existing automatic traffic control system with a timer that can adjust to the existing traffic condition and update based on that?

Our suggested method intends to provide a computer vision-based traffic signal controller that can adjust to the existing traffic condition. It makes use of actual footage from the Surveillance cameras installed at road intersections to determine the traffic density by measuring the number of cars at the light and altering the length of the green signal accordingly. The vehicles are separated into four categories: automobiles, trucks, buses, and bikes to calculate the correct calculation of the green light time. The system employs YOLOv5 to compute the number of automobiles before modifying the traffic signal's timer in line with the suitable direction of vehicle density. In contrast to a static system, traffic is cleared significantly more swiftly thanks to this, maximizing the periods whenever the green light is on. Unwanted delays, traffic, and waiting times are reduced as a result, which also reduces fuel consumption and emissions.

The main contribution of our proposed system are as follows:

- Proposing a better solution for timer based automatic traffic control system.
- Using the existing Surveillance cameras on the road to collect the traffic info and update the signals accordingly.
- Using YOLOv5, which is 88% smaller and 180% faster than YOLOv4, to compute the number of automobiles before modifying the traffic signal's timer in line with the suitable direction of vehicle density.
- Achieving almost perfect result by counting the number of vehicles, identifying the types of vehicles, measuring the speed of each vehicle and feeding proper value to the server.

II. LITERATURE REVIEW

de Frías, C. J., Al-Kaff, A., Moreno, F. M., Madridano, Á., and Armingol, J. M. describe a cooperative system for monitoring traffic flow based on the use of unmanned aerial vehicles (UAVs) in [10]. A lightweight semantic neural network was suggested by this system, which uses RGB photos as input to produce a segmented image of the vehicles and their 2D actual positions. Then, each car receives this information so that it may fully understand its environment. The suggested approach has undergone multiple experiments with a variety of scenarios and conditions, and the outcomes demonstrate the system's effectiveness and robustness while also demonstrating how it may be used for traffic monitoring.

PJ Navarathna, VP Malagi proposed some important technology and answers to the many issues that the lack of digitalization was causing the populace in [9]. It addresses problems with the city's infrastructure, public security, and safety, and offers the best remedies. It places a strong emphasis on using AI in conjunction with other technologies, such as the Internet of Things (IoT), deep learning, machine learning, pattern recognition, big data analytics, and cloud infrastructures, to create a fully functional smart city.

Karimi, A., and Duggirala, P. S. presented a new approach to formalizing and implementing traffic rules in [12]. The DMV driver manual for California is used as a working example. Its method offers a clear mapping between the handbook's rules and their formal model, as well as between the model and implementation. They use the Clingo programming language to explicitly model the traffic regulations in the Answer Set Programming (ASP) logic programming paradigm to show the effectiveness of this method. They later incorporate these regulations into CARLA, a computerized testing ground for autonomous automobiles. By correctly applying the right-of-way rules for autonomous cars in real-time, they replicate the behavior of autonomous vehicles at four-way and three-way uncontrolled junctions. As a result, as compared to CARLA's default FIFO controller, the behaviors of autonomous vehicles under our controller are more realistic.

Manjón Prado, Javier examined the escalating issue of traffic congestion, as well as various potential solutions, such as traffic simulation software and autonomous vehicles in [11]. The new choice for recreating a traffic scenario is Unreal Engine. Using a click-and-drag editor, a tool has been developed that enables users to build and alter road tracks and add autonomous car models to them. Different scenarios are used to evaluate the simulator while important data is collected. Finally, the results are discussed together with suggestions and directives for future improvement.

Khushi used video processing to fix the congestion problem in [1]. Before being delivered to the servers, the footage from the live stream is processed. There, a CPP-based algorithm is utilized to generate the outcome. The dynamic algorithm demonstrated a 35

Renjith Soman employed ANN and a fuzzy controller to suggest a smart traffic signal system in [2]. Images obtained from cameras placed at traffic locations are used by this system. Before further normalization, the image is initially transformed to a grayscale version. Following that, the segmented image is passed through an ANN to count the automobiles regardless of size, and the output is then used in a control scheme to set periods for the green and red lights using crisp output. Results had a 2

Chabchoub, A., Hamouda, A., Al-Ahmadi, S., and Cherif, A. designed a smart traffic light controller using fuzzy logic and image processing with MATLAB, to control movement in two ways, aided by a camera and auto sensors. in [3]. The Fuzzy logic has two inputs and six outputs designed, the console input is the number of cars on each road and the time of the assumed red, yellow and green signal according to the vehicles congestion. The simulation result is similar to the proposed control unit, as it deals with the lights simultaneously according to the number of cars in each branch of the road, which leads to the use of all the time to operate the stoplights.

A. A. Zaid, Y. Suhweil, and M. A. Yaman suggested a traffic signal that can be adjusted to the prevailing traffic conditions using fuzzy logic [4]. For the major and secondary driveways in this system, two fuzzy control systems with three inputs and one output are used. Simulated traffic conditions were improved for lower traffic density using VISSIM and MATLAB.

Al-Majhad, H. G., Bramantoro, A., Syamsuddin, I., Yunianta, A., Basori, A. H., Prabuwono, A. S., and Barukab, O. M. aims to benefit from the current Riyadh road infrastructure and apply the Internet of Things paradigm for detecting traffic congestion with Everything as a Service approach. in [5]. By using the services, the users are able to identify the exact location where congestion occurs and an alternate solution can be provided easily. To achieve this, Business Process Execution Language is embedded as a supporting framework layer. This approach clearly defines how the services are executed through the proposed models.

Agrawal, Aditi, and Rajeev Paulus proposes Fuzzy Adaptive Control System (FACS) that uses fuzzy logic to decide the phase sequence and green-time for each lane based on sensed input parameters in [6]. It is designed with an aim to improve traffic clearance at isolated intersection especially in peak traffic hours of the day along with giving precedence to emergency vehicle as soon as it is detected and also assist pedestrian passage thus reducing their waiting time at the intersection. Performance of the proposed Fuzzy Adaptive Control System (FACS) is evaluated through simulations and compared with Pre-Timed Control System (PTCS) and Traffic Density-based Control System (TDCS) at a busy intersection with lanes leading to offices, schools and hospitals. Simulation results show significant improvement over PTCS and TDCS in terms of traffic clearance, immediate addressing of the emergency vehicle and giving preference to pedestrian passage at the intersection.

Adbeb, Tesfanesh, Di Wu, and Muhammad Ibrar presented heuristic algorithms called Congestion-Free Path (CFP) and Optimize CFP (OCFP), in SD-VANET architecture in [7]. The proposed algorithms address the traffic congestion issue and also provide a feasible path (less end-to-end delay) for a vehicle in VANET. They used the NS-3 simulator to evaluate the performance of the proposed algorithms, and for generating a real scenario of VANET traffic; we use the SUMO module. The results show that the proposed algorithms decrease road traffic congestion drastically compared to exiting approaches.

Saili Shinde examined the various traffic signal control system techniques in [8]. This study notes that each method has a similar architecture, which includes selecting input data, extracting traffic parameters from it, processing it, figuring out density, and updating parameters. The numerous traffic signal control system methods are examined in this reference. This paper shows that the architecture of each approach is similar and consists of picking data input, obtaining traffic data from it, analyzing it, calculating intensity, and changing parameters. The first method involves the employment of VANETS to collect information on each vessel's position, which is subsequently sent via installed GPS to the nearest Intelligent Traffic Light (ITL). Additionally, these ITLs will update the vehicles around with statistics. In the event of incidents, information would be provided to drivers so they could choose an alternative path to avoid traffic. This method is not workable because of the high deployment costs. The second technique uses a transceiver to capture each car's token utilizing infrared sensor-based embedded systems. Vehicles' data lines regularity tags can be used to identify them in an emergency and let vehicles pass. By adopting this technique, red light violations are discovered. However, this strategy is not flexible because infrared sensors need to be visible. The third methodology employs a fuzzy logic approach and two probabilistic reasoning controllers, one of which lengthens the green period of a route at an intersection and the other of which optimizes the signal. The outgoing and incoming lines' record cameras act as sensors to collect data from input sources. The controller then uses the data gathered by these detectors to determine the best course of action and minimize the goal function. The fourth method uses fuzzy reason, and the system's inputs include the number of vehicles and the average speed of movement in each route. Using sensors placed along the route, it is possible to calculate the number of cars and the cruising pace of the flow of traffic. The fifth approach makes use of spaced-apart photoelectric sensors to collect data and transmit it to the traffic panel, which weighs each road individually and adjusts the traffic light as necessary. However, maintenance is quite expensive. Video image processing is employed to detain the data in the sixth technique. Dynamic noise removal and numerous morphological processes are conducted to get a clear illustration of the automobile. A fresh rectangle is generated, and the vehicle calculation is increased each time a new car enters the subject of attention. Although the algorithm is simple to use, it does not account for shadow overlapping and occlusion.

Golubev, Kirill, Aleksandr Zagarskikh, and Andrey Karsakov described a brand-new framework for agent-based traffic modeling that combines various traffic model classes into a single vehicle agent and lets the user choose a particular model for each supported class in [13]. The framework was created with the help of the Unreal Engine 4 gaming engine, which enables users to simulate realistic vehicle interaction and create realistic visualizations for visual analysis and computational steering.

A. *Limitations of Related Works*

- The constraints of a vehicular network scenario that new communication, data, and resource approaches must address to increase overall network performance, such as low power, full storage, message's time-to-live, and bandwidth.
- SD VANET has insufficient performance evaluation, Uncertainty over the proposal's usefulness and correctness, as well as integration problems with fog, SDN, and VANET.
- No security analysis, no solution when connection loss with controller occurs. only weak security analysis.
- ANN and a fuzzy controller has hardware dependency, Sharing Difficulty, Overfitting Likely, ANN Solution Not Guaranteed.
- Fuzzy logic has two significant drawbacks: it cannot handle ambiguous data, and it cannot infer human thought.

III. Proposed System

Our proposed system by dynamically calculating the green signal period based on the volume of traffic at the signal hopes to reduce such instances. By doing this, it will be ensured that the way with greater traffic receives a longer green signal duration than the route with a relatively lower volume of traffic. The suggested solution uses image processing and object detection to calculate real-time traffic density from images captured by CCTV cameras placed at traffic intersections. Fig. 1 illustrates how this image is transmitted to the vehicle detecting algorithm that uses YOLOv5. To determine the traffic density, the number of each type of vehicle is counted, including automobiles, cyclists, buses, and trucks. This density, along with a few other variables, is used by the signal-switching algorithm to set the green signal timer for each lane. Accordingly, the red signal times are modified. To prevent a particular lane from going without traffic, there is a highest and lowest value for the green signal time.

So, the proposed system will work this way.

- The proposed system will use Closed Circuit Television (CCTV) for collecting live videos and snapshots installed at traffic signals.
- Snapshots from the CCTVs will be passed to servers at intersections for combining computer vision and image processing to calculate traffic density.
- For detecting vehicles, the YOLOv5 model will be used as shown before.
- A time scheduling algorithm will use these values and update the green and red signal times for each signal.

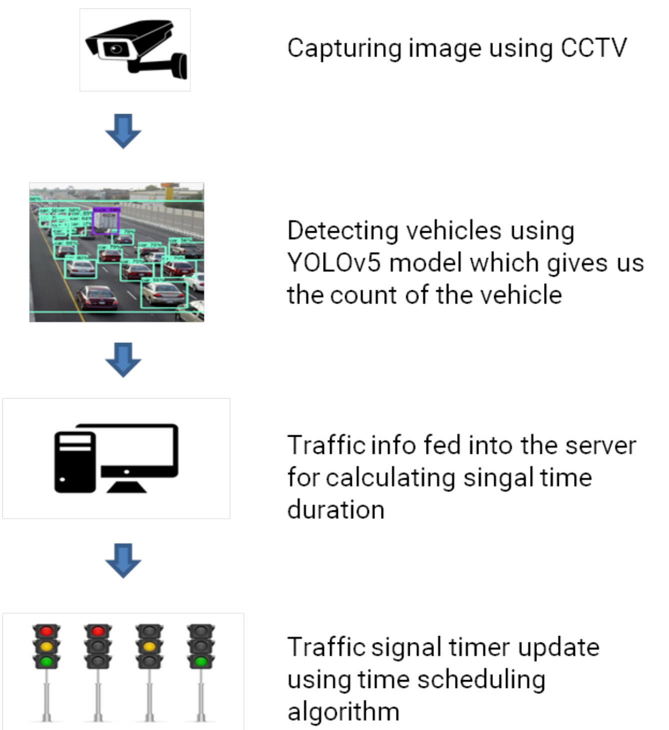


Fig. 1. Proposed system

Fig. 1 Shows the flow diagram of the proposed system.

A. Vehicle Detection System

The proposed system uses YOLOv5 (You Only Look Once) for automobile recognition. This is one of the most well-known object detection algorithms for its accuracy and speed. It can perceive vehicles of different modules and can provide the desired accuracy and processing time. A single neural network is applied to the entire image by the algorithm, which then divides it into regions and forecasts enclosing frames and possibilities for each region. The projected probabilities are used to weigh these bounding boxes. By downloading photos from Google and manually labeling them using the visual image annotation tool LabelIMG, the dataset for training the replica was created. Then, the system was taught by pre-guided credence obtained from the YOLO site. By altering the 'classes' variable, the number of yield synapses in the ultimate coating was placed to be identical to the number of modules the system is intended to perceive. The replica was taught until the loss was much lower and didn't appear to be decreasing after making these configuration adjustments. The OpenCV library was then used to detect vehicles using these weights that were entered into the program. Following the model has been filled and a figure has been fed into it, it outputs the results in a JSON format, or as pairs of keys and values, where keys are labels and the values are the labels' confidence and coordinates.

Fig. 2 shows the test result of our vehicle detection system. Applying the YOLOv5 model, we can detect the different kinds of cars, in packages and with corresponding labels.

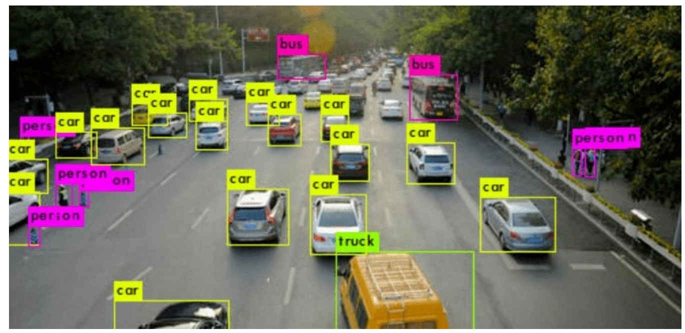


Fig. 2. Result of vehicle detection system

B. Signal Updating

The servers at junctions will receive images from the Closed-circuit- circuit television and process them to calculate the traffic volume. The traffic surveillance module's dynamic traffic data is used by the Signal Shifting Technique to set the green signal schedule and update other lights' red signal timers. The method uses previously observed vehicle data that is provided in JSON format. Following that, the cumulative number of automobiles in each class is determined using this data. The program then identifies the lane with the highest car density. The data is subsequently transferred to the server after being located in the more crowded lane. The computer then extends the signal's green light by a few seconds where a high vehicular density was observed.

The default timing is established for the primary signal of the very first loop when the program is initially executed. Two separate lines are activated here, one to handle vehicle detection for each lane, and another to lever the timer of the existing signal. While the green signal timer is running for the current signal, the system takes a snapshot of the next lane and the red-light time for the other lights, and the green-light time for the next signal, respectively. As the timer and the vehicle detection are handled by two different lines, the timer does not interfere. When, the green light time of the existing signal becomes zero, the timer for the next signal starts. The cycle of counting timers and detecting vehicles keep going in the same manner for each signal.

The green light duration for each signal can be measured by this:

Here, Gdur is the duration of green signal time The number of vehicles is the total number of cars of each type detected by the vehicle detection system Average time to cross intersection = Distance⁄(Average speed of the vehicle) The number of lanes is the total number of tracks available in the intersection

The typical speed of vehicles will be measured in advance and will be stored on the server. Depending on the class of vehicles, the speed will be different. Using the vehicle detection system, we will detect the type of vehicles as well, and according to that data, we will use the average speed values to measure the average time to cross the intersection. The signals do not shift in the direction with the highest density initially; instead, they do so in cycles. This is in line with the existing structure, which sees the signals convert to green sequentially in a predictable fashion without requiring inhabitants to adjust their behavior or create any perplexity. The yellow signs have also been taken into account, and the signal sequence is the same as it is in the existing system.

After the signal turns green, the picture is snapped after some time. This provides the structure a period to decode the figure, compute the number of automobiles in every class present in the figure, determine the time for the green light, and then set the timers for both the existing signal and the next signal's red signal based on that calculation. The best green light duration based on the number of cars of each type at a signal was determined using the bandwidth of automobiles at starting and their sprint times. From that, an estimated of the typical time each type of vehicle requires to cross a junction was found.

$$G_{dur} = \frac{\sum_{vehicletype}(TotalVehicles * AverageCrossingTime)}{TotalLanes + 1} \quad (1)$$

Here,

G_{dur} is the duration of the green signal time

TotalVehicles is the total number of cars of each type detected by the vehicle detection system

AverageCrossingTime = $\frac{Distance}{AverageVehicleSpeed}$

TotalLanes is the total number of tracks available in the intersection

The typical speed of vehicles will be measured in advance and will be stored on the server. Depending on the class of vehicles, the speed will be different. Using the vehicle detection system, we will detect the type of vehicles as well, and according to that data, we will use the average speed values to measure the average time to cross the intersection. The signals do not shift in the direction with the highest density initially; instead, they do so in cycles. This is in line with the existing structure, which sees the signals convert to green sequentially in a predictable fashion without requiring inhabitants to adjust their behavior or create any perplexity. The yellow signs have also been taken into account, and the signal sequence is the same as it is in the existing system.

After the signal turns green, the picture is snapped after some time. This provides the structure a period to decode the figure, compute the number of automobiles in every class present in the figure, determine the time for the green light, and then set the timers for both the existing signal and the next signal's red signal based on that calculation. The best green light duration based on the number of cars of each type at a signal was determined using the bandwidth of automobiles at starting and their sprint times. From that, an estimated of the typical time each type of vehicle requires to cross a junction was found.

C. Simulation

To show the process we have created two different simulations using Unreal Engine 4 from scratch. One of which shows the current traffic system and the other shows our proposed system. It helps with system visualization and comparison with the current static system. Unreal Engine (UE) is a free-to-use 3D computer graphics game engine developed by Epic Games. The next-gen real-time 3D content and experiences are made possible by Unreal Engine, which is used by game developers and producers across industries. Unreal Engine is a full set of development tools for games, broadcast and live event production, education and simulation, architectural visualization and automotive, television content and linear film generation, and other real-time functions.

A four-way intersection with four traffic lights is present there. A timer that shows how much time is left until a signal turns between green-yellow, yellow-red, and red-green is situated on top of each signal. The quantity of vehicles that have passed through the intersection is also shown next to each light. Different types of vehicles arrive from all directions. To make things even, we had to keep the environment and the traffic flow similar for each simulation. And for simplicity, we have set the speed of each class of vehicles to the average speed of their class.

Factors to consider (same for both simulations): 1. Number of lanes 2. Traffic flow 3. Speed of vehicle 4. Lag suffered by vehicles during startup 5. Similar road, intersection, and the same distance

1) 1st simulation (traditional system): In the traditional system, every phase of the traffic signals uses a fixed timer that does not update or vary in response to the actual congestion on the road. So, one particular direction with heavier traffic is given the same amount of green signal time as the other direction, which has lesser traffic than the other or in some cases, no traffic at all, which is inconvenient.

Fig. 3. Traditional traffic system (1)

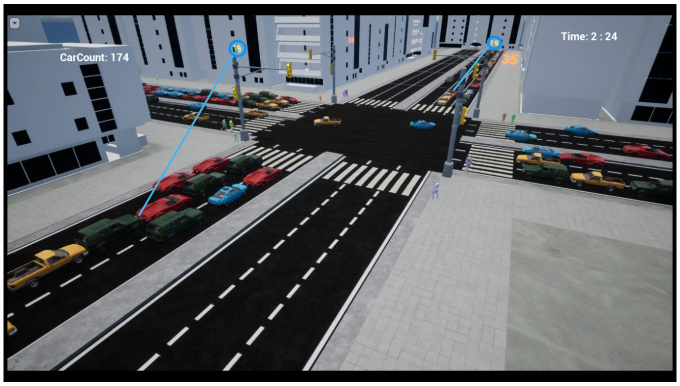


Fig. 4. Traditional traffic system (2)

Fig. 3 and Fig. 4 show the simulation of the traditional traffic system. In this simulation, we will see that, for every signal, the timer is fixed. Every signal is given the same amount of green signal time, in this case, 20 seconds each. But the flow of traffic in each direction is not the same, which doesn't allow proper traffic flow in the intersections as intended.

2) 2nd simulation (proposed system): By automatically calculating the green signal period based on the volume of traffic at the signal, the approach we've suggested tries to mitigate such issues. By doing so, it will be ensured that the route with greater traffic receives a longer green signal duration than the route with a relatively lower volume of traffic.

This simulation shows how our proposed system works. As the traffic flow of each lane is not the same, our proposed system updates the signal based on real-time traffic.

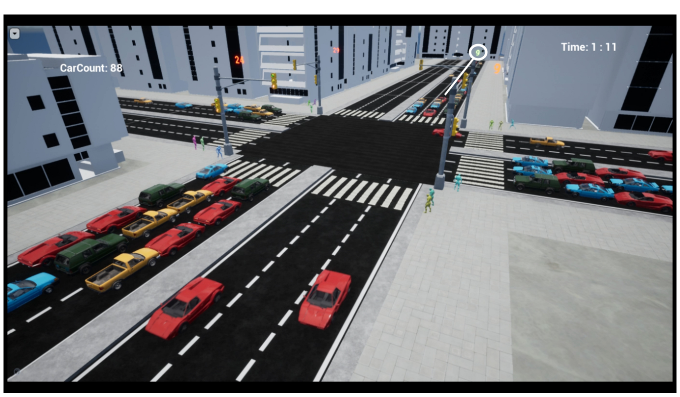


Fig. 5. Proposed traffic system (1)

Fig. 6. Proposed traffic system (2)

Fig. 5 shows the lane with a lower traffic flow is assigned a less amount of green signal time, 10 seconds in this case.

Fig. 6 shows the lane with a higher traffic flow is assigned a large amount of green signal time, 40 seconds in this case. This value can vary from 10 seconds minimum to 70 seconds maximum in our simulation based on the real-time traffic.

IV. Result and Analysis

Using several sample photographs with various numbers of cars, our vehicle detection method was put to the test, and it was found that the detection performance ranged between 65 and 75 percent. Although sufficient, this is not ideal. The absence of a pertinent dataset is the primary cause of the lower accuracy. The system's accuracy can be improved by feeding the model data from actual traffic camera footage.

Two simulations were executed multiple times to assess the proposed method with the established system, and from those simulations, some data was acquired.

In both cases, we have took into account the factors to consider,

- Number of lanes
- Traffic flow
- Speed of vehicle
- Lag suffered by vehicles during startup
- Similar road, intersection, and the same distance

We ran both simulations for an hour and recorded the number of vehicles that passed through the junction over three minutes three-minute periods in each system.

TABLE I. Total Cars Passed in the Traditional System

No	Time	Lane 1	Lane 2	Lane 3	Lane 4	Total
1	3	67	36	48	75	226
2	6	56	44	60	59	219
3	9	76	57	34	56	223
4	12	47	68	63	52	230
5	15	87	31	51	43	212
6	18	42	77	48	55	222
7	21	65	69	46	44	224
8	24	42	71	55	59	227
9	27	33	82	66	33	214
10	30	52	63	61	41	217
11	33	47	67	72	40	226
12	36	79	45	53	49	226
13	39	83	49	51	42	225
14	42	62	72	66	11	211
15	45	93	37	51	33	214
16	48	97	42	25	65	229
17	51	106	32	43	37	218
18	54	56	77	46	38	217
19	57	65	79	53	28	225
20	60	70	75	49	30	224

TABLE II. Total Cars Passed in the Proposed System

No	Time	Lane 1	Lane 2	Lane 3	Lane 4	Total
1	3	87	56	72	63	278
2	6	73	68	66	75	282
3	9	95	55	68	69	287
4	12	61	78	72	64	275
5	15	74	76	60	69	279
6	18	88	63	81	52	284
7	21	97	67	66	58	288
8	24	81	77	52	71	281
9	27	102	64	62	51	279
10	30	107	59	51	58	275
11	33	85	82	58	52	277
12	36	97	58	66	61	282
13	39	77	71	81	56	285
14	42	73	75	71	68	287
15	45	91	70	59	67	287
16	48	80	84	71	46	281
17	51	66	92	72	46	276
18	54	78	75	76	52	281
19	57	107	55	61	57	280
20	60	82	69	60	69	280

TABLE III. Comparison of Passing Cars in 2 Models

No	Time	Total Car Passed(Current)	Total Car Passed(Proposed)	Difference
1	3	226	278	52
2	6	219	282	63
3	9	223	287	64
4	12	230	275	45
5	15	212	279	67
6	18	222	284	62
7	21	224	288	64
8	24	227	281	54
9	27	214	279	65
10	30	217	275	58
11	33	226	277	51
12	36	226	282	56
13	39	225	285	60
14	42	211	287	76
15	45	214	287	73
16	48	229	281	52
17	51	218	276	58
18	54	217	281	64
19	57	225	280	55
20	60	224	280	56
Total		4429	5624	1195

From Table I, we get the number of cars passing through the intersection every 3 minutes for 1 hour in the traditional system.

The result is shown in Fig. 7, where the number of vehicles passing through the intersection is between 210 and 230.

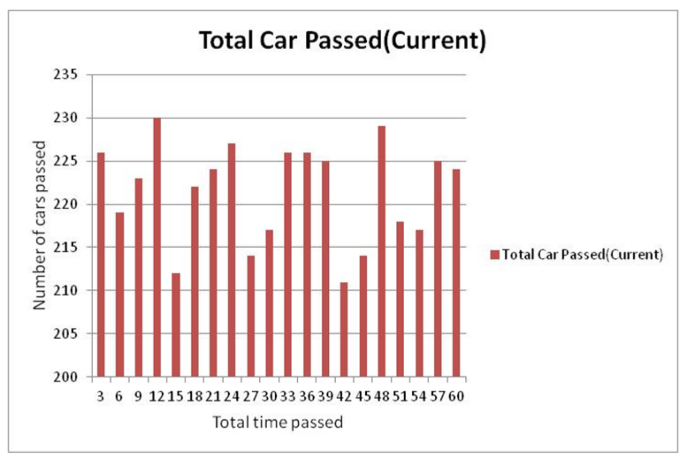


Fig. 7. Number of cars passed in the traditional system

From Table II, we get the number of cars passing through the intersection every 3 minutes for 1 hour in our proposed system.

The result is shown in Fig. 8, where between 270 and 290 automobiles traveled past the crossing.

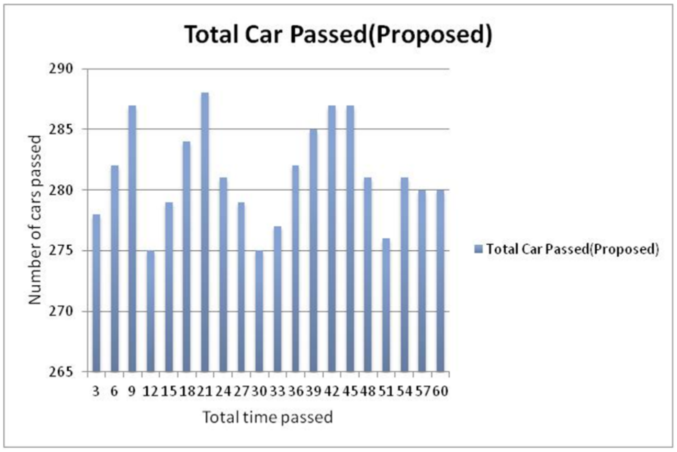


Fig. 8. Number of cars passed in the proposed system

From Table III, we can see the comparison of the result of the simulation we have run for both systems. The result, in this case, is quite satisfactory.

The result is shown in Fig. 9, where 40-80 more vehicles are being passed through the intersection in our proposed system compared to the traditional traffic system.

Fig. 9. Comparison of the number of cars passed in both systems

Here are 20 different results of 1 hour of simulation of both systems. Every 3 minutes, the traditional system lets 210-230 vehicles pass through the intersection whereas our proposed system is letting 270-290 vehicles pass through the intersection, which is a noticeable improvement.

TABLE IV. Comparison between 2 Models

Parameters	Traditional system	Proposed system
Average cars passed per minute	74	94
Total cars passed in 1 hour	4429	5624

By comparing the simulations while maintaining the same set of conditions, the outcome of Table IV was obtained.

As we can see, keeping all the conditions alike, the current system was able to pass 4429 vehicles total in 1 hour, whereas our proposed system was able to pass 5624 vehicles in same amount of time, which is 1195 more vehicles than current system.

Thus, our proposed system will increase the performance by 26.98126%

V. Conclusion

In conclusion, the suggested program guarantees that, by adaptively modifying the green signal timing according to the density at the light, the route with more vehicles is given a green light time for a longer period in comparison to the route with less traffic. Due to fewer unnecessary delays, less traffic, and shorter wait times, there will be less fuel use and pollution.

In regards to the number of vehicles passing the intersection, the system performs around 23 percent better than the current method, which is a substantial improvement. With more calibration and authentic surveillance data for the training set, this system has the potential to perform even more effectively. Additionally, the suggested method has some benefits over other intelligent traffic management systems that are already used, like Pressure Pads and Infrared Sensors. The system may be installed for a very low cost because it uses CCTV footage from traffic signal cameras and typically doesn't need any new hardware because these cameras are already present at intersections with considerable traffic.

There might only be a small amount of alignment required. Maintenance expenses are also decreased when compared to other traffic enforcement methods like "pressure mats", which frequently face stress for their installation on roadways where these are continuously imposed to tremendous stress. To allow improved traffic control, the proposed method can therefore be connected with Surveillance cameras in large cities

VI. Future Work

To improve traffic management and reduce congestion, the project could be extended further to include features like these:

1) Detecting violations of traffic rules: Vehicles that are breaking traffic rules can be identified easily by video streaming and capturing images of vehicles with number plates. It is feasible to recognize the automobiles that are skipping red lights in the picture or video stream by establishing a transgression barrier and recording the plate number of the image and whether that line has been crossed when the light is red. Similarly, lane changing of vehicles might also be detected. These can be accomplished through image processing methods or backdrop removal.

2) Detecting violence or accidents on road: Accidents can be detected easily and necessary steps can be taken quickly. Intersections commonly see catastrophic collisions as a result of the presence of numerous hazardous collapses, like crashes due to angular and turning left. To safeguard property and lives while simultaneously minimizing traffic and delays, it is crucial to accurately and quickly identify accidents at crossings. The list of automobiles that remain immobile for a lengthy amount of time in an inhospitable area, such as the center of the road, might be modified to exclude parked cars from this restriction.

Also if any occurrence happens by any vehicle, it can be detected and tracked. Also, the probability of the vehicle taking a specific route can be predicted in advance.

3) Detecting emergency vehicles: Vehicles like ambulances or fire service can be detected and the timer can be adapted accordingly to make a quick passage for the emergency vehicle. The model can be programmed to recognize emergency vehicles in addition to other types of cars, allowing it to adjust timers to give them precedence and enable them to cross the signal as soon as possible.

REFERENCES

[1] Khushi, "Smart Control of Traffic Light System using Image Processing", 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), 2017, PP. 99–103.

[2] Soman, Renjith and Radhakrishnan, K, "Traffic Light Control and Violation Detection Using Image Processing", Traffic, vol. 8, No. 4, 2018.

[3] Chabchoub, Abdelkader and Hamouda, Ali and Al-Ahmadi, Saleh and Cherif, Adnen, "Intelligent traffic light controller using fuzzy logic and image processing", International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, No. 4, 2021.

[4] Suhweil, Yazan and Al Yaman, Musa and others, "Smart controlling for traffic light time",2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), IEEE, 2017, PP. 1–5.

[5] Al-Majhad, Hailah Ghanem and Bramantoro, Arif and Syamsuddin, Irfan and Yunianta, Arda and Basori, Ahmad Hoirul and Prabuwono, Anton Satria and Barukab, Omar M, "Intelligent traffic light controller using fuzzy logic and image processing", International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, No. 4, 2018.

[6] Agrawal, Aditi and Paulus, Rajeev, "Smart Intersection Design for Traffic, Pedestrian and Emergency Transit Clearance using Fuzzy Inference System", International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, No. 3, 2021.

[7] Adbeb, Tesfanesh and Wu, Di and Ibrar, Muhammad, "Software-defined networking (SDN) based VANET architecture: Mitigation of traffic congestion", International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, No. 3, 2020.

[8] Shinde, Ms Saili and Jagtap, SR, "Intelligent Traffic Management Systems", A Review, vol. 2, No. 9, 2016.

[9] Navarathna, Pramathi J and Malagi, Vindhya P, "Artificial intelligence in smart city analysis",2018 International conference on smart systems and inventive technology (ICSSIT), IEEE, 2018, PP. 44–47.

[10] de Frías, Carlos Justo and Al-Kaff, Abdulla and Moreno, Francisco Miguel and Madridano, Ángel and Armingol, José María, "Artificial intelligence in smart city analysis",2020 IEEE Intelligent Vehicles Symposium (IV), IEEE,vol. 10, 2020, PP. 33–38.

[11] Manjón Prado, Javier, "Using Unreal Engine as an engineering tool for traffic simulation and analysis",Universitat Politècnica de Catalunya, vol. 10, 2020.

[12] Karimi, Abolfazl and Duggirala, Parasara Sridhar, "Formalizing traffic rules for uncontrolled intersections",2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS), IEEE, 2020, PP. 41–50.

[13] Golubev, Kirill and Zagarskikh, Aleksandr and Karsakov, Andrey, "A framework for a multi-agent traffic simulation using combined behavioural models", Procedia Computer Science, Elsevier, vol. 136, 2018, PP. 443–452.

Emotion Detection from Text and Sentiment Analysis of Ukraine Russia War using Machine Learning Technique

Abdullah Al Maruf[1], Zakaria Masud Ziyad[2], Md. Mahmudul Haque[3], Fahima Khanam[4]
Department of Computer Science and Engineering, Bangladesh University of Business and Technology
Mirpur, Dhaka, 1216, Bangladesh

***Abstract*—In the human body, emotion plays a critical function. Emotion is the most significant subject in human-machine interaction. In economic contexts, emotion detection is equally essential. Emotion detection is crucial in making any decision. Several approaches were explored to determine emotion in text. People increasingly use social media to share their views, and researchers strive to decipher emotions from this medium. There has been some work on emotion detection from the text and sentiment analysis. Although some work has been done in which emotion has been recognized, there are many things to improve. There is not much work to detect racism and analysis sentiment on Ukraine -Russia war. We suggested a unique technique in which emotion is identified, and the sentiment is analyzed. We utilized Twitter data to analyze the sentiment of the Ukraine-Russia war. Our system performs better than prior work. The study increases the accuracy of detecting emotion. To identify emotion and racism, we used classical machine learning and the ensemble method. An unsupervised approach and NLP modules were used to analyze sentiment. The goal of the study is to detect emotion and racism and also analyze the sentiment.**

***Keywords*—Emotion detection; racism; sentiment analysis; social media; machine learning; ensemble; Ukraine-Russia**

I. Introduction

Emotion is a strong feeling caused by one's circumstances and conditions, moods, interpersonal connections, or pleasure and dissatisfaction. The emotional experience includes perceptions of the world, cognitive capacities, behavioral reactions, metabolic anomalies, and instrumental activity. Emotions are difficult to define because they are a fleeting state of mind. Images, speech, facial expressions, textual information, emoticons, and other kinds of expression may all be used to determine emotion. Textual data is essential for research [1]. Massive amounts of text-based data have been generated regularly in recent years via social media and conversations such as messenger, Whatsapp, Twitter, and other means [2]. The progression of digital communications and its popularity, particularly virtual networking, keeps individuals interested in how they connect and communicate amongst themselves. People have become accustomed to expressing feelings lightly and intuitively through social media communication and the simplicity of responses. People use social media to keep up with what's happening in the world and to share their opinion and feedback via likes, comments, and shares, among other things [3]. Today's most popular social media platforms are Facebook, Twitter, and Instagram. People visit Facebook to keep in touch with friends, family, and loved ones, learn about what's happening worldwide, and express what's important to them, according to Facebook's vision and mission. People incline to be more verbose on Facebook, yet posts go through more simple "likes" than lengthy comments. Since February 2017, Facebook has included additional capabilities that allow users to express their specific feelings in reaction to a post, such as the ability to mention "love" or "sadness" instead of liking a post [4]. People use Twitter to express their thoughts, feelings, and views through short messages or tweets at any time. Individuals' emotional states of mind, such as joy, worry, and hopelessness, are captured overtly or indirectly within those short messages along with bigger communities, such as the viewpoints of people in a particular country [5][6]. There are five ways to emotion identification from text, including keyword-based, lexical/corpus-based, learning-based, hybrid-based, and deep learning-based approaches, but each has its limitations[7]. Specifically, identifying object words from tweets is named object-oriented feature to perform sentiment—Bi-gram, uni-gram model with object-oriented feature effective better [8]. Emotion was discovered and recognized using machine learning and deep learning methods. There are also several classifications for identifying emotions; some of them are the k-nearest neighbor(KNN) algorithm, Support vector machine(SVM), decision tree(DT), random forest(RF), linear discriminant analysis(LDA), etc. DT is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, KNN is a method for classifying objects based on closest training examples in the feature space, LDA is a method used in statistics, pattern recognition, and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events and SVM analyze data with recognizing patterns used for classification and regression analysis. Based on the lexical approach, Real-time emotional analysis was carried out, and the data was obtained from online social media[9]. Sentiment analysis may be performed using a deep learning model that has been pre-trained, as well as unsupervised algorithms such as Valence Aware Dictionary and sEntiment Reasoner(VADER), textblob, and k-means clustering. It's also possible to gauge sentiment using lexicons. The sentiment was analyzed using large-scale data from Twitter and a machine learning-based technique. Sentiwordnet and sentiment are sentic computing-based public lexicons [10]. Emotion is identified from audio sources following machine learning techniques and categorized into six basic emotions. Auto weka performed the best outrun compared to SVM, KNN, and multi-layer perceptron (MLP) [11]. Finding out mental instability RF shows (87%) accurate results from the manually surveyed datasets [12].The basic facts are that domain adaptability and accuracy are the main constraints for emotion recognition from text. The advantages of the suggested approach are an improvement in emotion detection accuracy. The dataset was compiled from a number of sources. The dataset will aid researchers in overcoming this industry's problem of domain adaptation. The research project also analyzes the war between Russia and Ukraine's tweets for mood. The topic has fewer datasets that are readily available. Therefore, gathering a dataset is the main contribution to this field. The desired model to address accuracy issues is ensemble model learning. The main goal of the research was to increase accuracy, thus we utilized a variety of ensemble techniques. We proposed a unique technique where ensemble and classical machine learning are applied. We also used an unsupervised approach to analyze the sentiment. In our work, we will show an approach that detects emotion, analyzes sentiment, and detects racism. We used various methods and techniques to identify emotion, racism, and sentiment analysis. To increase the speed and quality of the learning process, we applied machine learning and ensemble method using parameters. The following is the structure of this paper: Some recent studies in this field are addressed in Section 2. Section 3 details our study technique. Section 4 contains the results, and We'll talk about our strengths and weaknesses and provide suggestions for the future. Then, we eventually finished the paper.

II. Related Work

According to a case study [12], a 3-turn dialogue detects and distinguishes emotions. In three turns, it contains a collection of human emotions on Twitter. They also talk about the Amiens system, which detects emotions in text messages. The algorithm applied the Long short-term memory (LSTM) model to recognize human emotions based on the in-depth reading. The system's primary input is a mix of word2vec and doc2vec embedding. Then, for that issue, utilize the most current Bi-Long Memory Short Term category, which used word embedding as input and predicted human emotion. When the Amiens score is 0.7185, the output results reveal considerable variations in f points above the model's base. As a potential future project, they plan to expand hybrid approaches through emotion handling and emotion lexicons management. A case study showed the interaction of emotion by applying hybrid and machine learning methods for six basic emotion categories. A comparison study of speaker-independent and dependent recognition was represented [13]. As different formats of emotion detection systems exist, the text identification system is one of them. A machine learning-based automated system was implemented to understand a textual form of data to analyze its class. SVM was performed with a 63.5% accuracy rate in the following system of automatic text identification [14]. Linguistic-related emotion identification has some complexity, and little research has been done on it. For every individual language, the data preprocessing step becomes the most challenging part. Preprocessing steps followed tokenization, segmentation, and other extraction methods to filter raw data for better applicability. A system was implemented to perform a hybrid approach to classify six basic emotions for Punjabi words [15]. Urdu language-based work focused on commercial-related emotion detection [16]. The algorithm support vector classifier(SVC), KNN, RF, and Naive Bayes(NB) was applied to datasets (Smartphone and Sports), and SVC performed best among them with accuracy above 80%. NB achieved the goal of detecting Bangla text-related emotion. As a complex linguistic analysis, three emotions category presented in this process [17]. Facial expression means a lot of things about the mentality of a person. The machine learning-based method SVM algorithm was used to detect the actual emotion of a person from an image. Framework and UI environment constructed in their system [18]. In the present world, the most used handheld electronic device is a smartphone. A self-automated system named "iself" is represented in their work where it can detect the user's emotions from their smartphone data. The system's structure describes how emotion detection happened by analyzing smartphone users' data [19]. Social media is a great source for mass data collection. Five preprocessing techniques were used for Arabic language emotion detection contextual research. The minimal sequential optimization(SMO) classifier performed better than NB and SF in classifying six basic emotions distinguished from the text [20]. Emoticons can be an expression of emotion. The study reviewed the Arabic tweets and labeled them into the four emotion categories. Testing was performed using SVM and MNB algorithms for developing the system[21].Deep learning techniques were used for detecting emotion from the Persian text document. Preprocessing technique word2vec was used for normalizing the dataset. NB, DT, and SVM algorithms were executed to classify emotion, and 10-fold cross-validation was performed to evaluate the performance. The SVM algorithm scored the highest accuracy in their system [22]. A system overviewed in the research is SEDAT, which can detect emotion from Arabic social media tweets in real-time. Convolutional Neural Network(CNN)-LSTM neural network technique was applied to construct the system. Their work shows that the system is better accurate than the TeamUNCC system [23]. Arabic language-based sentiment analysis work combined three Arabic Steemers (ISRI, Light Steemer, snowball). The research outcome exhibits a comparison between deep learning and machine learning where CNN, SVM, NB, and MLP are exerted in implementation [24]. Emotion classification and social media are almost connected in parallel cause it is a rich data sources. In this case, comparative model analysis on different datasets is highlighted. TME model acquires maximum accuracy than other models, and WME remains close to TME [25]. From youtube users' text comments, emotion identification was conducted in this research. Unsupervised machine learning technique executed to determine emotion category from YouTube comment data corpus.The system achieved an average of 92.75% precision, close to another existing system that applied the SVM algorithm[26]. The microblog-based emotion classification task is represented by using deep learning techniques. CNN model was selected for this purpose, though the experiment was held in four phases. Sina Weibo was the data source for the experiment. This study exposed CNN as the maximum accurate (97.60%) for detecting emotion from microblogs [27]. Another research is a deep learning-based emotion recognition system. It also demonstrates how excellent deep learning is at detecting text emotion. Data preparation includes sentence segmentation and word embedding. SVM and LSTM are employed to categorize seven different emotion groups. Among these, LSTM has a higher efficiency (94.7%) than SVM [28]. Most of the existing sentiment analysis and emotion detection research is performed on deep learn ing or machine learning-based unique techniques. This study presents a system that converts multi-label classification problems into a single binary classification and then solves the problem by applying the deep learning method. The proposed system model Binary Neural Network (BNET) shows the best accuracy in evaluation measured in (multi-label accuracy) Jaccard [29]. Although several textual-based emotion recognition systems exist, the researcher concentrated on discovering emotional states from poetry text in this study. The suggested method is capable of detecting 13 different emotion classes. The system recognizes the context and value of tokens. Comparative investigation demonstrates that the suggested model outperforms the CNN-BiLSTM model in this system[30]. Emotion-based research is becoming more popular since computer machines can predict likely outcomes. Several research and records used a variety of procedures and strategies to identify particular human emotions. Previous work had various drawbacks; thus, the suggested system employed the layered LSTM approach to identify emotion. The suggested system outperformed the LSTM and SVM approaches, reaching a 99.2% accuracy rate [31]. Another research represents EmoDet2, a system that innovates deep learning architecture for identifying emotion from the text.The suggested system used two different algorithms: BERT and BiLSTM. EmoDet2 has been created by applying the ensemble method. The system F1 score (0.75) outperforms the baseline model, which was encouraging. SEMEVAL-2019 dataset had the best accuracy in recognizing emotion from the text, allowing the EmoDet2 system to outperform the baseline model[32]. A suggested method was given in another study that used a machine-learning approach to identify emotions. The dataset is divided into two types, one binary, and the other multi-class. The system was programmed with seven machine-learning models for the experiment. The sentiment is divided into two categories: joyful and unpleasant. The LR-SGD and the tf-idf classifier produced ideal results in a comparison study. The voting classifier beat all others, obtaining 84 percent accuracy using both TF and tf-idf[33]. According to a study by [34], social media is getting a lot of attention in today's world. Public and private opinions on various topics are constantly expressed and distributed via social media platforms. One of the most prominent social media platforms is Twitter. Twitter is a social media platform that allows buddies, relatives, and colleagues to interact and keep in touch by exchanging short, frequent messages. Twitter is the main microblogging website that allows users to post status updates (known as" tweets"). These tweets occasionally reflect thoughts on a variety of issues [35]. Twitter is a real-time microblogging platform that allows individuals or groups to express their opinions on a topic and have them appear on a timeline. Web search apps and real-world applications such as current global trends and world events, and extracting the most up-to-date information regarding occurrences use microblog data for analysis and conclusion-making. Sentiment Analysis and Opinion Mining are two types of text mining that involve the analysis of sentiments, opinions, and emotions and the assessment of the text's content. When evaluating people's thoughts, feelings, assessments, attitudes, and responses to services, goods, organizations, personalities, events, themes, and issues, and their qualities, sentiment analysis is another name for opinion mining. Sentiment and subjectivity are strongly influenced by the context and domain in which they occur. It is not only due to language changes but also to the dual meaning of feelings of the same term in various domains. The processes of extracting nontrivial patterns and intriguing information from unstructured script texts are called opinion mining and sentiment analysis, respectively [36]. Before the internet, it took a long time for information about a company's stock price, direction, and general attitudes to spread among individuals. Web technology has ushered in a new era of rapid information transmission and retrieval. Applying positive or negative information about a company, product, person, or other entity may be as simple as clicking a mouse or utilizing microblogging services like Twitter [37]. Social media platforms have grown in importance as a forum for political debate worldwide. Users may use Twitter to send tweets, which are short communications of up to 140 characters. The number of people using Twitter is steadily increasing. Over 100 million active users throughout the world send over 250 million tweets every day, according to the business. In several fields, including the financial market, politics, and social movements, sentiment in Twitter data has been utilized for prediction or assessment. Emerging events or news are frequently followed by a surge in Twitter activity, offering a unique chance to assess the relationship between stated public mood and political outcomes. Furthermore, sentiment analysis may be used to investigate how these occurrences impact public perception. It provides a fresh and contemporary perspective on the dynamics of the election process and public opinion to the general public, the media, lawmakers, and academics. They present a method for real-time analysis of popular sentiment toward presidential candidates, as stated on Twitter during the 2012 U.S. election [38]. Many readers value social media platforms such as Twitter and Facebook because they enable users to easily discuss and express their thoughts on various issues and send messages worldwide. Every day, millions of quick messages (tweets) are shipped on Twitter, making it one of the most prominent microblogging and social networking services. The challenge of assessing users' tweets in terms of emotions, ideas, and viewpoints in a wide range of activities and domains is addressed by Twitter sentiment analysis [39]. Twitter sentiment analysis is a tricky issue, even though sentiment analysis has recently garnered significant popularity in various fields. Businesses use sentiment analysis to look at customer reviews of their services, while the government and other organizations use it to observe public health and forecast political trends, among other things. Manual procedures were frequently used for this before the advent of social networks. Manual techniques were frequently used for this before the beginning of social networks [40]. Customers who want to buy a quality product based on user reviews at reliable online retailers can take advantage of the service offered by the Twitter Sentiment Analysis for Business Project. Additionally, it benefits commercial enterprises that wish to accelerate their firm's growth by revamping goods or services following customer demands and preferences [41]. One of the most crucial tasks in text mining is automatic sentiment analysis of texts, which aims to establish if a given document has a positive, negative, or neutral attitude. Nowadays, It has gotten a lot of attention because opinion mining on microblogging sites is widespread. Both public and commercial companies increasingly rely on the ability to determine whether a written document has a favorable or unfavorable impression [42]. SVM, LR, NB, and KNN are the techniques employed for sentiment analysis of how people feel about moving the nation's capital. As measured by the performance evaluation algorithm's accuracy, precision, recall, and F-measure findings, the support vector machine fared better than the other three algorithms. The administration will receive an overview of public opinion from the perspective of data from social media through sentiment analysis of the dialogue surrounding the relocation of the nation's capital [43]. With the development of Web 2.0 and the rising popularity of social media platforms like Face book, Twitter, and Google+, users can now exchange information and, as a result, have a say in the material published on these platforms[44]. Many analysts, businesspeople, and politicians use blogs, microblogs, social networks, and other types of websites as a massive source of information to grow their businesses by taking advantage of the copious amounts of text produced by users who provide ongoing feedback on the prominence of a particular subject through emotions, viewpoints, and feedback. The last ten years have seen a growth in sentiment analysis methods, mainly when applied to tweets, in both the academic and industrial areas [45]. Twitter sentiment analysis is a useful tool for various activities using Twitter-based analysis. Twitter sentiment analysis using attentional-graph neural networks is done by the Attentional-graph Neural Network-based Twitter Sentiment Analyzer (AGN-TSA). A three-layered neural network used by AGN-TSA combines information from the tweet's text and its user connections [46]. Recurrent neural networks (RNNs) are also used to examine sentiment in tweets. The technique can categorize tweets with an accuracy rate of 80.74% while considering a binary task, following testing 20 various design strategies [47]. Furthermore, A brand-new unsupervised learning framework built on Concept-based and Hierarchical Clustering is suggested for Twitter Sentiment Analysis. Serial ensembles combine common hierarchical clustering techniques, including single linkage, complete linkage, and average linkage algorithms. Additionally, TF-IDF performs better than the Boolean method compared to other feature representation techniques also examined [48]. One of the most intriguing study areas is sentiment analysis from Twitter [49]offered a comparative emotive analysis from a different linguistic standpoint. Several models, such as PLMs, RoBERTa, and BERT, were used to evaluate each language separately. For fine-tuning, four Nigerian languages were pooled in the training dataset. Among the models tested in the proposed system, the AfriBERTa model fared the best. At present, politics and virality in social media is a common phenomenon. This study [50] overviewed sentimental analysis from Twitter in Greece, Spain, and the United Kingdom. The dataset originated from tweets of parliament members and politicians. The experimental models are distinguished into two types multilingual and Monolingual. A comparative analysis was conducted between these regions. Another study [51] used NB Classifier to determine the tweet's word semantic analysis. The proposed system collects data from live-fetched twitter according to user input. After analyzing the stored user data, the visual heat is represented through Google Maps. There is a variety of methods that may be used to analyze sentiment. Both automatic and hybrid approaches were used for the experiment. Using Deep learning [52] performed, the analysis of sentiment on stemmed Turkish Twitter user data. Twitter API was used to create the dataset. The data was preprocessed by eliminating tokens, id numbers, and punctuation marks, among other things. The training dataset was followed by three distinct techniques: shift, shuffle, and hybrid. As DL algorithms, CNN, RNN, and Han architecture were used. Compared to previous systems, the suggested system can acquire a modest advantage. The research [53]presented a one-of-a-kind effort on sentiment inspection based on online meal delivery. Food reviews are well-known for their commercial worth. It isn't just about Twitter users; it's also about evaluating business competitors. A comparative analysis was done on Twitter user data connected to food delivery to categorize users' attitudes and customer reviews regarding the organization. Lexicon-based categorization was used to differentiate sentiment polarity.

The literature focuses on the research's shortcomings, gaps, and advancements. A lot of studies have been done on the broad topic of emotion detection from text. But something needs to be improved. Accuracy is one among them. Therefore, the suggested model is created to increase accuracy. There isn't much work left over from the Russia-Ukraine war. This study looks into how people feel about going to war. The research also examined racist behavior. Since we lacked a labeled dataset, we used the Python library to evaluate the data and used unsupervised learning to determine sentiment. Racism originating from the war is detected via semi-supervised learning. The literature demonstrates that sentiment varies according to the domain.

III. Methodology

Social media sites like Facebook detect emotion. Another dataset was obtained from Kaggle and was used to identify racism. Finally, we collected a dataset from Twitter. The Twitter dataset is about the conflict between Russia and Ukraine. This dataset was used to identify racism and assess public opinion on the Russia-Ukraine conflict. The datasets' source and application are given in Table I. Overview of the entire system is shown in Fig. 1. As previously stated, dataset 1 was acquired via social media and manually classified. After that the dataset is divided into two parts. During the training phase, 18000 data were trained, whereas 3000 data were tested. Fig. 2 depicts the number of classes in dataset 1. Fig. 3 illustrates dataset 2, where 0 denotes a non-racist speech, and one represents a racist utterance. Despite the recent and rising interest in utilizing Twitter to study human behavior and attitudes, the capacity to use Twitter data for social science research and scientific research still has a long way to go[54]. Twitter provides an API, and by using the API, data can be collected for research [55]. We used Twitter API to collect tweets about the Ukraine-Russia crisis. Our dataset 3 had a total of 16,208 records. We used some keywords to find the exact data we needed. Fig. 4 depicts the stages involved in gathering data from Twitter. The ensemble approach is used to address the challenges with emotion detection accuracy. The ensemble technique combines a variety of models. The accuracy rises when the ensemble approach is used. First, we used the suggested model to analyze our dataset, and the accuracy was higher than with the old approach. Later, we applied the model to additional datasets used by academics. The accuracy of the model greatly increases whenever we apply it to new datasets.

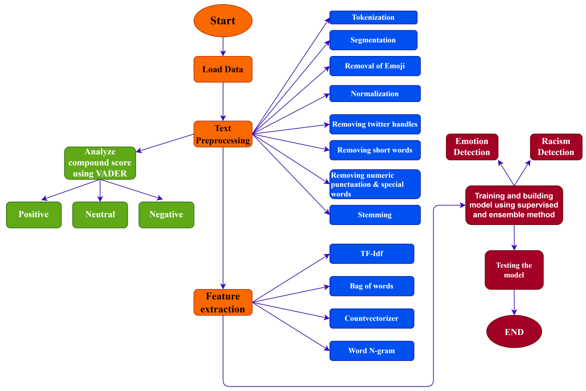


Fig. 1. Overview of the entire system

TABLE I. DATASET SUMMARY

Dataset	Application	Source
Dataset 1	Detect Emotion	Social media(self made)
Dataset 2	Rasicm Detection	Kaggle
Dataset 3	Sentiment Analaysis Racism Analysis	Twitter(Self made)

A. *Text Representation*

1) Text preprocessing: Text preprocessing prepares text data for machines to use for analysis and prediction. Data preprocessing is an essential aspect of machine learning. The algorithms and models need to forecast the result accurately. Besides lowering the extracted feature space, preprocessing can increase classification accuracy [56]. The neat text and some NLP packages are used to text preprocess. The text prprocessing technique applied in our work:

Removing Numeric, punctuation, and special character: We deleted any superfluous punctuation, digits, or symbols to enhance the dataset and replaced them with space. Also, to make the dataset more algorithmically implementable.

Removing Twitter Handles (@user): A was created to eliminate the undesirable pattern of text from the tweets. It takes two arguments, the original and refined strings, which remove the text patterns. The method will give the desired string after removing unwanted text patterns in the output. To withdraw the design from our data, we used the function.

Removing short words: Short and useless words are deleted in this stage. Words with a length of three or fewer were eliminated from this list. As like hmm, oh, and so forth.

Removal of emoji,URL and hashtag: We must remove emojis from text to gain accurate output in our work. Therefore, we also eliminated URLs and hashtags from our text.

Segmentation: Text segmentation is the technique of splitting written material into valuable components, including words, sentences, or themes. We applied the segmentation technique. First, any kinds of commas, dots, and hyphens were deleted.

Tokenization: The tokenizer breaks up the input text into little chunks known as tokens. For example, if elements of a word are more prevalent than the word itself, there can be more tokens than words.

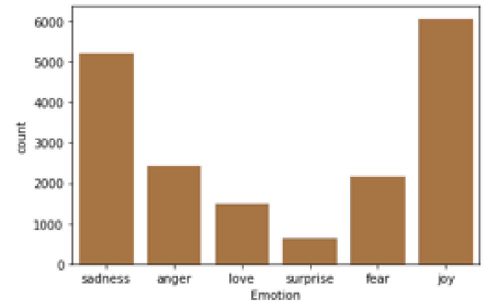


Fig. 2. Number of each class in the dataset emotions

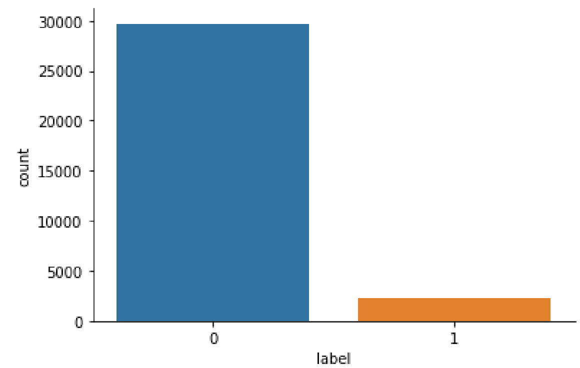


Fig. 3. Number of class in the dataset of racism detection

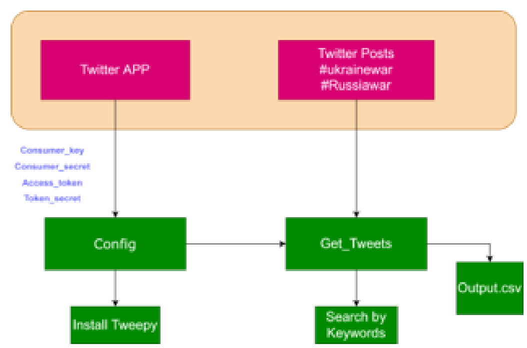


Fig. 4. Process of data collection from Twitter

Streaming: Data streaming is sending a continuous stream of data to extract useful information. A data stream is a collection of pieces arranged in time. Stemming is the process of removing suffixes (",", "ly", "es", "s", and so on) from a word using a set of rules. For Example, play, playing, and player, are the different variations of the word – "play".

Lemmatization: It seems the term yet ensures that it retains its meaning. Lemmatization has a pre-defined vocabulary that records the context of words and checks the word in the dictionary while decreasing the term's length. Twitter users occasionally utilize lengthy words, such as "loooovvveee, greeeeat", when they purposely type or add more letters that

TABLE II. EXAMPLE OF BAG-OF-WORDS FEATURE

She	He	Good	Girl	Will	Guy	
1	1	2	1	0	0	B1
0	0	1	0	1	1	B2

repeat more frequently. We replace the long and needless character

2) *Feature extraction:* The most crucial phase in detecting emotion is feature selection, which influences the task's overall outcome. Therefore, it is critical to choose features carefully to improve classification because better feature selection leads to accurate prediction. We used various features to analyze our procedure data after completing the pre-processing step. We applied several feature extraction techniques to find the suitable one for our proposed system. Combining several strategies to achieve the best potential result is beneficial [17]. Feature extraction techniques applied:

Bag-of-Words Features: An approach for extracting characteristics from text documents is a Bag of Words. Machine learning algorithms may be trained using these features. It develops a vocabulary of all the unique terms in the training sets. Take a corpus, for example, shown in Table II (a textual compilation) named A comprising B documents b1, b2, ... bB, with X unique tokens retrieved from the corpus A. The list of X tokens (words) will be formed, and B X X will determine the size of the bag-of-words matrix Y. In document B(i), the frequency of the token is represented in each row of the matrix Y. For instance, suppose we have two documents-

B1: She is a good girl. He is also good.

B2: Will is a good guy.

It starts by building a vocabulary from the documents' unique terms. ["She", "He", "good", "girl", "Will", "guy"] Here, B=2, X=6. The 2 X 6 matrix Y will be expressed as follows:

The scenario above shows the training characteristics, which include the frequency of each word on each page. This's strategy known as the bag-of-words approach since it's based on the number of occurrences rather than the sequence or order of words.

Word n-grams: Several component evaluation approaches are available for AI-based, more aesthetic implementation, such as n-gram, tf-idf, count vectorizer, and word integration[57]. The n-grams feature combines n consecutive words or characters that are considered beneficial for categorizing texts. Here, we looked at the performance of unigram and bi-gram to acquire the optimum model. To achieve the entire situation, we integrated the bi-gram and tri-gram features, providing a substantially superior outcome in our study. We also observed scikit-learn to explore features [58]. A series of N words or characters is referred to as an n-gram. Consider the following sentence: **"Emotion detection from text".**
Unigram model(n=1)
["Emotion", "detection", "from", "text"]
Bigram model(n=2)
["Emotion detection", "detection from", "from text"]

In natural language processing, n-grams are an essential but fundamental notion. N-grams play a large part in our system since so many applications need to extract insights from text. The n-gram parameter's range affects the outcome.

Tf-idf Vectorizer: The term frequency multiplied by the inverse document frequency is known as tf-idf. The tf-idf vectorizer [59], which evaluates the presence of a word in a document rather than using only raw numbers, is an excellent example of a suitable input representation [60]. The term frequency refers to the number of times a specific word occurs in a conferred document. On the other hand, Inverse document frequency considers all that include that term. **Term Frequency (TF):** First, let's define the term "frequent" (TF). It is a metric for the number of times a phrase, t, appears in a document, d:

$$tf_{t,d} = \frac{n_{t,d}}{Numberoftermsinthedocuments} \tag{1}$$

Here, The number of times the phrase "t" appears in the document "d" is represented by n in the numerator. As a result, each paper and word would be assigned a unique TF value. We used the same vocabulary as in the Bag-of-Words model to demonstrate how to compute the TF for B2. B2: Will is a good guy.

Here,

- Vocabulary: 'She', 'He', 'good', 'girl', 'Will', 'guy'
- Number of words in B2 = 5
- TF for the word 'Will' by equation (1)= 1/5 Similarly,
- TF('She',B2) = 0/5
- TF('He',B2) = 0/5
- TF('good',B2) = 1/8
- TF('girl',B2) = 0/8
- TF('guy',B2) = 1/8

Inverse Document Frequency (IDF): The IDF is a metric that determines how essential a phrase is. Therefore, the IDF value is required since simply computing the TF is insufficient to comprehend the significance of words.

$$idf_t = log\frac{numberofdocuments}{numberofdocumentsnumberofdocumentswithterm``t"} \tag{2}$$

he IDF values for all the words in B2 are calculated by equation 2:

IDF('Will') = = log(2/1)

TF-IDF: The TF-IDF score for each word in the corpus is now computed. Higher-scoring words are more important, whereas lower-scoring words are less significant. Calculate the tf-idf score:

$$(tf_idf)_idf)_idf)_idf)_{t,d} = tf_{t,d}\, idf_t \tag{3}$$

Countvectorizer: The scikit-learn module in Python provides a fantastic feature called count vectorizer. The count vectorizer employs a bag-of-words method that avoids textual structures. Instead, only word counts are used to extract information. So, First and foremost, each document will be converted to a vector format. Then, the vector's input measures how many times each word appears in the content [61]. Finally, we combined n-grams with Countvectorizer techniques [62].

B. Proposed Work

We proposed a model for detecting emotion in text. Our model will also see racism. We'll look for and examine racism in the context of the Ukraine-Russia conflict. The public's viewpoint toward the competition is also investigated.

1) Detect emotion from text: The dataset description and data preparation techniques, such as text preprocessing and feature extraction, have previously been covered. The dataset is divided into two steps for identifying emotion. During the training phase, 18000 data points were learned, whereas 3000 data points were tested. We used classical machine learning as well as an ensemble model. In training the model to recognize emotion, many classifiers were utilized. Six primary emotions are detected by the system (anger, fear, surprise, joy, love, and sadness). We'll go through the classifier that we used in our system later. Fig. 5 presents the architecture for emotion recognition.

2) Racism detection: Racism was detected using data from Kaggle (dataset 2) and Twitter (dataset 3). The Kaggle dataset has labels, whereas the Twitter dataset does not. We combined the two datasets and utilized text preparation and feature extraction techniques. The boosting approach, as well as a machine learning classifier, were used. One-third of the data is tested, and the rest is trained. The best prediction model was determined after using all approaches and algorithms. The model investigated racism in dataset 3, which concerned the Ukraine-Russia war. In the suggested process for detecting racism, the following algorithm was utilized to categorize tweets as racist or non-racist.

Algorithm 1 Racism Detection Tresholds

if $prediction \geq 0.3$ **then**

  $label \leftarrow 1$

else

  $label \leftarrow 0$

end if

Where 1 implies racism and 0 denotes non racism. The framework for detecting racism is depicted in Fig. 6.

3) Sentiment analysis: The present research's ultimate job is sentiment analysis. Dataset 3 was used to investigate the sentiment. We want to look at how people feel about the Ukraine-Russia conflict. Dataset 3 is obtained from Twitter and is on the conflict between Ukraine and Russia. On Twitter, a tweet is a microblog message. It's only allowed to be 140 characters long. Most tweets include text, embedded URLs, photos, and usernames. There are also misspellings in them. As a result, several preprocessing processes were performed on the tweets to eliminate unnecessary data. The reasoning is that cleaner data is more suitable for mining and feature extraction, resulting in more accurate results. We didn't need to use all of the preprocessing approaches outlined above since VADER can understand the sentiment of a text that incorporates emoticons, slang, conjunctions, capital phrases, punctuation, and other idioms. We avoided misspellings, short words, and unusually lengthy letters.

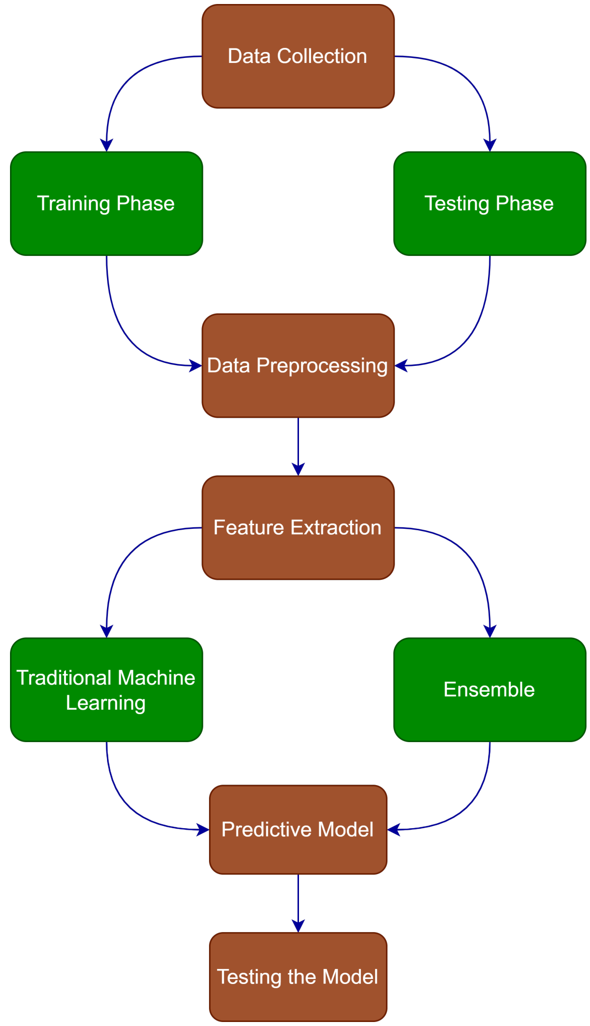


Fig. 5. Emotion detection architecture

Algorithm 2 Sentiment Score Range

if $compoundscore \geq 0.05$ **then**

  $sentiment \leftarrow$positive

else if $compound\_score \leq -0.05$ **then**

  $sentiment \leftarrow Negative$

else if then

  $sentiment \leftarrow Neutral$

end if

We compute a compound score to analyze sentiment. After assessing the sentiment, we determined the positive, negative, and neutral phrases used in the war. The result and discussion section will provide the top terms used negatively

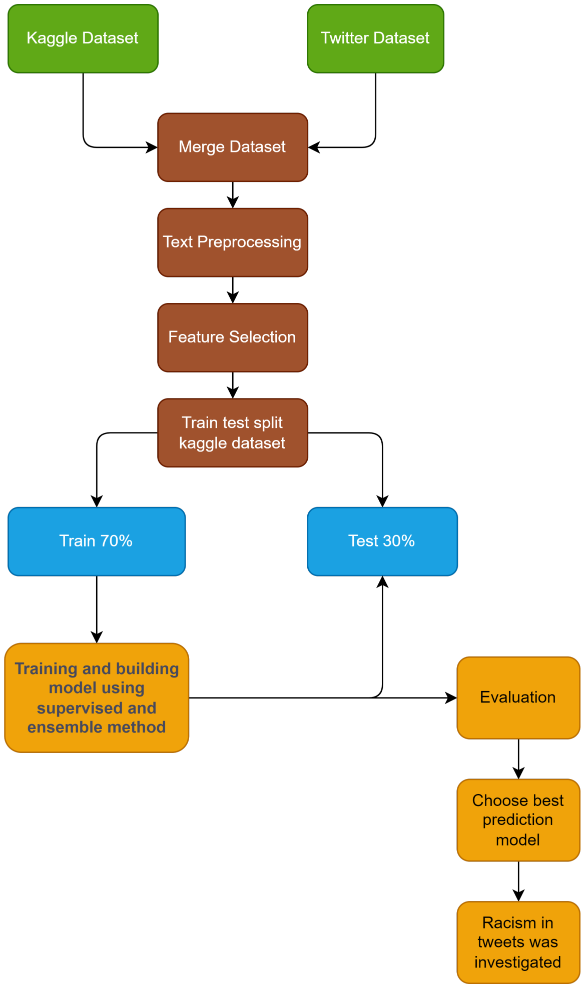


Fig. 6. Racism detection architecture

in the conflict. In addition, positive and neutral words will be highlighted.

C. Proposed Model

1) Supervised machine learning classifiers: The learning process is divided into training and testing phases. First, training data samples were used as input, and then the learning algorithm was applied to learn the feature and build the learning model [63]. Finally, for prediction for the data or test, the learning model executes in the testing phase. The approach of supervised learning aims to train the machine. For each of the algorithm, we used parameters based on need. Fig. 7 depicts the model in detail. The dataset is first split into two phases: training and testing. The data were prepared using a feature engineering technique. In order to forecast the label, the baseline method was used. The test data was finally employed to predict the value and contrast it with its actual value. Finally, a number of evaluation criteria are used to gauge performance. Some of the used supervised algorithms in this system:

Decision Trees: A DT [64] is a classifier defined as a recursive split of the instance space. The DT consists of nodes that form a rooted tree, a dispersed tree with no incoming

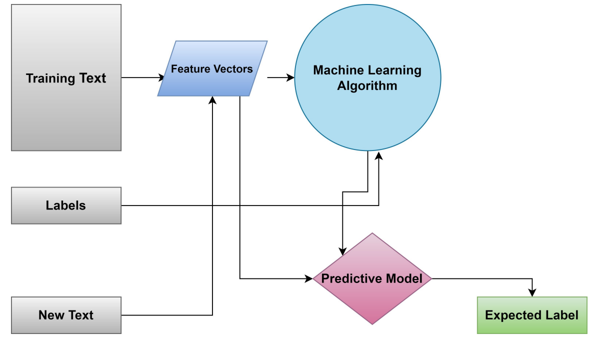


Fig. 7. Supervised learning model architecture

edges, and a root node. Each node has one incoming advantage. An internal node has outgoing edges, whereas leaves are the other nodes. Based on the input values, each internal node in a decision tree separates into multiple sub-spaces. In the simplest case, each test evaluates a single attribute, and the occurrence space is partitioned according to the attribute's value. Every node has a label representing the property being tested, and its branches have starting values. Mini_samples_split, max_depth and class_weight were used as parameter.

Linear Regression: The purpose of LR, a portion of the regression algorithm family, is to discover the interrelationships and dominance of variables. Regression analysis is used to forecast a target variable; predicting an attribute from a limited set is one topic of classification.LR [65] pertains to the supervised learning algorithms area as well.

Naive Bayes: Another supervised learning approach and statistical classification method is Bayesian classification [66]. The Bayesian classification's primary goal is to tackle problems with prediction and solve prediction difficulties. This classification may mix observable data and delivers effective learning techniques. Learning techniques and algorithms have been simplified to understand and analyze using Bayesian classification.

Logistic Regression: LR is the most powerful statistical and data mining technique [67]. LR has several key benefits, including the ability to generate probabilities naturally and the ability to handle multinomial classification issues. A further advantage is that far too many LR model analysis techniques rely on the same fundamentals as linear regression. We determine the likelihood that the output variable's perception corresponds to the proper category using LR [68]. The LR model is used to categorize emotions from inside the input text [69]. With training and testing sets, LR classifies text into several emotion categories [70]. Logistic regression is a broader version of linear regression. Consider the linear regression equation below:

$$x = \mathrm{x}_0 + \mathrm{x}_1\mathrm{z}_1 + \mathrm{x}_2\mathrm{z}_2 + ... + \mathrm{x}_n\mathrm{z}_n \quad (4)$$

Here x is the response variable, and the predictor variables are Z1, Z2, Z3, and Zn. So if we apply a sigmoid function to equation 1, we will get a logistic function.

$$l = \frac{1}{1 + e^{-(\mathrm{x}_0 + \mathrm{x}_1\mathrm{z}_1 + \mathrm{x}_2\mathrm{z}_2 + ... + \mathrm{x}_n\mathrm{x}_n)}} \quad (5)$$

Random Forest: RF is among the technique in Machine Learning used to process the data in vast quantities and widely used algorithm based on the results of a DT formed throughout training [71] [72]. The RF algorithm combines numerous trees with training data to generate greater accuracy. The concentrated output from every DT seems to be the forest's output. Each tree has a different set of features and uses typical tree-building methods, and these properties generate the nodes and leaves [73].

Support Vector Machine: SVM is a pattern identification and categorization technique that is pretty simple to use [74]. SVM is among the most extensively utilized cutting-edge machine learning technology (SVM). The SVM approach is related to supervised learning, which requires feature extraction and produces desired results. SVM has the benefit of being very easy to execute and scaling vast amounts of data more efficiently than neural networks [75]. The kernel approach allows SVMs to effectively conduct nonlinear classification by implicitly translating their inputs within high-dimensional feature regions [76]. SVM finds the biggest and best margin hyperplane to categorize the text into numerous emotion classifications. Choosing an appropriate kernel function to get better results is critical because it determines the transformed feature space in which the training set instances will be categorized. Some well-known kernels were used:

$$Linear : K(\mathrm{A}_i, \mathrm{A}_j) = \mathrm{A}_i^T \mathrm{X}_j \tag{6}$$

$$Polynomial : K(\mathrm{A}_i, \mathrm{A}_j) = (\gamma \mathrm{A}_i^T \mathrm{X}_j + r)^d, \gamma > 0 \tag{7}$$

The kernel parameters are

$$\gamma, r, d$$

$$\mathrm{A}_i$$

is a training vector that the function maps into a high-dimensional space.The linear kernel used sigmoid in the research and the n_samples were 100.

2) Ensemble method: Ensemble learning is a procedure that involves integrating many models or classifiers to address a specific issue. Both machine learning and deep learning techniques can benefit from the ensemble approach. It is followed to attain the intended aim and improve system performance. We applied bagging and boosting in our suggested method to discover the outcome. Though many existing systems use the ensemble approach, we attempted to construct this system using a machine learning-based ensemble model. AdaBoost, Gradient boost, and XGBoost have been used to develop and evaluate overall system performance. AdaBoost uses a recursive technique to help poor classifiers better by learning from their mistakes. Although XGBoost is identical to AdaBoost, it outperforms AdaBoost in highly condensed and sophisticated data and system optimization. Ensemble learning is commonly implemented using decision trees, which aid in the solution of quantitative issues. Ensemble learning computes the final classification based on the ensembled findings given by decision trees rather than relying on a single decision tree's predictive analysis and outcome. In our ensemble learning, we used a variety of machine learning classifiers, including LR, SVM, RF, and others, in addition to the decision tree, to train and analyze the model and find the best output. We used

TABLE III. HYPERPARAMETERS OF THE XGBCLASSIFIER

Types	Values
learning_rate	0.2
max_depth	10
n_estimators	200
subsample	0.8

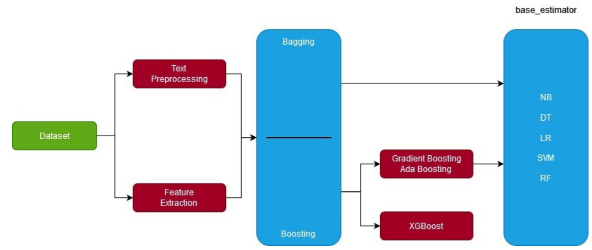


Fig. 8. Proposed ensemble model

several n estimators (10,50,100) to improve performance because the ensemble approaches heavily depend on parameters. Fig. 8 depicts the architecture and base estimator used in our ensemble model. Soft voting and Hard voting classifiers are two types of voting classifiers used to create the results [77]. The test preprocesses ing approach and feature extraction are completed initially. When we deployed the XGB classifier of the boosting ensemble, we employed the grid search parameter tuning strategy to choose the optimum hyperparameters. Table III displays the XGB classifier's hyperparameters.

3) VADER: VADER is an analytical program that uses both analyses of lexical and rules-based to analyze sentiment. It generally employs a combination of strategies that classify lexical characteristics according to their semantic orientation, which might be positive or negative. This approach not only detects the polarity of positive or negative attitudes but also determines how the sentiment is [78]. This tool is consistent and has a higher accuracy ratio for performing sentiment analysis from social media. There are several advantages of VADER. It works well with social media content while quickly summarizing various topics. No need for training data though it includes defined vocabulary for analyzing sentiment. Because of its speed and precision, it may be utilized for asynchronous data. The VADER packages display a positive, neutral, or negative value from individual tweets [79]. We used VADER to analyze the sentiment. First, a compound score is calculated based on these values. Next, a compound score is calculated based on these values. The compound score thresholds were used to categorize the sentiment into negative, positive, and neutral, as indicated in Algorithm 2. However, VADER performs better for social media data which is unsupervised learning.

IV. RESULT AND DISCUSSION

The accuracy of all the models is calculated. After finding the best model, we compute each class's precision, recall, and f-score. Combining word-n-gram with a count vectorizer is the most outstanding result to detect emotion. We employ bigrams and trigrams in our implementation. The accuracy could be determined as described in the equation since accuracy was chosen as the assessment criteria, where TP = True Positive,

TABLE IV. CLASSIFIER PERFORMANCE OF SUPERVISED MODEL

Classifier	Training Result	Testing Result
Logistic Regression	0.99	0.88
Decision Tree	0.99	0.86
Naive Bayes	0.96	0.76
SVC	0.98	0.86

TABLE V. BOOSTING METHOD PERFORMANCE (EMOTION DETECTION)

Algorithm	Training Result	Testing Result
ADABoost	0.99	0.89
GradientBoost	0.96	0.78
XGBoost	0.97	0.90

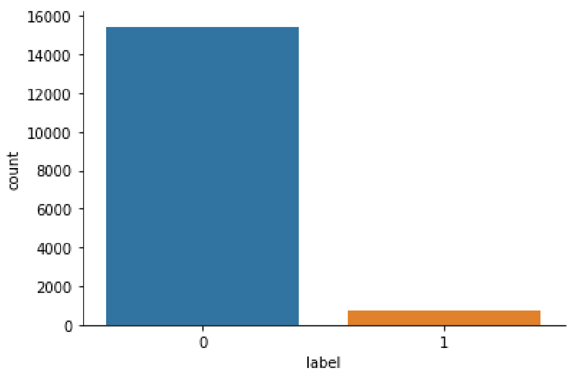


Fig. 9. Racism found in dataset 3

TN =True Negative, FP = False Positive, and FN = False Negative.

The performance of the base model was shown in Table IV. The best testing outcome came from LR. The testing results are the same for DT and SVC. However, NB has the worst results, and overfitting problems are a big deal here. From Table V, XGBoost achieves the desired outcome. Gradient boosting performs the poorest out of all methods, with Adaboosting marginally outperforming the baseline method. The performance of the supervised classifiers for identifying emotion is displayed in Table IV. Since we utilized the supervised model and ensemble model bagging and boosting, Table V shows boosting model performance, while Table VII shows all classes' precision, recall, and f-score. Table VI shows the overall performance of our applied mode to detect emotion. Our technology outperforms the prior method when it comes to seeing emotion. Finally, we compared our model to some of the most recent works in emotion recognition. Table VIII compares our model with others. The accuracy of the proposed ensemble model is higher than that of earlier systems, as demonstrated in Table VIII. Accuracy for gloves, SVM, and LR is under 80%. C-BiLSTM, however, achieved 88% accuracy. The suggested technique, meanwhile, achieves 90% accuracy.

We only use the F-score to detect racism. We used F-score instead of accuracy since we can see how unbalanced our dataset 2 is in Fig. 3. Compared to the values with labels:1, we can observe that the deals with label:0 have a far higher number of differences. As a result, if we use accuracy as our assessment criterion, we may meet a significant proportion of false positives. The percentage of our relevant findings is referred to as precision. The rate of relevant total results accurately categorized by our algorithm is called recall. We are constantly faced with a trade-off between accuracy and recall, with high precision resulting in low memory and vice versa. F-core is calculated using equation 5. Both bag-of-words and tf-idf perform best when it comes to detecting racism. Table IX shows the results of the racism detection.

$$\text{F-score} = 2 * \frac{Precision + Recall}{Precision * Recall} \tag{8}$$

TABLE VI. OVERALL PERFORMANCE OF APPLIED METHOD

Model	Testing Result	Training Result
Supervised	0.88	0.99
Bagging	0.89	0.99
Boosting	0.90	0.97

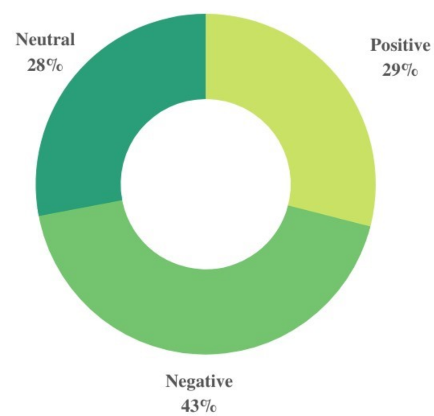


Fig. 10. Sentiment result

TABLE VII. PRECISION, RECALL, F-SCORE OF ALL CLASSES

Emotion	Precision	Recall	F-score
anger	0.90%	0.90%	0.90 %
fear	0.90%	0.83%	0.86 %
joy	0.90%	0.93%	0.91 %
love	0.80%	0.71%	0.75 %
sadness	0.93%	0.94%	0.94 %
surprise	0.67%	0.74%	0.72 %

TABLE VIII. COMPARISON OF RESULT WITH OTHER PAPERS

Ref.	Method/Algorithm	Result
[2]	Glove	F1-69.3 %
[14]	SVM	Accuracy-63.5 %
[30]	C-BiLSTM	F.score-88 %
[33]	LR-SGD	Accuracy-79 %
Proposed system	Ensemble (Boosting)	Accuracy-90.45%

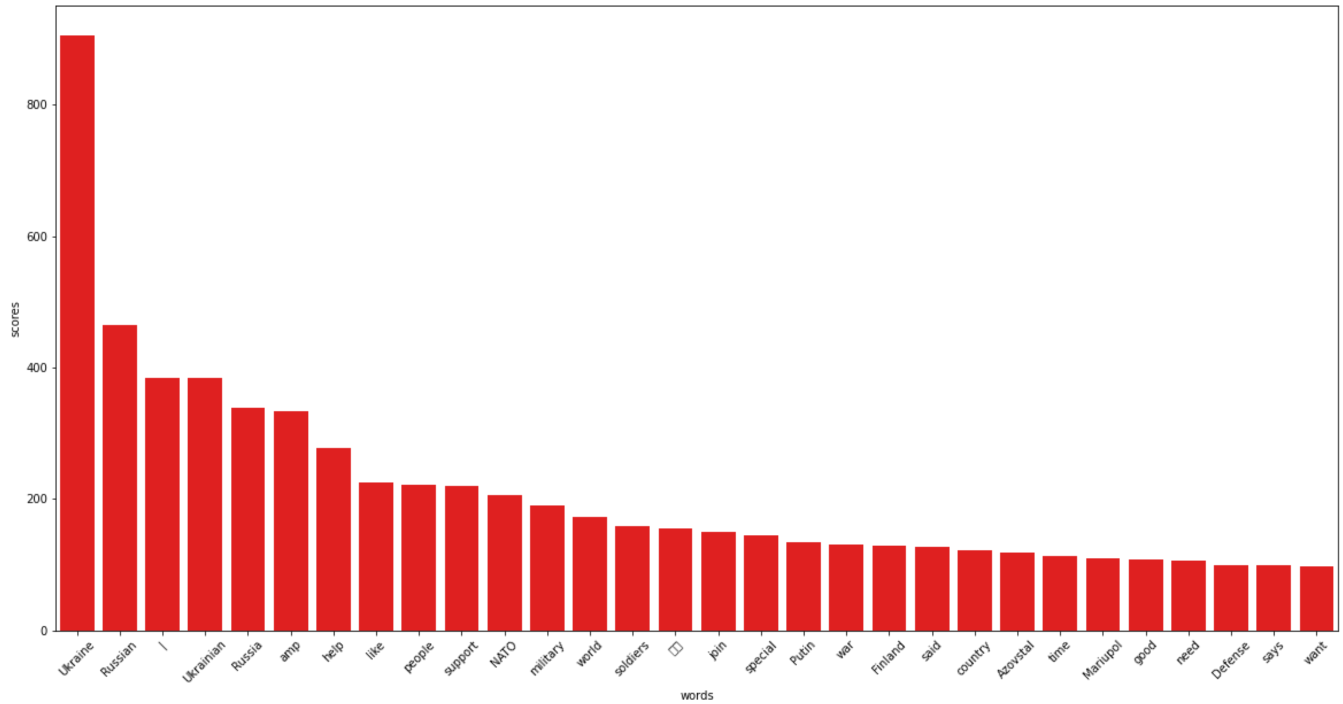


Fig. 11. Most common positive words

TABLE IX. RACISM DETECTION ACCURACY

Algorithm / Model	Feature Extraction	F-score
Logistic Regression	Bag-of-Words	0.552
Logistic Regression	Tf-Idf	0.560
XGBoost	Bag-of-Word	0.563
XGBoost	Tf-Idf	0.547

We chose our best model for identifying racism and used it to examine racism in tweets on the Ukraine-Russia conflict. According to our research, people are less racist regarding this war.

The outcome is shown in Fig. 9. The figure shows that to examine the racism in dataset 3, people are less racist about the war—15459 sentences that are non-racist and 749 sentences that are racist. Finally, when it comes to sentiment analysis, the results show that negative sentiment was identified in 7002, the positive feeling was found in 4639 tweets, and neutral attitude in 4567. Fig. 10 depicts the sentiment discovered in dataset 3.

A. Discussion

We integrate the best model of our used procedures into our website after obtaining it. When a user registers on the website, a unique user name is assigned to them. Then, our system can discern their emotion using that person's textual data. The person's emotional state is kept in the database after the emotion is detected. The system can only store an individual's feelings and restore their output over time. As a result, anyone can examine their feelings throughout time. The system can display a person's emotions at any time. However, in the future, we will expand the system to automatically analyze a person's feelings and warn the user of any emotional changes. A future study could look into the dynamic alterations.

We have had some issues with the application of detecting racism.

The accuracy of racism detection was low, as seen in the results section. We discovered that an unbalanced dataset caused the problem after further investigation. We detected more nonracist text in our dataset than racist text. When a dataset is unstable, it refers to the data dominance space vs. the dispersed data. Unbalanced learning challenges can be divided into two categories. There are two types of relationships:

1) between classes and
2) inside classes.

- Between the classes: This refers to the unequal distribution of data samples between two classes.
- Inside the class: when the classes of data differ between more than one idea, the unbalanced problem is considered inside the class.

As a result, our dataset issue was sandwiched between the problem of class imbalance. Oversampling and undersampling approaches are two techniques for sorting this problem [80]. Future studies can handle these issues. If we have a sizeable balanced dataset, we can use deep learning in the future. In addition, we may address the issue of data imbalance to achieve better results. The globe is in the grip of the Ukraine-Russia conflict. We attempted to overserve the sentiment of the general public. Geographical sentiment analysis can help decision-makers understand a particular country's sentiment. It will aid in taking any measures necessary to end the problem. Positive, negative, and neutral words are illustrated in Fig.

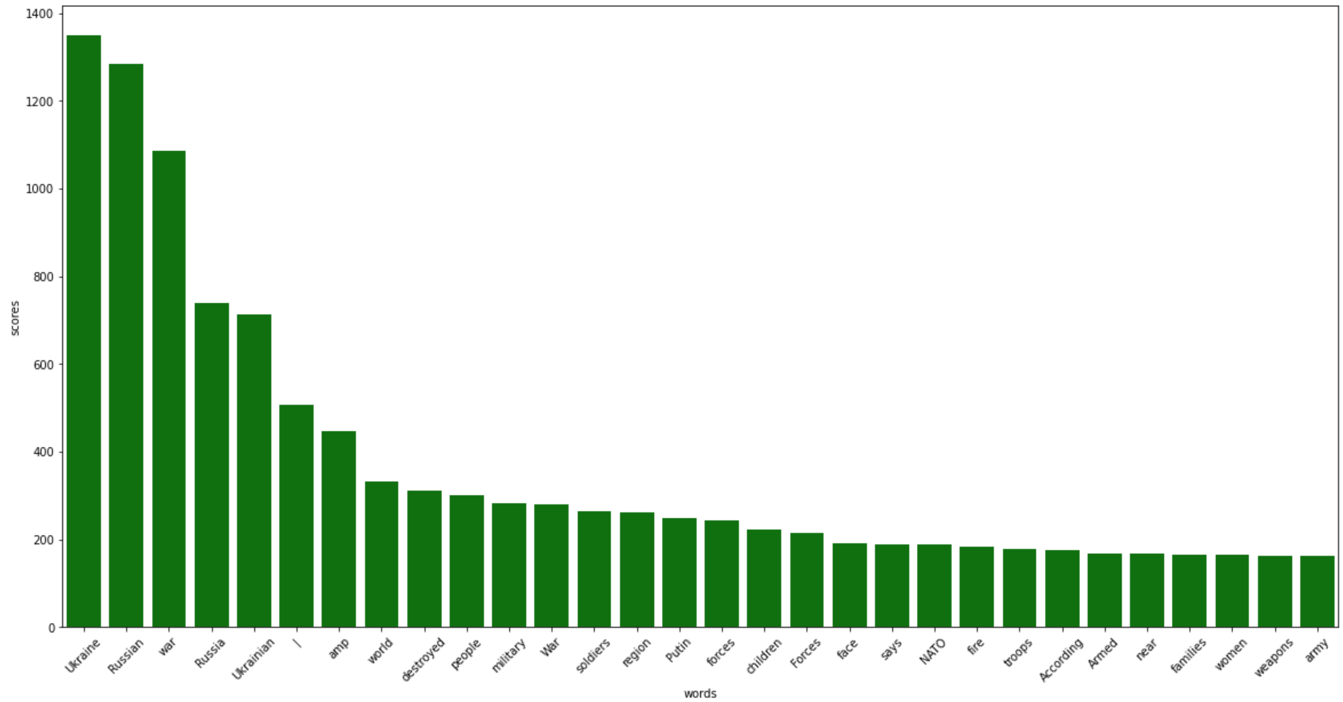


Fig. 12. Most common negative words

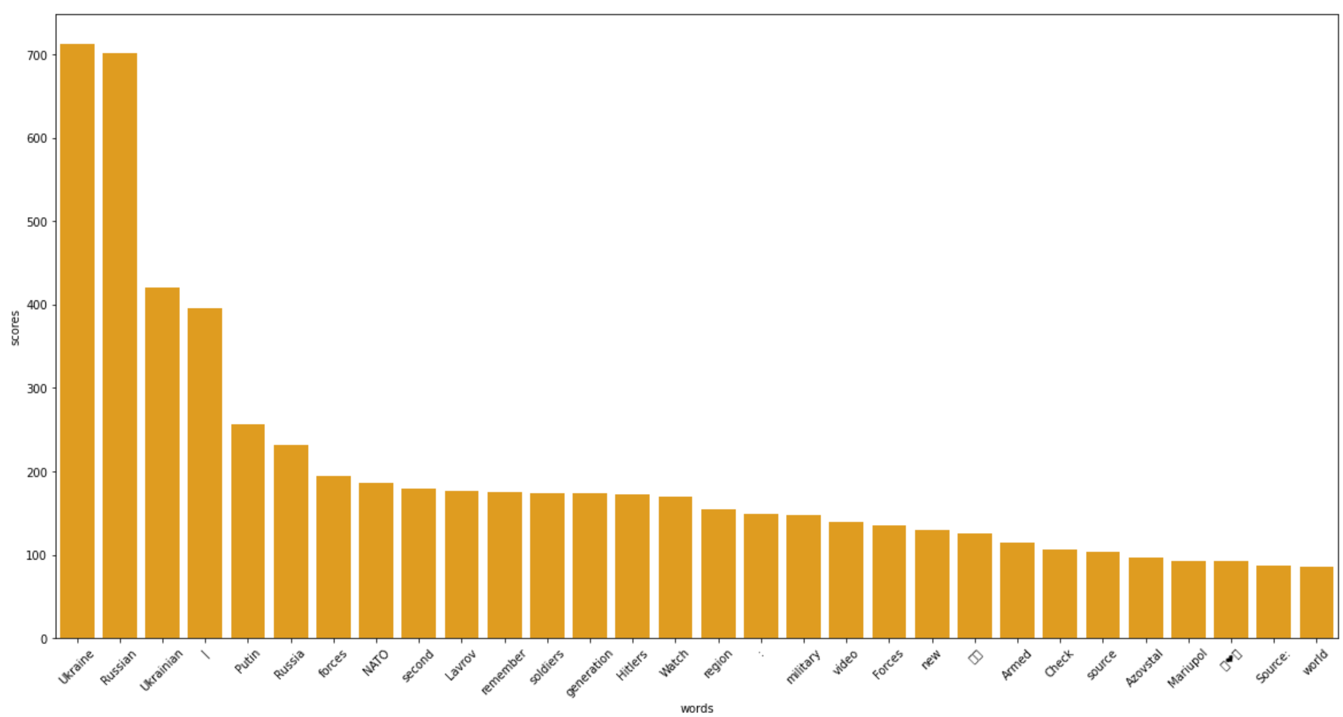


Fig. 13. Most common neutral words

11, 12, and 13. The most frequently used words are Ukraine, Russia, and Ukrainian, which can be used in positive, negative, or neutral contexts. Every term has a mixed connotation, meaning that some people will interpret it positively while others will interpret it negatively. Therefore, more research is required, and each word can be given as a bigram or trigram. The most common words in the study are visually depicted. Therefore, a thorough analysis is advised.

V. Conclusion

This study was designed to accomplish several goals. This study first tries to assess an individual's emotional trajectory through time and to recognize emotion from textual format using machine learning techniques. The ensemble technique works best for detecting emotion. However, there were various difficulties in determining the best accuracy from our dataset. We decided on the ensemble model's optimal accuracy by integrating several data preparation methods. XGBclassifier demonstrated 90% accuracy in identifying the mood in the text.

There are several applications for emotion detection, including marketing, customer reviews, the detection of psychological instability, and others. XGBclassifier and Logistic Regression perform best in identifying racism. However, the ability to recognize racism has to be improved. In our study, we looked at how individuals felt about the conflict between Ukraine and Russia. Nearly half of the Twitter users had unfavorable feelings (43%). The study of this conflict had not received much attention from artificial intelligence researchers. The professional can remedy the problem if we further investigate this. No system is flawless, after all, and we are fully conscious of the constraints of our system. We are looking to implement this system using multiple or large datasets, and deep learning techniques will be applied in the future. Our system comparatively works better than the previous systems to detect emotion. The novelty and future development have already been explored. We hope that this study will contribute to further research and influence the field of emotion recognition, racism detection, and sentiment analysis.

Author Contributions

Abdullah Al Maruf: Implementation; Writing: Review and editing; Writing: Original draft; Data collection.Zakaria Masud Jiyad: Writing: Review and editing; Writing: Original draft; Data collection. Md. Mahmudul Haque::Writing: Review and editing; Writing: Original draft; Data collection.Fahima Khanam: Supervision

References

[1] R. Rahman *et al.*, "Detecting emotion from text and emoticon," *London Journal of Research in Computer Science and Technology*, 2017.

[2] U. Rashid, M. W. Iqbal, M. A. Skiandar, M. Q. Raiz, M. R. Naqvi, and S. K. Shahzad, "Emotion detection of contextual text using deep learning," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, 2020, pp. 1–5.

[3] S. A. Salam and R. Gupta, "Emotion detection and recognition from text using machine learning," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 6, pp. 341–345, 2018.

[4] C. Pool and M. Nissim, "Distant supervision for emotion detection using facebook reactions," *arXiv preprint arXiv:1611.02988*, 2016.

[5] M. Hasan, E. Agu, and E. Rundensteiner, "Using hashtags as labels for supervised learning of emotions in twitter messages," in *ACM SIGKDD workshop on health informatics, New York, USA*, vol. 34, no. 74, 2014, p. 100.

[6] M. Hasan, E. Rundensteiner, and E. Agu, "Emotex: Detecting emotions in twitter messages," 2014.

[7] S. N. Shivhare, S. Garg, and A. Mishra, "Emotionfinder: Detecting emotion from blogs and textual documents," in *International Conference on Computing, Communication & Automation*. IEEE, 2015, pp. 52–57.

[8] B. Le and H. Nguyen, "Twitter sentiment analysis using machine learning techniques," in *Advanced Computational Methods for Knowledge Engineering*. Springer, 2015, pp. 279–289.

[9] S. Kuamri and C. N. Babu, "Real time analysis of social media data to understand people emotions towards national parties," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2017, pp. 1–6.

[10] R. Pandarachalil, S. Sendhilkumar, and G. Mahalakshmi, "Twitter sentiment analysis for large-scale data: an unsupervised approach," *Cognitive computation*, vol. 7, no. 2, pp. 254–262, 2015.

[11] P. J. Manamela, M. J. Manamela, T. I. Modipa, T. J. Sefara, and T. B. Mokgonyane, "The automatic recognition of sepedi speech emotions based on machine learning algorithms," in *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*. IEEE, 2018, pp. 1–7.

[12] V. U. Kumar, A. Savithri, M. J. Bhavani, A. M. Priya, K. V. S. B. Jahnavi, and N. D. N. Lakshmi, "Finding psychological instability using machine learning," in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*. IEEE, 2020, pp. 1–4.

[13] A. D. Souza and R. D. Souza, "Affective interaction based hybrid approach for emotion detection using machine learning," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2019, pp. 337–342.

[14] S. Koppu, K. Lakshmanna, P. Venugopal, R. K. Poluru *et al.*, "Automated category text identification using machine learning," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. IEEE, 2020, pp. 1–4.

[15] L. Sana, K. Nasir, A. Urooj, Z. Ishaq, and I. A. Hameed, "Bers: Bussiness-related emotion recognition system in urdu language using machine learning," in *2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC)*. IEEE, 2018, pp. 238–242.

[16] S. Grover and A. Verma, "Design for emotion detection of punjabi text using hybrid approach," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 2. IEEE, 2016, pp. 1–6.

[17] S. Azmin and K. Dhar, "Emotion detection from bangla text corpus using naive bayes classifier," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*. IEEE, 2019, pp. 1–5.

[18] M. Healy, R. Donovan, P. Walsh, and H. Zheng, "A machine learning emotion detection platform to support affective well being," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 2694–2700.

[19] B. Sun, Q. Ma, S. Zhang, K. Liu, and Y. Liu, "iself: Towards cold-start emotion labeling using transfer learning with smartphones," *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 4, pp. 1–22, 2017.

[20] O. Rabie and C. Sturm, "Feel the heat: Emotion detection in arabic social media content," in *The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014)*. Citeseer Kuala Lumpur, Malaysia, 2014, pp. 37–49.

[21] W. A. Hussien, Y. M. Tashtoush, M. Al-Ayyoub, and M. N. Al-Kabi, "Are emoticons good enough to train emotion classifiers of arabic tweets?" in *2016 7th International Conference on Computer Science and Information Technology (CSIT)*. IEEE, 2016, pp. 1–6.

[22] S. S. Sadeghi, H. Khotanlou, and M. Rasekh Mahand, "Automatic persian text emotion detection using cognitive linguistic and deep learning," *Journal of AI and Data Mining*, vol. 9, no. 2, pp. 169–179, 2021.

[23] M. Abdullah, M. Hadzikadicy, and S. Shaikhz, "Sedat: sentiment and emotion detection in arabic text using cnn-lstm deep learning," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2018, pp. 835–840.

[24] M. Baali and N. Ghneim, "Emotion analysis of arabic tweets using deep learning approach," *Journal of Big Data*, vol. 6, no. 1, pp. 1–12, 2019.

[25] Y. Rao, H. Xie, J. Li, F. Jin, F. L. Wang, and Q. Li, "Social emotion classification of short text via topic-level maximum entropy model," *Information & Management*, vol. 53, no. 8, pp. 978–986, 2016.

[26] M. Hajar *et al.*, "Using youtube comments for text-based emotion recognition," *Procedia Computer Science*, vol. 83, pp. 292–299, 2016.

[27] D. Xu, Z. Tian, R. Lai, X. Kong, Z. Tan, and W. Shi, “Deep learning based emotion analysis of microblog texts,” *Information Fusion*, vol. 64, pp. 1–11, 2020.

[28] M. Karna, D. S. Juliet, and R. C. Joy, “Deep learning based text emotion recognition for chatbot applications,” in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. IEEE, 2020, pp. 988–993.

[29] M. Jabreel and A. Moreno, “A deep learning-based approach for multi-label emotion classification in tweets,” *Applied Sciences*, vol. 9, no. 6, p. 1123, 2019.

[30] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and S. Khan, “Classification of poetry text into the emotional states using deep learning technique,” *IEEE Access*, vol. 8, pp. 73 865–73 878, 2020.

[31] D. Haryadi and G. P. Kusuma, “Emotion detection in text using nested long short-term memory,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.

[32] H. Al-Omari, M. A. Abdullah, and S. Shaikh, “Emodet2: Emotion detection in english textual dialogue using bert and bilstm models,” in *2020 11th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2020, pp. 226–232.

[33] A. Yousaf, M. Umer, S. Sadiq, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, “Emotion recognition by textual tweets classification using voting classifier (lr-sgd),” *IEEE Access*, vol. 9, pp. 6286–6295, 2020.

[34] A. Sarlan, C. Nadam, and S. Basri, “Twitter sentiment analysis,” in *Proceedings of the 6th International conference on Information Technology and Multimedia*. IEEE, 2014, pp. 212–216.

[35] A. Go, L. Huang, and R. Bhayani, “Twitter sentiment analysis,” *Entropy*, vol. 17, p. 252, 2009.

[36] K. K. Pawar, P. P. Shrishrimal, and R. R. Deshmukh, “Twitter sentiment analysis: A review,” *International Journal of Scientific & Engineering Research*, vol. 6, no. 4, pp. 957–964, 2015.

[37] T. Rao, S. Srivastava *et al.*, “Analyzing stock market movements using twitter sentiment analysis,” 2012.

[38] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, “A system for real-time twitter sentiment analysis of 2012 us presidential election cycle,” in *Proceedings of the ACL 2012 system demonstrations*, 2012, pp. 115–120.

[39] O. Adwan, M. Al-Tawil, A. Huneiti, R. Shahin, A. A. Zayed, and R. Al-Dibsi, “Twitter sentiment analysis approaches: A survey,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 15, pp. 79–93, 2020.

[40] M. Bibi, W. Aziz, M. Almaraashi, I. H. Khan, M. S. A. Nadeem, and N. Habib, “A cooperative binary-clustering framework based on majority voting for twitter sentiment analysis,” *IEEE Access*, vol. 8, pp. 68 580–68 592, 2020.

[41] J. N. Rao, Y. DurgaMalleswari, K. T. Krishna, C. V. Sai, and Y. R. Sri, “Twitter sentiment analysis for different business applications,” *Journal ofAdvanced Research in Dynamical and Control Systems*, vol. 12.

[42] E. S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, O. S. Siordia, and E. A. Villaseñor, “A case study of spanish text transformations for twitter sentiment analysis,” *Expert Systems with Applications*, vol. 81, pp. 457–471, 2017.

[43] E. Sutoyo and A. Almaarif, “Twitter sentiment analysis of the relocation of indonesia’s capital city,” *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1620–1630, 2020.

[44] D. Jovanoski, V. Pachovski, and P. Nakov, “Sentiment analysis in twitter for macedonian,” *arXiv preprint arXiv:2109.13725*, 2021.

[45] M. Pota, M. Ventura, R. Catelli, and M. Esposito, “An effective bert-based pipeline for twitter sentiment analysis: A case study in italian,” *Sensors*, vol. 21, no. 1, p. 133, 2020.

[46] M. Wang and G. Hu, “A novel method for twitter sentiment analysis based on attentional-graph neural network,” *Information*, vol. 11, no. 2, p. 92, 2020.

[47] S. C. Nistor, M. Moca, D. Moldovan, D. B. Oprean, and R. L. Nistor, “Building a twitter sentiment analysis system with recurrent neural networks,” *Sensors*, vol. 21, no. 7, p. 2266, 2021.

[48] M. Bibi, W. A. Abbasi, W. Aziz, S. Khalil, M. Uddin, C. Iwendi, and T. R. Gadekallu, “A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis,” *Pattern Recognition Letters*, vol. 158, pp. 80–86, 2022.

[49] S. H. Muhammad, D. I. Adelani, I. S. Ahmad, I. Abdulmumin, B. S. Bello, M. Choudhury, C. C. Emezue, A. Aremu, S. Abdul, and P. Brazdil, “Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis,” *arXiv preprint arXiv:2201.08277*, 2022.

[50] D. Antypas, A. Preece, and J. C. Collados, “Politics and virality in the time of twitter: A large-scale cross-party sentiment analysis in greece, spain and united kingdom,” *arXiv preprint arXiv:2202.00396*, 2022.

[51] J. Gautam, M. Atrey, N. Malsa, A. Balyan, R. N. Shaw, and A. Ghosh, “Twitter data sentiment analysis using naive bayes classifier and generation of heat map for analyzing intensity geographically,” in *Advances in Applications of Data-Driven Computing*. Springer, 2021, pp. 129–139.

[52] H. A. Shehu, M. H. Sharif, M. H. U. Sharif, R. Datta, S. Tokat, S. Uyaver, H. Kusetogullari, and R. A. Ramadan, “Deep sentiment analysis: a case study on stemmed turkish twitter data,” *IEEE Access*, vol. 9, pp. 56 836–56 854, 2021.

[53] S. K. Trivedi and A. Singh, “Twitter sentiment analysis of app based online food delivery companies,” *Global Knowledge, Memory and Communication*, 2021.

[54] T. H. McCormick, H. Lee, N. Cesare, A. Shojaie, and E. S. Spiro, “Using twitter for demographic and social science research: tools for data collection and processing,” *Sociological methods & research*, vol. 46, no. 3, pp. 390–421, 2017.

[55] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, 2010.

[56] Y. Bao, C. Quan, L. Wang, and F. Ren, “The role of pre-processing in twitter sentiment analysis,” in *International conference on intelligent computing*. Springer, 2014, pp. 615–624.

[57] P. Shah, P. Swaminarayan, and M. Patel, “Sentiment analysis on film review in gujarati language using machine learning,” *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, p. 1030, 2022.

[58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[59] F. Pedregosa *et al.*, “sklearn. feature_extraction. text. tfidfvectorizer,” 2013.

[60] S. R. D’souza and K. Sonawane, “Sentiment analysis based on multiple reviews by using machine learning approaches,” in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019, pp. 188–193.

[61] A. Dadhich and B. Thankachan, “Sentiment analysis of amazon product reviews using hybrid rule-based approach,” in *Smart Systems: Innovations in Computing*. Springer, 2022, pp. 173–193.

[62] S. C. Eshan and M. S. Hasan, “An application of machine learning to detect abusive bengali text,” in *2017 20th International conference of computer and information technology (ICCIT)*. IEEE, 2017, pp. 1–6.

[63] V. Nasteski, “An overview of the supervised machine learning methods,” *Horizons. b*, vol. 4, pp. 51–62, 2017.

[64] O. Z. Maimon and L. Rokach, *Data mining with decision trees: theory and applications*. World scientific, 2014, vol. 81.

[65] C.-J. Lin, R. C. Weng, and S. S. Keerthi, “Trust region newton methods for large-scale logistic regression,” in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 561–568.

[66] J. H. Friedman and N. I. Fisher, “Bump hunting in high-dimensional data,” *Statistics and computing*, vol. 9, no. 2, pp. 123–143, 1999.

[67] M. Maalouf, “Logistic regression in data analysis: an overview,” *International Journal of Data Analysis Techniques and Strategies*, vol. 3, no. 3, pp. 281–299, 2011.

[68] T. Rymarczyk, E. Kozłowski, G. Kłosowski, and K. Niderla, “Logistic regression for machine learning in process tomography,” *Sensors*, vol. 19, no. 15, p. 3400, 2019.

[69] F. M. Alotaibi, “Classifying text-based emotions using logistic regression,” 2019.

[70] M. Z. Asghar, F. Subhan, M. Imran, F. M. Kundi, S. Shamshirband, A. Mosavi, P. Csiba, and A. R. Várkonyi-Kóczy, "Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content," *arXiv preprint arXiv:1908.01587*, 2019.

[71] K. Nugroho, E. Noersasongko, A. Z. Fanani, R. S. Basuki *et al.*, "Improving random forest method to detect hatespeech and offensive word," in *2019 International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2019, pp. 514–518.

[72] M. Z. Asghar, F. Subhan, M. Imran, F. M. Kundi, S. Shamshirband, A. Mosavi, P. Csiba, and A. R. Várkonyi-Kóczy, "Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content," *arXiv preprint arXiv:1908.01587*, 2019.

[73] F. Livingston, "Implementation of breiman's random forest machine learning algorithm," *ECE591Q Machine Learning Journal Paper*, pp. 1–13, 2005.

[74] M. Jain, S. Narayan, P. Balaji, A. Bhowmick, R. K. Muthu *et al.*, "Speech emotion recognition using support vector machine," *arXiv preprint arXiv:2002.07590*, 2020.

[75] T. Pfister, "Emotion detection from speech," *Gonville & Caius College*, p. 60, 2010.

[76] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, pp. 381–386, 2020.

[77] O. Sagi and L. Rokach, "Ensemble learning: A survey wiley interdisciplinary reviews: Data mining and knowledge discovery," 2018.

[78] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and vader sentiment," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 122, 2019, p. 16.

[79] K. Mishra, I. Kandasamy, V. Kandasamy WB, and F. Smarandache, "A novel framework using neutrosophy for integrated speech and text sentiment analysis," *Symmetry*, vol. 12, no. 10, p. 1715, 2020.

[80] M. S. Shelke, P. R. Deshmukh, and V. K. Shandilya, "A review on imbalanced data handling using undersampling and oversampling technique," *Int. J. Recent Trends Eng. Res*, vol. 3, no. 4, pp. 444–449, 2017.

An Analytical Model of Induction Motors for Rotor Slot Parametric Design Performance Evaluation

Ahamed Ibrahim Sithy Juhaniya[1], Ahmad Asrul Ibrahim[2], Muhammad Ammirrul Atiqi Mohd Zainuri[3], Mohd Asyraf Zulkifley[4]
Dept. Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia

***Abstract*—Induction motors are commonly used in most electricity generation due to their low investment cost. However, the performance of the induction motors for different applications highly depends on rotor design and machine geometry. For example, changing rotor bar height and width varies the rotor resistance and reactance, thereby leading to variation of the motor efficiency. A parametric study on rotor slot geometry parameters such as opening height, rotor slot depth, and rotor slot width, is carried out to investigate the effect of the parameters on the efficiency of a squirrel cage induction motor. The study is based on analytical model that considers a general-purpose squirrel cage induction motor with the specification of 5.5 kW, 60 Hz, and 460 V. The analytical model is developed and simulated within the MATLAB software environment. The effects of each parameter variation toward efficiency of the induction motor are investigated individually as well as all together using a 4D scatter plot. Results show that the efficiency can be improved up to 0.1% after designing a suitable setting of rotor slot parameters from the initial settings.**

***Keywords*—Analytical model; efficiency; induction motor; rotor slot parameters**

I. Introduction

Electrical machine is widely used for electromechanical energy conversion either to operate as generator or motor [1]. The electrical machines can be categorized as motors and generators depending on their energy conversion mechanism. They are considered electrical motors when they convert electrical energy into mechanical energy. Among various types of electrical motors, DC motors, synchronous motors, and induction motors are more popular for applications in residential and industrial [2], [3]. On one hand, DC motors require commutation mechanisms in their operation, leading to high maintenance cost. On the other hand, the synchronous motors are very expensive due to the use of high-cost permanent magnets; furthermore, they require starting mechanisms. Therefore, the induction motor is a promising option for electromechanical energy conversion applications [4] because this type of electrical machines is a low cost, low maintenance, and self-starting mechanism. Although the induction motors are favorable for many applications; they suffer from some disadvantages particularly low efficiency [5]. Efficiency is an important indicator to show the performance of an electrical motor. It can be improved by modeling the electrical motor using optimal parameter settings [6]. Among the design parameters of the electrical motor, rotor slot geometry parameters provide more effect to its performance. This condition is mainly due to the rotor slot design, which determines the resistance and leakage reactance of the electrical motor.

Several works have been carried out in the past few years to improve the performance of induction motors by redesigning the rotor slot. In [2], various shapes of rotor bar were considered to determine the best shape that provides the highest efficiency with minimum overall cost. The effect of using different materials specifically aluminium and copper for the rotor bar were studied in [7]. Flux2D software was used in [8] to carry out a parametric study of rotor slot design using a finite element method (FEM) technique to showcase their effect on the behavior of induction motors, such as flux distribution and power losses. The FEM technique was also used in [9] to study the impact of changing rotor slot parameters toward the induction motor performance, total harmonic distortion and starting characteristics of the motor. In [10], the FEM technique was used to analyze the electromagnetic characteristics of induction motors using various shapes of rotor bars to obtain the best design. Apart from studying normal rotor bar shapes, the effect of broken rotor bar and different stator slot opening settings were tested using FEM technique to evaluate the induction motor performance [6], [11]. Investigation on the rotor bar shapes was also carried out in [12], but they focused on generator rather than motor applications.

In the previous works, the investigation of changing the rotor slot geometry parameters was mainly based on 2D analysis instead of 3D analysis, which provides less accurate results. The FEM technique, which is based on a commercial software, is not flexible because of computational burden particularly for the parametric study at the initial design process. Furthermore, the important performance parameters, such as efficiency, could not be appropriately addressed. Therefore, this work is carried out to investigate the effect of changing rotor slot parameters toward efficiency of the induction motor using an analytical model. In the following section, the basic principles of induction motor is discussed to provide better understanding before the analytical model to calculate efficiency can be constructed. Then, a development of the analytical model is explained in Section III. In Section IV, results on varying rotor slot parameters are showcased and discussed. Finally, a conclusion of the research work is drawn in Section V.

II. Efficiency of Induction Motor

Induction motors are used in most applications to fulfill either industrial and domestic needs known as workhorse of modern industry [13]. The induction motor can be divided into two types: 1) wound rotor induction motor and 2) squirrel cage induction motors. The squirrel cage type is widely used in the industrial applications because of its idiosyncratic characteristics, such as simplicity, robustness, easy maintenance, and low cost [14]. Therefore, a squirrel cage induction motor model is used in this work.

Operational principles of squirrel cage induction motor are very much similar to transformers. Two magnetic circuits in transformers, referred to as primary and secondary windings, are isolated by a small air-gap. The circuits can be represented as two relatively movable windings which are called stator and rotor in the induction motor [14], [15]. The stator winding of the three-phase squirrel cage induction motor is connected to the three-phase AC source either in a star or delta configuration. Current is supplied to the rotor circuit by the process of electromagnetic induction [16] from the magnetic field of stator circuit. The induction motor is called a singly excited machine because the rotor has no external source. The induction motors are considered self-starting motors because they do not require any external means during the starting condition [17]. Induction motor normally operates at a speed that slightly lower than its synchronous speed; therefore, it is known as asynchronous motor. The difference between operating speed and synchronous speed is known as slip speed of the motor.

Understanding the power losses of the induction motor is important before its efficiency can be calculated. Power flow diagram, as shown in Fig. 1, is used to illustrate the power losses of induction motor. Power transfer from stator to rotor is referred as the airgap power loss, P_{ag}. P_{ag} is calculated by taking into account stator copper losses, P_{sc} and iron losses, P_{iron} as expressed in (1). On the other hand, the conversion power loss, P_{conv} is the total electrical output power from the rotor after subtraction of all types of electrical losses. P_{conv} is calculated by subtracting the rotor copper losses from the airgap power loss. Rotor copper losses, P_{rc} are copper losses generated inside the rotor due to the rotor internal resistance. The output power of the motor is obtained by considering all the losses due to the rotational motion of the rotor shaft. The losses related with rotation motion of the shaft consist of mechanical and ventilation power losses, P_{mv} and stray power losses, P_{stray}.

$$P_{ag} = P_{input} - P_{sc} - P_{iron} \tag{1}$$

$$P_{conv} = P_{ag} - P_{rc} \tag{2}$$

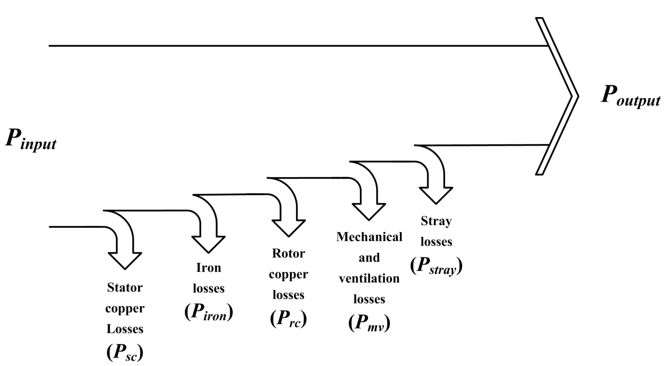


Fig. 1. Power flow diagram of induction motor [15], [16], [17]

Total power loss of induction motor can be calculated as follows:

$$P_{loss} = P_{sc} + P_{rc} + P_{iron} + P_{stray} + P_{mv} \tag{3}$$

The value of P_{mv} can be obtained depending on the number of poles, as shown in Table I. In this work, a four-pole induction motor is considered; therefore, the mechanical and ventilation losses are 1.2% of the output power.

TABLE I. MECHANICAL AND VENTILATION LOSSES, P_{mv} [18]

No. of poles	P_{mv}
2	$0.03 \times P_{out}$
4	$0.012 \times P_{out}$
6,8	$0.08 \times P_{out}$

The stator copper losses, P_{sc} due to resistance in the stator winding can be calculated using stator current, I_s and stator resistance, R_1 as expressed in (4). On the contrary, the rotor copper losses, P_{rc} due to resistance in the rotor winding can be obtained from the rotor current, I_r and rotor resistance, R_2 using expression in (5). The values of R_1 and R_2 can be derived from stator and rotor slot modeling subsections that will be discussed in detail subsequently in the induction motor modeling section.

$$P_{sc} = 3(I_s)^2 R_1 \tag{4}$$

$$P_{rc} = 3(I_r)^2 R_2 \tag{5}$$

The iron losses, P_{iron} are the summation of stator tooth iron losses, P_{st}, stator yoke iron losses, P_{sy}, and tooth flux pulsation core losses, P_{tp}, as follows:

$$P_{iron} = P_{st} + P_{sy} + P_{tp} \tag{6}$$

The stator tooth iron losses, P_{st} can be calculated as follows:

$$P_{st} = \; K_T P'(f/50)^{1.3} B_{TS}^{1.7} W_T \tag{7}$$

where f refers to frequency, K_T represents the core loss augmentation due to mechanical machining, and P' is the specific weight losses in W/kg at 1.0 T and 50 Hz. The value of P' is normally between 2 and 3 W/kg as used by the manufacturer [11]. B_{TS} is the stator tooth flux density, and W_T is the weight of the stator tooth. In this model, B_{TS} and W_T are set to 1.55 T and 4 kg, respectively.

The stator yoke iron losses, P_{sy} can be determined using the following formula:

$$P_{sy} = K_Y P'(f/50)^{1.3} B_C^{1.7} W_Y \tag{8}$$

where K_Y stands for the mechanical machining factor of the stator yoke that normally takes the value in the range between 1.6 and 1.9. B_C and W_Y represent yoke flux density of stator and weight of the stator, respectively.

The tooth pulsation core losses, P_{tp} are neglected because the value is remarkably smaller than the iron losses related with stator tooth and stator yoke.

The stray losses, P_{stray} include the core losses associated with rotor surface and space harmonic cage losses. In most applications, the stray losses are considered by the manufacturers as 1% of the expected output power of the motor [19] and can be expressed as follows:

$$P_{stray} = 0.01 \times P_{out} \tag{9}$$

The efficiency of induction motor, η can be obtained from the total power losses, P_{loss} and output power, P_{out} using the following expression:

$$\eta = \frac{P_{out}}{(P_{out} + P_{loss})} \tag{10}$$

III. Modeling of Induction Motor

Fig. 2 shows an equivalent circuit of the induction motor as referred to the stator side using the concept of transformer modeling. The stator winding resistance, R_1 represents copper losses in the stator, and the stator winding reactance, X_1 represents leakage flux of the stator. The magnetization reactance, X_m represents magnetization effect of the induction motor. R_2 and X_2 are rotor resistance and rotor leakage reactance, respectively.

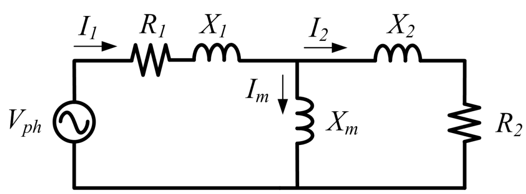


Fig. 2. Equivalent circuit of induction motor

An induction motor can be modelled using equations derived from the per phase equivalent circuit in Fig. 2. In this work, a 5.5 kW, 460 V, and 60 Hz squirrel cage induction motor is considered; its specifications are tabulated in Table II. Single-cage type slots are used for rotor and stator slots.

TABLE II. Motor Specifications

Parameters	Values
Output power(kW)	5.5
Frequency(Hz)	60
Rated voltage (V)	460
Power factor	0.83
No. phase	3
No. poles	4

Fig. 3 shows a quarter section of an induction motor model where all the parameters should be obtained to fulfill the design requirement. Parameter settings of the induction motor are presented in Table III. The parameter stack length refers to the total length of lamination stack, which provides rigidity to the stator. Stack length excludes the end ring length.

A. Stator Slot Design

Fig. 4 illustrates a shape of stator slot and its geometry parameters that can be used to calculate the stator resistance (R_1) and stator leakage reactance (X_1). This work provides focus on the rotor slot parameters; thus, all stator slot dimensions

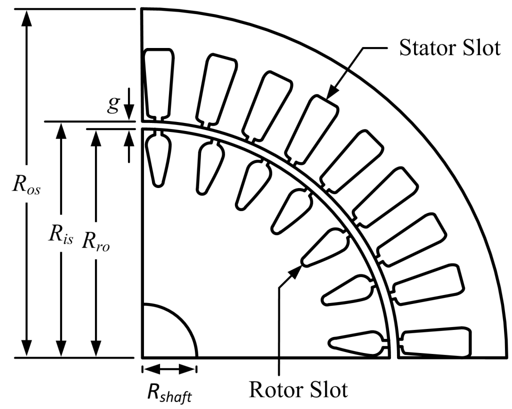


Fig. 3. Quarter section of an induction motor

TABLE III. Induction Motor Parameter Settings

Parameters	Dimension
Stator outer radius (R_{os})	90 mm
Stator inner radius (R_{is})	56 mm
Rotor outer radius (R_{ro})	55.65 mm
Rotor shaft radius (R_{shaft})	17.5 mm
Airgap length (g)	0.35 mm
Stack length	132 mm
No. stator slots (N_s)	36
No. rotor slots (N_r)	28

are fixed as shown in Table IV. R_1 and X_1 can be calculated using the following expressions:

$$R_1 = \frac{L_c W}{A_{co} a_1} \times R_{co} \tag{11}$$

$$X_1 = \frac{2\mu_o \omega L_c W^2}{pq(\lambda_s + \lambda_{ds} + \lambda_{es})} \tag{12}$$

where L_c is the stator coil length, W is the number of turns in stator windings, A_{co} is the area of a conductor, a_1 is the number of parallel current paths, R_{co} is the resistivity of the copper, p is the number of poles, and q is the number of stator slot per pole per phase; λ_s, λ_{ds} and λ_{es} are the stator slot, differential and end ring coefficients, respectively.

The stator slot coefficient (λ_s) can be calculated as follow:

$$\lambda_s = \left[\frac{2H_s}{3\,(B_{s1} + B_{s2})} + \frac{2H_w}{(B_{os} + B_{s1})} + \frac{H_{os}}{B_{os}}\right] \times \left(\frac{1 + 3\beta}{4}\right) \tag{13}$$

β is known as chording factor, which takes the values in range as follows:

$$\frac{2}{3} \leq \beta \ \leq \ 1 \tag{14}$$

The stator differential coefficient (λ_{ds}) is calculated using the following expression:

$$\lambda_{ds} = \frac{0.9\tau_s q^2 {K_w}^2\,(0.18\sin\phi + 1.24)}{1.21g(1+K_{st})} \times \left[1 - 0.033\ \frac{{B_{os}}^2}{g\tau_s}\right] \quad (15)$$

$$\phi = \pi\,(6\beta - 5.5) \quad (16)$$

where τ_s is the slot pitch of the stator, K_w accounts for winding factor, and K_{st} is the saturation factor. The stator end ring coefficient (λ_{es}) can be obtained from the following:

$$\lambda_{es} = 0.34\ \frac{q}{L}\,(L_{es} - 0.64\ \beta\tau) \quad (17)$$

where L is the stack length, L_{es} is the stator end ring length, and τ is the pole pitch. R_1 and X_1 are calculated using the values given in the Table IV, and they take the values 0.94 and 2.16 Ω, respectively.

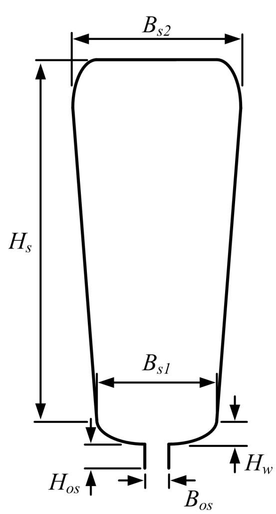


Fig. 4. Design of stator slot

TABLE IV. STATOR SLOT PARAMETER SETTINGS

Parameters	Length(mm)
B_{os}	2.5
B_{s1}	5.4
B_{s2}	9.2
H_{os}	2
H_w	1.5
H_s	20.4

B. Rotor Slot Design

The geometry parameters for rotor bar design are shown in Fig. 5. Resistance and reactance of the rotor can be derived from the rotor bar design parameters. The rotor bar design parameters are rotor slot opening height (Hor), rotor slot height (Hr), rotor slot tooth width (Btr), rotor slot opening width (Bor), rotor slot lower width (B_{r1}), and rotor slot upper width (B_{r2}). In this study, Bor is fixed to a constant value at 1.5 mm because the number of rotor slot remains the same throughout the work, causing constant magnetization effect [11]. In addition, B_{r1} and B_{r2} depend on the Hor, Hr and Btr as expressed in (18) and (19).

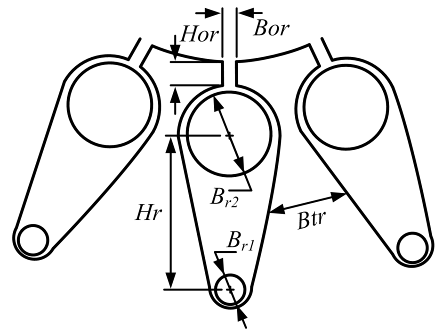


Fig. 5. Design of rotor slot

$$B_{r1} = \frac{\pi(Bor - 2Hor) - N_r Btr}{\pi + N_r} \quad (18)$$

$$B_{r2} = B_{r1} + 2Hr\tan(\frac{\pi}{N_r}) \quad (19)$$

Then, rotor resistance (R_2) and rotor leakage reactance (X_2) can be calculated using the following expressions:

$$R_2 = \frac{4m(WK_w)^2 R_{co}}{N_r}\left[\frac{LK_R}{A_b} + \frac{L_{er}}{2A_{er}\sin{(\frac{\pi p}{N_r})}^2}\right] \quad (20)$$

$$X_2 = \frac{4m(WK_w)^2\omega\mu_o L}{N_r(\lambda_r K_x + \lambda_{dr} + \lambda_{er})} \quad (21)$$

where m is the number of phases, K_w is the winding factor, K_R is skin effect coefficient for resistance, A_b is the area of rotor slot, L_{er} is the rotor end ring length, A_{er} is the area of rotor end ring, and K_x is the skin effect coefficient for rotor reactance. λ_r, λ_{dr} and λ_{er} are the permeance values of the rotor slot, differential and end ring coefficients, respectively.

The end ring and skin effects on the resistance and reactance of the rotor are also considered in the rotor resistance and reactance calculation. Equation (20) represents total resistance of the rotor including end ring part resistance and resistance due to the rotor slot [20].

IV. RESULTS AND DISCUSSION

This work focuses on the design of rotor slot, where Bor is fixed. Table V shows the dimension constraints of rotor slot to ensure that it does not violate the considered size of the induction motor as given in Table III. As a reference, the initial values of Hor, Btr and Hr are set to 0.5 mm, 4.25 mm and 14 mm, respectively the same as that in [11]. Fig. 6 depicts the variation of efficiency over increasing motor speed. It clearly shows that the motor synchronous speed is 1800 rpm at 60 Hz frequency. On the contrary, the efficiency at the rated speed (1757 rpm) is observed at 89.7%. The effect of changing the rotor slot parameters (Hor, Hr and Btr) on motor efficiency will be discussed in detail in the following subsections:

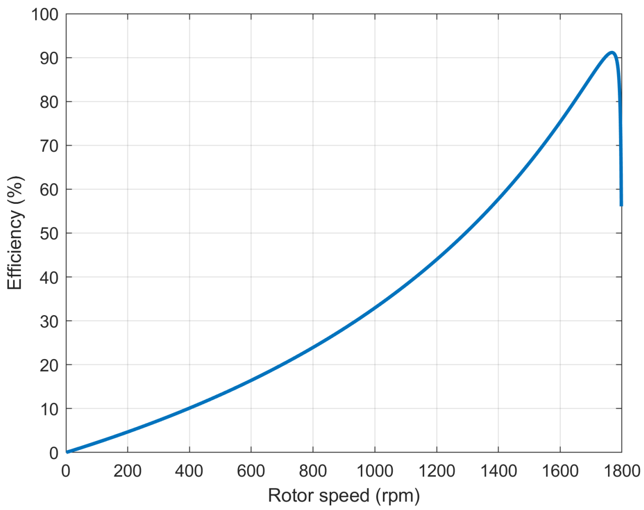


Fig. 6. Performance of the induction motor at initial settings

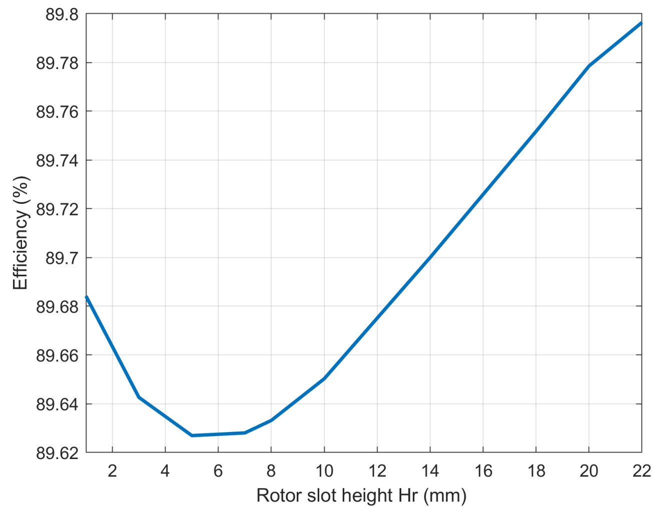


Fig. 7. Individual variation of efficiency when changing Hr

TABLE V. ROTOR SLOT GEOMETRY PARAMETER SETTINGS

Parameters	Length(mm)	
	Min	Max
Hor	0.5	5
Btr	2	5.5
Hr	1	22

A. Effect of Changing Rotor Slot Height

In this case study, the rotor slot height, Hr is varying within the limits as given in Table V, whereas the other parameters are maintained in the initial condition. Fig. 7 shows the variations of efficiency when changing Hr between 1 and 22 mm. Evidently, the efficiency increases with increasing Hr value from the initial point at 14 mm. A significant improvement in the effective area of the rotor slot can be obtained at higher value of Hr which reduces the rotor resistance. Copper losses related to the rotor slot also decreased due to the lower resistance, resulting in higher efficiency. When reducing Hr below 14 mm, the efficiency decreases up to 89.63% at $Hr = 5$ mm, and then rebounds. In this case, the effect of increasing the rotor resistance continues when decreasing Hr. However, the lower value of Hr causes the depth of the rotor slot to reduce, leading to the flux lines coupled with the stator lines. As a result, the leakage reactance of the rotor is reduced; therefore, the efficiency is improved. At a certain point, the leakage reactance becomes more dominant than the effect of increment on the rotor resistance (might be due to less effect on the effective area of the rotor slot). As can be seen from the figure, the highest value of efficiency (89.79%) can be obtained when Hr is at the upper bound, at 22 mm. At this setting, the efficiency can be improved by 0.09% as compared with the initial condition.

B. Effect of Changing Rotor Slot Opening Height

Fig. 8 shows the variations of efficiency when the rotor slot opening height, Hor varies between 0.5 and 5 mm. In this case study, Hr and Btr are maintained at the initial condition at 14 and 4.25 mm, respectively. As shown in the figure, efficiency slightly decreases until $Hor = 2.5$ mm, and then, it steeply increases to the highest value at 89.715% when $Hor = 5$ mm. The efficiency can be improved up to 0.015% compared with the initial condition for individual Hor settings. It is important to notice that Hor provides less impact on the efficiency than Hr. This finding is mainly because the Hor is located very close to the stator surface area in which the flux lines through the Hor are almost coupled with the stator flux lines. Therefore, significant improvement in efficiency is difficult to obtain when changing Hor.

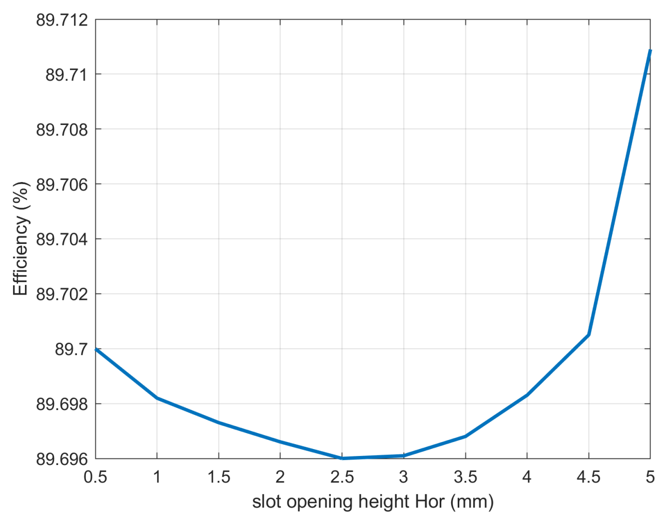


Fig. 8. Individual variation of efficiency when changing Hor

C. Effect of Changing Rotor Slot Tooth Width

Fig. 9 depicts the variation of efficiency when varying the rotor slot tooth width, Btr. In this case study, Btr is varied between 2 and 5.5 mm, whereas Hr and Hor are maintained at the initial condition at 14 and 0.5 mm, respectively. The range of Btr is based on the number of rotor slot in Table III. The figure shows that variations of efficiency have completely different pattern from the previous results, where the efficiency drastically drops in the middle because reactance and resistance become dominant at this value. The highest efficiency at 89.78% can be achieved when Btr is set to 3.75 mm. The efficiency can be improved by 0.08% as compared with the initial condition.

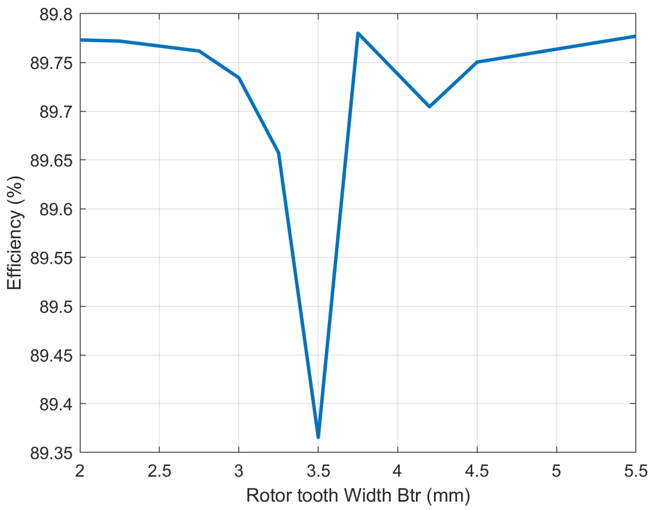


Fig. 9. Individual variation of efficiency when changing Btr

D. Effect of Changing All Selected Parameters

In previous subsections, the effect of individual parameters on efficiency have been discussed, but they could not explain the overall effect of changing Hr, Hor, and Btr simultaneously. Therefore, a 4D plot is used in this subsection to explain the effect of changing all selected rotor slot parameters. Fig. 10 shows a 4D scatter plot of efficiency when changing all the three parameters. The figure shows that the efficiency reaches maximum value at 89.8% when Hr, Hor and Btr at the maximum settings or more specifically at 22, 5, and 5-5.5 mm, respectively as highlighted by red circle. When all parameters are at maximum settings, the overall area covered by the rotor slot is increased and leads to a significant reduction in rotor resistance. Therefore, copper losses can be reduced, and the efficiency is improved. Among the three parameters, Hr provides the highest influence to the efficiency improvement, followed by Btr as shown by arrows in the figure.

V. Conclusion

This paper presents an analytical approach to investigate the influence of rotor slot geometry parameters on motor efficiency. An analytical model of three-phase squirrel cage induction motor was developed within a MATLAB environment for the investigation purposes. Several case studies were carried out and they can be categorized into individual and amalgamated changing of rotor slot parameters. In the individual case studies, the effects of changing each individual rotor slot geometry parameters namely, rotor slot height (Hr), rotor

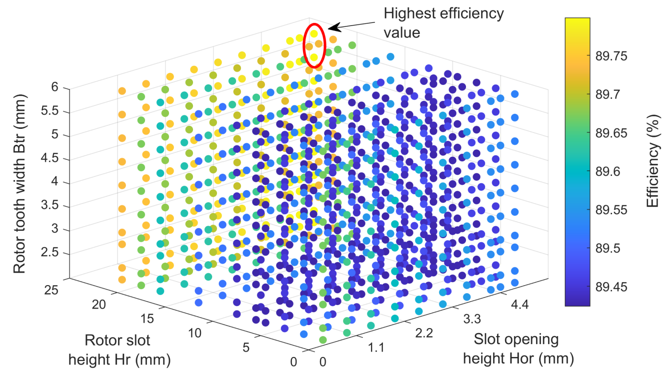


Fig. 10. 4D Scatter Plot of Efficiency when Changing All Parameters

slot opening height (Hor) and rotor slot tooth width (Btr) were analyzed. The results show that the maximum efficiency can be obtained up to 89.79%, 89.715%, and 89.78% (0.09%, 0.015%, and 0.08% improvement from the initial condition) when changing the individual parameter of Hr, Hor and Btr, respectively. However, the results cannot sufficiently explain how much the improvement can be achieved when considering all parameters. Therefore, the three parameters were changed and analyzed together in the amalgamated case study. In this case, the efficiency can be increased up to 0.1% compared with the initial condition or achieve the maximum value at 89.8%. It clearly shows that the performance of a squirrel cage induction motor can be improved substantially when the selected rotor slot geometry parameters are set appropriately. However, this approach is limited to only three parameters as presented in this paper. Thus, an optimization approach can be used in the future to solve the problem when considering more parameters.

Acknowledgment

The authors would like to thank Universiti Kebangsaan Malaysia for funding the research under grant GGPM-2019-031.

References

[1] A. Boglietti, A.M. Elrefaie, O. Drubel, A.M. Omekanda, N. Bianchi, E.B. Agamloh, M. Popescu, A.D. Gerlando and J.B. Bartolo. "Electrical machine topologies: Hottest topics in the electrical machine research community," IEEE Industrial Electronics Magazine, vol. 8, no. 2, pp. 18-30, 2014.

[2] Q. Ding, Z. Yang, X. Sun, Q. Zhao and H. Zhu. "Analysis of rotor slot width influence on a bearingless induction motor," Computers and Electrical Engineering, vol. 81, pp. 1-16, 2019.

[3] M. Sellami and S. Tounsi. "Control of axial flux DC motor with permanent magnet dedicated to electric traction," International Journal of Electrical Components and Energy Conversion, vol. 1, no. 1, pp. 44-48, 2015.

[4] R.J. Swaraj and T. Archana. "Comparison of electric motors for electric vehicle applications," International Journal of Research in Engineering and Technology, vol. 6, no. 9, pp. 12-17, 2017.

[5] W. Zhikun, C.T.Z.E. Wood, H. Shaojia, W. Hongtao and X. Tao. "Challenges faced by electric vehicle motors and their solutions," IEEE Access, vol. 9, pp. 5228-5249, 2020.

[6] E. Makhetha, M. Muteba and D.V. Nicolae. "Effect of rotor bar shape and stator slot opening on the performance of three phase squirrel cage induction motors with broken rotor bars," Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa, 2019, pp. 463-468.

[7] K.A.I. Daut, K. Anayet, N. Gomesh, M. Asri and Y.M. Irwan. "Comparison of copper rotor bars with aluminium rotor bars using FEM software: A performance evaluation," International Conference on Computer and Electrical Engineering, 2009, pp. 453-456.

[8] V.A. Galindo, X.M. Lopez-Fedez, J.A.D. Pinto and A.P. Coimbra. "Parametric study of rotor slot shape on a cage induction motor," International Symposium on Electromagnetic Fields in Electrical Engineering, 2002, pp. 190-195.

[9] K.N. Gyftakis and J. Kappatou. "The impact of the rotor slot number on the behaviour of the induction motor," Advances in Power Electronics, pp. 837010, 2013.

[10] J. Kappatou, K. Gyftakis and A. Safacas "FEM study of the rotor slot design influences on the induction machine characteristics," Advanced Computer Techniques in Applied Electromagnetics, vol. 30, pp. 247-252, 2007.

[11] E. Maloma, M. Muteba and D. Nicolae. "Effect of rotor bar shape on the performance of three phase induction motors with broken rotor bars," International Conference on Optimization of Electrical and Electronic Equipment, 2017, pp. 364-369.

[12] A. Leicht and K. Makowski. "Influence of shape and material of rotor bars on performance characteristics of single-phase self-excited induction generators," International Journal for Computation and Mathematics in Electrical and Electronic Engineering, vol. 38, no. 4, pp. 1235-1244, 2019.

[13] I. Daut, K. Anayet and A. Fauzi. "Development of copper rotor of AC induction motor," Australian Journal of Basic and Applied Sciences, vol. 4, no. 12, pp. 5941-5946, 2010.

[14] J. Singh, K. Singh and H. Kaur. "Designing of three phase squirrel cage induction motor for good efficiency," International Journal of Engineering and Innovative Technology, vol. 5, no. 7, pp. 40-45, 2016.

[15] A.G. Yetgin, M. Turan, B. Cevher, A.İ. Çanakoğlu and A. Gün. "Squirrel cage induction motor design and the effect of specific magnetic and electrical loading coefficient," International Journal of Applied Mathematics Electronics and Computers, vol. 7, no. 1, pp. 1-8, 2019.

[16] J.M. Apsley, S. Williamson, A.C. Smith and M. Barnes. "Induction motor performance as a function of phase number," IEE Proceedings: Electric Power Applications, vol. 153, no. 6, pp. 898-904, 2006.

[17] M.A. Iqbal and V. Agarwal. "Investigation & analysis of three phase induction motor using finite element method for power quality improvement," International Journal of Electronic and Electrical Engineering, vol. 7, no. 9, pp. 901-908, 2014.

[18] I. Boldea and S.A. Nasar. "The induction machines design handbook," Taylor and Francis Group, 2010.

[19] D.J. Kim, J.H. Choi, Y.D. Chun, D.H. Koo and P.W. Han. "The study of the stray load loss and mechanical loss of three phase induction motor considering experimental results," Journal of Electrical Engineering and Technology, vol. 9, no. 1, pp. 121-126, 2014.

[20] A. Boglietti, A. Cavagnino and M. Lazzari. "Geometrical approach to induction motor design," The 33rd Annual Conference of the IEEE Industrial Electronics Society, 2007, pp. 149-156.

BBVD: A BERT-based Method for Vulnerability Detection

Weichang Huang[1], Shuyuan Lin[2]*, Chen Li[3]
College of Information Science and Technology / the College of Cyber Security, Jinan University, Guangzhou, China[1]
School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China[2,3]

***Abstract*—Software vulnerability detection is one of the key tasks in the field of software security. Detecting vulnerability in the source code in advance can effectively prevent malicious attacks. Traditional vulnerability detection methods are often ineffective and inefficient when dealing with large amounts of source code. In this paper, we present the BBVD approach, which treats high-level programming languages as another natural language and uses BERT-based models in the natural language processing domain to automate vulnerability detection. Our experimental results on both SARD and Big-Vul datasets demonstrate the good performance of the proposed BBVD in detecting software vulnerability.**

***Keywords*—Vulnerability detection; BERT; software security**

I. INTRODUCTION

Software vulnerability is security flaw that exists in software [1]. When developers develop software, security issues such as logic error, buffer overflow, array out-of-bound access and other errors can easily occur in the source code. Vulnerability can be exploited for privilege escalation, leakage of secret data, denial of service and many other types of attacks [2], which put the integrity and availability of the software and the computer system at risk. The size and complexity of modern software are increasing dramatically, which also diversifies the types and causes of vulnerability. Thus, the traditional vulnerability detection methods are facing new challenges.

Traditional source code vulnerability detection depends on manual code auditing [3], relying on the expertise of the security personnel to review the software source code. However, in the face of the massive amount of code in complex software, the workload of manual audit vulnerability is large and the effectiveness of detection relies strongly on the amount of a priori knowledge of security personnel. It is unrealistic to rely entirely on manual audits to discover existing types and possible variants of vulnerability types. Now, automated vulnerability detection is becoming an important supplement to manual audits.

There are many similarities between high-level programming languages and natural language [4], [5]. High-level programming languages such as C/C++ that inherit the syntax of natural language have a defined syntax and semantics. Long sentences in natural language consist of words and phrases. A programming language also consists of a series of instructions. If these instructions are treated as words in a natural language, different combinations of instructions produce more complex operations, which correspond to long sentences in natural language. The choice of different words, the difference in word placement and the contextual connection all affect the semantics of sentences in natural language. In the case of programming language, the placement and combination of instructions also affect the logic of the programming language. Therefore, considering these similarities, we can use language models in NLP for software vulnerability detection.

The BERT-based model allows for parallelized computation in the model. Also, multiple experimental results show that BERT-based models [6]–[9] have outperformed existing models in natural language processing, including recurrent neural network (RNN) architectures. Therefore, BERT-based language models are more effective than RNNs and we choose to use BERT-based models to detect software vulnerability. Our contributions in this paper are as follows:

- We propose a method called BBVD for detecting vulnerability in C/C++ source code by using BERT-based models.
- We verify that the BERT-based language models used in the BBVD approach can be migrated to areas outside of the natural processing domain such as vulnerability detection.
- Our experimental results on both SARD [10] and Big-Vul [11] datasets show that BBVD outperforms the state-of-the-art method such as SySeVR [12].

The rest of the paper is organized as follows: Section I presents the related works on BERT-based models and software vulnerability detection methods. Section III describes the proposed BBVD method and the flow for the software vulnerability detection. Section IV describes the experimental setup of datasets and BERT-based models used in our work. Section V reports the experimental results and performance of the BERT-based models. Section VI concludes our paper.

II. BACKGROUND AND RELATED WORK

A. BERT beyond NLP

In recent years, some classical methods have yielded good performance using model optimization strategies to process data containing noise and outliers, such as graph-based [13], [14] and information- and cue-based [15], [16] strategies. Furthermore, neural networks can be more efficient by fusing attention mechanisms when processing large amounts of input information. These networks using attentional mechanisms have been used in a wide range of tasks such as image text matching [17], visual sentiment analysis [18], video question

*Corresponding authors.

answering [19], generative adversarial task [20], multi-task travel route planning [21], etc.

Attention mechanisms are also used in BERT to focus each element in the input sequence on other elements. The paper [6] demonstrates that BERT pre-training facilitates almost all types of NLP tasks (except generative models). The question-and-answer (QA) domain is an important area of NLP. BERT in the field of QA is more effective than the previous methods, and the experimental results of some papers [22]–[25] in question retrieval and answer determination show that a significant improvement can be achieved by using BERT or attention mechanisms. The application of BERT to the reading comprehension task also had a huge impact on the various original techniques [26]. The effectiveness of works [27] using BERT in document retrieval tasks similarly demonstrates the usefulness of BERT in the field of information retrieval. In addition to these domains, BERT has also shown significant enhancements in areas such as text summarization [28], text classification [29], [30] etc.

The BERT can also be migrated to domains other than natural language processing. Inspired by pre-training models in the BERT model, Kanade et al. [31] proposed the CuBERT (CodeUnderstandingBERT) model designed with two pre-training tasks of predicting masked tokens and whether two logical lines of code are related to each other in a contextual sentence. The CodeBERT model proposed by Feng et al. [4] used bimodal data of natural and programming languages. They pre-trained CodeBERT in six programming languages and showed optimal performance on the NL-PL downstream task. The above two code pre-training models only consider codes as token sequences, ignoring the structural information in the codes. Guo et al. [32] proposed the GraphCode-BERT model using the data flow information of codes in the pre-training process, which significantly improved the performance of four downstream tasks, including code search, code clone detection, code translation and code summarization.

There have been studies using the BERT model to detect software vulnerability. Ziems et al. [33] fine-tuned the 100,000 C/C++ source code files on a pre-trained model known as BERT base and tested the fine-tuned model with over 100 types of vulnerability. Their work shows the feasibility of BERT for vulnerability detection. Although our approach is also based on BERT related model, there are several differences compared to their approach: (i) Their input is source code files without comments. Ours is the code slice that can better represent the syntax and semantic information of the program; (ii) They use pre-trained model which based on an English Wikipedia dataset containing 2.5 billion words, we use the C/C++ code slice to pre-train and then generate a pre-trained model; (iii) They consider only BERT, we have experimented on other BERT-based models; (iv) They only experimented on the synthetic dataset SARD, we do experiments on both the synthetic dataset SARD and the real dataset Big-Vul.

B. Vulnerability Detection Methods

Source code vulnerability detection methods can be divided into two categories: static and dynamic methods. Static methods do not require the execution of source code and can detect vulnerability by combining information from source code files, control flow graph [37], program dependency graph [42], LLVM IR [38], etc. Dynamic methods obtain the execution path and data flow of program crashes by executing the code file to obtain the heap, stack, registers and other information of the actual running program. Typical dynamic methods are fuzz testing and taint analysis, these methods applied to large project source code files have problems such as low path coverage and path explosion, which require a lot of computational resources. Therefore, we focus on static methods.

Static detection methods can be divided into two categories: traditional methods and machine learning based methods. Traditional detection methods are mainly based on code similarity and pattern match. Code similarity-based detection is suitable for vulnerability caused by code clone. It has a high false negative when the cloned code itself is not vulnerable but other code fragments are flawed. There are four types of code similarity-based detection methods: text-based, lexicon-based [34]–[36], syntax-based [39]–[41] and semantic-based [42]–[45]. Text-based and lexicon-based methods are simple to implement and can detect the source code of almost all programming languages. However, recognizing only textual information or lexical information will lose the syntactic and semantic information of the source code, which leads to a low accuracy rate. Syntax-based and semantic-based methods can identify syntactic and semantic information of programs with high detection accuracy. But as program size increases, the subtree matching algorithm for detecting abstract syntax tree similarity and the subgraph matching algorithm for detecting program dependency graph similarity grow dramatically in time and space complexity.

In contrast, pattern-based approaches require a large number of experts to define rules for matching vulnerability manually. The subjectivity of experts may lead to high false positives and high false negatives. For a specific vulnerability, the corresponding vulnerability patterns can be extracted. The pattern matching method describes these vulnerability patterns in a specific syntax. Then the source code is compared and detected after a series of processing. Methods that use pattern matching to detect vulnerability include PMD, Coverity Prevent [46], etc.

Researchers have started applying machine learning to detect software vulnerability due to increased computer computing power. Yamamoto et al. [47] applied machine learning algorithms such as naive bayes model to the vulnerability descriptions of the National Vulnerability Database (NVD) [48] for vulnerability detection. Their proposed method shows good prediction performance on a dataset containing more than 6000 vulnerable source codes. Toloudis et al. [49] combine techniques related to text analysis and principal component analysis to process the vulnerability descriptions of the NVD and perform experiments on a large dataset containing 70,678 vulnerable source code. Spanos et al. [50] use classical machine learning methods on the same dataset to predict software vulnerability, the experimental results achieved an accuracy of about 80%.

As the number of public vulnerabilities grow, it becomes possible to use deep learning models to extract abstract feature representations from vulnerability samples. Not all information in the source code file is helpful for software vulnerability detection. Too much irrelevant information may interfere with

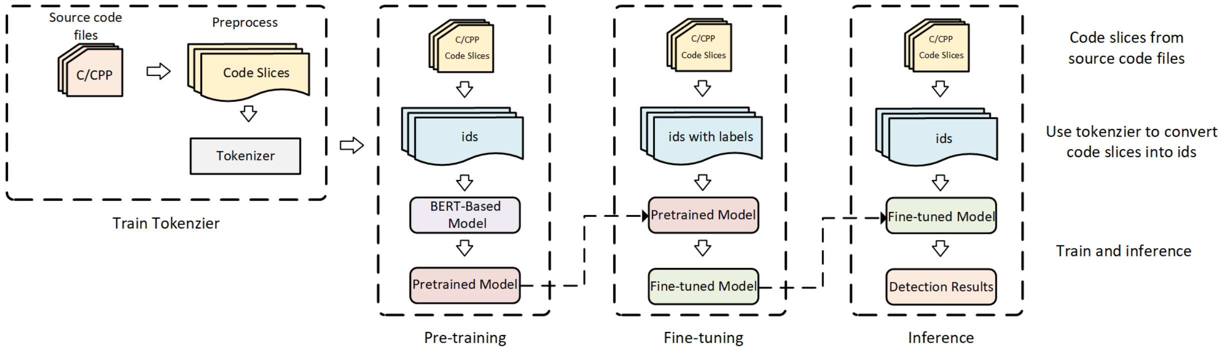


Fig. 1. Framework of BBVD

prediction. Program slicing techniques are widely used in vulnerability detection tasks which can obtain code fragments related to vulnerability. Li et al. [5] proposed the concept of code gadgets to serve as an intermediate representation of source code for obtaining semantic information related to vulnerability. They predicted the existence of vulnerability related to API and library function calls in programs based on code gadgets and BiLSTM. Li et al. [12] proposed code slices based on code gadgets combined with program dependency graphs. The code slices can represent the vulnerability syntax and semantic features of the program and fed into the BiGRU model to obtain vulnerability detection results.

In addition to using models in NLP to detect software vulnerability, some studies use graph neural networks to detect vulnerability. The Devign model proposed by Zhou et al. [51] uses an abstract syntax tree as a skeleton. The model incorporates the control flow and data flow information to generate a joint graph representation. A gated graph neural network and a convolutional neural network model are used for graph classification to detect vulnerable code. Chakraborty et al. [52] proposed ReVeal using a combination of gated graph neural networks and multilayer perceptrons. Their experiments showed that adding multilayer perceptron modules to gated graph neural networks helps in vulnerability detection.

III. Design of BBVD

In this section, we present a brief overview of the BBVD method. Fig. 1 illustrates the detection pipeline which trains from scratch. We divide the whole process into four phases: training the tokenizer, pre-training phase, fine-tuning phase and inference phase. In the first stage, we extract code slices from the source code files and then train the tokenizer based on the tokens in the code slices. The second stage converts the code slices into ids using the tokenizer and pretrains on the ids using a BERT-based model. The ids are token indices, numerical representations of tokens building the sequences that will be used as input by BERT-based models. The third stage labels the code slices and defines the downstream task as the classification for fine-tuning. The fourth stage uses datasets not identified by the model for inference.

A. Train Tokenizer Phase

In the stage of training tokenizer, we need to extract code slices from the source code files. Specifically, code slices are syntax-based vulnerability candidates (SyVCs) and semantics-based vulnerability candidates (SeVCs) which proposed by Li et al. [12]. In order to get the code slices, we need to use the open-source code analysis tool joern to parse the source code files and obtain the corresponding control flow graph and program dependency graph. Then, parsing out the call graph of function based on the CFG and PDG to extract array-related, pointer-related, api-related and arithmetic-related code slices. Furthermore, keeping the original name of keywords in the programming language and mapping the user-defined variable and function names one-to-one to symbolic names (e.g. "variable_0", "variable_1", "func_0", "func_1") to reduce the interference of user-defined function names and variable names. Fig. 2 shows different types of code slices extracted from the same source file. Finally, a tokenizer is generated based on these code slices.

B. Pretrain Phase

During the pre-training phase, the code slices of the pre-training dataset are transformed into ids using the tokenizer generated in the first phase. Fig. 3 shows an example of converting code slices and ids to each other. The ids are filled or truncated according to the max position embeddings in the model network parameter config. That is, when the length of ids is less than the max position embeddings, special token ids (e.g. "$\langle pad \rangle$" correspond id is 1) are used to fill in the right side of ids. When the length of ids exceeds the max position embeddings, the redundant part to the right of the ids is truncated. The BERT-based model trained with unsupervised learning to learn the context of code slices and their syntax and semantics. We use language masking model approach to train the BERT-based model. The language masking model masks some of the input locations based on probabilities (the original token is replaced with "$\langle mask \rangle$") and trains the model to predict the masked parts using the remaining parts. Meanwhile, the model is optimized based on the original masked tokens and the predicted ones.

C. Finetune and Inference Phase

After pre-training is completed, we need to perform the downstream task on the pre-trained model. Since vulnerability detection can be regarded as a binary classification, we define the downstream task as a classification task. Similarly, the code slices extracted from the fine-tuned dataset are transformed

```
void CWE121_Stack_Based_Buffer_Overflow__CWE806_char_declare_memcpy_03_bad()
{
    char * data;
    char dataBuffer[100];
    data = dataBuffer;
    if(5==5)
    {
        /* FLAW: Initialize data as a large buffer that is larger than the small
        memset(data, 'A', 100-1); /* fill with 'A's */
        data[100-1] = '\0'; /* null terminate */
    }
    {
        char dest[50] = "";
        /* POTENTIAL FLAW: Possible buffer overflow if data is larger than dest *
        memcpy(dest, data, strlen(data)*sizeof(char));
```

Raw Source Code

```
void CWE121_Stack_Based_Buffer_Overflow__CWE806_char_declare_memcpy_03_bad() 23
char * data ; 25
char dataBuffer [ 100 ] ; 26
data = dataBuffer; 27
if ( 5 == 5 )  28
memset ( data , 'A' , 100 - 1 ); 31
data [ 100 - 1 ] = '\0'; 32
memcpy ( dest , data , strlen ( data ) * sizeof ( char ) ); 37
dest [ 50 - 1 ] = '\0'; 38
printLine ( data ); 39
void printLine (const char * line) 10
if ( line != NULL )  12
printf ( "%s\n" , line ); 14
```

API Call Related Code Slice

```
void CWE121_Stack_Based_Buffer_Overflow__CWE806_char_declare_memcpy_03_bad()
char * data ; 25
data = dataBuffer; 27
memset ( data , 'A' , 100 - 1 ); 31
data [ 100 - 1 ] = '\0'; 32
memcpy ( dest , data , strlen ( data ) * sizeof ( char ) ); 37
dest [ 50 - 1 ] = '\0'; 38
printLine ( data ); 39
void printLine (const char * line) 10
if ( line != NULL )  12
printf ( "%s\n" , line ); 14
```

Pointer Related Code Slice

```
void CWE121_Stack_Based_Buffer_Overflow__CWE806_char_declare_memcpy_03_bad() 23
char dataBuffer [ 100 ] ; 26
data = dataBuffer; 27
memset ( data , 'A' , 100 - 1 ); 31
data [ 100 - 1 ] = '\0'; 32
memcpy ( dest , data , strlen ( data ) * sizeof ( char ) ); 37
dest [ 50 - 1 ] = '\0'; 38
printLine ( data ); 39
void printLine (const char * line) 10
if ( line != NULL )  12
printf ( "%s\n" , line ); 14
```

Array Related Code Slice

Fig. 2. Code slices

```
void func_0 ( ) wchar_t * variable_0 ; variable_0 = NULL ; while ( 1 ) wchar_t * variable_1 = ( wchar_t * ) malloc ··· wprintf ( variable_2 " % ls \ n ", variable_4 )
```

Encode code slice to ids

```
0, 437, 279, 67, 20, 270, 269, 268, 67, 20, 272, 268, 67, 21, 271, 268, 67, 21, 273, 357, 271, 285, 270, 268, 67, 22, 269, 268, 67, 21, 273, 270, 299, 419, 20, 2, 1 ··· 1
```

Decode ids to code slice with special tokens

```
<s>void func_0 ( ) wchar_t * variable_0 ; variable_0 = NULL ; while ( 1 ) wchar_t * variable_1 = ( wchar_t * ) malloc ···  " % ls \ n ", variable_4 )</s><pad>···<pad>
```

Fig. 3. Interconversion between the ids and code slices

into ids using the trained tokenizer. In order to make the classification task work properly, we need to add the appropriate labels to the ids at this phase. Label 0 means the code slice is a normal sample and label 1 means a vulnerable sample. Then, the fine-tuning dataset is split into a training set and an evaluation set. The pre-trained model is fine-tuned on the training set and evaluated on the evaluation set. Finally, generating the fine-tuned model. Using the fine-tuned model can inference whether the source code file is vulnerable or not. Repeat the steps of extracting code slices and transforming them into ids. Input the ids into the fine-tuned model can get the detection results.

IV. Experiment Setup

We designed follow three Research Questions (RQs).

- RQ1: Do the BERT-based models work better than the RNN-based models for vulnerability detection?
- RQ2: Do the use of irrelevant datasets in the pre-training and fine-tuning phases interfere with the detection performance of the model?
- RQ3: Do the different value of max embedding lengths (e.g., the max lengths of ids) have an influence on the detection effect of BERT-based models?

We implement BERT-based models by using the huggingface [53] with PyTorch 1.8.0. The computer running the experiment has an NVIDIA GeForce GTX 2080TI GPU and 16 Intel(R) Xeon(R) Gold 5117 CPUs running at 2.00 GHz.

A. *Datasets*

We experiment on two vulnerability datasets: SARD and Big-Vul. SARD is a synthetic dataset in which most of the testcases are written for academic and research purposes to reproduce specific vulnerability. SARD contains a batch of testcases composed of code in C/C++. These testcases are classified into different CWEs (Common Weakness Enumeration) according to the associated vulnerability types. They are classified into good, bad, and mixed types according to whether the code file is a vulnerability, among which mixed means that the code file contains both the vulnerability and the fix code. For the testcases of bad and mixed types containing vulnerability, the additional information about the line number where the vulnerability is located is also given. The entire sard dataset consists of 10,682 cpp source code and 24,633 c source code files.

Big-Vul dataset is a real dataset which collects testcases from the commit records related to CVEs(Common vulnerability and Exposures) that have appeared in the git repositories of 310 well-known C/C++ open source projects between 2002 and 2019 including Linux, Chrome, etc. All the data is provided in the form of CSV and each testcase containing information, such as the project name, associated CVE number, description, code content, and function differences before and after the commit. All vulnerability are categorized into ten types such as denial of service attacks, memory corruption, and privilege escalation. Each testcase labeled as vulnerable is one of the combinations of these types. To facilitate extracting code slices, the samples in Big-Vul need to be filtered. The function differences in commit contain the corresponding code lines which change after the patch. We select the data where the number of code lines change by less than 100. The total number of filtered data is 16,608, of which 16,436 are c source code files and 172 are cpp source code files. Due to the complexity of the vulnerability types in the Big-Vul dataset, we called the memory corruption-related vulnerability data from filtered Big-Vul dataset as Big-Vul-MemCorr and the entire filtered Big-Vul dataset as Big-Vul-All.

B. Models

We conducted experiments on five models. Specifically, this experiment contains two RNN-based models (BiGRU and BiLSTM) and three BERT-based models (RoBERTa, DistilBERT and MobileBERT). Table I shows some important network structure-related related parameters of BERT-based models used in this paper.

TABLE I. NETWORK STRUCTURE RELATED PARAMETERS

Model	Hidden Size	Hidden Layers	Attention Heads	Intermediate Size	Max Position Embeddings
RoBERTa	768	12	12	3072	514
DistilBERT	768	6	12	3072	512
MobileBERT	512	24	4	512	512

RoBERTa. This model mainly based on BERT with several adjustments: longer training time with larger batch size and more training data; longer training sequence; dynamic adjustment of masking mechanism.

DistilBERT. It's proposed for the most popular BERT pre-training model with a 40% reduction in model size and 60% faster inference operations while retaining 97% of the performance.

MobileBERT. It compresses the BERT model which reduces the model size by a factor of three to four and increases the speed by a factor of four to five with little loss of effect, allowing a variety of NLP applications to be easily deployed on mobile.

C. Evaluation Metrics

The effectiveness of vulnerability detection can be evaluated using the following metrics: accuracy (*A*), precision (*P*), recall (*R*), F1-score (*F1*) and Mathews Correlation Coefficient (*MCC*). True positive (*TP*) indicates the number of samples with vulnerability detected. False Positive (*FP*) indicates the number of samples without vulnerability but detected as such. True Negative (*TN*) indicates the number of samples without vulnerabilit detected. False Negative (*FN*) denotes the number of samples with vulnerability but detected as not having vulnerability. The metric $A = \frac{TP+TN}{TP+FP+TN+FN}$ is the number of correctly predicted samples out of all the samples. The metric $P = \frac{TP}{TP+FP}$ represents all the samples that are declared to be vulnerable but what percentage of them are actually vulnerable. The metric $R = \frac{TP}{TP+FN}$ represents all the samples that are actually vulnerable but what percentage declared vulnerable. The metric $F1 = 2 * \frac{P*R}{P+R}$ is used to measure test accuracy, which is a weighted average of the precision and recall. The *F1* score is 1 when it's best and on 0 when it's worst. The $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ returns a value between negative one and positive one. A coefficient of positive one represents a perfect prediction, zero means no better than random prediction and negative one indicates total disagreement between prediction and actual label.

V. EXPERIMENTS AND RESULTS

In this section, we conduct experiments on SARD, Big-Vul-All, Big-Vul-MemCorr datasets based on the design of BBVD described in section III. Comparative experiments are conducted using the models and evaluation metrics described in section IV.

A. Preprocess

1) Code slices:: For the SARD, Big-Vul-All and Big-Vul-MemCorr datasets, we extracted four types of code slices, corresponding to the API function call, array usage, pointer usage and arithmetic expression, respectively.

- The API function call related code slices: We extract 135,145 from SARD, 68,686 from Big-Vul-All and 6,381 from Big-Vul-MemCorr.
- Array usage related code slices: We extract 56,346 from SARD, 11,566 from Big-Vul-All and 4,953 from Big-Vul-MemCorr.
- Pointer usage related code slices: We extract 318,912 from SARD, 161,596 from Big-Vul-All and 4,525 from Big-Vul-MemCorr.
- The arithmetic expression related code slices: We extract 7,810 from SARD, 16,842 from Big-Vul-All and 5,192 from Big-Vul-MemCorr.

Since there are still duplicates in the above extracted code slices, we need to de-duplicate these data. For the SARD dataset, the API function call related code slices are reduced from 135,145 to 69,432; the array usage related code slices are reduced from 56,346 to 24,680; the pointer related code slices are reduced from 318,912 to 128,492; the arithmetic related code slices are reduced from 7,810 to 6,956. The total number of non-duplicate data in the SARD dataset is 229,560, of which 44,447 are vulnerable and 185,113 are non-vulnerable. For the Big-Vul-All dataset, the API function call related code slices are reduced from 68,686 to 29,865; the array usage related code slices are reduced from 11,566 to 6,151; the pointer related code slices are reduced from 161,596 to 101,804; the arithmetic related code slices are reduced from 16,842 to 15,476. The total number of non-duplicate data in the Big-Vul-All dataset is 153,287, of which 19,908 are vulnerable and 133,379 are non-vulnerable. Table II describes the number of extracted code slices for the SARD, Big-Vul-All and Big-Vul-MemCorr datasets.

2) Tokenizer:: Tokenization is an important step in natural language processing which breaks long texts such as sentences, paragraphs, and articles down into word-based data structures for subsequent processing and analysis work. In order to translate code slices into vectors that can be recognized by the BERT-based models, we need a tokenizer to translate them into vectors which can be recognized by BERT-based models.

TABLE II. NUMBER OF CODE SLICES EXTRACTED FROM DATASETS

Datasets	Slices Type	Total Slices	De-duplicated Slices	Vul Slices	Non-Vul Slices
SARD	API Related	135145	69432	12786	56646
SARD	Array Related	56346	24680	7195	17485
SARD	Pointer Related	318912	128492	11680	116812
SARD	Arithmetic Related	7810	6956	12786	56646
Big-Vul-All	API Related	68686	29865	4919	24946
Big-Vul-All	Array Related	11566	6151	325	5826
Big-Vul-All	Pointer Related	161596	101804	13406	88398
Big-Vul-All	Arithmetic Related	16842	15476	1258	14191
Big-Vul-MemCorr	API Related	6381	2064	538	1526
Big-Vul-MemCorr	Array Related	4953	1067	204	863
Big-Vul-MemCorr	Pointer Related	4525	3154	657	2587
Big-Vul-MemCorr	Arithmetic Related	5192	2911	440	2471

Specifically, we used two tokenizers in experiments, one based on byte-pair encoding and the other based on wordpiece. DistilBERT and MobileBERT model use tokenizer based wordpiece. RoBERTa use tokenizer based byte-pair encoding.

- Tokenizer based byte-pair encoding (BPE). BPE is an algorithm for encoding based on byte pairs. The algorithm is described as a cascading iterative process in which the most frequent pair of characters in a string is replaced by a character that does not appear in the string.
- Tokenizer based wordpiece. The wordpiece method is very similar to BPE in general, except that when selecting characters for merging, BPE uses the highest frequency, while wordpiece uses the highest probability.

B. Experiments for RQ1

To answer RQ1, we use SySeVR to perform the vulnerability detection on the SARD and Big-Vul datasets, respectively. We use the detection results as a benchmark. The BERT-based models are then used to perform the detection on the same datasets and the detection results are compared with the benchmark. In detail, the BERT-based models used in RQ1 is the same as that described in section IV. All BERT-based models used in this paper are pretrained for *10 epochs* with *learning_rate=1e-04* and fine-tuned for *10 epochs* with *warmup_steps=1000, learning_rate=1e-05, weight_decay=0.1*. The loss functions used for both pre-training and fine-tuning phase are cross-entropy loss.

TABLE III. DETECTION RESULTS OF SARD (THE METRICS UNIT: %)

Model	Dataset	A	P	R	F1	MCC
BGRU(SySeVR)	SARD	95.16	**89.63**	86.36	87.97	84.96
BLSTM(SySeVR)	SARD	95.06	88.65	86.99	87.81	84.72
RoBERTa	SARD	**95.42**	85.25	**93.90**	**89.37**	**86.63**
DistilBERT	SARD	95.39	88.01	89.73	88.86	85.97
MobileBERT	SARD	95.02	84.56	92.56	88.38	85.36

Table III shows the detection results of 2 RNN-based models and 3 BERT-based models on the SARD dataset. The entire SARD dataset is divided into a training set and a test set after randomly sorted, with four-fifths of the training set and one-fifth of the test set. The RNN-based model is trained with the training set and tested with the test set. The BERT-based model is pre-trained with the entire SARD dataset, fine-tuned with the training set, inferred with the test set, and the inference results are compared with the benchmark.

In terms of F1, and the MCC metric, all the three BERT-based models outperform RNN-based models. Among the BERT-based models, RoBERTa has the best result. The corresponding F1 and the MCC metric is 93.90%, 89.37%, and 86.63%, respectively. All the BERT-based models outperform RNN-based models in accuracy metrics except MobileBERT. As for the precision metric, although the RNN-based model exceeds the BERT-based models, the recall rate of BERT-based models outperform RNN-based models. That is, the BERT-based model has a lower miss detection rate. Also, considering the metrics F1 and MCC, the overall detection effect of the BERT-based model is better than that of the RNN-based model.

TABLE IV. DETECTION RESULTS OF BIG-VUL-ALL (THE METRICS UNIT: %)

Model	Dataset	A	P	R	F1	MCC
BGRU(SySeVR)	Big-Vul-All	70.69	**13.09**	**14.30**	**13.67**	-3.94
BLSTM(SySeVR)	Big-Vul-All	72.41	12.90	12.18	12.53	-3.82
RoBERTa	Big-Vul-All	83.96	9.97	9.32	9.63	0.85
DistilBERT	Big-Vul-All	**84.23**	11.16	10.32	10.72	**2.09**
MobileBERT	Big-Vul-All	83.97	9.24	8.46	8.83	0.07

Table IV shows the detection results on the Big-Vul-All dataset. Compared to the detection results on the synthetic dataset SARD, the detection results show a huge drop. This means that real vulnerabilities have complex syntax and semantic infomations and it is impractical to experiment on Big-Vul-All, which covers multiple vulnerabilities. Therefore, we focus on one vulnerability type called memory corruption and perform experiments on the Big-Vul-MemCorr dataset.

TABLE V. DETECTION RESULTS OF BIG-VUL-MEMCORR (METRICS UNIT: %)

Model	Dataset	A	P	R	F1	MCC
BGRU(SySeVR)	Big-Vul-MemCorr	69.90	42.55	26.08	32.34	15.11
BLSTM(SySeVR)	Big-Vul-MemCorr	70.50	44.73	29.56	35.60	18.12
RoBERTa	Big-Vul-MemCorr	71.94	48.87	37.82	**42.64**	**24.82**
DistilBERT	Big-Vul-MemCorr	66.78	39.46	**38.26**	38.85	16.06
MobileBERT	Big-Vul-MemCorr	**72.18**	**49.13**	24.78	32.94	19.39

Table V shows the results of several models on Big-Vul-MemCorr. Because the vulnerability types in Big-Vul-MemCorr dataset all fall into one category, the detection results are much better than those for Big-Vul-All. Of these five models, the one works best is the RoBERTa model. Its accuracy reaches 71.94%, precision reaches 48.87%, recall reaches 37.82%, F1 reaches 42.64% and MCC reaches 24.82%. Such results also illustrate that, despite the complexity of the real vulnerability types, it is feasible to use the BERT-based model for detection if only one vulnerability type is targeted.

C. Experiments for RQ2

To answer RQ2, we use different combinations of three datasets (SARD, Big-Vul-All and Big-Vul-MemCorr) described in section IV in the pre-training phase and fine-tuning phase. Similarly, the BERT-based model is also shown in sectionIV. The combination of datasets has the following five combinations (COMBs).

- COMB1: Pre-training phase and fine-tuning phase using SARD dataset.

- COMB2: Pre-training phase using Big-Vul-MemCorr dataset and fine-tuning phase using Big-Vul-MemCorr dataset.
- COMB3: Pre-training phase using Big-Vul-All dataset and fine-tuning phase using Big-Vul-MemCorr dataset.
- COMB4: Pre-training phase using SARD dataset and fine-tuning phase using Big-Vul-MemCorr dataset.
- COMB5: Pre-training phase using Big-Vul-All dataset and fine-tuning phase using SARD dataset.

TABLE VI. DETECTION RESULTS OF COMBS (THE METRICS UNIT: %)

Max Length	COMBs	Metrics	RoBERTa	DistilBERT	MobileBERT
512	COMB1	A	**95.42**	95.39	95.02
		P	85.25	**88.01**	84.56
		R	**93.90**	89.73	82.56
		F1	**89.37**	88.86	88.38
		MCC	**86.63**	85.97	85.36
512	COMB2	A	71.94	66.78	**72.18**
		P	48.87	39.46	**49.13**
		R	37.82	**38.26**	24.78
		F1	**42.64**	38.85	32.94
		MCC	**24.82**	16.06	19.39
512	COMB3	A	**69.42**	67.38	69.18
		P	**43.16**	39.60	42.19
		R	34.34	**34.78**	31.73
		F1	**38.25**	37.03	36.22
		MCC	**18.49**	15.21	16.73
512	COMB4	A	**70.02**	64.14	62.82
		P	**42.64**	35.44	32.75
		R	25.21	**36.52**	33.04
		F1	31.69	**35.97**	32.90
		MCC	**14.88**	11.08	7.19
512	COMB5	A	**95.34**	95.09	95.16
		P	**94.11**	92.62	90.54
		R	82.41	82.59	**85.29**
		F1	**87.87**	87.32	87.84
		MCC	**85.29**	84.50	84.88

Table VI shows the results of using different combinations of datasets in the pre-training and fine-tuning phases. For the comparison of COMB1 and COMB5 or the comparison of COMB2 and COMB4, we can see that using the same dataset for the pre-training and fine-tuning phases is helpful for the vulnerability detection. All three BERT-based models have slightly higher F1 on COMB1 than on COMB5 and on COMB2 than on COMB4. In the comparison between COMB2 and COMB3, the two COMBs use different datasets in the pre-training phase. The COMB2 uses Big-Bul-MemCorr, while COMB3 uses Big-Vul-All. Both the RoBERTa and DistilBERT models have better F1 and MCC metrics on COMB2 than COMB3. This illustrates that the dataset used in the pre-training phase can affect the detection results. Big-Vul-All covers multiple vulnerability types, while Big-Vul-MemCorr covers only one vulnerability type. If the model learns too many vulnerability features in the pre-training phase it may affect the detection results of a single vulnerability.

D. Experiments for RQ3

To answer RQ3, we set the maximum sequence length for the BERT-based model to two lengths: 512 and 1024.

Table VII shows the results of the model for vulnerability detection using different maximum sequence lengths. For SARD, the total number of tokens in code slices less than or equal to 512 is 142,303 and less than or equal to 1024 is 144,526. For Big-Vul-MemCorr, the total number of tokens in code slices less than or equal to 512 is 2,430 and less than or equal to 1024 is 3,135. If the maximum sequence length is set to 512, the code slice will be truncated when the number of tokens in the code slice is greater than 512. As a result, these code slices lose some of their syntactic and semantic information. However, the results on COMB2 show that it is not the case that the larger the maximum sequence length, the better the detection. The model does not detect as well at a maximum sequence length of 1024 as it does at a maximum sequence length of 512. At a maximum sequence length of 1024, the F1 of the RoBERTa model on COMB2 is about 5 percent lower than at a maximum sequence length of 512.

TABLE VII. DETECTION RESULTS WITH DIFFENENT MAX LENGTHS OF IDS (THE METRICS UNIT: %)

Max Length	COMBs	Metrics	RoBERTa	DistilBERT	MobileBERT
512	COMB1	A	**95.42**	95.39	95.02
		P	85.25	**88.01**	84.56
		R	**93.90**	89.73	82.56
		F1	**89.37**	88.86	88.38
		MCC	**86.63**	85.97	85.36
1024	COMB1	A	95.40	**95.57**	95.25
		P	93.33	**95.34**	93.72
		R	**83.52**	82.41	82.35
		F1	88.16	**88.40**	87.67
		MCC	85.52	**86.04**	85.02
512	COMB2	A	71.94	66.78	**72.18**
		P	48.87	39.46	**49.13**
		R	37.82	**38.26**	24.78
		F1	**42.64**	38.85	32.94
		MCC	**24.82**	16.06	19.39
1024	COMB2	A	**69.30**	68.22	64.62
		P	**43.36**	40.93	39.05
		R	36.95	34.34	**50.43**
		F1	39.90	37.35	**44.02**
		MCC	**19.58**	16.39	19.10

VI. CONCLUSION

In this paper, we propose BBVD by studying BERT-based models for software vulnerability detection. Specifically, the proposed BBVD uses C/C++ code for pre-training, fine-tuning, and inferencing to detect vulnerability. Our results show that these BERT-based models outperform existing RNN-based models such as BiGRU and BiLSTM for vulnerability detection. There are still issues that can be investigated, such as detection results on real datasets are not as good as on synthetic datasets. Future work should use more real vulnerability datasets to improve the effectiveness of the BERT-based models in detecting vulnerability.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of P. R. China (Nos. 62002068).

REFERENCES

[1] Dowd M, McDonald J, Schuh J. The art of software security assessment: Identifying and preventing software vulnerability[M]. Pearson Education. 2006: 1-1129.

[2] Plate H, Ponta S E, Sabetta A. Impact assessment for vulnerability in open-source software libraries[C]. IEEE International Conference on Software Maintenance and Evolution. 2015: 411-420.

[3] Bacchelli A, Bird C. Expectations, outcomes, and challenges of modern code review[C]. International Conference on Software Engineering. 2013: 712-721.

[4] Feng Z, Guo D, Tang D, et al. CodeBERT: A pre-trained model for programming and natural languages[C]. Findings of the Association for Computational Linguistics. 2020: 1536-1547.

[5] Li Z, Zou D, Xu S, et al. VulDeePecker: A deep learning-based system for vulnerability detection[J]. Network and Distributed System Security Symposium. 2018: 1-15.

[6] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. In Proceedings of NAACL-HLT. 2019: 4171-4186.

[7] Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.

[8] Sun Z, Yu H, Song X, et al. Mobilebert: A compact task-agnostic bert for resource-limited devices[J]. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 2158-2170.

[9] Sanh V, Debut L, Chaumond J, et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter[J]. arXiv preprint arXiv:1910.01108, 2019.

[10] Accessed Oct 15, 2022. Software assurance reference dataset (SARD). https://samate.nist.gov/SARD/

[11] Fan J, Li Y, Wang S, et al. A C/C++ code vulnerability dataset with code changes and CVE summaries[C]. International Conference on Mining Software Repositories. 2020: 508-512.

[12] Li Z, Zou D, Xu S, et al. SySeVR: A framework for using deep learning to detect software vulnerability[J]. IEEE Transactions on Dependable and Secure Computing. 2021: 1-15.

[13] Lin S, Xiao G, Yan Y, Suter D, Wang H. Hypergraph optimization for multi-structural geometric model fitting[C]. In Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 8730-8737.

[14] Lin S, Luo H, Yan Y, Xiao G, Wang H. Co-clustering on Bipartite Graphs for Robust Model Fitting[J]. IEEE Transactions on Image Processing. 2022, 31: 6605-6620.

[15] Lin S, Wang X, Xiao G, Yan Y, Wang H. Hierarchical representation via message propagation for robust model fitting[J]. IEEE Transactions on Industrial Electronics. 2020, 68(9): 8582-8592.

[16] Yang H, Lin S, Cheng L, Lu Y, Wang H. SCINet: Semantic Cue Infusion Network for Lane Detection. IEEE International Conference on Image Processing. 2022, 1811-1815.

[17] Huang F, Zhang X, Zhao Z, et al. Bi-directional spatial-semantic attention networks for image-text matching[J]. IEEE Transactions on Image Processing. 2018, 28(4): 2008-2020.

[18] Xu J, Li Z, Huang F, et al. Visual sentiment analysis with social relations-guided multiattention networks[J]. IEEE Transactions on Cybernetics. 2020, 52(6): 1-13.

[19] Liu Y, Zhang X, Huang F, et al. Cross-Attentional Spatio-Temporal Semantic Graph Networks for Video Question Answering[J]. IEEE Transactions on Image Processing. 2022, 31: 1684-1696.

[20] Huang F, Jolfaei A, Bashir A K. Robust multimodal representation learning with evolutionary adversarial attention networks[J]. IEEE Transactions on Evolutionary Computation. 2021, 25(5): 856-868.

[21] Huang F, Xu J, Weng J. Multi-task travel route planning with a flexible deep learning framework[J]. IEEE Transactions on Intelligent Transportation Systems. 2020, 22(7): 3907-3918.

[22] Yang W, Xie Y, Lin A, et al. End-to-end open-domain question answering with BERTserini[C]. Conference of the North American Chapter of the Association for Computational Linguistics. 2019: 72-77.

[23] Li C, Pang B, Liu Y, et al. Adsgnn: Behavior-graph augmented relevance modeling in sponsored search[C]. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 223-232.

[24] Pang B, Li C, Liu Y, et al. Improving Relevance Modeling via Heterogeneous Behavior Graph Learning in Bing Ads[C]. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 3713-3721.

[25] Bi S, Li C, Han X, et al. Leveraging Bidding Graphs for Advertiser-Aware Relevance Modeling in Sponsored Search[C]. Findings of the Association for Computational Linguistics: EMNLP. 2021: 2215-2224.

[26] Qu C, Yang L, Qiu M, et al. BERT with history answer embedding for conversational question answering[C]. International ACM SIGIR conference on research and development in information retrieval. 2019: 1133-1136.

[27] Zhan J, Mao J, Liu Y, et al. An analysis of BERT in document ranking[C]. International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1941-1944.

[28] Grail Q, Perez J, Gaussier E. Globalizing BERT-based transformer architectures for long document summarization[C]. Conference of the European chapter of the association for computational linguistics: Main volume. 2021: 1792-1810.

[29] Munikar M, Shakya S, Shrestha A. Fine-grained sentiment classification using BERT[C]. Artificial Intelligence for Transforming Business and Society. 2019, 1: 1-5.

[30] Sun C, Qiu X, Xu Y, et al. How to fine-tune bert for text classification?[C]. China National Conference on Chinese Computational Linguistics. 2019: 194-206.

[31] Kanade A, Maniatis P, Balakrishnan G, et al. Learning and evaluating contextual embedding of source code[C]. International Conference on Machine Learning. 2020: 5110-5121.

[32] Guo D, Ren S, Lu S, et al. Graphcodebert: Pre-training code representations with data flow[J]. arXiv preprint arXiv:2009.08366, 2020.

[33] Ziems N, Wu S. Security Vulnerability Detection Using Deep Learning Natural Language Processing[C]. Conference on Computer Communications Workshops. 2021: 1-6.

[34] Baker B S. A program for identifying duplicated code[J]. Computing Science and Statistics. 1993: 49-49.

[35] Ducasse S, Rieger M, Demeyer S. A language independent approach for detecting duplicated code[C]. International Conference on Software Maintenance. 1999: 109-118.

[36] Gitchell D, Tran N. Sim: A utility for detecting similarity in computer programs[J]. ACM Sigcse Bulletin. 1999, 31(1): 266-270.

[37] Allen F E. Control flow analysis[J]. ACM Sigplan Notices. 1970, 5(7): 1-19.

[38] VenkataKeerthy S, Aggarwal R, Jain S, et al. Ir2vec: Llvm ir based scalable program embeddings[J]. ACM Transactions on Architecture and Code Optimization. 2020, 17(4): 1-27.

[39] Kim Y C, Cho Y Y, Moon J B. A plagiarism detection system using a syntax-tree[C]. International Conference on Computational Intelligence. 2004, 1: 23-26.

[40] Jiang L, Misherghi G, Su Z, et al. Deckard: Scalable and accurate tree-based detection of code clones[C]. International Conference on Software Engineering. 2007: 96-105.

[41] Baxter I D, Yahin A, Moura L, et al. Clone detection using abstract syntax trees[C]. International Conference on Software Maintenance. 1998: 368-377.

[42] Ferrante J, Ottenstein K J, Warren J D. The program dependence graph and its use in optimization[J]. ACM Transactions on Programming Languages and Systems. 1987, 9(3): 319-349.

[43] Liu C, Chen C, Han J, et al. GPLAG: detection of software plagiarism by program dependence graph analysis[C]. International Conference on Knowledge Discovery and Data Mining. 2006: 872-881.

[44] Pham N H, Nguyen H A, Nguyen T T, et al. Complete and accurate clone detection in graph-based models[C]. International Conference on Software Engineering. 2009: 276-286.

[45] Sheneamer A, Roy S, Kalita J. A detection framework for semantic code clones and obfuscated code[J]. Expert Systems with Applications. 2018, 97: 405-420.

[46] Kim Y, Kim M, Kim Y J, et al. Industrial application of concolic testing approach: A case study on libexif by using CREST-BV and KLEE[C]. International Conference on Software Engineering. 2012: 1143-1152.

[47] Yamamoto Y, Miyamoto D, Nakayama M. Text-mining approach for estimating vulnerability score[C]. International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security. 2015: 67-73.

[48] Accessed Oct 15, 2022. National Vulnerability Database. https://doi.org/10.18434/M3436

[49] Toloudis D, Spanos G, Angelis L. Associating the Severity of vulnerability with their Description[C]. International Conference on Advanced Information Systems Engineering. 2016: 231-242.

[50] Spanos G, Angelis L, Toloudis D. Assessment of vulnerability severity using text mining[C]. Proceedings of the Pan-Hellenic Conference on Informatics. 2017: 1-6.

[51] Zhou Y, Liu S, Siow J, et al. Devign: Effective vulnerability identification by learning comprehensive program semantics via graph neural networks[J]. Advances in Neural Information Processing Systems. 2019, 32: 1-11.

[52] Chakraborty S, Krishna R, Ding Y, et al. Deep learning based vulnerability detection: Are we there yet[J]. IEEE Transactions on Software Engineering. 2021, 48(9): 3280-3296.

[53] Wolf T, Debut L, Sanh V, et al. Transformers: State-of-the-art natural language processing[C]. Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020: 38-45.

Hybrid Deep Learning Architecture for Land Use: Land Cover Images Classification with a Comparative and Experimental Study

Salhi Wiam[1], Tabiti Khouloud[2], Honnit Bouchra[3], SAIDI Mohamed Nabil[4], KABBAJ Adil[5]
Research Laboratory in Information Systems, Intelligent Systems and Mathematical Modeling,
National Institute of Statistics and Applied Economics,
Rabat, Morocco[1,2]
LPRI Multidisciplinary Research and Innovation Laboratory, Moroccan School of Engineering Sciences EMSI,
Casablanca, Morocco[3]
Research Laboratory in Information Systems, Intelligent Systems and Mathematical Modeling,
National Institute of Statistics and Applied Economics, Rabat, Morocco[4,5]

***Abstract*—Deep Learning algorithms have become more popular in computer vision, especially in the image classification field. This last has many applications such as moving object detection, cancer detection, and the classification of satellite images, also called images of land use-land cover (LULC), which are the scope of this paper. It represents the most commonly used method for decision making in the sustainable management of natural resources at various geographical levels. However, methods of satellite images analysis are expensive in the computational time and did not show good performance. Therefore, this paper, on the one hand, proposes a new CNN architecture called Modified MobileNet_V1 (MMN) based on the fusion of MobileNet_V1 and ResNet50. On the other hand, it presents a comparative study of the proposed model and the most used models based on transfer learning, i.e. MobileNet_V1, VGG16, DenseNet201, and ResNet50. The experiments were conducted on the dataset Eurosat, and they show that ResNet50 results emulate the other models.**

Keywords—Deep Learning; image classification; land use-land cover; MobileNet; ResNet; satellite images

I. Introduction

Nowadays, the artificial intelligence domain is knowing an important development, especially in image classification. It is one of the most common challenges in computer vision [1]. Indeed, it refers to the categorization of images into one of many predetermined classes. A single image can be classified into a number of different categories. However, manually inspecting and classifying images can be time-consuming, especially when the images have a large size. Thus, automating the process would be extremely beneficial [2].

Image classification is used in several domains [3], [4] e.g. medicine, videos surveillance, economy, and agriculture especially for the classification and categorization of Land Use-Land Cover (LULC) images. Moreover, at the local, regional, and national levels, the LULC classification plays an important role in program planning, management, and monitoring [5]. On the one hand, The LULC information helps in understanding the land occupation issues, and on the other hand, it is crucial in the constitution of policies and programs that are necessary for development planning. Furthermore, it is important to track the LULC pattern's evolution throughout time in order to ensure long-term development. To accomplish sustainable urban growth and to control the random expansion of cities, urban development authorities need such planning models that allow every available piece of land to be used in the most reasonable and optimal way possible. It requires knowledge of the area's current and previous LULC. In addition, the LULC maps can be used to track changes in our ecosystem and environment [6], [7], [8].

Currently, image processing of LULC using deep learning algorithms is gaining more attention [9], [10], [11]. Since, they outperformed the classical approaches, even if the interpretation of LULC images requires good agricultural experience.

The main objective of this work is to study the efficiency of deep convolutional neural network (CNN) architectures for LULC image classification. Thus, our work aims to solve the following research questions: (1) Is there any deep learning technique that consistently outperforms the other ones? (2) Is it possible to use deep learning to correctly categorize LULC images and surpass previous methods? (3). What is the highest level of accuracy that DL can attain with LULC images ?

The contributions of our paper are the following: we design a comparison between different fine-tuned DL architectures (VGG16, DenseNet201, ResNet50 and MobileNet_V1) and a Modified MobileNet_V1 (MMN) architecture that combines MobileNet_V1 and ResNet50 on several levels namely: performance, and amount of parameters etc. Since there is a lack of LULC image datasets, the different models were evaluated on the Eurosat dataset [4] of LULC images and on the same dataset using data augmentation techniques. The rest of this paper is organized as follows: The second section presents the literature review, the section three provides the used methods and materials. Section four exposes the experimental results. Section five provides the discussion, and the paper is concluded in the last section.

II. Literature Review

A. LULC Classification Datasets

There are several LULC classes, such as developed or urban areas, farmland, and wooded areas, and so on. LULC maps have many applications such as conservation of natural resources, entry of GIS data, delineation of tax and property boundaries, etc. Unfortunately, the existing and published datasets are rarely labeled, which makes the classification process very challenging. Thus, some datasets have been proposed, but each of them has a disadvantage. For example, the UC Merced dataset [12], [13], [14], [15], [16], which has been suggested by [16], contains 21 land use land cover classes with 100 images per class and is extracted from USGS National Map Urban Area Imagery. However, this dataset is a bad choice for a comparison based on deep learning models because it has a very small number of images.

Although PatternNet [17] and NWPU-RESISC [18] have high resolutions, they share the same issue of having 100 images per class. Additionally, SAT-6, which was developed by [19], has six classes: barren land, trees, grassland, roads, buildings, and water, and has a resolution of 1 m/pixel and a size of 28×28 and the images were produced using imagery from the National Agriculture Imagery Program (NAIP). Moreover, the AID Dataset, which was first presented by [15], includes 30 classes with 200-400 images each with a size of 600×600.

B. LULC Classification Methods

Several methods have been proposed in the field of LULC classification. There are the traditional methods based on color analysis, shape and texture, or a combination of all this information [20], [21], but their effectiveness in the real world is still constrained. To solve the limitations of the old methods, numerous modeling tools, including dynamic, statistical methodologies, have been employed to create accurate simulations that consider economic, spatiotemporal and social factors [22]. Remote sensing imaging classification, anomaly detection, and prediction issues can be resolved using machine learning modeling. Maximum likelihood classifiers, Markov chain models, support vector machines, Markov chain models and other machine learning algorithms have been historically used to classify images [23], [7], [8], [11], [24]. But, in recent years, with the advancement in the performance of computing units, several approaches in the context of deep learning have been used. In order to provide superior performance for the treatment and classification of images (categorical image mapping and classification using convolutional neural networks (CNN)), deep learning models outperform traditional models in extracting spatial characteristics at various levels of remote sensing images [25]. A method in [10] presents an architecture called Joint Deep Learning (JDL), which integrates a multilayer perceptron (MLP) and a convolutional neural network (CNN) and is implemented via a Markov process.

In [6], Scale Sequence Joint Deep Learning (SS-JDL) is introduced as a new DL method for LULC classification. The effectiveness of this SS-JDL method has been tested on the digital aerial photography of three complex and heterogeneous landscapes. Thus, In [26] a simple discriminative CNN (D-CNN) method is proposed to improve the performance of remote sensing image scene classification. Furthermore, the spectral and spatial information content of remote sensing images is captured and utilized in [27] and employing a multi-attention method using a bidirectional long-term memory network. To address the imbalance in LULC classes, a recent work in [28] recommended the use of oversampling, while in [29], the authors suggest using consensus-based collaborative multi-label learning to balance labels. To explain the model predictions in [30], the authors tested various interpretable artificial intelligence methodologies while using the DenseNet model [31]. The framework for learning deep representations of spectral bands for the same purpose is finally proposed in [32].

The deep CNNs utilized in this paper were improved using a pre-trained network in the context of deep learning. The ILSVRC-2012 image classification challenge dataset served as the main source of pre-training data for the networks. Even though these pre-trained networks were developed on images from a completely unrelated field, the features generalized effectively [33], [34]. As a result, it was discovered that the pre-trained networks were appropriate for classifying remote sensing images. Consequently, we use for comparison some deep CNN models like VGG16 [35], ResNet50 [36], MobileNet_V1 [37] and DenseNet201 [31] which represent the state of the art for the classification of the introduced LULC classes.

III. Materials and Methods

Recent studies have proved that DL algorithms are very efficient in image processing and computer vision [38]. They have been applied in various remote sensing imaging modalities with high performance [39] in segmentation, detection and classification. Although these technologies have shown promising results in remote sensing, they require a lot of data. Motivated by the success of DL and remote sensing image processing, our work will present a comparative study between a new CNN architecture that combines the MobileNet_V1 architecture with the concept that characterizes the ResNet50 architecture and different DL architectures (VGG16, ResNet50, MobileNet_V1 and DenseNet201) (application in the LULC domain).

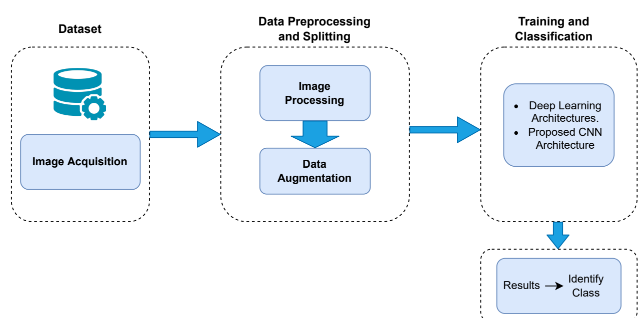


Fig. 1. Process of LULC Classification

Fig. 1 illustrates the used processes to compare the various models. The first step is image acquisition. Then, the preprocessing and splitting of data. Finally, training, and classification.

A. LULC classification

LULC data indicates the extent of an area covered by forests, wetlands, impervious surfaces, agriculture, and other land and water types [40]. It refers to the categorizing or classification of human activities and natural landscape elements over a period of time using scientific methodologies.

With this information, it is possible to better understand the effects of natural phenomena and human use of the landscape [40]. Maps can be also used to assess urban growth, model water quality issues, predict flood and storm surge impacts, track wetland loss and potential sea level rise impacts, prioritize conservation efforts, and compare land-cover changes with environmental effects or linkages in socioeconomic changes like population growth [23].

In the literature, there exist several approaches of features extraction from remote sensing images. In this context, CNNs are a sort of neural network [16], that is considered as the state of the art of image classification approaches [4] due to its good results.

B. Data Preprocessing

Even if some DL architectures can handle 13 bands of images as an input, they can not use the TIFF files. Therefore, all the images should be converted to RGB format. In order to avoid over-fitting, data augmentation was applied after the preprocessing and only to the training data. Additionally, some geometrical transformations were included i.e. rescaling, rotations, shifts, shears, zooms, and flips. We also used multiple augmentation approaches to build a new image from each input image.

C. Training and Classification

1) Convolutional neural network: CNN is not merely a deep neural network with many hidden layers. It is a deep network that analyses and recognizes images based on the brain's visual cortex. This is how CNN differs from previous neural networks in terms of idea and function [41]. It also gets its name from an operation called convolution, a linear mathematical action involving matrices. CNN has several layers, including a convolutional layer, a non-linearity layer, a pooling layer, and a fully-connected layer [1].

The pooling and non-linearity layers do not contain any parameters, while the convolutional and fully-connected layers do. In terms of machine learning issues, CNN performs exceptionally well. Specifically, in applications dealing with high resolution images [42]. CNN collects features from images automatically by constructing many convolution layers [43] resulting in a feature hierarchy. The shallower front convolution layer employs a smaller perceptual domain, allowing it to learn some features of the local image, whereas the deeper back convolution layer employs a broader perceptual domain, allowing it to learn more features [1].

2) Deep learning architectures:

a) VGG16: Simonyan and Zisserman proposed the Visual Geometry Group (VGG) CNN architecture in 2014, and it won the ILSVR competition. VGG16 enhances the decision function by replacing the huge filters with numerous 3×3 filters one after the other and 2×2 for max pooling, resulting in a more discriminative decision function (decrease in the number of parameters). VGG16 is easy to understand and utilize because it only has two layers: convolution and pooling [35], [44]. There are a total of 16 weighted layers, as indicated by the number 16.

b) ResNet50: ResNet50 is a deep residual network created by K. He, X. Zhang, S. Ren, and J. Sun in 2015, and it won the ILSVRC 2015. It's a model that tackles the problem of disappearing gradients (gradient tends to zero quickly) by using a novel notion called skip connection (stacking convolution layers and adding the original input to the convolution block's output). ResNet50 is made up of five stages, each with a convolution and identity block and three convolution layers. ResNet50 uses images with a resolution of 224×224 pixels and has 50 residual networks [45], [36].

c) MobileNet_V1: MobileNet_V1 is a 28-layer architecture that accepts input images that are $224 \times 224 \times 3$ pixels in size. It introduced the notion of Depthwise Separable Convolution, in which two 1D convolutions with two kernels are used instead of a single 2D convolution to minimize the model's size (fewer parameters) and complexity (fewer multiplications and additions). MobileNet_V1 also adds two new global hyperparameters (width multiplier and resolution multiplier) that allow model developers to trade latency or accuracy for speed and compact size based on their needs [37], [46].

d) DenseNet201: The dense convolutional network (DenseNet201) is a CNN with 201 depth layers that takes a 224×224 input image. DenseNet201 is a ResNet improvement that incorporates dense layer connections. Each layer receives more inputs from the all preceding layers and delivers its own feature maps to the layers below it. Concatenation is used by each layer to obtain "collective knowledge" from all previous layers. DenseNet outperforms ordinary networks by reducing processing requirements, reducing the number of parameters, encouraging feature reuse, and improving feature propagation [31], [47].

D. Proposed Architecture Modified MobileNet_V1 (MMN)

The input layer, convolutional layers, pooling layers, fully connected layers, and the output layer are the five layers that make up a CNN model. Furthermore, a CNN model can be trained from end to end to enable feature selection, extraction and prediction or classification. It's difficult to figure out how a network understands and analyzes an image. However, features extracted by layers of a network have been demonstrated to outperform human-built features [48].

Our model is a combination of the MobileNet_V1 and ResNet50 architectures. We took the same architecture of the MobileNet_V1 which uses the notion of depthwise separable convolution, and we optimized it by adding the notion of skip connection, which characterizes the ResNet50 model. It means that we add the output of a layer to the next one. This is to try to improve the results of MobileNet_V1 by providing an alternate gradient path, and it has been experimentally proven that these additional paths are frequently beneficial for model convergence.

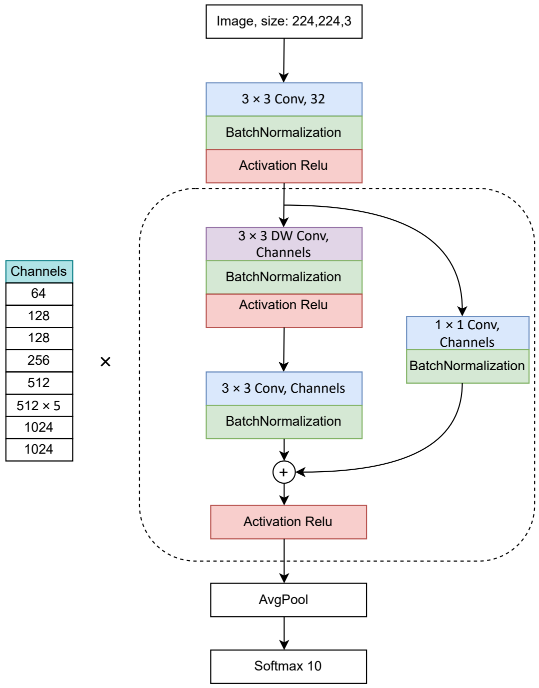


Fig. 2. Architecture of proposed CNN MMN

The proposed CNN MMN has the following architecture (Fig. 2):

- Input layer: the inputs are images with dimension (244×244).
- Convolutional layers: a convolution is a linear process that involves the input and a set of weights. It's made for two-dimensional inputs, with multiplication taking place between a two-dimensional array of weights (filters) and an array of input data. We have three sorts of layers in the proposed architecture MMN : one with a 3×3 size filter and the same padding, one with a 1×1 size filter and the same padding, and one with a 1×1 filter and valid padding.
- Depthwise separable convolution : This layer was employed in our architecture with the same padding and a 3×3 filter.
- Pooling layers: a technique for subsampling feature maps by aggregating the presence of features across feature map patches. Average pooling and maximal pooling are two different types of pooling algorithms. To calculate the average value in each patch for each feature map in the proposed design, we used avg-pooling.
- BatchNormalization: This method involves refocusing and rescaling layer inputs to normalize them and makes artificial neural networks faster and more stable.
- After each BatchNormalization, we employed Rectified Linear Unit Layers (ReLU).
- Fully connected layers: they treat input data as a single vector and produce a single vector as output.

IV. Results

A. The Used Dataset

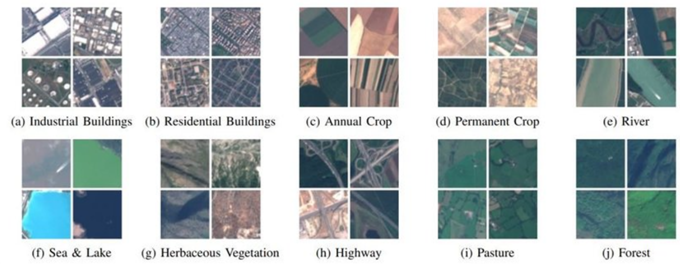


Fig. 3. Dataset classes [4]

TABLE I. Eurosat Data Distribution

Class name	Number of images
Annual Crop	3000
Forest	3000
Herbaceous Vegetation	3000
Highway	2500
Industrial	2500
Pasture	2000
Permanent Crop	2500
Residential Buildings	3000
River	2500
Sea and Lake	3000

In this study, We used the EuroSAT database of Sentinel-2 images of differents European cities. It contains 27000 of 64×64 images. It covers several classes [4] i.e., Industrial Buildings, Residential Buildings, Sea and Lake, Herbaceous Vegetation, Annual Crop, Permanent Crop, River, Highway, Pasture, Forest (Fig. 3, Table I).

TABLE II. Sentinel-2 Bands, Wavelength, and Resolution [4]

Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 – Coastal aerosol	0.443	60
Band 2 – Blue	0.490	10
Band 3 – Green	0.560	10
Band 4 – Red	0.665	10
Band 5 – Vegetation Red Edge	0.705	20
Band 6 – Vegetation Red Edge	0.740	20
Band 7 – Vegetation Red Edge	0.783	20
Band 8 – NIR	0.842	10
Band 8A – Vegetation Red Edge	0.865	20
Band 9 – Water vapour	0.945	60
Band 10 – SWIR – Cirrus	1.375	60
Band 11 – SWIR	1.610	20
Band 12 – SWIR	2.190	20

Sentinel-2 data is multispectral, with 13 bands covering the visible, near-infrared, and shortwave infrared spectrum (Table II). Since these bands are available in various spatial resolutions ranging from 10 m to 60 m, the images can be classified as high-medium resolution. Although data from other satellites with higher resolution (1 m to 0.5 cm) is available. Sentinel-2 data is free and has a long revisit duration (5 days), making it a good choice for LULC monitoring.

B. Evaluation Metrics

After training the various architectures, the final step is to evaluate the performance of the used architectures. Among the various classification performance properties, our research uses the following benchmark metrics: Accuracy (ACC), Precision (PRE), Recall, F1-score (F1) [49].

– Accuracy :
Accuracy is the proportion of true results among the total number of cases examined.

$$ACC = \frac{(TP\ +\ TN)}{(TP+\ FP+\ FN+\ TN)} \tag{1}$$

– Precision:
Precision is the proportion of predicted positives that are actually positive.

$$PRE = \frac{(TP)}{(TP+\ FP)} \tag{2}$$

– Recall:
Recall is the proportion of actual positives that are correctly classified.

$$PRE = \frac{(TP)}{(TP+\ FN)} \tag{3}$$

– F1-score:
The F1-score is a number between 0 and 1 and is the harmonic mean of precision and recall.

$$PRE = 2 \times \frac{(precision \times recall)}{(precision+\ recall)} \tag{4}$$

Where: TP stands for: True Positive. FP: False Positive.TN: True Negative, and FN: False Negative.

C. Results without Data Augmentation

Our research was conducted on a publicly available image dataset of LULC images (Eurosat Dataset [4]) and on the same dataset using data augmentation. During the experimental study, 80% of images were used for training and 20% for testing and validation, and this operation was repeated 20 times randomly, the average of the results was retained. This using the following experimental parameters for classification: All images in the dataset were 64×64 pixels, with the exception of the proposed model image, which was scaled to 224×224 pixels. Thus for the data augmentation techniques, we trained the different models with an image size of 224×224. We used a batch size of 32 and a total of 20 epochs to train the models. For optimization, Adam is used with $\beta 1$ = 0.9, $\beta 2$ = 0.999, and the learning rate is set to 0.0001. As a result, a ReduceLROnPlateau with a min_lr of 1e-20 and val_accuracy as the monitor is employed. We utilized the ReLU to train a fully connected layer, and we changed the last dense layer in all models to yield 10 classes corresponding to the distinct classes in the Eurosat database [4], rather than the 1000 classes used by ImageNet. The results of the Eurosat image classification with the following architectures (Proposed Architecture, Fine–tuning the top layers of VGG16, ResNet50, DenseNet201, and MobileNet_V1) are shown in this section. Several experiments were carried out to evaluate the performance and resilience of each given model, based on the cited metrics in Section IV-B and the confusion matrix.

TABLE III. DIFFERENT METRICS OF THE PROPOSED ARCHITECTURE MMN

Metrics	Training	Testing
Accuracy	99.09%	94.01%
F1-Score	99.08%	94.02%
Recall	98.96%	93.9%
Precision	99.2%	94.8%
Loss	0.03	0.18

1) Proposed Architecture Modified MobileNet_V1 (MMN): According to Table III, we observe that the accuracy of the training data is remarkably high compared to testing, with values of 99.09% and 94.01% respectively, and this is the case for all the other metrics (F1-score, Recall, Precision). However, the loss metric has been significantly increased from training to testing with values of 0.03 and 0.18, respectively.

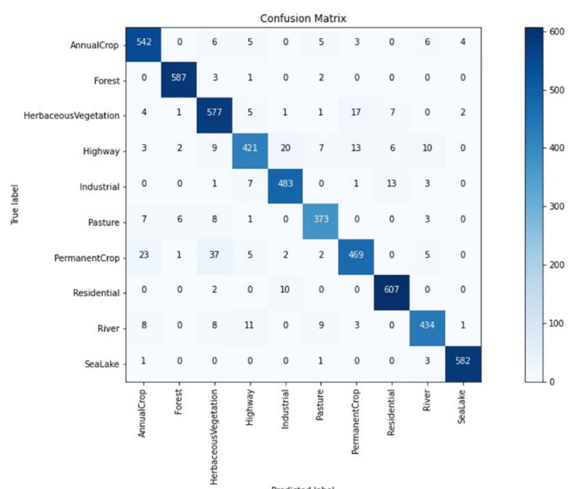


Fig. 4. Confusion matrix of proposed architecture MMN

From the confusion matrix (Fig. 4), we note that the most recognized class is “Residential” with 607 images. Therefore, for “Pasture” the model was able to identify just 373 images.

TABLE IV. DIFFERENT METRICS OF VGG16

Metrics	Training	Testing
Accuracy	100%	96.9%
F1-Score	100%	96.87%
Recall	100%	96.84%
Precision	100%	96.89%
Loss	2.8e-05	0.18

2) VGG16: Table IV shows the Accuracy, F1-score, Recall, Precision, and loss of VGG16. Indeed, the accuracy decreased from training to testing with a différence of -3.1%. The same was noticed in the other metrics except for the loss which increased from the value of 2.8e-05 for training to 0.18 for testing.

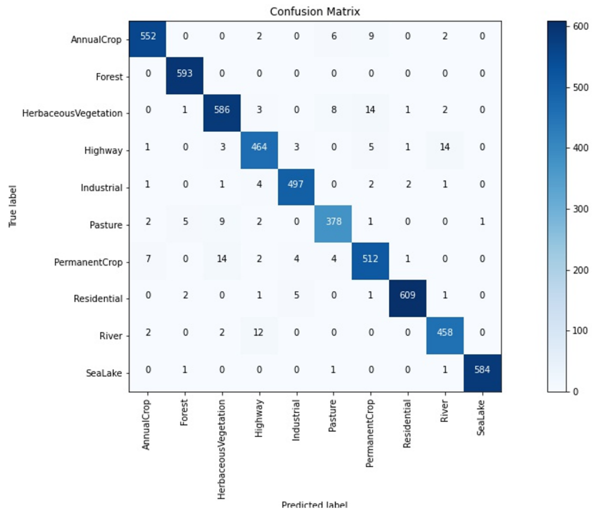


Fig. 5. Confusion matrix of VGG16

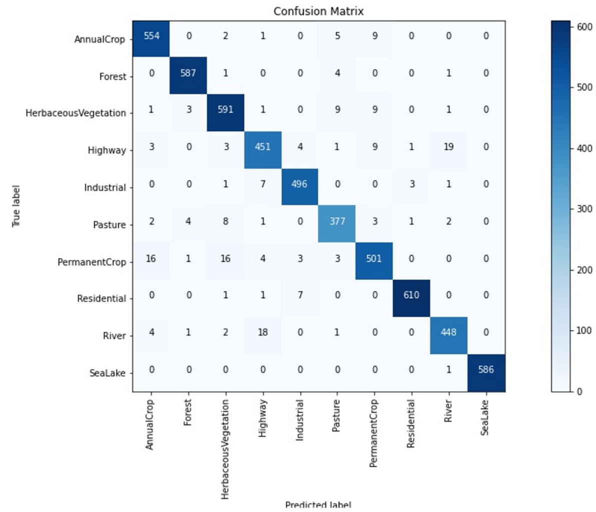


Fig. 6. Confusion matrix of ResNet50

Fig. 5 shows the confusion matrix of the VGG16 model, which illustrates that 609 images were correctly labeled as the “Residential” class, but only 378 images for “Forest” were classified by the model.

TABLE V. Different Metrics of ResNet50

Metrics	Training	Testing
Accuracy	100%	96.31%
F1-Score	100%	96.34%
Recall	100%	96.28%
Precision	100%	96.4%
Loss	1.55e-05	0.17

3) ResNet50: Table V presents the obtained results with the Resnet50 classifier. We observe an increase in the accuracy of the training compared to the testing where its values are equal to 100%, 96.31% respectively, and it is the same for the other metrics. The value of F1-score for the training is equal to 100%, while for the testing it is 96.34%. We also observe that the recall and the precision have increased from the training (100%, 100%, respectively) in comparison with the testing (96.28%, 96.4%, respectively). For the loss function, its value is equal to 1.55 e-05 for training and 0.17 for testing.

As shown in Fig. 6, the most identifiable class is “Residential”, whereas the model was capable to predict only 377 images as “Pasture”.

TABLE VI. Different Metrics of MobileNet_V1

Metrics	Training	Testing
Accuracy	99.91%	76.31%
F1-Score	99.89%	76.37%
Recall	99.87%	74.96%
Precision	99.91%	77.88%
Loss	0.01	0.9

4) MobileNet_V1: Table VI illustrates the obtained results of MobileNet_V1. The table shows a remarkable difference between the results of the different metrics for the training data compared to the testing data with values of 99.91%, 99.89%, 99.87%, 99.91% and 0.01 (Accuracy, F1-score, Recall, Precision, and Loss, respectively) for the training and for the testing (76.31%, 76.37%, 74.96%, 77.88% and 0.9). According to these results we can distinguish that the model overfit.

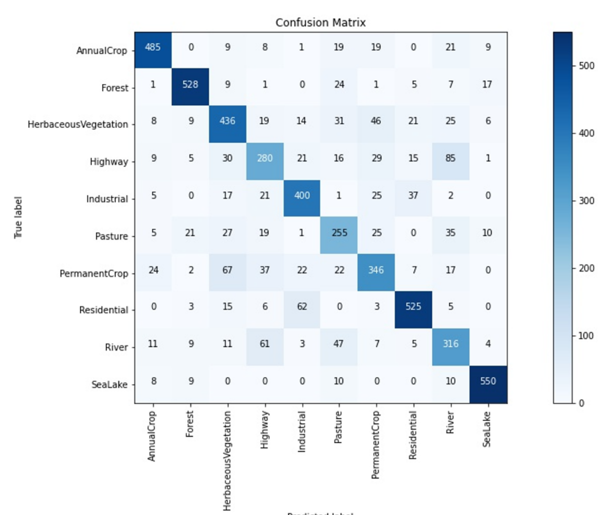


Fig. 7. Confusion matrix of MobileNet_V1

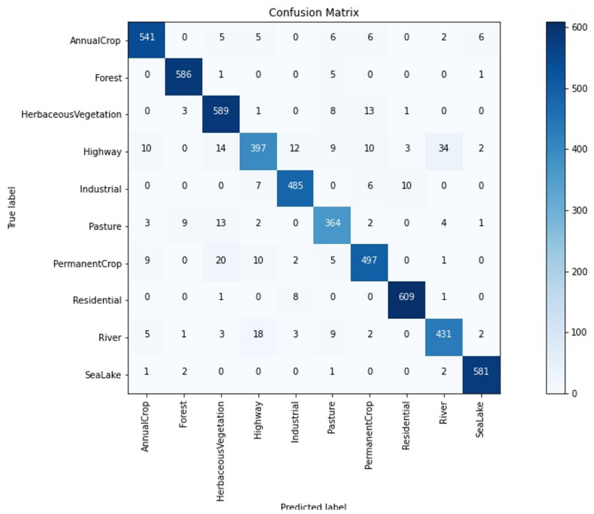


Fig. 8. Confusion matrix of DenseNet201

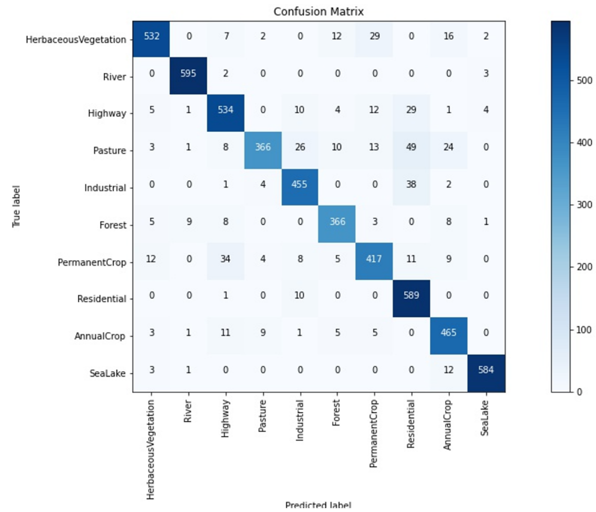


Fig. 9. Confusion matrix of proposed CNN MMN

TABLE VIII. THE RESULTS OF THE PROPOSED CNN ARCHITECTURE MMN WITH DATA AUGMENTATION

Metrics	Training	Testing
Accuracy	91.89%	90.79%
F1-Score	91.89%	90.79%
Recall	90.52%	89.75%
Precision	93.34%	91.91%
Loss	0.23	0.27

The confusion matrix (Fig. 7) indicates that the model can correctly predict 550 images of the "SeaLake" class and just 255 images of the "Pasture" class.

TABLE VII. DIFFERENT METRICS OF DENSENET201

Metrics	Training	Testing
Accuracy	100%	94.07%
F1-Score	100%	94.11%
Recall	100%	93.98%
Precision	100%	94.24%
Loss	8.72e-05	0.25

5) DenseNet201: As shown in Table VII, we can observe that accuracy, F1-score, recall and precision have remarkably high values for training (100%, 100%, 100%, 100%) compared to testing (94.07%, 94.11%, 93.98%, 94.24%), as well as for loss which has decreased from testing (0.25) to training (8.72e-05).

Concerning the confusion matrix (Fig. 8), the model was able to correctly identify 609 images of the class "Residential", however only 364 images were correctly labeled as "Pasture".

D. Results with Data Augmentation

In this section, we present the different results of the trained models, but this time using data augmentation techniques on the same dataset.

1) Proposed Architecture Modified MobileNet_V1 (MMN): According to Table VIII, we can see that the different metrics do not vary remarkably from training to testing. There is a small decrease in the accuracy, the F1-score, the recall, and the precision. However, the loss function was increased from 0.23 in training to 0.27 in testing.

TABLE IX. DIFFERENT METRICS OF VGG16

Metrics	Training	Testing
Accuracy	96.87%	96.01%
F1-Score	96.87%	96.09%
Recall	96.49%	95.71%
Precision	97.26%	96.5%
Loss	0.09	0.12

TABLE X. DIFFERENT METRICS OF RESNET50

Metrics	Training	Testing
Accuracy	99.97%	97.59%
F1-Score	99.97%	97.61%
Recall	99.97%	97.59%
Precision	99.97%	97.63%
Loss	0.001	0.17

From the confusion matrix (Fig. 9), we note that the model correctly recognizes 595 images of the class "River",the model was also able to identify only 366 images of "Forest" and "Pasture".

2) VGG16: Table IX shows the accuracy, f1-score, recall, precision and loss of the VGG16 model on the training data and on the testing data. We notice that for all the metrics, there is an insignificant difference between training and testing.

Concerning the confusion matrix (Fig. 10), the model was able to classify 600 images of "Residential" as well as 367 images of the "Forest" class.

3) ResNet50: As shown in Table X, the different evaluation metrics decreased from training to testing, with values of more than 97% for testing and more than 99% for training. The loss function showed a remarkable increase between the two cases with 0.001 for training data and 0.17 for testing data.

For the confusion matrix (Fig. 11), the most recognized class by ResNet50 is the "SeaLake" class with 599 images,

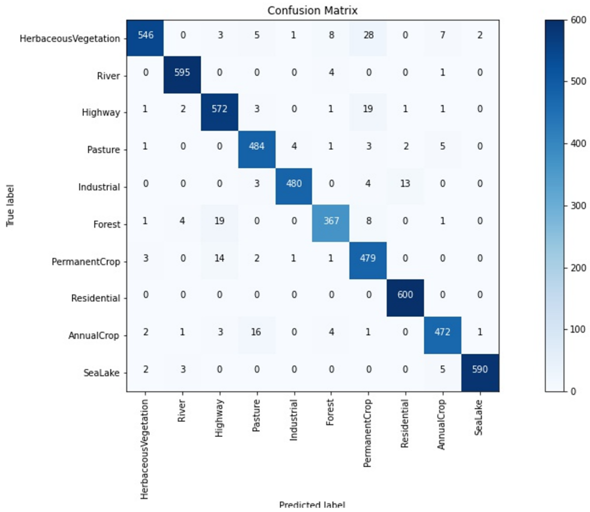


Fig. 10. Confusion matrix of VGG16

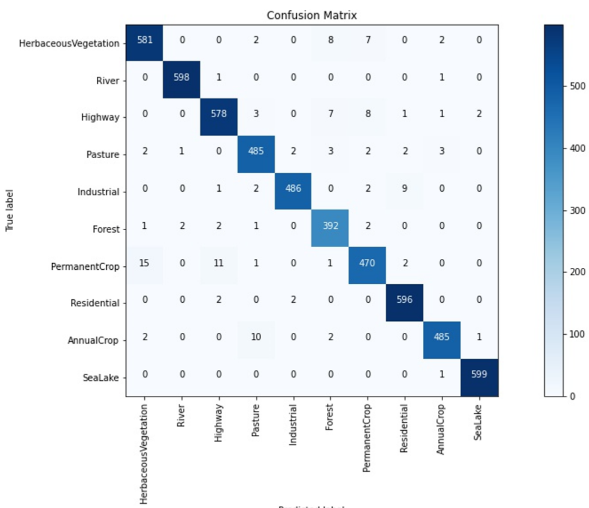


Fig. 11. Confusion matrix of ResNet50

TABLE XI. Different Metrics of MobileNet_V1

Metrics	Training	Testing
Accuracy	97.18%	96.59%
F1-Score	97.17%	96.64%
Recall	97.03%	96.56%
Precision	97.32%	96.73%
Loss	0.08	0.12

therefore the least classified class is the “Forest” with just 392 images.

4) *MobileNet_V1:* Table XI represents the evaluation result of MobileNet_V1 architecture. It shows that there is a small difference between training and testing i.e. accuracy, f1-score, recall, precision and loss.

As shown in Fig. 12, the model was able to recognize correctly 596 images of the class “River” and only 375 for the class “Forest”.

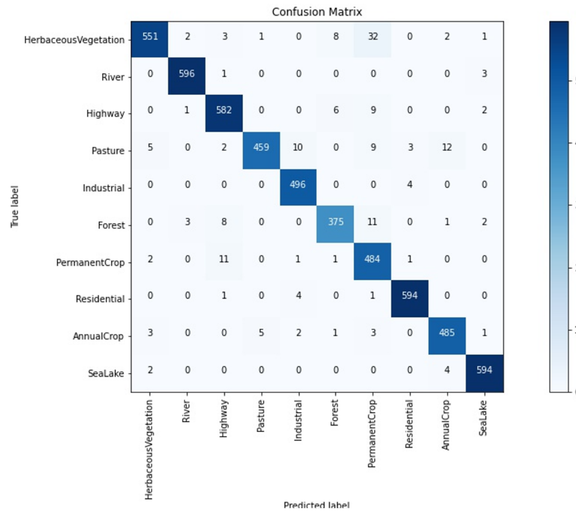


Fig. 12. Confusion matrix of Mobilenet_V1

TABLE XII. Different Metrics of DenseNet201

Metrics	Training	Testing
Accuracy	98.77%	97.24%
F1-Score	98.74%	97.27%
Recall	98.62%	97.18%
Precision	98.87%	97.36%
Loss	0.03	0.09

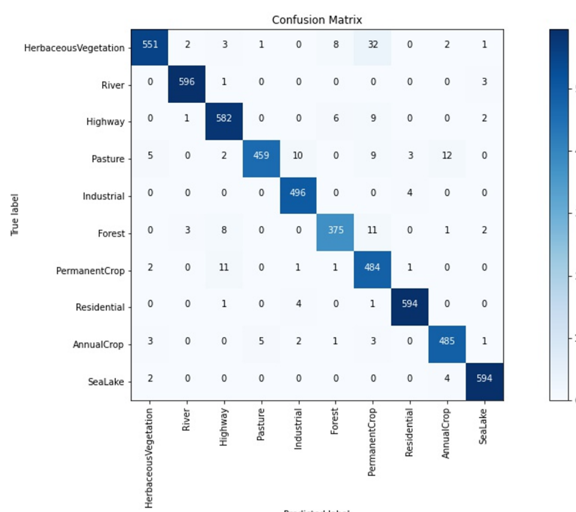


Fig. 13. Confusion matrix of DenseNet201

5) *DenseNet201:* All metrics have a minimal variation between training and testing, as shown in Table XII, with the exception of loss, which has a small increase between training and testing.

Fig. 13 shows that 593 images were correctly marked as “SeaLake” and “River”, but the model was able to identify just 389 images of “Forest”.

E. Results on Real Images

The satellite image was acquired in the surroundings of Berrechid, Morocco, in May 2022, we took two patches from it to test several classifiers.

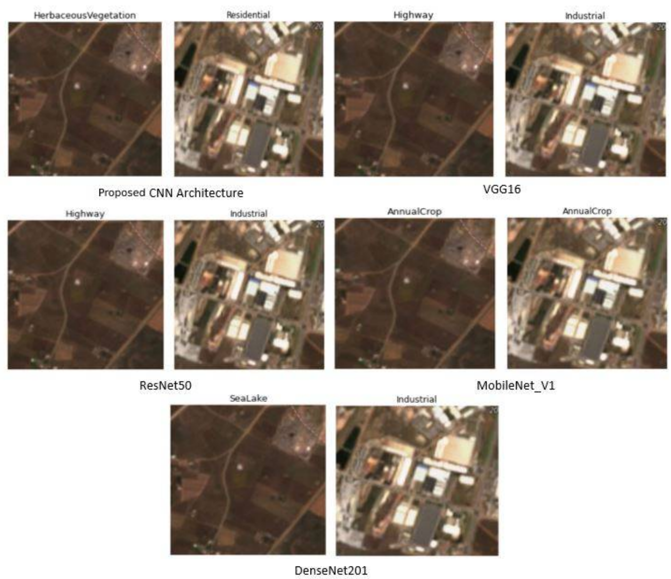


Fig. 14. The results of the different models

The Fig. 14 shows the results of the different models on two real images. We notice that for the proposed architecture MMN, the model gave as classification for the first image "Herbaceous Vegetation" and for the second image "Residential". For VGG16 and ResNet50, we got "Highway" for the first image which is clearly wrong classification, and "Industrial" for the second, while MobileNet_V1 has classified the first and second image as "AnnualCrop", which can not be possible for the second one. Finally, for DenseNet201, the model gave an erroneous classification for the first image (SeaLake) and classified the second as "Industrial".

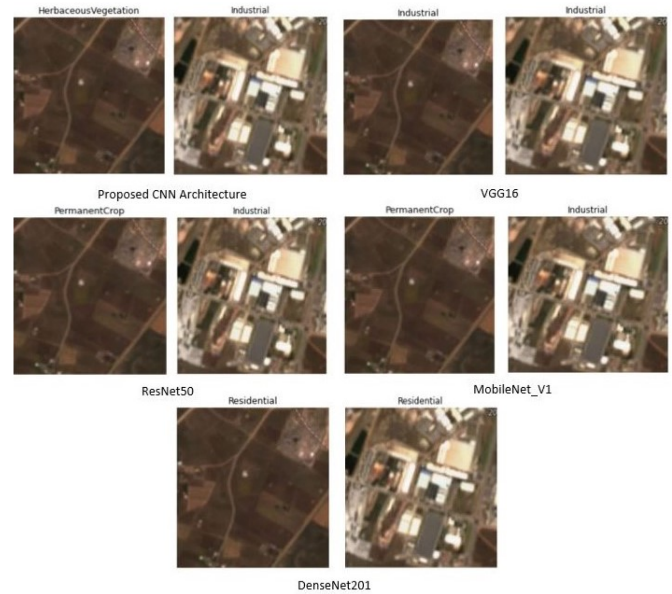


Fig. 15. The results of the different models with data augmentation

The obtained results with data augmentation are shown in the Fig. 15. We can observe that the proposed method identified the first image as "Herbaceous Vegetation" and the second as "Industrial", while VGG16 classified the 2 images as "Industrial" which can't be true for the first image. For ResNet50 and MobileNet_V1, they recognized the first image as "Permanent Crop" and the second as "Industrial". DenseNet201 identified the first and second image as "Residential", it gave wrong classification for the first image.

V. Discussion

In this study, we addressed the multiclass classification of the LULC domain, based on the Eurosat dataset [4]. The training was applied using transfer learning, in order to identify the best architecture. The experimental study proved that the neural networks architectures are very useful for LULC image classification since they have shown high performance.

TABLE XIII. Evaluations Metrics of Different Architectures

Models	Accuracy	F1-Score	Recall	Precision	Loss	parameters number
MMN	94.01%	94.02%	93.9%	94.8%	0.18	6402506
VGG16	96.9%	96.87%	96.84%	96.89%	0.18	15768906
ResNet50	96.31%	96.34%	96.28%	96.4%	0.17	27787658
MobileNet-_V1	76.31%	76.37%	74.96%	77.88%	0.9	3239114
DenseNet-201	94.07%	94.11%	93.98%	94.24%	0.25	22259786

Table XIII compares the performance of each architecture based on the mentioned metrics in Section IV-B, and the number of parameters. From the results, it can be seen that accuracy, F1-score, recall and precision when using the

MobileNet_V1 model are remarkably lower than the other models, with values of 76.31%, 76.37%, 74.96% and 77.88%, respectively, and this is due to the minimum number of parameters (3.2 M). Furthermore, the DenseNet201 and the proposed model showed the same performance except for the loss function, which has a value of 0.18 for the proposed architecture and 0.25 for DenseNet201. This with a too small number of parameters (6.4 M) in comparison with DenseNet201 (18.35 M), It is due to the fact that the extracted features from the previous layer have been added to the next layer. The main advantage of our architecture is that it needs a small number of parameters.

Both the VGG16 and ResNet50 architectures perform well, with 96% accuracy, F1-score, recall, and precision, which is expected considering ResNet50's depth and VGG16's usage of small filters. It can be seen that the MMN architecture is almost as accurate compared to the other architectures while being smaller in terms of parameters.

TABLE XIV. EVALUATIONS METRICS OF DIFFERENT ARCHITECTURES WITH DATA AUGMENTATION

Models	Accuracy	F1-Score	Recall	Precision	Loss	parameters number
MMN	90.79%	90.79%	89.75%	91.91%	0.27	6402506
VGG16	96.01%	96.09%	95.71%	96.5%	0.12	15768906
ResNet50	97.59%	97.61%	97.59%	97.63%	0.17	27787658
MobileNet-_V1	96.59%	96.64%	96.56%	96.73%	0.12	3239114
DenseNet-201	97.24%	97.27%	97.18%	97.36%	0.09	22259786

Table XIV shows the obtained results of the different architectures with the use of data augmentation. Based on the presented results, ResNet50 and DenseNet201 showed good performance with an accuracy of more than 97%. Thus, we notice that data augmentation has improved the results for both architectures. Then we have the MobileNet_V1 and VGG16 with an accuracy of more than 96%. We also have for the loss function, 0.09 for the DenseNet201 model, 0.12 for VGG16 and MobileNet_V1, and an increase for ResNet50 with 0.17. Therefore, we can see that MobileNet_V1 with a lower number of parameters was able to achieve such remarkable results compared to the other models. Generally, this is expected since data augmentation is one of the best techniques for reducing overfitting. Table XIV shows that the MMN architecture was able to achieve an accuracy of 90.79% which is a reduced value compared to other architectures. This is most likely due to the model's limited capacity, which prevents it from learning all of the patterns in the data, or because some architectures are very sensitive to data augmentation.

VI. CONCLUSION AND PERSPECTIVES

In this work, we evaluated the performance of the automated methods used to classify LULC images into 10 classes on the Eurosat dataset using four DL architectures (VGG16, ResNet50, MobileNet_V1 and DenseNet201) and a proposed CNN architecture MMN based on a combination of the two methods MobileNet_V1 and ResNet50. Comparing the results, we found that our architecture proved good performance but performs poorly using data augmentation compared to the other architectures. Therefore, our model needs to be adjusted to get better performance on the dataset with data augmentation. We also found that the performance of the DenseNet201 model and the proposed architecture is equivalent but the proposed model involves fewer parameters, which means it requires less memory consumption. In the experiment, we found that with and without the use of data augmentation, the VGG16 and ResNet50 models perform well (accuracy greater than 96%) compared to the other architectures. Moreover, we noticed that MobileNet_V1 can efficiently classify LULC images in the case of data augmentation, but with the original dataset its results were poor.

The proposed model will be improved by applying parameter tuning and studying the loss function. In addition, the used dataset in the training phase has a high impact on the efficiency of the model, thus it is very important to test the model with another dataset. Since each model has its own method to extract features, which highly influences the classification accuracy, we aim to study deeply the layers of feature extraction in order to improve the proposed model efficiency.

ACKNOWLEDGMENTS

This research is supported by National Institute of Statistics and Applied Economics, Rabat, Morocco.

REFERENCES

[1] X. Jiang, Y. Wang, W. Liu, S. Li, and J. Liu, "Capsnet, cnn, fcn: Comparative performance evaluation for image classification," *International Journal of Machine Learning and Computing*, vol. 9, no. 6, pp. 840–848, 2019.

[2] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 558–567.

[3] M. I. Lakhal, H. Çevikalp, S. Escalera, and F. Ofli, "Recurrent neural networks for remote sensing image classification," *IET Computer Vision*, vol. 12, no. 7, pp. 1040–1045, 2018.

[4] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.

[5] S. Mubako, O. Belhaj, J. Heyman, W. Hargrove, and C. Reyes, "Monitoring of land use/land-cover changes in the arid transboundary middle rio grande basin using remote sensing," *Remote Sensing*, vol. 10, no. 12, p. 2005, 2018.

[6] C. Zhang, P. A. Harrison, X. Pan, H. Li, I. Sargent, and P. M. Atkinson, "Scale sequence joint deep learning (ss-jdl) for land use and land cover classification," *Remote Sensing of Environment*, vol. 237, p. 111593, 2020.

[7] S. E. Jozdani, B. A. Johnson, and D. Chen, "Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification," *Remote Sensing*, vol. 11, no. 14, p. 1713, 2019.

[8] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y.-A. Liou, A. Rahman *et al.*, "Land-use land-cover classification by machine learning classifiers for satellite observations—a review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020.

[9] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 166–177, 2019.

[10] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson, "Joint deep learning for land cover and land use classification," *Remote sensing of environment*, vol. 221, pp. 173–187, 2019.

[11] R. Nijhawan, D. Joshi, N. Narang, A. Mittal, and A. Mittal, "A futuristic deep learning framework approach for land use-land cover classification using remote sensing imagery," in *Advanced computing and communication technologies*. Springer, 2019, pp. 87–96.

[12] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *arXiv preprint arXiv:1508.00092*, 2015.

[13] K. Nogueira, O. A. Penatti, and J. A. Dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, 2017.

[14] O. A. Penatti, K. Nogueira, and J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 44–51.

[15] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[16] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.

[17] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 197–209, 2018.

[18] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[19] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "Deepsat: a learning framework for satellite imagery," in *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, 2015, pp. 1–10.

[20] X. Bian, C. Chen, L. Tian, and Q. Du, "Fusing local and global features for high-resolution scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2889–2901, 2017.

[21] G. Cheng, J. Han, L. Guo, and T. Liu, "Learning coarse-to-fine sparselets for efficient object detection and scene classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1173–1181.

[22] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sensing of Environment*, vol. 241, p. 111716, 2020.

[23] A. Panda, A. Singh, K. Kumar, A. Kumar, A. Swetapadma *et al.*, "Land cover prediction from satellite imagery using machine learning techniques," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2018, pp. 1403–1407.

[24] M. M. Aburas, M. S. S. Ahamad, and N. Q. Omar, "Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review," *Environmental monitoring and assessment*, vol. 191, no. 4, pp. 1–28, 2019.

[25] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[26] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns," *IEEE transactions on geoscience and remote sensing*, vol. 56, no. 5, pp. 2811–2821, 2018.

[27] G. Sumbul and B. Demİr, "A deep multi-attention driven approach for multi-label remote sensing image classification," *IEEE Access*, vol. 8, pp. 95 934–95 946, 2020.

[28] D. Koßmann, T. Wilhelm, and G. A. Fink, "Towards tackling multi-label imbalances in remote sensing imagery," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 5782–5789.

[29] A. K. Aksoy, M. Ravanbakhsh, T. Kreuziger, and B. Demir, "A novel uncertainty-aware collaborative learning method for remote sensing image classification under multi-label noise." *CoRR, vol. abs/2105.05496*, 2021.

[30] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 103, p. 102520, 2021.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[32] U. Chaudhuri, S. Dey, M. Datcu, B. Banerjee, and A. Bhattacharya, "Interband retrieval and classification using the multilabeled sentinel-2 bigearthnet archive," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9884–9898, 2021.

[33] P. Aggarwal, N. K. Mishra, B. Fatimah, P. Singh, A. Gupta, and S. D. Joshi, "Covid-19 image classification using deep learning: Advances, challenges and opportunities," *Computers in Biology and Medicine*, p. 105350, 2022.

[34] W. Liu, C. Li, M. M. Rahaman, T. Jiang, H. Sun, X. Wu, W. Hu, H. Chen, C. Sun, Y. Yao *et al.*, "Is the aspect ratio of cells important in deep learning? a robust comparison of deep learning methods for multi-scale cytopathology cell image classification: From convolutional neural networks to visual transformers," *Computers in biology and medicine*, vol. 141, p. 105026, 2022.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[37] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[38] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches," *arXiv preprint arXiv:1803.01164*, 2018.

[39] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2020.

[40] P. C. Pandey, N. Koutsias, G. P. Petropoulos, P. K. Srivastava, and E. Ben Dor, "Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers—a review of the state of the art," *Geocarto International*, vol. 36, no. 9, pp. 957–988, 2021.

[41] P. Kim, "Convolutional neural network," in *MATLAB deep learning*. Springer, 2017, pp. 121–147.

[42] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*. Ieee, 2017, pp. 1–6.

[43] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[44] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-vgg16 cnn model for big data places image recognition," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018, pp. 169–175.

[45] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, and P. De Geus, "Malicious software classification using transfer learning of resnet-50 deep neural network," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2017, pp. 1011–1014.

[46] K. Cai, X. Miao, W. Wang, H. Pang, Y. Liu, and J. Song, "A modified yolov3 model for fish detection based on mobilenetv1 as backbone," *Aquacultural Engineering*, vol. 91, p. 102117, 2020.

[47] S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through

glass diffusers using densely connected convolutional networks," *Optica*, vol. 5, no. 7, pp. 803–813, 2018.

[48] O. Mohamed, E. A. Khalid, O. Mohammed, and A. Brahim, "Content-based image retrieval using convolutional neural networks," in *First International Conference on Real Time Intelligent Systems*. Springer, 2017, pp. 463–476.

[49] M. Fatourechi, R. K. Ward, S. G. Mason, J. Huggins, A. Schloegl, and G. E. Birch, "Comparison of evaluation metrics in classification applications with imbalanced datasets," in *2008 seventh international conference on machine learning and applications*. IEEE, 2008, pp. 777–782.

APPENDIX

APPENDIX A. TERMS AND THEIR ABBREVIATIONS

LULC: Land Use and Land Cover.
DL: Deep Learning.
CNN: Convolutional Neural Network.
GIS: Geographic Information System.
MMN: Modified MobileNet_V1.

A Real-Time Open Public Sources Text Analysis System

Chi Mai Nguyen[1], Phat Trien Thai[2], Van Tuan Nguyen[3], Duy Khang Lam[4]
Viettel High Technology Industries Corporation
Ho Chi Minh City, Vietnam

***Abstract*—With the emergence of digital newspapers and social media, one can easily suffer from information overload. The enormous amount of data they provide has created several new challenges for computational and data mining, especially in the natural language processing field. Many pieces of research focusing on the information extraction process, such as named entity recognition, entity linking, and text analysis methodologies, are available. However, there is a lack of development for a system to unify all these advanced techniques. The current state-of-the-art systems are either semi-automatic or can only handle short-text documents. Most of them are not real-time or have a long lag. Some of them are domain restricted. Many of them only focus on a single source: Twitter. In this work, we proposed a system that can automatically collect, extract, and analyze information from public source text documents, like news and tweets. The system can be used in different domains, such as scientific research, marketing, and security-related domains.**

***Keywords*—named entity recognition; entity linking; text analysis system; data mining; natural language processing**

I. Introduction

We live in an age of information overload. The explosive growth of digital newspapers and online social networks creates enormous amounts of text data daily. This situation creates new challenges for computational and natural language processing (NLP) [1], [2], [3], [4], [5], [6]. While many pieces of research in the information extraction process and text analysis methodologies are available, there is a lack of development for a system to unify all these. This work proposed a system that can automatically collect, extract, analyze, and monitor information from text documents, such as named entities. The system takes in unstructured text from open public sources, like digital newspapers and Twitter, employs some information extraction processes, and records the information analysis results. It is fully automatic and has a real-time monitoring feature.

The system is not limited to any data domain. It can be applied in the news report system to automatically support reporters in detecting popular keywords. The proposed system can detect trending topics from scientific papers for research purposes. Trends are also crucial for analysts, marketing professionals, and retailers who want to monitor their and competitors' online products. Also, some government-run organizations can benefit from the proposed system by having security information about hot events related to entities from digital news or social media.

Many existing systems have applications comparable to ours, but they still have a few limitations. Some do the text analysis task on a single Twitter data stream [7], [8], [9], [10], while others only serve on their specific domains [11], [12], [13]. Also, many systems are not real-time as Trend Miner, [10], or not fully automatic as LRA Crisis Tracker, [12]. In addition, several studies analyze documents using various entity recognition methods to extract entities [9], [14], [15]. Still, they do not mention how to address the issue of many different entities co-referring to the same real-world object. To overcome these drawbacks, we propose a system with the below contributions:

- It can automatically collect data from Twitter and digital newspapers.
- It can process both long and short text.
- It employs an extra Entity Linking after the Named Entity Recognition module to map the extracted entities to their unique identities.
- It is capable of real-time processing.

The proposed system comprises six modules deployed as micro-services. First, the Data Stream collects raw documents from Twitter and multiple news sources. Next, named entities are extracted from the text by the Named Entity Recognition module. Once entities have been mapped with their identity by Entity Linking, they are fed into the Entity Tracking module to count the frequency within a time step. The Trend Detection module will take the calculated frequency and perform a trending test computation. Finally, The Entity Monitoring module will keep track of all the entities' occurrence frequencies and notify users if abnormal rising trends are detected.

A more detailed description of the system is included in Section III. Section II presents a brief overview of some current state-of-the-art event monitoring systems. Experiment results and evaluation are shown in Section IV. The summary and future works are discussed in Section V.

II. Related Works

There are a significant number of systems that analyze document streams. The majority of previous works focus on social media, such as Twitter. These include TwitterStand [16], TwitterMonitor [17], and Jasmine [18], which focus on detecting trending keywords that correspond to global or local events.

LRA Crisis Tracker[1] is a system that tracks armed group activities and conflict-related incidents in the remote border region encompassing the northeastern Democratic Republic

[1]https://crisistracker.org/

of Congo and the eastern Central African Republic. It uses data from crowd-sourcing. Domain experts must re-examine the data before feeding it to the system. In contrast to the Crisis Tracker system, the proposed system is fully automatic.

Redites [7] is a system that monitors events based on Twitter's tweets. The system will decide if a new tweet contains a new event whenever it is streamed. The event is then classified. Only events in security categories are analyzed and monitored. Redites is fully automatic. However, it can only handle short text documents like tweets from Twitter. The proposed system can handle both long- and short-text documents.

Social Sensor[2] tracks and monitors predefined events on social media [19]. The proposed system automatically extracts the events from news and tweets. Moreover, Social Sensor can only handle short text documents like tweets.

Trend Miner[3] is similar to our proposed system; however, it does not focus on real-time aspects and can only handle short-text documents [20].

Cheng et al. [10] proposed an early warning system to detect COVID-19 outbreaks based on Twitter's data. The system has a 6-27 days lag and is only applicable for short-text analysis.

Epitweetr [11] is a system that detects outbreaks and public health threats using Twitter's data. The system is fully automatic and provides additional support for public health experts in detecting and identifying the geo-location of an outbreak. However, compared to the proposal, the system is restricted by data domain and can only handle short-text documents.

Goh et al. [12] proposed an approach using structured data and unstructured clinical notes to predict and diagnose sepsis. This approach is, however, restricted by the data domain and is not a fully automatic system.

A real-time system [9] was built to ingest the stream of all tweets and identify clusters of event-related entities on a minute-by-minute basis. The system first extracts named entities, hashtags and internal knowledge graph entities from each tweet. It keeps only trending entities detected by its internal Trend Detection system known as Twitter Trends. Then, a weighted graph is naturally constructed with the entities as nodes and their cosine similarities as edge weights. Community detection algorithms have been used to detect the sub-graphs (clusters) based on their links to others. To solve the Event Evolution problem, they add a layer of Cluster Linking to form cluster chains. Finally, clusters are ranked based on the aggregate popularity of their entities and persisted in internal stores for future use.

EveSense [8] is an Event Detection application that detects real-life events and related trending topics from the Twitter stream and allows users to find interesting events that have recently occurred.

News Monitor [15] is a scalable real-time framework for analyzing and exploring news articles. The system collects news from many RSS news sources and automatically extracts the main content and other metadata. News Monitor is analogous to our approach compared to many news services, such as Event Registry [21]. However, it cannot handle the entity ambiguity problem when data diversity increases.

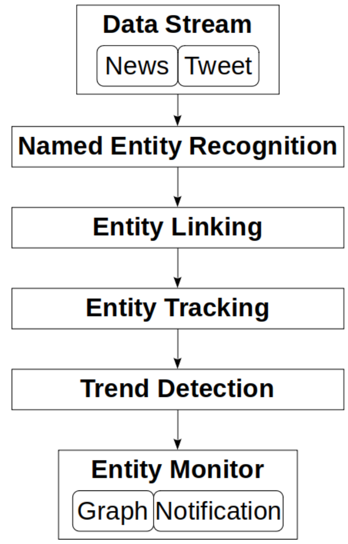


Fig. 1. System diagram

Our system can handle both long texts like news and short texts like tweets. It is also capable of analyzing these data sources simultaneously. The design, thus, can explore the potential relationship between the information flows to produce more consistent, accurate, and valuable information. Moreover, the system has an Entity Linking module to assign a unique identity to entities mentioned in the text, which helps robust join and union operations that can integrate information about entities [22].

III. Main Components

The system is engineered in modules and is extensible. Each module is a separate service that can be modified or removed without affecting the whole system. A new module can also be added easily. Fig. 1 presents an overview of the system.

A. Data Stream

The Data Stream module handles the task of streaming posts from social media (such as tweets from Twitter) and news from digital newspapers. The module uses Scrapy[4] and Automation Browser[5] for its streaming process. This module is updated accordingly to the newspapers and social media structures.

B. Named Entity Recognition

The Named Entity Recognition (NER) [23], [24] module handles the task of extracting and classifying named entities in

[2]http://socialsensor.iti.gr/
[3]http://www.trendminer-project.eu/
[4]https://scrapy.org/
[5]https://axiom.ai/

Fig. 2. Named entity recognition process

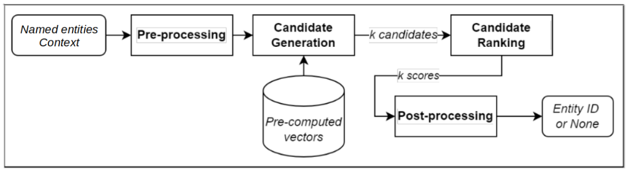


Fig. 3. Entity linking process

unstructured text into predefined categories. Two deep learning approaches, namely Bidirectional Encoder Representations from Transformers (BERT) [25], [26] and Bidirectional Long Short-Term Memory (BiLSTM) [27], [28], [29], were used in this module. Each sentence from an unstructured text document will be tokenized and vectorized using the BERT model. The result vectors are fed to the BiLSTM model for detecting and classifying the named entities and their categories, as shown in Fig. 2.

C. Entity Linking

The Entity Linking (EL)[30] module links the named entities extracted from the NER module to their corresponding predefined unique identity. For example, the entities “US”, “USA”, and “United States” have the same identity as “United States of America”. Whereas “New York” is the name for both a city and a state in the United States of America, they have two separate identities. Fig. 3 shows how the entity linking process works in this module. The core of this module is BERT deep learning approach [31], [32].

D. Entity Tracking

The Entity Tracking module handles the task of tracing and tracking the named entities from each unstructured document. The module computes and keeps a record of the occurrence frequency of each named entity in a defined time step. The entities are then ranked by their occurrence frequencies. Top-ranked entities are fed to a trend detection module to examine if there is an abnormal increasing trend. This work is a real-time process. Data is streamed and processed simultaneously.

E. Trend Detection

The Trend Detection module uses the Mann-Kendall approach [33], [34], [35], [36] to detect an abnormal increasing trend in the occurrence frequency of an entity. The Entity Monitoring module will notify the users if a rising trend is seen.

F. Entity Monitoring

The Entity Monitoring module controls both the Entity Tracking module and the Trend Detection module. It handles the task of monitoring the entities and notifying users if there is an anomaly, as shown in Fig. 4.

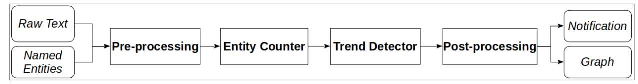


Fig. 4. Entity monitoring process

IV. Experiments and Evaluation

We have implemented and tested our system. We manually labelled 927 text documents with 20000 sentences and 14 different entity categories to train and evaluate the NER module. After training, we used a set consisting of 1500 sentences for testing the trained module. The final f1-score of the module is 91.41%. The EL module is trained and tested on a dataset consisting of 4475 entities, 1244 of which were manually labelled. The recall rate of the EL module is 97.2%. The system is capable of monitoring up to 3000 entities simultaneously.

Currently, the system can automatically collect and process news from 63 newspapers and public tweets in English from Twitter. We tested the system on a dataset of news articles and tweets collected from mentioned sources from July 15, 2022 to October 20, 2022. We chose July 21, 2022 as a virtual ”current date” and let the system simulate the process of extracting and analyzing data. The results of the analysis are visualized as graphs.

The systems will automatically monitor the top-ranked entities when nothing is specified, as mentioned in Section III-D. Fig. 5 and Fig. 6 show the example graphical presentations of the tracking and the monitoring processes over multi entities, respectively. As a default, the system will monitor the top 100 entities; however, we chose to showcase only five of them in the figures for readability. Each subgraph in Fig. 5 presents the time series of the occurrence frequency of an entity, i.e., the number of news articles and tweets mentioning the entity during some predefined time step. Each subgraph in Fig. 6 presents the monitoring process of an entity. The blue line presents the occurrence frequency of an entity, the yellow line shows the plot of the normalized Mann-Kendall values of that entity, and the vertical red lines indicate the time when abnormal rising trends are detected.

The system also allows users to choose the monitored entities and sources freely. Fig. 7 shows an example graphical presentation of the monitoring process of one specific entity. The upper graph presents the time series of the occurrence frequency of the entity. In the lower graph, the yellow line shows the plot of the computed normalized Mann-Kendall values of the entity, and the vertical red lines indicate the time when abnormal rising trends are detected. Each red line presents an anomaly. Fig. 8 shows a monitoring process of three entities of interest on Twitter, and Fig. 9 shows the same process on digital newspapers.

Moreover, the system can detect if two (or more) entities simultaneously have an abnormal rising trend. The upper graph of Fig. 10 shows the plots of two occurrence frequency time series of the two entities in comparison. The lower graph’s orange and blue lines present the computed normalized Mann-Kendall values of the two entities. At the same time, the vertical red lines indicate the time when both entities have

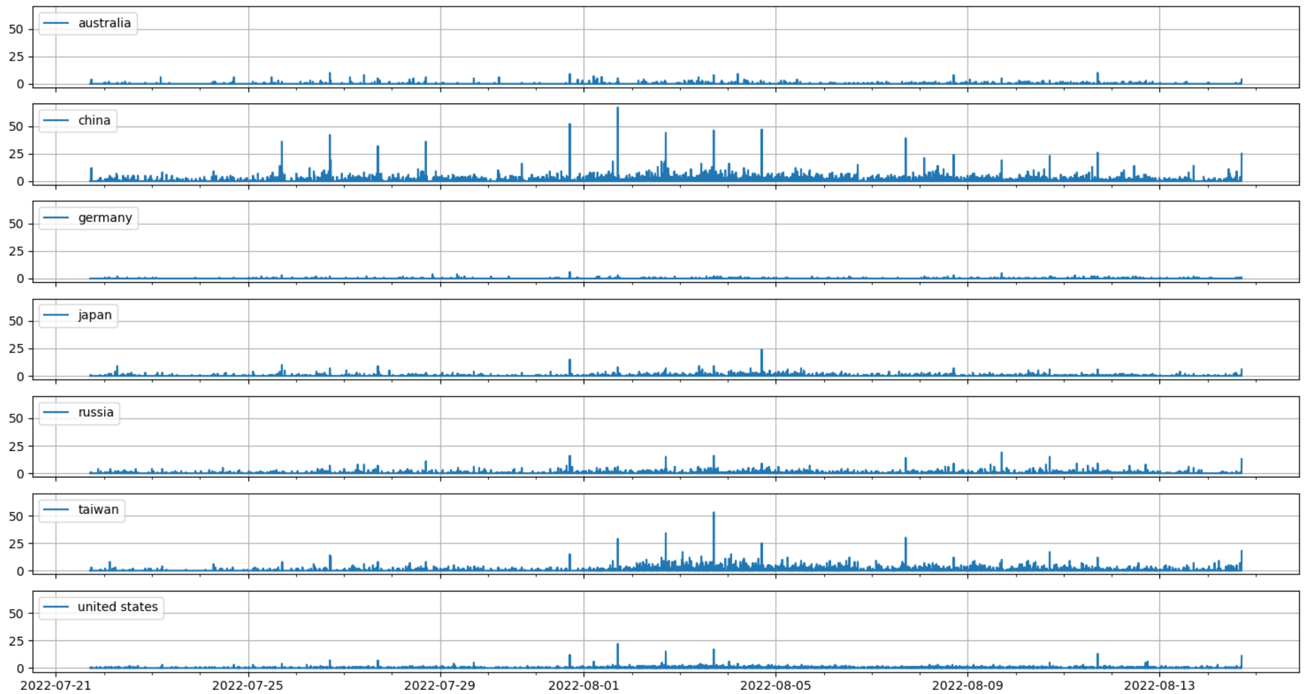


Fig. 5. Example of multi-entities tracking

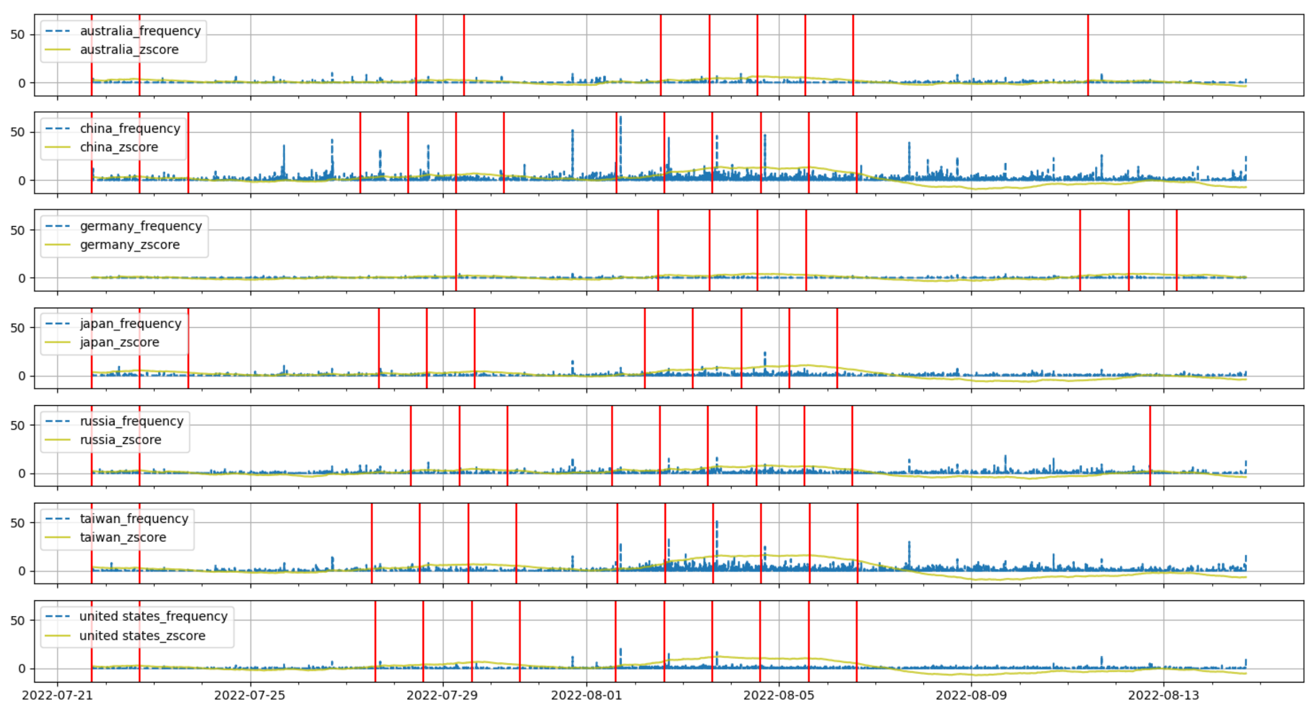


Fig. 6. Example of multi-entities monitoring

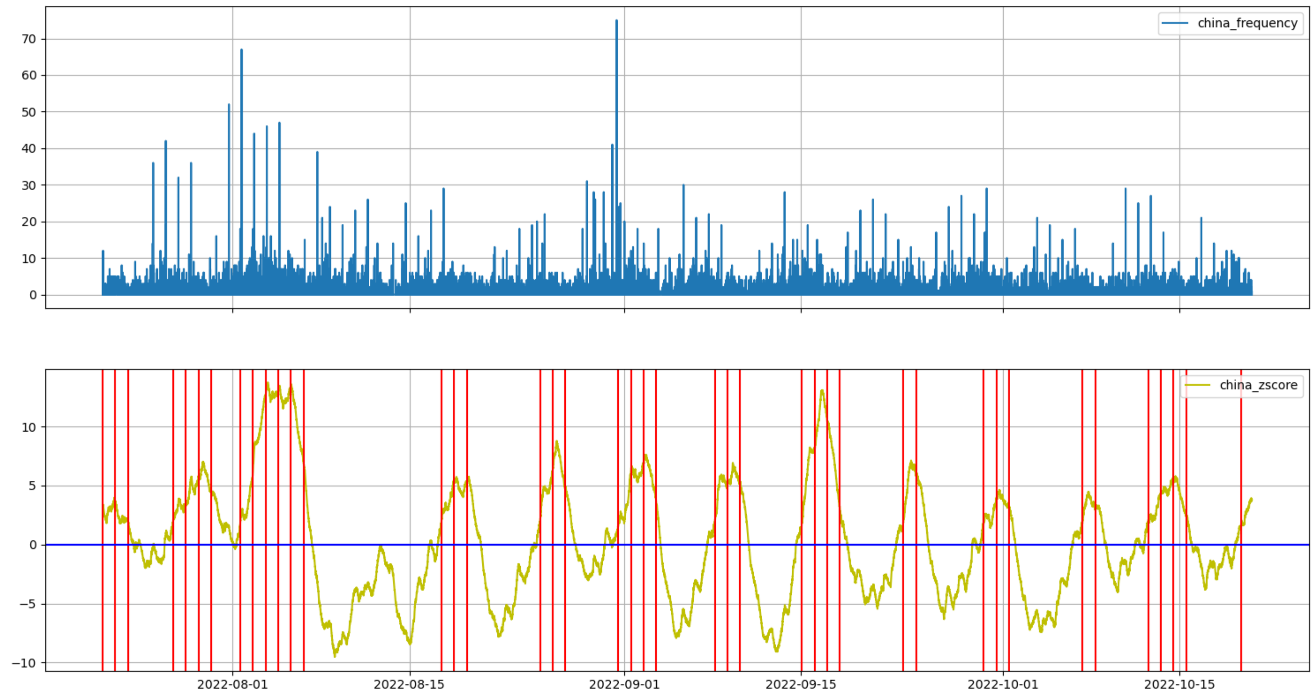


Fig. 7. Example of an entity monitoring process

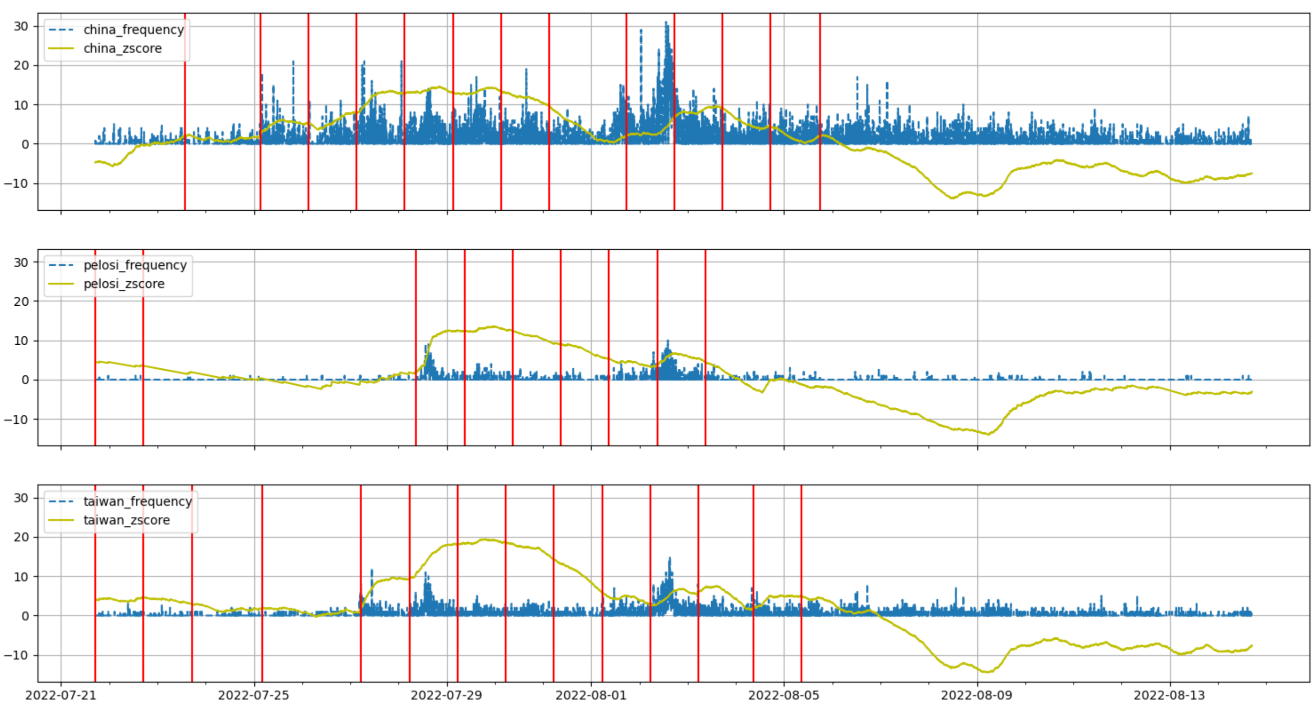


Fig. 8. Example of selected entities monitoring on twitter

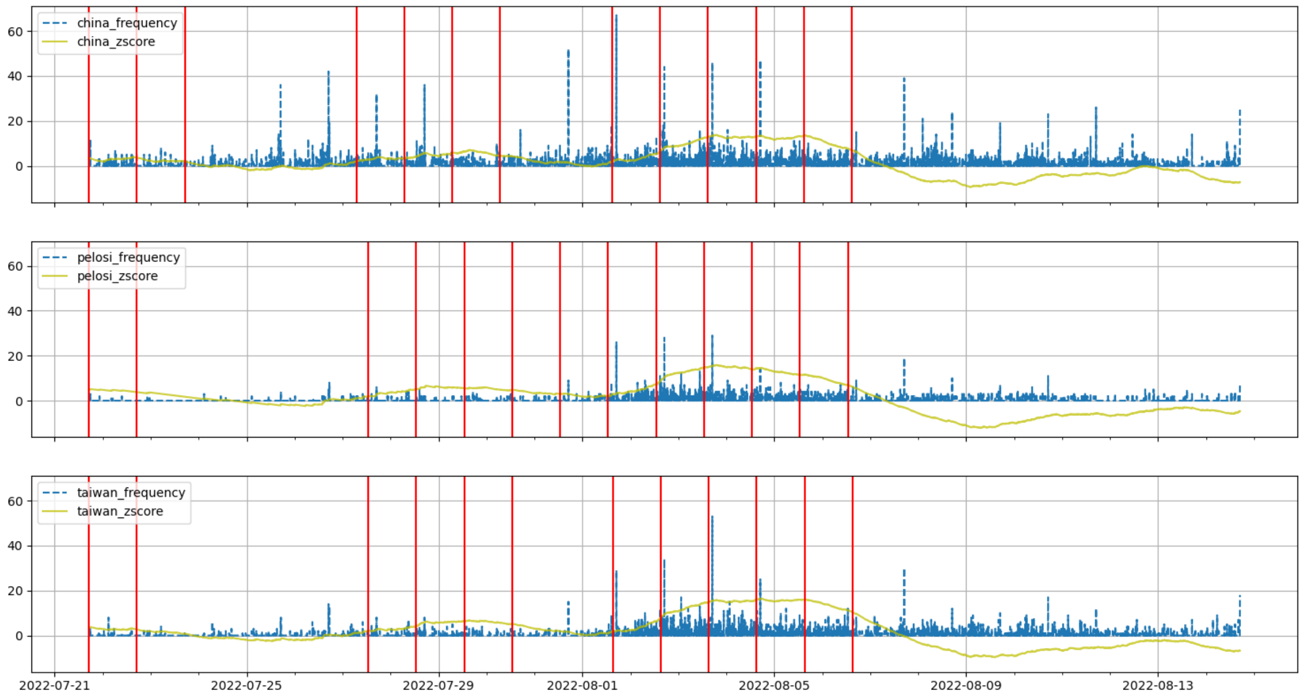


Fig. 9. Example of selected entities monitoring on news

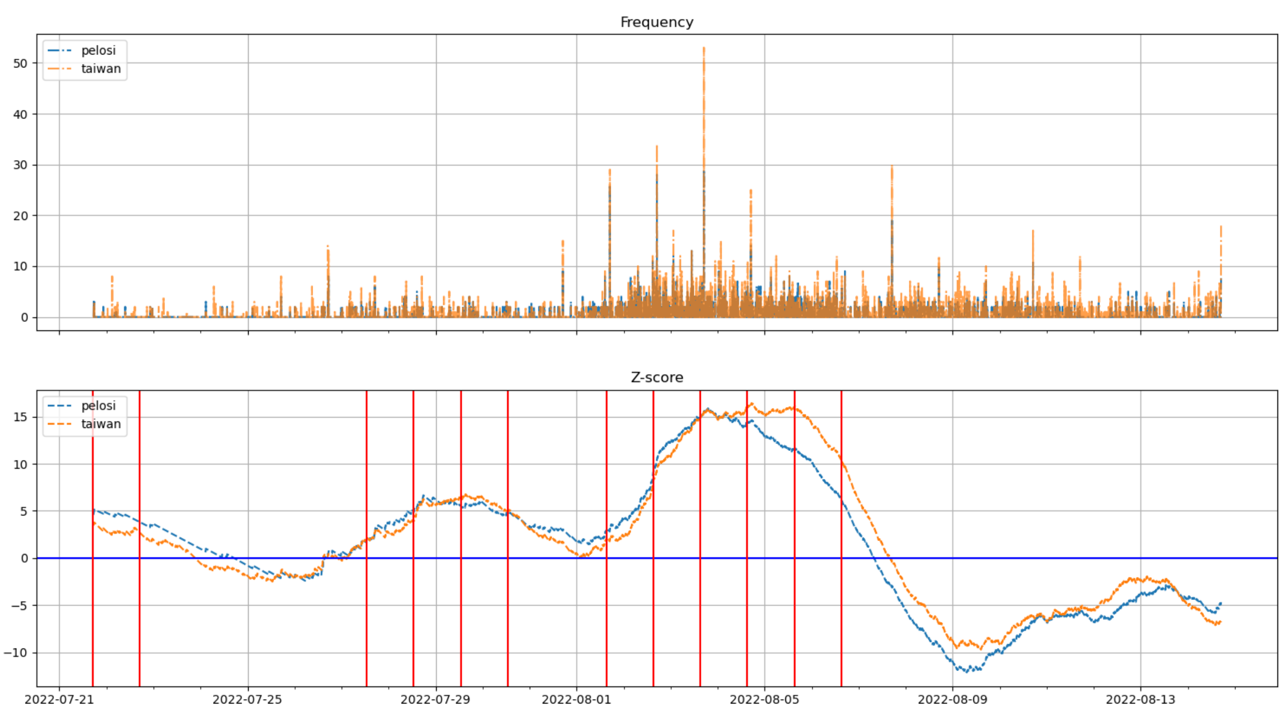


Fig. 10. Example of trend comparison between two entities

abnormal rising trends.

In all graphs, the y-axis presents values, and the x-axis is time.

V. Conclusions

We presented a system capable of automatically collecting, extracting, and analyzing information from open sources such as digital newspapers and Twitter. The system provides significant help for online exploratory analysis. Since the system is flexible, it can be used in different domains, such as scientific research, marketing, and security-related domains. As stated, the system is extensible. Possible future works may include but are not limited to (i) an event extraction module that detects and classifies possible events in which entities are involved, (ii) a sentiment analysis module that determines emotion about events, (iii) a language detection module that handles multi-lingual data, (iv) machine translator module that translates documents into a specific language, (v) a graphical user interface that is friendly for users in different domains, and (vi) a visualization improvement for the data and the outcome analysis.

References

[1] K. Jaseena, J. M. David *et al.*, "Issues, challenges, and solutions: big data mining," *CS & IT-CSCP*, vol. 4, no. 13, pp. 131–140, 2014.

[2] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[3] K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.

[4] H. A. Schwartz and L. H. Ungar, "Data-driven content analysis of social media: A systematic overview of automated methods," *The ANNALS of the American Academy of Political and Social Science*, vol. 659, no. 1, pp. 78–94, 2015.

[5] L. S. Lai and W. M. To, "Content analysis of social media: A grounded theory approach," *Journal of Electronic Commerce Research*, vol. 16, no. 2, p. 138, 2015.

[6] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia Tools and Applications*, pp. 1–32, 2022.

[7] M. Osborne, S. Moran, R. McCreadie, A. Von Lunen, M. Sykora, E. Cano, N. Ireson, C. Macdonald, I. Ounis, Y. He *et al.*, "Real-time detection, tracking, and monitoring of automatically discovered events in social media," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 37–42.

[8] Z. Saeed, R. Abbasi, and I. Razzak, *EveSense: What Can You Sense from Twitter?*, 04 2020, pp. 491–495.

[9] M. Fedoryszak, B. Frederick, V. Rajaram, and C. Zhong, "Real-time event detection on social data streams," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, jul 2019.

[10] I. Cheng, J. Heyl, N. Lad, G. Facini, and Z. Grout, "Evaluation of twitter data for an emerging crisis: an application to the first wave of covid-19 in the uk," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021.

[11] L. Espinosa, A. Wijermans, F. Orchard, M. Höhle, T. Czernichow, P. Coletti, L. Hermans, C. Faes, E. Kissling, and T. Mollet, "Epitweetr: Early warning of public health threats using twitter data," *Eurosurveillance*, vol. 27, no. 39, p. 2200177, 2022.

[12] K. H. Goh, L. Wang, A. Y. K. Yeow, H. Poh, K. Li, J. J. L. Yeow, and G. Y. H. Tan, "Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare," *Nature communications*, vol. 12, no. 1, pp. 1–10, 2021.

[13] B. Alkouz, Z. Al Aghbari, and J. H. Abawajy, "Tweetluenza: Predicting flu trends from twitter data," *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 273–287, 2019.

[14] A. Al-Laith and M. Shahbaz, "Tracking sentiment towards news entities from arabic news on social media," *Future Generation Computer Systems*, vol. 118, pp. 467–484, 2021.

[15] A. Saravanou, N. Panagiotou, and D. Gunopulos, *News Monitor: A Framework for Querying News in Real Time*, 03 2021, pp. 543–548.

[16] J. Sankaranarayanan, H. Samet, B. Teitler, M. Lieberman, and J. Sperling, "Twitterstand: News in tweets," 01 2009, pp. 42–51.

[17] M. Mathioudakis and N. Koudas, "Twittermonitor: Trend detection over the twitter stream," 06 2010, pp. 1155–1158.

[18] K. Watanabe, M. Ochi, M. Okabe, and R. Onai, "Jasmine: A real-time local-event detection system based on geolocation information propagated to microblogs," 10 2011, pp. 2541–2544.

[19] L. M. Aiello, G. Petkos, C. Martin, D. Corney, S. Papadopoulos, R. Skraba, A. Göker, I. Kompatsiaris, and A. Jaimes, "Sensing trending topics in twitter," *IEEE Transactions on multimedia*, vol. 15, no. 6, pp. 1268–1282, 2013.

[20] D. Preoţiuc-Pietro and T. Cohn, "A temporal model of text periodicities using gaussian processes," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 977–988.

[21] G. Leban, B. Fortuna, J. Brank, and M. Grobelnik, "Event registry: learning about world events from news," 04 2014, pp. 107–110.

[22] W. Shen, J. Wang, and J. Han, "Entity linking with a knowledge base: Issues, techniques, and solutions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, pp. 443–460, 2015.

[23] B. Mohit, "Named entity recognition," in *Natural language processing of semitic languages*. Springer, 2014, pp. 221–245.

[24] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50–70, 2020.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[26] C. Liang, Y. Yu, H. Jiang, S. Er, R. Wang, T. Zhao, and C. Zhang, "Bond: Bert-assisted open-domain named entity recognition with distant supervision," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 1054–1064.

[27] S. Cornegruta, R. Bakewell, S. Withey, and G. Montana, "Modelling radiological language with bidirectional long short-term memory networks," *arXiv preprint arXiv:1609.08409*, 2016.

[28] S. Zhang, D. Zheng, X. Hu, and M. Yang, "Bidirectional long short-term memory networks for relation classification," in *Proceedings of the 29th Pacific Asia conference on language, information and computation*, 2015, pp. 73–78.

[29] G. Liu and J. Guo, "Bidirectional lstm with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[30] D. Rao, P. McNamee, and M. Dredze, "Entity linking: Finding extracted entities in a knowledge base," in *Multi-source, multilingual information extraction and summarization*. Springer, 2013, pp. 93–115.

[31] L. Wu, F. Petroni, M. Josifoski, S. Riedel, and L. Zettlemoyer, "Scalable zero-shot entity linking with dense entity retrieval," *arXiv preprint arXiv:1911.03814*, 2019.

[32] A. Hamdi, E. Linhares Pontes, E. Boros, T. T. H. Nguyen, G. Hackl, J. G. Moreno, and A. Doucet, "A multilingual dataset for named entity recognition, entity linking and stance detection in historical newspapers," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 2328–2334.

[33] H. Mann, "Non-parametric test against trend. econometrical, 1945 (13): 245-259."

[34] M. G. Kendall, "Rank correlation methods." 1948.

[35] R. O. Gilbert, *Statistical methods for environmental pollution monitoring*. John Wiley & Sons, 1987.

[36] F. Wang, W. Shao, H. Yu, G. Kan, X. He, D. Zhang, M. Ren, and G. Wang, "Re-evaluation of the power of the mann-kendall test for detecting monotonic trends in hydrometeorological time series," *Frontiers in Earth Science*, vol. 8, p. 14, 2020.

Analysis of EEG Signals in a Patient with Spastic Cerebral Palsy Undergone Dolphin-Assisted Therapies

Oswaldo Morales Matamoros[1], Erika Yolanda Aguilar del Villar[2], Abril Pérez Sánchez[3],
Jesús Jaime Moreno Escobar[4] and Ricardo Tejeida Padilla[5]
Escuela Superior de Ingeniería Mecánica y Eléctrica, Zacatenco,
Instituto Politécnico Nacional, México[1,2,3]
Escuela Superior de Turismo, Instituto Politécnico Nacional, México[4,5]

***Abstract*—Cerebral palsy is a group of developmental disorders that affects a certain percentage of population motivating the development of several types of therapies ranging from conventional where physical therapies are included to some *alternative therapies* such as the dolphin-assisted therapies (DAT), in order to improve the quality of life of patients suffering these disorders. To find scientific evidence of the DAT effectiveness, in this work is developed a four-stage first-order cybernetic model: i) Signal Acquisition, ii) EEG Processing, iii) EEG Exploring and iv) Healthcare Informatics System (HIS-DAT), in order to explore the electroencephalographic signals behavior from a patient with Infantile Spastic Cerebral Palsy undergone DAT, as well as bioacoustic signals emitted by a female bottlenose dolphin via specialized transducers or passive sensors for aquatic environments, by using nonlinear mathematical tools. We found that the Power Spectrum of signals from EEG and Hydrophone yield similar densities along all DAT and the child's brain activity increases 3-fold higher-frequency when the therapist-dolphin pair interacts with the patient. These findings are supported by the Self-Affine Analysis outcomes, pointing out the emergence of negative correlations from the patient's brain activity during the whole session of DAT but the greatest changes occurred During DAT.**

***Keywords*—Brain–computer interfaces; healthcare system; non-linear dynamics; assisted therapies; infantile spastic cerebral palsy; analysis of EEG signals; power spectrum; self-affine analysis**

I. Introduction

Artificial intelligence (AI) generally is entitled with the development of techniques and devices to perform tasks that involve human interpretation and decision-making. Recently has raised many advances in medical AI applications owing the substantially enhanced computing power of modern computers and the vast amount of digital data available for collection and utilization. AI technologies can ingest, analyze, and report large volumes of data across different modalities to detect disease and guide clinical decisions for health services management and patient care treatments [1]. For instance, in [2] processed electroencephalography (EEG) signals to detecting emotional states tested the effects of meditation music therapy to stabilize mental states in groups of three to anger, calmness, happiness and sadness. They collected 120 emotion signals by using an Emotive 14-channel EEG headset. Therefore, tools developed in AI can be applied to measure and evaluate the effectiveness of therapies carried out to improve the life-quality of patients suffering one or several developmental disorders, based on EEG signals collected from the patients' brain activity undergone assisted therapies.

Cerebral palsy is a group of developmental disorders diagnosed by their disturbance of the subject's movement and posture, caused by early nonprogressive lesions of the central nervous system [3]. In 1862, William Little connected this disorder to a lack of oxygen during birth and described it as a disorder that appeared to strike children in the first year of life, affected developmental skill progression, and did not improve over time. In 1897, Sigmund Freud suggested that cerebral palsy might be rooted in the brain's development in the womb and related aberrant development to factors influencing the developing fetus. Birth asphyxia alone was thought to be the cause of cerebral palsy until the 1980s, when biomedical research found this etiology to be less likely and only one of many with potential to result in cerebral palsy [4].

Cerebral palsy can be defined according to the anatomical site of the brain lesion (cerebral cortex, pyramidal tract, extrapyramidal system, or cerebellum); clinical symptoms and signs (spasticity, dyskinesia [dystonic and choreo-athetotic forms], or ataxia); topographical involvement of extremities (diplegia, quadriplegia, or hemiplegia); timing of presumed insult (prepartum, intrapartum, or postneonatal); and classification of degree of muscle tone (isotonic, hypotonic, or hypertonic) [5], [6].

The main symptoms of cerebral palsy are disorders of movement and posture, but more recently other symptoms have been included in the definition: disturbances of sensation and perception, global or specific cognitive difficulties, communication disorders, behavioral disorders, and seizures [3], [7]. There is no cure for this lifetime condition, but therapy, education, and technology can maximize each child's potential by improving functional abilities and quality of life.

The diagnosis of cerebral palsy is made largely through clinical observations, whose major signs collectively can identify delayed motor milestones, abnormal neurologic examination, persistence of primitive reflexes, and abnormal postural reactions [4]. The Gross Motor Functional Classification System was developed by Rosenbaum and collaborators in 2002 to help professionals anticipate gross motor abilities and severity of cerebral palsy, according to five levels: Level I: able to walk without restrictions; Level II: able to walk without devices; Level III: able to walk with mobility devices; Level IV: self-mobility with limitations; and Level V: self-mobility severely limited [4], [8].

More than 50% of children with cerebral palsy and mental retardation suffer from epilepsy, mostly revealed in the first 4–5 years of life. Epilepsy usually occurred in children with accompanying vision and hearing disorders, and the inability to walk [9]. Besides, children with cerebral palsy often have the following disorders: irregular sleep patterns, lower urinary tract dysfunction, gastrointestinal disorders, drooling, aspiration, swallowing, osteoporosis, malnutrition, inadequate growth, gastroesophageal reflux, impaired language skills or cognitive limitations, among others [10].

Spasticity refers some degree of loss of motor control, abnormal tone, and weakness, which commonly can lead to contractures, scoliosis, and hip subluxation or dislocation. The child with cerebral palsy needs frequent position changes and needs to be out of his or her bed and/or wheelchair frequently at home and at school. Scoliosis may not only progress more rapidly but may continue after skeletal maturity occurs. Progressive hip adduction and flexion is also common, leading to femoral anteversion, subluxation, deformities of the femoral head, and hip dislocation, which can progress to degeneration, sometimes with pain [10].

Treatment of spasticity involves systematic rehabilitation, if necessary, assisted by pharmacotherapy, physiotherapy or surgical interventions, offering benefit, at least in the short term. Most physical therapies are based on the principles of neuroplasticity, patterning, postural balance, muscle strengthening, or stretching [9].

The total degree of overflow brain damage from the initial insult, may not be realized until the preschool or school age period. One of the most limiting comorbidities of cerebral palsy is mental retardation. Children with cerebral palsy that suffer mental retardation can have self-abusive behavior such as biting, head banging, or scratching. These behaviors may stem from anxiety that results from being unable to communicate needs such as hunger or pain. However, children with cerebral palsy who have normal intelligence may have some degree of coexisting visual perceptual abnormalities and learning disabilities that can affect overall development and learning. The assessment of intellectual functioning is often challenging in children with motor impairment because the majority of tests used require verbal and motor responses [10].

Confirming the diagnosis allows the family to plan for long-term treatments and management options that may be needed by the child. Having a diagnosis of cerebral palsy may qualify the child for increased insurance benefits, admission into handicapped preschool or school programs, and assistance in the form of federal assistance. The prediction of the future deficits or abilities at the time of diagnosis often is a *wait and see* situation [4].

Moreover, we want to highlight the importance of trying to improve the functional abilities and quality of life of children with cerebral palsy in developing countries owing. For instead, in [5] it argued that people with *Cerebral palsy* has a high risk for discrimination, neglect, and abuse, particularly those in rural settings owing low school attendance and lack to access to professional assistive. Indeed, individuals with cerebral palsy have a right not only to inclusion but also to full participation in society and pursuit of their hopes and aspirations.

There are several *alternative* therapies used by families to complement systematic rehabilitation (assisted by pharmacotherapy, physiotherapy or surgical interventions), in order to try to improve their children's quality of life. Although there is a little scientific evidence about the effectiveness of these therapies, for some parents it is essential to explore each and every possible intervention that may be available for their children. Regardless their effect, these alternative therapies may offer the parent a feeling of participation or control, but unfortunately they usually cost large sums of money. Some examples of these alternative therapies are the following: massage therapy, horse riding, hyperbaric oxygen, conductive education, patterning, transcutaneous electrical stimulation, chiropractic manipulation, or dolphin-assisted therapy [10].

Since 1971, the *alternative* therapies called dolphin-assisted therapies (DAT) have been proposed to try to improve the quality of life of patients suffering psychological and neurological disorders such as cerebral palsy [11], [12], including Infantile Spastic Cerebral Palsy. In these patients are manifested early nonprogressive lesions of their central nervous system produced during pregnancy, as well as a severe motor delay in the early stages of their development owing their brain undergoes changes in its structure and organization. Hence, these patients have difficulty making more precise movements such as taking a spoon or making a clamp with their fingers and they also lack muscle coordination in some movements.

DAT are based on interaction between live captive dolphins and patients who suffer from various conditions, and this kind of interaction is supervised for both the dolphin trainer and the child therapist. However, an important differentiator that highlights this sort of alternative therapies is the intelligence of these animals, described and being recognized since the development of the Minoan civilization in Greece (3000 BC and 200 BC), and being studied over the years by various researchers [13], [14]. Like in humans, the dolphins' brain is very big regarding their body, therefore communication between both species can be possible [15], [16], [17].

There is evidence [18], [19], [20] about the abilities of dolphins to learn simple tasks with greater efficiency than some primates, allowing to observe the dolphins' ability to transmit the acquired knowledge and complex behaviors learned to other members of their species even for generations [21], [22], [23]. Besides, dolphins have been able to modify their behavior to do the tasks in a more efficient way, showing their interest to be part of the assisted therapies [15], [24]. In part, these dolphin's capabilities make it a better candidate to be part of assisted therapies.

In summary, cerebral palsy affects a certain percentage of the population motivating the development of several types of therapies ranging from conventional wherein physical therapies are included to some *alternative therapies* assisted by animals such as DAT, in order to improve the quality of life of patients with these developmental disorders. To find scientific evidence of the effectiveness of DAT, in this work, based on information technologies, has been developed a Healthcare Informatics System (HIS-DAT), as the last stage of a four-stage first-order cybernetic model, to explore biosignals or biomarkers for measuring mainly quality of patient's medical care. Our HIS-DAT explores the behavior of electroencephalographic (EEG) signals in addition with bioacoustic signals by using nonlinear mathematical tools.

To look for scientific evidence from exploring EEG signals in a patient with Spastic Cerebral Palsy undergone DAT by means of our HIS-DAT, we make use of several tools from both computer science, in particular with parallel computational systems, and non-parametric or non-linear mathematical algorithms for the EEG-signal processing, in order to find emerging patterns at different time scales in the brain activity of a child with spastic cerebral palsy.

Therefore, this article is organized as follows. Section II describes our method designed to collect and process EEG signals. Section III presents the main results obtained after processing both bioacoustic and EEG biosignals. Section IV discusses the results obtained from both the Power Spectral Density and Self-Affine analysis in terms of changes detected in patient's brain Before, During and After DAT. Finally, Section V provides the authors' main thoughts.

II. Method

A. Case Study

During Dolphin-Assisted Therapies (DAT), the fundamental participants are therapist, dolphin's trainer, patient with Infantile Spastic Cerebral Palsy (ISCP) and female bottlenose dolphin. In this work, the cerebral activity of the patient with ISCP and dolphin are measured and explored to determine with scientific evidence the cause-effect correlation between them during DAT. It is worth emphasizing that only one child with ISCP was used as a case study, which we will refer to hereinafter as *Patient*. Characteristics to be taken into account for this research are following:

- Patient with ISCP: Analysis of the cerebral electric fields resulting from interaction between neurons during process of DAT.
- Female bottlenose dolphin: Bioacoustic signal emitted for communication between frequencies of 200 Hz and 150 kHz. Added to these the *clicks* emitted for its echolocation ranging 20 kHz to 150 kHz.

B. Signal Acquisition Tools

Electroencephalography measures the spontaneous electrical activity of the cerebral *cortex*, i.e. the surface layer of the brain. This phenomenon is not due to the simple addition of the action potentials of the 1×10^{11} neurons that make up the brain. The rhythmic variations in the observed waves owed a large population of neurons oscillating coherently. In other words, they all generate a signal of the same frequency. The presence of these rhythms indicates brain activity at the level of millions of neurons acting together synchronously. The form of the EEG signals depends on the age and alertness of patient, changing markedly during childhood. In a healthy brain, this activity is very similar in the different regions of the brain, so there will be no appreciable differences between the different areas of the cortex, called *lobes*. The real-time recording of brain electrical activity lacks a reproducible pattern, which resembles a *chaotic* signal of very small amplitude, around 10-100 µV so it is not difficult to confuse it with a random noise. For this reason, normally studies of the EEG signals focus on its power spectrum via an analysis in frequency bands, but we also introduce a fractal characterization [25] to find emerging patterns from a quantitative nonlinear point of view.

EEG signals are measured as a voltage difference between two electrodes. In order to be able to measure such a small signal, the contact between electrode and skin must be correct. Often the skin receives a preparatory treatment, namely oily skin and dead cells are removed from the surface to arrange a conductive gel or paste to improve contact, in order to measure the impedance between the electrode and the skin. To achieve correct results, the impedance should not be much above 5 $k\Omega$.

Usually there are two configurations for the placement of the electrodes: i) in a referential way and ii) in a bipolar way. The first one is also called *monopolar configuration*, the reference is the same for all electrodes, i.e. when two EEG signals are recorded with a 2-channel EEG module, two active and one reference electrodes would be needed in addition to the ground electrode. Whereas for *a bipolar configuration* each channel has its own reference, so that to record two signals, i.e. four electrodes are necessary in addition to the ground electrode. The ground electrode that is added in both configurations is intended to reduce the common mode, e.g. the annoying interference of the electrical network [26].

Generally an EEG signal is characterized by means of its power spectrum. Traditionally the spectrum of an EEG signal is divided into four frequency bands. Besides, an EEG signal is going to be made up of components from all bands to a greater or lesser extent, but under certain conditions one or the other prevails. An EEG signal can be described as follows:

1) δ Band (0.5-3.5 Hz): Low frequency and high intensity waves (a few hundred µV). They occur in young children and adults only in a state of deep sleep, unconsciousness or situations that increase intercranial pressure such as brain tumors.
2) θ Band (3.5-7.5 Hz): These waves of amplitude less than 20 µV occur during the maturation process in the entire cerebral cortex, although it predominates in the occipital and temporal region and is faster in the frontal area. There are prevailed in children 5-7 aged and are associated in adults and adolescents with creative thoughts, stress or psychic disorders.
3) α Band (7.5-12.5 Hz): These waves of amplitude less than 20 µV are known as *alpha rhythm* and are referred to bursts of 20-100 µV of amplitude and great periodicity at those frequencies predominant over the occipital region but appearing throughout brain cortex. Average frequency of the alpha rhythm in an adult population is about 10.2 Hz. These waves are also associated with states of relaxation, inactivity and are very evident in the absence of visual stimuli[27].
4) β Band (12.5-30 Hz): These small amplitude signals, below 20 µV, are quite common and predominate during adulthood. There are usually divided into low beta, medium beta and high beta. The low beta rhythm is usually localized in the frontal and

occipital lobes and the other two are less localized. There are more irregular than the alpha rhythm and are associated with psychophysical activity, states of agitation, alertness or mental activity when solving problems.

To record the patient's brain activity or Human-Computer Interaction (HCI) in bands δ, θ, α and β, we use the ThinkGear ASIC Module v1.0 (TGAM1) passive sensor developed by the NeuroSky company (see Fig. 1), measuring the frequency bands of brain waves and generating raw time series, i.e. raw sound signals [28], [29]. These sound signals were recorded from the patient with ISCP by the first frontopolar electrode (F_{P1}) of 10-20 System of Electrode Placement [30].

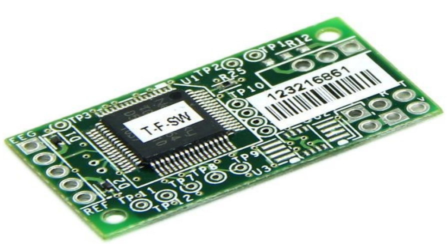


Fig. 1. Signal acquisition tools: Sensor ThinkGear ASIC module v1.0 (TGAM1)

Fig. 2. Signal acquisition tools: Hydrophone 8103 of Brüel & Kjaer

This sensor enables to construct fractal time series to be compared with spectrogram obtained from power analysis of bioacoustic signals, in order to insight relevant information about the dynamics pertaining to the behavior of the patient's brain Before, During and After DAT.

To collect dolphin's bioacoustic signals, a Brüel & Kjaer 8103 Hydrophone and its coupling attachments are used to sound card that allows communication with computer equipment. This small and highly sensitive transducer is used to perform absolute sound measurements in frequency ranging from 0.1 Hz to 180 kHz with a reception sensitivity of -211 dB referred to 1V/μPa. It has high sensitivity relative to its size and good overall characteristics, making it generally applicable for both industrial and educational use. The high-frequency response of Type 8103 is especially valuable when conducting acoustic investigations of marine animals and in measurement of pressure distribution patterns by ultrasound. It is also useful for cavitation measurements. Fig. 2 shows the main external characteristics of the type 8103 hydrophone.

The applied Signal Acquisition Tools are described in detail in [31], while the mathematical tools used in this work are explained in [32].

C. Stimuli

Sonar of dolphins in conjunction with the resonance characteristics of humans are contributing factors to the current interest of this type of alternative therapies by researchers and foundations devoted to the treatment of neurological ailments, mainly in children [15], [16] with developmental disorders .

In [18] is argued that a dolphin orients its sonar by itself to the cranial area of the patient during DAT.

Hence, there is an increment of the Power Spectral Density when dolphin is closer to the patient, and thus a certain resonance frequency is yielded in specific parts of the patient's body (see Table I).

TABLE I. RESONANCE FREQUENCY RANGE OF THE PATIENT'S BODY

Human body parts	Range of Resonance Frequency (Hz)
Arms and shoulders	[2,8]
Chest	[2,12]
Hips	[2,8]
Lower back	[6,12]
Abdomen	[2,14]
Head	[8,27]

Therefore, the main issue is that dolphin's sonar produces resonance of different parts of the human body, yielding effects on the nervous system of the patient via psychological changes [33]. This process is attributed to the generation of endorphin in the brain, bringing about the activation of collagen molecules throughout the body.

Positive effects of DAT have been observed since 1978 in [34], [35], who proposed an experimental model of assisted-therapy to be a starting point for complementing the systematic rehabilitation, highlighting relevant factors in the effectiveness of these alternative therapies such as the type of patient condition, the water temperature, and the time required in a day. Each DAT is different, according to the condition to be treated as well as the patient specific needs [15], [16]. The conclusion of these works is that patient improvement depending only on testimonials explained by parents of the patient to therapists and trainers.

For this reason, some authors in [36] refer to DAT such as a *scientific Heresy* owing there had not a study where the brain activity of patients were measured and explored during DAT from computational and mathematical points of view.

D. Methodology

1) Model: The Healthcare Informatics System developed in this works relies on information technology, allowing to

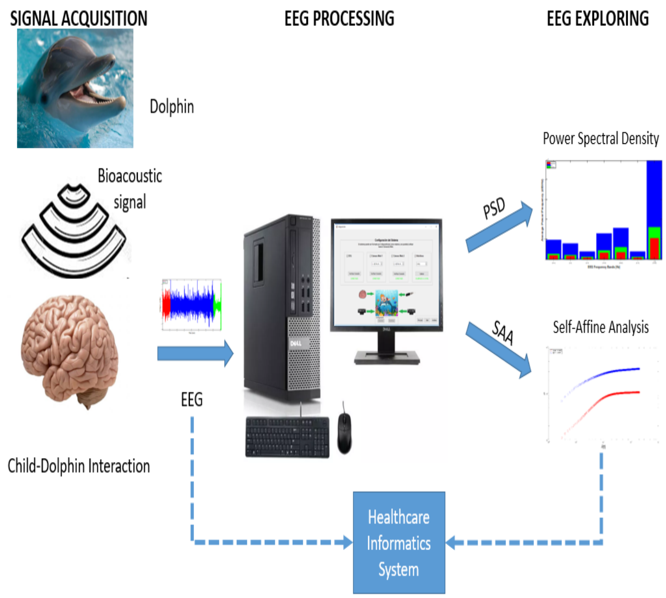


Fig. 3. Four-stage first-order cybernetic model of the healthcare informatics system or HIS-DAT

organize and analyze health (EEG signals) records to measure and evaluate healthcare (effectiveness) outcomes.

Our Healthcare Informatics System is included in a model designed for measuring and assessing the effectiveness of DAT in a patient with ISCP. This model can be summarized in a four-stage first-order cybernetic model (see Fig. 3): i) *Signal Acquisition*, ii) *EEG Processing*, iii) *EEG Exploring* and iv) *Healthcare Informatics System (HIS-DAT)*. The input of this system is the *Signal Acquisition* stage wherein acoustic activity of the dolphin's sonar stimulates the patient's brain activity, considered the Child-Dolphin Interaction. In the *EEG Processing* stage, biosignals acquired in parallel in the previous stage are digitized and transformed into the frequency space. Subsequently, an *EEG Exploring* stage is performed as the output of the system, summarizing the behavior of biosignals by means of both a Power Spectral Density and a Self-Affine Fractal Analysis. Finally, in the fourth stage the Child-Dolphin Interaction is contrasted with non-parametric mathematical methods applied, in order to assess how effective the assisted therapies were for the patient, i.e. HIS-DAT focuses on making sense of the feedback or Input / Output interaction of the entire DAT-system.

2) Procedure: DAT can be carried out both in a tank and in open sea, contributing to the effectiveness of this kind of alternative therapies. There is no a standardized procedure for this type of assisted-therapies, since they vary to adapt to both patient and location conditions. Hence, depending on the sort of facility, DAT are performed differently, but there are prevailing factors [37], [38].

It is noteworthy mentioning that National Polytechnic Institute of Mexico by means of its Ethics Committee in document number D/1477/2020 approved this methodology for sampling and treating not only patients with ISCP but also female bottlenose dolphin.

In Delfiniti-Ixtapa facilities everything starts with an interview to the parents in order to learn more on their children, including the infants' diagnosis and behavior. This is vitally important in children whose communication skills are limited, and helps adjust therapies to the needs of each patient [39], [40].

Already in the water two specialists are present during the process: i) dolphin trainer and ii) therapist. Dolphin trainer supervises the behavior of dolphin while therapist places dolphin in certain parts of the patient body, especially those points that are located on the patient skull. Since a patient is introduced into the water, his/her motor exercises begin to stimulate the movement of the joints and effort in the muscles for carrying out a treatment focused on diseases that attack the central nervous system, regardless of the type of therapies [41], [42], [43]. Moreover, these exercises are performed as a warm-up before DAT and generate trust between patient and therapist, a crucial element during therapies. For patients with motor severity, water provides support in the body enabling them to perform more postures. Besides, water helps exercise muscles and joints, activating blood circulation to induce the desired state of relaxation.

The introduction of dolphin into the procedure is at discretion of the human specialist, taking place until the patient's confidence enables it. In [16] describes how many times in a day dolphin accepted the presence and the interaction of patient, before and after being trained for that purpose. After training dolphin, patient displayed feelings of joy, acceptance and a state-reflective positive [44].

The Delfiniti-Ixtapa personnel have developed their own Intervention Methodology for performing DAT (*IM4pDAT*). Thus, in the interaction with the dolphin, the patient is guided to caress it with complete movements that stimulate orientation and movement in three dimensions: length, height and depth [44]. To perform effectively DAT, the cooperation of both parties is stimulated by rewards, which is known as conditioned behavior (see Fig. 5). If *IM4pDAT* are done correctly, patient is rewarded by caresses, games and activities with dolphin, while dolphin cooperation is rewarded by attention and food as prizes.

Fig. 4 shows a Therapist who indicates to trainer where to orient the dolphin's nose to specific points on the patient's body until a light-touch is performed. Starting with the upper part of the chest where the thymus is located, a gland that develops T lymphocytes and responsible for the cellular immunity of the human body [45]. The dolphin's sonar emission can be perceived by therapist, either by waves in the water or by a sensation in the area where the body is in contact.

Next points to be in touch with dolphin are the frontal lobe located on the patient's forehead, and later in nape of the neck where the medulla oblongata is located, this latter communicates the peripheral nervous system and the spinal cord with the brain. Subsequently, therapist orients dolphin to the middle of head, in order to locate it in pituitary gland, a gland that is part of endocrine system responsible for producing hormones that carry instructions and information among several cell groups. Dolphin is oriented towards occipital and parietal of patient in an upright position, where it is usual that certain discomfort occurs when dolphin is approached to left hemisphere of patient's brain. In [46] it is established left hemisphere is dominant for mathematical comprehension and language. In Delfiniti-Ixtapa a therapies last from 15 to 30 minutes, according the patient's needs [44].

Fig. 4. Dolphin-assisted therapies in mexico: development

Fig. 5. Dolphin-assisted therapies in mexico: dolphin feeding

As mentioned before, DAT are alternative therapies increasingly recommended by foundations to support children with neurological disorders and rehabilitation centers, in addition to being recommended by the Family Integral Development System (DIF, in Spanish) in Mexico [16], [47].

Nonetheless, there are bad practices when DAT are performed, as in all human activity, since it is still a business goal that increases in popularity over the years. Hence, DAT have been seen as a procedure focused mainly on animal rights and without any scientific evidence on its effectiveness. However, in [16], [48], [15] it as also encouraged the emergence of new research to improve techniques to be carried out, in order to assess the effectiveness of this method in the treatment of several disorders, diseases and ailments.

To understand the origin of benefits of DAT and try to offer scientific evidence about its effect in children with ISCP, a Human-Computer Interaction model is developed for the frequency-analysis in sessions lasting 8 minutes, proposing two sub-types of experiments. On the one hand, one session at a pool without a dolphin for 6 minutes measuring patient's brain activity in four exercises:

1) Before therapy in the water (1 min).
2) During therapy in the water without a therapist (2 min).
3) During therapy in the water with the therapist (2 min).
4) After a water therapy (1 min).

On the other hand, one session at the dolphin's tank with salt water, previously informing the completion of DAT and measuring patient's brain activity in five exercises:

1) Before DAT (1 min).
2) During DAT without therapist and without dolphin (2 min).
3) During DAT without therapist and with dolphin (2 min).
4) During DAT with therapist and with dolphin (2 min).
5) After DAT (1 min).

The above mentioned experiments are designed to independently compare the effect in DAT under different factors such as water, wind or dolphins [49]. To verify the expected outcomes for the experimentation, the raw-results were acquired using *IM4pDAT*.

E. Data Acquisition and Analysis

The correlation analysis between bioacoustic signals emitted by dolphin and an EGG signal from patient with ISCP is performed from a software developed in the programming-language MATLAB because of being a programming language oriented to design scripts, i.e. an interpreted language made-up a set of predesigned functions for execution of programs in mathematical language. To perform only the program's final stage, i.e. analysis of collected data, minimum requirements of MATLAB R2016b release are demanded.

Data-acquisition process during DAT requires the following devices:

- Sensor ThinkGear ASIC Module v1.0 (TGAM1).
- 2 Web with Standard or SD Resolution, although it supports high-definition or HD.
- USB-DAQ 8/2-ch production analysis card.
- Charging to Delta Tron converter with integrated TEDS.
- BNC (male) to 10-32 UNF (female) adapter.
- Brüel Miniature hydrophone & Kjær Model 8103 with 10-32 UNF cable (male).

Besides the analysis of biosignals performed in this research, a system programmed in Matlab R2022a is also developed in order to create folders on local disk of computer equipment used for this analysis. These folders store samples collected in each analysis step. Integrated development environments or IDE's are designed for each part of the process to be managed intuitively by users, so that the program execution is divided into modules to provide the user graphical access to system control. Likewise, each module is divided into functions with specific goals for successfully running of the program by dividing the process into sub-sequential processes. Each function is executed by indirect calls, allowing sequential execution of same, as well as transfer of data among them. Fig. 6 and 7 show both IDE Configuration for signal acquisition and the one for storing Basic patient information, respectively.

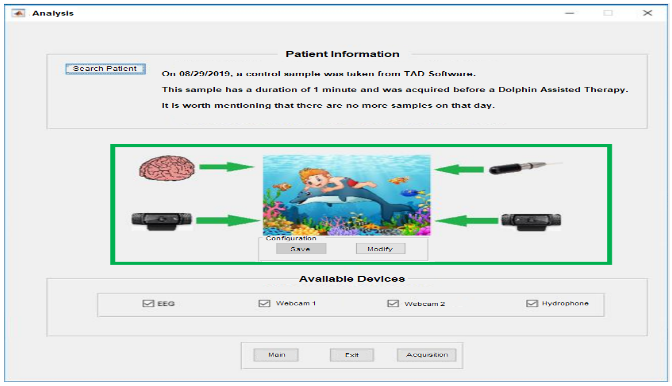


Fig. 6. Data analysis: configuration for signal acquisition

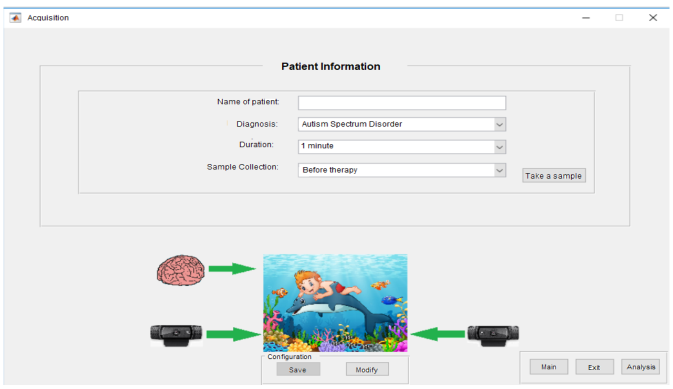


Fig. 7. Data analysis: basic patient information

For the statistical analysis of time series, in this work an Self-Affine (or Fractal) Analysis was performed. A fractal object is scale-invariant or self-affine when it remains invariant under an anisotropic transformation scale (different scales in all directions). Despite their differences, on a transformation scale, directions are not completely independent. If, when zooming in, one of coordinate axes is transformed into a factor b, $x \rightarrow bx$, the rest of coordinate axes must be rescaled into a factor $b\ x_i\ \rightarrow b^{ai}x_i$, in order to preserve the invariant set. Exponent ai is called the Hurst exponent (or dynamic scaling exponent) and tell us what the degree of anisotropy or correlation of the set is [50].

Biosignals, specifically EEG signals, representing a fractal system having self-affinity at a macro level and at a micro level. Small changes create a huge variety of patterns at both micro and macro levels. Fractal analysis makes it possible to determine fractional dimension and detect properties of self-affinity in objects or time series subject to investigation and possessing complex characteristics [51], [52].

In this work, we apply Self-Affine Analysis as theoretical basis for evaluating dynamics of EEG signals of patient with ISCP owing this mathematical tool gives us a notion of the behavior of cerebral signals. From EEG time series it is possible to explore the behavior of many biological complex systems at different time-scales. Expected relationship between values of EEG at time t and at time $t + \tau$ is a correlation relationship in brain activity. Otherwise, a stationary time series has correlation when it depends only on time interval between two observations and falls to zero fast enough while increasing, reflecting the fact that influence of past values decreases with considered delay [50], [53], [51], [52].

Thus, evaluation of Hurst exponent is a first step in recognition and characterization of complex dynamics in EEG time series. This analysis makes it possible to distinguish a random behavior from a non-random one.

Likewise, a EEG time series with some level of pre-visibility will show positive self-correlation. Contrarily, a EEG time series with negative self-correlation has no level of pre-visibility. A Hurst exponent ranging $0.5 < H \leq 1$ corresponds to EEG time series showing a period of growth followed by another analog. This means that one increase is more likely to be followed by a similar one; this is known as positive self-correlation [53].

While values located at $0 < H < 0.5$ correspond to a behavior of one period of growth followed by another of decrease, or vice versa, there is more probability that next period is below average; it has a negative correlation. In addition, $H = 0.5$ corresponds to a random movement; an increase can be followed by a decrease or by a similar one. Finally, the movements do not display any memory, it has auto-correlation equal to zero, namely $H = 0$. All this will give us information about the behavior of patient with ISCP when his/her brain is excited or not by a bioacoustic signal emitted by the dolphin [53].

III. Results

This section is divided into three main parts. The first describes how the performance or test operation of biosignal acquisition equipment is validated. The second explains the Power Spectrum Density results. Finally, the third shows the Self-Affine Analysis results.

A. *Performance Validation*

A data capture was realized using all the sensors for testing the system operation. This test lasted 5 minutes with the dolphin presence to capture the bioacoustic signal by using the 8103 hydrophone. Additionally, a control subject without any diagnosed condition was positioned on the side of the tank to perform the EEG analysis on it.

Once data acquisition process was completed, *Analyzing* IDE option corresponding to *FFT* option was selected in order to verify whether the use of the hydrophone is part of the saved settings. Hence, an indirect call is made to a function to graph the raw-data, both from the hydrophone and from the EEG. Firstly it is set-up the following: i) the hydrophone's sensitivity, ii) the total sample, iii) the analysis time (in this case, 5 minutes), and iv) the folder for the patient as the destination path for the outcomes. Besides the raw-data collected are plotted to yield the time series corresponding to the Volts referred to Pascals (V/Pa) captured by the hydrophone in each $1\ /f_s$ seconds (see Fig. 8). Images generated by webcams are saved as image-files in the patient's folder.

A DAT session was studied from the average frequencies obtained for the histogram of this Experiment, both from the EEG and the hydrophone 8103. These frequencies were estimated in an expected range for a patient in a relaxed state and a dolphin with little activity, respectively.

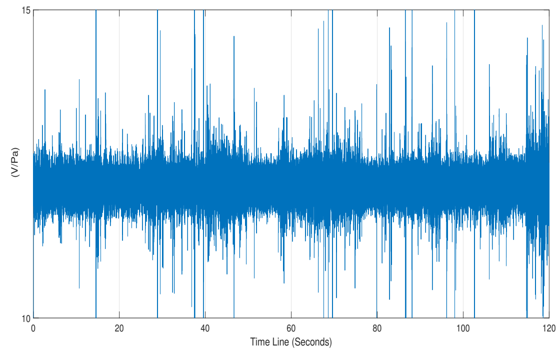


Fig. 8. Time series of the hydrophone resulting from the test

B. Power Spectrum Density

The power density was estimated via an indirect call to a function by adjusting both the minimum frequency to which the hydrophone is sensitive ($f_{min} = 0.1$ Hz) and the sampling frequency ($f_s = 96000$ samples per second). The *pwelch* function is applied inside the MATLAB *Signal Processing Toolbox* to obtain the Welch transform using a kaiser window of size (2/f_{min})xf_s. In Fig. 9 is shown the graph resulting from the estimation of the power density expressed in decibels (dB) and referred to Hertz, for each of the frequencies captured by the hydrophone 8103 during the experimentation.

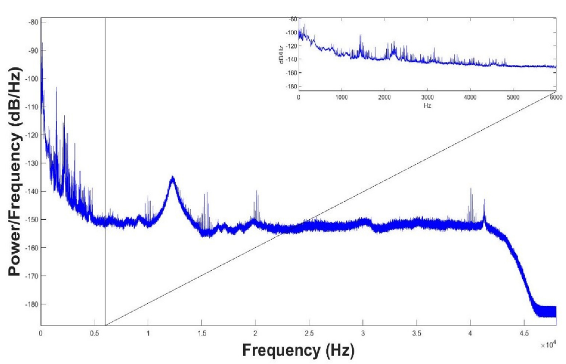


Fig. 9. Periodogram of the hydrophone resulting from the test

Then, the *spectrogram* function is applied inside the MATLAB *Signal Processing Toolbox* with a *kaiser* window of size (2/0.1)xf_s to calculate the Fourier transform in short-time of the frequencies captured every 1 /f_s seconds. This is required to define a normalized-vector to the maximum frequency estimated by the hydrophone. As a result, the spectrogram of Fig. 10 is constructed, depicting a heat-map where each column contains the estimate of the power density of frequencies captured while running the experiment. Fig. 10 and 11 show a spectrogram of biosignals from patient's brain activity and dolphin's echolocation, respectively. These graphs are stored with the patient's name as *image files.jpg* in the selected folder as the *Patient Data*.

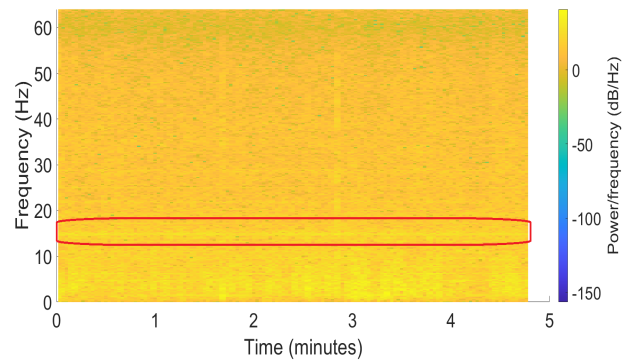


Fig. 10. Spectrogram resulting from the test from EEG

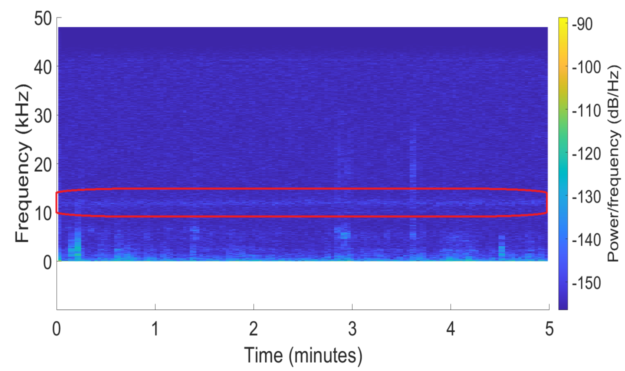


Fig. 11. Spectrogram resulting from the test from hydrophone

Likewise, Fig. 10 and 11 show three-dimensional graphs where not only whole power given by a frequency is observed, but also the moment when a sample manifested certain frequency. Lighter areas point out higher power of a frequency at a certain moment and, on the contrary, dark areas indicate lower power. Thus, after analyzing these two-spectrograms it is found (in red) a clear activation of the brain of patient with ISCP in $\beta = 12$ Hz frequency band, which are related to a state of alert and conscious attention, at the same time when dolphin emits at approximately 12 KHz. This means an indirect activation of the child's brain caused by dolphin's echolocation during DAT.

Fig. 12 shows the resulting histogram from the average of the power of frequencies captured during the experimentation, in intervals of two octaves. Although the hydrophone 8103 is sensitive until 180 KHz, its maximum frequency can be represented as nineteen octaves. For this experimentation only the first-twelve octaves were considered, ranging from 0 to 4096 Hz due to a greater occurrence of the highest powers from seven to ten octaves.

For the data generated from the EEG signals, we also apply the same functions for graphing and analyzing the hydrophone,

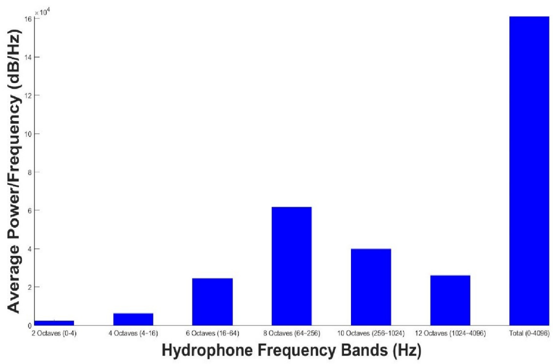


Fig. 12. Histogram of the hydrophone resulting from the test

and performing a similar analysis to the hydrophone without plotting in octaves but plotting in frequency bands of brain activity, such as δ (0.5-4.0 Hz), θ (4.0-8.0 Hz), α (8.0-12.0 Hz), beta β (12.0-30.0 Hz)[54].

Once data processing is finished, a new IDE is opened (see Fig. 13(a)) to choose between viewing the outcomes of the sensor *EEG* (see Fig. 13(b)) or the *Hydrophone* (see Fig. 13(c)). If it is wanted to return to the *Analysis* IDE, it must be pressed the return button. The high-quality versions are located in the same folder selected as the *Patient Data*.

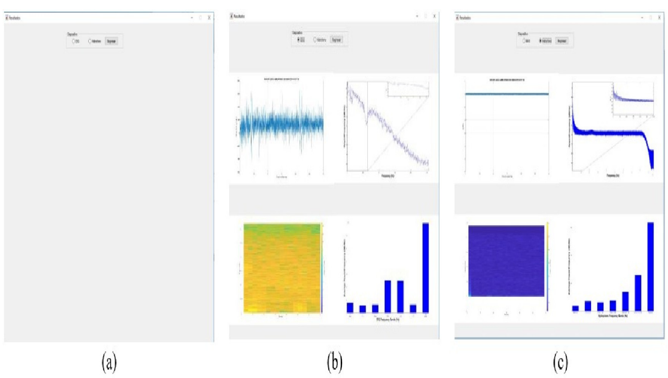


Fig. 13. Results of ISCP patient. (a) window of the Integrated Development Environment (IDE), (b) fast fourier transform for EEG signal, and (c) fast fourier transform for hydrophone biosignal

A frequency analysis during DAT is carried out from the signals captured in this cetacean via the 8103 hydrophone and the EEG signal corresponding to a DAT performed on a patient with ISCP (see Fig. 14 [55]).

In Table II is shown the recorded outcomes corresponding to the average of frequencies obtained from the EEG signals. High-frequencies ($\alpha + \beta$ bands) refer to an attention state and low-frequencies ($\delta + \theta$ bands) correspond to a relaxation state in a human. In the case of water therapy without a dolphin, it is observed a decrease at the end of therapy of 72.77% in the power of the patient's brain activity for low-frequencies and 85.33% for high-frequencies. Contrarily, in the case of DAT was found an increase of 9.73% and 6.85%, respectively, at the end of therapy. In both sub-types of experimentation, the

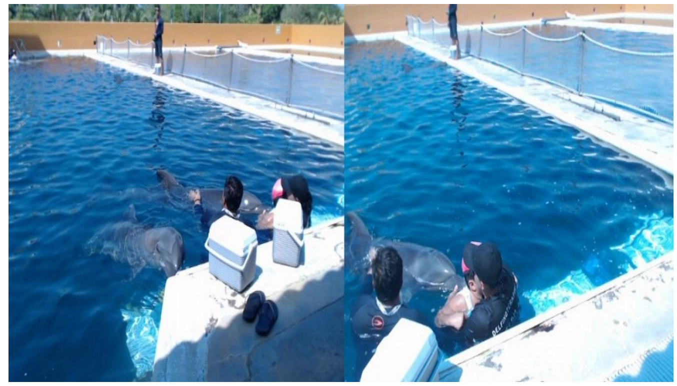


Fig. 14. DAT given to a patient with spastic cerebral palsy in Delfiniti-Ixtapa, Mexico

greatest power is recorded when the main factor of change is introduced in each therapy, being the therapist in the case of therapy in the pool and the dolphin in the case of assisted therapy. Thereby, during DAT the greatest power is recorded between both sub-types of experiments. The increase observed is 160.3% in low-frequencies and 143.84% in high-frequencies regarding the frequencies recorded Before DAT.

C. Self-Affine Analysis

Fig. 15 and 16 show the main outcomes for RAW-EEG brain activity. Fig. 15 shows the cerebral activity of patient with ISCP undergo DAT with dolphin and therapist. Fig. 16 presents the results in terms of voltage fluctuations regarding brain activity, estimated by a Self-Affine Analysis. The three-curves depict the voltage fluctuation behavior at different time-scales: i) in red color the sample taken Before DAT, ii) in blue color the sample taken During DAT, and iii) in green color the sample taken After DAT. The resulting values of the Hurst exponent are: i) Before DAT $H = 0.4133$, ii) During DAT $H = 0.3898$, iii) and After DAT $H = 0.4462$; in all cases the behavior of patient's brain activity is antipersistent or displayed negative correlations ($0 < H < 0.5$)., i. e. there is a constant change in the patient's brain activity

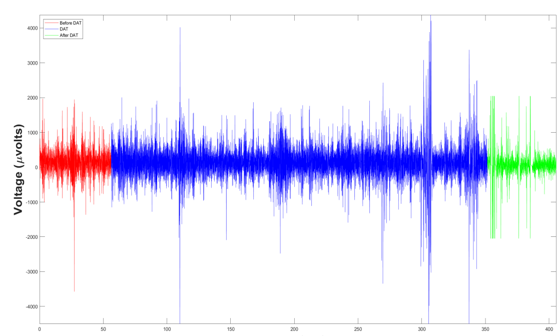


Fig. 15. Main results for RAW EEG brain activity of the child with ISCP: i) in red color Before DAT, ii) in blue color During DAT, and iii) in green color After DAT

TABLE II. RESULTS FROM PATIENT WITH ISCP, EXPERIMENT CARRIED OUT ON JANUARY 25, 2020. THE BOLD BARS INDICATE THE BEST THERAPEUTIC EFFICIENCY IN TERMS OF POWER DENSITY

Therapy	Bands	Before	During	After	$\delta+\theta+\alpha+\beta$ Before	$\delta+\theta+\alpha+\beta$ During	$\delta+\theta+\alpha+\beta$ After
Pool without therapist	$\delta+\theta$	423.65	256.56	115.36	1097.75	563.28	214.28
	$\alpha+\beta$	674.1	306.72	98.92			
Pool with therapist	$\delta+\theta$	423.65	365.08	115.36	1097.75	771.99	214.28
	$\alpha+\beta$	674.1	406.91	98.92			
DAT without dolphin and therapist	$\delta+\theta$	262.48	222.89	288.02	596.98	649.42	645.43
	$\alpha+\beta$	334.5	426.53	357.41			
DAT without dolphin and with therapist	$\delta+\theta$	262.48	217.5	288.02	596.98	477.77	645.43
	$\alpha+\beta$	334.5	260.27	357.41			
DAT with dolphin and therapist	$\delta+\theta$	262.48	**683.26**	288.02	596.98	**1498.9**	645.43
	$\alpha+\beta$	334.5	**815.64**	357.41			

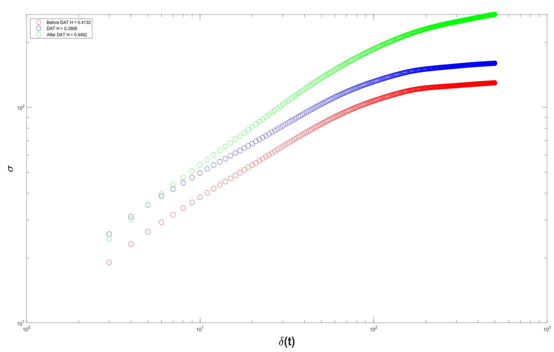


Fig. 16. Results form self-affine analysis describing the voltage fluctuation behavior at different time-scales: i) in red color before DAT, ii) in blue color during DAT, and iii) in green color After DAT

IV. DISCUSSION

From Table II, it can be observed that the greatest power occurs both at low-frequencies ($0.5 - 8.0\ Hz \rightarrow \delta + \theta$) and at high-frequencies ($8.0 - 30\ Hz \rightarrow \alpha + \beta$) when performing DAT with a therapist working on the patient. In addition, when frequencies ranging from 0.5 to 30.0 Hz ($\delta+\theta+\alpha+\beta$), DAT gets the greatest power during therapy. Moreover, a simulated DAT in a pool is working as a relaxant since it has a tendency to reduce the power of brain activity. While tendency of DAT is to enter with a certain power, it increases almost 3-fold the power and produces a greater effect than initial therapy, as well as it keeps high-frequencies responsible for concentration.

In [56] and in [57] have been pointed out that a complete conventional therapy should increase the brain's highest-frequencies in conditions, e.g. Attention Deficit Hyperactivity Disorder since it makes patient's concentration be kept, an effect that does occur in DAT.

Concerning to outcomes resulting from the Self-Affine Analysis, it is observed: i) Before DAT $H = 0.4133$, ii) During DAT $H = 0.3898$, and iii) After DAT $H = 0.4462$. Thus, in the full-session of DAT the patient's brain activity displays only negative correlations at different time-scales, i.e. the patient's brain activity increases and decreases all time long, but the greatest changes occurred During DAT due to it yields the lowest Hurst exponent ($H = 0.3898$), i.e. the most negative correlation displayed.

Finally, it is worth mentioning that it is not possible to compare the results obtained in this article with those yielded by other related articles on the topic owing non-parametric and non-linear mathematical methods such as this work are not used and are only limited in many cases in the opinions of parents or simply basic descriptive statistics such as the mean or standard deviation. Besides, this work contains just a single case study, which gives a first approximation of what happened with the brain activity of a patient with cerebral palsy during DAT, but it is the basis for future research, which should have at least both control and intervention subjects in order to strength our findings.

V. CONCLUSIONS

Family members of patients with Infantile Spastic Cerebral Palsy (ISCP) appeal to both conventional and alternative therapies such as dolphin-assisted therapies (DAT). This work trials to find scientific evidence about the effect of DAT in a child with ISCP for improving his functional abilities and quality of life. Therefore, we develop a four-stage first-order cybernetic model: i) Signal Acquisition, ii) EEG Processing, iii) EEG Exploring and iv) Healthcare Informatics System (HIS-DAT) to measured the brain activity of this patient by collecting and exploring two passive-sensor types of biosignals: i) Electroencephalographic and ii) Sonar echolocation of a female bottle-nose dolphin.

We found that the Power Spectrum of signals from EEG and Hydrophone signals yield similar densities along all DAT and increases 3-fold higher-frequency brain activity when the therapist-dolphin pair interacts with the patient. Increments in higher-frequencies, ranging from 12 to 30 Hz ($\alpha + \beta$), corresponding to concentration activities can help not only in patients with ISCP but also in other disorders such as Attention Deficit Hyperactivity Disorder. These findings are supported by the Self-Affine Analysis outcomes, pointing out the emergence of negative correlations from the patient's brain activity during the whole session of DAT but the greatest changes occurred During DAT. Usage of a single child as a case study shows first results of how this patient's brain activity behaves when this brain has neuronal disability and it is auto-stimulated by dolphin's sonar.

The proposed HIS-DAT model stores and retrieves all applicable information regarding patients undergone DAT. HIS-DAT also allows doctors, therapists, dolphins' trainers, and parents to have quick, easy, and efficient access to changes in patients' brain activity Before, During and After DAT in order to evaluate its effectiveness.

ACKNOWLEDGMENT

The research described in this work was carried out at Superior School of Mechanical and Electrical Engenieering (Escuela Superior de Ingeniería Mecánica y Eléctrica) of the Instituto Politécnico Nacional, Campus Zacatenco. It should be noted that this research is part of doctoral thesis entitled *Modelo sistémico para determinar la eficacia de una terapia asistida por delfines* supported by Erika Yolanda Aguilar del Villar, work directed by Dr. Oswaldo Morales Matamoros and Dr. Ricardo Tejeida Padilla, in adition with *Modelo de sistema viable para organizaciones de turismo de salud, basado en la delfinoterapia.* supported by Abril Pérez Sánchez, work directed by Dr. Oswaldo Morales Matamoros and Dr. Jesús Jaime Moreno Escobar.

REFERENCES

[1] S. Secinaro, D. Calandra, A. Secinaro, V. Muthurangu, and P. Biancone, "The role of artificial intelligence in healthcare: a structured literature review," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 125, Dec. 2021. [Online]. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01488-9

[2] N. V. Kimmatkar and B. V. Babu, "Novel approach for emotion detection and stabilizing mental state by using machine learning techniques," *Computers*, vol. 10, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2073-431X/10/3/37

[3] L. Bottcher, "Children with Spastic Cerebral Palsy, Their Cognitive Functioning, and Social Participation: A Review," *Child Neuropsychology*, vol. 16, no. 3, pp. 209–228, Apr. 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/09297040903559630

[4] M. W. Jones, E. Morgan, J. E. Shelton, and C. Thorogood, "Cerebral Palsy: Introduction and Diagnosis (Part I)," *Journal of Pediatric Health Care*, vol. 21, no. 3, pp. 146–152, May 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0891524506004068

[5] A. Colver, C. Fairhurst, and P. O. D. Pharoah, "Cerebral palsy," *The Lancet*, vol. 383, no. 9924, pp. 1240–1249, Apr. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0140673613618358

[6] C. Fanizza, R. Maglietta, G. Buscaino, R. Carlucci, M. Ceraulo, G. Cipriano, R. Grammauta, V. Renò, F. C. Santacesaria, L. Sion, and E. Papale, "Emission rate of acoustic signals for the common bottlenose and striped dolphins in the gulf of taranto (northern ionian sea, central-eastern mediterranean sea)," in *2018 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, 2018, pp. 188–192.

[7] R. Carlucci, A. A. Bas, R. Maglietta, V. Renò, C. Fanizza, A. Rizzo, R. Crugliano, and G. Cipriano, "Site fidelity, residency and habitat use of the risso's dolphin grampus griseus in the gulf of taranto (northern ionian sea, central-eastern mediterranean sea) by photo-identification," in *2018 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, 2018, pp. 173–177.

[8] X. Qing, S. Liu, G. Qiao, Y. Dong, S. Ma, and D. He, "Acoustic propagation investigation of a dolphin echolocation pulse at water-sediment interface using finite element model," in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, 2018, pp. 1–4.

[9] M. Sadowska, B. Sarecka-Hujar, and I. Kopyta, "Cerebral Palsy: Current Opinions on Definition, Epidemiology, Risk Factors, Classification and Treatment Options," *Neuropsychiatric Disease and Treatment*, vol. 16, pp. 1505–1518, 2020.

[10] M. W. Jones, E. Morgan, and J. E. Shelton, "Primary Care of the Child with Cerebral Palsy: A Review of Systems (Part II)," *Journal of Pediatric Health Care*, vol. 21, no. 4, pp. 226–237, Jul. 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0891524506004160

[11] S. Birch, "Dolphin sonar pulse intervals and human resonance characteristics," in *Proceedings of the 2nd International Conference on Bioelectromagnetism (Cat. No.98TH8269)*, 1998, pp. 141–142.

[12] Z. Zhaohui and P. Yuan, "Study on the characteristics of dolphins click signal," in *2020 5th International Conference on Communication, Image and Signal Processing (CCISP)*, 2020, pp. 76–80.

[13] G. Cipriano, S. Gatto, C. Cherubini, R. Crugliano, F. C. Santacesaria, S. Bellomo, C. Fanizza, E. Papale, G. Buscaino, R. Maglietta, P. Ricci, and R. Carlucci, "Variation in the emission rates of striped dolphin's vocalizations recorded in the gulf of taranto (northern ionian sea, central mediterranean sea)," in *2022 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, 2022, pp. 236–241.

[14] J. Liu, X. Yang, C. Wang, and Y. Tao, "A convolution neural network for dolphin species identification using echolocation clicks signal," in *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2018, pp. 1–4.

[15] A. of Marine Mammal Parks & Aquariums, "Delfín nariz de botella," *Alliance of Marine Mammal Parks & Aquariums*, 2017. [Online]. Available: https://www.ammpa.org/sites/default/files/files/animalfactsheets/AMMPA-DolphinFactSheet-SPANISH-PRINT.pdf

[16] Delfiniti, "Delfinoterapia," *TERAPIAFISICA*, 2017. [Online]. Available: https://www.terapia-fisica.com/delfinoterapia/

[17] nosinperro.es, "Intervención asistida con animales," *nosinperro.es*, 2020. [Online]. Available: https://www.youtube.com/watch?v=y2nyS3meprM

[18] S. Birch, "Dolphin sonar pulse intervals and human resonance characteristics," in *Proceedings of the 2nd International Conference on Bioelectromagnetism (Cat. No.98TH8269)*. IEEE, 2002.

[19] E. Wille, "Dolphin assisted psychotherapy: An adjunct to evidence-based treatment for substance abuse disorders," *carf International*, 2017. [Online]. Available: http://www.carf.org/Dolphin-Assisted-Therapy/

[20] H. Sugimatsu, J. Kojima, T. Ura, R. Bahl, V. S. Sagar, and R. Chauhan, "Real-time automatic estimation of the number of migrating ganges river dolphins (platanista gangetica) during the acoustic census by using a mobile four-hydrophone array system," in *OCEANS 2017 - Anchorage*, 2017, pp. 1–7.

[21] C. K. A. Lim, W. C. Chia, and S. W. Chin, "A mobile driver safety system: Analysis of single-channel eeg on drowsiness detection," in *2014 International Conference on Computational Science and Technology (ICCST)*, Aug 2014, pp. 1–5.

[22] C. Lack, "Obsessive-compulsive disorder: Evidence-based treatments and future directions for research," *World journal of psychiatry*, vol. 2, pp. 86–90, 12 2012.

[23] T. Wichmann, "Detail zooming in artificial intelligence world modeling," in *SoutheastCon 2015*, April 2015, pp. 1–5.

[24] H. Sugimatsu, J. Kojima, S. Nam, T. Ura, R. Bahl, V. S. Sagar, and R. Chauhan, "Improvement of the video camera system mounted on a balloon for supporting the visual census of river dolphins," in *OCEANS 2018 MTS/IEEE Charleston*, 2018, pp. 1–6.

[25] S. Sakuraba, H. Kobayashi, S. Sakai, and K. Yokosawa, "Alpha-band rhythm modulation under the condition of subliminal face presentation: Meg study," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 6909–6912.

[26] C. Haddix, A. F. Al-Bakri, W. Besio, and S. Sunderam, "A comparison of eeg alpha rhythm detection by tripolar concentric ring electrodes and conventional disk electro des," in *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2018, pp. 068–072.

[27] J. Bhattacharya, P. Kanjilal, and S. Nizamie, "Decomposition of posterior alpha rhythm," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 6, pp. 738–747, 2000.

[28] K. G. Li, M. I. Shapiai, A. Adam, and Z. Ibrahim, "Feature scaling for eeg human concentration using particle swarm optimization," in *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Oct 2016, pp. 1–6.

[29] C. O. S. Jimenez, H. G. A. Mesa, G. Rebolledo-Mendez, and S. de Freitas, "Classification of cognitive states of attention and relaxation using supervised learning algorithms," in *2011 IEEE International Games Innovation Conference (IGIC)*, Nov 2011, pp. 31–34.

[30] J. A. I. R. Silva, F. E. Suarez Burgos, and S.-T. Wu, "Interactive

visualization of the cranio-cerebral correspondences for 10/20, 10/10 and 10/5 systems," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016, pp. 424–431.

[31] J. J. Moreno Escobar, O. Morales Matamoros, R. Tejeida Padilla, L. Chanona Hernández, J. P. F. Posadas Durán, A. K. Pérez Martínez, I. Lina Reyes, and H. Quintana Espinosa, "Biomedical signal acquisition using sensors under the paradigm of parallel computing," *Sensors*, vol. 20, no. 23, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/23/6991

[32] O. M. Matamoros, J. J. M. Escobar, R. Tejeida Padilla, and I. Lina Reyes, "Neurodynamics of patients during a dolphin-assisted therapy by means of a fractal intraneural analysis," *Brain Sciences*, vol. 10, no. 6, 2020. [Online]. Available: https://www.mdpi.com/2076-3425/10/6/403

[33] M. L. M. Duarte and M. de Brito Pereira, "Vision Influence on Whole-Body Human Vibration Comfort Levels," *Shock and Vibration*, vol. 13, no. 4-5, pp. 367–377, 2006. [Online]. Available: http://www.hindawi.com/journals/sv/2006/950682/abs/

[34] D. E. Nathanson and S. de Faria, "Cognitive improvement of children in water with and without dolphins," *AnthrozoÃ¶s*, vol. 6, no. 1, pp. 17–29, 1993.

[35] D. E. Nathanson, D. de Castro, H. Friend, and M. McMahon, "Effectiveness of short-term dolphin-assisted therapy for children with severe disabilities," *AnthrozoÃ¶s*, vol. 10, no. 2-3, pp. 90–100, 1997.

[36] P. Weisleder, "Unethical prescriptions: Alternative therapies for children with cerebral palsy," *Clinical Pediatrics*, vol. 49, no. 1, pp. 7–11, 2010, pMID: 19628756.

[37] Y. Cai, N. K. H. Chia, D. Thalmann, N. K. N. Kee, J. Zheng, and N. M. Thalmann, "Design and development of a virtual dolphinarium for children with autism," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 208–217, 2013.

[38] Liling, Dupeng, and Zhangzhaohui, "Bottlenose dolphin echolocation clicks characteristics acquisition and analysis," in *2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2019, pp. 1–4.

[39] S. Birch, "Dolphin sonar pulse intervals and human resonance characteristics," in *Proceedings of the 2nd International Conference on Bioelectromagnetism (Cat. No.98TH8269)*, Feb 1998, pp. 141–142.

[40] K. B. and M. D., "Dolphin assited therapy: Evaluation of the impact in neuro-sensory-motor functions of children with mental, behavioural and neurodevelopmental disorders," *Revista Argentina de Clínica psicológica*, vol. 29, no. 4, pp. 292–307, 2020.

[41] G. Richard, van der Steen Steffie, C. R. F.A., V. Theo, and E.-S. Marie-Jose, "Verbal interactional synchronization between therapist and children with autism spectrum disorder during dolphin assisted therapy: Five case studies," *Animals*, vol. 9, no. 10, p. 716, 2019.

[42] B. Kreiviniene, D. Mockeviciene, Z. Kleiva, and V. VaiSvilaite, "The psychosocial effect of therapeutic activities with dolphins for children with disabilities," *SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference*, vol. 3, p. 94, 05 2019.

[43] R. E. Griffioen and M.-J. Enders-Slegers, "The effect of dolphin-assisted therapy on the cognitive and social development of children with down syndrome," *Anthrozoös*, vol. 27, no. 4, pp. 569–580, 2014.

[44] D. de México S.A. de C.V., "Delfinoterapia (tad)," *Delfiniti Ixtapa.*, 2019.

[45] G. R. Brandan Nora and K. Sofía, "Linfocitos t," *Universidad Nacional del Nordeste*, 2019. [Online]. Available: https://med.unne.edu.ar/sitio/multimedia/imagenes/ckfinder/files/files/Carrera-Medicina/BIOQUIMICA/linfot.pdf

[46] N. M. Rubio, "Hemisferio cerebral izquierdo: partes, características y funciones," *Psicología y Mente*, 2019. [Online]. Available: https://psicologiaymente.com/neurociencias/hemisferio-cerebral-izquierdo

[47] R. R. Noticias, "Promueve dif la terapia asistida por delfines para niños con discapacidad," *RMX Región Noticias*, 2016. [Online]. Available: http://rmxnoticias.com/locales/promueve-dif-la-terapia-asistida-por-delfines-para-ninos-con-discapacidad/

[48] P. G. Yeves, "¿qué ocurre con la experimentación con animales?" *La mente es maravillosa*, 2019. [Online]. Available: https://lamenteesmaravillosa.com/que-ocurre-con-la-experimentacion-con-animales/

[49] A. Maujean, C. A. Pepping, and E. Kendall, "A systematic review of randomized controlled trials of animal-assisted therapy on psychosocial outcomes," *Anthrozoös*, vol. 28, no. 1, pp. 23–36, 2015.

[50] C. Peng, J. Mietus, J. Hausdorff, S. Havlin, H. Stanley, and A. Goldberger, "Long-range anti-correlations and non-gaussian behaviour of the heartbeat," *Phys Rev Lett*, vol. 70, p. 1343, 1993.

[51] A. S. Balankin, "Dynamic scaling approach to study time series fluctuations," *Phys. Rev. E*, vol. 76, p. 056120, Nov 2007.

[52] F. J, *Fractals*. Plenum Press, 1988.

[53] A. Bunde, S. Havlin, J. Kantelhardt, T. Penzel, J. Peter, and K. Voigt, "Correlated and uncorrelated regions in heart-rate fluctuations during sleep," *Phys Rev Lett*, vol. 85, p. 3736, 2000.

[54] B. Senevirathna and P. Abshire, "Spatio-temporal compressed sensing for real-time wireless eeg monitoring," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–5.

[55] J. Kantelhardt, S. Tismer, F. Gans, A. Schumann, and P. T., "Scaling behavior of eeg amplitude and frequency time series across sleep stages," *EPL (Europhysics Letters)*, vol. 112, no. 1, p. 18001, October 2015.

[56] M. Arns, C. K. Conners, and H. C. Kraemer, "A Decade of EEG Theta/Beta Ratio Research in ADHD: A Meta-Analysis," *Journal of Attention Disorders*, vol. 17, no. 5, pp. 374–383, Jul. 2013. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1087054712460087

[57] A. Caye, J. M. Swanson, D. Coghill, and L. A. Rohde, "Treatment strategies for ADHD: an evidence-based guide to select optimal treatment," *Molecular Psychiatry*, vol. 24, no. 3, pp. 390–408, Mar. 2019. [Online]. Available: http://www.nature.com/articles/s41380-018-0116-3

Dynamic Light Settings as Data Augmentations for Automated Scratch Detection

GRAVE Valentin[1], FUKUDA Osamu[2], YEOH Wen Liang[3],
OKUMURA Hiroshi[4], YAMAGUCHI Nobuhiko[5]
Graduate School of Science and Engineering, Saga University, Saga, Japan

***Abstract*—The manufacture of plastic parts requires a rigorous visual examination of its production to avoid the shipment of some that would be defective to its customers. In an attempt to ease the detection of scratches on plastic parts, the prototype of a computer-assisted visual inspection system was developed. The aim of this paper is to introduce how we explored ways to design a semi-automatic system comprising of a lamp whose orientations and intensities help in revealing irregularities on subjects that would have been missed with a unique light configuration. This process was qualified as "hardware data augmentation". The pictures collected by our system were then used to train several convolutional neural networks (YOLOv4 algorithm/architecture). Finally, the performances of their models were confronted to evaluate the effects of the different light settings, and deduce which parameters are favourable to capture datasets leading to robust defect detection systems.**

***Keywords*—Augmentation technique; deep neural network; image processing; light emission; object detection**

I. Introduction

The appearance of a product is the first and foremost characteristic that will attract, or repel, a customer. One of the key factors for an "appealing aspect" is the lack of flaws in the final product. Indeed, if defects are visible on an item, it will be considered poorly made. Therefore, it is essential for a manufacturer to carefully design its processes so that irregularities become as rare as possible.

Companies have been striving to produce quality products by implementing verification procedures in crucial stages of their production. In the plastic industry, for example, it is frequent for operators to be tasked with observing the surfaces of plastic parts under some light and comparing them with typical flawless pieces as validation standards. These pieces are moved to different positions, so the incident light may expose defects that would otherwise have remained invisible. However, visually inspecting parts is a fairly laborious and time-consuming burden for a human and the accuracy of its outcome may be unreliable. Among the few factors that can alter the judgment of an operator, their lack of experience or assiduity may certainly be the most common [1].

With the advent of artificial intelligence (AI), mostly deep learning techniques, applied to the detection of elements in pictures or video streams, industries have been trying to develop systems that would automatically notify the presence of defective items in their production lines. Nowadays, the popular method to implement such an AI is via the usage of Convolutional Neural Network (CNN) models [2], [3], [4], whose algorithms and architectures enable them to extract features from pictures (or videos) and classify their content through pattern matching.

Although current CNNs are quite efficient in detecting familiar categories of objects within pictures, they are still perfectible when used for specific tasks in specialized fields. For example, when integrated as monitoring tools on production lines of plastic or metallic parts, the scarcity of defective training data may lead to poor model performances. A substantial and diverse dataset is often required for a deep learning algorithm, such as CNNs, to produce an efficient model [5]. Moreover, another problem comes from the fact that an operator has to move and orientate the pieces in different positions to reveal flaws that would not be noticeable with a unique point of view and light incidence, since these movements are hardly ever performed automatically by an inspection system.

It is not rare for studies to explore ways that would expand limited datasets or boost the efficiency of deep learning models, but these researches generally involve software solutions revolving around image processing [6], [7]. In our case, we looked into a hardware method to enhance the quality of artificial networks' trainings through the acquisition of datasets whose elements have undergone physical data augmentations, as opposed to artificial data augmentations. These physical "data augmentations" come as fluctuations of both the brightness and the orientation of a light exposing a target (i.e. the object to be represented in the sets). This concept also aims to be useful to help models in detecting defects, as the visibility of the details of an item may vary with the intensity and the angle of incidence of its illuminating lamp. These multiple lighting conditions act as the movements an operator performs during their inspection: moving items into different positions and adjusting the brightness to facilitate their judgment, but unlike these manual inspections, our concept aims to perform these operations automatically.

II. Related Work

Researchers have investigated numerous means of performing data augmentations for both the expansion of insufficient assortments of pictures and for the creation of robust models. Our study investigates yet another novel approach to collecting and exploiting data targeting computer-assisted visual inspection of surfaces.

In deep learning, the size and diversity of a dataset are critical to creating a decent model [5], in case of a restricted set of training samples circumventing these limitations may be required. Therefore, resorting to data augmentation is a common preprocessing step in deep learning [6], [7] (geometric transformation, noise addition, mix of images, etc.). Granted that the affinity of the new samples estimates them close to the originals, this artificial inflation of a dataset's population permits reducing the risks of a model overfit while improving its feature extraction capabilities, thus enhancing its adaptability to new data [8].

Data augmentation involving synthetic manipulations of the illumination of a picture is not unusual, whether it be by modifying their "general" brightness [9], or by adding a "synthetic" light source and shifting its position [10]. However, the majority of previous works revolve around software augmentation to generate their data, which often results in unnatural representations that may not be relevant to the models.

Moreover, some softwares and appliances aiding in the inspection of plastic parts already exist, such as the SavvyInspector™ [11] from Savvy Optics Corp., but they often do not include automatic image processing technologies nor consider different lighting configurations. Several studies also aimed at facilitating the examination of surfaces, but their experiments often did not take the orientation of the light into consideration [12], [13], [14].

Our study focused on the creation of a simple system that made possible the acquisition of data that are physically "augmented" through the different light settings used during their capture. This method is useful to generate specimens that are representative of occurrences a model will face in a live setting. Pictures of defective plastic parts were collected to train models, and the performances of these models on common test samples have subsequently been compared.

III. Image Capturing System

In current industrial environments, visual inspections necessitating rigorous examinations are still predominantly performed manually. So as to expose as many potential defects as possible on products, the common procedure is to scrutinise them under multiple lighting configurations. Operators are required to adjust the positions of the objects' surfaces, such as in Fig. 1, so the angles of incidence of the rays emitted by the light source can reveal deficiencies [15]. They may also modify the brightness of the source, since it has an obvious influence on the global visibility of the subjects they observe. Nevertheless, an equilibrium is to be found: while a dim light makes it difficult to distinguish details -if anything-, a bright light can also alter vision by flooding a defective area, covering its flaws with interfering glares.

In order to observe the effects distinct illumination settings of samples have on the efficiency of a CNN model's training, we have developed a semi-automatic device equipped with a light that automatically rotates around a subject while taking pictures as it reaches a designated set of positions. Its diagram is illustrated in Fig. 2 and its prototype is shown in Fig. 3. However, the current system version still needs the user to change the intensity of the light manually.

This system enables the capture of pictures with subjects illuminated from multiple angles and with different lighting brightness. The current prototype allows 256 levels of luminosity and has an amplitude of about 270°, since the support

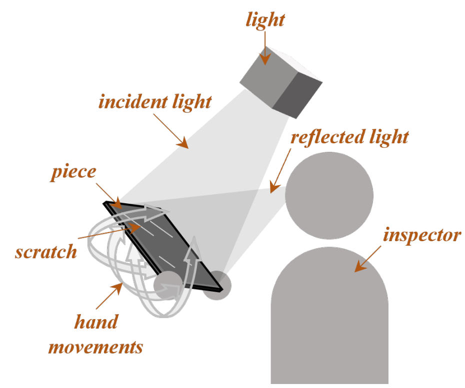


Fig. 1. Diagram of a manual inspection

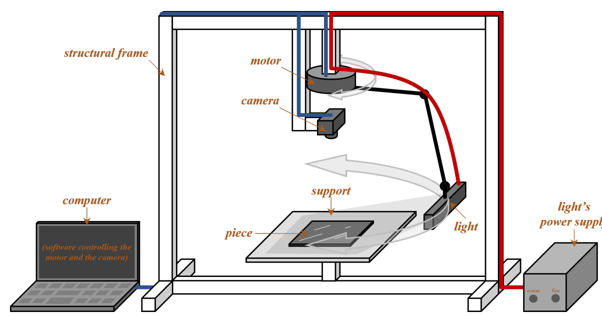


Fig. 2. Diagram of our automatic inspection setup

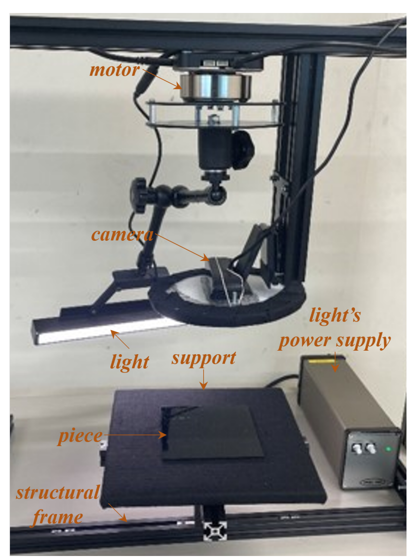


Fig. 3. View of the actual prototype

of the camera makes a full revolution impossible (as it gets in the way of the arm holding the bar light as it rotates). Because the target observed persists over a given set of light configurations, we qualified the process by which the variety of images was captured using our system as "hardware data augmentations", as opposed to a regular collection of pictures that would have been taken in a unique light setting. The augmentations thus engendered by our system took various forms: appearances/disappearances of scratches, variations of their shapes as well as shifts of their colors (or that of their supports). Several images were thereby collected to train multiple models on data having different light exposures. Their incorporations into datasets are detailed further in a subsequent section.

A. Hardware Components

Our device measures 50 by 50 by 50cm. Its constituting elements, visible on the photograph in Fig. 3, are the following:

- A camera that collects shots of plastic pieces. Created by Logitech, the C920 is a 1080p webcam that permits a stream of 30 frames per second.
- An electrical motor that orientates the light. The KM-1U is manufactured by Keigan Motor. It can be controlled programmatically and has a step resolution of about 0.05°.
- A light that illuminates the subjects that are observed. This light bar, model TLB245x25-22WD-4, is made by Aitec System. It can accept up to 24V and 0.3A, which results in a maximum power of 7.2W.
- An adjustable power supply that regulates the light intensity. The TPDP1B-2450NCW is made by Aitec System. It has 256 (16*16) gradations and it can output up to 24V±0.5V and 1.85A.

Thanks to its 256 gradations, the power supply made it possible to fine-tune the luminosity level of the light. Since the latter can accept up to 7.2W, each gradation of the power supply represents a fluctuation of 0.028125W (7.2W / 256).

Finally, to run the software that operates our prototype and to carry the CNN trainings, we used a laptop whose operating system was Windows 10 and which included an Intel i7-8750H @2.20GHz CPU, an NVIDIA GTX 1060 GPU and 16GB of RAM.

B. Operating Software

A custom software had to be developed to enable a user to control the acquisition device we created. Considering that the operating system of our working station was Windows, we thought that designing a program in C# on Visual Studio was the most practical. The User Interface (UI) of the current version is shown in Fig. 4. This UI is divided into three main sections:

- **Configuration**: in this column are listed the elements related to the establishment of connections with the external peripherals (motor and camera) and the creation of the base file structure where captures will be stored.

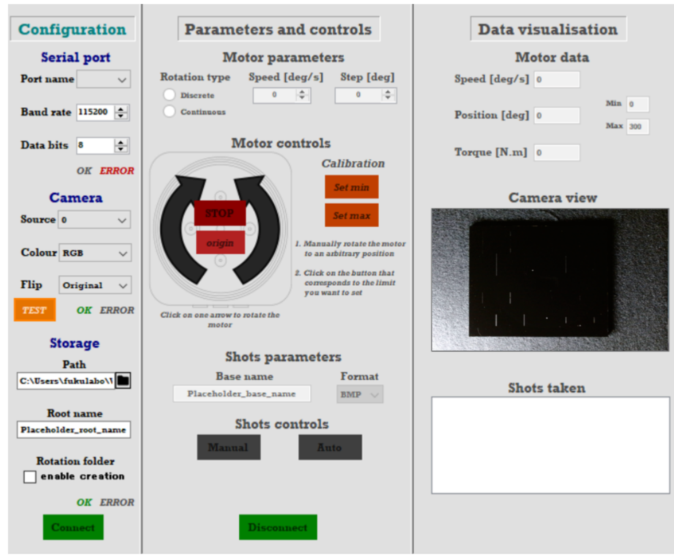


Fig. 4. Interface of our image acquisition software

- **Parameters and controls**: this column contains the components that enable the user to parameterize the motor, select how the captures are to be taken (automatically or manually) and control the motor manually (if desired).
- **Data visualisation**: data coming from the system (motor and camera) are shown within this space, so the user can monitor the acquisition process.

This program allowed us to capture a lot of pictures with ease and speed. Indeed, within the **Parameters and controls** division, it is possible to specify the angular spacing appropriate between two successive captures and to define the movement of the camera as "automatic". However, as stated previously, the current prototype necessitates the intensity of the light to be set manually. Despite its resulting semi-automatic nature, the device still fulfils the incommodious task of rotating the light on its own, only leaving a simple punctual intensity tuning to the operator, which makes for a prompt acquisition process.

IV. Experiments

In this section, we will specify the content of the datasets that were gathered thanks to our system and how they were used for experimental purposes. The objective of the tests was to assess the performances of CNN models when the parameters of physical light sources varied during the acquisition of the population of their training/validation sets. From here onward, since each couple of training and validation collections contain similar data, to facilitate the reading, they will be referred to as a single entity: "training datasets".

A. Experimental Conditions

The variables of our experiments were as follows:

- **Independent variables**: The effects of different illumination settings, such as their brightness and their orientation, were observed.

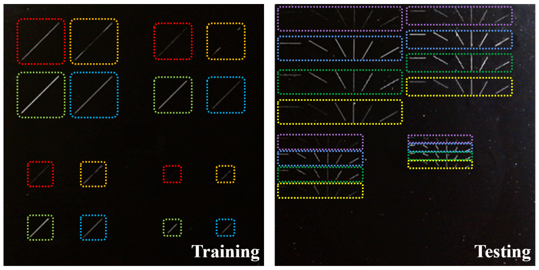


Fig. 5. Samples of training and testing pieces with their lines distributions

- **Dependent variables**: The detection abilities of the different models were measured thanks to the number of True Positives (TP), False Negatives (FN) and False Positives (FP) they output when handling testing data. We also evaluated them through their precision and recall, which are metrics derived from TP, FN and FP.
- **Control variables**: Throughout our experiments, the sizes of the training and validation datasets have been kept at respectively 40 and 5 images. Every model has been trained for 3000 iterations in Darknet, a neural network framework used to train YOLO [16] models. Furthermore, every model has been tested on the same set of 24 pictures to ensure a consistent and reliable baseline to confront the results.

B. *Dataset Compositions*

We formed 5 datasets based on photographs of six different plastic parts created by cutting sheets of acrylic resin into 100 by 100mm squares. Lines have been etched onto the surfaces of these pieces with a laser engraver, so that they mimic scratches that can usually be found on items on a production line, for example. Every piece comes with its own pattern of lines, carved with determined sets of laser parameters (power, speed and repetition), repeated in four different areas but with decreasing lengths. The lines within the training pieces were 20mm, 15mm, 10mm and 5mm long, while that of the testing pieces were 7mm, 5.25mm, 3.50mm and 1.75mm long. The aim was to obtain synthetic flaws that had different aspects. A visual explanation of the line distributions can be found in Fig. 5, in which features highlighted with a common color were traced with the same laser configurations.

The pictures have all been collected by our semi-automatic visual inspection system. In order to minimize the risks of getting biased datasets, the pictures have been taken at night time when no external source could interfere with the light exposure of the pieces intended by the system.

1) Training data's characteristics: Four pieces comprising 16 scratches each, listed in Fig. 6, were used to take 784 photographs. This amount of data is due to the number of combinations possible with:

- 4 pieces whose scratches are tilted: 0°, 45°, 90°, 135°
- 7 light intensities: 0.45W, 0.9W, 1.35W, 1.8W, 2.25W, 2.7W, 3.15W
- 28 light positions: 0°, 10°, 30°, ..., 250°, 260°, 270°

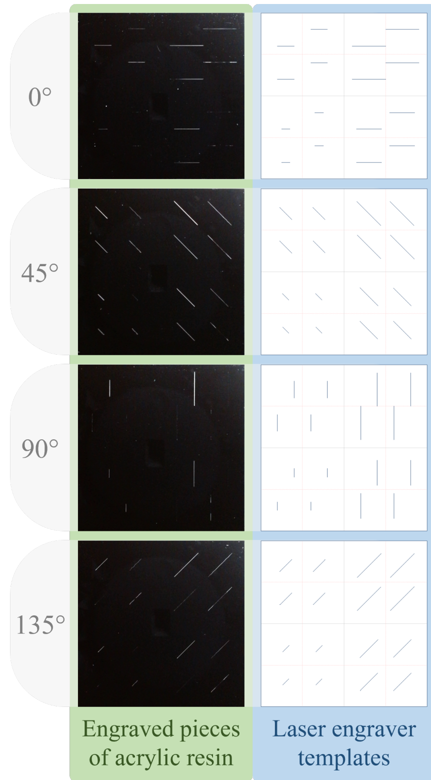


Fig. 6. Base pieces used for training, with the tilt value of their scratches

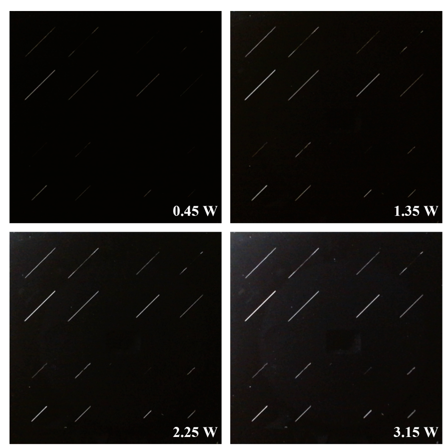


Fig. 7. Training pieces with different light intensities

Fig. 7 presents one of the plastic parts under different illuminations, while Fig. 8 shows the same sample brightened from distinct angles.

Before being incorporated into any training datasets, these data had to be annotated. Since some of them had scratches difficult to see clearly, especially those taken in low brightness, it was decided to create an algorithm that created copies of the images and turned their pixel black if their respective luminance values were below a given threshold, and white otherwise. This way, we were able to use the black and white pictures to consistently classify what scratch should be

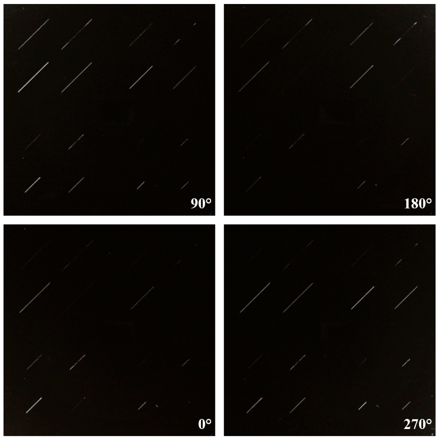


Fig. 8. Training pieces with different light orientations

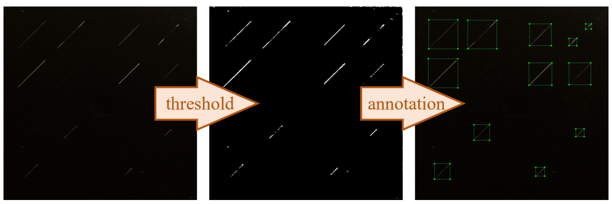


Fig. 9. Scratch annotation based on their luminance value

annotated (white pixels) or which one should be considered "invisible" (black pixels), as shown in Fig. 9. In our case, a value of 30 for the luminance was chosen, as it was below this point a scratch became hardly perceptible to the naked eye.

2) *Test data's characteristics:* Two pieces comprising 96 scratches each, listed in Fig. 10, were used to take 24 photographs. This amount of data is due to the number of combinations possible with:

- 2 pieces: test_A, test_B
- 4 light intensities (in watts): 0.45, 1.322, 2.25, 3.122
- 3 light positions (in degrees): 0, 100, 200

3) *Overview of the datasets:* Four training datasets were assembled, whose structures are summarised in Table I. The formations of those batches aimed at getting models trained on images captured in specific light conditions, so as to try to perceive which setting may be the most efficient, when opposing their results. In addition, 1 testing dataset of 24 distinct pictures was gathered to assess the performances of the models. During the computation of the metrics, to emphasise the hypothetical discrepancies between the models (especially the brightness represented in their training images), it has been considered as 7 sub-datasets: 4 organised by light intensity, outlined in Table II, and 3 organised by light orientation, depicted in Table III. The performances of the models are confronted and reviewed in the next section.

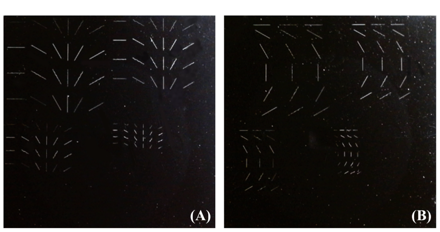


Fig. 10. Testing pieces

TABLE I. Overview of the Training Datasets' Contents

Dataset name	Light intensities (in watts)	Light orientation (in degrees)	Size
TRAIN_intLow_oriQuad	0.45 / 0.9 / 1.35	0 / 90 / 180 / 270	Train: 40 Valid: 5
TRAIN_intHigh_oriQuad	2.25 / 2.7 / 3.15	0 / 90 / 180 / 270	Train: 40 Valid: 5
TRAIN_intMix_oriQuad	0.45 / 0.9 / 1.35 1.8 2.25 / 2.7 / 3.15	0 / 90 / 180 / 270	Train: 40 Valid: 5
TRAIN_intMix_oriMix	0.45 / 0.9 / 1.35 1.8 2.25 / 2.7 / 3.15	0... *(increments of 10°)* ..270	Train: 40 Valid: 5

TABLE II. Overview of the Testing Sub-Datasets, Organised by Light Intensity

Dataset name	Light intensities (in watts)	Light orientation (in degrees)	Size
TEST_intVeryLow	0.45	0 / 100 / 200	6
TEST_intLow	∼1.322	0 / 100 / 200	6
TEST_intHigh	2.25	0 / 100 / 200	6
TEST_intVeryHigh	∼3.122	0 / 100 / 200	6

C. CNN Trainings

Finally, once the data was collected, annotated and put together, they were used to train models. For our project, the neural network framework named Darknet, and most especially its algorithm/network architecture YOLOv4 [16], has been chosen to assist in the generation of models. This end-to-end learning algorithm is well-known in the field of image processing, its speed and accuracy held it as a reference among CNNs. Every one of the 4 datasets was fed into the algorithm to output its corresponding model, so experiments could be conducted to analyze their respective abilities to find defects within pictures. The training phases were not started from scratch, **yolov4.conv.137** pre-trained weights were chosen as a base. Since the evolution of the average training loss curves indicated that the models had converged by 3000 iterations, a training session stopped when this count was reached.

TABLE III. Overview of the Testing Sub-Datasets, Organised by Light Orientation

Dataset name	Light intensities (in watts)	Light orientation (in degrees)	Size
TEST_ori0	0.45 / ∼1.322 / 2.25 / ∼3.122	0	8
TEST_ori100	0.45 / ∼1.322 / 2.25 / ∼3.122	100	8
TEST_ori200	0.45 / ∼1.322 / 2.25 / ∼3.122	200	8

TABLE IV. METRICS RESULTING FROM THE TESTS CONDUCTED ON EACH MODEL, ON THE ENTIRE TESTING DATASET (24 PICTURES)

Model	TP	FN	FP	Precision	Recall
TRAIN_intLow_oriQuad	597	1707	2	0,997	0,259
TRAIN_intMix_oriQuad	447	1857	9	0,98	0,194
TRAIN_intHigh_oriQuad	502	1802	6	0,988	0,218
TRAIN_intMix_oriMix	644	1660	3	0,995	0,28

V. INTERPRETATION OF THE RESULTS

As a first step to confront the ability of the models to spot irregularities, we had trained *TRAIN_intLow_oriQuad*, *TRAIN_intMix_oriQuad* and *TRAIN_intHigh_oriQuad* so tests could be conducted with the brightness as an independent variable, as their pictures were taken with similar light orientations. Based on the 24 photographs of test plastic pieces, the total number of True Positives (defects found), False Negatives (defects unexposed) and False Positives (regular areas inferred as abnormal) were counted. Then, the precisions and recalls were derived from these TPs, FNs and FPs.

The precision, whose formula is shown in (1), describes the proportion of "positive identifications" that were actually correct.

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

The recall, whose formula is shown in (2), describes the proportion of correct "positive identifications" that were successfully identified.

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

The values of the various metrics are transcribed in the Table IV. If the header of this table is coloured green, it translates a positive metric for classification of a model, while a red-coloured header is for a negative one. When a metric's value is written in green, it means that this value is the best of those recorded; if it is red, it is the worst (it takes into account the positive/negative nature of the metric).

When analysing the measures, it has been noticed that the outcomes of the models were relatively close to one another, nonetheless, *TRAIN_intLow_oriQuad* was the most efficient and *TRAIN_intMix_oriQuad* was the less effective. In order to examine the impact of light orientation, it was attempted to increase the diversity of the less reliable dataset: instead of only including images of pieces that were enlightened from a source positioned at either 0°, 90°, 180° or 270°, we decided to pick from 28 angles. In this way, *TRAIN_intMix_oriMix* has been constituted, as *TRAIN_intMix_oriQuad* was the most lacking. The TPs, FNs, FPs, precision and recall obtained are also reported in Table IV. From this new array of values, it has been noted that *TRAIN_intMix_oriMix* not only outperformed *TRAIN_intMix_oriQuad*, it also seemed to slightly surpass the results of the other models.

Though there were differences between the models, they were rather limited and based on a single set of data, which did not reflect their effectiveness at handling data according to their attributes. It was conjectured that it would be pertinent to underline the capabilities of the models at processing images taken in specific light conditions. Thus, the former test dataset was split into several subgroups consisting of images sorted either by intensity (Table II) or by orientation (Table III). For these new comparisons, it was decided to restrict the metrics to the average of True Positives: within the scope of our experiments, the significance of this measure makes it sufficient to gain a decent appreciation of the quality of a neural network's training.

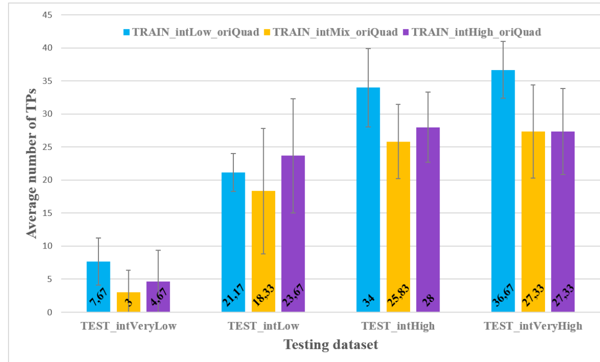


Fig. 11. Comparisons of the average number of TPs on the test subsets arranged by brigthness, grouped by light intensities within their training data

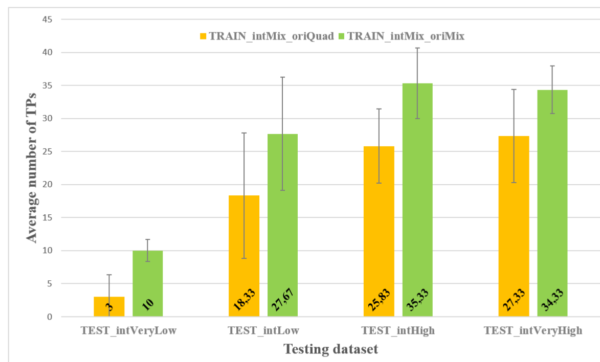


Fig. 12. Comparisons of the average numbers of TPs on the test subsets arranged by brigthness, grouped by light orientations within their training data

The new arrangements of data corroborated the previous experience. Indeed, for the sets containing photographs classified by their luminosity, *TRAIN_intLow_oriQuad* was more efficient than *TRAIN_intHigh_oriQuad* and even more than *TRAIN_intMix_oriQuad*, as shown in Fig. 11. *TRAIN_intMix_oriMix* obtained more satisfactory results than its counterpart trained with four different light orientations, yet again, as visible in Fig. 12. The reason why the models struggled to detect scratches in the case of very low light intensities is that even though most of them were barely discernible, they were annotated as ground truths within the test data, which led to a lot of False Negatives (and, conversely, hardly any True Positives) during the tests that included them. The purpose of these data was to challenge the models on confounding specimens.

Furthermore, similar observations were made for the sets containing photographs classified by the angular positions of

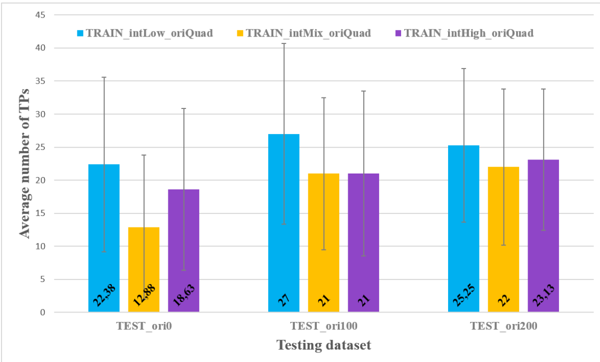


Fig. 13. Comparisons of the average number of TPs on the test subsets arranged by orientation, grouped by light intensities within their training data

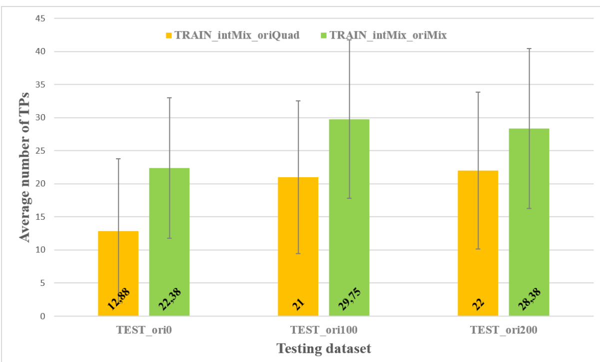


Fig. 14. Comparisons of the average numbers of TPs on the test subsets arranged by orientation, grouped by light orientations within their training data

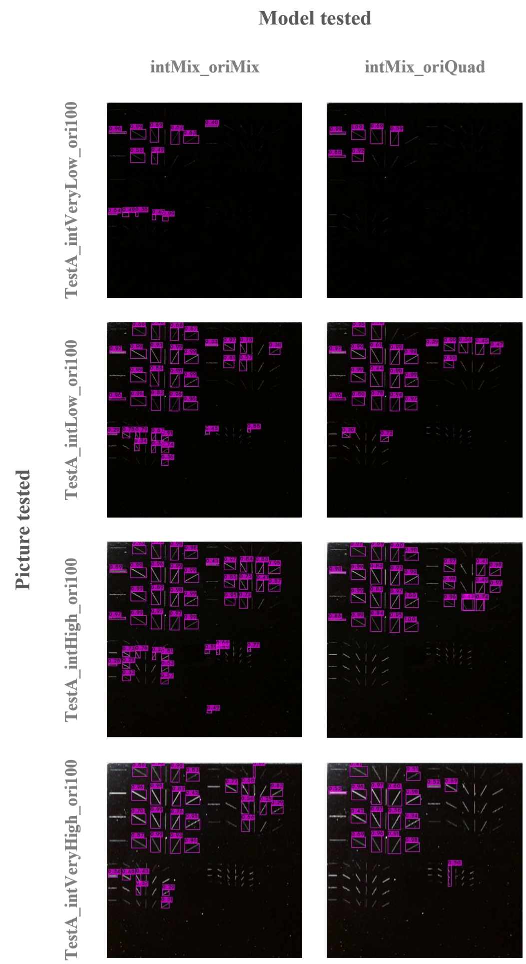


Fig. 15. A few inferences produced by *intMix_oriQuad* and *intMix_oriMix*

their light source, as shown in Fig. 13 and Fig. 14.

A few examples of images with their corresponding inferences, output by *TRAIN_intMix_oriQuad* and *TRAIN_intMix_oriMix*, are displayed in Fig. 15.

VI. Conclusion

Our study involved the conception of an automatic system that facilitates the acquisition of pictures intended for training/testing datasets, while adding physical data augmentation in the form of variations of light brightness and orientation.

This device was used to gather our own images, which were assembled into datasets that were fed to CNNs. These neural networks generated multiple models, each trained on photographs taken under specific lighting conditions. These models were then opposed to one another to observe if the parameters of their illumination had some effects on their ability to handle new data.

It has been noted that both the independent variables (light intensity/light position) that were experimented on influenced the outcome of the training processes. For instance, a model based on data whose illumination varied, but was limited to four positions, obtained worse results than another model based on data whose illumination also fluctuated, but had a broader array of angular positions.

Since the pieces involved in this research have been created to suit a specific scenario: scratches detection on plastic parts, our “hardware data augmentation” technique might only be relevant to this test case. For that reason, future work reiterating the experiments, with larger and more intricate datasets, is expected to confirm (or disprove) the conclusions drawn from the current study. Additionally, the implementation of an automatic light intensity controller, along with the incorporation of YOLO into our software, are objectives envisaged to make our system a fully independent computer-assisted visual inspector.

References

[1] T.L. Johnson, ”How and why we need to capture tacit knowledge in manufacturing: Case studies of visual inspection”, *Applied Ergonomics* Vol.74, pp.1–9, 2019.

[2] V. Hoskere, ”Vision-based Structural Inspection using Multiscale Deep Convolutional Neural Networks”, *arXiv*, 2018.

[3] S.Y. Wang, ”A computer vision based machine learning approach for fatigue crack initiation sites recognition”, *Computational Materials Science* Vol.171, 2020.

[4] I. Kuric, ”Visual Product Inspection Based on Deep Learning Methods”, *Advanced Manufacturing Processes*, pp.148–156, 2020.

[5] A.R. Ajiboye, ”Evaluating the effect of dataset size on predictive model using supervised learning technique”, *International Journal of Software Engineering & Computer Sciences (IJSECS)*, Vol.1, pp.75–84, 2015.

[6] C. Shorten, ”A survey on Image Data Augmentation for Deep Learning”, *Journal of Big Data*, 6:60, 2019.

[7] R. Gontijo-Lopes, ”Tradeoffs in Data Augmentation: An Empirical Study”, *International Conference on Learning Representations*, 2021.

[8] L. Taylor, ”Improving Deep Learning Using Generic Data Augmentation”, *arXiv*, 2017.

[9] I. Kandel, ”Brightness as an Augmentation Technique for Image Classification”, *Emerging Science Journal*, Vol.6 - No.4, pp.881–892, 2022.

[10] D. Crispell, ”Dataset Augmentation for Pose and Lighting Invariant Face Recognition”, *arXiv*, 2017.

[11] D. Takaki, ”Objective scratch and dig measurements”, *Savvy Optics Corp.*, 2012.

[12] G. Fu, ”A deep-learning-based approach for fast and robust steel surface defects classification”, *Optics and Lasers in Engineering*, Volume.121, pp.397–405, 2019.

[13] L. Song, ”Weak Micro-Scratch Detection Based on Deep Convolutional Neural Network”, *IEEE Access*, vol.7, pp.27547–27554, 2019.

[14] L. Yang, ”Steel Plate Surface Defect Detection Based on Dataset Enhancement and Lightweight Convolution Neural Network”, *Machines (MDPI)* 10, 523, 2022.

[15] C.J. Barr, ”Influence of surface texturing on scratch/mar visibility for polymeric materials: a review”, *Journal of Materials Science*, pp.1221–1234, 2017.

[16] J. Redmon, ”You Only Look Once: Unified, Real-Time Object Detection”, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.779–788, 2016.

Multi-Task Multi-User Offloading in Mobile Edge Computing

Nouhaila Moussammi[1], Mohamed El Ghmary[2], Abdellah Idrissi[3]
Department of Computer Sciences, Faculty of Sciences, Mohammed V University in Rabat, Morocco[1,3]
Department of Computer Sciences, FSDM Sidi Mohamed Ben Abdellah University, Fez, Morocco[2]

***Abstract*—Mobile Edge Computing (MEC) is a new method to overcome the resource limitations of mobile devices by enabling Computation Offloading (CO) with low latency. This paper proposes a multi-user multi-task effective system to offload computations for MEC that guarantees in terms of energy, latency for MEC. To begin, radio and computation resources are integrated to ensure the efficient utilization of shared resources when there are multiple users. The energy consumed is positively correlated with the power of transmission and the local CPU frequency. The values can be adjusted to accommodate multi-tasking in order to minimize the amount of energy consumed. The current methods for offloading aren't appropriate when multiple tasks and multiple users have high computing density. Additionally, this paper proposes a multi-user system that includes multiple tasks and high-density computing that is efficient. Simulations have confirmed the Multi-User Multi-Task Offloading Algorithm (MUMTOD). The results in terms of execution time and energy consumption are extremely positive. This improves the effectiveness of offloading as well as reducing energy consumption.**

Keywords*—*Time execution; energy consumption; computation offloading; mobile edge computing

I. Introduction

Mobile devices such as tablets and smart phones are becoming more and more popular. This has led to increasing the number of mobile apps, such as augmented Reality, natural language processing and interactive online gaming [1]. These kinds of mobile apps are typically resource-hungry and latency-sensitive, which means they require greater computing capacity and use more energy, but they have serious time constraints. Mobile devices are constrained in size and are limited in resources, such as CPU-cycle frequency and energy consumption, memory. Computing offloading is a viable solution. Mobile users are now able to perform their computing tasks to devices with the right computing resources or a computing server by using the computation offloading.

A new paradigm of computing called MEC was suggested to solve this issue. It is created to relieve mobile devices of massive computing workloads [2]. MEC offers CC capabilities at the edge of the cellular network, near mobile devices. MEC permits mobile apps to run directly on the device or on the MEC servers [3]. The MEC model offers high bandwidth, low latency and high computing speed when compared to traditional MCC. This is because of the smaller distance between mobile devices and the edge servers [4]. Recent MEC paradigm developments [5] shift computation-intensive tasks away from mobile devices and toward nearby MEC servers. Among these advancements is single-user computation. The majority of these technologies have a common objective which is to reduce energy use, allocate radios and increase computational resources, decrease costs and/or meet the delays required by mobile IoT networks. MEC is a rapidly developing computing model, extends cloud and the services, it provides to the edges of the network for applications that are resource-intensive and require the highest level of performance, MEC network computation offloading is a viable method to allow mobile apps that are resource-intensive. The SMD is a limited source of energy to execute tasks and this is a major issue. Additionally when offloading is utilized specific parts of applications that require computational power are divided into multiple, mutually exclusive offloadable tasks [6]. We believe that mobile users from multiple locations are able to offload computation tasks that they have duplicated to network edge servers and then share the results between them. Edge computing is the process of transferring processing and storage toward the edges of connected devices. Edge computing isn't located in devices.

Edge computing relies on the offloading of computation. It is first utilized in cloud computing and later in edge computing. To transfer computation tasks to servers devices may use computing offloading. In certain situations, all computation tasks are not able to transfer to servers because of limitations on network connectivity restrictions on network connectivity [7]. It is crucial to swiftly determine how many tasks need to be transferred to servers and which ones should remain local. Only when computation tasks are properly offloaded can achieve the QoS and enhance the QoE.

In this paper, we will discuss the allocation of resources, computation offloading to enhance the efficiency of multi-user multi-task offloading computation for MEC. This is the reason behind our research: The most significant factors that impact the effectiveness of multi-user MEC systems are the computational resources on the edge, wireless channel radio resources, and the offloading of computation of mobile device user's tasks. Then, it is essential to develop an approach to these problems. MEC system lets users transfer their application data to the MEC server through the wireless channel. This paper describes an integrated model for resource allocation. The model is based on multi-user, multi-task computing environment that permits MEC within mobile IoT networks. The goal is to decrease the consumption of energy within the computational latency constraints.

This paper is focused on the following aspects: A model that incorporates resource allocation can be described as an optimization issue to reduce the energy consumed under the constraints of computation latency in multi-user, multiple-task MEC systems that are used in mobile IoT networks.

The MUMTOD algorithm is developed to take the most efficient offloading choice for the tasks of computation of every mobile device user in the MEC system. The remainder of the paper is based on these steps: Section II discusses related research regarding computation offloading. Our system model multi-user multitask offloading of computations and problem formulation in section III. Our proposed solution in Section IV. Section V includes simulations to demonstrate our model for computation offloading. The paper concludes in Section VI.

II. Related Work

Task offloading is a crucial method to overcome the limitations of storage for edge computing as well as computing power within the IoT network. Edge devices can outsource some or all of its computing functions to an edge computing server. This can increase the speed of processing, conserve energy, and decrease the time to respond. The research continues to find the most efficient optimization strategy for various situations. MEC is a crucial element of devices. The MEC offers a wide range of storage and computation capabilities to mobile devices. The majority of these research address the problem of offloading computations to a single user such as [8]. Others address offloading of computation in multi-user environments[9].

The researchers [10] looked at three different types of offloading decisions including partial offloading, complete offloading, as well as local execution. The goal of this study is to optimize computational resources and decrease the amount of time needed to complete an independent task within the MEC system. A dynamic offline strategy for single-user computation offloading has been suggested using both deterministic and random methods[11]. This method takes into account offloading computation and resource scheduling to establish the most efficient schedule offloading policy. This method aims to decrease energy consumption while also satisfying the requirements for delay. A scenario where users receive the list of computation tasks to be offloaded. Each task must be handled by the MEC server within a fixed time frame[12]. The proposed optimization problems aim at the reduction of energy usage, the total processing delays as well as the insatiable processing workload.

The authors of [13] investigated multi-user edge computing scenarios that are based on orthogonal frequency division multiple access (OFDMA) and time division multiple access (TDMA). The limitation of computing delay implies that the optimal resource allocation algorithm should be convex. This resolves the issue by minimizing the weighting and power consumption of mobile devices. To simplify the process, a less efficient resource allocation algorithm is recommended for cloud systems with small capacity. The issue of offloading in multi-user situations was investigated by [14]. The authors proposed a simple search algorithm that consists of two segments to determine an optimal conditional time. A method of coordinate descent was designed to improve mode selection.

While the work [15] focused on one server system that fixed local computational resources, as well as the transmission capabilities, this is the first time investigating the real MEC system where some users are subscribers with the highest priority for services that offload computation. This research is an important expansion of the research [15]. Task offloading serves two primary objectives: to decrease the time required to execute applications that run on user equipment and to save energy . Efficiency of task offloading is a multi-fold issue. MEC servers might not be as efficient as cloud servers, therefore it is crucial to assign the tasks. Transferring offloading of user equipment to the MEC servers should be carried out in this way the MEC servers make use of their resources in a responsible manner.

The MEC Servers hosting on different Base station might not share the identical set of drivers Services, as well as the computational load on MEC servers could also be affected Variable as the passage of time. The optimal allocation of resources to relieve load Tasks, fair and steady load distribution among the MEC servers. The seamless coordination between edge and cloud is just one of the aspects that make up seamless integration. The issue is the difficulty in using of the MEC of the MEC services efficiently [16]. Another aspect The size of the defined remains a problem which is the Tasks that need to be delegated. The QoE requirements will determine the tasks that need to be delegated. The process of running applications can be difficult. It is also possible that the tasks be a matter of distinct priority, the impact of computing overhead, progress and other dependencies[17]. The scheduling, selection, location and the management of work are all a part of the process. A myriad of issues could arise from the task of offloading. This is why the key to optimal task delegation is to select the best correct task to offload to the correct MEC server at the correct time to allow for meeting. The most efficient performance of an app running when using the MEC utilizes resources efficiently. It is clearly an optimization that is multi-objective problem is considered to be NP hard [18]. In the context of this debate, task offloading could be defined Uploading the entire module to an application that contains the Calculation, data required and other libraries that are dependent on it can be delivered to remote locations server, and then receiving the result of the computation from the remote server. It is crucial to talk about the process of offloading, scheduling is crucial to offload tasks, which is a method for executing a list of tasks for a specific number of computer resources that can be utilized to achieve a goal [19].

In [20] an energy-saving, dynamic scheduling strategy and offloading method developed to cut down on energy use and speed up the completion of applications. The problem could be turned into a problem of energy efficiency by reducing task dependence and the time limit for completion. It is broken down into three sub-tasks: controlling the frequency of the clock, transmission power allocation, and calculation offload selection. Wang and coworkers addressed the MEC offloading problem. Wang et al. suggested a new framework for offloading based on deep reinforcement that can be used to automatically identify the most efficient way to load for different scenarios based on the specifics of each offloading job. This reduces the total delay in service [21], and the others concentrated on management of resources in a multi-user MEC system. The research did not offer any insights into the effect of channel-specific information on the devices' consumption of energy. To address these issues we will focus on task offloading process within multi-user MEC systems and provide an energy efficient stochastic framework. The optimization issue is solved with an effective algorithm that doesn't require prior knowledge of the time of task completion or any statistics about the channel state.

The authors of [22] looked at the possibility that the wireless energy transfer was the main driver of MCC , and developed an energy-efficient framework to increase the probability of computing while balancing the limitations of delay and energy. A majority of these researches relied on the assumption or prediction of the task's timing or channel status. The load-shedding of traffic from IoT devices and the quality of wireless channels are complex and are difficult to predict.

The work [23] focused on a partial computation offloading method, it optimizes computational and communication resources employing the dynamic voltage scaling, an algorithm design is included in the description of this method. The main goal of the method is to reduce energy consumption and delay in execution. This can be accomplished by taking into account computational time, transmission power and the offloading ratio.

An Markov chain theory-based search algorithm that is one-dimensional is described in [24]. The algorithm is presented in [24]. It determines the best strategy for offloading stochastic computations in a single-user MEC system. The perspective of scheduling flow shops is further developed by the work of [23] that has created an optimization issue for task-offloading scheduling and the allocation of power for transmission in an MEC system, which an energy harvesting technique is employed to cut down on the power consumption and energy consumption for data transmissions to enable offloading computation. It uses a simple algorithm to determine the best offloading option for every time slot [24].

III. System Model and Problem Formulation

This research aims to decrease the use of energy in multi-user edge computing devices that are multi-tasking by deploying the concept of offloading computation. This section is dedicated to a system model adopted during our research.

This paper investigates the concept of a multi-user as well as a multi-tasking system as illustrated in Fig. 1.

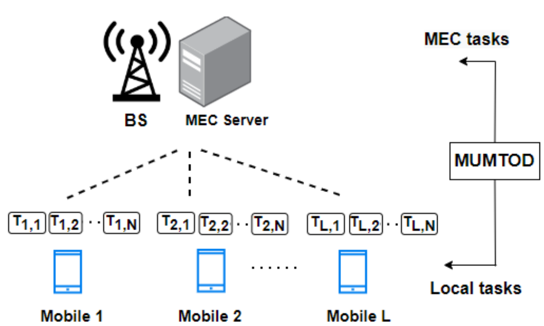


Fig. 1. System model and problem formulation

Take $U = \{1, 2, .., L\}$ as a set of mobile devices. Each mobile is assigned $T = \{1, 2, .., N\}$ which is a set of tasks that must be accomplished, which are linked with a base station. Let's start with the introduction of communication in a MEC, Our environment includes L users. Each user has to complete computation tasks. Let $x_{i,j} \in \{0, 1\}$ represents the number of integers that are used to compute offloading decisions for the task j of the user of a mobile device i. Particularly if $(x_{i,j} = 0)$, the task j of the user of a mobile device i will be performed locally. in contrast, $(x_{i,j} = 1)$ the task j of the user of mobile devices i is transferred to the MEC server through the wireless channel.

This is why we have chosen $X = \{x_{1,1}, x_{1,2}, .., x_{L,N}\}$, as the offloading profile to handle the computing tasks of all users of mobile devices. Frequency Division Multiple Access is (FDMA) is a multi-access technique with higher performance. Every user will be provided with just a portion of the bandwidth that the system provides. This is , the data uplink rate of every mobile user i is as follows:

$$r_i(p_i) = Blog(1 + p_i\frac{h^2}{\omega_0}) \tag{1}$$

which B is the bandwidth of the channel. p_i represents the mobile's i transmission power, h is the channel's gain, and ω the channel's noise power. If all mobile IoT devices decide to delegate their computing tasks to the wireless access channel at once in a computation offloading time, there is the limit on data rate R.

$$\sum_{i=1}^{L}\sum_{j=1}^{N} x_{i,j} r_i \leq R \tag{2}$$

Every smartphone user i is equipped with N computing tasks. The tasks can be completed local on the device, or remotely on the MEC server through wireless communication. We further define $u_{i,j}$ as the user of the mobile device i who is requesting the task of computation j to be executed.

Every computation task j is identified by $(C_{i,j}, D_{i,j})$ in which $C_{i,j}$ refers to the number of CPU cycles needed to complete the task of computation j and $D_{i,j}$ is the total size of data that is able to be offloaded. The time required to transfer the result from the MEC server to the mobile device users is not taken into account in this study. This is because the result is usually smaller than input data [25]. All computations are performed locally using a mobile device for local execution. the total time of the task i that is executed locally can be expressed as:

$$T_{i,j}^{loc} = \frac{C_{i,j}}{F_i^{loc}} \tag{3}$$

where F_i^{loc} represents the computing capacity of mobile i. Different mobile devices may have varying computational capabilities. The energy required to perform the task locally is determined as follows:

$$E_{i,j}^{loc} = y_i C_{i,j} \tag{4}$$

y_i is the value of the energy consumed by CPU cycles. It is determined by [24]. In MEC Server: A user of a mobile device i chooses to transmit computation task j to an edge server through the wireless channel. The time required for task execution when offloading is determined by time of transmission which is the amount of time required for users of mobile devices to offload the task of computation and also the amount of time needed for the computation task to run in the MEC servers which is task execution.

Furthermore, the energy usage for the offloading process is only measured by the cost of communication for offloading the task information to the MEC server. The total amount of offloading time which is the sum of the transmission time and execution time and energy used in the transmission are calculated using (1).

$$T_{i,j}^{up+exec} = \frac{D_{i,j}}{r_i} + \frac{C_{i,j}}{F_i^s} \quad (5)$$

$$E_{i,j}^{up} = p_i \frac{D_{i,j}}{r_i} \quad (6)$$

F_i^s represents the computing power, measured in the form of CPU cycles per (s) of the edge server which was assigned to the user i. We suppose that all users of mobile devices are part of the edge computing server's computation resources.

This section presents also an optimization problem to achieve efficacy in computation offloading in a multi-user multi-task MEC system. Every user of a mobile device is accountable for the execution time as well as the energy consumed. In addition, the allocation of available compute and radio resources on the edge server is also managed. The problem of offloading is described as a constrained optimization problem:

$$\text{(P1)} \qquad \min_x \sum_{i=1}^{L} \sum_{j=1}^{N} (E_{i,j}^{loc} + E_{i,j}^{up})$$

$$\text{s.t}$$

$$\text{(C1)} \qquad \sum_{i=1}^{L} \sum_{j=1}^{N} x_{i,j}.r_{i,j} \leq R \qquad \forall i,j$$

$$\text{(C2)} \qquad \sum_{i=1}^{L} \sum_{j=1}^{N} x_{i,j}.F_i^s < F \qquad \forall i,j$$

$$\text{(C3)} \qquad x_{i,j} \in \{0,1\} \qquad \forall i,j$$

$$\text{(C4)} \qquad \sum_{i=1}^{L} \sum_{j=1}^{N} (E_{i,j}^{loc} + E_{i,j}^{up}) < E^{max}$$

The goal of the optimization problem is to reduce the amount of energy consumed by mobile device users by using task offloading. The capacity for data rates is the constraint C1. The top limit of the CPU's frequency is shown by the constraint C2. Constraint C3 indicates that task-offloading decision variables are binary variables. Constraint (C4) means that both the Edge server and mobile device use less energy than the maximum energy consumption E^{max}.

IV. Proposed Solution

This section will introduce our solution (MUMTOD), a multi-user multi-task offloading decision algorithm. It addresses the optimization issue.

This section explains the multi-user multitask offloading decision algorithm. It provides the specifics of the procedure to get the most optimal MEC algorithm for computing offloading. Each computation task is initiated by using $x_{i,j} = 0$ that indicates local execution.

Every mobile device transmits the specifications of every computation task, including $\{C_{i,j}, D_{i,j}, y_i, p_i\}$ along with the computational capability F_i^{loc} to the MEC server. The MEC server then determines the mobile device's uplink data rate using the equation (1). MEC server then determines the most optimal solution for the offloading of computations decision by using the constraint (C2). In the end, MEC server sends the decision to each mobile device , which reducing the energy used in the entire system .The algorithm 1 describes the procedure of offloading multitask multiuser computation the decision.

Algorithm 1 Multi-User Multi-Task Offloading Decision Algorithm

1: **Initialization:** Each mobile device user i initializes the offloading decision for each computation task with $x_{i,j} = 0$
2: **for all** each mobile user i ($i \in U$) **do**
3:   **for all** each computation task j ($j \in T$) **do**
4:     **Transmit** the computational capability of each mobile user F_i^{loc} and the requirements of computation task $\{C_{i,j}, D_{i,j}, y_i, p_i\}$ to the edge server
5:   **end for**
6: **end for**
7: Based on equation (1): **Calculate** the uplink data rate r_i, for every device.
8: **Utilize** Equation (2) to determine the most efficient computation offloading decision value $x_{i,j}$ for all computation tasks. This will reduce energy consumption overall. **Send** the decision value $x_{i,j}$to each device

TABLE I. Simulation Parameters

Parameter	Value
Number of mobile device users	100
The background noise	10^{-9}
Task generation rate (seconds)	10
Computational load (millions of CPU cycles)	200
Number of computation tasks	100
Data entry size (MB)	10

To demonstrate and assess the model, a simulation of two various scenarios was conducted. They are:

Computation Offloading (CO): The CO policy permits all mobile device users to perform their computations locally on the device or via a MEC server. It is based on the model of optimization described in [26].

The Proposed Solution (PS): considers computation, communication, and the impact on the consumption of energy.

V. Simulation Results and Discussion

In this section in this section, we'll first outline the outcomes of the algorithm we propose and verify its effectiveness. Then, we'll demonstrate the results obtained by the parameters of the system.

The algorithm MUMTOD was implemented within Eclipse IDE version 2022- 03 (4.23.0) and written in the Java programming language. One application was identified as [27] for the experiment as shown in Table I.

The simulation parameters are described as follows, we are considering N=100 mobile devices and 100 independent computation tasks which can be run locally or remotely via

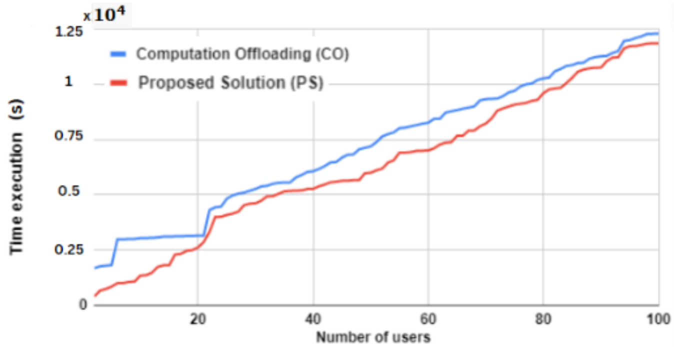


Fig. 2. Execution time for different mobile devices L=100

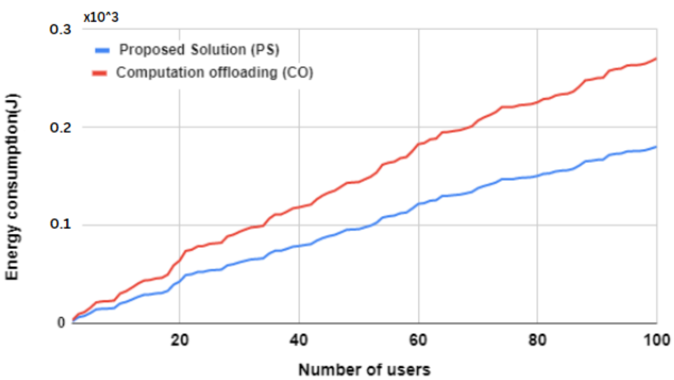


Fig. 3. Energy consumption for different mobile users L=100

the edge server. The following sections provide a summary of the results. The time of simulation was measured in (s) and the scenarios were constructed using a predetermined number 100 tasks.

In Fig. 2, computation tasks are measured in relation to the number of users on the mobile devices. The total time to execute for both scenarios is displayed in the figure in two curves. the execution of our model is nearly as fast or quicker than the time required by the computation offloading scenarios in the case of 24 mobile devices owners. Although as the number of mobile users is growing and our model is more efficient than the scenario that uses computation offloading. This is because the communication channels shared by all users are overloaded. The time required to improve communication speeds due to the growing number of mobile devices has also increased.

Fig. 3 illustrates the amount of energy used to complete the computation tasks in two scenarios. This is then compared to the number of users. It is observed from Fig. 3 that energy consumption significantly increases when the number of users of mobile devices increases the model proposed uses less energy than the alternatives. As the number of mobile devices grows the gap gets bigger. Our model can also help reduce the energy usage.

Offloading scenarios consume more energy than other model. Every mobile device user and information transmitted are competing for the limited resources of communication.

VI. Conclusion

This paper proposes a multi-user multi-task offloading algorithm to support MEC in the mobile user. A problem of optimization was designed to determine near-optimal offloading choices for every mobile user. This is done in order to reduce the energy consumed by mobile users of devices. A thorough process was employed to develop an efficient offloading algorithm that could be used to solve the optimization issue. Additionally,our model was more efficient than other computational offloading scenarios in terms of execution time as well as energy consumption which use different simulation settings and has an improved selection of tasks to offload. Our model is able to support multi-user, multiple-task computation offloading within MEC to mobile IoT networks.

For our future work, a secure computation offloading model that ensures the security of edge computing on mobile devices will be examined. This layer safeguards the data transmitted from cyberattacks.

References

[1] K. Panetta et al., " A comprehensive database for benchmarking imaging systems ", IEEE transactions on pattern analysis and machine intelligence, vol. 42, no 3, p. 509-520, 2018.

[2] L. Jiao, H. Yin, H. Huang, D. Guo, et Y. Lyu, " Computation offloading for multi-user mobile edge computing ", IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), p. 422-429, 2018.

[3] Y. Wen, W. Zhang, et H. Luo, " Energy-optimal mobile application execution: Taming resource-poor mobile devices with cloud clones ", proceedings IEEE Infocom, p. 2716-2720, 2012.

[4] J. Luo, X. Deng, H. Zhang, et H. Qi, " QoE-driven computation offloading for edge computing ", Journal of Systems Architecture, vol. 97, p. 34-39, 2019.

[5] P. Mach et Z. Becvar, " Mobile edge computing: A survey on architecture and computation offloading ", IEEE communications surveys & tutorials, vol. 19, no 3, p. 1628-1656, 2017.

[6] M. El Ghmary, et al. Efficient multi-task offloading with energy and computational resources optimization in a mobile edge computing node. International Journal of Electrical & Computer Engineering (2088-8708), 9(6), 2019.

[7] Y., Hmimz, et al.Bi-objective optimization for multi-task offloading in latency and radio resources constrained mobile edge computing networks. Multimedia Tools and Applications, 80(11), 17129-17166, 2021.

[8] M. El Ghmary, Y. Hmimz, T. Chanyour, A. Ouacha, et M. O. C. Malki, " Multi-task Offloading and Computational Resources Management in a Mobile Edge Computing Environment ", 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), p. 1-7, 2020.

[9] I. A. Elgendy, W. Zhang, Y.-C. Tian, et K. Li, " Resource allocation and computation offloading with data security for mobile edge computing ", Future Generation Computer Systems, vol. 100, p. 531-541, 2019.

[10] M. El Ghmary, et al.Time and resource constrained offloading with multi task in a mobile edge computing node. International Journal of Electrical & Computer Engineering (2088-8708), 10(4), 2020.

[11] Z. Ning, P. Dong, X. Kong, et F. Xia, " A cooperative partial computation offloading scheme for mobile edge computing enabled Internet of Things", IEEE Internet of Things Journal, vol. 6, no 3, p. 4804-4814, 2018

[12] T., Chanyour, et al. Delay-aware and user-adaptive offloading of computation-intensive applications with per-task delay in mobile edge computing networks. International Journal of Advanced Computer Science and Applications, 11(1), 2020.

[13] C. You, K. Huang, H. Chae, et B.-H. Kim, " Energy-efficient resource allocation for mobile-edge computation offloading ", IEEE Transactions on Wireless Communications, vol. 16, no 3, p. 1397-1411, 2016.

[14] S. Bi et Y. J. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading", IEEE Transactions on Wireless Communications, vol. 17, no 6, p. 4177-4190, 2018.

[15] Y., Hmimz,et al. Energy efficient and devices priority aware computation offloading to a mobile edge computing server. In 2019 5th International Conference on Optimization and Applications (ICOA) pp. 1-6. IEEE, 2019.

[16] H. Guo, J. Liu, H. Qin, et H. Zhang, " Collaborative computation offloading for mobile-edge computing over fiber-wireless networks ", in GLOBECOM 2017-2017 IEEE Global Communications Conference, p. 1-,2017.

[17] Y. Mao, C. You, J. Zhang, K. Huang, et K. B. Letaief, " A survey on mobile edge computing: The communication perspective ", IEEE communications surveys & tutorials, vol. 19, no 4, p. 2322-2358, 2017.

[18] K. Wang, X. Wang, X. Liu, et A. Jolfaei, " Task offloading strategy based on reinforcement learning computing in edge computing architecture of internet of vehicles ", IEEE Access, vol. 8, p. 173779-173789, 2020.

[19] P. Mach et Z. Becvar, " Mobile edge computing: A survey on architecture and computation offloading ", IEEE communications surveys & tutorials, vol. 19, no 3, p. 1628-1656, 2017.

[20] S. Guo, B. Xiao, Y. Yang, et Y. Yang, " Energy-efficient dynamic offloading and resource scheduling in mobile cloud computing ", in IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, p. 1-9, 2016.

[21] J. Wang, J. Hu, G. Min, W. Zhan, Q. Ni, et N. Georgalas, " Computation offloading in multi-access edge computing using a deep sequential model based on reinforcement learning ", IEEE Communications Magazine, vol. 57, no 5, p. 64-69, 2019.

[22] C. You, K. Huang, et H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer", IEEE Journal on Selected Areas in Communications, vol. 34, n 5, p. 1757-1771, 2016.

[23] J. Bi, H. Yuan, S. Duanmu, M. Zhou, et A. Abusorrah, " Energy-optimized partial computation offloading in mobile-edge computing with genetic simulated-annealing-based particle swarm optimization ", IEEE Internet of Things Journal, vol. 8, no 5, p. 3774-3785, 2020.

[24] J. Li, H. Gao, T. Lv, et Y. Lu, " Deep reinforcement learning based computation offloading and resource allocation for MEC ", in 2018 IEEE Wireless communications and networking conference (WCNC), p. 1-6, 2018.

[25] X. Chen, " Decentralized computation offloading game for mobile cloud computing ", IEEE Transactions on Parallel and Distributed Systems, vol. 26, no 4, p. 974-983, 2014.

[26] K. Zhang et al., " Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks ", IEEE access, vol. 4, p. 5896-5907, 2016.

[27] L. Huang, X. Feng, L. Zhang, L. Qian, et Y. Wu, " Multi-server multi-user multi-task computation offloading for mobile edge computing networks ", Sensors, vol. 19, no 6, p. 1446, 2019.

Performance Comparison of Multiple Neural Networks for Fault Detection of Sensors Array in Oil Heating Reactor

Mai Mustafa[1], Sawsan Morkos Gharghory[2], Hanan Ahmed Kamal[3]
Electronics and Communication Engineering Department, Cairo University, Giza, Egypt[1,3]
Computers and Systems Department Electronics Research Institute, cairo, egypt[2]

***Abstract*—Fault detection is an important issue for early failure revelation and machine components preserving before the damage. The processes of fault detection, diagnosis and correction especially in oil heating reactor sensors are among the most crucial steps for reliable and proper operation inside the reactor.The fault detection in sensors array of heating reactor is considered as an important tool to guarantee that the controller can take the best possible action to insure the quality of the output. In this paper, fault detection for the temperature sensor in oil heating reactor using different types of faults with different levels is addressed. Multiple approaches based on Neural Network (NN)s such as the classical Fully Connected Neural Network (FCNN), Bidirectional Long Short Term Memory network (BiLSTM) based on Recurrent Neural Network (R.N.N.) and Convolutional Neural Network (CNN) are suggested for this purpose. The suggested networks are trained and tested on real dataset sequences taken from sensors array readings of real heating reactor in Egypt. The performance comparison of the suggested networks is evaluated using different metrics such as "confusion matrix", accuracy, precision, etc. The various NN are simulated, trained and tested in this paper using MATLAB software 2021 and the advanced tool of "DeepNetworkDesigner". The simulation results prove that CNN outperforms the other comparative networks with classification accuracy reached to 100% with different levels and different types of faults.**

Keywords*—Fault detection; sensor array; oil heater reactor; confusion matrix; neural network; recurrent neural network; convolution neural network; bidirectional short term memory

I. Introduction

Sensor faulty readings is considered one of the major source of industries inefficiencies, detection and diagnosis have a strong ability to improve the operational precision and processes stability inside factories and reduce any inefficiencies in the production cycle. Therefore the objective of fault detection and diagnosis is to reduce the production losses, while assuring the safety to both human and devices.

With the high number of sensors used in monitoring industrial operation, it is very time consuming and a large number of labor is needed to follow whether they work properly or not.

Automatic fault detection is one stage of multi-stage processes system to detect-diagnose-correct any fault at sensors array in a complex control systems. Literature review to some techniques was presented to detect the faults such as serial principal component analysis (SPCA) [1], Decision Tree, Random Forest, Nearest Neighbors [2] Support-Vector Machine (SVM) [3], [4], Fuzzy Deep N.N (FDNN) [5], principal component analysis (PCA) [6], independent component analysis (ICA) [7], Serial Principal Component Analysis [8], lossless compression method [9], KNN rules [10], Kalman Filter [11], and hidden Markov models (HMM) [12],

and other methods [13], [14], [15] , [16], [17], , [18], [19] and [20].

In oil heating reactor many sensors are used to monitor and control the processes inside it because any malfunctions in processes inside the heating reactor may lead to the non-quality production and high cost.

Therefore, it is very crucial to detect any sensor anomalies as early as possible to avoid any product defect.

The stages of detection, diagnosis and correction for a faulty sensor reading using N.N-based classification are shown in Fig. 1 and are illustrated as follows:

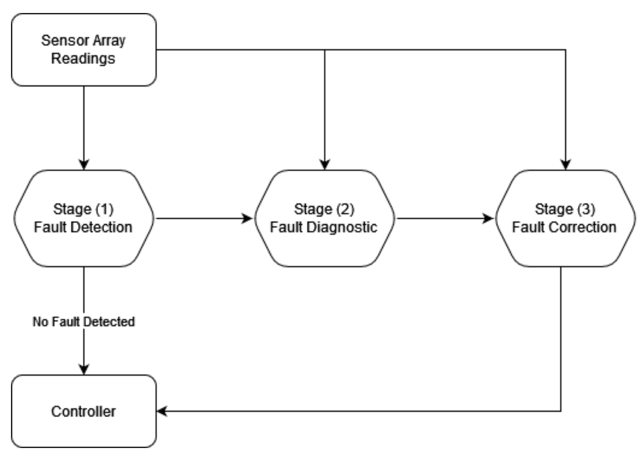


Fig. 1. Sensor array reading possessor

1) Fault detection: in this stage, the sensor array processor will review all sensors readings by using a trained NN based classification to determine if there is a fault in the sensor readings or not and define the faulty sensor if there is a fault.
2) Fault diagnosis: in which the sensor array processor will use the number of the faulty sensor from the previous stage to compare its reading with all other sensors readings to determine the type of the fault using trained classification NN.

3) Fault correction: in this stage, the sensor array processor will utilize the type of fault from the previous stage and the other sensor readings to predict the true reading value of this sensor and send the corrected value to the controller.

This papers presents three N.Ns for fault detection of temperature sensor from sensors array in oil heating reactor and evaluate their performance through comparison using different metrics. The suggested networks are Fully connected neural network (FCNN),Bidirectional long short term memory(BiLSTM) based on recurrent neural network(RNN) and Convolution neural network (CNN). Where, from the simulation results comparison will be used to assess the best performance one of the suggested networks to be recommended in future work in detecting the faulty sensor.The remaining of this paper is organized as follows:

Section II introduces oil heating reactor and description to its physical functions.
Faults types and their mathematical models are displayed in Section III.
Section IV discuss the suggested networks for fault detection, data sets used and how the data is Prepared to be used
Section V is show the construction of the NNs, training parameter and training/test data.
The simulation results of comparing the suggested networks with their performance evaluation using different metrics are displayed in Section VI.
Finally, Section VII concludes the work and discusses the future work.

II. Oil Heating Reactor and its Physical Function Description

This section will discuss the physical construction of the reactor, sensor array, and control system. The physical system consists of heating element for oil before it is separated in later stages.the System also has multiple numbers of sensors - for monitoring the temperature and the flow of natural gas - which are illustrated in the following section. The construction of oil heating reactor and the schematic diagram of its control system are shown in Fig. 2 and Fig. 3.

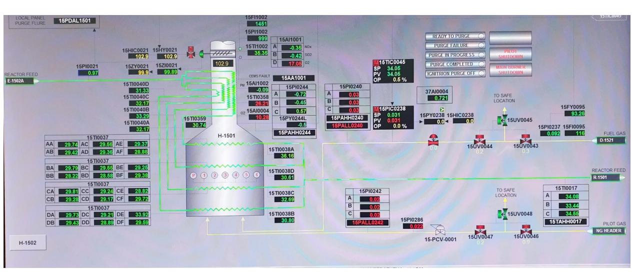


Fig. 2. Oil heating reactor

1) Sensor array: The sensor array consists of 37 sensors in total divided into:

i- 1 oil flow Sensor
ii- 34 Temperature sensors
iii- 1 Gas flow Sensor
iv- 1 Gas Pressure Sensor

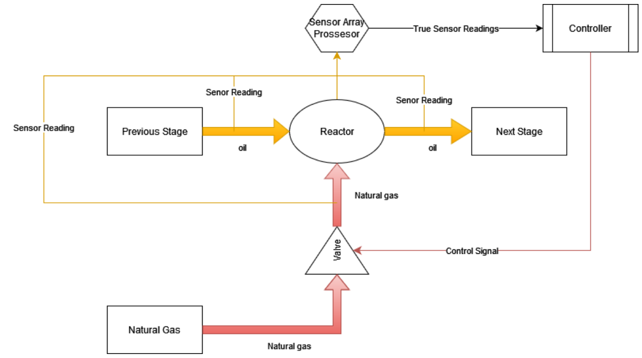


Fig. 3. Control system schematic diagram

All these sensors are installed in various places to monitor the temperature and flow of oil and natural gas through the reactor to help the controller to control the flow of natural gas that will be burned to heat the oil that exits the reactor to be at specific temperature.

2) Heating element: The heating element consists of two stages:-

i- Preheating: where the exhaust of heater is used to preheat the oil.
ii- Main heating:burning natural gas to heat oil passing through some pipes.

III. Fault Types Mathematical Models

This section will discuss the mathematical model of the various types of faults [21], [22], [23], [24].

A. The drift fault

As shown in Fig. 4,

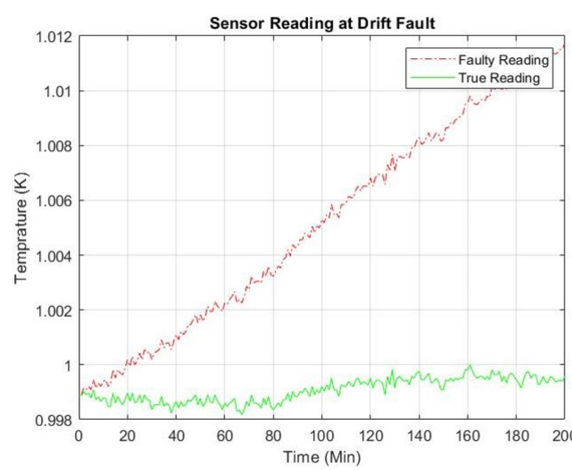


Fig. 4. The drift fault sensor reading

the value of sensor reading will increase in positive feedback loop, so the faulty sensor reading X will be modeled by the following equation:

$$X = X_o + \delta X.t \quad (1)$$

Where $\frac{\delta X}{\bar{X}}$ called fault level, X is the sensor reading, $\bar{X}$ is the average of Sensor readings, t is the time stamp, and X_o is the true reading.

B. The Bias Fault

As shown in Fig. 5,

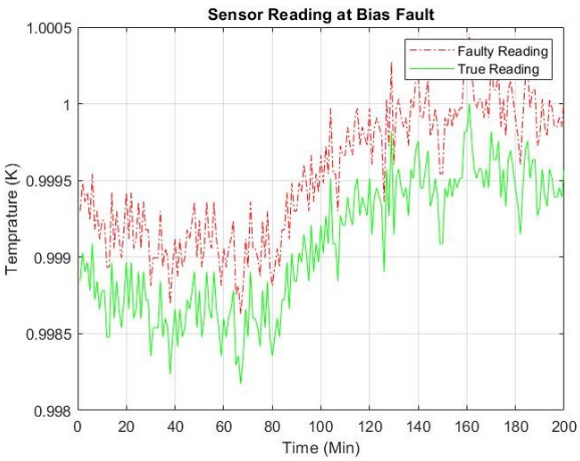


Fig. 5. The bias fault sensor reading

the value of sensor reading will increase by constant value, so the faulty sensor reading will be modeled by the following equation:-

$$X = X_o + \delta X \tag{2}$$

Where $\frac{\delta X}{\bar{X}}$ called fault level, X is the sensor reading, $\bar{X}$ is the average of sensor readings, and X_o is the true reading.

C. The Precision Degradation Fault

As shown in Fig. 6,

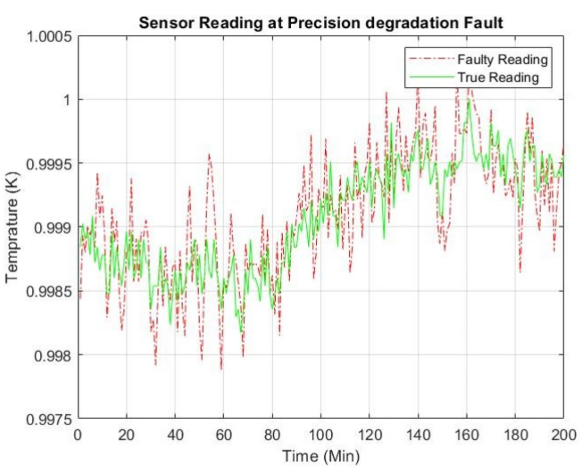


Fig. 6. The precision degradation fault sensor reading

the value of sensor readings will be mixed with white Gaussian noise (WGN), so the faulty sensor reading will be corresponding to the following equation:

$$X = X_o + \delta X.W.G.N \tag{3}$$

Where $\frac{\delta X}{\bar{X}}$ called fault level, X is the sensor reading, $\bar{X}$ is the average of sensor readings, WGN is white Gaussian noise with value varying form (0 to 1), and X_o is the true reading.

D. The Spike Fault

As shown in Fig. 7,

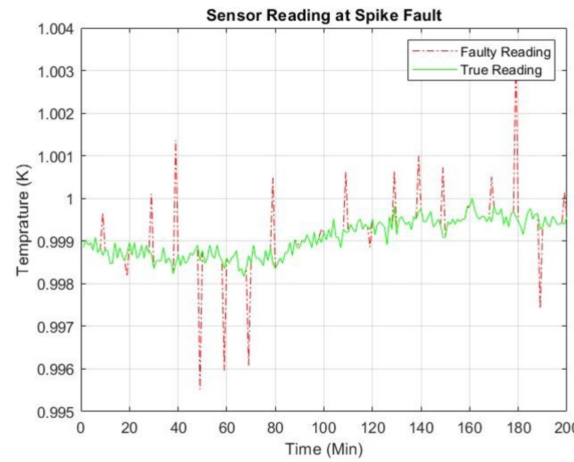


Fig. 7. The spike fault fault sensor reading

the value of sensor readings will be mixed with spike noise [22], [23], [24], so the faulty sensor reading will be modeled according to the following equation:

$$X = X_o + \delta X.W.G.N\delta(t - t_0) \tag{4}$$

Where $\frac{\delta X}{\bar{X}}$ called fault level, X is the sensor reading, $\bar{X}$ is the average of Sensor readings, WGN is white Gaussian noise with value varying form (0 to 1), t is time Stamp X_o is the true reading, and t_0 is the time of the spike.

E. The Stuck Fault

As shown in Fig. 8

the value of sensor reading will be stuck in a certain value, so the faulty sensor reading will be calculated according to the following equation:

$$X = \delta X \tag{5}$$

Where $\frac{\delta X}{\bar{X}}$ called fault level and X is the sensor reading, and $\bar{X}$ is the average of sensor readings.

IV. The Proposed Networks for Classification-Based Fault Detection

In this paper, fault detection stage is implemented by NNs based-classification. After training the networks on series dataset of sensors readings,the network classifies the readings into faulty or non-faulty. The suggested NNs are described in the following section.

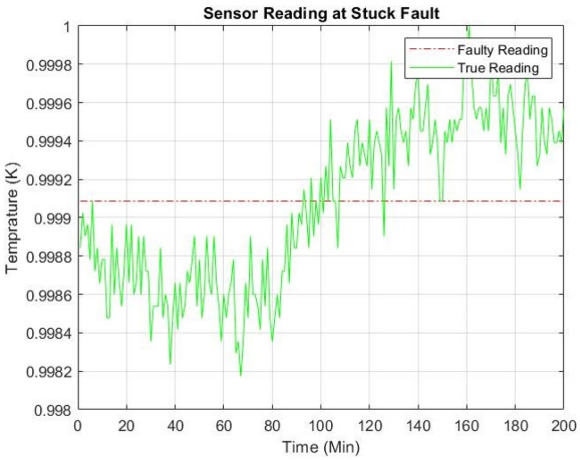


Fig. 8. The stuck sensor reading

A. Description of Suggested Networks

This part is dedicated to show the construction of the used NNs in this Paper.

1) FCNN: A FCNN [25] consists of a series of fully connected layers as demonstrated in Fig. 9.

A fully connected (FC) layer is a function from multiple inputs to one output.

Each output dimension depends on each input.

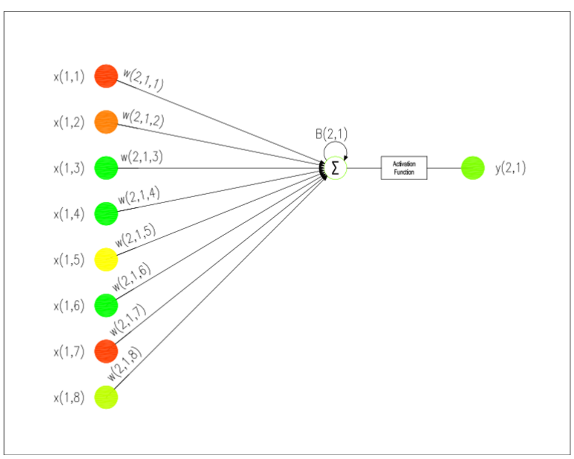


Fig. 9. FCNN construction

2) BiLSTM: BiLSTM network has the ability to learn bidirectional long-term dependencies between time steps of time series data which is helpful for a network to learn at each time step from the total time series data.

BiLSTM handles input sequences in two directions using two sub-layers.

It includes two recurrent network layers, in which the first one processes the sequence of inputs in forwards direction while the second processes the inputs sequence in backwards. Both layers are connected to the same output layer. Thus, BiLSTM network captures the total information about preceding and future sequence of data points [26], [27], [28], [29].

BiLSTM network utilizes the past states to determine the output of the current states as shown in Fig. 10.

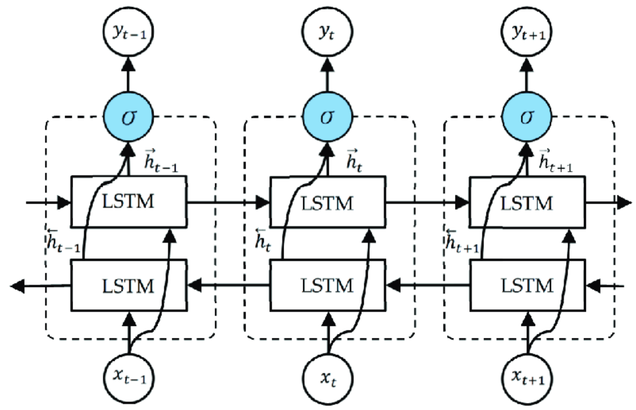


Fig. 10. BiLSTM NN construction

3) CNN: The structure of CNN is depicted in Fig. 11, which consists of convolution layers, pooling layers, and FCL where these layers adaptively learn spatial hierarchies of features through the back-propagation learning [30], [31], [32] and [33].

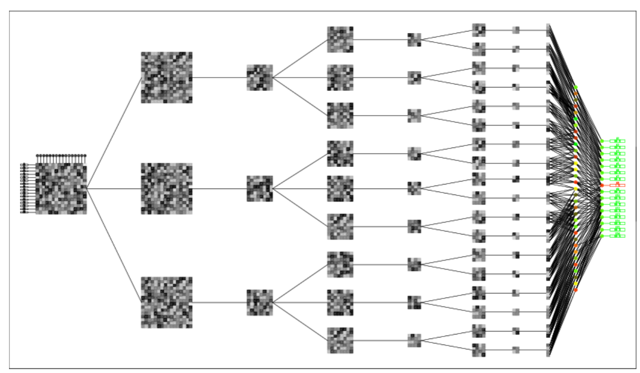


Fig. 11. CNN construction

B. Data sets

The real non-faulty data set collected from sensors array of real oil heating reactor in Egypt over a period of one month is used for training the suggested networks.

This stage of the work in this paper focuses on the temperature sensors readings of 34 sensors arranged in a matrix of 42660 rows and 34 columns.

The Faulty data is created using the different types of faulty models mentioned previously. Thus, the suggested network will be trained for classifying the sensor readings into faulty or non-faulty class.

C. Processing Data

The following procedures are conducted for preparing the data-set for classification:

1- The data will be divided into test and training data, in which the test data is 25% from the total data.
2- Choosing the fault type that will be used to test the NN, where the faulty data is detected from this type.
3- Choosing the fault level that will be used to test the NN, where the faulty data is detected from this fault level.
4- Label both of the training and test data without and with the corresponding type of fault and its level by using two classes to be inputs to the suggested N.Ns

V. Construction of Used NN

In this paper, the performance of three types of neural networks based-classification (FCNN, BiLSTM and CNN) are compared for detecting the faulted temperature sensor using various fault types with different levels .

A. FCNN

FCNN used in this paper is consist of the following layers as shown in Fig. 12:

- Input layer consist of 34 input nodes; one for each sensor reading.
- Hidden layer has 70 nodes.
- Output layer consist of 1 output (label) node for the (+ev) result which means that the sensor is faulty and (-ev) result which means that the sensor is not faulty.

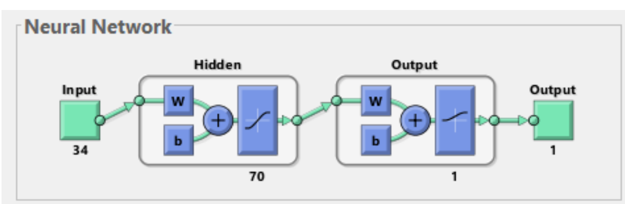


Fig. 12. Proposed FCNN construction

The training parameter used in the FCNN will be as follows:-

- Scaled conjugate gradient back propagation “trainscg” [34] is used as the training function.
- Mean Square Error “mse” [35] is used to evaluate the training result by calculating the error between actual and network outputs.
- Maximum of 1000 epoch will be applied.
- 6 validation checks will be used.

The FCNN have 31995 non-faulty readings and 31995 faulty readings.

In test phase, 10665 faulty readings and 10665 non-faulty readings are used for testing each sensor.

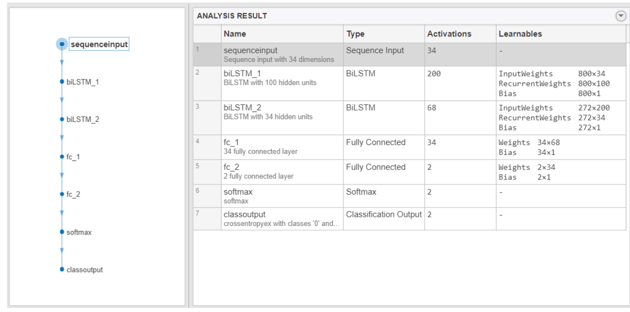


Fig. 13. Proposed BiLSTM RNN construction

B. BiLSTM

BiLSTM is used to classify a Sequence of data as shown in Fig. 13. The BiLSTM network consists of sequence input layer, two BiLSTM layers, two fully connected layers, softmax layer and finally classification layer.

1) The data sequences of sensors readings is divided into batches; each batch consists of 15 consecutive readings.
2) Thus, the input data to BiLSTM network consists of array of size (15x1) to each sensor reading where for all sensors are (15x34).
3) 2133 non-faulty and 2133 faulty readings are used for training the network, while 711 readings from each of the faulty and non-faulty data are utilized for test phase.

- There are 34 input nodes; one for each sensor; 15 consecutive reading.
- The BiLSTM NN uses 200 BiLSTM nodes in one hidden layer.
- The BiLSTM NN uses 68 BiLSTM nodes in one hidden layer.
- The BiLSTM NN uses 34 nodes in FCL.
- The BiLSTM NN uses 2 output nodes; one for (+ve) Faulty sensor and one for (-ev) non faulty sensor.

The training parameter used in the FCNN will be as follows:

- “adam” is used as the training function.
- 40 epochs maximum will be used.
- 30 iterations per epoch applied.

For the BiLSTM training there is 2133 non-faulty reading runs and 2133 faulty reading runs.

For each sensor test there is 711 faulty reading runs, and 711 the non-faulty runs.

C. CNN

The inputs to CNN must be of two dimension matrix as the gray scale picture shown in Fig. 15. The data sequences of sensors readings are divided into batches; each batch consists of 15 consecutive readings. Thus, the input data to CNN network consists of matrix of size (15x34).

Fig. 14. Input picture to CNN

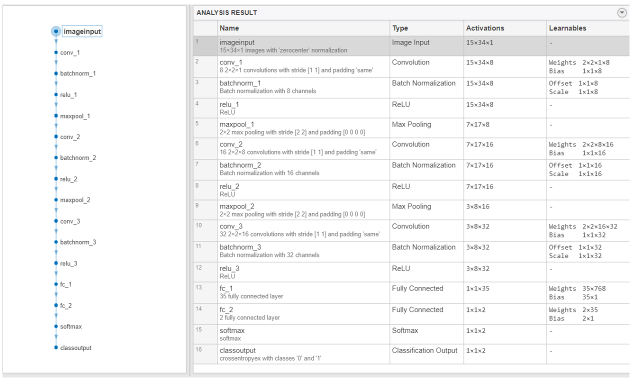


Fig. 15. Proposed CNN construction

Consequently, 2133 non-faulty and 2133 faulty readings are used for training the network while 711 faulty readings and 711 non-faulty readings are utilized for test phase.

The CNN construction is shown in Fig. 14.

so there is 2133 non-faulty and 2133 faulty training runs and 711 non-faulty and 711 faulty test run each consist of (15X34) Reading.

- The CNN has 15×34×1 images as input.
- The CNN has 8 (2×2) convolutions.
- The CNN has (2X2) Max Pooling.
- The CNN has 16 (2×2) convolutions.
- The CNN has (2X2) Max Pooling.
- The CNN has 32 (2×2) convolutions.
- The CNN has (2X2) Max Pooling.
- The CNN has 35 nodes in F.C.L.
- The CNN has 2 nodes in F.C.L.
- The CNN has Softmax Layer.
- The CNN has 2 output nodes one for (+ve) Faulty sensor and one for (-ev) non faulty Sensor.

The training parameter used in the CNN will be as follows:

- The Optimization algorithm named stochastic gradient descent with momentum "sgdm" is used for more efficient neural network weights during the network training [36] is used as the Training Function.

TABLE I. Network Parameters Comparison

Parameter	FCNN	BiLSTM N.N	CNN
Input Type	Value	Matrix	Picture
Number Of Layers	3	5	11
Input Layer	34 Node	34 Node	1 Node
Layer-2	F.C	BiLSTM	Convolution
Layer-3		BiLSTM	Max Pooling
Layer-4		F.C	Convolution
Layer-5			Max Pooling
Layer-6			Convolution
Layer-7			Max Pooling
Layer-8			F.C
Layer-9			F.C
Layer-10			Soft Max
Output Layer	2 node	2 node	2 node
Maximum Epoch	1000	40	100
Validation Check	6		
Iteration/epoch	1	30	16
Training Points	63990	4266	4266
Training Function	trainscg	adam	sgdm
Test Points	21330	1422	1422

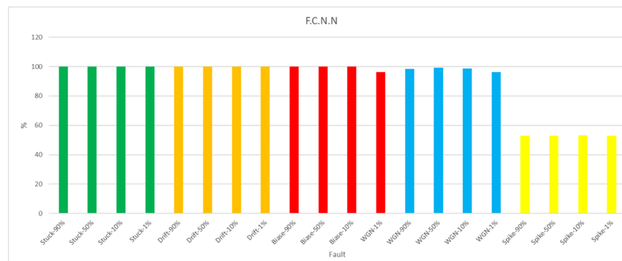


Fig. 16. FCNN performance

- 100 epoch maximum is used.
- 16 iteration per epoch is used.

D. Parameter comparison

Summary combining the comparison between the different neural networks constructions is shown in Table I.

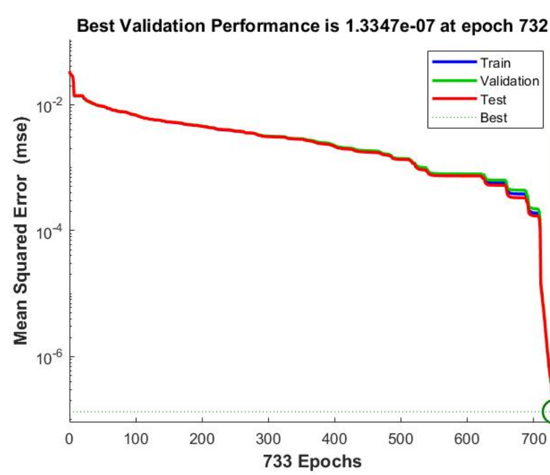


Fig. 17. FCNN training result

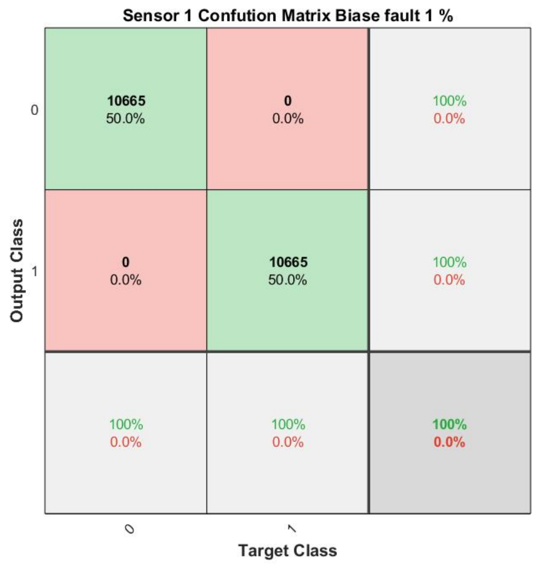


Fig. 18. FCNN confusion matrix

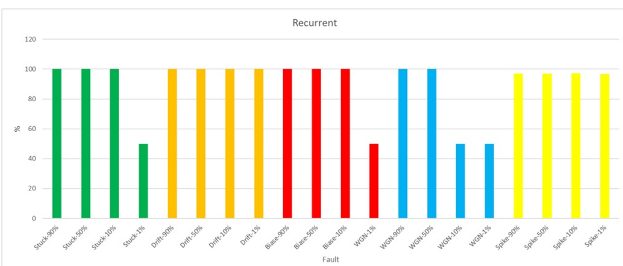


Fig. 19. BiLSTM performance

VI. Network Results Comparison

This section is dedicated to discuss the results of training the N.Ns used in this paper.

A. Performance Metrics for the Suggested Network

The performance of the suggested N.Ns is compared and evaluated using the following metrics [37]:

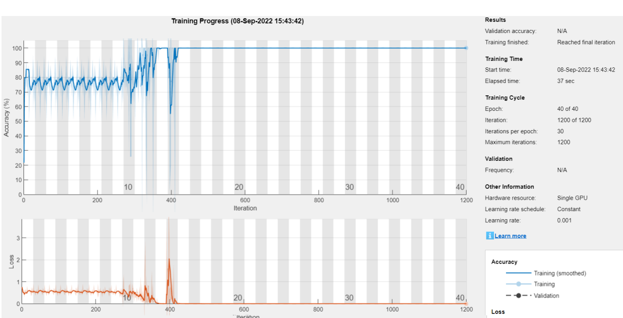


Fig. 20. BiLSTM training result

Fig. 21. BiLSTM confusion matrix

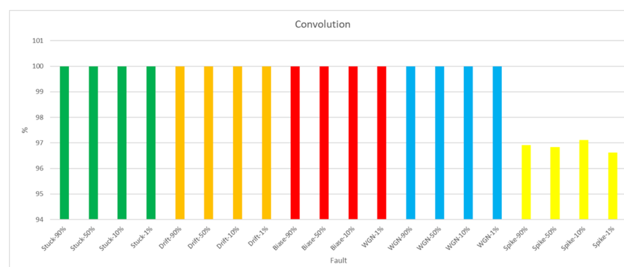


Fig. 22. CNN performance

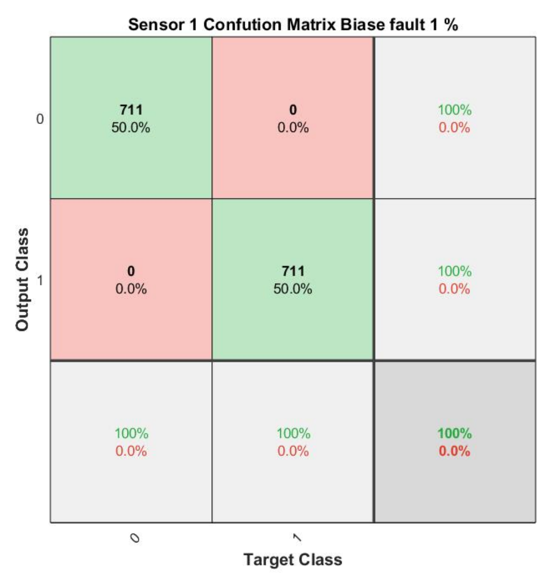


Fig. 23. CNN confusion matrix

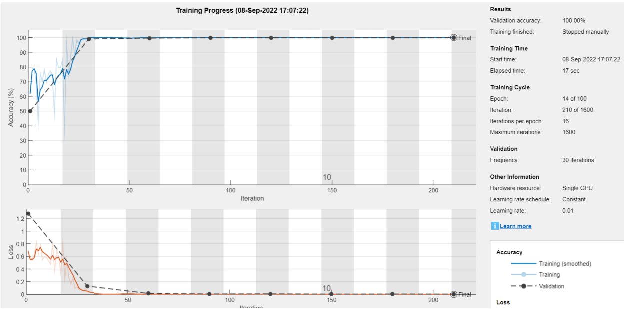


Fig. 24. CNN training result

	Stuck	Drift	Biase	WGN	Spike
Sequence	50	100	50	50	97
Convolution	100	100	100	100	97
Classic	100	100	100	96	53

	Sequence	Convolution	Classic
Stuck	50	100	100
Drift	100	100	100
Biase	50	100	100
WGN	50	100	96
Spike	97	97	53

Fig. 25. 1% fault level comparison for all NNs

1- Accuracy: It is the ratio of correct predictions to total predictions.
2- Precision: the ratio of correct fault detection to all fault predictions.
3- Recall:It is the ratio of correct fault detection to all true fault readings.
4- F-Score:It is the harmonic mean of precision and recall

The results of NNs classification include four types of data samples which are described as follows:

I) True positive (TP): is the number of samples that the network label as faulty while they are in-fact a faulty readings.
II) True negative (TN): is the number of samples that the network label as non-faulty while they are in-fact a non-faulty readings.
III) False positive (FP): is the number of samples that the network label as faulty while they are in-fact a non-Faulty readings.
IV) False negative (FN): is the number of samples that the network label as non-Faulty while they are in-fact a faulty readings.

The aforementioned metrics according to the four types of classified data are calculated by the given equations as follows:

1- Accuracy

$$= \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

2- Precision

$$= \frac{TP}{TP + FP} \tag{7}$$

3- Recall

$$= \frac{TP}{TP + FN} \tag{8}$$

4- F-Score

$$= \frac{1}{\frac{1}{Percition} + \frac{1}{Recall}} \tag{9}$$

5- False positive rate (FPR)

$$= \frac{FP}{TN + FP} \tag{10}$$

6- False negative rate (FNR)

$$= \frac{FN}{TP + FN} \tag{11}$$

B. Results and Discussion Simulation

The suggested networks are trained by all types of fault with different levels which are: [1%, 10%, 50% and 90%]. The performance results of the suggested networks for detecting all types of fault with there different levels in metrics of accuracy, precision and recall are given in Table II:

C. Results

The following results is obtained after the training and testing of the N.Ns

1) FCNN performance: The comparison of classification accuracy of FCNN for detecting the different fault types with its different levels is shown in Fig. 16.

As shown above the FCNN is very weak in detecting spike fault with accuracy of 50% while its good in detecting the other faults with their different levels where the results of detection are acceptable with accuracy ranges from 96% to 100%.

While the mean square error versus the number of epochs in training FCNN is shown in Fig. 17, confusion matrix is shown in Fig. 18.

2) BiLSTM performance: The comparison of classification accuracy of BiLSTM network for detecting the different fault types with its different levels is shown in Fig. 19.

As shown above, the accuracy of BiLSTM in detecting Basie, stuck and P.D.E faults is weak at fault levels (1%).On contrast, the accuracy results are acceptable with other faults.

While the network accuracy and the losses versus the number of epochs in training BiLSTM network are shown in Fig. 20, confusion matrix is shown in Fig. 21.

TABLE II. NN RESULTS COMPARISON

NN TYPE	Fault Type	Fault Level	Accuracy	Precision	F-score
CNN	Bias	90%	100%	100%	100%
CNN	Bias	50%	100%	100%	100%
CNN	Bias	10%	100%	100%	100%
CNN	Bias	1%	100%	100%	100%
CNN	Drift	90%	100%	100%	100%
CNN	Drift	50%	100%	100%	100%
CNN	Drift	10%	100%	100%	100%
CNN	Drift	1%	100%	100%	100%
CNN	Spike	90%	96.9%	100%	100%
CNN	Spike	50%	96.8%	100%	100%
CNN	Spike	10%	97.1%	100%	100%
CNN	Spike	1%	96.6%	100%	96.8%
CNN	Stuck	90%	100%	100%	96.7%
CNN	Stuck	50%	100%	100%	97%
CNN	Stuck	10%	100%	100%	96.5%
CNN	Stuck	1%	100%	100%	100%
CNN	P.D.E	90%	100%	100%	100%
CNN	P.D.E	50%	100%	100%	100%
CNN	P.D.E	10%	100%	100%	100%
CNN	P.D.E	1%	100%	100%	100%
BiLSTM	Bias	90%	100%	100%	100%
BiLSTM	Bias	50%	100%	100%	100%
BiLSTM	Bias	10%	100%	100%	100%
BiLSTM	Bias	1%	50%	0%	0%
BiLSTM	Drift	90%	100%	100%	100%
BiLSTM	Drift	50%	100%	100%	100%
BiLSTM	Drift	10%	100%	100%	100%
BiLSTM	Drift	1%	100%	100%	100%
BiLSTM	Spike	90%	96.9%	100%	96.8%
BiLSTM	Spike	50%	96.8%	100%	96.7%
BiLSTM	Spike	10%	97.1%	100%	97%
BiLSTM	Spike	1%	96.6%	100%	96.5%
BiLSTM	Stuck	90%	100%	100%	100%
BiLSTM	Stuck	50%	100%	100%	100%
BiLSTM	Stuck	10%	100%	100%	100%
BiLSTM	Stuck	1%	50%	0%	0%
BiLSTM	P.D.E	90%	100%	100%	100%
BiLSTM	P.D.E	50%	100%	100%	100%
BiLSTM	P.D.E	10%	50%	0%	0%
BiLSTM	P.D.E	1%	50%	0%	0%
F.C.N.N	Bias	90%	100%	100%	100%
F.C.N.N	Bias	50%	100%	100%	100%
F.C.N.N	Bias	10%	100%	100%	100%
F.C.N.N	Bias	1%	100%	100%	100%
F.C.N.N	Drift	90%	100%	100%	100%
F.C.N.N	Drift	50%	100%	100%	100%
F.C.N.N	Drift	10%	100%	100%	100%
F.C.N.N	Drift	1%	100%	100%	100%
F.C.N.N	Spike	90%	53.1%	100%	11.6%
F.C.N.N	Spike	50%	53.1%	100%	11.5%
F.C.N.N	Spike	10%	53.2%	100%	11.9%
F.C.N.N	Spike	1%	53.1%	100%	11.7%
F.C.N.N	Stuck	90%	100%	100%	100%
F.C.N.N	Stuck	50%	100%	100%	100%
F.C.N.N	Stuck	10%	100%	100%	100%
F.C.N.N	Stuck	1%	100%	100%	100%
F.C.N.N	P.D.E	90%	98.5%	100%	98.4%
F.C.N.N	P.D.E	50%	99.2%	100%	99.2%
F.C.N.N	P.D.E	10%	98.6%	100%	98.6%
F.C.N.N	P.D.E	1%	96.4%	99.8%	96.2%

3) CNN performance: The comparison of classification accuracy of CNN for detecting the different fault types with there different levels is shown in Fig. 22.

As shown above the result of the CNN is accepted even at spike fault and low level faults.

While the network accuracy and the losses versus the number of epochs in training CNN network are shown Fig. 23, confusion matrix is shown in Fig. 24.

4) The networks accuracy comparison at 1% fault level:: The comparison between all NN performance at various fault types and 1% fault level is shown in Fig. 25.

As shown in this figure, the results demonstrate that the CNN has the best performance in detecting all fault types and levels.

VII. CONCLUSION

In this paper, different neural networks are suggested for detecting the fault in temperature sensor of sensors array in oil heating reactor.

From the given results in this paper, the networks performance in metrics of Accuracy, Precision and Recall are varying depending on the network type, fault type, and fault level.

The CNN performance for fault detection outperforms the other networks, and it is the best so far to detect even very low level faults such as 1% fault level with accuracy reached to 97%.

This results show that the CNN is the best candidate to be implemented for fault diagnosis in future works, that will be focused on fault diagnosis and fault elimination.

REFERENCES

[1] Y. Xu, R. Meng, and Z. Yang, "Research on micro-fault detection and multiple-fault isolation for gas sensor arrays based on serial principal component analysis," *Electronics*, vol. 11, no. 11, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/11/1755

[2] M. Hojabri, S. Kellerhals, G. Upadhyay, and B. Bowler, "Iot-based pv array fault detection and classification using embedded supervised learning methods," *Energies*, vol. 15, no. 6, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/6/2097

[3] Z. Wang, L. Yao, and Y. Cai, "Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine," *Measurement*, vol. 156, p. 107574, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224120301111

[4] H. Han, X. Cui, Y. Fan, and H. Qing, "Least squares support vector machine (ls-svm)-based chiller fault diagnosis using fault indicative features," *Applied Thermal Engineering*, vol. 154, pp. 540–547, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359431118337992

[5] S. U. Jan, Y. D. Lee, and I. S. Koo, "A distributed sensor-fault detection and diagnosis framework using machine learning," *Information Sciences*, vol. 547, pp. 777–796, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025520308422

[6] S. Gajjar, M. Kulahci, and A. Palazoglu, "Real-time fault detection and diagnosis using sparse principal component analysis," *Journal of Process Control*, vol. 67, pp. 112–128, 2018, big Data: Data Science for Process Control and Operations. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152417300677

[7] I. Guney, E. Kilic, O. Ozgonenel, M. Ulutas, and E. Karadeniz, "Fault detection in induction motors with independent component analysis (ica)," *IEEE Bucharest Power-tech*, vol. 1, 06 2009.

[8] Y. Xu, R. Meng, and Z. Yang, "Research on micro-fault detection and multiple-fault isolation for gas sensor arrays based on serial principal component analysis," *Electronics*, vol. 11, p. 1755, 05 2022.

[9] Y. Kim, K. M. Jeon, Y. Kim, C.-H. Choi, and H. K. Kim, "A lossless compression method incorporating sensor fault detection for underwater acoustic sensor array," *International Journal of Distributed Sensor Networks*, vol. 13, p. 155014771774784, 12 2017.

[10] Y. Rui-jun, D. Dan-feng, and C. Yan, "Fault detection of gas sensor arrays based on knn rules," 11 2019, pp. 74–78.

[11] F. Wibowo and W. Putra, "Sensor array fault detection technique using kalman filter," 12 2019, pp. 124–128.

[12] A. Kouadri, M. Hajji, M.-F. Harkat, K. Abodayeh, M. Mansouri, H. Nounou, and M. Nounou, "Hidden markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems," vol. 150, 2020, pp. 598–606. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148120300112

[13] X. Liu, D. Pei, G. Lodewijks, Z. Zhao, and J. Mei, "Acoustic signal based fault detection on belt conveyor idlers using machine learning," *Advanced Powder Technology*, vol. 31, no. 7, pp. 2689–2698, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921883120301898

[14] S. Basangar and B. Tripathi, "Literature review on fault detection of equipment using machine learning techniques," in *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2020, pp. 62–67.

[15] L. Janssen and I. Lopez Arteaga, "Data processing and augmentation of acoustic array signals for fault detection with machine learning," *Journal of Sound and Vibration*, vol. 483, p. 115483, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022460X20303151

[16] Q. Mao, Y. Wang, and S. Huang, "Baseline-free sensor fault detection for piezoelectric array," *Journal of Intelligent Material Systems and Structures*, vol. 33, p. 1045389X2110188, 05 2021.

[17] *Framework for Reliable Fault Detection with Sensor Data*, 09 2022, pp. 41–76.

[18] J. Yang, Y. Chen, and Z. Sun, "A real-time fault detection and isolation strategy for gas sensor arrays," 05 2017, pp. 1–6.

[19] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar array fault detection using neural networks," 05 2019, pp. 196–200.

[20] E. Namigo, "Fault detection using neural network," *Journal of Theoretical and Applied Physics*, vol. 1, p. 70, 02 2017.

[21] S. Huang, K. Tan, P. Er, and T. Lee, *Fault Types and Modelling*, 03 2020, pp. 7–18.

[22] L. Li, G. Liu, L. Zhang, and Q. Li, "Sensor fault detection with generalized likelihood ratio and correlation coefficient for bridge shm," vol. 442, 2019, pp. 445–458. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022460X18307430

[23] X. Luo, K. Fong, Y. Sun, and M. Leung, "Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system," *Energy and Buildings*, vol. 186, pp. 17–36, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778818329207

[24] U. Saeed, S. U. Jan, Y.-D. Lee, and I. Koo, "Fault diagnosis based on extremely randomized trees in wireless sensor networks," *Reliability Engineering and System Safety*, vol. 205, p. 107284, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095183202030781X

[25] Y. Liu, X. Yan, C.-a. Zhang, and W. Liu, "An ensemble convolutional neural networks for bearing fault diagnosis using multi-sensor data," *Sensors*, vol. 19, no. 23, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/23/5300

[26] S. Jeong, M. Ferguson, R. Hou, J. P. Lynch, H. Sohn, and K. H. Law, "Sensor data reconstruction using bidirectional recurrent neural network with application to bridge monitoring," *Advanced Engineering Informatics*, vol. 42, p. 100991, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034619305646

[27] S. M. Gharghory, "A hybrid model of bidirectional long-short term memory and cnn for multivariate time series classification of remote sensing data," *Journal of Computer Science*, vol. 17, no. 9, pp. 789–802, Sep 2021. [Online]. Available: https://thescipub.com/abstract/jcssp.2021.789.802

[28] Y. Tong, P. Wu, J. He, X. Zhang, and X. Zhao, "Bearing fault diagnosis by combining a deep residual shrinkage network and bidirectional lstm," *Measurement Science and Technology*, vol. 33, no. 3, p. 034001, dec 2021. [Online]. Available: https://dx.doi.org/10.1088/1361-6501/ac37eb

[29] T. Han, R. Ma, and J. Zheng, "Combination bidirectional long short-term memory and capsule network for rotating machinery fault diagnosis," *Measurement*, vol. 176, p. 109208, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224121002256

[30] Y. Sun, H. Zhang, T. Zhao, Z. Zou, B. Shen, and L. Yang, "A new convolutional neural network with random forest method for hydrogen sensor fault diagnosis," *IEEE Access*, vol. 8, pp. 85 421–85 430, 2020.

[31] X. Lu, P. Lin, S. Cheng, Y. Lin, Z. Chen, L. Wu, and Q. Zheng, "Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph," *Energy Conversion and Management*, vol. 196, pp. 950–965, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890419307332

[32] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mechanical Systems and Signal Processing*, vol. 133, p. 106272, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S088832701930487X

[33] C. Wu, P. Jiang, C. Ding, F. Feng, and T. Chen, "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network," *Computers in Industry*, vol. 108, pp. 53–61, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016636151830513X

[34] P. K. Vadla, Y. V. R. Naga Pawan, B. P. Kolla, and S. L. Tripathi, "Accurate detection and diagnosis of breast cancer using scaled conjugate gradient back propagation algorithm and advanced deep learning techniques," pp. 99–112, 2021.

[35] K. Das, J. Jiang, and J. Rao, "Mean squared error of empirical predictor," *Annals of Statistics*, vol. 32, 07 2004.

[36] W. Ilboudo, T. Kobayashi, and K. Sugimoto, "Robust stochastic gradient descent with student-t distribution based first-order momentum," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, pp. 1–14, 12 2020.

[37] D. Powers and Ailab, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation," *J. Mach. Learn. Technol*, vol. 2, pp. 2229–3981, 01 2011.

The Impact of Peer Code Review on Software Maintainability in Open-Source Software: A Case Study

Aziz Nanthaamornphong[1], Thanyarat Kitpanich[2]
College of Computing, Prince of Songkla University, Phuket, Thailand

***Abstract*—Recently, open-source software (OSS) has become a considerably popular and reliable source of functionality corrections. OSS also allows software developers to reduce technical debt in software development. However, previous studies have shown that the main problem within OSS development is the lack of systematic processes and formal documents related to system development, such as requirements, designs, and testing. This feature of OSS development causes problems in the software quality, such as those related to security and maintainability. In this research, the authors focused on the software's maintainability because this attribute has to greatest potential to reduce the cost and increase the productivity of the software development process. There is currently no existing research that examines whether OSS developers pay attention to software maintainability. To better understand how OSS developers improve software maintainability, this research aims to answer the question: "Are developers interested in software maintainability under the modern code review of open-source software projects?" To answer the research question, the authors investigated the code review process in which the OSS developers changed the code based on a review of code comments related to maintenance and collected the sub-characteristics associated with software maintainability from the existing literature. The authors examined the review comments from two OSS projects: Eclipse and Qt. The results suggest that the number of code revisions due to maintenance issues was moderate and that the OSS developers tend to improve source code quality. This direction could be observed from the increasing number of modifications on given maintenance-based comments over the years. Therefore, an implication of this is the possibility that OSS project developers are interested in software maintainability.**

***Keywords*—Open-source software; software maintainability; code review**

I. Introduction

Open-source software (OSS) has received considerable attention from users worldwide, especially among organizations and commercial industries that widely utilize OSS. Notably, OSS is often a key to organizational success for entrepreneurs. OSS can create the opportunity for organizational growth, enhance performance, increase the options for software use, and decrease operating costs. Moreover, users from around the world can use OSS freely without fees, allowing software developers to access the OSS code repository [1] and jointly develop software [2]. OSS projects are being continuously developed and improved by software developers with expertise and experience in system development. Therefore, OSS has become popular and credible in terms of effectiveness and functionality.

A previous study [3] noted that OSS development lacks systematic processes or procedures and formal documents related to system development, such as requirements, design, testing, etc. Considering the aforementioned problems, most software developers working on OSS projects solve relevant problems when they are reported by users or software developers on the same team. One frequent problem is the occurrence of a defect or bug in the system. When the source code is frequently modified, the size of the code increases, and it generally becomes more complex. Moreover, some source codes can be difficult to understand. These issues can lead to serious problems, such as increases in time and development costs. Additionally, software developers and commercial entrepreneurs who want to develop software often cannot comprehend old codes that are complex and poorly described. Thus, accessibility may be poor, and call functions may fail. In addition, poor results or inaccurate information may not meet user expectations, and possible vulnerabilities in the system may lead to the pirating of information or viruses. All of these issues can cause software project development to ultimately fail. The quality of the source code of OSS projects could be improved [4], in particular, the relevant software security and maintainability of these projects [5].

Software maintainability can improve product quality at a low cost and ensure that software development meets the intended objective. Therefore, software maintainability is one of the most important steps in the software development life cycle because most software developers spend approximately 40-50% of their time identifying defects or errors during the development process or after product delivery [6]. Notably, maintainability can solve OSS quality issues and prevent problems that may affect the system in the future. The code review process is an important part of the software development process that ensures the creation of high-quality software and the implementation of successful OSS projects [7]. Thus, many OSS projects use "peer code review" or "modern code review" [8] as a guideline for developing and improving code.

Peer code review is commonly used in software engineering for quality control [9]. Generally, a code change must be reviewed by the software developers on the same development team or reviewers other than the code editor. Generally, the code review process does not have a fixed format, but it prevents problems associated with patch files and helps avoid bugs or errors that can influence the long-term applicability of the software. Moreover, the comments from reviewers can be used to identify bugs or defects in the source code, improve the source code, modify the source code to meet certain standards, and improve code readability [10]. However, some comments may not be directly related to software quality improvements.

A review of studies on software development in OSS projects indicated that the comments provided via peer code review have not been extensively investigated. Developers should review comments related to software maintainability in OSS projects. These comments are often collected in the Gerrit system, which provides services related to the code review process and facilitates communication and the exchange of ideas about software development among software developers. In this study, two OSS projects: Eclipse (https://eclipse.org/) and Qt (https://www.qt.io/), are investigated. These projects were selected because they have different purposes and have already been completed. Therefore, extensive comment information is available for both projects. The projects provide the opportunity for researchers in the field of software engineering to access this information and study code reviews. The main research questions posed in this study are as follows:

1) How many sub-characteristics are related to software maintainability?
2) Are the developers of OSS projects interested in software maintainability under the peer code review?
3) What characteristics of software maintainability have the developers considered?

In order to answer these questions, text mining was applied to identify the sub-characteristics related to software maintainability that appear in comments provided by peer code review, and the latent Dirichlet allocation algorithm (LDA) was applied to develop new characteristics related to software maintainability. Quantitative data analysis was performed to identify the trends of developers in software maintainability for OSS projects. In addition to answering the main research questions, the results were used to determine the factors that affect the number of comments related to maintainability, and statistical analysis was conducted to answer the following research questions:

1) Does the size of the reviewer pool influence the comments related to maintainability?
2) How many comments related to maintainability are given by code reviewers each day?

The results of this study will provide empirical evidence for the research community regarding the software quality of OSS projects. Moreover, this study will shed light on the importance of software maintainability and help software developers understand sub-characteristics related to software maintainability, which usually appear in OSS projects.

The remainder of this paper is organized as follows. Section II provides the necessary background and related work. Section III describes the research methodology. Section IV shows the results of this study. Section V discusses the results. Section VI concludes the paper and outlines our future work.

II. Related Work

In this section, the theories and methodologies used to conduct the research are summarized, including the literature review, which provides the basic framework of this study.

A. Peer Code Review for OSS Projects

Code review is a way to reduce the risk of error and improve the quality of software by checking the source code to identify defects or bugs that may occur during software development. Moreover, such a review can standardize the source code [11]. When organizations or projects that require software development use this code review process, it ensures that the developed code will meet the specifications or needs of the user and that the impacts of source code modifications are minimized. Therefore, code review has been a best practice in software engineering for over 35 years [12]. "Peer code review," or "modern code review," is applied by open-source organizations and communities to improve the quality of source code [13]. The source code is standardized by deleting duplicate functions and removing irrelevant or unnecessary code. Moreover, the code review process allows detailed comments from experienced reviewers that can improve the source code. Additionally, peer code review facilitates the sharing of knowledge about system development [11] because code review services often utilize specific communication tools. The most common tools are CodeFlow, Gerrit, Collaborator, Crucible, Review Board, and Upsource.

B. Software Maintainability

ISO/IEC 25010 is the model used by organizations around the world because it has set the software quality standard for code review based on system and software requirements. Furthermore, ISO/IEC 25010 was selected in this study because it includes both quality evaluation for a system or software and the corresponding effects on stakeholders or users. The international standards that are used to assess quality include requirement definition and quality measurement and evaluation [14].

ISO/IEC 25010 defines software maintainability as "the degree of effectiveness and efficiency with which a product or system can be modified to improve it, correct it or adapt it to changes in the environment, and in requirements." In a previous study on software maintainability, Ghosh et al. [15] identified the 40 sub-characteristics that impact maintainability. These sub-characteristics are used as keywords in this study to obtain comments related to software maintainability.

C. Literature Review

Bakar and Arsat [16] studied the factors that affect the quality of OSS by measuring the quality of the source code using McCall's Quality Factor model. This model considers maintainability, accuracy, reliability, efficiency, and ease of use. The results indicated the relationships among lines of code and the complexity of the source code based on correlation analysis. The most influential factors were correctness, maintainability, efficiency, and usability. Related research on complexity measurements for source code generally uses Cyclomatic Complexity (CC) as an indicator, although there are newer indicators as well. For example, Walden et al. [17] analyzed code and used complexity metrics to predict the possible density gaps in 14 web applications. The study results could be used by software developers in OSS projects to create a simple system structure and clearly illustrate the development of software, especially security, because the number of gaps was decreased. Norick et al. [18] studied OSS projects developed in C/C++ to determine whether the number of revisions affected the quality of the source code. This analysis used the CC of each coded function and the density of comments as representatives of source code quality. However, the study did not find a correlation between the number of software developers and the software quality, and the average software quality was satisfactory for all the studied OSS projects. Schmidt and Porter [19] studied the software development process of the "Skoll" OSS project and of closed-source projects. They found that the community of OSS developers could control long-term maintainability procedures and development costs, and these characteristics created confidence and a level of acceptance in the quality of the software. Specifically, users had confidence in the software and corresponding system. Such practices are easier to implement in OSS projects compared to closed-source projects because the disclosure of sources allows software developers who are interested in OSS projects to voluntarily participate in quality improvement tasks.

Successful OSS projects usually participate in peer code reviews to assess the quality of the software [20]. The code review process begins with the software developers creating patch files, which add value to a project. Rigby [21] stated that using supportive tools for peer code review could affect the progress of OSS, and there was a subsequent push for software development that facilitates code review. As noted by Bosu and Carver [7], review tools have been widely used to study comments via platforms such as ReviewBoard. The researchers analyzed the number of participants in the project, the number of requests for code review, and the response time after a request for code review was submitted by comparing MusicBrainz Server and Asterisk. However, Bacchelli and Bird [22] studied motivation and challenge in searching for defects in Microsoft projects using the CodeFlow tool, which records data from code reviews. The comments aiming to guide code modification during the peer code review process were investigated. The results showed that reviewers with expertise in system development and reading and reviewing source code provided the most useful comments for developers.

Baysal et al. [10] sstudied the factors that affect the modification of patch file defects by extracting code review information for the WebKit project from Bugzilla. Most reviewers denied large patch files because developers modified or developed source code that was too complex, included unnecessary components, and/or affected other functions or modules. Tao et al. [23] proposed a method for developers to code patch files that were accepted by reviewers. They verified the causes of rejected patch files by collecting rejection information for Eclipse and Mozilla patch files. The information was collected and stored based on an online survey of 246 software developers. Considering the issues that caused a code fix to be accepted or rejected, the community can determine how to optimize the peer code review process.

fBosu et al. [10] searched for ways to improve the efficiency of peer code review, which was useful for system development. They analyzed 1.5 million comments from Microsoft projects to identify the characteristics of useful and useless comments. Some comments that software developers received contained incorrect information or comments unrelated to system development, and software developers took time to respond to these questions and discuss why they did not modify the source code accordingly. Moreover, the study identified the factors that were correlated to beneficial comments, such as review experience, organizational experience, and being on the same team as the code developers. Czerwonka and Greiler [24] found that up to 50% of the comments provided via the peer code review process were related to software maintainability and that very large patch files decreased the benefit of reviewer comments and increased the review time by six hours to a week. Nonetheless, they found that reviewer comments are beneficial and reflect the overall view of software development. Such a review system promotes cooperation in problem resolution and expedites the development and improvement of software. Such systems provide a standardized framework for researchers to better understand peer code review.

Maintainability is a very important factor because software maintenance accounts for 66% of the life cycle and 40-80% of the cost of development [25]. Code smells are among the most important problems that must be urgently solved. A code smell is any characteristic in a source code, such as a poorly designed structure or poor organization, that can lead to larger issues. These factors decrease the quality of software maintainability. Wagey et al. [26] presented a model of software maintainability and applied it to six OSS projects. The results showed that the model improved the structures of the OSS projects. Refactoring [27], a process that involves shortening lines of code or combining similar functions to reduce the complexity of the code [28], has also been used to improve software functionality. Rizvi and Khan [29] used regression analysis to develop a class-level maintainability estimation model called MEMOOD, which facilitated development and reduced the time required to improve or edit code before delivery. This model used the features of understandability and modifiability, which were calculated based on object-oriented metrics, to estimate the maintainability of the code considering its size and structural complexity. Khan and Khan [30] developed a model that was used to measure the quality of analyzability. Multiple linear regression was conducted for the structural design of a system in which complexity might affect the quality of software maintainability, especially its analyzability. However, they found that other factors, such as the size of a file or program, may also make maintainability difficult.

III. Research Methodology

This study analyzes the comments related to software maintainability from the peer code review of OSS projects using text mining techniques. The main procedure consists of two parts, as described in the subsequent subsections.

A. Development of Software for Searching and Storing Data from Gerrit

Software was developed for searching and storing data associated with Eclipse and Qt comments that were collected by Gerrit from 2017 to 2021. Before storing the data from Gerrit, tables were designed in the database to support the data used in this study. The stored information is as follows:

- ID - the unique number for each data entry;
- Patch_numbers - the number of patches;

- Created_on - the review request date;
- Uploader - the uploader's ID (Gerrit user name);
- Author - the code author;
- Reviewers - the reviewer's ID (Gerrit user name);
- File - the name of the file with the review request;
- Line - the number of changed lines of code;
- Message - the comment message; and
- Kind - the status of the changed code, such as trivial rebase, no code change or rework.

B. Analyzing a Dataset

In this step, the sub-characteristics related to software maintainability that appear in comments provided via the peer code review of OSS projects were searched using a text-mining technique. This technique also verifies that the comments were related to maintainability. This step is shown in Fig. 1, and each step is shown as follows.

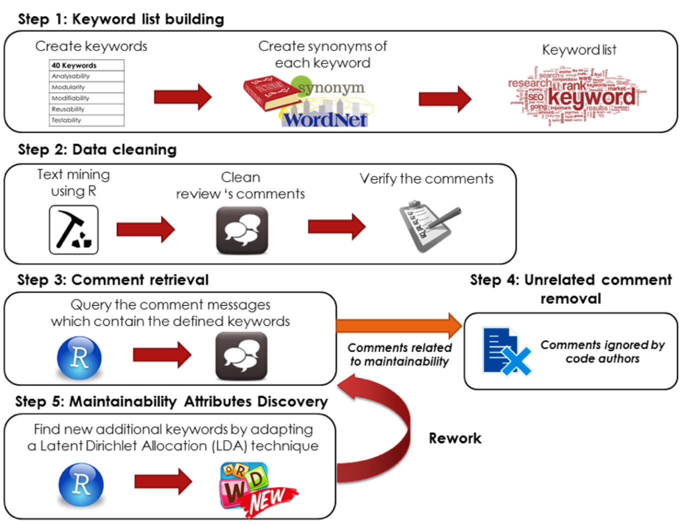


Fig. 1. The data analysis process

Keyword List Building - A set of keywords that are related to software maintainability, such as modularity, an-analyzability, and testability, was constructed. These keywords were selected based on the sub-characteristics that are related to the definition of "software maintainability". Previously, Ghosh et al. [31] used these collected keywords to search for comments related to software maintainability. However, they found that the reviewers of OSS projects typically did not use words in comments that matched these keywords. Therefore, the sub-characteristics related to software maintainability are converted to their root forms to reflect the basic meaning of the term; for example, "modifiability" was shortened to the root "modify". Then, synonyms for each root word, such as "edit", "improve", "solve", and "amend" (for "modify"), were also added to the search set. The words with the same meaning as these keywords were identified using WordNet, a website that provides synonyms for various terms.

Data Cleaning - In this study, R was used for data cleaning. Specifically, the text mining (tm) package was used in this step. In data cleaning, the stop word function was employed to remove terms that were not significant without changing the meaning of the sentence. For example, pronouns, conjunctions, and prepositions usually appear in each comment but could generally be removed. Moreover, whitespace, numbers, and special characters, such as special symbols (,), , *, #, and ! were deleted. Then, word stemming was performed by removing prefixes and suffixes. Words with the same root, such as "compatible" and "compatibility," have similar meanings; therefore, "ible" and "ibility" could be removed to obtain the root "compat". Another example is "connection", "unconnected", "connective" and "connecting", for which the root is "connect". Moreover, letters were converted from upper case to lower case to reduce the number of required indices and the processing time.

Comment Retrieval - In this step, a command set was developed to search for comments with keywords or meanings related to software maintainability. At least one word from the constructed keyword list must be matched, and the search results were converted into a file format for easy analysis. In order to search for comments related to software maintainability, we developed a command in R that connects to a MySQL database and searches based on various subcommands. For example, a set of commands is developed to find the word "modify," and words with the same meaning as "modify" appear in the text of comments in the database.

In order to search for comments containing keywords, we created a command set for each keyword for the convenience of searching and storing. After the comments related to software maintainability were obtained, each comment was reviewed to improve credibility. This check was manually performed to verify that each comment contained a keyword or related term. If the text in a comment did not match a keyword or related word, we deleted the text from the database. After verification, the results were stored in table format, and relations were preserved. The table shows the number of terms, either keywords or related words, that appear in comments and are related to software maintainability.

Unrelated Comment Removal - We analyzed the results from step 3 and determined the number of comments in Eclipse and Qt related to software maintainability. These results could help answer the question of whether the developers are interested in software maintainability in OSS projects based on peer code review. To determine whether a comment was considered in code modifications, a command set was established in R by pairing comments related to software maintainability and comments from Gerrit. Notably, most comments from Gerrit included both the comments from reviewers and the comments of software developers who responded. This code modification assessment considers the type of code modification, such as trivial rebase or rework.

Additionally, the words in the reply related to code modifications, such as "done", "finished", or "edited" (as shown in Fig. 2), were considered. After checking the comments based on this procedure, the comments with no replies were removed from the database.

Maintainability Attribute Discovery - Thirty-three sub-

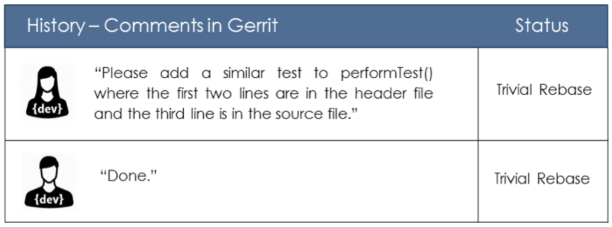


Fig. 2. Example of comment related to maintainability and responses of code authors

characteristics of software maintainability were selected, and other sub-characteristics of software maintainability hidden in comments related to Eclipse and Qt projects were searched using the LDA. This algorithm could be applied to search for topics related to maintainability. The result of the LDA in R presented the frequency of occurrence of words that were related to the keywords, and we were required to determine whether these words should be new keywords. If the answer was yes, the term was added to the keyword set, and steps 3 and 4 were repeated until there were no additional keywords.

IV. Results

Based on our research questions (described in Section I), the results of the analysis are presented.

A. How Many Subs-Characteristics are Related to Software Maintainability?

In Ghosh [31], 40 sub-characteristics related to maintainability were identified and studied. The following criteria were used to select sub-characteristics: 1) the definition of the sub-characteristic must be relevant to the definition of "software maintainability" according to ISO/IEC 25010, and 2) a sub-characteristic keyword cannot be a common or general word. The sub-characteristics of this study that did not meet those criteria were as follows:

1) Compliance,
2) Conciseness,
3) Delivery,
4) Documentation,
5) Impact Analysis,
6) Programming Language; and
7) Self-descriptiveness.

The obtained sub-characteristics that met these criteria were converted into verbs or nouns by removing the suffix, and the final keywords were easy to search for in the comments of reviewers on OSS projects. The results of keyword determination are shown in Table I.

TABLE I. The List of Keywords

No.	sub-characteristics	Keyword	No.	sub-characteristics	Keyword
1	Accuracy	Accuracy	18	Implementation	Implement
2	Adaptability	Adapt	19	Instrumentation	Instrument
3	Analyzability	Analyze	20	Integrability	Integrate
4	Augmentability	Augment	21	Localization	Localization
5	Availability	Available	22	Modifiability	Modify
6	Changeability	Change	23	Modularity	Modular
7	Completeness	Complete	24	Perfectiveness	Perfect
8	Complexity	Complex	25	Portability	Portable
9	Comprehensibility	Comprehension	26	Readability	Read
10	Consistency	Consistency	27	Reusability	Reuse
11	Correctability	Correc	28	Simplicity	Simple
12	Durability	Durable	29	Stability	Stable
13	Efficiency	Efficient	30	Standardization	Standard
14	Effort	Effort	31	Testability	Test
15	Expandability	Expand	32	Traceability	Trace
16	Extensibility	Extension	33	Understandability	Understand
17	Flexibility	Flexible			

We then reviewed the keyword search results to identify potential words that should not be considered keyword synonyms. The criteria for synonym consideration are as follows:

1) Cannot be a common term used in the programming language, such as class, function, feature, method, object, or parameter, or a general computer-related term, such as system, computer, software, file, or command.
2) For duplicate synonyms, only synonyms from three root words are selected.
3) For different synonyms, all words from three roots are used as keywords.

The search process for the keywords and synonyms that met these criteria is shown in Table II. The software developers that use Eclipse and Qt can consider all characteristics of software maintainability to improve and edit their source codes. Therefore, in addition to the 33 characteristics identified by the LDA, the frequent terms in comments related to software maintainability were also searched. The LDA results in Table III contain keywords and synonyms such as change, adjust, test, and improve. Therefore, to choose the new maintainability characteristics from the LDA, we considered the most frequently used words in each group.

Considering the maintainability characteristic produced by the LDA, some groups did not contain appropriate words. However, Eclipse and Qt contained two new characteristics related to software maintainability: cohesion and duplicate. To define the new maintainability characteristics, the authors used comments provided by the reviewers to search and identify these two maintainability characteristics. The words that typically appear with "cohesion" and "duplicate" had a relationship value of 0.6. Thus, those keywords were associated with software maintainability at least 60% of the time. The words most highly correlated with the new types of maintainability characteristics in comments are considered to define the maintainability characteristics and group the synonyms (as shown in Table IV).

1) The cohesion module identifies the modules that work together by connecting the components of the modules. These components may work at the same time or have the same input but different operations.
2) The duplicate function combines functions with similar properties or procedures to reduce the duplication of functionality, which can potentially cause confusion in calling.

TABLE II. THE LIST OF KEYWORD AND SYNONYMS

No.	Keyword	Synonym
1	Accuracy	accuracy, exact, truth, certainty, precision, propriety, rectitude, validity, sure, definite, inevitable, rigorous, evident, categorical, explicit, just, lawful
2	Adapt	adapt, adjust, modulate, alter, fine, shape, regulate
3	Analyze	analyze, analysis, diagnose, assay, delineate, muse, anatomize
4	Augment	augment, amplify, spread, inflate, escalate, dilate, enhance, accumulate, increase, suffuse, raise, mount, aggrandize, splay, accrete
5	Available	available, accessible, handiness, obtainable, satisfactory, convenient, usable, benefit
6	Change	change, vary, remodel, permute, convert, transform, revamp, purge, reform
7	Complete	complete, full, absolute, plenary, finish, utter, flawless
8	Complex	complex, complicate, sophisticated, elaborate, manifold, labyrinthine, multiple, confuse, entangle, mix, muddle, discursive, tangle, intricate, bewildered, imbroglio, intricacy, jumble, obscurity
9	Comprehension	comprehension, finality, inference, conclusion, notion, realization, savvy
10	Consistency	consistency, coherence, pertinacity, adhesion, invariability, tenacity
11	Correct	correct, rectify, fit, favorable, appropriate, worthy, deserve, suitable, due, rightful, infallible, redress, regularize
12	Durable	durable, lasting, hardy, imperishable, substantial, permanent, long, immune, indissoluble, enduring, strong
13	Efficient	efficient, able, capable, competent, proficient
14	Effort	effort, endeavor, fighting, might, stamina, energy, strength, activity
15	Expand	expand, add, accretion, accrue, develop, flatten, grow, prosper, thrive
16	Extension	extension, flare, connect, continue, broaden, enlarge, widen, magnify, prolong, elongate, protract, proliferate
17	Flexible	flexible, elastic, dexterous, limber, resilient, springy, pliable, stretch
18	Implement	implement, execute, accomplish, achieve, resource, utilize, apply
19	Instrument	instrument, equipment, accessory, apparatus, machinery, appliance, device, fgadget
20	Integrate	integrate, gather, collect, compile, assemble, embody, coordinate, cooperate, harmonize, consolidate, sticking, combination, joining, amassing, hoard, compound
21	Localization	localization, limitation, narrowing, restriction, stint, definition, circumscription
22	Modify	modify, customize, edit, improve, solve, repair, amend, qualify
23	Modular	modular, configuration, component, constituent, ingredient, composition, complement, element, procedure, segment
24	Perfect	perfect, excellent, ideal, immense, keen, superb, wonderful, terrific, fantastic, splendid, magnificent, superior, entirety, consummate, faultless
25	Portable	portable, mild, soft, slight, lightweight, feathery, weak, mushy, flimsy
26	Read	read, peruse, pronounce, extrapolate, imagine, speculate, surmise, see, view, interpret, transliterate, look
27	Reuse	reuse, recycle, reiterate, rehash, reclaim, exploit, revise
28	Simple	simple, ease, expedient, facile, easy
29	Stable	stable, still, steady, invariable, constancy, endurance, firmness, fastness, indissolubility, sturdily
30	Standard	standard, formula, archetypal, representative, typical, characteristic, degree, tier, quality, criterion, measurement, imperative, touchstone, property
31	Test	test, verify, check, prove, evaluate, examine, assess, attempt, experiment, inspect, trial, proof, tryout, investigate
32	Trace	trace, pursue, tag, trail, detect, search, seek, probe, meaning, significance, hint, consequence, point, follow, behave, track, extract, transcribe
33	Understand	understand, explain, fathom, grasp, knowledge, perceive

TABLE III. THE LDA RESULTS

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
change	use	add	work	method
convert	sure	adjust	example	tool
parameter	suggest	file	make	class
patch	implement	test	code	review
index	cohesion	improve	model	package
name	resource	worth	duplicate	commit

TABLE IV. NEW KEYWORDS

No.	Keyword	Synonyms
1	Cohesion	cohesion, sticky, tough, gummy, leathery, tenacious
2	Duplicate	duplicate, repetitive, double, copy, transcript, counterpart, facsimile, reproduce, mimeograph, repeat, replicate

B. Are the Developers of OSS Projects Interested in Software Maintainability Under the Peer Code Review?

The Qt and Eclipse projects had totals of 309,165 and 108,357 comments, respectively. Thirty-five keywords were identified based on the characteristics that were relevant to maintainability. These keywords included 33 previously used terms and two terms obtained from the LDA. A script was developed in program R for text mining. The results indicated that 39,638 and 93,629 comments were related to maintainability in the Qt and Eclipse projects, respectively, based on the keywords (36.58% and 30.28% in Fig. 3). In both projects, the number of comments related to software maintainability was low compared to the total number of comments.

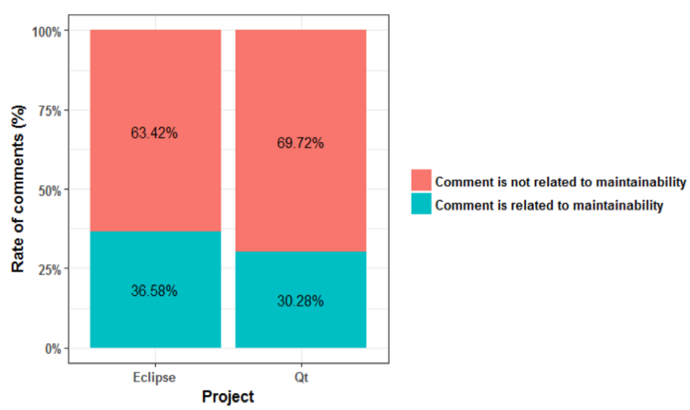


Fig. 3. Rate of comments are related to maintainability and comments are not related to maintainability

The keywords were used 64,616 and 149,610 times in the comments of the Eclipse and Qt projects, respectively. In order to determine the extent to which software developers of Eclipse and Qt prioritized software maintainability, the comments related to software maintainability and code editing were analyzed. Overall, 36,975 and 94,408 comments led to source code modifications for Eclipse and Qt, representing 57.22% and 63.10% of all comments, respectively. Fig. 4 shows that the comments provided to the software developers of Eclipse and Qt resulted in source code edits 50-60% of the time.

We found that software developers of Eclipse and Qt edited

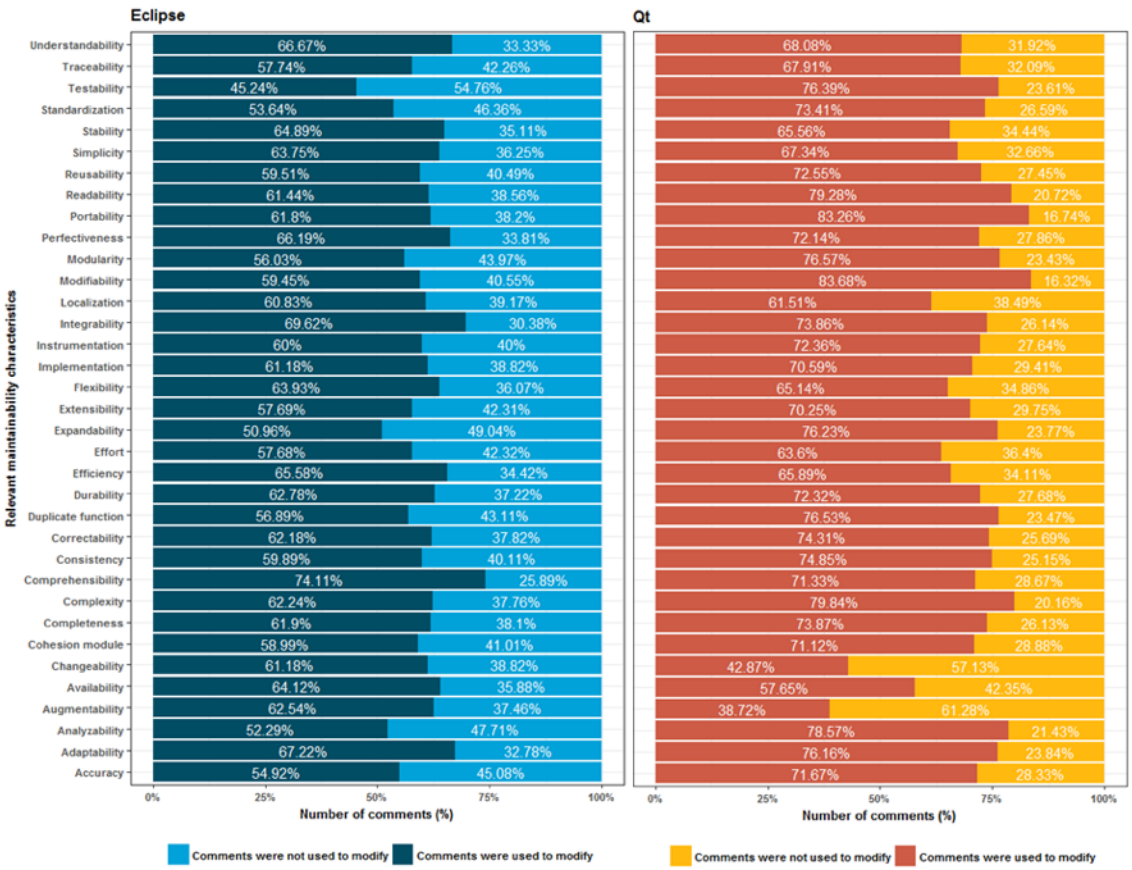


Fig. 4. Ratio of comments to modification code

the source code according to the comments related to software maintainability approximately 50-60% of the time. This result provides empirical evidence suggesting that the software developers of Eclipse and Qt prioritized software maintainability in more than 50% of source code modifications.

The Pearson correlation coefficient was determined in this study at a 95% confidence interval to obtain the relationship between comments related to maintainability and comments related to code changes in these two OSS projects. Based on an analysis in R, the Pearson correlation coefficients of Eclipse and Qt were 0.995 and 0.993, respectively. These coefficients indicate that the number of comments related to software maintainability and the number of comments related to code modifications were highly related. In other words, when the reviewers of Eclipse and Qt provided more software maintainability comments, the software developers made more source code edits.

C. What Characteristics of Software Maintainability have the Developers Considered?

The five most common comments related to software maintainability were investigated, and the number of code modifications related to those comments was determined. Fig. 5 illustrates that the software developers of OSS projects make the most source code edits based on comments that discuss readability, and the rates of related changes are 61.44% for Eclipse and 72.14% for Qt.

D. Does the Size of the Reviewer Pool Influence the Comments Related to Maintainability?

In the two OSS projects, most of the comments that led to code edits were made by reviewers and other software developers. Therefore, the number of comments was compared to the number of comments related to software maintainability each month. Moreover, the sizes of the groups or teams of software developers were considered to determine how many software developers reviewed the code.

R was used to calculate correlation coefficients and obtain the relationship between the number of software developers who provided comments related to software maintainability and the number of comments related to software maintainability. The annual Pearson correlation coefficients in the study period were 0.78, 0.88, 0.72, 0.53, and 0.75 for Eclipse and 0.85, 0.86, 0.57, 0.005, and 0.17 for Qt (as shown in Fig. 6 and 7).

Fig. 6 shows that the group size of reviewers and the number of comments related to software maintainability were highly related in the Eclipse project. Fig. 7 illustrates that the group size of reviewers and the numbers of comments related to the project maintainability in Qt were highly correlated.

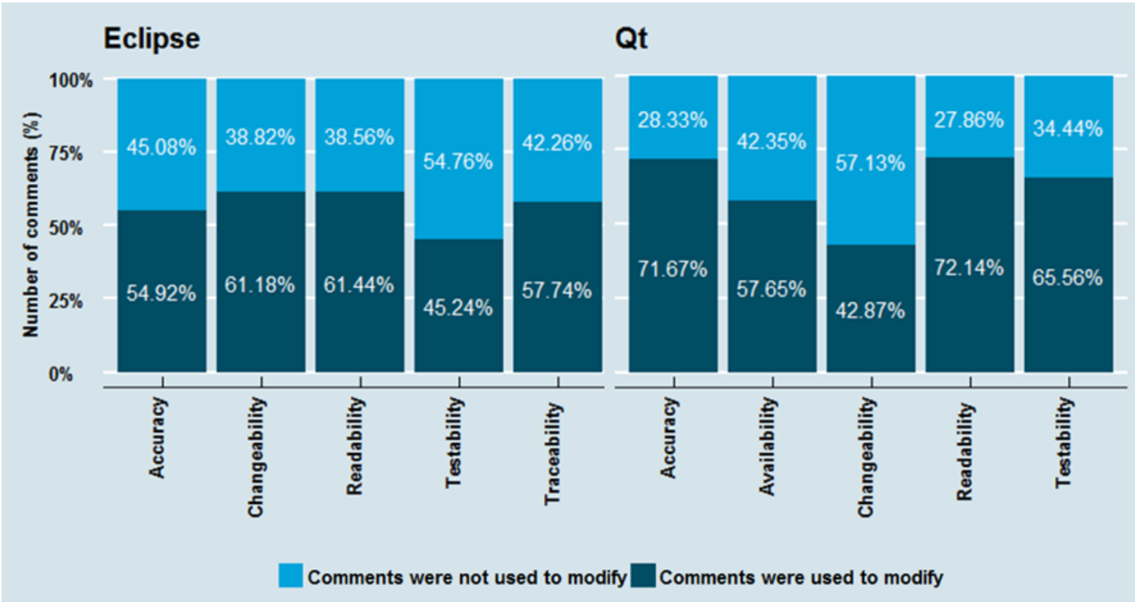


Fig. 5. The percentage of addressed comments related to sub-characteristics

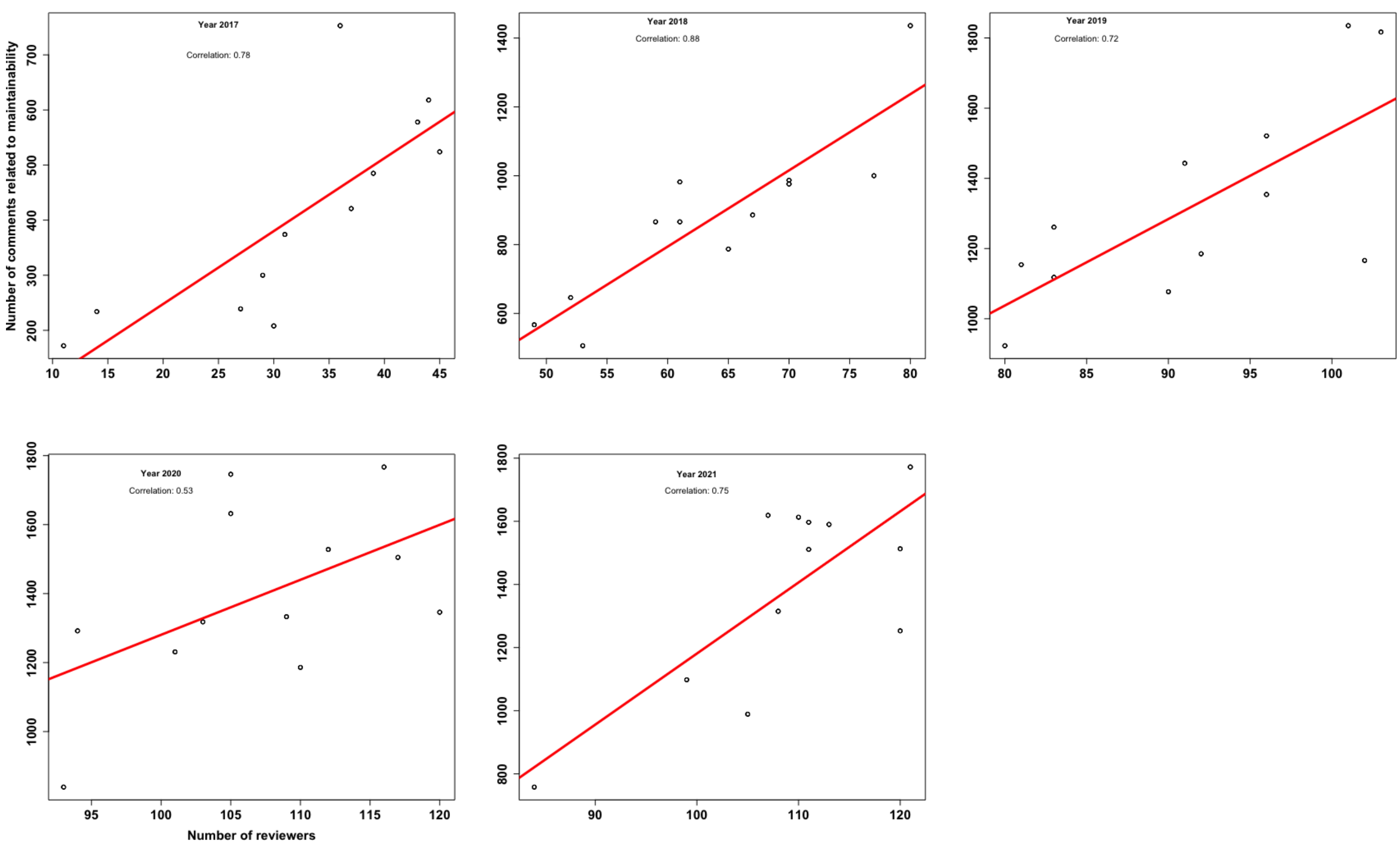


Fig. 6. The relationship between the number of reviewers and comments related to maintainability of the Eclipse project

The data were further analyzed to determine the trend in the number of reviewers. Fig. 8 shows that the number of reviewers who provide comments related to maintainability has increased annually. As observed in the Eclipse and Qt projects, the new generation of software developers is interested in developing high-quality software and providing comments that improve

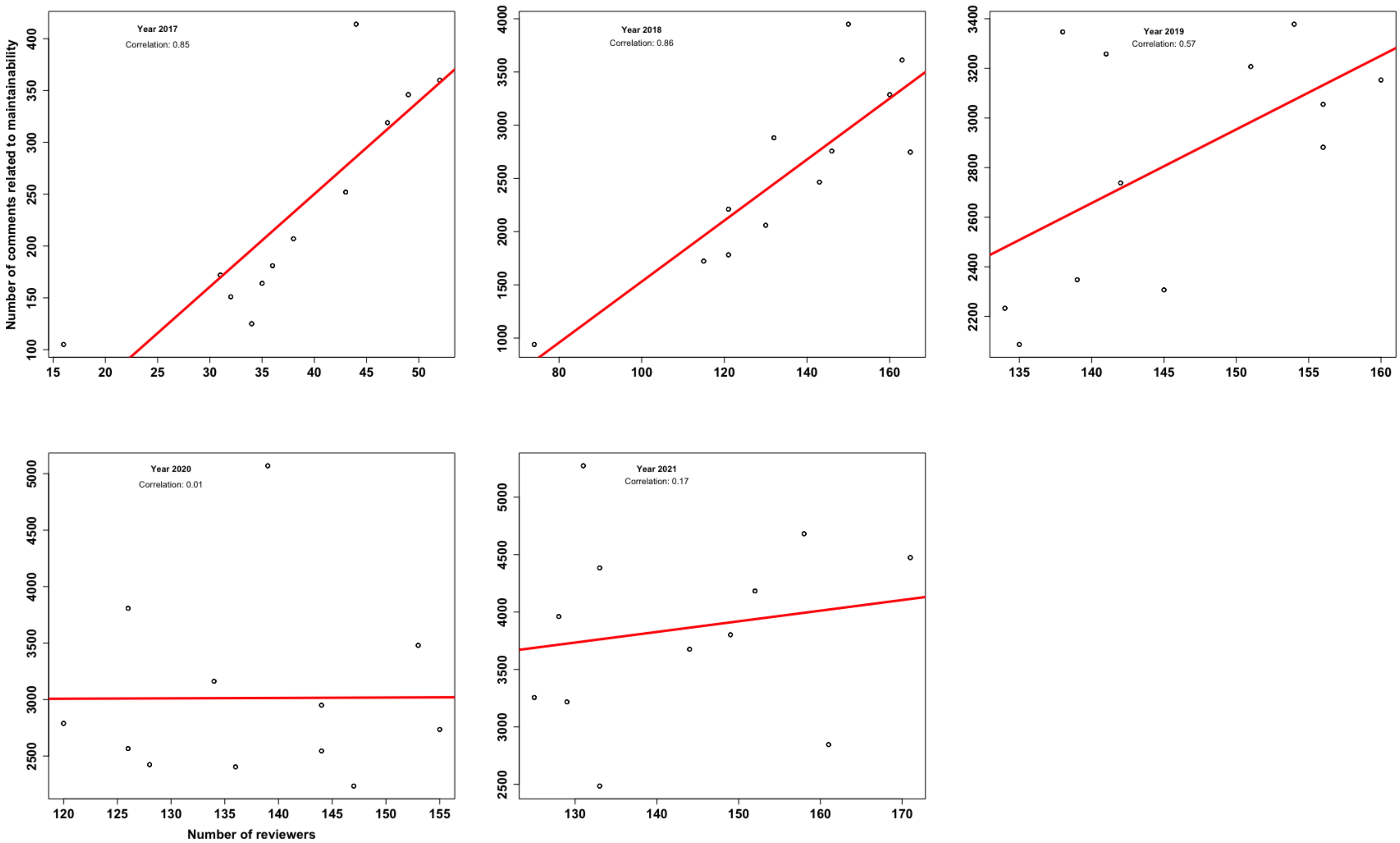


Fig. 7. The relationship between the number of reviewers and comments related to maintainability of the Qt project

the quality of OSS projects.

E. How Many Comments Related to the Maintainability are Given by Code Reviewers each Day?

The working times of reviewers for each project were evaluated based on ANOVA and posthoc comparisons considering Tukey's honest significant difference (HSD), which compares the average of multiple data pairs. Tukey's HSD was used because hypothesis testing with ANOVA considers multiple averages. Therefore, the results do not indicate which data pair is different. Moreover, Tukey's HSD is a popular statistical test in software engineering research. In this study, the results of the posthoc Tukey's HSD indicate how many comments were made by reviewers on each day of the week on average.

R was used to perform ANOVA at a 95% confidence level for the following hypotheses.

H0: The average number of comments related to software maintainability was the same on each day of the week.

H1: The average number of comments related to software maintainability differed on each day of the week.

The results indicate that Eclipse has a p-value of 0.000649 and Qt has a p-value of 0.0105. The p-values of both projects are lower than 0.05, so H0 is rejected. Therefore, it can be concluded that the average number of comments related to maintainability provided by reviewers each day varies. To assess the differences, Tukey's HSD is calculated.

R was used to perform the variance for Tukey's HSD calculation at a confidence level of 95%, and the hypotheses of the test are as follows.

H0: The average number of comments related to software maintainability is the same each day.

H1: The average number of comments related to software maintainability is different each day.

Based on the variance of Tukey's HSD, the average number of comments from Monday to Saturday does not vary for Eclipse. While there is a significant difference between the comments related to the maintainability on Sunday compared to those made throughout the rest of the week, the difference is not significant. Moreover, the values on Saturday and Sunday are similar. This finding may indicate that the Eclipse code reviewers usually review code and provide comments related to software maintainability from Monday to Saturday rather than on Sunday. Based on the variance of Tukey's HSD for Qt, there are no significant differences between the number of relevant comments on weekdays and weekends. Therefore, it can be concluded that the reviewers of Qt usually review code and provide comments related to software maintainability every day.

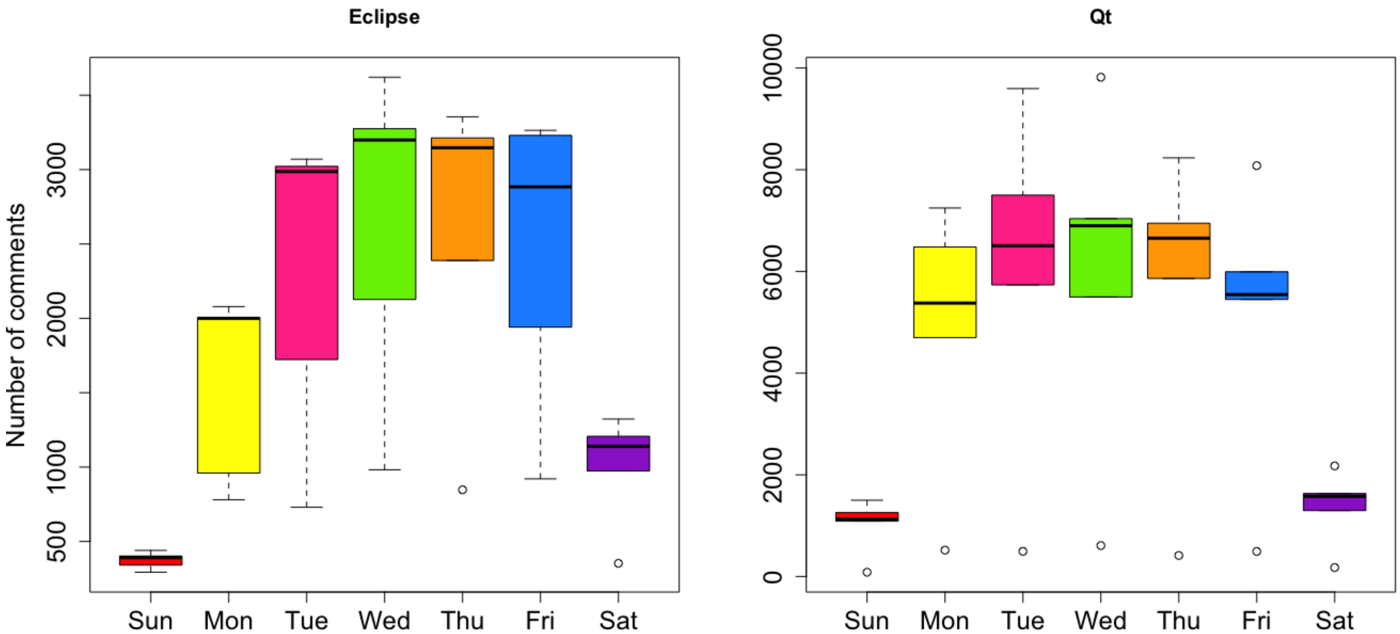


Fig. 8. Boxplot of comments related to the maintainability per day of the week

V. Discussion and Threats to Validity

This section discusses the results and threats to the validity of this study.

A. Discussion

The comments related to software maintainability from the reviewers of OSS projects account for approximately 30% of all comments, and some comments are related to the development of the main functions and subfunctions of the system and to other characteristics, such as stability [32]. In addition to maintainability, stability is a very important characteristic because users around the world want to be sure that OSS can maintain their confidentiality and integrity and prevent accessibility issues or allow software editing without permission. Moreover, published software must be made available. Therefore, software developers who focus on stability mainly try to find possible gaps in code or vulnerable code changes. The results of this study indicate that most comments led to modifications of the code, which is a good sign for software quality. Therefore, it is predicted that the software developer communities of the Eclipse and Qt OSS projects will continue to focus on maintainability in the future.

The results of this study and the applied methodology are of interest in assessing software maintainability. This study differs from other studies in that software engineering researchers have typically investigated Eclipse and Qt based on other maintainability topics and methods, such as using indicators to verify source code and check maintainability. For example, Yamashita et al. [33] checked for files that were vulnerable to size increases and analyzed the complexity of the software. Additionally, Caglayan et al. [34] used indicators to predict defects in OSS projects, and Counsell et al. [35] calculated the maintainability index based on coupled factors. Therefore, the results of this study related to "software maintainability" may differ from the results of other studies based on various factors, such as the methods, processes, tools, and environments of OSS projects.

B. Threats to Validity

Threats to validity can reduce the accuracy and reliability of studies and lead to inaccurate results. This section presents the threats to the validity of the current study. The threats in this study can be categorized into three classes as follows.

Construct Validity - This research compiled OSS comments from the code review system called "Gerrit". The storage structure of Gerrit was studied to design and develop searching and storage software. The collected data were sets of text or characters, and some data, such as symbols and figures, may be incomplete. However, the collected data were considered acceptable for analyzing the quality of OSS projects, especially maintainability, which was the objective of the research. In the research process, R was used for text mining, data processing, and data analysis, as well as to search for comments related to software maintainability. Moreover, R was used to conduct various statistical methods, such as variance analysis, correlation analysis, and regression analysis. Therefore, other programs or tools were used for text mining, and the results may be different.

Internal Validity - This study presents the comments related to software maintainability from the code review processes of Eclipse and Qt. In the comment search process, keywords and synonyms related to each sub-characteristic of software maintainability were selected to provide the support framework for searching. Only comments with one or more of the selected keywords or synonyms were analyzed, and irrelevant comments or those that did not meet specific criteria were removed. Result verification was performed by manually reviewing all comments related to software maintainability before processing. However, this manual may have an inherent bias, and we attempted to minimize this bias by repeating the verification process many times. Moreover, expert code reviewers were asked to verify the results by reading the comments (approximately 30% of all comments). Currently, there is no tool available for the automatic review of comments.

External Validity - The results of this study only provide empirical evidence for the two OSS projects considered. Therefore, the findings are not applicable to every OSS project because each project differs in internal structure, and the code review and software development processes also vary. However, the results of this study can provide preliminary guidelines for other studies related to the quality of similar OSS projects.

VI. Conclusion

The objective of this research was to analyze the comments associated with the Eclipse and Qt projects as part of the code review process. The analytical results suggest that the number of changes based on maintainability is moderate. This result can improve the awareness of software developers regarding the importance of code modification related to maintainability. Moreover, the empirical data in this research can benefit source code improvements made by software developer communities in OSS projects. Additionally, the code review trends illustrated in this study can improve decision-making processes among software developers of OSS projects, who should prioritize software maintainability. In addition, these developers should be ready to perform code editing in situations or environments that may change in the future.

Although the results of this research provide only basic conclusions, the existing information is sufficient for guiding the development and improvement of OSS quality. Moreover, the findings can be used by researchers who are interested in software maintainability to study related topics, such as plug-in development, using Gerrit. This comment-related approach could improve code modification and extend beyond maintainability to other topics, such as readability and testability. Additionally, tools or processes could be developed to assess the edits of software developers in OSS projects according to reviewer comments in Gerrit. If such a script or program was developed to automatically check code edits in the Gerrit framework, the reviewers who provide the comments could be notified to facilitate better communication within the software development community and allow for faster subsequent reviews.

References

[1] P. C. Rigby, D. M. German, and M. A. Storey, "Open source software peer review practices: a case study of the apache server," in *Proceedings of the 30th international conference on Software engineering*. ACM, 2008, pp. 541–550.

[2] M. Aberdour, "Achieving quality in open-source software," *IEEE Software*, vol. 24, no. 1, pp. 58–64, 2007.

[3] V. Tiwari and R. Pandey, "Open source software and reliability metrics," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 1, no. 10, pp. 808–815, 2012.

[4] B. Norick, J. Krohn, E. Howard, B. Welna, and C. Izurieta, "Effects of the number of developers on code quality in open source software: a case study," in *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 2010, pp. 1–1.

[5] Y. Zhou and J. Davis, "Open source software reliability model: an empirical approach," in *ACM SIGSOFT Software Engineering Notes*, vol. 30, no. 4. ACM, 2005, pp. 1–6.

[6] G. S. Walia and J. C. Carver, "Using error information to improve software quality," in *Proceedings of the 2013 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. IEEE, 2013, pp. 107–107.

[7] A. Bosu and J. C. Carver, "Peer code review in open source communities using reviewboard," in *Proceedings of the ACM 4th annual workshop on Evaluation and usability of programming languages and tools*. ACM, 2012, pp. 17–24.

[8] M. B. Zanjani, H. Kagdi, and C. Bird, "Automatically recommending peer reviewers in modern code review," *IEEE Transactions on Software Engineering*, vol. 42, no. 6, pp. 530–543, 2016.

[9] R. A. Baker Jr, "Code reviews enhance software quality," in *Proceedings of the 19th international conference on Software engineering*. ACM, 1997, pp. 570–571.

[10] A. Bosu, M. Greiler, and C. Bird, "Characteristics of useful code reviews: An empirical study at microsoft," in *Proceedings of the 12th Working Conference on Mining Software Repositories*. IEEE, 2015, pp. 146–156.

[11] M. Bernhart, A. Mauczka, and T. Grechenig, "Adopting code reviews for agile software development," in *Proceedings of the 2010 Agile Conference (AGILE 2010)*. IEEE, 2010, pp. 44–47.

[12] M. Fagan, "Design and code inspections to reduce errors in program development," in *Software pioneers*. Springer, 2002, pp. 575–607.

[13] P. Thongtanunam, C. Tantithamthavorn, R. G. Kula, N. Yoshida, H. Iida, and K. Matsumoto, "Who should review my code? a file location-based code-reviewer recommendation approach for modern code review," in *Proceedings of the 2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*. IEEE, 2015, pp. 141–150.

[14] International Organization For Standardization, "Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – System and software quality models," *ISO/IEC 25010:2011*, 2011.

[15] S. Ghosh and A. K. Rana, "Comparative study of the factors that affect maintainability," *International Journal on Computer Science and Engineering*, vol. 3, no. 12, pp. 3763–3769, 2011.

[16] N. S. A. A. Bakar and N. Arsat, "Investigating the factors that influence the quality of open source systems," in *Proceedings of the 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*. IEEE, 2014, pp. 1–6.

[17] J. Walden, M. Doyle, G. A. Welch, and M. Whelan, "Security of open source web applications," in *Proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2009, pp. 545–553.

[18] B. Norick, J. Krohn, E. Howard, B. Welna, and C. Izurieta, "Effects of the number of developers on code quality in open source software: a case study," in *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, 2010, pp. 62–62.

[19] D. C. Schmidt and A. Porter, "Leveraging open-source communities to improve the quality & performance of open-source software," in *Proceedings of the 1st Workshop on Open Source Software Engineering at ICSE 2001*, 2001, pp. 1–5.

[20] J. Asundi and R. Jayant, "Patch review processes in open source software development communities: A comparative case study," in *Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007*. IEEE, 2007, pp. 166–171.

[21] P. C. Rigby, "Understanding open source software peer review: Review processes, parameters and statistical models, and underlying behaviours and mechanisms," Ph.D. dissertation, 2011.

[22] A. Bacchelli and C. Bird, "Expectations, outcomes, and challenges of modern code review," in *Proceedings of the 2013 international conference on software engineering*. IEEE, 2013, pp. 712–721.

[23] Y. Tao, D. Han, and S. Kim, "Writing acceptable patches: An empirical study of open source project patches," in *Proceedings of the 2014 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2014, pp. 271–280.

[24] J. Czerwonka, M. Greiler, and J. Tilford, "Code reviews do not find bugs: how the current code review best practice slows us down," in *Proceedings of the 37th International Conference on Software Engineering*. IEEE, 2015, pp. 27–28.

[25] R. Glass, "Frequently forgotten fundamental facts about software engineering," *IEEE Software*, vol. 18, no. 3, pp. 112–111, 2001.

[26] B. C. Wagey, B. Hendradjaya, and M. S. Mardiyanto, "A proposal of software maintainability model using code smell measurement," in *Proceedings of the 2015 International Conference on Data and Software Engineering (ICoDSE)*. IEEE, 2015, pp. 25–30.

[27] I. Kádár, P. Hegedűs, R. Ferenc, and T. Gyimóthy, "A manually validated code refactoring dataset and its assessment regarding software maintainability," in *Proceedings of the 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, 2016, pp. 1–4.

[28] K. Stroggylos and D. Spinellis, "Refactoring–does it improve software quality?" in *Proceedings of the 5th International Workshop on Software Quality (WoSQ'07: ICSE Workshops 2007)*. IEEE, 2007, pp. 10–10.

[29] S. Rizvi and R. A. Khan, "Maintainability estimation model for object-oriented software in design phase (MEMOOD)," *Journal of Computing*, vol. 2, no. 4, pp. 26–32, 2010.

[30] S. A. Khan and R. A. Khan, "Analyzability quantification model of object oriented design," *Procedia Technology*, vol. 4, pp. 536–542, 2012.

[31] S. Ghosh and A. K. Rana, "Comparative study of the factors that affect maintainability," *International Journal on Computer Science and Engineering*, vol. 3, no. 12, p. 3763, 2011.

[32] A. Bosu, J. C. Carver, M. Hafiz, P. Hilley, and D. Janni, "Identifying the characteristics of vulnerable code changes: An empirical study," in *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, 2014, pp. 257–268.

[33] K. Yamashita, C. Huang, M. Nagappan, Y. Kamei, A. Mockus, A. E. Hassan, and N. Ubayashi, "Thresholds for size and complexity metrics: A case study from the perspective of defect density," in *Proceedings of the 2016 IEEE International Conference on Software Quality, Reliability and Security (QRS)*. IEEE, 2016, pp. 191–201.

[34] B. Caglayan, A. Bener, and S. Koch, "Merits of using repository metrics in defect prediction for open source projects," in *Proceedings of the 2009 ICSE Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development*. IEEE, 2009, pp. 31–36.

[35] S. Counsell, X. Liu, S. Eldh, R. Tonelli, M. Marchesi, G. Concas, and A. Murgia, "Re-visiting the'maintainability index'metric from an object-oriented perspective," in *Proceedings of the 41st Euromicro Conference on Software Engineering and Advanced Applications*. IEEE, 2015, pp. 84–87.

Bi-LSTM Model to Recognize Human Activities in UAV Videos using Inflated I3D-ConvNet

Sireesha Gundu[1], Dr.Hussain Syed[2]
Research Scholar[1], Associate Professor[2]
School of Computer, Science and Engineering, VIT-AP University
Andhra Pradesh[1,2]

***Abstract*—Human activity recognition in aerial videos is an emerging research area. In this paper, an Inflated I3D-ConvNet (Inflated I3D) and Bidirectional Long Short-Term Memory (Bi-LSTM) based human action recognition model in UAV videos have been proposed. The initial module was pre-trained using the Kinetics-400 video dataset, which consisted of 400 classes of human activities and around 400 video clips for each class culled from real-world and arduous YouTube videos. The proposed inflated I3D-ConvNet which was built on 2D-ConvNet inflation learns and extracts spatio-temporal features from aerial video while leveraging the architectural design of Inception-V1. The proposed model employs Bi-LSTM architecture for human action classification on the Drone-Action dataset which is a smaller benchmark UAV-captured video dataset. This model considerably improves the state-of-the-art results in activity classification using the SoftMax classifier and retains an accuracy of about 98.4%.**

Keywords*—*2D-ConvNet; Bi-LSTM; drone-action; inception-V1; inflated I3D-ConvNet; Kinetics-400

I. Introduction

Drone technology has advanced rapidly in recent decades. Unmanned aerial vehicles (UAVs), or drones, are proving very useful in areas that humans cannot reach or cannot reach quickly and effectively. Drone technology has much greater potential, scope, and size when used in a variety of settings, including businesses, governmental institutions, military installations, etc. They take little effort, energy, or time and can travel to even the most remote locations where fewer people are needed.

In the next generation of drones, the key characteristics of size, autonomy play, and propulsion are crucial. The type of drone is determined by the technologies used to operate it. Multirotor is the most prevalent type of drone extensively used by most of the professionals. Video surveillance, aerial photography, etc. are some of the multirotor drone applications. Multirotor is easy to manufacture and most economic to fly. With a multirotor, you can position and frame the camera more precisely for good aerial shots. The most popular and commonly used multirotor aircraft are quad-copters. Multi-rotors have many drawbacks such as limited speed, reduced endurance, and less flying time. With light cargo, the maximum flight time for a multirotor is 20 to 30 minutes. Due to this, multirotors are not appropriate for applications involving long-distance travel, inspection, or mapping.

Understanding how a human body is articulated in an image and identifying human movement in videos taken by these types of drones [1] is a challenging research problem. Sometimes in many situations such as those including visual blur, perspective distortion, low-resolution scenes, occlusions, etc., it is highly challenging to discern human motions.

With the swift advancement of deep learning algorithms and neural network architectures, numerous great accomplishments have been made in image recognition tasks on various datasets [2], [3], [4]. The CNN model pre-trained on ImageNet [5] has already achieved significant performance in image recognition. The same methods can be adopted in the existing action recognition mainstream models for extracting temporal and spatial features from frames of RGB video and optical flow subsequently. Although great video-based action recognition performance can be attained, our research work discovered that pre-train on still image dataset ImageNet is not an optimal preference. Since still images differ from video sequences to a significant extent, the novel dataset known as Kinetics for human action recognition [6] was released. This dataset is appropriate for pre-training the architectures for video recognition. The outcome of the experiment [7] shows that every convolutional neural network architecture pre-trained on this dataset outperforms the architectures pre-trained on ImageNet. In action recognition, an important factor along with image recognition is sequence modeling. In sequence modeling problems, compared with RNN (Recurrent Neural Network) and unidirectional LSTM [8], bidirectional LSTM [9], [10], [11] is broadly used architecture because it can address the most resilient problems such as vanishing gradient and exploding gradient up to some extent.

The majority of existing techniques suffer from a high false alarm rate. Furthermore, while these techniques perform well on simple datasets, their performance is restricted when dealing with real-world scenarios. To address these issues, we present in this paper a robust and efficient model that learns visual features from a sequence of frames by integrating them as spatiotemporal information from raw video. The following are the key contributions of current work:

- For drone-based outdoor surveillance networks, we propose a reliable ConvNet-based sophisticated paradigm. For deep spatiotemporal feature extraction, we use a pre-trained inflated inception-V1 architecture, followed by a sequential learning method that outperforms state-of-the-art approaches in terms of accuracy.
- The proposed framework employs a multi-layer Bi-LSTM architecture. Because of the forward and backward passes applied at each layer of the LSTM model, this improves the capabilities of effective learning. As

a result, the final trained model is not limited to the data used for training, but it can also be used in video surveillance networks.

- In this paper, we proposed a novel network that combines the advantage of a sophisticated ConvNet feature extractor pre-trained on Kinetics and robust sequence modeling tools such as Bidirectional LSTM. Initially, like I3D Inception 3D-CNN [12], [13], [14] pre-trained on Kinetics is exercised as the feature extractor. Next, to capture high-level temporal features, the feature vectors created by I3D are input into a bidirectional LSTM network. Finally, to create useful predictions of these high-level features, a SoftMax classifier is used. We can train the network on an end-to-end basis.

The rest of the paper is structured into sections as follows: Section II presents the overview of the existing models for action recognition, Section III gives the problem definition, Section IV describes the proposed methodology, Section V reports the datasets adopted for the proposed work, experimental results, performance evaluation, comparison with the state-of-the-art models and discussion. Section VI concludes the proposed work. Section VII provides the future work. Finally, Section VII gives the acknowledgments.

II. Literature Survey

A. 3D-ConvNet for spatio-temporal feature extraction

Since 3D-CNN was first introduced for action recognition tasks, it has always been a prevalent research method. Using this method, spatial and temporal data are directly extracted from raw surveillance video. C3D [15] is one more predominant research effort. Using a simple linear classifier on four main-stream benchmarks, C3D can outperform other algorithms by learning spatial-temporal properties from 3D-CNN. Recently, large-size datasets such as Kinetics and ActivityNet [16] have been introduced. After pre-training on a large video dataset, a few other works incorporating a 3D-CNN model outperform leading-edge models. Before 3D-CNN came into existence, due to its ability to directly extract spatial characteristics using 2D convolution, which is more effective and exact than creating features by hand, 2D-CNN models totally rule the image recognition domain.

The previous research study says that for the action recognition task, very deep 2D-CNN models can be outperformed by very shallow and Kinetics pre-trained 3D-CNN models to a great extent. A novel model called inflated I3D-ConvNet has been proposed that enhances kernels of inception module [17] such as convolution and pooling in GoogleNet [18] into 3D. In the UCF-101 dataset [19], this model has achieved maximum accuracy.

B. Bi-LSTM for Action Recognition

In sequential modeling tasks like human action recognition in videos, RNN is a very powerful and widely used network architecture. LSTM (Long Short-Term Memory) is an RNN-based network that is extensively used in video-based action recognition for learning motion features. This also can be used to prevent gradient explosion problems and gradient vanishing problems up to some extent during the training process. Another research work [20] has suggested a video as an ordered sequence based on LSTM architecture. In this, the underlying CNN network generates the output features which are fed to the LSTM network yielding an exceptional performance in the UCF-101 dataset. Another model LRCN [21] is very broad in both spatial & temporal dimensions. In this, we utilized CNN for spatial feature learning and LSTM for temporal feature learning. Also, the LRCN model has different input and output lengths.

Bidirectional RNNs are initially addressed in [22], later for speech recognition tasks, the work [23] has been proposed by using the bidirectional RNN to outperform the unidirectional RNN. Next, bidirectional LSTMs were introduced to predict frame-wise phoneme classification [24], network-wide traffic speed [25], etc. In terms of prediction bidirectional LSTMs give good results than unidirectional LSTMs. In various tasks involving videos like object segmentation in video [26], video-super resolution [27], fine-grained action detection [28] and spatio-temporal feature learning for gesture recognition [29] bidirectional LSTM architecture has been leveraged. Another research work [30] extends the idea of coupling the convolutional module with the RNN module. This architecture leverages the element-wise max pooling and BiConvLSTM and includes the temporal encoding in forward temporal directions and backward temporal directions as well because in most heterogeneous datasets accessing the forthcoming details from the present state is very much advantageous for accurate predictions.

CNNs are capable of extracting local features whereas Bi-LSTMs are very good at handling long-term dependencies. Unlike LSTMs (which utilize only past information), when complete time-series sequence data is at hand Bi-LSTMs make use of both past and future information which permits the network to make more accurate predictions. In sequential data, to capture the different temporal local dependencies, this model employs different sizes of convolutional kernels. Datasets such as WISDM, UCI-HAR, and PAMAP2 are adopted to calculate the performance of the proposed multiple human activities identification model. Merits and demerits of the existing research works are described in the below Table I.

It is possible to reach a conclusion by utilising some of the constraints and potential improvements learned from previous research. Existing approaches were insufficiently effective. This is primarily due to the algorithm's complexity and potential. Our proposed approach is used to fill gaps in the existing scientific literature. For this, we introduced a novel model named I3D-BiLSTM that employs inflated I3D-CNN pre-trained on the Kinetics dataset and bidirectional LSTM architectures for learning spatio-temporal features of adjacent video frames. This approach achieved considerably highest performance in terms of accuracy compared to existing works.

III. Problem Definition

Human activity recognition in a fixed camera-based surveillance system is relatively one of the well-studied areas of interest whereas human activity recognition in UAV-based surveillance systems is comparatively less studied [31], [32], [33]. This area of research has drawn a great deal of attention

TABLE I. MERITS AND DEMERITS OF RELATED WORKS

Method	Dataset	Merits	Demerits
C3D [15]	UCF-101	Effective approach for spatiotemporal feature learning using deep 3D ConvNets	Achieved less accuracy
ActivityNet [16]	ActivityNet	Provides largescale video benchmark for human activity understanding	Needs more categories and samples per category to unveil new challenges in understanding and recognizing human activities in present scenarios
Inception-v3 [17]	ILSVR 2012	Enhances the kernels of inception module such as convolutional and pooling in GoogleNet into 3D	Lack of verification of ensemble result on the test set as the test and validation set error tends to correlate very well
GoogleNet [18]	ILSVRC14	Improved utilization of the computing resources inside the network	Improvement needed in classification of groudtruth predictions
UCF-101 [19]	UCF-101	Large benchmark dataset consists of unconstrained videos	Less number of classes and categories, below baseline action recognition results
CNN+LSTM [20]	Sports-1M+ UCF-101	Yeilds good performance using CNN and LSTM for video classification	Deeper integration of the temporal sequence information into the CNNs is required
LRCN [21]	Flickr30k+ COCO2014	Achieved good results in CNN special learning and LSTM temporal feature learning with different input and output lengths	Needs more investigation on perceptual problems with time-varying visual input or sequential outputs
BRNN [22]	TIMIT	Efficient estimation of the conditional posterior probability of complete symbol sequences without making any explicit assumption about the shape of the distribution	Need more investigation on classification
DBLSTM-HMM hybrid [23]	TIMIT speech	Outperforms both RNNs and unidirectional LSTM	Experiments in LSTM learning algorithms, output error functions, and improved generalisation is required
Frame-wise Phoneme classification [24]	TIMIT speech	Outperforms both RNNs and unidirectional LSTM	Experiments in LSTM learning algorithms, output error functions, and improved generalisation is required
SBU-LSTM [25]	loop detector data+INRIX data	Best prediction performance for the traffic speed on different types of traffic network in both accuracy and robustness	Requires more attention in learning and interpreting spatial features towards graph-based structure
ConvGRU+ConvLSTM [26]	DAVIS+ FBMS+ SegTrack-v2+ FT3D	Improved encoding techniques used in spatio-temporal evolution of objects in a video for motion segmentation	More research is required on promising algorithms like Instance-level video object segmentation
BRCN for multi-frame SR [27]	25 YUV format videos+Set1+Set2	Powerful temporal dependency modeling for SR videos containing complex motions, achieved better performance and faster speed	1) The model has a very shallow architecture (i.e., only 3 layers), so its performance is poor 2) Due to the low-resolution images the computational complexity grows with the upsampled spatial size 3) Due to the lack of large-scale video SR dataset, model cannot directly learned from scratch (raw videos)
Multi-stream CNN +Bidirectional LSTM [28]	MPII Cooking 2+new MERL Shopping	Effective use of pixel trajectories rather than stacked optical flow as input to the motion streams, leading to a significant improvement in results	Extraction of pertinent data is required to increase more accuracy
MoIWLD [29]	Hockey Fight +BE-HAVE +Crowd Violence	Modified sparse-representation-based classification model is used to control the reconstruction error of coding coefficients and minimize the classification error for violence detection in videos	Requires more training to improve the accuracy
BiConvLSTM [30]	Hockey Fights+Movies+Violent Flows	Leverages the element-wise max pooling and BiConvLSTM, includes the temporal encoding in forward temporal directions and backward temporal directions for the detection of violence in videos	Further investigation is needed for the best results on more dynamic and heterogeneous datasets

in recent years as drones are becoming more popular and extremely beneficial in different applications such as military, agriculture, search and rescue, security monitoring systems, etc [34]. Human activity recognition in aerial videos is a challenging task because of various factors such as environmental background, camera motion, object distances and variety of cameras, longitude, the latitude of the UAV flight, etc. [35], [36], [37]. Another challenging task is human activity recognition in a video (sequence of frames) where each frame has different variations such as spatio-temporal features loss, background clutter, occlusion, low inter-class and high intra-class variances for some classes, background lighting change, image distortion, and a person pose variation [38], [39], [40], [41]. We need to develop a sophisticated model that addresses these issues and accurately classifies the human activities in UAV captured video datasets.

IV. PROPOSED METHODOLOGY

A. 3D-ConvNet

In the realm of computer vision, Convolutional Neural Network (CNN/ConvNet) is one of the well-known deep learning neural networks that can handle massive volumes of data in the early 1980s. It is a feed-forward neural network with multiple layers that agglomerates many convolutional layers called grouping layers, hidden layers, and activation layers on top of each other in a particular order. This sequential design allows ConvNet to learn hierarchical attributes using the multi-layers of artificial neurons. An artificial neuron or perceptron is the function that evaluates the weighted sum of one or more inputs and passes this value to the non-linear function called the activation function which produces the output as an activation value.

Initially, 2D-ConvNet was pre-trained on ImageNet for image recognition. When an image is passed as an input to the 2D-ConvNet, it generates different activation functions that are directed to the next layers. During the convolution, a kernel/filter of fixed size should be applied to the input image to obtain convolved features. The initial layer extracts primitive information such as diagonal or horizontal edges. The next layer takes hold of the output from the previous layer and extracts more intricated features such as amalgamated edges or corners. Hence, the deeper layers of the network can extract even more complex features such as faces, objects, etc. The terminal convolutional layer generates the activation map. Based on this, the confidence score (between 0 and 1) will be produced by the classification layer to classify the image. 2D-ConvNet architectures pre-trained on ImageNet and other datasets failed to recognize the objects in the images or videos which are having different angles and different illumination conditions. Later on, Carreira and Zisserman introduced a new 3D-ConvNet architecture pre-trained on Kinetics for human action recognition to extract spatio-temporal features in videos.

In this work, we proposed a novel architecture shown in Fig. 1 where each short video clip classification is treated as an action recognition problem. For this subject, 3D-CNN has traditionally been a popular strategy for facilitating spatio-temporal learning. A substantial amount of training data is often required to train a 3D-CNN from scratch. Carreira and Zisserman introduced a new model that inflates a pre-trained 2D-CNN on ImageNet architecture along the temporal dimension to produce an inflated I3D-ConvNet. They used the Kinetics human action dataset to train the inflated I3D-ConvNet to address the action recognition challenge. Based on their work, we exercised Inception-v1 I3D for our challenge and fine-tune it on our Drone-Action dataset [42]. In this paper, we have chosen the pre-trained weights of the RGB stream as the initial weights are publicly accessible for Inception-v1 I3D.

In this model architecture, we manifested the evolution of 3D-ConvNet from ImageNet-based 2D-ConvNet architecture. Like standard conventional networks, 3D-ConvNet is a natural video modeling approach with spatio-temporal filters explored previously in [43], [44], [45]. Inflating 2D-ConvNets into 3D implies converting successful image classification models such as deep 2D-ConvNets into 3D-ConvNets by inflating all the filters and pooling kernels of size 'n × n' into 'n × n × n' with a supplementary temporal dimension. One important characteristic of these models is that they precisely generate hierarchical illustrations of the spatio-temporal data. Compared to 2D-ConvNet, it has many parameters due to the added kernel dimension, which makes training very difficult but results have proven that they have produced promising results for evaluation on the larger benchmark dataset (e.g. Kinetics). Fig. 2 depicts the feature extraction process for an inflated I3D-ConvNet architecture.

We used a modest modification of C3D for this research, which contains 14 convolutional layers, 27 BatchNormalization layers, 27 ReLU layers, 13 DepthwiseConv layers, 4 ZeroPadding layers, one GlobalAveragePooling layer, and A top layer with 4 fully connected layers. As in the original version, short clips of 244 × 244 pixels each with a crop of 112 × 112 pixels make up the model's inputs. After all convolutions and completely connected layers, we used batch normalization. Another variation of the original model is, that instead of using a temporal stride of 1 stride 2 is used. This moderates the memory usage and permits greater batch sizes. When fully connected layers without weight bindings have been constructed, batch normalization is very crucial at this stage.

B. LSTM

Long Short-Term Memory (LSTM) deep learning architecture was first developed by Hochreiter and Schmidhuber in 1997. It is a gated Recurrent Neural Network (RNN) architecture. RNN network works on the present input by considering the previous output as feedback and stores the information in its memory for future prediction. Unlike RNN, it is capable of handling long-term dependencies. LSTM networks address the following issues that RNN fails to solve:

- Preserving the information for a long period or handling long-term dependencies to predict the current output.
- Fine control over carrying forward the important information and forgetting the unnecessary past information.
- Removal of exploding and vanishing gradient problem which occurs during the network's training phase, backtracking is used to reduce loss and achieve the best outcomes.

C. LSTM Architecture

LSTM works very similar to RNN but it has a key feature that differs it from RNN that is it preserves information for a long period for future cell processing. Three gates make up an LSTM cell: Forget gate, Input gate, and Output gate. The internal process of an LSTM cell is depicted in the following Fig. 3:

It has a memory pool that has two key vectors of states.

1) Short-term state: It is also known as a hidden state. This state keeps the output at the current time step. The short-term state of the previous timestamp is represented with S_{t-1} and the current timestamp is S_t.
2) Long-term state (L_{t-1}): It is also known as the cell state. This state reads and rejects the information while passing through the network which is meant for long-term storage. The long-term state of the previous timestamp is represented with L_{t-1} and the current timestamp is L_t.

Here, the cell state contains all timestamps and information as well. As depicted in the figure, the decision of reading, writing, and storing depends on the activation functions whose outputs are in between (0, 1).

1) Forget gate: This is the first state in the cell of the LSTM network. This gate decides whether to store the information of the previous timestamp or forget it. The Forget gate Eq. (1) is as follows:

$$F_t = \sigma(C_t * U_f + S_{t-1} * W_f) \tag{1}$$

Where,

C_t = Current timestamp t input,
U_f = Weight connected with the input,
S_{t-1} = Previous timestamp's short-term state or hidden state,
W_f = The weight matrix for the short-term state.

Next, the activation function i.e. the $sigmoid$ function is applied to it which gives the value of f_t in between (0, 1). It is then multiplied by the previous timestamp's long-term state, as shown in the calculations below by using Eq. (2) and (3).

$$L_{t-1} * f_t = 0, \qquad if \quad f_t = 0 \tag{2}$$

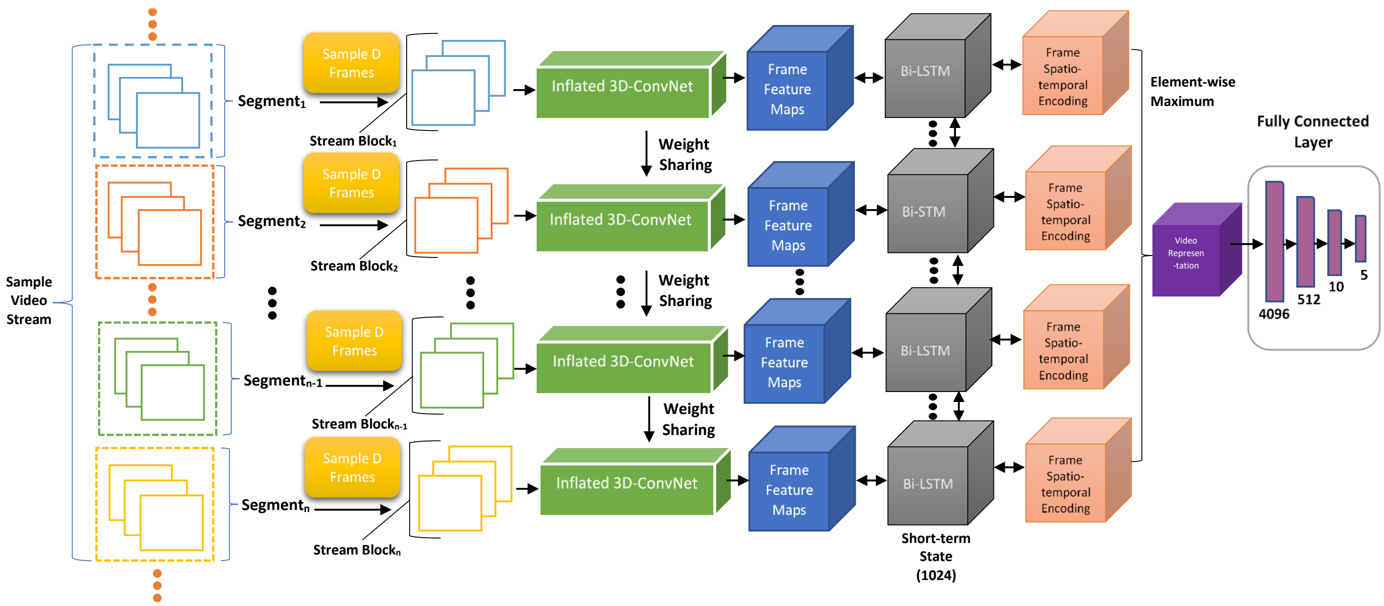


Fig. 1. Framework of the proposed Inflated I3D-ConvNet-BiLSTM for action recognition

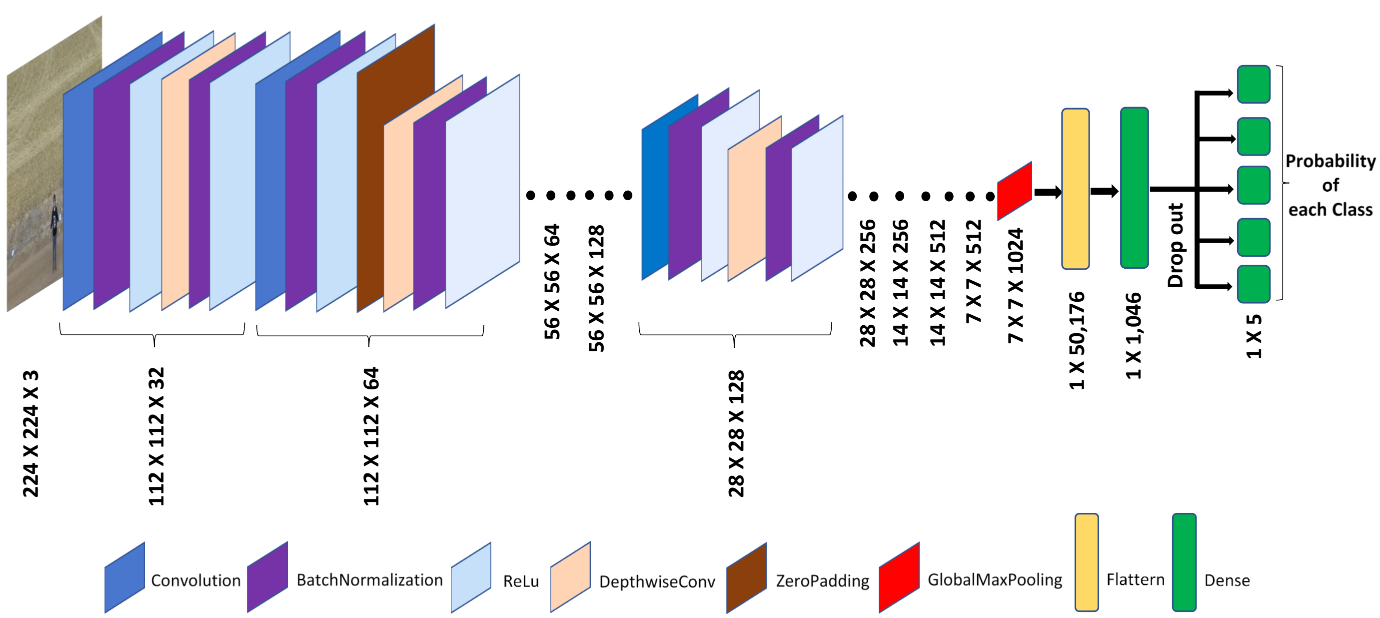


Fig. 2. The feature extraction workflow of Inflated I3D-ConvNet architecture

$$L_{t-1} * f_t = L_{t-1}, \qquad if \quad f_t = 1 \tag{3}$$

If the f_t value is 0 it forgets everything, otherwise, it forgets nothing.

2) Input gate: This is employed to manage the input value flow and quantifies the importance of the latest information in the cell. The Input gate Eq. (4) is as follows:

$$I_t = \sigma(C_t * U_i + S_{t-1} * W_i) \tag{4}$$

Where,

C_t = Current timestamp t input,
U_i = Matrix of weights associated with the input,
S_{t-1} = Previous timestamp's short-term state or hidden state,
W_i = The weight matrix for the short-term state.

Next, the activation function is then subjected to the *sigmoid* function, resulting in the value of 'I' at timestamp 't'. The value lies between (0, 1).

In LSTM, the latest information also known as new infor-

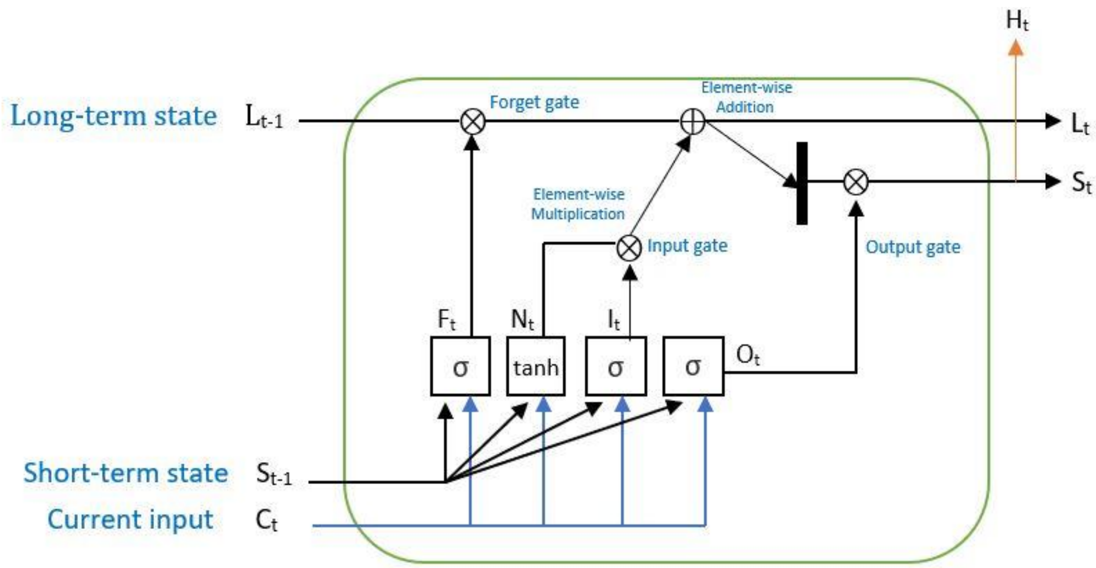


Fig. 3. Long short-term memory architecture

mation in Input Gate is used to update the long-term state.

Latest Information (or New Information):

$$N_t = tanh(C_t * U_c + S_{t-1} * W_c) \tag{5}$$

This latest information shown in Eq. (5) is a function of the short-term state at the previous timestamp 't-1' & input 'C' at the timestamp 't'. This information is necessary to pass it through the long-term state. After the activation function, $tanh$ is applied over it, therefore the value of the latest information lies between (-1, 1).

If N_t is negative, this information is subtracted from the long-term state; if N_t is negative, this information is added to the long-term state at the current timestamp. The following Eq. (6) is an updated equation to add N_t to the long-term state.

$$L_t = F_t * L_{t-1} + I_t * N_t \tag{6}$$

Where L_{t-1} represents the long-term state at the current timestamp and others represent previously determined values.

3) Output gate: This is used to control the cell utilization for calculating the output activation of the LSTM unit and is used to determine the generation of the output from the current internal long-term state to the next short-term state. The Output gate Eq. (7) is as follows:

$$O_t = \sigma(C_t * U_o + S_{t-1} * W_o) \tag{7}$$

This equation is similar to Forget gate and Input gate. The output value of this equation lies between 0 and 1 when the $sigmoid$ activation function is applied to it.

Next, the following Eq. (8) will be used to compute the current short-term state O_t and $tanh$ of the updated long-term state:

$$H_t = O_t * tanh(L_t) \tag{8}$$

That is, the short-term state is the function of present output and $tanh$ of the long-term state. Then apply the SoftMax activation function to the short-term state L_t to get the output of the present timestamp using the following Eq. (9):

$$Output = SoftMax(S_t) \tag{9}$$

In this, the token with the highest score in the output is the prediction.

D. Bidirectional LSTM

Bidirectional LSTM or Bi-LSTM is an extension of the LSTM model. Bi-LSTMs, unlike baseline LSTMs (trains a model in unidirectional i.e., forward direction and use only past information), train the model in two directions called forward direction and backward direction as depicted in the following Fig. 4.

In the forward direction, the model learns a sequence of inputs from past to future whereas in the backward direction model learns from future to past when the complete sequence of time-series data is available. This denotes that the calculation of the output frame at timestamp 't' depends on the preceding frame at time 't-1' and the subsequent frame at time 't+1' because it performs processing in both directions.

This approach uses two hidden states, one for the forward pass and another for the backward pass, to preserve the past and future information. To enable the network to make more accurate predictions, these states should be combined and this mechanism is called merging. This can be achieved through the functions called sum, averaging, multiplication, and concatenation. Among these functions concatenation is the default mechanism.

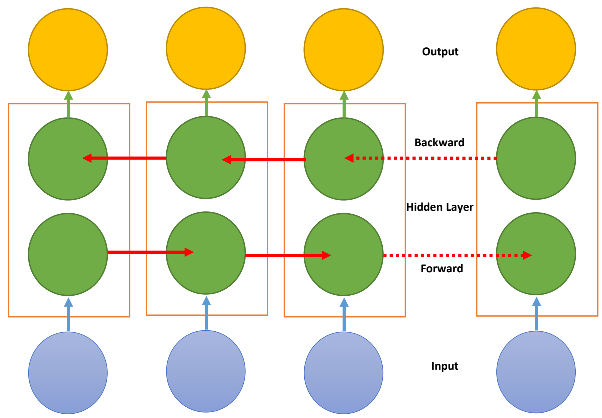


Fig. 4. Bidirectional LSTM architecture

V. Experimental Results and Discussion

First, we briefly described the datasets that are utilized to evaluate the models in this segment. Secondly, we provided the classification outcomes after implementing the Inflated I3D-ConvNet model with Bi-LSTM for recognizing human actions. In the performance evaluation section, charts of model accuracy and loss are provided. We finally reported the comparison results of existing models with the proposed model in the following section. All these experiments are performed on the anaconda platform.

A. Overview of the Datasets

1) Kinetics-400: The below Table II presents an overview of the Kinetics-400 dataset.

TABLE II. Summary of Kinetics-400 Dataset

Feature	**Value**
Actions	400
Clips per class	min 400
Total clips	306, 245
Repetitions per clips	3-5
Each clip length	10 Sec

In this paper, we used a human action video dataset called Kinetics-400. It contains 400 human action classes, with at least 400 video clips per action, each clip lasting 10 seconds, for a total of 306,245 videos. Each clip in this dataset is retrieved by first searching YouTube for suggestions and then using AMT (Amazon Mechanical Turkers) to determine if the clip contains an action. At least three or more out of five confirmations were needed to accept the clip. Next, we ran de-duplication on the dataset and verified that the clips were retrieved from each video and that the clips did not share common video footage. Finally, the noise reduction and overlap of the class were reviewed.

Rather than activities or events, the dataset is focused on a broad range of classes and human actions. Some of the human actions are drinking, drawing, pumping fist, laughing, etc., human-human actions such as shaking hands, hugging, kissing, etc., as well as human-object actions such as washing dishes, opening a present, mowing the lawn, etc.

Some actions in the Kinetics-400 dataset are fine-grained and require emphasis on an object to categorize, e.g., playing various kinds of wind instruments. Some other actions require temporal reasoning to categorize e.g., various kinds of swimming. There are also different Parent-Child groupings like Personal Hygiene (cutting nails, washing hands, Brushing teeth, . . .), Music (trombone, playing drums, violin, . . .), Cooking (peeling, frying, cooking, . . .), Dancing (macarena, ballet, tap, . . .), etc.

2) Drone-Action: The dataset consists of 13 actions, each action having 10-20 clips, a total of 240 clips, and 66919 frames. All videos in the dataset were delivered in HD format (1920 $\times$ 1080 resolution) and recorded at 25 fps, with an average duration of 11.15 seconds for each action. Midway through a wheat field, actions were recorded on an unsettled road while the hover and follow modes were active. The actions were divided into three categories called following, side-view, and front-view based on the camera movement and the viewpoint. The video recording camera was 10 MP, A 3DR SOLO rotor-craft served as the unmanned aerial vehicle, and the camera was a GoPro HERO 4 Black with 5.4 mm, IR CUT anti-fish eye replacement, and a 3-axis Solo gimbal. The flight ended between 8 and 12 metres low altitude and was slow. The following Table III provides an overview of the Drone-Action dataset.

TABLE III. Summary of Drone-Action dataset

Feature	**Value**
Actions	13
Actors	min 10
Clips	240
Clips per class	10-20
Repetitions per class	5-10
Mean clip length	11.15 Sec
Total duration	44.6 min
Frames	66919
Frame rate	25 fps
Resolution	1920 $\times$ 1080
Camera motion	Yes (hover and follow)
Annotation	Bounding box

In this research work, out of 13 action classes, only 5 classes [clapping, waving hands, walking (front, back, left, right), jogging (front, back, left, right), running (front, back, left, right)] with a minimum of 20 video clips for each action class and an average duration of 11.15 seconds for each action were utilized for classification of human actions.

B. Results

Initially, to make your I3D-ConvNet network a strong feature extractor, pre-train it with the Kinetics-400 video dataset. Referring to the RGB I3D model [12] and using the open-source I3D model, we attempt to emulate the results to achieve

97.6% accuracy on the Drone-Action dataset with RGB modality. With reference to the previous work, to represent high-level temporal features, we integrated the Bi-LSTM network with I3D and used the Drone-Action dataset to validate the proposed model and attempt to obtain an accuracy of 98.4% which outperforms the other mainstream models and also the original I3D model.

Below Table IV provides instances of sequences that correlate to various types of actions and scenarios in the Drone-Action dataset and Table V shows some of the sample results with a percentage of accuracy after validating the model using the Drone-Action dataset.

C. Performance Evaluation

The evaluation measure we have chosen for this experiment was accuracy. The action classifier returns scores based on this the accuracy was evaluated. The Drone-Action dataset was divided into train-to-test split sets. Each train-to-test split set was randomly generated at a ratio of 70:30.

The two measures which we considered to assess the proposed model's accuracy trained on Kinetics-400 are as follows:

- Firstly, freeze the network weights and utilize it to generate features for the Drone-Action video dataset, and then train the SoftMax classifier for the Drone-Action dataset classes that use train data to evaluate a test set.
- Secondly, fine-tune the network for classes of the Drone-Action dataset by utilizing its training data and re-evaluating its test sets.

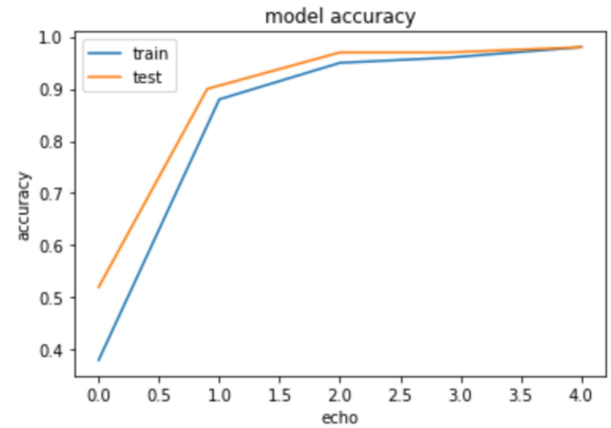
This clearly says that the proposed inflated I3D-ConvNet architecture with Bi-LSTM perks from pre-training on the Kinetics-400 dataset and also after pre-training on the Kinetics dataset training only the model's last layers yield much better performance. The proposed model's model accuracy and model loss per epoch is depicted in the Fig. 5.

1) Comparison with classical models: We compare the performance of I3D models to earlier classical approaches in Table VI, on HMDB-51 & UCF-101 datasets and the bar chart of the various model's average accuracies is shown in the following Fig. 6. We provide the findings while pre-training on miniKinetics and the complete Kinetics dataset.

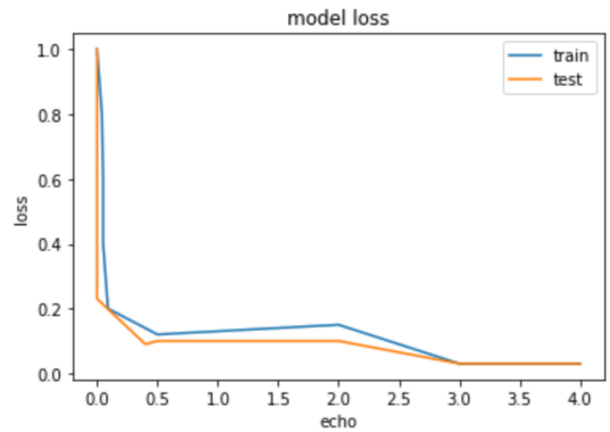
D. Discussion

All activities were correctly recognised with a high classification performance using the proposed approach. The bidirectional LSTM ability to use past and future inferences in the signal allows for accurate differentiation of the dynamic activities like clapping, jogging, wavinghands, running, walking.

The results of this work are comparable to the best related works using the same datasets. In contrast to earlier 3D-ConvNets (C3D) models, Kinetics' pre-trained I3D models



(a) Model accuracy



(b) Model loss

Fig. 5. Performance of the proposed model

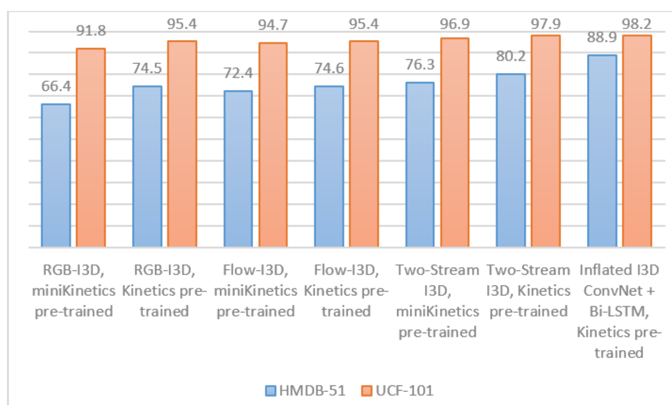


Fig. 6. Bar chart of the average accuracies of the various models

demonstrate a far higher difference. This could be due to Kinetics higher quality, but it could also be due to I3D's superior architecture. The I3D RGB stream on HMDB-51 improves after switching from miniKinetics to Kinetics pre-training, indicating that 3D-ConvNets may permit a massive volumes of information to understand robust motion characteristics. The two streams achieve equal performance after Kinetics pre-

TABLE IV. DRONE-ACTION DATABASE: INSTANCES OF SEQUENCES CORRESPONDING TO VARIOUS TYPES OF ACTIONS AND SCENARIOS

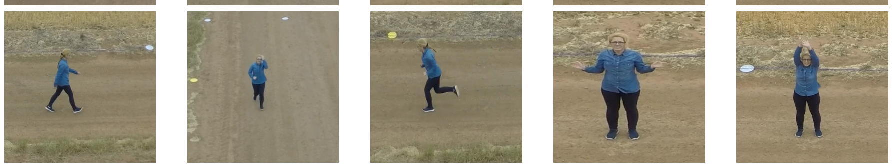
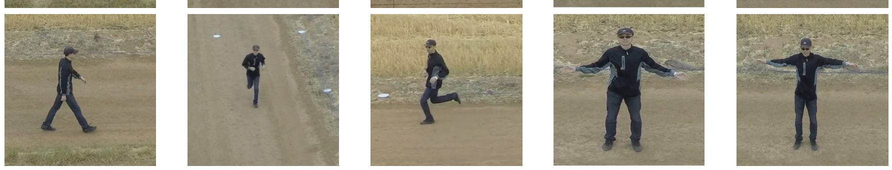
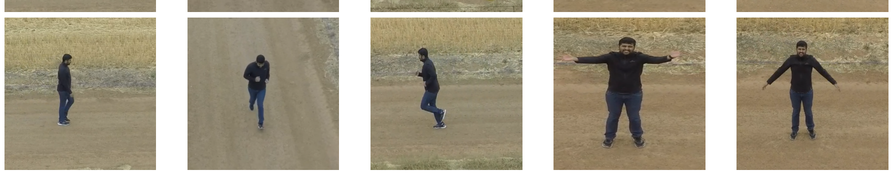
Actor	walking	jogging	running	clapping	waving hands
Actor 1					
Actor 2					
Actor 3					
Actor 4					
Actor 5					
Actor 6					
Actor 7					
Actor 8					
Actor 9					
Actor 10					

TABLE V. THE ACTION CLASSES OF THE DRONE-ACTION DATASET

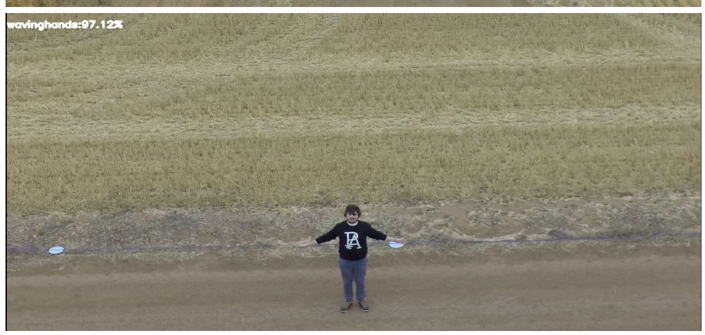
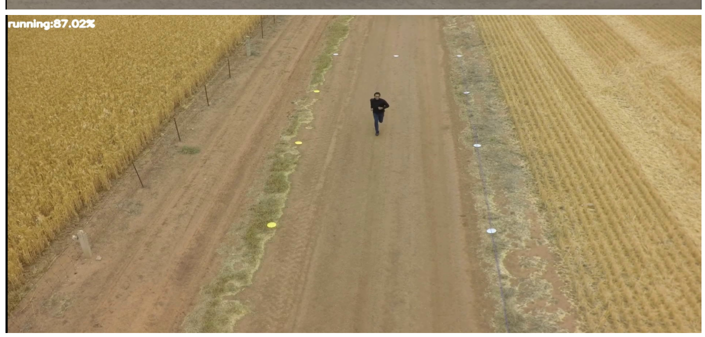
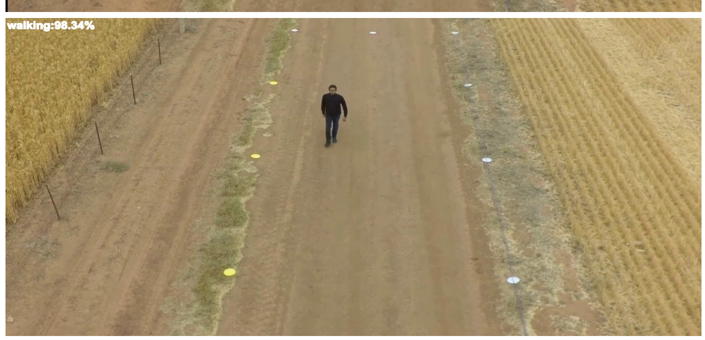
Action	Result
clapping	
jogging	
waving hands	
running	
walking	

TABLE VI. Comparison with Classical Models Averaged over Three Train/Test Splits on HMDB-51 & UCF-101 Datasets

Model	HMDB-51	UCF-101
RGB-I3D, miniKinetics pre-trained	66.4%	91.8%
RGB-I3D, Kinetics pre-trained	74.5%	95.4%
Flow-I3D, miniKinetics pre-trained	72.4%	94.7%
Flow-I3D, Kinetics pre-trained	74.6%	95.4%
Two-Stream I3D, miniKinetics pre-trained	76.3%	96.9%
Two-Stream I3D, Kinetics pre-trained	80.2%	97.9%
Inflated I3D-ConvNet + Bi-LSTM, Kinetics pre-trained	88.9%	98.2%

training, but they're still good together. When their predictions are averaged, the results range from 74.6% to 80.2%.

The average accuracy over the three common train/test splits was used to compare our techniques. When pre-trained on Kinetics, a model from our RGB-I3D or RGB-Flow lineup outperforms any model combination or any previously published performance by any model. Our proposed model significantly improves on previous models, bringing overall performance to 88.9% on HMDB-51 and 98.2% on UCF-101 respectively.

VI. Conclusion

In this paper, we introduced a new architecture known as inflated I3D-ConvNet with Bi-LSTM pre-trained on the Kinetics-400 video dataset. First, on the Kinetics-400 dataset, we initially model the low-level features of successive frames using a pre-trained I3D model, and then we learn high-level special features using a Bi-LSTM network. The leading performance of inflated I3D-ConvNet pre-trained on the Kinetics-400 video dataset also corroborates that this novel model is more advanced and efficient than other mainstream models. The proposed model can serve as an alternative solution to the data loss and long-term dependency problems that arise as the training data size increases. This architecture uses various convolutional filter sizes by capturing different local dependencies to enhance feature extraction and also can able to classify simple activities like waving hands, walking, clapping, jogging, and running with outstanding accuracy. The suggested model uses the inflated process as a domain adaptation job for 3D-ConvNets, and the input structure of split video segments outperforms the standard sequential input structure.

VII. Future Work

In the proposed model we utilized Kinetics-400 large-scale and high-quality video dataset which includes 400 human action classes for pre-training the architecture. As a future plan, we plan to rerun all the experiments using other benchmark datasets such as Kinetics-600 and Kinetics-700 that cover 600 & 700 human action classes consequently with and without pre-training, explore inflating operation on 3D-ConvNets and also examine the efficacy of implementing our inflated I3D-ConvNets into two and three-stream architectures.

Acknowledgments

We would like to extend our profound gratitude and sincere appreciation to everyone on the Kinetics dataset project, especially Tim Green and Brain Zhang for creating data that can be used for experiments and also we would like to express our profound gratitude to Asanka G.Perera, Yee Wei Law and Javaan Chahl for sharing `Drone-Action` dataset for our experiments.

References

[1] V. Krassanakis, M. P. Da Silva, and V. Ricordel, "Monitoring human visual behavior during the observation of unmanned aerial vehicles (Uavs) videos," *Drones*, vol. 2, no. 4, pp. 1–19, dec 2018.

[2] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 633–659, 2013.

[3] S. Abu-El-Haija, N. Kothari, J. Lee, P. Natsev, G. Toderici, B. Varadarajan, and S. Vijayanarasimhan, "Youtube-8m: A large-scale video classification benchmark," *CoRR*, vol. abs/1609.08675, 2016. [Online]. Available: http://arxiv.org/abs/1609.08675

[4] S. Gundu, H. Syed, and J. Harikiran, "Human detection in aerial images using deep learning techniques," in *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2022, pp. 1–10.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[6] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev *et al.*, "The kinetics human action video dataset," *arXiv preprint arXiv:1705.06950*, 2017.

[7] K. Hara, H. Kataoka, and Y. Satoh, "Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?" *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 6546–6555, 2018.

[8] R. L. Abduljabbar, H. Dia, and P. W. Tsai, "Unidirectional and bidirectional LSTM models for short-term traffic prediction," *Journal of Advanced Transportation*, vol. 2021, 2021.

[9] A. Mihanpour, M. J. Rashti, and S. E. Alavi, "Human action recognition in video using db-lstm and resnet," in *2020 6th International Conference on Web Research (ICWR)*. IEEE, 2020, pp. 133–138.

[10] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM with CNN Features," *IEEE Access*, vol. 6, pp. 1155–1166, nov 2017.

[11] S. S. Saha, S. S. Sandha, and M. Srivastava, "Deep convolutional bidirectional lstm for complex activity recognition with missing data," in *Human Activity Recognition Challenge*, 2021, vol. 199, pp. 39–53.

[12] J. Carreira and A. Zisserman, "Quo Vadis, action recognition? A new model and the kinetics dataset," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 4724–4733, 2017.

[13] Y. Huang, Y. Guo, and C. Gao, "Efficient Parallel Inflated 3D Convolution Architecture for Action Recognition," *IEEE Access*, vol. 8, pp. 45 753–45 765, 2020.

[14] X. Wang, Z. Miao, R. Zhang, and S. Hao, "I3d-lstm: A new model for human action recognition," in *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 3. IOP Publishing, 2019, p. 032035.

[15] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3d convolutional networks,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[16] F. Caba Heilbron, V. Escorcia, B. Ghanem, and J. Carlos Niebles, “Activitynet: A large-scale video benchmark for human activity understanding,” in *Proceedings of the ieee conference on computer vision and pattern recognition*, 2015, pp. 961–970.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, pp. 1–9, 2015.

[19] K. Soomro, A. Roshan Zamir, and M. Shah, “UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild,” Tech. Rep., 2012. [Online]. Available: http://crcv.ucf.edu/data/UCF101.php

[20] J. Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, “Beyond short snippets: Deep networks for video classification,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, pp. 4694–4702, 2015.

[21] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, 2017.

[22] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[23] A. Graves, N. Jaitly, and A.-r. Mohamed, “Hybrid speech recognition with deep bidirectional lstm,” in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, 2013, pp. 273–278.

[24] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM networks,” *Proceedings of the International Joint Conference on Neural Networks*, vol. 4, pp. 2047–2052, 2005.

[25] Z. Cui, R. Ke, Z. Pu, and Y. Wang, “Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction,” pp. 1–11, 2018. [Online]. Available: http://arxiv.org/abs/1801.02143

[26] P. Tokmakov, K. Alahari, and C. Schmid, “Learning Video Object Segmentation with Visual Memory,” *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 4491–4500, 2017.

[27] Y. Huang, W. Wang, and L. Wang, “Video Super-Resolution via Bidirectional Recurrent Convolutional Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1015–1028, 2018.

[28] B. Singh, T. K. Marks, M. Jones, O. Tuzel, and M. Shao, “A Multi-stream Bi-directional Recurrent Neural Network for Fine-Grained Action Detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 1961–1970, 2016.

[29] T. Zhang, W. Jia, X. He, and J. Yang, “Discriminative Dictionary Learning with Motion Weber Local Descriptor for Violence Detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 696–709, 2017.

[30] A. Hanson, K. Pnvr, S. Krishnagopal, and L. Davis, “Bidirectional convolutional lstm for the detection of violence in videos,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

[31] S. Zaghbani and M. S. Bouhlel, “Mask rcnn for human motion and actions recognition,” in *International Conference on Soft Computing and Pattern Recognition*. Springer, 2020, pp. 1–9.

[32] M. Ding, N. Li, Z. Song, R. Zhang, X. Zhang, and H. Zhou, “A lightweight action recognition method for unmanned-aerial-vehicle video,” in *2020 IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE)*. IEEE, 2020, pp. 181–185.

[33] W. Sultani and M. Shah, “Human action recognition in drone videos using a few aerial training examples,” *Computer Vision and Image Understanding*, vol. 206, p. 103186, 2021.

[34] B. Mishra, D. Garg, P. Narang, and V. Mishra, “Drone-surveillance for search and rescue in natural disaster,” *Computer Communications*, vol. 156, pp. 1–10, apr 2020.

[35] M. Teutsch and W. Krüger, “Detection, segmentation, and tracking of moving objects in UAV videos,” *Proceedings - 2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012*, pp. 313–318, 2012.

[36] M. Barekatain, M. Martí, H.-F. Shih, S. Murray, K. Nakayama, Y. Matsuo, and H. Prendinger, “Okutama-action: An aerial view video dataset for concurrent human action detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 28–35.

[37] A. G. Perera, Y. W. Law, T. T. Ogunwa, and J. Chahl, “A Multiviewpoint Outdoor Dataset for Human Action Recognition,” *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 5, pp. 405–413, oct 2020.

[38] K. V. V. Subash, M. V. Srinu, M. Siddhartha, N. S. Harsha, and P. Akkala, “Object detection using ryze tello drone with help of mask-rcnn,” in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2020, pp. 484–490.

[39] M. Cristani, R. Raghavendra, A. Del Bue, and V. Murino, “Human behavior analysis in video surveillance: A Social Signal Processing perspective,” *Neurocomputing*, vol. 100, pp. 86–97, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2011.12.038

[40] Z. Cheng, L. Qin, Q. Huang, S. Yan, and Q. Tian, “Recognizing human group action by layered model with multiple cues,” *Neurocomputing*, vol. 136, pp. 124–135, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2014.01.019

[41] L. Shao, L. Ji, Y. Liu, and J. Zhang, “Human action segmentation and recognition via motion and shape analysis,” *Pattern Recognition Letters*, vol. 33, no. 4, pp. 438–445, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2011.05.015

[42] A. G. Perera, Y. W. Law, and J. Chahl, “Drone-action: An outdoor recorded drone video dataset for action recognition,” *Drones*, vol. 3, no. 4, pp. 1–16, dec 2019.

[43] S. Ji, W. Xu, M. Yang, and K. Yu, “3d convolutional neural networks for human action recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012.

[44] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[45] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler, “Convolutional learning of spatio-temporal features,” in *European conference on computer vision*. Springer, 2010, pp. 140–153.

Aspect-based Sentiment Analysis for Bengali Text using Bidirectional Encoder Representations from Transformers (BERT)

Moythry Manir Samia[1*], Alimul Rajee[2*], Md. Rakib Hasan[3], Mohammad Omar Faruq[4], Pintu Chandra Paul[5]
Dept. of Information and Communication Technology, Comilla University, Cumilla, Bangladesh[1,2,3,5]
Dept. of Accounting and Information Systems, Comilla University, Cumilla, Bangladesh[4]

***Abstract*—Public opinion is important for decision-making on numerous occasions for national growth in democratic countries like Bangladesh, the USA, and India. Sentiment analysis is a technique used to determine the polarity of opinions expressed in a text. The more complex stage of sentiment analysis is known as Aspect-Based Sentiment Analysis (ABSA), where it is possible to ascertain both the actual topics being discussed by the speakers as well as the polarity of each opinion. Nowadays, people leave comments on a variety of websites, including social networking sites, online news sources, and even YouTube video comment sections, on a wide range of topics. ABSA can play a significant role in utilizing these comments for a variety of objectives, including academic, commercial, and socioeconomic development. In English and many other popular European languages, there are many datasets for ABSA, but the Bengali language has very few of them. As a result, ABSA research on Bengali is relatively rare. In this paper, we present a Bengali dataset that has been manually annotated with five aspects and their corresponding sentiment. A baseline evaluation was also carried out using the Bidirectional Encoder Representations from Transformers (BERT) model, with 97% aspect detection accuracy and 77% sentiment classification accuracy. For aspect detection, the F1-score was 0.97 and for sentiment classification, it was 0.77.**

Keywords*—Sentiment analysis; Bengali sentiment analysis; Aspect-Based Sentiment Analysis (ABSA); Bengali ABSA; deep learning; Bidirectional Encoder Representations from Transformers (BERT)

I. Introduction

Online communication has grown extremely prevalent all around the world due to the internet. Social media is used by about 51% of the world's population overall. They use social media on a daily basis for an average of 2.5 hours [1]. Social media platforms, for example, Twitter and Facebook have made it simple to convey comments. Online newspapers have replaced traditional newspapers, and readers now even share their opinions about many news stories in the news portal's comment section. People express their perspectives and ideas on a wide range of topics, including politics, business, and international issues. As a result, a significant volume of data is created daily and contains many insightful opinions on a variety of subjects on those online sites. For many widespread applications, it's vital to comprehend opinions from user comments. For instance, the quality of goods or services can be improved by studying client feedback from comments on e-commerce sites. Also, People frequently voice their opinions on many national issues in democratic nations like Bangladesh, India, USA on social media or news websites, which must be taken into consideration for a country's development. Studying all the text manually and trying to understand what the individuals are feeling is difficult. So, all of this data must be automatically studied to provide useful information as it is quite beneficial to understand the viewpoints of the general public on a particular topic. ABSA is a popular method for obtaining such useful data.

Sentiment analysis, often referred to as opinion mining, is a technique for determining if a given expression is negative, positive, or indifferent. There are three layers of sentiment analysis to consider, the first one to evaluate is the document level, the second is the sentence level, and the third is the level for aspect. At document level, the text is analyzed to see if it has a favorable or unfavorable sentiment. The polarity of every sentence is determined by sentence level. These two layers of analysis don't indicate what people liked and didn't like in particular. This shortcoming of earlier layers is resolved by the third layer. Sentiment analysis is known as ABSA on the aspect label. The main objective of ABSA is to find out the correlation of a piece of text with a set of aspects and deduce their sentimental polarities in the process. The most comprehensive version for document-level SA to retrieve key information is ABSA. ABSA consists of two key tasks. 1) Extricating the specific aspects addressed in the text. 2) For each aspect, identify the sentiment's polarity [2]. For example, "The fabric has decent quality, but the price is excessive." This evaluation focuses on two factors: 'price' and 'fabric'. 'Positive' feeling is indicated with the 'price' aspect, whereas a 'negative' feeling is shown with the 'fabric' aspect. Aspect categories are stated specifically inside this review. However, reviews can be implicit at times. For instance, "I asked the waiter three times for tomato sauce, but he didn't bring any". The implicit aspect here is a restaurant's 'service'.

Bengali is spoken by around 226 million people worldwide, with a large proportion of them using the internet [3]. There are very few ABSA datasets available in Bengali language. Not many Bengali-based ABSA works are available, to the extent that we are aware, and the insufficiency of Bengali ABSA datasets is one of the key barriers to the development of Bengali ABSA. The present Bengali ABSA research is based on a variety of deep learning and machine learning strategies, for instance: Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN). BERT is well-known for resolving many shortcomings of these machine learning and deep learning techniques. In a self-supervised fashion BERT learns the language model by utilizing Transformer's encoder. But to the best of our knowledge, BERT was never used in any prior ABSA works for Bengali text.

As we previously discussed, there isn't much research done on ABSA in Bengali. We committed our work to the Bengali language by making the following contributions:

- For implementing ABSA in Bengali text, we collected heterogeneous Bengali data from news portals, YouTube channels and comments from social media.
- We annotated the collected dataset in five aspects: Technology, Corruption, Economy, Politics, and Sports, and three sentiment categories: Neutral, Positive, and Negative.
- We performed aspect detection and sentiment classification using our dataset by implementing BERT, which provided the best performance in terms of Bengali ABSA to the best of our knowledge.

The paper is organized as follows. The literature review is covered in Section II, and the research procedures for our experiment are described in Section III. We present and analyze the experimental results in Section IV. Finally, Section V provides the conclusion along with potential future directions.

II. Literature Review

A fast-growing and improving sub-field of Natural Language Processing (NLP) is sentiment analysis. Subramaniam *et al.* conducted a sentiment analysis-based survey in their research [4]. They studied the IMDB dataset and observed that sequence neural models had the best results, but they come at a high cost in terms of classifier complexity. By comparison, CNN had a preferable performance with less complexity over models of sequence neural. YouTube videos were used for a study on sentiment analysis by Alhujaili *et al.* in their research work [5]. The development of sentiment analysis is being accelerated by cutting-edge deep learning techniques like LSTM and BERT. Althobaiti suggested an automated technique for extracting coarse-grained hate speech and offensive language from Arabic tweets [6]. BERT was contrasted with two traditional machine learning methods SVM and Logistic Regression (LR). Along with the textual content of the tweets, the use of sentiments and text-based emoji descriptions was also looked at. They had an F1-score of 84.3% for offensive language detection, 81.8% for hate speech detection, and 45.1% for fine-grained hate-speech recognition by applying BERT. BERT was used by Nijhawan *et al.* for sentiment classification in their research [8]. Experiments revealed that, despite its comparatively simpler structure, the BERT model outperforms many prominent models in this subject.

The majority of Bengali sentiment analysis works are based solely on determining the text's polarity. Recurrent Neural Network (RNN) with Long Short-term Memory (LSTM) was used to analyze sentiment in Bengali data to avoid long-term reliance by Ahmed *et al.* [3]. Demonstration of the benefits of tuned hyper-parameter and how it can aid sentiment analysis on a dataset was given in their paper. An attention-based CNN algorithm was proposed by Basri *et al.* for analyzing sentiment on the Bang-lish Disclosure in their paper [10]. Islam *et al.* introduced two fresh datasets for sentiment analysis in Bengali in their research that was manually tagged [11]. $BERT_{BSA}$, which is a deep learning method of sentiment analysis for the Bengali language that surpassed all the previous models, was also introduced in their research.

ABSA was first discussed as a chapter in Liu, B.'s research paper. There were discussions of ABSA's methods and related issues [12]. The goal of SemEval2014's Task 4 was to advance the field of ABSA. Pontiki *et al.* developed and published benchmark datasets for ABSA with manually labeled reviews from the restaurant and laptop sectors [13]. Four subtasks, Aspect term extraction, Aspect category detection, Aspect term polarity, and Aspect category polarity—were used to analyze the 163 entries from 32 teams. Karimi *et al.* research on adversarial training in ABSA was discussed in their paper [14]. The studies with the proposed architecture show that using adversarial instances during network training enhances the efficiency of the common purpose BERT, while BERT which is in-domain and post- trained, brought better results for feature extraction and sentiment classification. Suciati *et al.* suggested a methodology to ensemble aspect extraction focused on Part-of-speech (POS) tagging, dependency parsing, and dense neural networks [7]. They found that deep learning integrated with conventional methods can yield better outcomes than lexicon-based approaches.The aspect-agnostic issue, which is pervasive in the context modeling of ABSA, was identified and discussed by Xing *et al.* [9]. They claimed that during context modeling, the semantics of the provided aspect should be taken into account as a new indication outside the context. In order to produce the aspect-aware masked states specifically designed for the ABSA work, they presented context encoders that are aspect-aware, such as Aspect-Aware BERTs (AABERTs), Aspect-Aware Graph Convolutional Networks (AAGCN), and Aspect-Aware LSTM (AALSTM).

Rahman *et al.* constructed two freely accessible datasets, 'Cricket' and 'Restaurant', to be used for the ABSA in Bangla, which function as a benchmark for the Bangla ABSA field [15]. It was the first ABSA work in Bengali. Customer reviews of restaurants make up one of the datasets, while manually labeled user comments on cricket make up the other. 2900 comments about cricket in five aspect categories are included in the 'Cricket' dataset, while 2600 restaurant reviews are included in the 'Restaurant' dataset. They designed an aspect category extraction model which is based on CNN architecture in the same year [2]. They used their suggested datasets [15] to compare their model with well-known machine learning techniques. For the 'Cricket' dataset, they achieved 51% F1-Score while KNN, SVM, and RF classification only achieved 35%, 34%, and 37% accuracy, accordingly. For the 'Restaurant' dataset, KNN, SVM, and RF classification achieved 42%, 38%, and 38% F1-Score, respectively, while CNN outperformed them with 64% F1-Score. In [16], Boidini implemented three stacked Auto-encoders (AEs) models based on the datasets from [15] to categorize aspects in the Bengali text. The stacking network's layers were individually trained to comprehend the encoded data of the layer that came before them. The three layers that were trained in a stacked manner were AE, Sparse AE (SAE), and Contractive AE (CAE). In comparison to the studies of Rahman *et al.* [2,15], all of the suggested models exhibit improved precision, F1 score, and recall. The CAE model especially generated the best F1-score on the 'Restaurant' and 'Cricket' datasets with 0.87 and 0.91 respectively. Based on the dataset provided by Rahman *et al.* [15], Haque *et al.* constructed the Bangla ABSA model [17]. The traditional supervised aspect classification machine learning approach was employed in this work with minimal preprocessing. They observed that if the data is less preprocessed, then the result has a higher F1 Score. To perform aspect extraction on the dataset provided by Rahman *et al.* [15], F.A. Naim presented a novel technique called PSPWA (Priority Sentence Part Weight Assignment) [1]. PSPWA is a phase of the data pre-processing process that leads to better performance. CNN and traditional supervised learning methods were utilized to show the performance. Compared to other learning algorithms, CNN performed best. CNN's F1-score for the 'Cricket' dataset was 0.59, and for the 'Restaurant' dataset it was 0.67.

The summary of this literature review is presented in Table I.

TABLE I. SUMMARY OF LITERATURE REVIEW

Author	Year	Dataset	Used Method	Result
Ahmed *et al.* [3]	2020	Prothom Alo, Cricket [15], etc.	LSTM	94%
Subramaniam *et al.* [4]	2021	IMDB dataset	RNN,CNN-LSTM, etc.	CNN-LSTM: 96%
Althobaiti [6]	2022	Data set from [20]	LR, SVM, BERT	Offensive language: 84.3%, Hate speech: 81.8%, Fine-grained hate speech: 45.1%
Suciati *et al.* [7]	2020	from PergiKuliner platform	SVM, LR, DT,ET	Food:GRU: 88.16%, BiLSTM: 88.16% Price:RF: 89.54%, Service: BiLSTM: 89.03% Ambience: BiLSTM: 84.78%
Nijhawan *et al.* [8]	2022	100042 tweets	RF,DT, LR,BERT	97.78%
Basri *et al.* [10]	2021	Dataset with 5000 short paragraphs	Attention -based CNN	Funny sentiment: 90% Multiclass data: 83%
Islam *et al.* [11]	2020	Prothom alo	Multi-lingual BERT	3-class: 60% 2-class: 71%
Pontiki *et al.* [13]	2014	Restaurant and laptop sectors reviews	CRF with features extracted	Laptop: 74.55% Restaurant: 84.01%
Karimi *et al.* [14]	2020	SemEval-2016, SemEval-2014	BERT	Laptop: 85.57% Restaurant: 81.50%
Bodini *et al.* [16]	2019	Cricket Restaurant [15]	AE, SAE, CAE	Cricket- CAE: 0.88 Restaurant- CAE: 0.87
Naim *et al.* [1]	2021	Cricket Restaurant [15]	CNN, RF, SVM etc.	Cricket-CNN: 0.59, RF: 0.41 Restaurant- CNN: 0.67, , RF: 0.35
Haque *et al.* [17]	2020	Cricket Restaurant [15]	SVM, RF, LR etc.	Cricket- SVM: 0.35, RF: 0.37 Restaurant- LR: 0.43, RF: 0.35
Rahman *et al.* [2]	2018	Cricket Restaurant [15]	CNN, RF, SVM etc.	Cricket-CNN: 0.51, RF: 0.37 Restaurant- CNN: 0.64 , RF: 0.38

III. METHODOLOGY

This work can be divided into two subtasks: determining a Bengali sentence's aspect and identifying sentiments. For the first subtask, five aspects can be defined: Technology, Corruption, Sports, Politics, and Economy. As for the second subtask, there will be three sentiment categories: Positive, Neutral, and Negative. BERT was used for both subtasks due to its exceptional features and efficiency. We began by collecting data, then labeled it and cleaned the dataset by removing stop words, punctuations, and decimal values. Following that, we preprocessed the dataset using tokenization and special token addition because BERT requires the data in a specific format. The training dataset was then fed into the BERT model, and the model was tested using the test dataset. Other deep learning models, such as CNN, Bidirectional LSTM (BiLSTM), LSTM, RNN and Gated Recurrent Unit (GRU) models, were also built, trained, and tested on our dataset to cross-validate the BERT model's results. Our model's performance is evaluated using a variety of performance metrics. The process as a whole is depicted in Fig. 1 below.

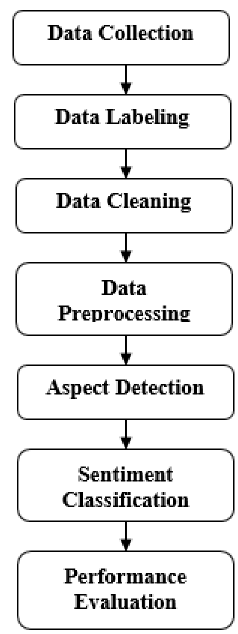


Fig. 1. Aspect-based sentiment analysis process

A. Data Collection

The dataset is a critical component of a machine learning algorithm since it determines the outcome based on its size and quality. Data was collected from social media, news portals, and Youtube comments of Bangladeshi users on five topics: Technology, Corruption, Sports, Politics, and Economy. The comments were stored in a Microsoft excel file.

B. Data Labeling

Some multi-lined sentences were simplified to single sentences by removing unnecessary sentences manually. Following that, 10042 sentences were left to be labeled. After that each author individually annotated the dataset. Each sentence was assigned to one of the five aspect categories—technology, corruption, sports, politics, and the economy—and was then annotated for sentiment polarity into three categories: positive, negative, and neutral. Each participant added an annotation to every comment. To figure out the ultimate aspect category and the polarity of a statement, voting percentage was calculated. In the final dataset, there are three columns for 10042 annotated sentences. One is for text or comments, the second one is for aspect and the final one is for sentiment category. Fig. 2 contains a sample of our dataset.

Text	Category	Polarity
রন্ধ্রে রন্ধ্রে জড়িয়ে পড়েছে যে দেশে দূর্নীতি ও সুদ ঘুষ কিছুতেই জেন লাগাম টানা যাচ্ছেনা.!!!	Corruption	Negative
ইলিশ রপ্তানি করে প্রচুর বৈদেশিক মূদ্রা অর্জন করা সম্ভব হচ্ছে	Economy	Positive
দুর্নীতি দমন কমিশন ক্ষমতাবানদের বিরুদ্ধেও স্বাধীন তদন্ত করবে এমনই আমাদের প্রত্যাশা	Corruption	Positive
গ্রামে আমরা টুজি নেটওয়ার্কই ঠিকমতো পাইনা, তারা আবার আসছে ফোরজি নিয়ে। যত্তসব।	Technology	Negative
আওয়ামীলীগ জনগণের জন্য কাজ করে বলেই তারা সফল।	Politics	Positive
আর কোন দিন বাংলাদেশ ক্রিকেট দল জয় এর মুখ দেখবে না যত দিন পাপন সভাপতি থাকবে।	Sports	Negative
কোন ব্যাংক থেকে এ ঋণ পাওয়া যাবে?	Economy	Neutral

Fig. 2. Sample of the labeled dataset

The Zipf's law was implemented in our dataset. According to Zipf's Law, there should be an inverse relationship between the frequency of various words in the dataset and their positions [18]. Fig. 3 demonstrates how Zipf's Law applies to our dataset. It suggests that the words in our dataset are not arbitrary.

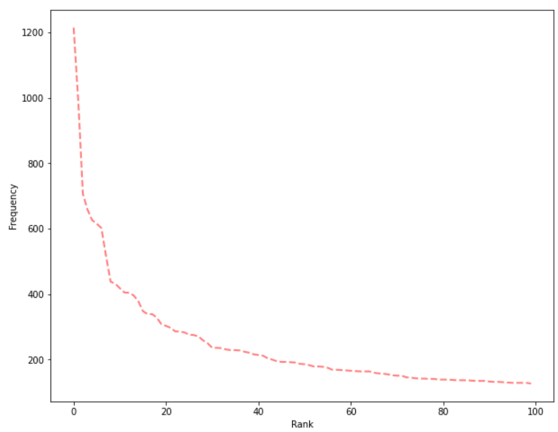


Fig. 3. Word frequency of our dataset using Zipf's law

Every effort was made for balancing the dataset, however for the chosen topics, people are more likely to comment on negative instances than positive ones, and comparatively less comments were without opinion. Because of this there are a few more negative comments with sentiment labeling and number of Neutral comments is comparatively less. As for the aspect labeling, all five aspects have nearly identical numbers of comments. Tables II and III display the dataset's statistics for aspect labeling and sentiment labeling, respectively.

TABLE II. Statistics of Aspect Labeling

Aspect category	No. of data
Technology	2003
Economy	2011
Corruption	2001
Sports	2021
Politics	2006
Total	**10042**

TABLE III. Statistics of Sentiment Polarity Labeling

Sentiment polarity	No. of data
Positive	3507
Negative	3909
Neutral	2626
Total	**10042**

C. Data Cleaning

NLP relies significantly on data preparation. To adapt the machine learning algorithm to the Bengali data, data cleansing is necessary. To clean the dataset, a few procedures were used.

Step 1: The dataset was first changed to lowercase.

Step 2: Several unnecessary decimal digits appeared in the text. So, they were taken out.

Step 3: We eliminated all types of punctuation, extra characters, and components.

Step 4: Bengali stop words come in a variety of forms. These cause issues when the data was examined. Therefore, all stop terms from the dataset were eliminated using a corpus of Bengali stop-words.

Fig. 4 displays the flow chart for data cleaning.

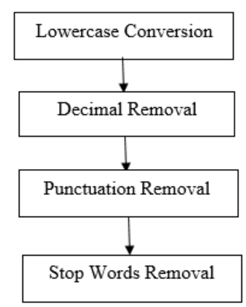


Fig. 4. Flow chart for data cleaning

D. Data Preprocessing

Text preprocessing is an important step in achieving better accuracy in Natural language processing tasks. The dataset should be preprocessed in a particular way for pre-trained models like BERT. Two steps were followed for preprocessing, tokenization and special token addition. The BERT model was trained to transform sentences into tokens. WordPiece tokenization was used in this case. This type of tokenization is useful when dealing with words that are not in our dictionary. After that special tokens, like [CLS], [SEP] were included in the data. The main dataset was then divided into two subsets: the training dataset and the test dataset. Data splitting was carried out in this research using the scikit-learn library's Train test split function.

E. Model Development

For both of our subtasks, aspect detection and sentiment classification, we used the BERT model. We began the aspect detection task by training the BERT model with training data and testing the model with test data. We used the same method to detect sentiment. By fully understanding both the left and right context, BERT aims to pre-train from the unlabeled text. Making use of Transformer's attention mechanism, BERT learns contextual relationships between words.

BERT: BERT is an acronym that stands for Bidirectional Encoder Representations from Transformers [20]. It is a free and open-source NLP framework. Transformer serves as an inspiration for BERT as it studies the connections between words in a sentence. Transformer comes with an encoder and a decoder that reads the input data and predicts the outcome. BERT only requires the encoder component of the transformer to complete its task. In order to clarify the meaning of vague terms in data, BERT attempts to provide context to the surrounding data. The BERT framework was pre-trained using a large volume of Wikipedia texts. BERT considers the context of each word of the text. A word's vector in early word embedding techniques was always the same, regardless of where the word appeared in the text. Contrary to them, BERT provides different vectors for the same words depending on their position. For input representation, the first sentence starts with a classification token, [CLS], and each subsequent sentence ends with a separation token, [SEP]. BERT includes two steps, pre-training and fine-tuning.

a. Pre-Training: In this step, the model is trained on unlabeled text data. Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) are associated with the pre-training task. MLM and NSP are applied together to reduce the total loss function of the two methods [26].

- **Masked Language Modeling (MLM):** The first step is to arbitrarily mask out 15% words in the input text data and replace them with [MASK] tokens [26]. The BERT attention-based encoder is then applied to the entire sequence, which only predicts the masked words based on the context that the other, non-masked words in the sequence provided. Next, the attention-based encoder of BERT processes the input. It only predicts masked words using the context given by the other non-masked words in the sentence.
- **Next Sentence Prediction(NSP):** NSP predicts the connection between two sentences [26]. The primary task of NSP is to determine whether the second sentence comes after the first sentence or not. If the second sentence comes after the first sentence, it is referred to as IsNext; otherwise, it is referred to as NotNext. The sentence is IsNext 50% of the time and NotNext 50% of the time.

b. Fine Tuning: BERT can be fine-tuned to perform a variety of language tasks. By simply placing a single layer over the base model, we can adjust the base model according to our own dataset [26]. The pre-trained parameters are used to initialize the BERT model at first. Then, using annotated data from the required tasks, all of the parameters are fine-tuned. Multilingual BERT is trained in 104 languages using Masked Language Modeling (MLM) [27], and Bengali is one of them. In our work, we used the multilingual BERT model to identify aspects and sentiments in order to gauge the scope of this language model. The pre-trained model weights and implementation for HuggingFace's Transformers have been made public, and it was used in our work. The two parts of BERT are encoding and decoding, respectively. Encoding is used to read text data, and decoding is used to make predictions. The classification layer is then placed over the encoding layer. In our work, in the classification layer, the loss function was computed using binary cross entropy, and the Adam optimizer was applied. The ktrain library was used to calculate the proper learning rate. The model was fed with training data. Then, the model was trained with the previously calculated learning rate, weight, and 10 epochs. Then, testing was done using the test dataset. Fig. 5 depicts how BERT processes a Bengali sentence.

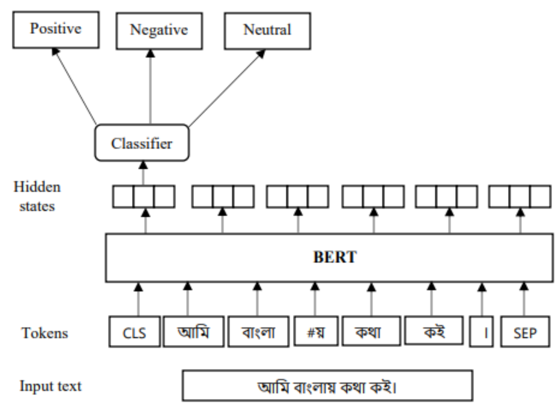


Fig. 5. BERT processing on Bengali text

Along with BERT, a number of deep learning models were also used in this ABSA work. Word2Vec [28] preprocessing technique was used in those models for feature extraction. Some brief descriptions of the algorithms are given below:

- **CNN:** CNN is a member of the family of ANN in deep learning. Using a variety of building pieces, including convolution layers, fully connected layers, and pooling layers, CNN is intended to adaptively and instinctively grasp spatial feature hierarchies by backpropagation [21].
- **RNN:** The outcome from the prior step is provided as input to the ongoing step in an RNN, a form of neural network [22]. In contrast to conventional neural networks, RNN input and output are interdependent. Many times, when forecasting the upcoming word of a statement, the prior words are necessary, and thus the preceding words must always be memorized. RNN was designed to address this issue by utilizing a Hidden Layer. The Hidden state is the most essential characteristic of RNN. It stores some sequence information in memory.
- **LSTM:** An RNN variant that can pick up dependency of order in predicting sequence issues is the LSTM network [23]. Every step uses the output from the previous one as its input in this technique. It resolved the problem of RNN long-term reliance, which is

that RNN can anticipate words looking at recent data but cannot predict words kept in long-term memory. The substantial number of memory cells and 4 neural networks, which are arranged in a chain pattern, make up the LSTM. The four components of a typical LSTM unit: Cell, Input gate, Output gate, Forget gate. Three gates regulate the data flow in and out of the cell, as well as the cell retains contents for arbitrary time periods. Time series with indeterminate duration can be examined, categorized, and predicted with the LSTM algorithm.

- **BiLSTM:** Two models are integrated into the bidirectional LSTM rather than training only one [24]. The input sequence is acquired by the first model, while its reverse is understood by the second model. Context from both sides enhances the performance of language-based models in numerous applications, including language translation, speech recognition, etc. Bi-Directional LSTMs are employed to use those features.
- **GRU:** GRU is a variant of RNN [19]. Similar to LSTM, it was created to address the vanishing gradient issue with RNN. It only has three gates, as opposed to the four in LSTM, and it doesn't keep track of the internal state of the cell. The hidden state of the GRU incorporates similar information as from the internal cell state of the LSTM recurrent unit. This set of information is sent to the subsequent GRU.

IV. Experiment and Result

A. Preparing Environment

Deep learning-based experiments need powerful CPU and GPU. In order to conduct experiments, a system with a core i5 processor with 8 GB of RAM was used. KERAS has been programmed in Python while TensorFlow has served as the backend.

B. Result

The results were evaluated for both subtasks using well-known evaluation matrices. They are F1-score, accuracy, recall, and precision. They are assessed using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Equations for accuracy, precision, recall, and F1-score are [25]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 * (Recall * Precision)}{Recall + Precision} \tag{4}$$

Both aspect detection and sentiment classification are multi-label classification. Tables IV and V display the findings of our analysis of the dataset for aspect extraction and sentiment classification.

TABLE IV. Aspect Detection Result

Algorithm	Accuracy	Precision	Recall	F1-score
BERT	**0.97**	**0.97**	**0.97**	**0.97**
BiLSTM	0.95	0.95	0.96	0.96
LSTM	0.94	0.95	0.94	0.95
GRU	0.88	0.94	0.90	0.92
RNN	0.71	0.85	0.75	0.78
CNN	0.91	0.93	0.92	0.93

TABLE V. Sentiment Classification Result

Algorithm	Accuracy	Precision	Recall	F1-score
BERT	**0.77**	**0.78**	**0.77**	**0.77**
BiLSTM	0.74	**0.78**	0.74	0.75
LSTM	0.73	**0.78**	0.74	0.75
GRU	0.62	0.77	0.64	0.69
RNN	0.52	0.68	0.51	0.58
CNN	0.63	0.74	0.67	0.69

Tables IV and V show that BERT got the highest score in both subtasks. Its F1-Score was 0.97 and it had a 97% accuracy rate for aspect detection. As for sentiment classification BERT accurately classified 77% sentiments, with an F1-Score of 0.77. BiLSTM achieved a 95% accuracy in aspect detection, closely followed by LSTM at 94%. Their respective F1 scores for these subtasks were 0.96 and 0.95. In terms of sentiment classification, BiLSTM predicted 74% sentiment correctly, while for LSTM it was 73%. CNN and GRU performed almost similarly in both subtasks. Their F1-score for those subtasks were same (0.75). Despite having the same F1-score in sentiment classification (0.69), CNN outperformed GRU in aspect detection (0.93). The subtasks with the poorest performance were completed by RNN. It achieved an F1-Score of 0.78 and an accuracy of 71% for aspect detection. Only about half of the sentiments were correctly predicted by RNN (accuracy: 52%). The RNN's F1-Score for detecting sentiments was 0.58.

C. Discussion

Based on the findings in the result section, the BERT model outperformed every other deep learning algorithm, in the both subtasks of ABSA, aspect detection and sentiment classification. They achieved the best scores in all the evaluation matrices due to its "self-attention" quality. Other classifiers' use of the word embedding method, which is not context-based, led them to miss some of the word's context and produce some inaccurate predictions. Contextual embedding, used by BERT, ensures that words are understood in relation to one another. It is one of the main factors influencing BERT's exceptional performance.

The results of LSTM and BiLSTM were close, but BiLSTM outperformed LSTM in terms of its ability to accurately detect more aspects and sentiments due to its capacity to understand the context from both sides of a statement, whereas LSTM only comprehends context from the beginning of a sentence. GRU and CNN performed comparably. RNN performed rather poorly as a result of vanishing gradient issues.

It can also be seen that aspect detection produces better results than sentiment classification since the dataset is evenly distributed in terms of aspects but slightly unbalanced in terms of sentiments.

D. Comparison with Previous Works

The results of this study were compared with four earlier Bengali ABSA based papers, [1], [2], [15], [17] and [18]. Table VI shows a comparison of our dataset to datasets used in their studies. Tables VII and VIII, respectively, provide performance comparisons for aspect detection and sentiment classification among them.

TABLE VI. Comparison of Bengali ABSA Datasets

Dataset	Total comments	Aspect category	Sentiment polarity
Cricket [15]	2958	Batting, Bowling, Team, Team management, Other	Positive Negative Neutral
Restaurant [15]	2053	Food, Price, Service, Ambiance, Miscellaneous	Positive Negative Neutral
BAN-ABSA [18]	9009	Politics, Sports, Religion, Others	Positive Negative Neutral
Our Dataset	10042	Technology, Economy, Corruption, Sports, Politics	Positive Negative Neutral

TABLE VII. Comparison of Aspect Detection Works for Bengali ABSA

Research	Dataset	Algorithm	Accuracy	F1-score
Rahman *et al.* [2, 15]	Cricket	RF	0.25	0.37
		SVM	0.19	0.35
		CNN	0.81	0.51
Rahman *et al.* [2, 15]	Restaurant	RF	0.30	0.33
		SVM	0.29	0.38
		CNN	0.83	0.64
Haque *et al.* [17]	Cricket	RF	-	0.37
		SVM	-	0.35
		LR	-	0.34
Haque *et al.* [17]	Restaurant	RF	-	0.35
		SVM	-	0.39
		LR	-	0.43
		RF	-	0.41
F. A. Naim [1]	Cricket	SVM	-	0.48
		CNN	-	0.59
F. A. Naim [1]	Restaurant	RF	-	0.35
		SVM	-	0.52
		CNN	-	0.67
Masum *et al.*[18]	BAN-ABSA	RF	-	0.65
		SVM	0.69	0.66
		CNN	0.79	0.75
		LSTM	0.77	0.76
		BiLSTM	0.80	0.78
This Research	Our Dataset	CNN	0.91	0.93
		RNN	0.71	0.78
		GRU	0.88	0.92
		LSTM	0.94	0.95
		BiLSTM	0.95	0.96
		BERT	**0.97**	**0.97**

Tables VI, VII, and VIII clearly show:

- Among all the datasets, the CNN algorithm gave the best accuracy and F1-score (0.93 and 0.93) in aspect detection for the dataset used in our research. A clear illustration of the performance of CNN algorithm in these datasets can be found in Fig. 6.

TABLE VIII. Comparison of Sentiment Classification Works for Bengali ABSA

Research	Dataset	Algorithm	Accuracy	F1-score
Masum *et al.*[18]	BAN-ABSA	SVM	0.65	0.41
		CNN	**0.72**	0.60
		LSTM	0.70	0.61
		BiLSTM	0.71	**0.62**
This Research	Our Dataset	CNN	0.63	0.69
		RNN	0.52	0.58
		GRU	0.62	0.69
		LSTM	0.73	0.75
		BiLSTM	0.74	0.75
		BERT	**0.77**	**0.77**

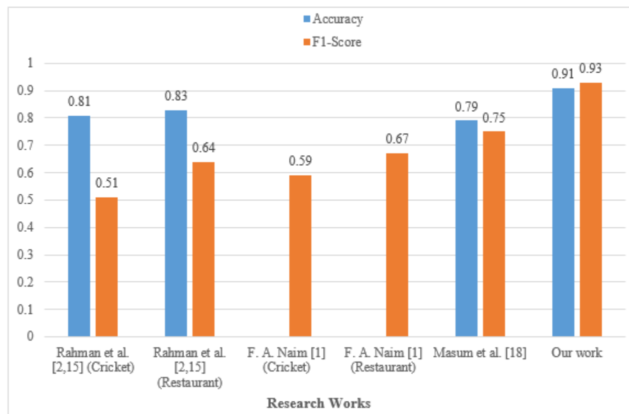


Fig. 6. Representation of CNN performance in research works for aspect detection

- It also can be seen that for the BiLSTM algorithm, that was applied by Masum *et al.* [18] and in our work for aspect detection, it scored better in this work with an accuracy score of 0.95 and F1-score of 0.96.
- To the best of our knowledge, among all Bengali ABSA works, for subtask 1, aspect detection, this research got the best accuracy score and the best accuracy and F1-score for BERT (0.97 and 0.97). Fig. 7 depicts the best aspect detection scores from the research works for clear comparison.
- To the best of our knowledge, this research received the best accuracy and F1-score for BERT (0.77 and 0.77) for subtask 2, sentiment classification, in comparison to the only other ABSA subtask 2 work that has been done in [18]. The comparison of the highest sentiment classification scores from the research works is shown in Fig. 8.
- When it comes to performance of the model, the dataset's richness and quality are important. Given that the dataset used in this research is larger than the Cricket, Restaurant, and BAN-ABSA datasets and includes more diversity, the same algorithms (e.g. CNN, BiLSTM) scored higher in this dataset.

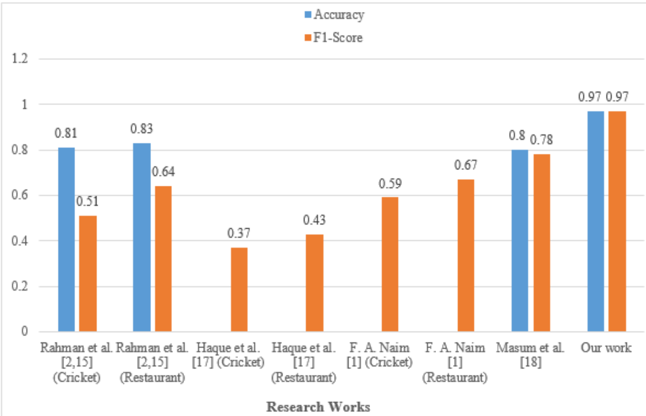


Fig. 7. Comparison of best performances of research works for aspect detection

Fig. 8. Comparison of best performances of research works for sentiment classification

V. Conclusion and Future Work

A. Conclusion

Since there are very few ABSA research works on Bengali text currently available, more Bengali language research in this field is still needed. Our model to perform ABSA was built by applying BERT, which is known for getting over the limitations of the algorithms used in prior Bengali ABSA works. Also, a new Bengali dataset was introduced that is annotated with five aspect categories (Technology, Politics, Sports, Economy, and Corruption) and three sentiment categories (Negative, Positive, and Neutral), and these aspects and categories are particularly popular in our everyday life. After comparing it to five additional models that were applied in this study, it is evident that our BERT-based model outperformed the others. Due to BERT's ability to understand words in relation to each other in a phrase, which CNN, RNN, GRU, LSTM, and BiLSTM lack, it received the highest score in both subtasks. For aspect detection, BERT accurately detected 97% of the aspects and the F1-Score was 0.97. As for sentiment analysis, the F1-Score was 0.77 and accuracy was 77%.

B. Future Work

Our future research will be concentrated on researching hybrid methodologies, which will integrate a number of models and practices to improve the efficacy of this research. The models used in this study adhere to supervised techniques. We'll use unsupervised techniques like Latent Dirichlet Allocation (LDA) in the future as well. We will focus more on reducing time complexity and memory space requirements. To more effectively train our models, we will conduct our study in the future using more sophisticated equipment. In order to enhance this dataset and add further preparation steps, we will add extra data before training models. Romanized Bengali and code-switched Bengali data will also be added to increase the diversity of the dataset. The sentiments can be classified into a wider variety of human emotions, including joy, sorrow, hate, and anxiety.

References

[1] F. A. Naim, "Bangla Aspect-Based Sentiment Analysis Based On Corresponding Term Extraction". International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021.

[2] M. A. Rahman, E. K. Dey, "Aspect Extraction from Bangla Reviews using Convolutional Neural Network." 7th International Conference on Informatics, Electronics & Vision and 2nd International Conference on Imaging, Vision & Pattern Recognition. 2018.

[3] A. Ahmed, M. A. Yousuf, "Sentiment Analysis on Bangla Text Using Long Short-Term Memory (LSTM) Recurrent Neural Network." Proceedings of International Conference on Trends in Computational and Cognitive Engineering. 2020.

[4] R. R. Subramaniam, N. Akshith, "A survey on sentiment analysis." 11th International Conference on Cloud Computing, Data Science and Engineering, 2021.

[5] R. F. Alhujaili, W. M.S. Yafooz, "Sentiment Analysis for Youtube Videos with User Comments: Review." International Conference on Artificial Intelligence and Smart Systems (ICAIS). 2021.

[6] M. J. Althobaiti, "BERT-based Approach to Arabic Hate Speech and Offensive Language Detection in Twitter: Exploiting Emojis and Sentiment Analysis." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 5. 2022.

[7] A. Suciati, I. Budi,"Aspect-Based Sentiment Analysis and Emotion Detection for Code-Mixed Review." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9. 2020

[8] T. Nijhawan , G. Attigeri, T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions." Journal of Big Data 9, Article number: 33, 2022.

[9] B. Xing, I. W. Tsang, "Out of Context: A New Clue for Context Modeling of Aspect-Based Sentiment Analysis." Journal of Artificial Intelligence Research 74, 2022.

[10] R. Basri, M.F. Mridha, M. A. Hamid, M. M. Monowar, "A Deep Learning based Sentiment Analysis on Bang-lish Disclosure." National Computing Colleges Conference (NCCC), 2021.

[11] K. I. Islam, M. S. Islam, M. R. Amin, "Sentiment analysis in Bengali via transfer learning using multi-lingual BERT." 23rd International Conference on Computer and Information Technology (ICCIT). 19-21 December. 2020.

[12] B. Liu, "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies, Vol. 5, No. 1. 2012

[13] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, et al, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis." Association for Computational Linguistics, pp 27-35. 2014.

[14] Akbar Karimi, Leonardo Rossi, Andrea Prati, "Adversarial Training for Aspect-BasedSentiment Analysis with BERT." 25th International Conference on Pattern Recognition (ICPR). 2020.

[15] M. A. Rahman, E. K. Dey, "Datasets for aspect-based sentiment analysis in bangla dataset." MDPI Journals, 2018, doi: 10.3390/data3020015.

[16] M. Bodini, "Aspect Extraction from Bangla Reviews Through Stacked Auto-Encoders." MDPI Journals, 2019, doi:10.3390/data4030121.

[17] S. Haque, T. Rahman, A. K. Shakir, M. S. Arman, K. B. B. Biplob, F. A. Himu, and, D. Das, "Aspect Based Sentiment Analysis In Bangla Dataset Based On Aspect Term Extraction." 2nd International Conference on Cyber Security and Computer Science, 2020.

[18] M. A. Masum, S. J. Ahmed, A. Tasnim, M. S. Islam, "BAN-ABSA: An Aspect-Based Sentiment Analysis dataset for Bengali and it's baseline evaluation." Proceedings of International Joint Conference on Advances in Computational Intelligence, pp 385–395. 2021.

[19] P. B.Weerakody,K. W. Wong, G. Wang, W. Ela, "A review of irregular time series data handling with gated recurrent neural networks." Neurocomputing, pp 161-178. 2021.

[20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding." NAACL-HLT, pp 4171–4186.2019.

[21] J. Alawad, B. Alburaidi, A. Alzahrani, F. Alflaj, "A Comparative Study of Stand-Alone and Hybrid CNN Models for COVID-19 Detection." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 5. 2021.

[22] A. Flores, H. Tito, D. Centty, "Recurrent Neural Networks for Meteorological Time Series Imputation." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 3. 2020.

[23] S. Alhagry, A. A. Fahmy, R. A. El-Khoribi, "Emotion Recognition based on EEG using LSTM Recurrent Neural Network." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10. 2017.

[24] M. Chihab, M. Chiny, N. M. H. Boussatta,Y. Chihab, M. Y. Hadi, "BiLSTM and Multiple Linear Regression based Sentiment Analysis Model using Polarity and Subjectivity of a Text." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 10. 2022.

[25] Z. Vujovic, "Classification Model Evaluation Metrics." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 6. 2021.

[26] K. Ghosh, A. Senapati, "Technical Domain Classification of Bangla Text using BERT." Proceedings of Intelligent Computing and Technologies Conference (ICTCon2021). 2021.

[27] T. H. V. Phan, P. Do, "BERT+vnKG: Using Deep Learning and Knowledge Graph to Improve Vietnamese Question Answering System." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7. 2020.

[28] A. Samih, A. Ghadi, A. Fennan, "ExMrec2vec: Explainable Movie Recommender System based on Word2vec." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 8. 2021.

Drought Forecasting in Alibori Department in Benin using the Standardized Precipitation Index and Machine Learning Approaches

Rodrigue B. W. VODOUNON[1], Henoc SOUDE[2], Ossénatou MAMADOU[3]
Department of Computer Science[1,2], Department of Physics[3]
Institute of Mathematics and Physics (IMSP), Porto-Novo, Benin

***Abstract*—Drought forecasting provides an early warning for the effective management of water resources to avoid or mitigate drought damage. In this study, the prediction of droughts is carried out in the department of Alibori in Benin republic using the standardized precipitation index (SPI) where two Machine Learning approaches were used to set up the drought prediction models which were Random Forest (RF) and Extreme Gradient Boosting (XGBOOST). The performance of these models was reported using metrics such as: coefficient of determination (R^2), root mean square error (RMSE), mean square error (MSE), and root mean absolute error (MAE). The results revealed that XGBOOST models gave better prediction performance for SPI 3, 6, 12 with coefficients of determination of 0.89, 0.83 and 0.99, respectively. The square root mean square error (RMSE) of the models gives 0.29, 0.40 and 0.07, respectively. This work demonstrated the potential of artificial intelligence approaches in the prediction of droughts in the Republic of Benin.**

Keywords*—*Droughts; forecasting; machine learning; SPI

I. Introduction

Droughts come after floods and represents one of the most dangerous natural disasters affecting many countries in the world, especially the countries of West Africa which includes Benin republic [1]. Generally caused by a lack control over precipitation, droughts are very complex and difficult to identify because of their ability to occur in any climate, anywhere on earth [2]. Losses caused by droughts between 1900 and 2013 worldwide are estimated at 135 billion dollars. Moreover, in 2018, a total of 281 climatic and geophysical events were recorded with 10,733 deaths and more than 60 million people affected worldwide [3]. According to [4] and [5], climate change would be a factor in the aggravation of drought situations and their frequency in the world. Several vital sectors such as agriculture sectors, food security, hydroelectricity production, human and animal health, industrial activities [2] are seriously affected by the adverse effects of droughts.

Benin republic, like the other West African countries, is not on the sidelines and has suffered the harmful consequences of climate change. According to [3], Benin has recorded more than half a dozen droughts with ever-increasing disasters, from the 1960s to date. One of the most impacted sectors of the economy is the agricultural sector, which contributes 1/3 of the Gross domestic product (GDP) and holds up to nearly 70% of jobs in Benin republic [6]. Despite the significant impacts of drought, Benin republic, like most African countries, has no tool to either monitor or predict it's occurrence.

According to their effects and duration, droughts are classified into three categories: meteorological, agricultural and hydrological droughts. Several indices have been developed to predict and forecast the different types of drought. Gokhan et al. [7] drew up in their bibliometric analysis an exhaustive list of the various indices used in the literature as well as the conditions in which to use them. We can mention: the Standardised Precipitation Evapotranspiration Index (SPEI) dedicated for determining the onset, duration and magnitude of drought conditions; the Standardized Precipitation Index (SPI) to monitor and predict droughts on several timescales including 3, 6 and 12 months corresponding to SPI 3, SPI 6 and SPI 12, respectively. The SPI 3 is dedicated for the meteorological drought prediction, the SPI 6 is used to predict agricultural drought and the SPI is dedicated to the prediction of hydrological drought.

Abhirup et al. [8] also used artificial neural networks to understand the effect of droughts in New South Wales (NSW) using the SPEI drought index. The performance results revealed a coefficient of determination of 0.86. On the other hand, Abhirup et al. [9] used Long Short Term Memory (LSTM) to predict SPEI at 1 and 3 months time scales in New South Wales. The results of this study revealed a coefficient of determination of more than 0.99 for SPEI 1 as for SPEI 3. Poornima et al. [10] on the other hand compared the predictive capacities of SPI and SPEI in China by the LSTM and the model ARIMA at time scales of 1, 3 and 6 months. In all cases, the LSTM outperformed the capabilities of the ARIMA model. Other studies such as [11] have used Random-Forest to predict SPI at 3 and 12 month timescales in the Haive River Basin in China. These models showed good prediction accuracy. Lotfirad et al. [12] also used Random-Forest to predict SPI and SPEI at time scales of 3, 12, 48 months in Iran. The results obtained showed that in temperate climates, such as northern Iran, the correlation coefficients of SPI and SPEI were 0.94, 0.95 and 0.81 at time scales of 3, 12 and 48 months, respectively while it was 0.47, 0.35 and 0.44 in arid and hot climates.

In spite of the multitude of works, Africans are less concerned and among them we can cite the work of Mulualem et al. [13] where they used Artificial Neural Networks to predict the Normalized Precipitation and Evapotranspiration Index (SPEI) for seven stations in the upper Blue Nile (UBN) basin in Ethiopia. The results obtained found that the coefficient of determination and root mean square error of the best model ranged from 0.820 to 0.949 and 0.263 to 0.428, respectively. As in the most studies on drought forecasting, they use satellite data sources without first assessing their reliability; satellite data may not reflect the reality.

The objective of this work is to propose a prediction tool of meteorological, hydrological and agricultural droughts in the Alibori department (BENIN), using the SPI 3 SPI 6 and SPI 12 indicators respectively. The first time we focus on using satellite sources. Indeed, in the concern to have a model close to reality, we carried out an analysis of similarities between satellite data and measured data. We then proposed two models based on the algorithms random forest and XGBoost. Finally, we presented a technique to improve the performance of the model based on the XGBOOST, by combining the different variables present in our dataset.

The rest of the document is organized as follows: Section II presents study area and drought indicators. Then we described our prediction approach in Section III. In Section IV we presented and discussed our results then we concluded with Section V

II. Background

In this section, we first presented our study area followed by the description of the indices chosen for the prediction of drought in our study area. Then, we presented the algorithms used for the implementation of the different models.

A. Study area

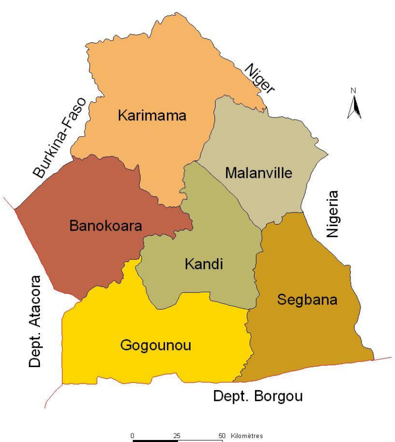


Fig. 1. Cartography of the alibori department (6)

Geographically located between $11°19'$ north latitude and $2°55'$ east longitude, Alibori is a department located in the northeast of the Republic of Benin. It is bordered to the north by the Republic of Niger, to the northwest by the Republic of Burkina Faso, to the east by the Federal Republic of Nigeria, to the west by Atacora and to the south by the department of Borgou [14]. Covering an area of $26\,242\,\mathrm{km}^2$ (nearly 23% of the national territory), Alibori is subdivided into six (06) municipalities as shown in Fig. 1 . These are: Malanville, Karimama, Segbana, Gogounou, Banikoara and Kandi, comprising 41 districts and 229 villages and city districts [14]. According to [6], the department of Alibori is the area most vulnerable and most at risk to droughts in Benin republic. [15] confirms this finding by stating that drought is more recurrent in the northern part of Benin, particularly in the municipalities of Karimama, Malanville and Segbana.

B. Standardized Precipitation Index (SPI)

The standardized precipitation index was developed in 1993 by Mc Kee, N.J. Doesken and J. Kleist of Colorado State University. Its role is not only to determine rainfall deficits for a given period [16] but also to improve the detection of drought episodes at both local and regional scales [17]. Its calculation is based solely on historical rainfall data. To calculate the SPI more efficiently, [2] suggests that it is desirable to have monthly precipitation data over a period of 20 to 60 years. The mathematical expression for calculating the standardized precipitation index is as follows:

$$\textbf{SPI} = \frac{P_i - P_m}{\sigma} \tag{1}$$

with:

$$\begin{cases} P_i : \text{the total rain of month i;} \\ P_m : \text{the average rainfall of the series on the time scale considered;} \\ \sigma : \text{the standard deviation of the series on the time scale considered.} \end{cases}$$

The seasons can be classified according to the values of the SPI as shown in Table I.

TABLE I. Classification of Seasons based on SPI Values [18]

SPI values	Season classes
2.0 and more	Extremely wet
from 1.5 to 1.99	Very humid
from 1.0 to 1.49	Moderately humid
from -0.99 to 0.99	Near normal
from -1.0 to -1.49	Moderately dry
from -1.5 to -1.99	Very dry
-2 and less	extremely dry

C. Machine Learning Approaches

1) Random forest: Random forests represent a supervised learning method based on decision trees [19]. This method generates multiple trees for prediction. Thus, at the end of the individual predictions, these are averaged to give a final prediction of the model. Each tree being weak, the aggregation of all the weak trees compensates for this shortcoming [20] and creates a more robust model. According to [21], the mean makes a random forest better than a single decision tree, thus improving its accuracy and reducing overfitting. The functioning of Random Forests is illustrated by Fig. 2.

2) Extreme gradient boosting: Extreme Gradient Boosting, also called XGBOOST, is an optimized implementation of the gradient boosting tree algorithm [22]. It is often the winning algorithm in competitions on Kaggle [23]. Unlike the Random Forest, the XGBOOST works sequentially. In other words, the algorithm is based on previous predictions to improve the results of other estimators. Thus, we rely on the predictions of the "weak learners" to build a "strong learner" [23]. XGBoost minimizes a regularized objective function (L1 and L2) that combines a convex loss function (based on the difference between predicted and target outputs) and a penalty term for model complexity. The training proceeds iteratively, adding new trees that predict the residuals or errors of the previous trees which are then combined with the previous trees to make the final prediction [24]. Fig. 3 further illustrates how the XGBOOST algorithm works.

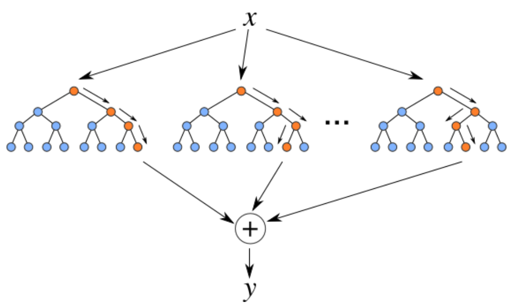


Fig. 2. Structure of a random forest (Random forest) [25]

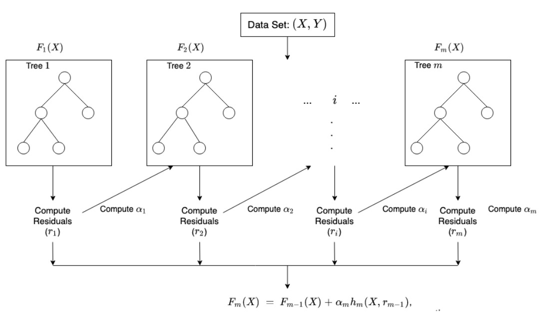


Fig. 3. How XGBOOST works [24]

3) Performance assessment: The mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and coefficient of determination (R^2) are five typical performance indicators used to assess the models performance. The following mathematical expressions illustrate how to calculate these metrics.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Actual - Predicted| \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Actual - Predicted\right)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Actual - Predicted\right)^2} \tag{4}$$

$$R^2 = 1 - \frac{Squared\ sum\ error\ of\ regression\ line}{Squared\ sum\ error\ of\ mean\ line} \tag{5}$$

with n : The number of samples.

III. Prediction Approach

A. Methodology

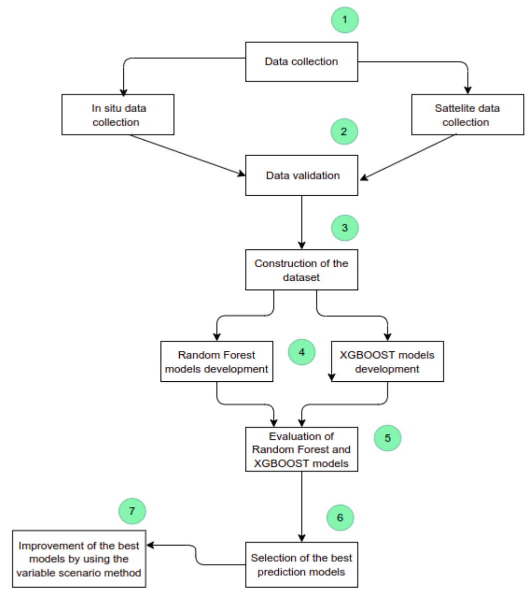


Fig. 4. Methodology of the study conducted

Our solution consists in predicting droughts in the department of Alibori using the drought index SPI$\{3, 6, 12\}$. The prediction is performed using machine learning algorithms, as shown in Fig. 4. First, we collected satellite data covering our entire study area over a period of 30years. Then, we carried out the validation of these data using observation data from a station of the AMMA-CATCH program (Multidisciplinary Analysis of the African Monsoon-Coupling of the Tropical Atmosphere and the Hydrological Cycle) [26]. After that, we built our dataset and our models based on the Random Forest and XGBOOST algorithms. Finally we selected the best of the models and tried to improve them using a variable scenario method described in [27]. It consists of identifying the most useful variables when predicting indicators. First, we determined the correlation between the variables using the Pearson correlation matrix, then we defined the different combinations of variables, starting with the most correlated variables. After construction of the variable scenarios, the models were retrained and re-evaluated on the basis of each variable scenario.

B. Data

The use of Machine Learning in the prediction of drought indicators requires data over a period of at least 30 years. Unfortunately, we do not have any data for such a period. So as an alternative, we therefore opted for the use of satellite data from NASA's Power Data Access Viewer platform(https://power.larc.nasa.gov/) which has data over a period of 36 years (1985-2021), covering the entirety of our study area and containing all the variables allowing the calculation of the SPI: Temperature, total precipitation, wind speed, wind direction, surface pressure, relative humidity, cloud amount, soil moisture, Top-Of-Atmosphere Shortwave Downward Irradiance and all sky and All Sky Surface Shortwave Downward Irradiance. In addition to this, its use is reported in several works [28], [29], [30] to quote only those.

1) Validation: It consists in evaluating the reliability of satellite data before using them. To this end, we considered the Bellefoungou station (Benin) [31] for which we already had in situ data over (08) years and we collected satellite data corresponding to this region and the same period. We then used the Taylor diagram and the Pearson correlation matrix to study the similarity between in situ and satellite data.The results of this validation phase are presented in Fig. 5 and 6. As shown in Fig. 5, there is a strong correlation (0.91) between observation data and satellite data. Fig. 6 also reveals that there is a strong correlation (0.91) between observation data (represented by the white circle on the x-axis) and satellite data (represented by the blue dot). This figure also shows that the difference in the standard deviation between the observation data and the satellite data is not very high (approximately equal to 0.3) and the error between these data is above 0.5. From all the above, it can be seen that there is a strong similarity between observational data and satellite data.

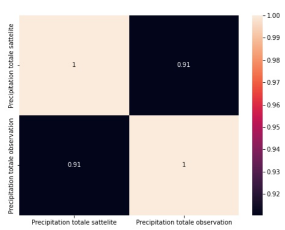


Fig. 5. Correlation matrix of the total precipitation variable

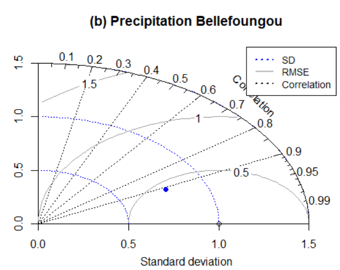


Fig. 6. Taylor diagram of the total precipitation variable

2) Construction of the dataset: By exploiting the precipitation data, the calculation of the SPIs at the time scales of 3, 6 and 12 months corresponding respectively to SPI 3, SPI 6 and SPI 12 has been implemented according to the formula 1 presented in the Section II.

After calculating the different SPIs, observations with missing data were removed. Once the datasets were purged of missing values, the data from the (06) municipalities were merged together in order to have a single dataset (Table II).

C. Model Building

We used the approaches described in the subsection II-C to implement models for the prediction of SPI 3 (meteorological drought), SPI 6 (agricultural drought), SPI 12 (hydrological drought). The different models have been implemented using the python programming language as well as different data science libraries such as: NumPy, Pandas, Scikit-learn, Matplotlib and Seaborn.

TABLE II. Details of the Satellite Dataset

VARIABLES	UNITS	TIME SCALE
Surface Pressure	KPa	Monthly
Total Precipitation	mm	Monthly
Air Temperature at 2 meters	C	Monthly
Relative Humidity at 2 meters	%	Monthly
Wind Speed at 2 meters	m/s	Monthly
Wind Direction at 2 meters	Degree (°)	Monthly
Soil Moisture	-	Monthly
Cloud Amount	%	Monthly
Top-of-Atmosphere Shortwave Downward Irradiance	MJ/m^2/day	Monthly
All Sky Surface Shortwave Downward Irradiance	MJ/m^2/day	Monthly

In order to effectively train and validate the developed models, the dataset was divided into two parts: the training data (77%) and the test data (23%). We also used GridsearchCV to obtain the best hyperparameters for the models (see Tables III and IV).

TABLE III. Random Forest Model Hyperparameter Values

Index	Hyperparameters	Values
SPI 3	n_estimators	70
	random_state	15
SPI 6	n_estimators	100
	random_state	9
SPI 12	n_estimators	100
	random_state	9

TABLE IV. XGBOOST Model Hyperparameter Values

Hyperparameters	Values
Subsample	0,7
colsample_bytree	1
eta	0,1
max_depth	10
n_estimators	100
booster	gbtree
verbosity	0

IV. Results and Discussion

We first evaluated the different models based respectively on Random-Forest and XGBOOST to identify the best models and then we improved them using variable scenario.

A. Random Forest Evaluation

Considering all the variables whose data were collected, prediction models for SPI 3, 6 and 12 were implemented and evaluated using the performance metrics presented in Section II. As it can be seen in the Table V, the models gave SPI 3, SPI 6 and SPI 12, respectively for performance R^2 of 0.832 (83%), 0.742 (74%) and 0.992 (99%).

After evaluating the models, it can be seen that the SPI 3 prediction model gives errors RMSE = 0.379, MAE = 0.276 and MSE = 0.143 respectively. As for the prediction model of SPI 6, we had the following errors: RMSE=0.479, MAE=0.344 and MSE=0.230. The SPI 12 prediction model gave the lowest errors with RMSE= 0.082, MAE = 0.041

and MSE = 0.006 values, respectively. This would mean that the variables considered are more relevant for predicting hydrological drought (SPI 12).

B. XGBOOST Evaluation

The Table VI presents the evaluation results of the prediction models of SPI 3, 6 and 12 based on the XGBOOST algorithm. This table shows good performance for all models with coefficient of determination R^2 respectively equal to: 88% for the SPI 3 prediction model, 81% for the SPI 6 prediction model and 99% for that of the SPI 12.

As for the errors made by the models, the results of the evaluation revealed that the prediction model of the SPI 3 gave the values RMSE = 0.317, MAE = 0.221 and MSE = 0.100, respectively. The SPI 6 prediction model gave the following results: RMSE = 0.423, MAE = 0.301 and MSE = 0.179. The SPI 12 prediction model provides the best performance and minimizes all errors the most with errors RMSE = 0.076, MAE = 0.041 and MSE = 0.005.

By comparing the performances of the Random Forest models and those of XGBOOST, we find that the XGBOOST algorithm is much more efficient in the case of the present study.

TABLE V. Random Forest Model Evaluation Result

Index	R^2	RMSE	MAE	MSE
SPI 3	0.832	0.379	0.276	0.143
SPI 6	0.742	0.479	0.344	0.230
SPI 12	0.992	0.082	0.041	0.006

TABLE VI. XGBOOST model Evaluation Result

Index	R^2	RMSE	MAE	MSE
SPI 3	0.882	0.317	0.221	0.100
SPI 6	0.817	0.423	0.301	0.179
SPI 12	0.993	0.076	0.041	0.005

C. Variable Scénario

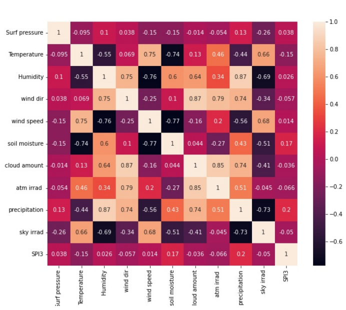


Fig. 7. Correlation matrix between input variables and SPI 3

According to the Fig. 7, the highest correlation is equal to 0.2 and this corresponds to the correlation between the total precipitation and the SPI 3. We also note a correlation of 0.17 between the humidity of the ground and SPI 3. Then follow weaker positive correlations respectively equal to 0.038, 0.026 and 0.014 corresponding respectively to the correlations between SPI 3 and surface pressure, relative humidity and wind speed. In addition to the positive correlations, some variables have a negative correlation with the SPI 3. These are the amount of cloud (-0.036), All Sky Surface Shortwave Downward Irradiance (-0.05), wind direction (-0.057), Top-Of-Atmosphere Shortwave Downward Irradiance (-0.066), Air temperature (-0.15). This correlation matrix made it possible to identify the variables that have a strong correlation with the SPI 3. This is the variable: total precipitation. It can be deduced that the total precipitation represents a relevant variable for the prediction of SPI 3 (meteorological drought). The same approach was adopted for the cases of SPI 6 (agricultural drought) and 12 (hydrological drought).

The Tables VII, VIII and IX respectively present the different variable scenarios according to the type of drought. Indeed, these scenarios are only combinations of the variables obtained from the Pearson correlation matrices. For the experimentation phase, they were successively used as input parameters to the XGBOOST models. These matrices made it possible to construct 11 scenarios per type of drought. The presence of a (✓) symbol in the cell of a column means that the variable concerned is part of the scenario corresponding to this column. For illustrative purposes, scenario 1 of the Table VII consists only of the Total precipitation variable. Scenario 3 consists of the variables: soil moisture, total precipitation and surface pressure.

TABLE VII. Scenarios of Variables for the Prediction of SPI 3 (Meteorological Drought)

Variables \ Scenarios	1	2	3	4	5	6	7	8	9	10	11
Soil moisture		✓	✓	✓	✓	✓	✓	✓	✓		
Total recipitation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Relative humidity				✓	✓	✓	✓	✓	✓	✓	✓
Wind direction								✓	✓	✓	✓
Surface pressure			✓	✓	✓	✓	✓	✓	✓	✓	✓
Cloud amount						✓	✓	✓	✓	✓	✓
Top-Of-Atmosphere Shortwave Downward Irradiance									✓	✓	✓
Wind speed					✓	✓	✓	✓	✓	✓	
All Sky Surface Shortwave Downward Irradiance							✓	✓	✓	✓	✓
Air temperature										✓	

TABLE VIII. Scenarios of Variables for the Prediction of SPI 6 (Agricultural Drought)

Variables \ Scenarios	1	2	3	4	5	6	7	8	9	10	11
Soil moisture	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Total precipitation		✓	✓	✓	✓	✓	✓	✓	✓	✓	
Relative humidity			✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind direction				✓	✓	✓	✓	✓	✓	✓	✓
Surface pressure					✓	✓	✓	✓	✓	✓	✓
Cloud amount						✓	✓	✓	✓	✓	✓
Top-Of-Atmosphere Shortwave Downward Irradiance							✓	✓	✓	✓	✓
Wind speed								✓	✓		
All Sky Surface Shortwave Downward Irradiance									✓	✓	
Air Temperature									✓		

Once the variable scenarios were established, the XGBOOST models were trained and validated by considering each variable scenario. Then, these models were evaluated using previously used performance metrics to identify not only the most accurate model but also the scenarios that provide better prediction performance. The performance results of the models according to each scenario of variables for each SPI

TABLE IX. Scenarios of Variables for the Prediction of SPI 12 (Hydrological Drought)

Variables \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
Soil moisture			✓	✓	✓	✓	✓	✓	✓	✓	✓
Total precipitation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Relative humidity		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind direction				✓	✓	✓	✓	✓	✓	✓	✓
Surface pressure						✓	✓	✓	✓	✓	✓
Cloud amount					✓	✓	✓	✓	✓	✓	✓
Top-Of-Atmosphere Shortwave Downward Irradiance							✓	✓	✓	✓	✓
Wind speed									✓	✓	✓
All Sky Surface Shortwave Downward Irradiance								✓	✓		
Air temperature										✓	

(SPI 3, SPI 6 and SPI 12) are presented in the Tables X, XI, XII. The Table X, reveals that scenario 9 is the one that provides the best results with a determination coefficient R^2 equal to 0.898 (89%), an RMSE of 0.29, an MAE of 0.211 and an MSE of 0.087. This table also reveals two important elements. First, it can be noticed that the more one adds a variable to a scenario, the more the precision of the model increases (even the variables having a weak correlation with the SPI 3) with the exception of the variables soil humidity and air temperature . Secondly, when we use as input parameters, scenario 2 which contains only the variables most correlated to SPI 3, the performance of the model deteriorates completely with a negative R^2 (-0.241). This actually means that the fact that the variables are the most correlated to the SPI 3, does not necessarily mean that they are the most essential for the prediction of the SPI 3. Moreover, a variable having a negative correlation (quantity of cloud for example) with SPI 3 can increase the predictive performance of the model. As for the Table XI, it shows that scenario 9 is still the one that provides the best results for the prediction of SPI 6 with a coefficient of determination R^2 equals to 0.830 (83%), an RMSE of 0.408, an MAE of 0.289 and an MSE of 0.167. This table also shows two other key pieces of information. We can also notice that the more we add a variable to a scenario, the precision of the model increases (even the variables having a negative correlation with the SPI 6) with the exception this time of the temperature variable of the air whose addition lowered accuracy and increased errors. Although the total precipitation and soil moisture variables constituting scenario 2 are the most correlated with SPI 6, the performance of the model was only 0.017 for this scenario. This actually shows that the fact that the variables are the most correlated to the SPI 6, does not necessarily mean that they are the most essential for the prediction of the SPI 6. Moreover, a variable having a negative correlation (wind speed per example) with SPI 6 can increase the predictive performance of the model.

The performance results of the XGBOOST model for the prediction of SPI 12 are presented in the Table XII. Unlike SPI 3 and SPI 6, this time it is scenario 8 that gives the best results with a determination coefficient R^2 equal to 0.994 (99%), an RMSE of 0.073, an MAE of 0.039 and an MSE of 0.005. Several other relevant information can be extracted from this table. Firstly, as in the case of SPI 3 and SPI 6, it can also be seen that the more a variable is added to a scenario, the more the precision of the model increases (even the variables having a negative correlation with the SPI 12) with the exception this time variables: Top-Of-Atmosphere Shortwave Downward Irradiance, wind speed, air temperature, the addition of which decreased accuracy and increased errors. Secondly, considering the three variables most correlated to SPI 12 which are: total precipitation, soil moisture and relative humidity forming scenario 3, the performances are already starting to be good (R^2 = 86%) . This is not the case for the prediction of SPI 3 and 6. This actually shows that the total precipitation, soil moisture and relative humidity have a strong influence on the prediction of SPI 12.Having a negative correlation with the SPI 12, only the addition of the variable All Sky Surface Shortwave Downward Irradiance slightly increased the performance of the model 0.994 (99%).

The variation of predicted SPI values and calculated SPI values is illustrated in Fig. 8, 9 and 10. As it can be seen, the predicted SPI (in blue) have a similar variation as the calculated SPIs 3 (in red). By considering these three figures, it can be noticed that the predicted SPI 12 is very close to the calculated SPI 12. This justifies the 99% accuracy of the SPI 12 prediction model. It can therefore be said that the XGBOOST is better suited to the prediction of hydrological drought in the department of Alibori.

TABLE X. Performance Results of the SPI 3 Prediction Model based on the XGBOOST Algorithm

Métrique \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
R^2	-0.0005	-0.241	0.200	0.449	0.526	0.636	0.646	0.665	0.898	0.884	0.874
RMSE	0.926	1.032	0.828	0.687	0.637	0.558	0,551	0.536	0.295	0.314	0.328
MAE	0.712	0.722	0.567	0.463	0.441	0.393	0.391	0.374	0.211	0.226	0.225
MSE	0.859	1.065	0.686	0.472	0.406	0.311	0.303	0.287	0.087	0.176	0.107

TABLE XI. Performance Results of the SPI 6 Prediction Model based on the XGBOOST Algorithm

Métrique \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
R^2	0.015	0.017	0.317	0.458	0.519	0.572	0.784	0.813	0.830	0.810	0.327
RMSE	0.985	0.984	0.189	0.730	0.688	0.649	0.460	0.428	0.408	0.432	0.814
MAE	0.792	0.722	0.136	0.532	0.480	0.452	0.315	0.294	0.289	0.304	0.656
MSE	0.971	0.969	0.035	0.534	0.474	0.421	0.212	0.183	0.167	0.187	0.663

TABLE XII. Performance Results of the SPI 12 Prediction Model based on the XGBOOST Algorithm

Métrique \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
R^2	0.663	0.780	0.860	0.896	0.913	0.994	0.993	0.994	0.993	0.993	0.993
RMSE	0.559	0.451	0.360	0.310	0.283	0.07	0.075	0.073	0.077	0.075	0.076
MAE	0.398	0.300	0.197	0.173	0.157	0.044	0.040	0.039	0.042	0.041	0.041
MSE	0.313	0.204	0.129	0.096	0.080	0.005	0.005	0.005	0.005	0.005	0.005

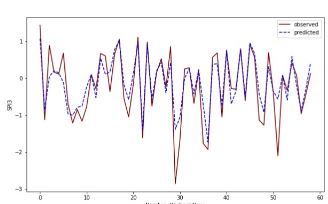


Fig. 8. Comparison of predicted SPI 3 and target SPI 3

V. Conclusion

This study consisted in designing prediction models of SPI 3 (meteorological drought), SPI 6 (agricultural drought) and SPI 12 (hydrological drought) in the department of Alibori based respectively on the Random forest and Extreme Gradient Boosting (XGBOOST) algorithms. Performance results

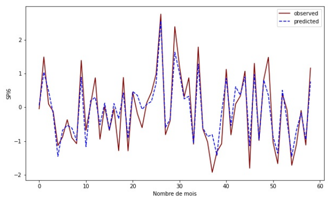


Fig. 9. Comparison of predicted SPI 6 and target SPI 6

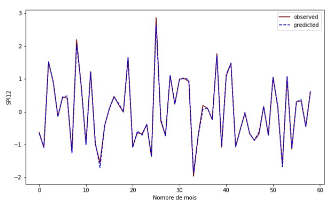


Fig. 10. Comparison of predicted SPI 12 and target SPI 12

indicated that XGBOOST performs better than Random forest in predicting SPI 3, SPI 6 and SPI 12 with respectively the coefficients of determination R^2 équal to 88%, 81% and 99%. To further improve the performance of XGBOOST models, 11 variable scenarios were designed. These scenarios were then used respectively to train, validate and evaluate the models in order to identify the scenarios that allow good prediction performance. At the end of this improvement step, the best performances obtained are respectively 89% for SPI 3, 83% for SPI 6 and 99% for SPI 12. The use of this technique based on the scenarios of variables allowed to deduce that the air temperature was not relevant for the prediction of the SPI 3. Similarly the variables air temperature and wind speed were not essential for the prediction of SPI 12. Future prospects consist of using other drought indices like SPEI in order to compare their results with those obtained in the present study in order to see the drought index that lends itself to drought prediction in Benin in general and in the Alibori in particular).

VI. Perspectives

In the future works, historical data will be used to predict future droughts, which would be useful to meteorological decision makers in developing drought mitigation measures and actions.

References

[1] D. Tsegai, M. Medel, P. Augenstein, and Z. Huang, "La sécheresse en chiffres," Convention des Nations Unies sur la Lutte contre la Désertification (CNULCD), Côte d'Ivoire, Abidjan, Rapport de la 15ème Conférence des Parties (COP 15), Mai 2022.

[2] O. météorologique mondiale (OMM) et Partenariat mondial pour l'eau (GWP), *Manuel des indicateurs et indices de sécheresse*, ser. Programme de gestion intégrée des sécheresses, Integrated Drought Management Programme Tools and Guidelines Series 2, Genève, 2016.

[3] A. N. de Protection Civile (ANPC), "Plan national secheresse 2019-2024," Bénin, Tech. Rep., Janvier 2019.

[4] réseau action climat france. (2019) Le réchauffement va aggraver le risque de sécheresses dans certaines régions. [Online]. Available: https://reseauactionclimat.org/rechauffement-climatique-aggrave-secheresse/

[5] M. France. (2020) Sécheresses et changement climatique. [Online]. Available: https://reseauactionclimat.org/rechauffement-climatique-aggrave-secheresse/

[6] F. de recherche CIMA, C. W. C. pour l'OMM, and A. N. au Bénin, "Projet intégrer la gestion des inondations et des sécheresses et de l'alerte précoce pour l'adaptation au changement climatique dans le bassin de la volta," World Meteorological Organization, Volta Basin Authority, Global Water Patnership West Africa, Rapport de consultation national au Bénin, Janvier 2021.

[7] G. Yildirim, A. Rahman, and V. P. Singh, "A bibliometric analysis of drought indices, risk, and forecast as components of drought early warning systems," *Water*, vol. 14, no. 2, p. 253, 2022.

[8] A. Dikshit, B. Pradhan, and A. M. Alamri, "Temporal hydrological drought index forecasting for new south wales, australia using machine learning approaches," *Atmosphere*, vol. 11, no. 6, p. 585, 2020.

[9] A. Dikshit, B. Pradhan, and A. Huete, "An improved spei drought forecasting approach using the long short-term memory neural network," *Journal of environmental management*, vol. 283, p. 111979, 2021.

[10] S. Poornima and M. Pushpalatha, "Drought prediction based on spi and spei with varying timescales using lstm recurrent neural network," *Soft Computing*, vol. 23, no. 18, pp. 8399–8412, 2019.

[11] J. Chen, M. Li, and W. Wang, "Statistical uncertainty estimation using random forests and its application to drought forecast," *Mathematical Problems in Engineering*, vol. 2012, 2012.

[12] M. Lotfirad, H. Esmaeili-Gisavandani, and A. Adib, "Drought monitoring and prediction using spi, spei, and random forest model in various climates of iran," *Journal of Water and Climate Change*, vol. 13, no. 2, pp. 383–406, 2022.

[13] G. M. Mulualem and Y.-A. Liou, "Application of artificial neural networks in forecasting a standardized precipitation evapotranspiration index for the upper blue nile basin," *Water*, vol. 12, no. 3, p. 643, 2020.

[14] B. Alexandre, S. and A. Djabar, Dine, "Cahier des villages et quartiers de ville du departement de l'alibori," INSTITUT NATIONAL DE LA STATISTIQUE ET DE L'ANALYSE ECONOMIQUE (INSAE), Benin, RECENSEMENT GENERAL DE LA POPULATION ET DE L'HABITATION (RGPH-4), Août 2016.

[15] "Catastrophes naturelles et objectifs du millÉnaire pour le dÉveloppement au bÉnin: Les propositions alternatives de social watch bÉnin," SOCIAL WATCH BÉNIN, Benin, Rapport Alternatifs annuel des Organisations de société civile sur les OMD, Décembre 2010.

[16] R. Mehdaoui, E. MILI, and A. SEGHIR, "Caractérisation à l'aide du spi de la sécheresse climatique dans le bassin versant de ziz (sud-est, maroc)," *Eur Sci J*, 2018.

[17] A. Mokhtar, M. Jalali, H. He, N. Al-Ansari, A. Elbeltagi, K. Alsafadi, H. G. Abdo, S. S. Sammen, Y. Gyasi-Agyei, and J. Rodrigo-Comino, "Estimation of spei meteorological drought using machine learning algorithms," *IEEE Access*, vol. 9, pp. 65 503–65 523, 2021.

[18] S. M., H. M., and W. D., *Guide d'utilisation de l'indice de précipitations normalisé*, Genève, 2012, no. OMM-N° 1090.

[19] E. S. R. I. (Esri). (2020) Fonctionnement de l'outil régression et classification de forêt aléatoire. [Online]. Available: https://pro.arcgis.com/fr/pro-app/latest/tool-reference/geoai/how-random-forest-classification-and-regression-works.htm

[20] M. FERECATU. Apprentissage, réseaux de neurones et modèles graphiques. Paris, France. [Online]. Available: http://cedric.cnam.fr/vertigo/cours/ml2/coursForetsAleatoires.html

[21] D. Mwiti. Random forest regression: When does it fail and why? [Online]. Available: https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why

[22] B. Cayla. La star des algorithmes de ml : Xgboost. [Online]. Available: https://datacorner.fr/xgboost/

[23] Xgboost, le grand gagnant des compétitions. [Online]. Available: https://datascientest.com/xgboost-grand-gagnant-des-competitions-machine-learning-algorithme

[24] How xgboost works. [Online]. Available: https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html

[25] Random forest regression. [Online]. Available: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

[26] S. Galle, M. Grippa, C. Peugeot, I. B. Moussa, B. Cappelaere, J. Demarty, E. Mougin, G. Panthou, P. Adjomayi, E. Agbossou *et al.*, "Amma-catch, a critical zone observatory in west africa monitoring a region in transition," *Vadose Zone Journal*, vol. 17, no. 1, pp. 1–24, 2018.

[27] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index," *Journal of climate*, vol. 23, no. 7, pp. 1696–1718, 2010.

[28] A. M. Abdel-Hameed, M. E.-S. Abuarab, N. Al-Ansari, H. S. Mehawed, M. A. Kassem, H. He, Y. Gyasi-Agyei, and A. Mokhtar, "Winter potato water footprint response to climate change in egypt," *Atmosphere*, vol. 13, no. 7, p. 1052, 2022.

[29] O. A. Marzouk, "Assessment of global warming in al buraimi, sultanate of oman based on statistical analysis of nasa power data over 39 years, and testing the reliability of nasa power against meteorological measurements," *Heliyon*, vol. 7, no. 3, p. e06625, 2021.

[30] G. C. Rodrigues and R. P. Braga, "Evaluation of nasa power reanalysis products to estimate daily weather variables in a hot summer mediterranean climate," *Agronomy*, vol. 11, no. 6, p. 1207, 2021.

[31] M. Hounsinou, O. Mamadou, M. Wudba, B. Kounouhewa, and J.-M. Cohard, "Integral turbulence characteristics over a clear woodland forest in northern benin (west africa)," *Atmospheric Research*, vol. 268, p. 105985, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016980952100541X

Smart Monitoring System using Internet of Things: Application for Agricultural Management in Benin

Pélagie HOUNGUE[1], Romaric SAGBO[2], Gilles DAHOUE[3], Julien KOMACLO[4]
Institut de Mathématiques et de Sciences Physiques
Université d'Abomey-Calavi, Dangbo, Bénin[1,2]
Institut National Supérieur de Technologie Industrielle
Université Nationale des Sciences, Technologies, Ingénierie et Mathématiques, Lokossa, Bénin[3,4]

***Abstract*—One of the major tools of the new era of digital transformation is Internet of Things (IoT) through which, one can look forward to exploring the new technologies in the digital world as well as how they help in improving the real world. This work provides an overview of the approach used to deploy a surveillance system for monitoring any indoor space in general and specifically for agricultural spaces. The entire process starts after motion detection by motion sensors using Machine Learning techniques. This requires coverage and response processing algorithms implemented in the electronic chain. The electronic part of the system relies on the micro-controllers, sensors and communications between them. A mobile application has been developed to allow competent authorities to receive alerts for real-time intervention with the aim of preventing the destruction of crops slaughtered near herds passage. The monitoring system' synoptic diagram and its operation along with the power modules description are introduced. Prototype has been designed and performance evaluation performed to show the system' responsiveness.**

***Keywords*—Monitoring system; agricultural space; machine learning; herds passage; motion detection**

I. Introduction

A thriving society offers myriad benefits: joy, peace, and security. Yet, security continues to be critical. This is true whether the safe space is for people, items, or animals. It makes sense to send out a team of guards to monitor and police a safe space. Therefore, having a monitoring system [1], [2], [3] remains the best engagement solution for well-being. Moreover, thanks to the development of the Internet of Things (IoT), one can now keep in real-time, an eye out for its own assets remotely [4].

Despite the crucial importance and active participation of livestock in the Beninese economy, their producers are less protected from the serious threats that often accompany the search for animal feed, especially in the dry season, which leads to loss of the plantations. With this in mind, a monitoring system that is intended to be able to adapt itself to the user needs, has been designed. The service produced by the system, allows receiving alerts when intrusions occur. Recently, in [5], a multi-layered architecture that is useful for various stakeholders (breeding producers, farmers, and policymakers) has been proposed to support Smart Agriculture and Smart Livestock. Transhumance related issues have been targeted, as those issues gain further expansion in developing countries. Indeed, during the transhumance process, passages dedicated to animals, usually referred to as corridors may not be followed by the breeders. This can cause deadly clashes between herders and farmers. In this paper, the above issues are addressed by extending the architectural solution provided in [5], to the implementation of a monitoring system to detect intrusion when the cows are too close to the farmers field, as well as to send alert message on the phone of the authorities that can try to anticipate the conflicts.

Machine Learning is being intensively used to solve real-time computer vision and image processing problems [6], [7]. Our solution is built using Machine Learning techniques associated with OpenCV library [8], [9] and a pre-trained library of objects and animals from Coco (Common Object in Context) dataset [10], [11] that is large-scale object detection, segmentation, and captioning dataset. Thanks to this trained library, the proposed intrusion detection system, based on Raspberry Pi, will allow recognizing some objects and animals such as cows, oxen and persons, using DNN (Deep Neural Network) algorithm [12], [13].

The specific objectives of this work are: i) use IoT to set up a monitoring system that detects intrusion of cows using a training model, ii) carry out the processing of the captured image and ensure the identification of the cows or oxen, iii) trigger an alarm after the detection of cattle movement at the entrance of the field and send an alert message and details about the field' location to authorities who can take appropriate decision in real-time. Sending alerts to the farmers, is avoided to reduce the direct conflicts before authorities intervention.

The remaining of this paper is set as follows. Section II discusses problem statement and motivation. Section III presents the novel IoT-based intrusion detection system followed by section IV that discusses experimental results. Sections V and VI, respectively focus on related work and conclusion, followed by perspectives.

II. Problem and Motivation

Although the rules of arable land are carefully codified, the roads are delimited and respected by the sedentary, the level of destruction caused by the traditional movements of pasture increases day by day. Land degradation has been a constant challenge for farmers and ranchers as it accompanied the spread of agriculture in prehistoric times. It can create dangerous outcomes like death when large herds enter in neighboring fields, making it necessary for authorities to be alerted to their presence, anywhere in the land [5], [14]. In order to reduce deadly outcomes and large herds of animals roaming fields, pioneers need to lessen the conflict between agricultural experts. This work aims to alert authorities as soon as herds of livestock are advancing toward neighboring ranches. In today's world, surveillance systems are predominantly based on more traditional equipment, such as closed-circuit cameras. Unlike, the proposed system not only monitors herds' passage, but is also able to send the exact position of the concrete place where the intrusion takes place. In addition to this, the image processing is done to decipher whether the movement is actually animals or if it is men or otherwise.

III. IoT-based Detection System

Our monitoring system can be divided into two main parts, namely, the electronic component and the mobile application. The first part, dealing with electronic, covers the communication between the various modules and the electronic devices used. When the motion sensor at the entrance of the field, detects the movement of an object, the camera is activated and the object is identified. If there is a cow or oxen among these identified objects, an alarm will be triggered, when it is less than 2 meters from the system, positioned at the entrance of the agricultural domain. This is called pre-intrusion. However, if the cow crosses the border by entering in the domain within 1.5 meters from the system, the same treatment process is followed. Again, for cows and oxen, it sends an alarm message to authorities and assigns them the location of the sensor that detected the movement, reporting the intrusion.

The second important part of the work, is mobile application design. This application acts as intermediary between electronic system and users. Therefore, it receives alerts, displays alerts, displays the history of intruders, and allows administrators to know the status of components of the system.

This section highlights the system architecture along with its functional requirements. Moreover, electronic details have been underlined, as well as the OpenCV library description.

A. Intrusion Detection System Architecture

Our intrusion detection system is based on a client/server architecture. All processing tasks are performed through the server. It is therefore, the brain of the system, which consists mainly of two parts, as shown in Fig. 1.

- Electronic components consisting of motion sensor, ultrasonic sensor, camera, Raspberry Pi 3, NEO-6M GPS module.
- The mobile application obtains information sent by electronic system regarding the intrusion' detection.

B. Detection Processing Steps

Fig. 2 shows the process of pre-intrusion and intrusion detection in agricultural domain. First, when the PIR motion sensor detects motion at the entrance to the monitored area, it triggers the camera, which is in standby mode by default, to identify objects whose presence has been notified by the sensor. For cows and oxen, ultrasonic sensors assess the distance between the object and the system. If the distance between them is less than 1.5 meters, the system will sound an alarm (buzzer). This is the pre-intrusion detection. When the PIR sensor placed in the field, detects movement within the detection area, an ultrasonic sensor is activated and the distance between the system and the trigger object is also calculated. Then, when the distance between the system and the object is less than 1.5 meters, the camera will be activated to check if it is a cow/oxen. In the case of cow or oxen, the system sends notification to the authorities. This is an intrusion detection. If not, no action is required.

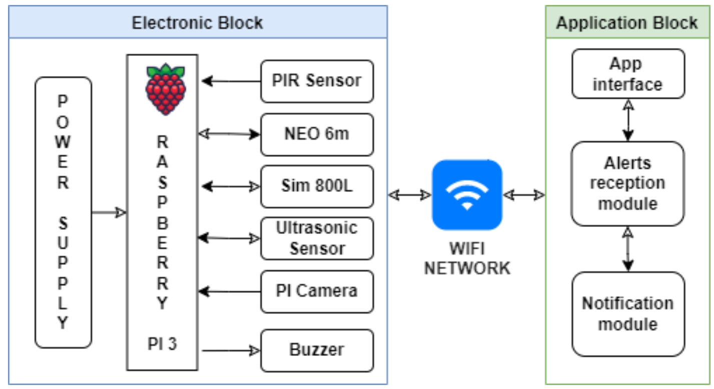


Fig. 1. System architecture

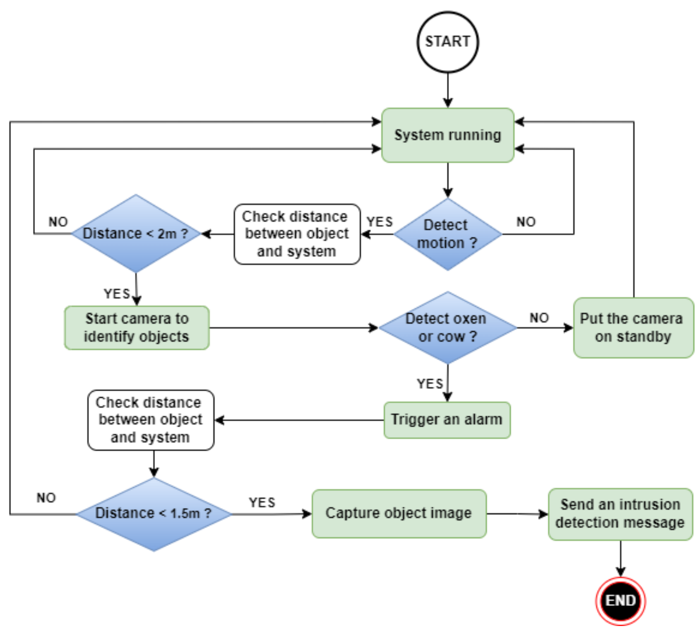


Fig. 2. System operations steps

C. OpenCV Library Usage

The recognition of objects, in particular an ox or a cow, was done using the OpenCV library [8], [9], which makes it possible to manipulate the objects and animals predefined in the Coco dataset [11]. The Coco (Common Object in Context) library is a large-scale object detection, segmentation, and captioning dataset. This library is designed from a large collection of images of each object and animal in different shapes and various landscapes. It allows the Raspberry Pi to identify 91 unique objects/animals and to provide a constantly updated confidence score. The OpenCV library along with Coco dataset are set up on the Raspberry.

Object recognition, specifically, cow and oxen, was performed using the OpenCV library [8], [9]. As library, it allows manipulating predefined objects and animals in the Coco dataset [11]. Coco (Common Object in Context) is a rich object detection, segmentation and captions dataset. It was designed from a large collection of images of each object and animal in different shapes and different landscapes. This allowed the Raspberry Pi to identify 91 unique objects/animals and provide a constantly updated confidence score. An OpenCV library along with a Coco dataset are set up on the Raspberry Pi.

D. Use Cases

Use case diagrams are provided in terms of actions and reactions, allowing to visualize system behavior from the user's point of view and give a complete overview of the functional behavior of a system. Fig. 3 shows "what" the proposed system provides by illustrating the possible interactions with the actors.

After authentication, authorities can:

- Comment on intrusions
- Visualize alerts
- Consult intrusions history

After authentication, administrator can manage the system as follows:

- Add or remove users in the system
- Check system status
- Consult intrusions history

E. Prototyping

Several tools were used in order to design the monitoring system. The following materials/tools are used to set up the prototype:

- Raspberry Pi 3, which is a computer reduced to its simplest form with a single ARM processor card, a bit larger than a credit card.
- Camera Pi, Raspberry Pi Night Vision Camera, that supports all versions of Raspberry Pi.
- PIR sensor, is a motion sensor used to determine whether a human or an animal has entered or left the module's detection field.
- Ultrasonic sensor that evaluates the distance between the object and the system.
- SIM800L GSM module, is the smallest and powerful GSM module which can automatically start and look for network connectivity. It allows SMS exchange, calls and data recovery in GPRS 2G+.
- The NEO-6M GPS module, is a GPS receiver, which has a high performance built-antenna to provide powerful satellite search capability.
- Solar panel kit, which is composed by a 10W 12V solar panel, a 5A 12/24V charge regulator and a 12V 12Ah battery. It is used to power on the monitoring system.

Fig. 4 shows the connections between system components. In fact, all the used components are directly connected to the main element, the Raspberry Pi. Main wires are used to connect various components and circuit boards. Each component has a specific role.

The system also includes a mobile application that allows getting intrusions notifications and consulting the history of intrusions. Thus, authority can consult the history of all the intrusions that take place in different agricultural areas and for which he has been alerted.

Fig. 5 shows what the surveillance system box looks like. At the top, 3 LEDs are positioned: one to signal the powering up of the device, the second to indicate an object detection and the last to highlight data transfer to the remote database. In the middle, the Pi camera and the motion sensor can be identified. Further down, an ultrasonic sensor is positioned to calculate the distance between the system and the detected object. In the profile view, you can easily distinguish the start button and the power plug.

Fig. 6 gives in one view, the proposed system from the implementation to the deployment. It shows: a) intrusion detection system prototype, b) system mounting on 1.5 meters high support covered by a solar panel, c) system deployed in real environment. The system is powered by a 30WATT solar panel, mounted on a 12V-9AH battery. The two elements are interconnected by a charge controller.

IV. Experimental Results and Validation

The designed prototype is experimentally evaluated and validated through the system deployment along a field in order to know the accuracy of the intrusion detection. The passage of cows is simulated in a real environment. A herd of ten cows circulate in front of the device, a hundred times. Each time, how accurately the device reported the intrusion is noted. For the simulation, three significant hours of the day are chosen to perform measurements over three days: Morning on day 1, noon on day 2, and evening on day 3. Fig. 7 shows the simulation results for three different time periods selected. Fig. 8 summarizes the measurements taken statistically.

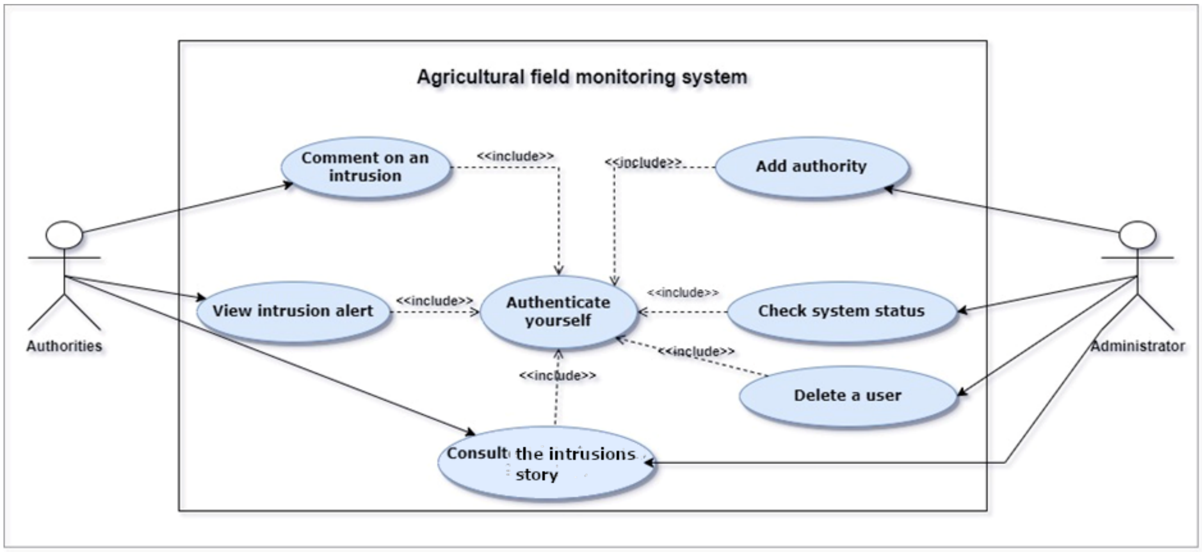


Fig. 3. System' Use Cases

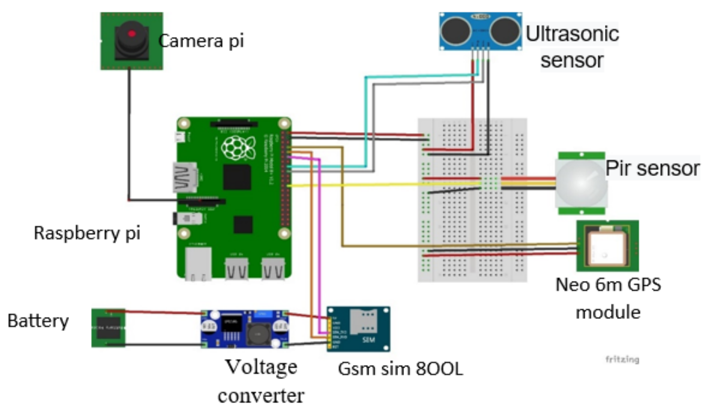


Fig. 4. Electronic circuit' wiring

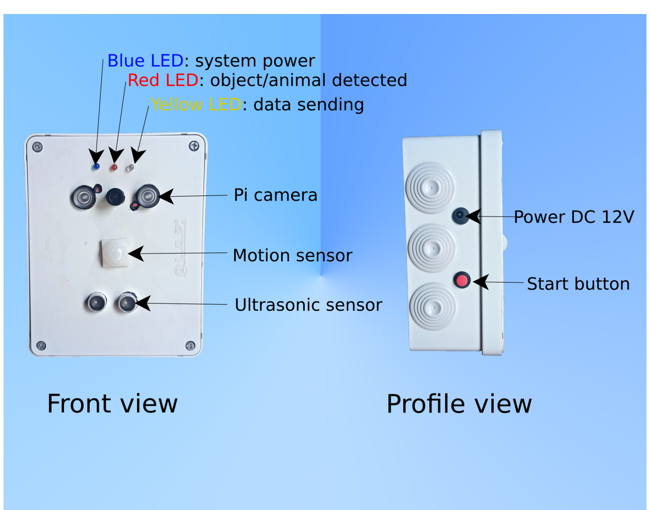


Fig. 5. Monitoring system box

Fig. 7 shows that the proposed system gives better results in the morning and noon, while Fig. 8 shows an accuracy between 0.5 and 1 with averages between 0.74 and 0.82 for morning and noon. Indeed, the evening measurements show an average of 0.53, with detection failing half the time along with a significant value of 0.29 for the standard deviation.

In conclusion, one can notice that noon measurements are more accurate than morning and evening measurements. Detection accuracy in the evening is not so high, although, the system box is equipped with a night vision camera. It turned out that this was due to the brightness and quality of the camera and the low processing power of the Raspberry Pi 3 micro-controller. Future version of the prototype should be equipped with a better camera to improve recognition quality using more powerful Raspberry Pi. This will reduce processing time for cow recognition from images sent by the camera.

V. Related Works

Many studies focused on different aspects of transhumance in Benin and worldwide. From this, it is first possible to draw some solutions about cows' movements monitoring along a specific way called corridor [15]. Secondly, some publications analyzed the effects of constantly greater cattle overflows on total biomass productivity [16]. Thirdly, other relevant research works focus on providing an IoT-based architecture to develop a Smart, Sustainable Agriculture platform as a solution [17]. Finally, some other studies discussed the importance of a good agricultural practice in contrast to forests establishment as a way to increase both rangeland' productivity and the appreciation of its use and management. From the outputs of these studies, one can be able to determine the importance of transhumance and its pipelines in Benin as well as in some countries which have the same challenges.

Endorsements have also been made about how a corridor can be materialized by identifying the factors that contribute to define transhumance routes and analyze the perception of transhumant herders on the determinants of these routes in

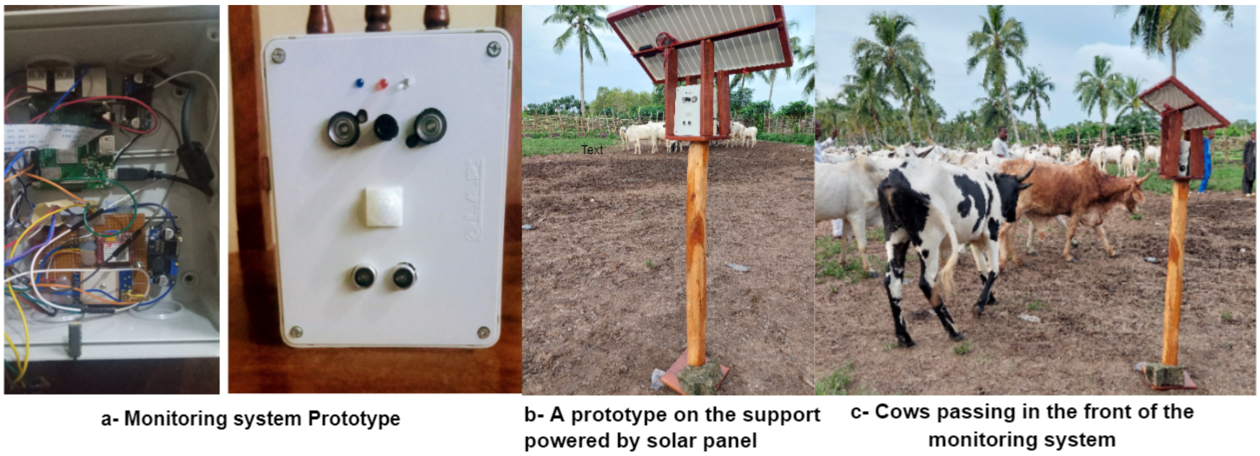


Fig. 6. Monitoring system in one view

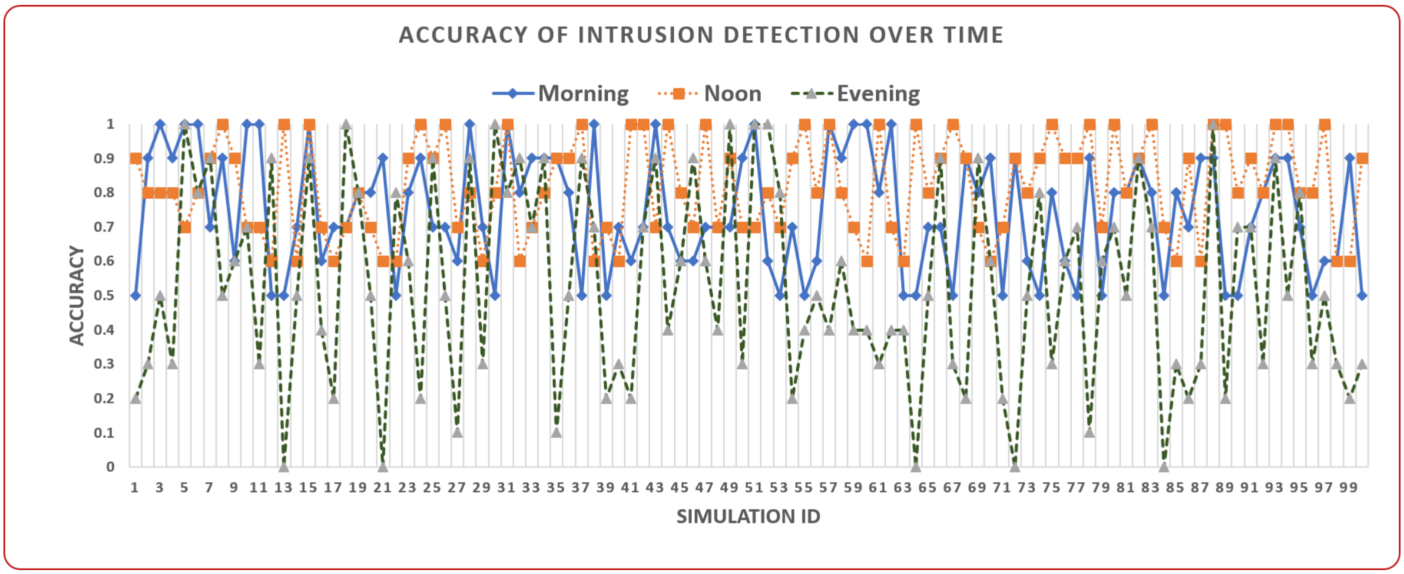


Fig. 7. Accuracy of intrusion detection over time

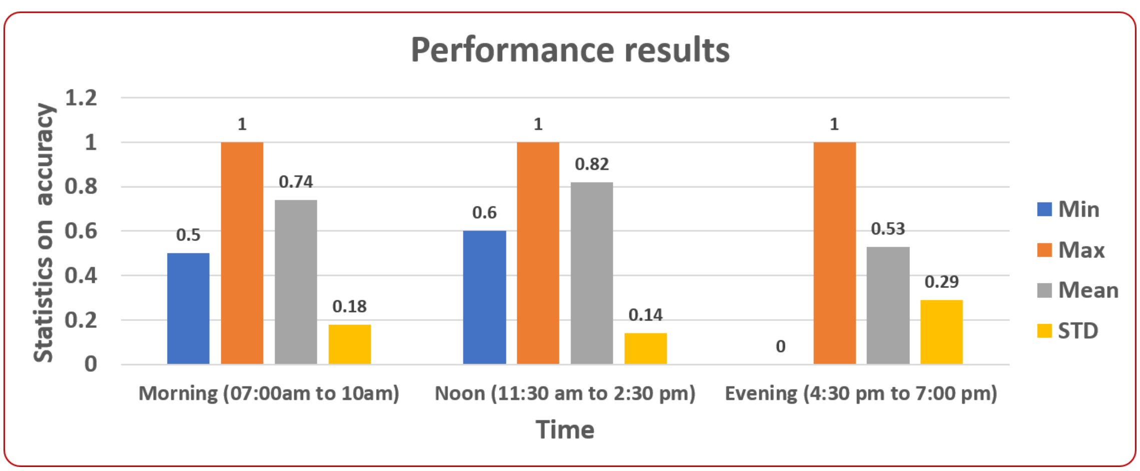


Fig. 8. Statistics on accuracy measured by the system

order to take better decisions in the management of grazed ecosystems [18], [19]. Other solutions are based on monitoring system [20] to manage selected physical conditions and behaviors of livestock across long distances and over large grazing areas in open pastures, as well as in fenced in areas [21], [22]. The solution proposed in [23] is to monitor cows using an IoT enabled sensor installed on the collar of each cow, which associated to the cow, a unique identifier. Livestock farming is assisted more and more by technological solutions such as a vision-based module that allows to automatically detect predators and distinguish them from other animals in order to prevent damages and respecting biodiversity [24]. Sensor technologies are also used to monitor animal health and ensure animal well-being in the fast changing conditions. In addition, authors of [25] try to show relevance of digital technologies for health and welfare monitoring usage and the management of livestock, kept in large pasture areas. Authors of [22] illustrate IoT technology for farmers and use sensors to gather and transfer data using parameters, like temperature, humidity and heartbeat. Once data are collected, they are sent to the Arduino Uno controller.

Moreover, authors of [26] have proposed a farmland surveillance-alert system using unmanned aerial vehicles for cattle presence detection on farmlands as a solution to curbing the problem of farm invasion and destruction. The system, upon detecting cattle presence, higher than a threshold level, sends SMS to farmer's selected number.

In [5], as a previous work, we have focused on a proposal of a multi-level architecture that is able to help through the implementation of a smart guidance system based on IoT Technologies, where herders can better control their livestock following the predefined corridors. Unlike all existing technological solutions, the current work is not intended to identify the cows before tracking them. This is explained by the fact that transhumants often do not adhere to this practice, because they fear that their livestock will be counted. Furthermore, the idea of automatically sending alerts to the farmer upon intrusion detection is rejected, as this does not resolve commonly recorded deadly conflicts. The designed system uses DNN Machine Learning algorithm, to detect intrusion and send alert to the authorities. The system achieved an average accuracy score that varies from 0.53 to 0.82 in real-life test environment.

VI. Conclusion

In this work, we succeeded in prototyping an intrusion detection system for spaces monitoring. The system consists of several modules, that each, plays an important role. The main objective of the project is to send the information collected after processing, to a remote database, specifically Firebase, when there is an intrusion in a monitored space. A test bed has been provided for livestock and agricultural domains. Upon detection, a generated alert is sent to a mobile application to allow visualization and decision-making by authorities. In future work, the prototype will be improved by putting the ultrasonic sensor on a servo motor in order to turn it in the direction of the moving object. Moreover, performance tests will be extended and will include, among other, the evaluation of the time taken to detect the intruding object.

Acknowledgment

We thank the African Center of Excellence in Mathematical Sciences and Applications (CEA-SMIA) for funding this work.

References

[1] C. Aquilani, A. Confessore, R. Bozzi, F. Sirtori, and C. Pugliese, "Review: Precision livestock farming technologies in pasture-based livestock systems," *Animal*, vol. 16, no. 1, p. 100429, 2022.

[2] A. Adnan, A. Muhammed, A. A. A. Ghani, A. Abdullah, and F. Hakim, "An intrusion detection system for the internet of things based on machine learning: Review and challenges," *Symmetry*, vol. 13, p. 1011, 2021.

[3] J. Zhang, F. Kong, Z. Zhai, S. Han, J. Wu, and M. Zhu, "Design and development of iot monitoring equipment for open livestock environment," *International Journal of Simulation: Systems, Science and Technology*, vol. 17, 2016.

[4] T. Todorov and J. Stoinov, "Expert system for milk and animal monitoring," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 01 2019.

[5] P. Houngue, R. Sagbo, and C. Kedowide, "An hybrid novel layered architecture and case study: Iot for smart agriculture and smart livestock," vol. 318 LNICST, 2020, pp. 71–82.

[6] D. N. Argade, S. D. Pawar, V. V. Thitme, and A. D. Shelkar, "Machine learning: Review," *International Journal of Advanced Research in Science, Communication and Technology*, 2021.

[7] A. A. Chaudhry, R. Mumtaz, S. M. H. Zaidi, M. A. Tahir, and S. H. M. School, "Internet of things (iot) and machine learning (ml) enabled livestock monitoring," 2020, pp. 151–155.

[8] G. Bradski, "The opencv library," *Dr. Dobb's Journal of Software Tools*, 2000.

[9] A. Zelinsky, "Learning opencv—computer vision with the opencv library," *IEEE Robotics and Automation Magazine*, 2009.

[10] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," 2014.

[11] T.-Y. Lin, G. Patterson, M. R. Ronchi, Y. Cui, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, L. Zitnick, and P. Dollár, "Coco - common objects in context," *COCO Dataset*, 2018.

[12] X. Yang, F. Li, and H. Liu, "A survey of dnn methods for blind image quality assessment," *IEEE Access*, vol. 7, pp. 123 788–123 806, 2019.

[13] R. Ravindran, M. J. Santora, and M. M. Jamali, "Multi-object detection and tracking, based on dnn, for autonomous vehicles: A review," *IEEE Sensors Journal*, 2021.

[14] R. V. C. Diogo, L. H. Dossa, S. F. U. Vanvanhossou, B. D. Abdoulaye, K. H. Dosseh, M. Houinato, E. Schlecht, and A. Buerkert, "Farmers' and herders' perceptions on rangeland management in two agroecological zones of benin," *Land*, 2021.

[15] P. Lesse, M. R. Houinato, J. Djenontin, H. Dossa, B. Yabi, I. Toko, B. Tente, and B. Sinsin, "Transhumance en république du bénin : états des lieux et contraintes," *International Journal of Biological and Chemical Sciences*, vol. 9, 2016.

[16] J. Ellison, K. Brinkmann, R. V. Diogo, and A. Buerkert, "Land cover transitions and effects of transhumance on available forage biomass of rangelands in benin," *Environment, Development and Sustainability*, 2021.

[17] E. Alreshidi, "Smart sustainable agriculture (SSA) solution underpinned by internet of things (IoT) and artificial intelligence (AI)," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019.

[18] B. O. K. GADO, I. T. IMOROU, O. AROUNA, H. S. IMOROU, and M. OUMOROU, "Déterminants des itinéraires de transhumance à la périphérie de la réserve de biosphère transfrontalière du w au bénin," *Journal of Applied Biosciences*, 2020.

[19] B. O. K. Gado, I. T. Imorou, O. Arouna, and M. Oumorou, "Caractérisation des parcours de transhumance à la périphérie de la réserve de biosphère transfrontalière du w au bénin," *International Journal of Biological and Chemical Sciences*, 2020.

[20] V. M. T. Aleluia, V. N. G. J. Soares, J. M. L. P. Caldeira, and A. M. Rodrigues, “Livestock monitoring: Approaches, challenges and opportunities,” *International Journal of Engineering and Advanced Technology*, 2022.

[21] E. S. Muhamed, M. Ting, and S. L. M. Belaidan, “Livestock health monitoring using iot technology for ethiopia,” *Journal of Advanced Research in Dynamical and Control Systems*, 2019.

[22] K. Shah, K. Shah, B. Thakkar, and M. H. Amrutia, “Livestock monitoring in agriculture using iot,” *International Research Journal of Engineering and Technology*, vol. 6, 2019.

[23] J. O. Isaac, “Iot - livestock monitoring and management system,” *International Journal of Engineering Applied Sciences and Technology*, 2021.

[24] V. del Castillo, L. Sánchez-González, A. Campazas-Vega, and N. Strisciuglio, “Vision-based module for herding with a sheepdog robot,” *Sensors*, vol. 22, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/14/5321

[25] A. Herlin, E. Brunberg, J. Hultgren, N. Högberg, A. Rydberg, and A. Skarin, “Animal welfare implications of digital tools for monitoring and management of cattle and sheep on pasture,” *Animals*, 2021.

[26] O. A. Adegbola, I. D. Solomon, and A. S. Oluwaseun, “A remote surveillance system for the detection of farmland invasion by cattle in sub-saharan african country,” *International Journal of Research and Review*, 2021.

Measurement Tool for Exposure Techniques in X-ray Ionizing Radiation Equipment

Edwin Arley Cortes Puentes[1], Andres Gomez Rodriguez[2], Fernando Martinez Santa[3]
Universidad Distrital, Francisco Jose de Caldas
Bogotá, Colombia

Abstract*—This article shows the development of an instrument for measuring the exposure parameters used to take radiographic studies in living beings; such as kilovoltage, current and time, since radiation protection is a fundamental pillar in the care of patients and operators of ionizing radiation equipment, it is necessary to calibrate these parameters in equipment that produce X-rays. For the manufacture of the measuring instrument is used an ESP32 microcontroller which is programmed using the Python syntax using the project micropython, in addition to current, distance and light sensors. The results of these measurements will be displayed through output devices such as organic light-emitting diode (OLED) displays, liquid crystal (LCD) displays, and a Web server, in order to perform the measurements safely from the control room and thus avoid exposure to radiation as much as possible. The kVp measurement performed in this article is for equipment operating at 60 [Hz], for high frequency equipment a new parameterization must be performed in order to obtain results as close to reality as possible. By using the web server for the transmission of measurement data, the radiation exposure was reduced and the calibration times of the equipment were improved. This article presents the measurements, and also the calculation of the error of each of the different exposure parameters of conventional X-ray equipments, such as kVp, mA, mAs and time. The errors obtained in the measurements were made assuming that the X-ray equipment used has a 0 error, i.e. assuming that the X-ray equipment is calibrated and that it is a standard equipment.

Keywords*—*Voltage; current; X-Ray; kVp; mA; mAs radiation; ESP32; OLED microcontroller

I. Introduction

The discovery of X-rays dates back to 1895 in the city of Hamburg, Germany and is due to experiments conducted by the physicist Wilhem Conrad Röntgen, this contribution has been of great importance and usefulness in both industry and medicine. Radiation is emitted energy that is transferred through space with or without influence on the atomic structure of matter, it is classified into non-ionizing radiation and ionizing radiation according to the effects produced by corpuscular contact [1]. Non-ionizing radiation includes ultraviolet (UV), infrared (IR) and microwave radiation. In cells, the possibility that they can generate heat decay has been considered, but it is not yet known whether they can generate microscopic effects. Ionizing radiation includes X-rays, gamma, alpha and beta rays. These types of radiation are capable of creating damage or not in human cells [2]. In our body, ionizing action is evidenced by chromosome breaks, where changes may include consequent deletions or abnormal translocations, these effects can be seen during cell division resulting in abnormal cell development or death. The action of X-rays on sex cells can generate alterations in the transmission of hereditary characters known as mutations. X-rays are electromagnetic radiation generated by the excitation of electrons in the internal orbit of an atom, with the ability to pass through opaque bodies. The wavelength range of X-rays is between 0.01 nm and 10 nm, which corresponds to an energy range of approximately 1 keV to 150 keV.

X-rays have been used extensively by humans in the fields of industry, veterinary medicine and medicine. In medicine they are used to obtain diagnostic images. As an example of its use we have panoramic and periapical equipment. Panoramic radiography is used to obtain a single image of the teeth and their supporting structures. Periapical radiography shows a more detailed image of the dental structure, it is used to explore the dental structure and the surrounding tissue. [3], mammograms: Mammography is one of the most important methods for detecting breast lesions. By means of this technique small anomalies can be identified which, if prevented in time, can be treated in a better way [4]. TACS “Computerized Axial Tomography” is a method that uses imaging technology for medical diagnosis, it allows to analyze the interior of the human body through millimetric transversal cuts to the cephalo-caudal axis, all this is done through the use of X-rays [5], C-arms, fixed and mobile X-ray equipment (they are used to obtain general X-rays of organs and bones), among others; These types of examinations, being non-invasive, represent a vital instrument for the analysis, study and care of animal and human health. Due to the imminent risk presented when using X-rays on living beings, it is essential to supervise and control the doses administered for such examinations.

However, in the process of diagnostic imaging, X-rays are used, which are dangerous, as it has been shown that they can generate negative eventualities to people. For this reason, and according to one of the pillars of radiation protection “as low as reasonably achievable” known as optimization, this indicates that the control elements of these devices must have the highest quality standards to obtain a good quality image with the lowest radiation dose [6].

Embedded systems work based on a processor, microprocessor, microcontroller or programmable logic devices, usually perform repetitive functions [7], therefore, we can find a variety of applications in practically all fields, from applications in bioengineering [8], as well as image processing [9], to industrial and educational levels as applications in industrial preventive maintenance [10] and control of didactic robots, respectively [11].

Since the beginning of mankind, human beings have sought their welfare in many ways, therefore any advance in technology has had approaches in medicine, embedded systems have not been alien to this, have been developed applications as simple as drug dispensers [12] and others a little more complex, as is the control of vital signs [13] to the physiological emulation of the lung [14].

Speaking specifically of applications focused on everything related to X-rays, we can find various studies such as FPGA-based ionizing radiation detection [15], sample alignment devices for the diffractometer [16].

Thanks to all these advances in electronics in terms of sensors and embedded systems applied to the medical field [17] , it is possible to perform increasingly faster and more precise measurements of X-ray exposure parameters (kVp,mAs,t) [18].

It is then intended to implement a system for measuring the exposure parameters kVp, mAs and time based on an ESP32 microcontroller, this in order to monitor and control the calibration of X-ray equipment and ensure the doses delivered to patients, thus this article is organized as follows: Section I gives an introduction on the parameters of exposure, the importance of X-rays and how embedded systems have been used in the area of medicine, Section II presents the design, which takes into account the electrical and mechanical characteristics that must be taken into account for the design of the same, Section III shows the data obtained from the measurements of the measurement system implemented, finally, Section IV presents some conclusions and recommendations that should be taken into account for future research.

II. Design

This section explores the electrical, mechanical and electronic characteristics to be taken into account for the development of the measurement equipment, such as connectivity, power, voltage range, current range, time, materials, among others.

A. Measurement of Exposure Parameters

When a specific radiographic study is required, it is necessary to vary the parameters of peak kilovolts $[kVp]$, current as a function of time $[mAs]$ and time $[t]$, because each patient has a different contexture, and each bone or organ of the body requires these parameters to be varied in order to obtain a good quality image that allows an accurate diagnosis.

1) Exposure parameters influence on the acquired image: Each exposure parameter varies the quality of the image to be acquired as follows:

kVp= Penetration ”More or less opaque an organ” The higher the kVp applied, the higher the penetration of the X-rays and vice versa. mAs= Definition ”Contrasts, grays” Allows to define the borders of the organs and to be able to distinguish between them [19]. t= Radiation exposure time. The longer an X-ray beam is the more penetration can be obtained in the image, however this parameter should be as low as possible without sacrificing the definition of the image, since the shorter the radiation exposure time of a patient the lower the risk of affecting the health of individuals.

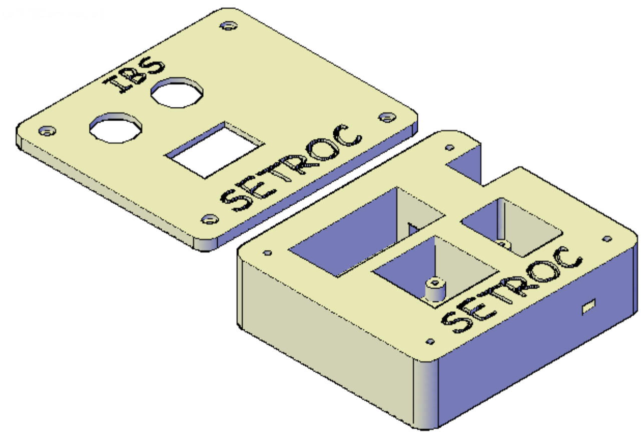


Fig. 1. kVp sensor box

B. Current Measurement

Conventional X-ray equipment generates a current in the range of $25[mA] to 500[mA]$, periapical and panoramic equipment uses lower values in the range of $3[mA] - 15[mA]$, the current in mammographs ranges from $20[mA] - 150[mA]$.

1) Current sensor INA219: The INA219 sensor allows current measurements from 0 to 400 [mA] and uses an I2C communication protocol compatible with the ESP32 microcontroller, which is why it is the current sensor module used by the measurement device proposed in this article.

C. High Voltage Measurement

Conventional X-ray equipment generates a voltage in the range of $30[kV] to 150[Kv]$ with variations of $1[kV]$, mammographs operate between $20[kV]$ to $35[kV]$ because air at these voltages is ionized [20] To measure these voltage magnitudes, there is specialized equipment such as a blider or voltage dividers that allow an indirect measurement [21]. The phosphor screens used to take radiographs emit an intensity of light that is proportional to the kVp applied, this feature is used so that the measuring instrument designed in this article obtains the measurement of [kVp] indirectly.

1) Light sensor for indirect kVp measurement: The *TSL-2591* sensor allows the measurement of the visible and infrared light spectrum separately, taking advantage of the light emanating from the phosphor screens and the sensor at the time of exposure, obtaining a digital constant that will be greater the higher the kVp applied and vice versa.

2) Sensor box design: Since the wavelength of light to be measured is in the visible light range, it is necessary to manufacture a box where the sensor and the phosphor screen are isolated from the light signals that may interfere with the measurement process. For this, the design is made in the AutoCAD software Fig. 1 and the .stl file is obtained, in this kVp sensor box is housed the light sensor, the phosphor screen that performs the measurement of kVp indirectly, plus a distance sensor installed in order to perform all measurements at the same distance and an OLED screen as user interface.

D. Time Measurement

Since the radiation from this equipment is ionizing, the exposure time is a parameter to be controlled, since the less time without sacrificing image quality the less radiation a patient will receive, it is necessary to measure the exposure time.

1) ESP32 internal timer: To perform the time measurement is used the *timer0* of the microcontroller *ESP32* that will use the signal emitted by the sensor *TSL2591* to start and end the trigger time count. This measurement is performed in the range of [ms], by tests performed the microcontroller can measure signals in the order of 15 [us].

E. Distance Measurement

Since the distance between the X-ray tube and the phosphor plate can be a factor in the measurement error, a distance sensor is used to ensure that the data acquisition is always at the same distance.

1) HCSR04 sensor: The HCS04 distance sensor is an ultrasonic type sensor that provides the possibility to measure distance in a range from 0 to 4 m with a sensitivity of 1 cm. It is installed next to the Kvp sensor in order to ensure that all measurements are made at the same distance and thus guarantee repeatability of the measurements made.

F. 20X4 OLED and LCD Display

After data acquisition, it is necessary to exchange information with the user, so that the user can acquire measurement information. Due to the practicality, size, low power consumption and connectivity with the ESP32, a 0.96" OLED screen will be used to display results on the kVp measurement device. In addition, the main control will use a 4x20 LCD coupled to an I2C module which facilitates the connection to the ESP32 for the same purpose of being a user interface.

G. Web Server

Using the Wifi connectivity provided by the ESP32 microcontroller, the measurement data obtained with the sensors is transmitted via a web server to a cell phone or laptop connected to the same network as the microcontroller. [22] This reduces the exposure to radiation and speeds up the verification and calibration process of the X-ray equipment, since the data transmission will be instantaneous via wifi.

H. ESP32

For the development of the measurement tool a development board is used that has integrated the ESP32 family microcontroller that are manufactured by espressif, because this family of microcontrollers has bluetooth and wifi connection [23], By having wifi it is possible to create a webserver for the exchange of information ESP32-Mobile device, which is an advantage, since this allows to reduce the risk of exposure to X-rays by being able to exchange information from the control room of the X-ray equipment with the study room where the measurement equipment is located. Communications with the sensors and output peripherals are handled by I2C communication protocol [24]. Fig. 2 shows a general diagram of the communications between the sensors, the microcontroller and the output peripherals. The flowchart shown in Fig. 3 shows the programming logic of the ESP32.

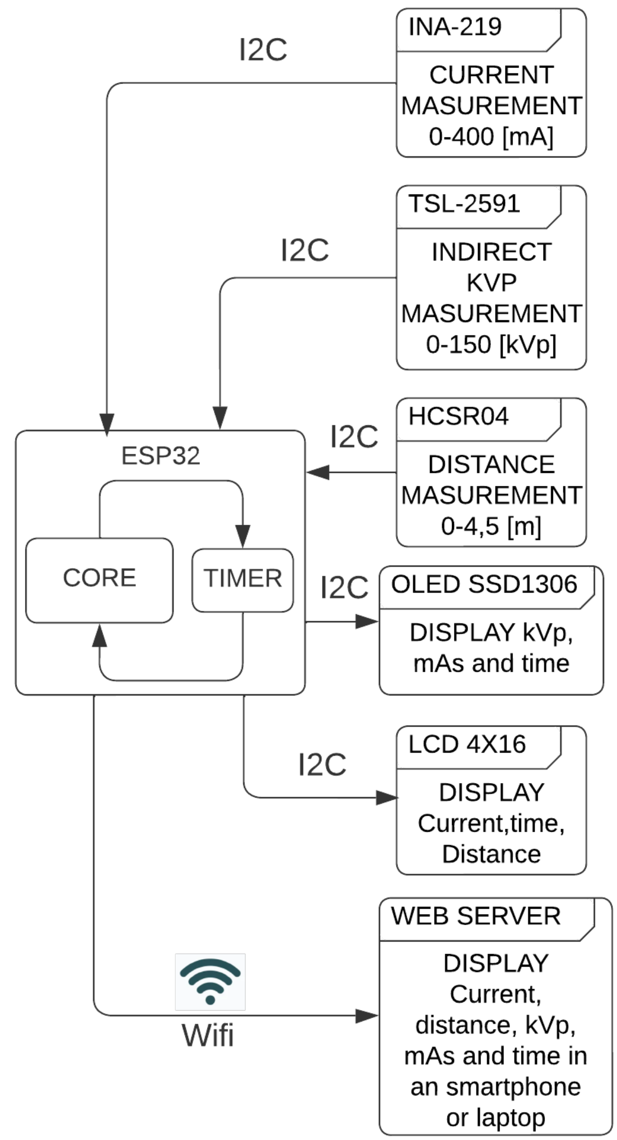


Fig. 2. General operating diagram

1) Control box design: In the control box is the current sensor (INA219), the LCD 4x16 and the ESP32 microcontroller, in turn is connected to the other sensors and peripherals output via I2C communications [25] with the sensor box kVp, ESP32 processes the data obtained by the sensors and output signals are issued to the OLED, LCD and web server. For the design of the control box the Autocad software is used and the .DWG file is obtained, with which the cuts are made on a laser cutting machine, the cuts are shown in Fig. 4 The design of the box is shown in the Fig. 5.

III. RESULTS

This section shows the results obtained from the measurements made with the measurement equipment to an Ameri-

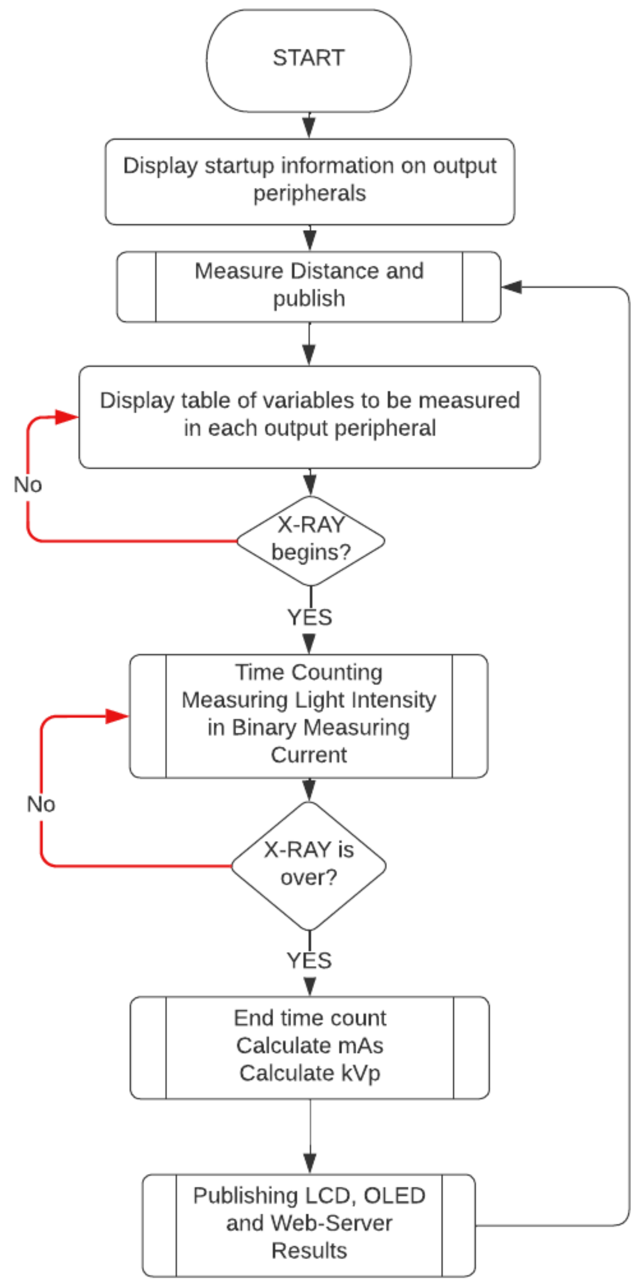


Fig. 3. General operating flow diagram

comp X-ray equipment model TXR-325, which operates in kVp from $30[kV]to150[Kv]$, in current from $0to500[mA]$ and in time from 0 to 5 [s], also are shown in this section the prototypes manufactured for the control box and sensor box kVp.

A. Case fabrication on 3D printer

Using a 3D printer, the kVp sensor box shown in the Fig. 6 is manufactured. First the design was made using the autoCAD software, the measures are 9x8x4 cm, once finished, it was exported to the 3D printer. The figure is recreated in a plastic, known as polylactic acid or PLA.

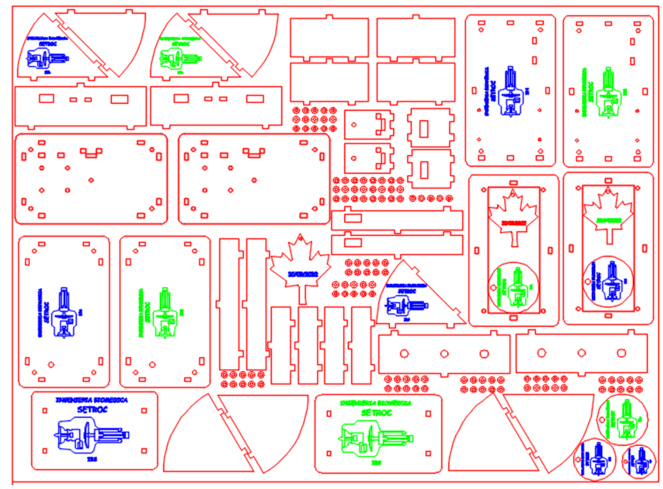


Fig. 4. Laser cuts

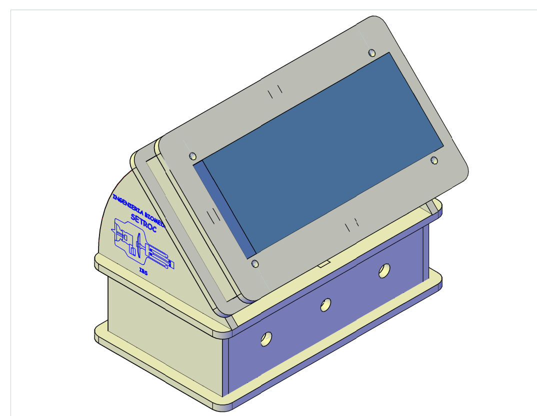


Fig. 5. Box Control

B. Laser Cutting Machine Control Box Manufacturing

Using a laser cutter, the kVp control box shown in the Fig. 7 was manufactured. First, the design was made using the autoCAD software, it was generated in 2D, the measurements are 7x12x3.5 cm, once finished, it was exported to the laser cutter. The figure is recreated in a 3 mm thick acrylic.

C. measurements Performed

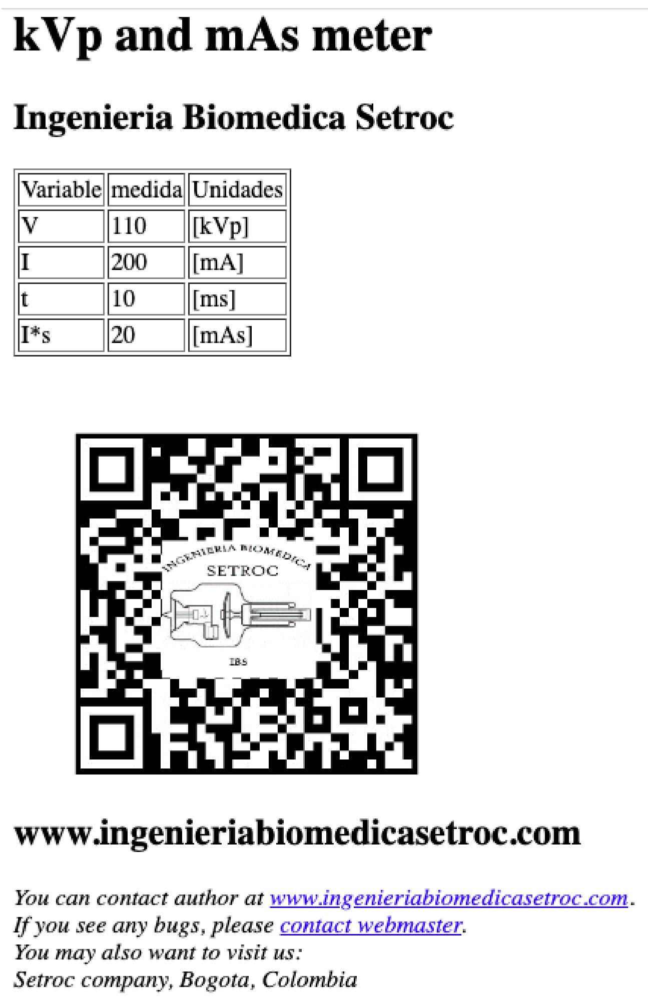
The Fig. 8 shows the complete measurement prototype, with which the measurements presented in this article were performed, the results are displayed on the OLED, LCD and a Web server that displays the results; the Fig. 9 shows an example of how the data is viewed from the control room on the Web server from a device that can be a laptop or a Smartphone, thus the measurement is performed safely without exposure to radiation.

Since there is no standard equipment, the measurements presented in the article are made assuming an error of zero, i.e. assuming that the Americomp model 325 X-ray equipment on which the tests were performed is the standard equipment and it is assumed that it is calibrated.

Fig. 6. Prototype box sensor

Fig. 7. Control box prototype

Fig. 8. Measuring equipment prototype



kVp and mAs meter

Ingenieria Biomedica Setroc

Variable	medida	Unidades
V	110	[kVp]
I	200	[mA]
t	10	[ms]
I*s	20	[mAs]

www.ingenieriabiomedicasetroc.com

You can contact author at www.ingenieriabiomedicasetroc.com.
If you see any bugs, please contact webmaster.
You may also want to visit us:
Setroc company, Bogota, Colombia

Fig. 9. Information published on web server

Similarly, the mean absolute error is calculated for each measured variable, using the formula in equation 1.

$$MAE = \frac{1}{n} * \sum_{i=1}^{n} |Y_i - X_i| \tag{1}$$

The analyzed measures are as follows:

1) Time measurement: .

The time measurement is performed non-invasively by taking the TSL2591 sensor signal as the trigger start and termination signal for the ESP32 internal timer that keeps track of the trigger duration time. The results of the measurements performed are shown in the Table I, where t_s is the set time and t_m is the time measured by the ESP32 internal timer.

The mean error for the calculated time is shown in equation 1, taking ts as Yi and tm as Xi.

MAE =2 ms

2) Current Measurement: The current measurement is performed invasively by placing the INA219 sensor in series with

TABLE I. TIME MEASUREMENTS

t_s	t_m	$Error$
100[ms]	101[ms]	1%
200[ms]	199[ms]	0.5%
300[ms]	299[ms]	0.33%
400[ms]	402[ms]	0.5%
500[ms]	501[ms]	0.2%
800[ms]	803[ms]	0.37%
1000[ms]	1005[ms]	0.5%

the secondary of the high voltage transformer and communicating the measurement to the ESP32 via I2C communication. The results of the measurements are shown in the Table II, where I_s is the set current and I_m is the measured current.

TABLE II. CURRENT MEASUREMENTS

I_s	I_m	$Error$
50[mA]	51[mA]	2%
55[mA]	53[mA]	3, 64%
60[mA]	57[mA]	5%
70[mA]	69[mA]	1, 43%
80[mA]	78[mA]	2, 5%
90[mA]	91[mA]	1, 1%
100[mA]	99[mA]	1%
150[mA]	147[mA]	2%

The calculated mean current error is shown in equation 1, taking Is as Yi and Im as Xi.

MAE =1,75 mA

3) mAs Measurement: Having the current and time measurements we can find the measured *mAs* using the equation 2, where I_m and t_m are the results of the current and time measurements respectively, the results obtained from the calculated mAs measurements are shown in the Table III where mAs_s is the $mAs_{selected}$ and $mAs_m$is the$mAs_{measured}$.

$$mAs = I_m * t_m \tag{2}$$

TABLE III. MAS MEASUREMENTS

mAs_s	mAs_m	$Error$
50[mAs]	51[mAs]	2%
100[mAs]	104[mAs]	4%
200[mAs]	203[mAs]	1, 5%
300[mAs]	302[mAs]	0, 67%
400[mAs]	397[mAs]	0, 75%
500[mAs]	496[mAs]	0, 8%

The mean error for the calculated mAs is shown in equation 1, taking mAss as Yi and mAsm as Xi.

MAE =2.83 mAs

Fig. 10 shows a diagram of the general operation of the kVp meter. The kVp measurement is done indirectly, to find the kVp value the microcontroller makes use of the digital value in binary number delivered by the visible light sensor and making use of the equation 3 calculates the measured kVp value. In the equation 3, N_{binary} is the binary number read by the sensor at the time of exposure and K is the constant found through testing. The results of the kVp measurements performed are shown in the Table IV where kVp_s is the kVp set in the X-ray equipment, kVp_m is the kVp measured with the measuring equipment.

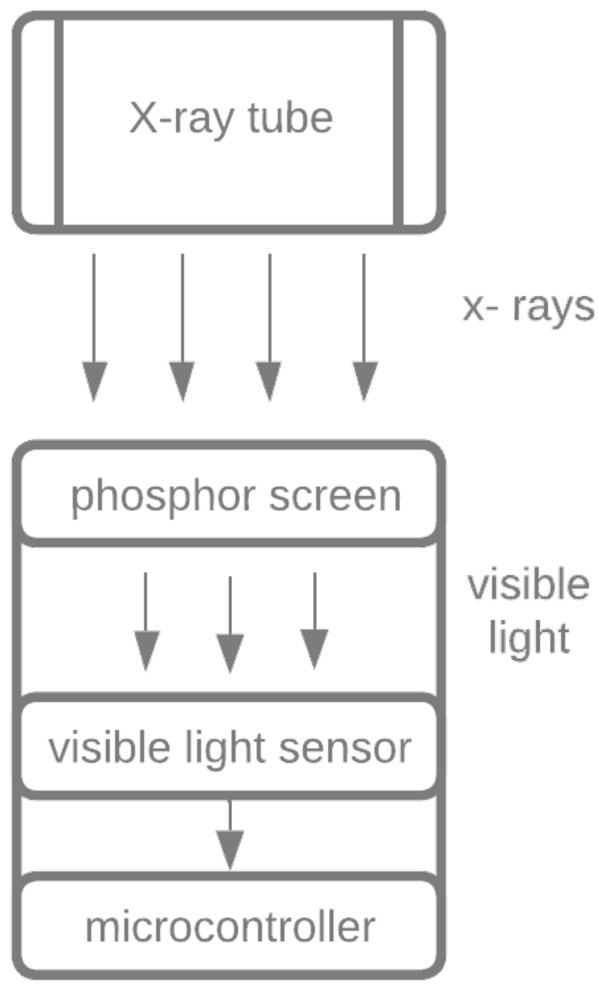


Fig. 10. General operation of kVp meter

$$kVp = N_{binario} * K \tag{3}$$

TABLE IV. KVP MEASUREMENTS

kVp_s	kVp_m	$Error$
50[kVp]	51[kVp]	2%
55[kVp]	53[kVp]	3, 64%
60[kVp]	57[kVp]	5%
70[kVp]	69[kVp]	1, 43%
80[kVp]	78[kVp]	2, 5%
90[kVp]	91[kVp]	1, 1%
100[kVp]	99[kVp]	1%
150[kVp]	147[kVp]	2%

The mean error for the calculated kVp is shown in equation 1, taking KVps as Yi and KVpm as Xi.

MAE =1,75 kVp

4) Distance measurement: The distance measurement made by the ultrasonic sensor between the collimator and the fabricated sensor is compared with the measurements taken with a flexometer, the results of the measurements made are shown in the Table V, it should be noted that the kVp measurements were made at a distance of 1[m].

TABLE V. DISTANCE MEASUREMENTS

$Distance$	$Distancie_m$	$Error$
10[cm]	9, 8[cm]	2%
20[cm]	19, 9[cm]	0, 5%
35[cm]	35, 1[cm]	0, 28%
50[cm]	49, 7[cm]	0, 60%
80[cm]	80, 2[cm]	0, 25%
90[cm]	89, 9[cm]	0, 11%
100[cm]	99, 8[cm]	0, 2%

The mean error for the calculated distance is shown in 1, taking distance as Yi and measured distance as Xi.

MAE =0,17 cm

The measurement instrument designed allows to have an impact in favor of the health of the operators and users of the machines in charge of generating the X-rays, due to the fact that the exposure to radiation decreases when carrying out the measurements in short times and also through of the webserver tool and an IP to transmit and view them in a control booth.

IV. LITERATURE REVIEW

The articles used as reference in this research are in common agreement on how important it is to minimize the time and distance of exposure to X-rays, finally, we can find concepts, methodologies and proposed applications focused on different areas such as biomedicine for means of embedded systems.

That is why applications have been proposed that seek to be less exposed to these rays, such as the one created through Atmega328p, with the limitation that it is only for Android devices, since it is implemented under an application or the one designed by means of the intel386EX for the real-time communication system of a moving vehicle but only for a closed network.

V. CONCLUSIONS

The impact of the media system designed in the field of medicine is very important and improves because it allows monitoring the X-ray equipment from a mobile device keeping a prudent distance from it, which generates a positive benefit for patients and operators, the first they could have a more optimal and efficient team, and the latter would avoid the recurring exposure to which they are immersed

It is common knowledge that exposure to X-rays generates short, medium and long term damage to health, therefore it is of vital importance to seek solutions that allow users and operators to mitigate these health consequences, that is why from our branch, we thought and implemented a measurement system that can be operated and monitored remotely.

The specification of the controller is vital for the desired application, in this case an ESP32 was selected, as it has the advantage of connecting to WIFI natively, which makes it ideal for Internet of Things (IoT) type projects such as the one developed, likewise its low cost is another added value.

This measuring equipment allows measurements in conventional equipment operating at 60 [Hz], since in high frequency equipment additional factors must be considered. By the fact that the tests showed the importance of having sensors with a high response speed, because to sample a signal with a frequency of $60[Hz]$, you must have a sensor module that performs the measurement and makes the transmission of information at least $2[ms]$, in order to reconstruct a signal.

For this type of measurements it is important to have sensors with a high response speed, since it is necessary to sample signals in the order of [ms], in the implementation of the proposed measurement system was one of the problems presented, because it was not possible to generate the expected signal, simply obtaining measurement points that did not indicate anything, the solution was a sensor with higher speed performance, thus the rectified signal was acquired.

As a future improvement for the proposed media system, it is recommended to use a sensor that is faster in terms of data processing, as this could make the data acquisition for the kVp variable more accurate.

For future research where it is required to implement a control or measurement prototype, it is recommended to have the appropriate protective equipment for work with X-rays.

It is important to keep the traceability of the error of the designed measurement system, for this, it is proposed as an improvement for the future, to use a standard equipment, since with a calibrated instrument we can know the real error of the equipment, and then compare it with our system.

ACKNOWLEDGMENT

I thank the ARMOS research group of the Universidad Distrital Francisco José de Caldas, especially the engineer Fernando Martinez Santa who as a teacher directed the project and also the company Ingenieria Biomedica Setroc for allowing us to run experimental tests under the direction of Daniel Cortes who provided his knowledge and experience in the area for the realization of this project.

REFERENCES

[1] M. V. D. Villca, “Historia de la radiología,” *Revista de actualización clínica Investiga*, vol. 37, p. 1787, 2013.

[2] A. Buzzi, “El descubrimiento de los rayos x,” *Buenos Aires, Argentina*, 2012.

[3] C. T. Venegas, “Historia de la medicina: un cincuentenario.–el descubrimiento de los rayos x,” *Revista de la Facultad de Medicina*, vol. 14, no. 1, pp. 60–63, 1945.

[4] F. Marquez, N. Acosta, T. Benavente, and M. Poma, “Dosis en mamografía: Estudio preliminar,” *Revista de Investigación de Física*, vol. 5, no. 01-02, 2002.

[5] C. Pereira, “Tomografía axial computada,” *XIII Seminario de Ingeniería biomédica*, 2004.

[6] W. S. Castrillón-Giraldo, J. M. Aramburo, and W. Jaramillo-Garzón, “Control de calidad en equipos de rayos x en intervencionismo,” *Revista Colombiana de Cardiología*, vol. 27, pp. 88–95, 2020.

[7] D. Pérez, “Sistemas embebidos y sistemas operativos embedidos,” *Lecturas en ciencias de la computación. Universidad Central de Venezuela, Vols.% i de% 2lSSN*, pp. 1316–6239, 2009.

[8] P. A. García, “Sistemas embebidos de tiempo real con aplicaciones en bioingeniería,” Ph.D. dissertation, Universidad Nacional de La Plata, 2019.

[9] L. Sugezky, M. Prieto, N. González, M. Giura, Y. Kuo, M. Trujillo, and J. Cruz, "Desarrollo e implementación de herramientas de simulación de modelos para sistemas embebidos," in *Congreso Argentino de Sistemas Embebidos*, 2016.

[10] W. Hasperué, C. A. Estrebou, G. Camele, P. López, M. Peña, R. Zambrano, L. C. Lanzarini, A. Fernández Bariviera, and M. Cerrada, "Procesamiento inteligente de la información: aplicaciones en bioinformática, trayectorias vehiculares, mantenimiento preventivo industrial y sistemas embebidos," in *XXIV Workshop de Investigadores en Ciencias de la Computación (WICC 2022, Mendoza)*, 2022.

[11] D. Azcurra, S. Rojo, and D. Rodríguez, "Arquitecturas de control para robots autónomos móviles didácticos basadas en sistemas embebidos," in *XIV Workshop de Investigadores en Ciencias de la Computación*, 2012.

[12] P. M. B. Bastidas, J. D. C. Vallejo, C. A. L. Narváez, D. M. Torres, and R. G. M. Perafán, "Dispensador automático de medicamentos sólidos para pacientes mediante uso de sistemas embebidos," *Boletín Informativo CEI*, vol. 7, no. 3, pp. 88–94, 2020.

[13] W. R. Ruiz Sánchez *et al.*, "Redes de sensores inalámbricos enfocadas a la medicina con énfasis en control de los signos vitales en pacientes adultos mayores," Master's thesis, PUCE, 2016.

[14] J. S. Agudelo Díaz and O. A. Cardoza Olano, "Emulación de la fisiología respiratoria del pulmón empleando un sistema embebido." 2022.

[15] C. R. A. N. LAVEAGA, D. A. A. F. JARAMILLO, and D. C. D. GALVÁN, "Desarrollo de un sistema embebido para detección de radiación ionizante para seguridad civil," 2016.

[16] D. S. Jiménez-Novoa, J. S. González-Bustamante, and W. Aperador-Chaparro, "Dispositivo de alineación de muestras para el difractómetro de rayos x con control de posición e interfaz de manipulación," *Respuestas*, vol. 22, no. 1, pp. 84–95, 2017.

[17] S. Widadi, M. K. Huda, I. Ahmad *et al.*, "Atmega328p-based x-ray machine exposure time measurement device with an android interface," *Journal of Robotics and Control (JRC)*, vol. 1, no. 3, pp. 81–85, 2020.

[18] J. Gázquez, F. Guillen, N. Novas, and M. Berenguel, "Diseño e implementación de un sistema de comunicación en tiempo real y tolerante a fallos para monitorización médica en vehículo en movimiento," *XXV Jornadas de Automática*, 2004.

[19] O. Lakhwani, V. Dalal, M. Jindal, and A. Nagala, "Radiation protection and standardization," *Journal of clinical orthopaedics and trauma*, vol. 10, no. 4, pp. 738–743, 2019.

[20] D. B. Silva, M. A. Santos, F. R. Barros, L. A. Santos *et al.*, "High precision instrumentation for measuring the true exposure time in diagnostic x-ray examinations," 2013.

[21] F. J. R. Campos, "Prácticas de alta tensión para fines didácticos en ingeniería," *Ingeniería e Investigación*, no. 13, pp. 45–54, 1986.

[22] L. García, J. M. Jimenez, J. Lloret, and P. Lorenz, "Wifi and lora energy consumption comparison in iot esp 32/sx1278 devices," in *SMART 2019, The Eighth International Conference on Smart Cities, Systems, Devices and Technologies*. IARIA XPS Press, 2019, pp. 26–31.

[23] I. Allafi and T. Iqbal, "Design and implementation of a low cost web server using esp32 for real-time photovoltaic system monitoring," in *2017 IEEE electrical power and energy conference (EPEC)*. IEEE, 2017, pp. 1–5.

[24] E. J. Carletti, "Comunicación-bus i2c," *Robots Argentina*, 2007.

[25] P. Bertoleti, *Proyectos com ESP32 y LoRa*. Editora NCB, 2019.

Encrypted Storage Method of Oral English Teaching Resources based on Cloud Platform

Tongsheng Si
School of Culture-Tourism and International Education, Henan Polytechnic Institute, Nanyang, China

***Abstract*—With the development of the times, the secure storage of educational resources has become one of the key security problems faced by colleges and universities. On the one hand, the cost of traditional resource storage is too expensive, on the other hand, its encryption and access efficiency are low. To solve this problem, this research takes the cloud platform serves as the main carrier for the encrypted storage of school teaching resources. On this basis, the convolutional neural network is encrypted and optimized, and the argmax algorithm is improved to improve the access efficiency of encrypted data. Finally, the effectiveness and superiority of the design method are compared and analyzed through the method of performance detection. The results show that the maximum consumption time of encryption and decryption of the encrypted storage model is no more than 20000ms, which is significantly less than that of the traditional model. The running time of the argmax output encryption module is 1.76ms and the running loss is 0.26 MB, which is less than that of the traditional model. It can be seen that the encrypted storage model has stronger encryption performance and access performance, and has a better application effect in the encrypted storage of oral English teaching resources with a large amount of access data and frequent updates.**

Keywords—Cloud platform; oral language; encryption; resource storage

I. Introduction

With the development of modern network technology, the cloud platform, as a new cloud service model, can provide more secure and efficient services for users who need computing and storage services through the deployment of a cloud machine learning model [1]. The development of cloud platform technology provides customers with services that need a high budget and high-tech support in the actual environment more conveniently and economically. Encrypted storage is one of the main types [2]. For modern colleges and universities, the security of teaching resources is one of the main security problems in the teaching process. Oral English teaching resources themselves have the characteristics of fast replacement speed and a large amount of stored data. As a network grafting service platform, the cloud platform has a stronger fit with the encrypted storage of oral English teaching resources [3]. As a special resource containing multilingual voice and text contrast materials, oral teaching resources are diversified and integrated. At the same time, the resource reserve is large and the real-time update speed is fast. Therefore, the encryption system for oral teaching resources needs higher data processing efficiency and a more robust and easy architecture. Applying cloud platform encrypted storage technology to the encrypted storage of oral English teaching resources can effectively improve the economy and quickness of the encrypted storage of school teaching resources, and provide teachers and students with better teaching and learning experience while improving security [4]. This research applies convolutional neural network, a robust deep learning model, to cloud platform encryption. On the one hand, it provides a theoretical path for the formation of an economic and practical teaching resource encryption system in colleges and universities, and on the other hand, it provides a new idea for the application research of deep learning algorithm in the field of data encryption.

II. Related Works

In recent years, the research on encrypted storage has developed more deeply from the perspective of encryption details and user experience. Li M et al. constructed an encrypted storage scheme using a searchable symmetric encryption method and used locally sensitive hash and bloom filter with high search accuracy. The results show that the encryption efficiency of this method is higher than that of the traditional method [5]. Xue K et al. proposed a method to protect encrypted cloud storage from EDOS attacks. This method reduces the resource consumption cost of cloud computing users and can comply with any access policy of cp-abe. The results show that this method has higher security and practicability [6]. Kumar G K's team proposed an encrypted storage algorithm using symmetric encryption keys for the same set of keys. The results show that the algorithm can reduce the data burden of cloud storage in the form of eliminating duplicate data, and can provide more efficient data access by using access policies [7]. Rao E et al. provided a secure data search method for encrypted user data. The results show that this method can provide users with different types of online data while ensuring users' privacy and data security [8].

On the other hand, the research based on cloud platform is gradually diversified. Farhadi H et al. used the cloud platform to quickly and automatically detect the time series image of measuring the burn area, so as to obtain the accurate information of the density and distribution of the combustion area in the large vegetation coverage area. The results show that this method has stronger dynamic real-time performance and can greatly improve the accuracy of the map of the combustion area [9]. Taking the cloud platform as the basic framework of smart education, Liu steam proposed a dual high-precision cooperation strategy to accurately perceive the needs of smart teaching and resource supply. The results show that this method not only improves the effectiveness of communication between teaching resources and learning needs, but also provides teachers and learners with a more personalized and secure data access strategy [10]. Liu P proposed an EV battery voltage evaluation strategy based on cloud platform, and used the spatial clustering method with noise to improve the calculation efficiency of outlier detection. The results show that this method not only maintains high recognition ability, but also reduces the complexity of calculation and has higher calculation efficiency [11]. Choi S et al. Built a digital twin data platform based on cloud platform to facilitate novices to build digital twin data schemes more quickly and easily. An example shows that this method is fast and effective [12]. This research applies the cloud platform, which has compatibility and diversified development direction, to the encrypted storage of teaching resources, and provides the school with a more economical and reliable security scheme of teaching resources through the encrypted storage service of cloud platform.

III. Design of Encrypted Storage Method of Oral English Teaching Resources on Cloud Platform

A. *Design of Cloud platform Storage Encryption Method based on Convolutional Neural Network Algorithm*

Modern teaching resource storage is divided into three main ways. One is that users graft and train the encryption storage module of teaching resources by themselves. For the school that lacks practical and technical talents in network and intelligent computing technology, the cost of designing the encryption storage system of teaching resources alone is too high, and it is technically impossible to achieve to a certain extent [13]; the second is to use it by renting a third-party encrypted storage model. Although this method can solve its own technical shortcomings, the maintenance cost is still relatively high, and the encrypted storage process is controlled by the third-party organization in the whole process, which has a certain risk; the third is to use cloud platform services for teaching resource storage [14]. For oral English teaching, which needs to constantly update a large number of oral interactive resources, cloud platform encrypted storage is a more secure, economical and real-time encrypted storage method. However, there are still some risks in cloud platform encrypted storage. This risk is mainly reflected in the infrastructure of cloud platform. The infrastructure of cloud platform encrypted storage is shown in Fig. 1.

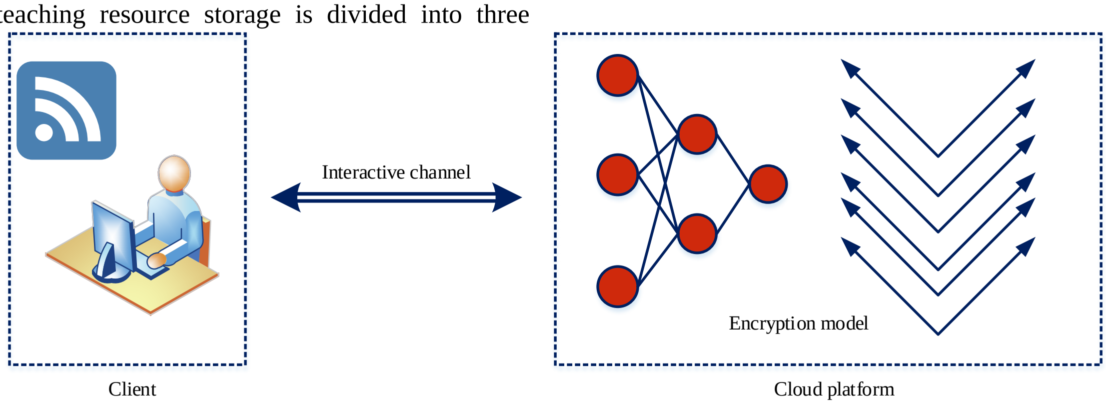


Fig. 1. Cloud platform based encrypted storage architecture

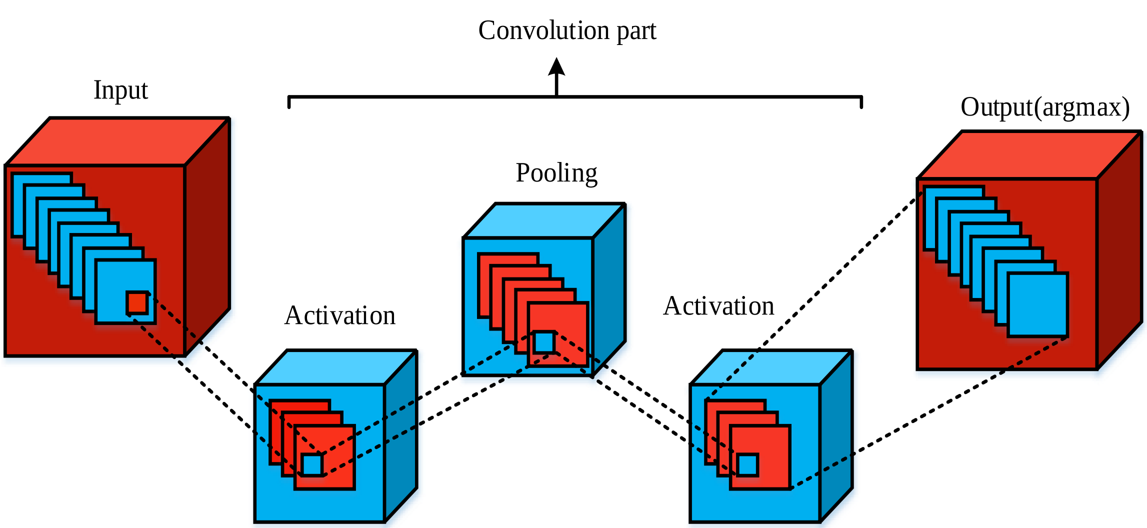


Fig. 2. Convolutional neural network

Fig. 1 shows the basic encrypted storage architecture of the cloud platform. It can be seen that users communicate with the cloud platform mainly through interactive channels. This way of user operation, channel communication, and cloud encrypted storage is easy to form the risk of resource leakage in two parts. One is the security and privacy protection of the cloud platform itself, that is, whether the encrypted storage model used by the cloud platform is reliable; the other is the data transmission security problem that may occur in the process of user interaction with the cloud platform. In this study, the two storage security issues are studied respectively. For the security and privacy protection of the cloud platform itself, this research optimizes the convolutional neural network algorithm based on the characteristics of encrypted storage of spoken English teaching resources. This is mainly because the convolutional neural network itself has strong deep learning ability and structural stability. For the college encryption application scenario, the convolutional neural network has lower application costs, is economical, and requires less maintenance consumption, Colleges and universities can implement the application without consuming a lot of technical support. The specific schematic diagram of the architecture is shown in Fig. 2.

The convolution neural network model in Fig. 2 is divided into three main parts: input part, convolution part and output part. The main input content of the input part of the optimized encrypted storage convolution neural network model in this study is to store the ciphertext used for encryption. This part of ciphertext can be divided into two specific ciphertexts, both of which are $w \times h \times c$ in size. The two ciphertexts are expressed as real part $\left[F(x)_R\right]$ and imaginary part $\left[F(x)_I\right]$ respectively. At this time, the number of plaintext slots of the system is n, and the convolution kernel parameter of plaintext of the system is k, At this time, the convolutional neural network encrypted storage model performs the operation by means of fast Fourier transform, and the real part operation result can be obtained:

$$\left[F(y)_R\right]=\left[F(x)_R\right]\otimes\left[F(f_i)_R\right]\oplus\left[F(x)_I\right]\otimes\left(-F(f_i)_I\right) \tag{1}$$

In formula (1), y represents the convolution result of the algorithm, x represents the input data, $\otimes$ represents homomorphic multiplication, and $\oplus$ represents homomorphic addition. Calculate the imaginary part of the ciphertext in the same way to obtain formula (2):

$$\left[F(y)_I\right]=\left[F(x)_R\right]\otimes\left[F(f_i)_I\right]\oplus\left[F(x)_I\right]\otimes F(f_i)_R \tag{2}$$

$F(f_i)_I$ and $F(f_i)_R$ represent the Fourier transform results of convolution kernel respectively. The cloud platform will automatically generate random variables r after calculation. The size of the random variable r is $w \times h$. On this basis, the cloud platform encrypts the results of the fast Fourier transform and sends the ciphertext to the user. The real encryption formula is shown in formula (3):

$$\left[F(y-r)_R\right]=\left[F(y)_R\right]\oplus\left[-F(r)_R\right] \tag{3}$$

In formula (3), $\left[F(y-r)_R\right]$ represents the real content of the ciphertext sent by the cloud platform to the customer, and $\left[-F(r)_R\right]$ represents the real fast Fourier transform encryption result. The imaginary part ciphertext can be obtained by adopting the same calculation method:

$$\left[F(y-r)_I\right]=\left[F(y)_I\right]\oplus\left[-F(r)_I\right] \tag{4}$$

In formula (4), $\left[F(y-r)_I\right]$ respectively represents the virtual part ciphertext content sent by the cloud platform to customers, and $\left[-F(r)_I\right]$ represents the virtual part fast Fourier transform encryption result. After the user decrypts the ciphertexts at both ends respectively, the two short ciphertexts can be combined to form a complete $F(y-r)$, and then the $(y-r)$ can be obtained by inverse fast Fourier transform. At this time, the cloud platform and the user can share the state of additivity. The common setting formula is:

$$\begin{cases} x^s = r \\ x^c = y - r \end{cases} \tag{5}$$

Where s represents the cloud platform, c represents the client, and r represents that the cloud platform will automatically generate random variables after operation. The plaintext convolution kernel parameters of the convolution neural network model are encrypted and stored. The calculation flow of the encryption process is shown in Fig. 3.

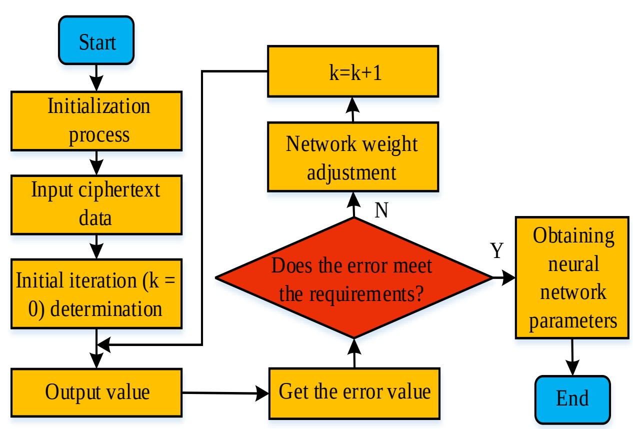


Fig. 3. Encrypted convolutional neural network flow

In the practical application of encrypted storage of teaching resources, due to the requirements of oral English teaching storage and real-time update, more complex operations are often required. In this case, the convolution cores of the convolution layer must participate in the independent input and convolution process, so as to obtain their own convolution results. Here, it can be assumed that there are c convolution cores and the same number of data channel inputs in the convolution layer. At this time, the system needs to treat each input and filtering calculation process as a separate data operation group in the ciphertext domain, carry out parallel operation through the operation method in a simple case, obtain c ciphertext individual results at the same time, and then accumulate a large number of ciphertext individual results to form the final overall ciphertext results. In addition to the convolution layer, this study also optimizes the activation layer of the encrypted storage product neural network model. Since the encrypted storage design conducted in this study requires the convolutional neural network model to perform nonlinear calculation, it is necessary to select the nonlinear function of the activation layer. The commonly used convolutional neural network activation functions include tanh function and relu function. The comparison of the function images of the two functions is shown in Fig. 4:

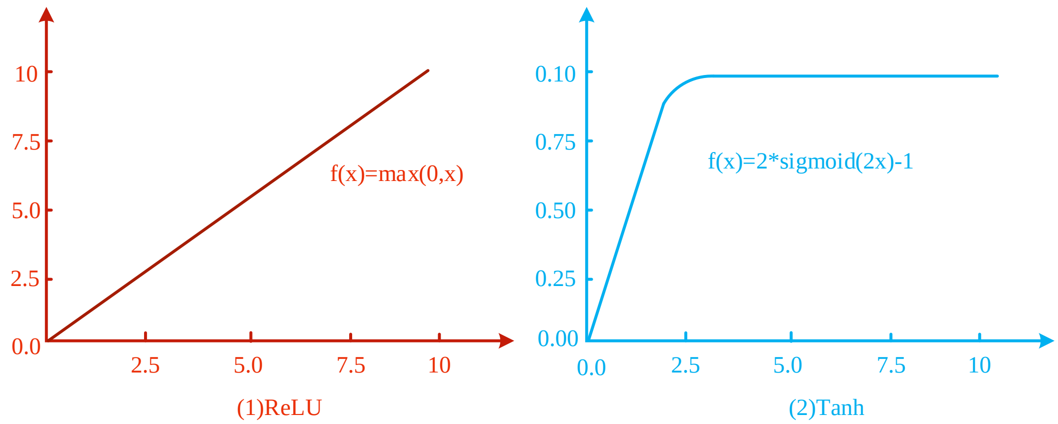


Fig. 4. Function image comparison

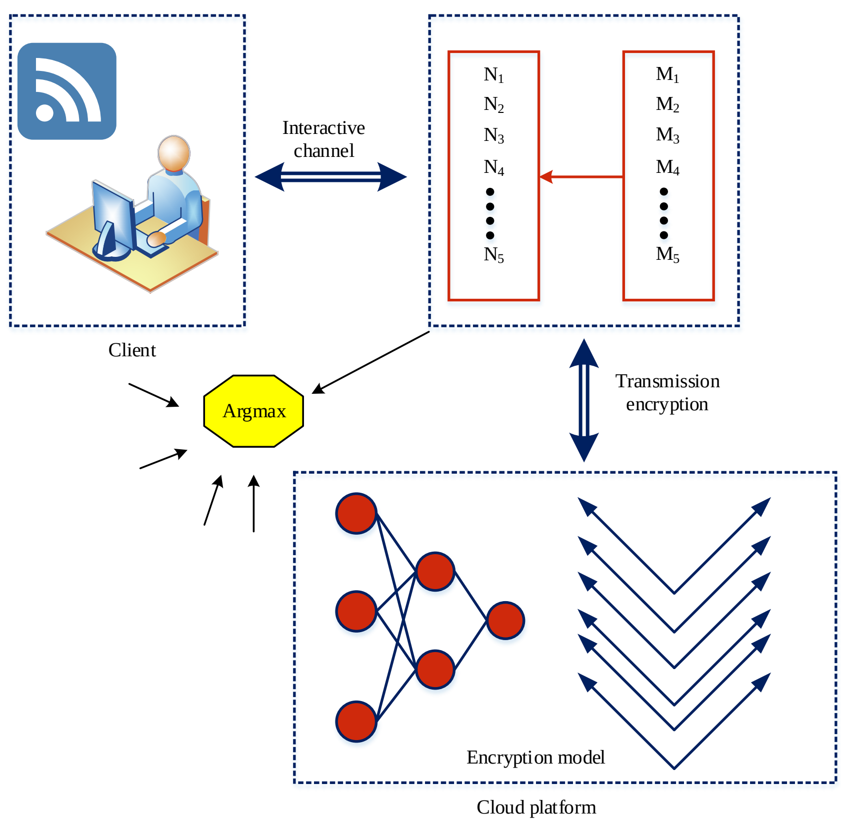


Fig. 5. Storage data transmission optimization

Tanh function formula is shown in formula (6):

$$\tanh(x) = 2 \cdot sigmoid(2x) - 1 \tag{6}$$

As shown in the calculation formula of Lu power division, it is more difficult to select and activate the Lu power division function (Reh) in the design process, so it is more difficult to activate the two kinds of operation formulas:

$$f(x) = \max(0, x) \tag{7}$$

Relu activation function can clear all nonpositive terms in the input data center, balance the size relationship between input data and output data on this basis, enable each neuron to be activated in the process of input and output, and continuously transfer this activation state downward. This method can help the model improve its input sparsity and model activity, and improve the operational performance of the model. Design of cloud platform output encryption method based on argmax algorithm

The ultimate purpose of the encrypted storage model of cloud platform oral English teaching resources designed in this study is to effectively ensure the security of teaching data in the two operations of customer storage and data transmission. Therefore, after the optimization design of the storage model, the research will also carry out the security design for the data transmission part of the cloud platform encrypted storage model, which is mainly optimized by the argmax algorithm, the overall optimization position is shown in Fig. 5:

It can be seen from Fig. 5 that the Argmax algorithm mainly acts on the transmission channel established when users communicate with the cloud platform on the channel, and improves data security by strengthening the reliability in the process of data input and output. In the cloud platform encryption storage algorithm designed in this study, cloud platform has encrypted the oral English teaching resources. When users need to take out or input data from the cloud platform storage, they need to access the data through the common protocol between the client and the cloud platform. This research mainly uses the optimized convolution neural network algorithm as the main storage encryption technology. The output of the last convolution layer in the convolution neural network is called Logits, which represents the weight difference between different categories in the input process, and the main function of the softmax layer is to convert Logits into the vector form representing the probability values of different categories. The softmax function is shown in formula (8):

$$f(x)_k = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}} \tag{8}$$

Where $i = 1, 2, \ldots, K$, K represents the length of Logits vector. The final classification result can be expressed as:

$$t = \arg\max(\log its) \tag{9}$$

The model connection between softmax layer and argmax algorithm is shown in Fig. 6.

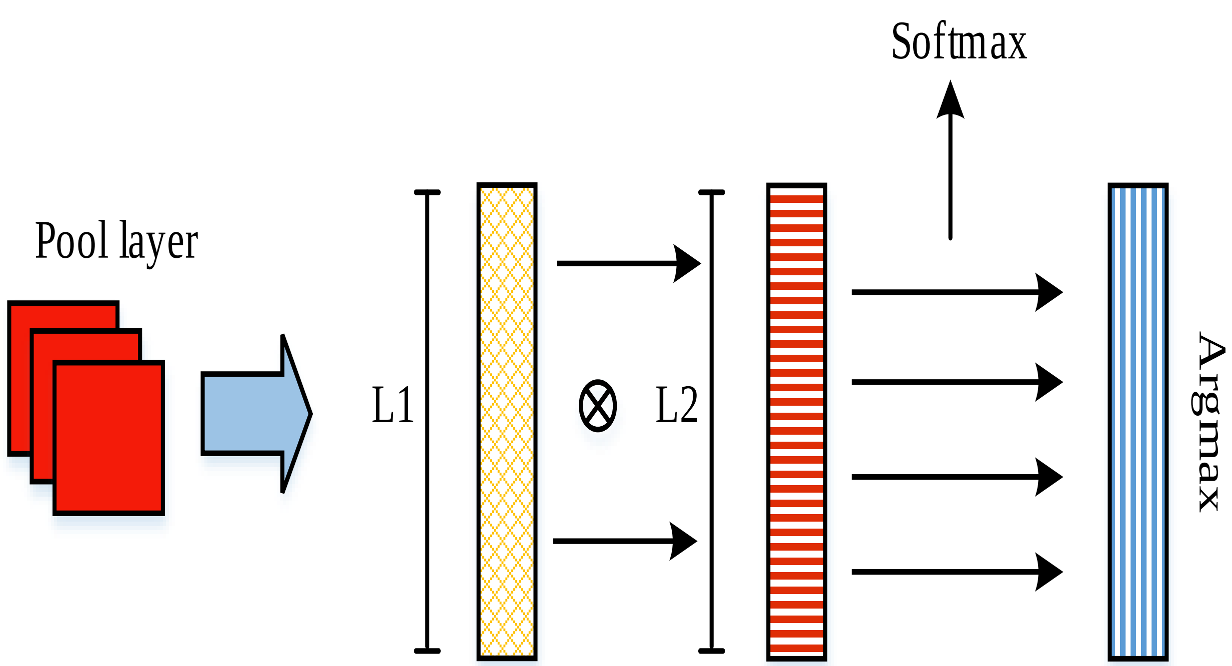


Fig. 6. Model connection

In this study, the improved argmax algorithm is used to encrypt the data access part. The Logits vector obtained from cloud platform encryption can be expressed as:

$$Enc(x)=\left(Enc(x_1),Enc(x_2),\cdots,Enc(x_n)\right) \tag{10}$$

Where x represents the generated random vector. The cloud platform extracts two ciphertexts arbitrarily from the ciphertext set and performs homomorphic subtraction calculation for the ciphertext to obtain $n\cdot(n-1)$ new ciphertexts. The specific matrix is shown in formula (11):

$$\left\{\begin{matrix} [x_1-x_2] & [x_1-x_3] & \cdots & [x_1-x_n] \\ [x_2-x_1] & [x_2-x_3] & \cdots & [x_2-x_n] \\ \cdots & \cdots & \cdots & \cdots \\ [x_n-x_1] & [x_n-x_2] & \cdots & [x_n-x_{n-1}] \end{matrix}\right\} \tag{11}$$

Since the optimized convolutional neural network encryption storage model used in this study adopts the combined encryption mode of real part ciphertext and imaginary part ciphertext, after obtaining the basic Logits vector matrix, it needs to be transformed into the form of double encryption matrix corresponding to the double ciphertext mode. The estimated size of the ciphertext matrix is n, and each individual ciphertext contains $n-1$ plaintext data different from each other, Ciphertext matrix itself has the dual commonalities of cloud platform and client. After the $n-1$ subhomomorphic subtraction, the cloud platform needs to perform random inter row conversion on the ciphertext matrix to obtain the row random matrix formed by the combination of double ciphertext matrices. The specific formula is shown in formula (12):

$$\left\{\begin{matrix} \left[\left(x_{\psi_1}-x_{\pi_{\psi 1}(2)}\right) \right. & \left(x_{\psi_1}-x_{\pi_{\psi_1}(3)}\right) & \cdots & \left.\left(x_{\psi_1}-x_{\pi_{\psi 1}(n)}\right)\right] \\ \left[\left(x_{\psi_2}-x_{\pi_{\psi 2}(1)}\right) \right. & \left(x_{\psi_2}-x_{\pi_{\psi_2}(3)}\right) & \cdots & \left.\left(x_{\psi_2}-x_{\pi_{\psi_2}(n)}\right)\right] \\ \cdots & \cdots & \cdots & \cdots \\ \left[\left(x_{\psi_1}-x_{\pi_{\psi_n}(1)}\right) \right. & \left(x_{\psi_n}-x_{\pi_{\psi_n}(2)}\right) & \cdots & \left.\left(x_{\psi_n}-x_{\pi_{\psi n}(n-1)}\right)\right] \end{matrix}\right\} \tag{12}$$

Where $(\psi_1,\ \psi_2,\ldots,\ \psi_n)$ represents a random arrangement of $(1,2,\cdots,\ n)$, and on this basis, $(\pi_{\psi 1}(2),\ \pi_{\psi 1}(3),\ldots,\ \pi_{\psi 1}(3))$ can be obtained by removing ψ_1 from the arrangement, while $(\pi_{\psi 2}(1),\ \pi_{\psi 2}(3),\ldots,\ \pi_{\psi 2}(n))$ represents the nonrandom arrangement after removing ψ_2 from the arrangement. After the cloud platform analogizes the matrix, it then randomizes the matrix of formula (12) with two matrices to obtain:

$$\left\{\begin{matrix} r_2^1 & r_2^1 & \cdots & r_n^1 \\ r_1^2 & r_3^2 & \cdots & r_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ r_1^n & r_2^n & \cdots & r_{n-1}^n \end{matrix}\right\},\left\{\begin{matrix} \varepsilon_2^1 & \varepsilon_2^1 & \cdots & \varepsilon_n^1 \\ \varepsilon_1^2 & \varepsilon_3^2 & \cdots & \varepsilon_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ \varepsilon_1^n & \varepsilon_2^n & \cdots & \varepsilon_{n-1}^n \end{matrix}\right\} \tag{13}$$

Formula (13) represents a random positive number r_j^i, $\varepsilon_j^i\in\left[0,r_j^i\right)$. The first ciphertext matrix and the row random matrix are multiplied by homomorphic multiplication, and then the matrix is interleaved with the second ciphertext matrix by homomorphic subtraction to obtain a fully random ciphertext matrix. The specific matrix is shown in formula (14):

$$\left\{\begin{matrix} \left[r_2^1\cdot\left(x_{\psi_1}-x_{\pi_{\psi_1}}(2)\right)-\varepsilon_2^1\right. & \cdots & \cdots & \left.r_n^1\cdot\left(x_{\psi_1}-x_{\pi_{\psi_1}}(n)\right)-\varepsilon_n^1\right] \\ \left[r_1^2\cdot\left(x_{\psi_2}-x_{\pi_{\psi_2}}(1)\right)-\varepsilon_1^2\right. & \cdots & \cdots & \left.r_n^2\cdot\left(x_{\psi_2}-x_{\pi_{\psi_2}}(n)\right)-\varepsilon_n^2\right] \\ \cdots & \cdots & \cdots & \cdots \\ \left[r_1^n\cdot\left(x_{\psi_n}-x_{\pi_{\psi_n}}(1)\right)-\varepsilon_1^n\right. & \cdots & \cdots & \left.r_{n-1}^n\cdot\left(x_{\psi_n}-x_{\pi_{\psi_n}(n-1)}\right)-\varepsilon_{n-1}^n\right] \end{matrix}\right\} \tag{14}$$

This operation process requires n sub homomorphic subtraction, homomorphic addition and homomorphic multiplication. When different types of homomorphic addition operations are carried out, it is necessary to change the setting mode in the operation process to ensure that the random number generated in the calculation process will not affect the symbol corresponding to the encrypted plaintext, that is, in the homomorphic encryption operation process, its influence on the symbol data is limited by the value range of the associated random number. After the cloud platform obtains the random ciphertext matrix, it can send the matrix information to the access user. The access user decrypts the ciphertext matrix using the common protocol to obtain the plaintext matrix. The specific matrix is shown in formula (15):

$$\left\{\begin{matrix} r_2^1\cdot\left(x_{\psi_1}-x_{\pi_{\psi_1}}(2)\right)-\varepsilon_2^1 & \cdots & \cdots & r_n^1\cdot\left(x_{\psi_1}-x_{\pi_{\psi_1}}(n)\right)-\varepsilon_n^1 \\ r_1^2\cdot\left(x_{\psi_2}-x_{\pi_{\psi_2}}(1)\right)-\varepsilon_1^2 & \cdots & \cdots & r_n^2\cdot\left(x_{\psi_2}-x_{\pi_{\psi_2}}(n)\right)-\varepsilon_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ r_1^n\cdot\left(x_{\psi_n}-x_{\pi_{\psi_n}}(1)\right)-\varepsilon_1^n & \cdots & \cdots & r_{n-1}^n\cdot\left(x_{\psi_n}-x_{\pi_{\psi_n}(n-1)}\right)-\varepsilon_{n-1}^n \end{matrix}\right\} \tag{15}$$

The access client finally counts and classifies the matrix plaintext symbols of each line, and obtains the corresponding real value at the cloud platform.

IV. Effect Analysis of Encrypted Storage Method of Oral English Teaching Resources on Cloud Platform

A. *Analysis of Storage Encryption Effect of Neural Network*

In the part of the effect analysis of the encrypted storage method of oral English teaching resources on the cloud platform, this study will be divided into two parts: cloud platform storage encryption and cloud platform storage and transmission encryption according to the cloud platform structure of oral English teaching resources. The cloud platform storage encryption part takes the optimized convolutional neural network storage encryption model as the main encryption storage tool. Therefore, this research will analyze the connection layer performance of the model, the running time and overhead of the model, and the efficiency of encryption and decryption. In order to explore the performance change of the model when the user input demand increases but the model output is stable, the performance of the model connection layer is analyzed. The performance analysis of the connection layer of the model is shown in Fig. 7.

Fig. 7 shows the operation performance changes of the connection layer under different input levels when the output level is fixed. Here, the operation performance changes are measured mainly by running time. As can be seen from Fig. 7, the operation efficiency of the model running time in the setting stage and the operation stage is significantly different. Under almost all input variables, the operation efficiency of the model connection layer in the operation stage is significantly higher than that in the setting stage. On this basis, observing the operation efficiency caused by the change of input variables, it can be found that when the value of input variables is between 2000 and 2500 and between 4000 and 4500, the operation efficiency of the connection layer has changed greatly. When the input variable value is between 2000 and 2500, the running time of the connection layer in the setting stage increases from 5.4ms to 7.8ms, while the running time of the connection layer in the setting stage increases from 2.9ms to 4.1ms, and the running efficiency decreases significantly. This is because the plaintext slot size of the model in this study is 2048, that is, a ciphertext in the model can support 2048 plaintext at the same time when it is most saturated, Therefore, when the amount of input data exceeds this amount, the operation efficiency of the model connection layer is greatly affected. In addition, in the stage where the value of the input variable is between 4000 and 4500, it is the node when the number of inputs exceeds twice the number of plaintext slots, so the impairment of operational efficiency is more obvious. To sum up, when the number of inputs does not exceed the number of plaintext slots, the performance of the connection layer of the model involved in this time is relatively reliable, but after exceeding, the performance will be affected to some extent, which is a normal phenomenon. In this study, the overall running time and running cost of the model is tested by distinguishing the storage encryption scheme from the convolution neural model. The convolution neural model selects the MNIST model and cifar-10 model respectively, and the storage encryption scheme selects minions and gazelle. The specific test results are shown in Fig. 8.

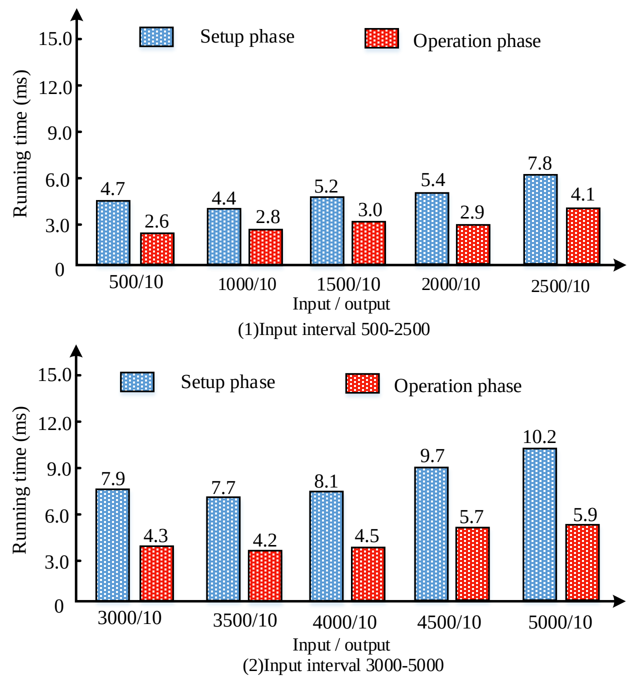


Fig. 7. Performance analysis of connection layer

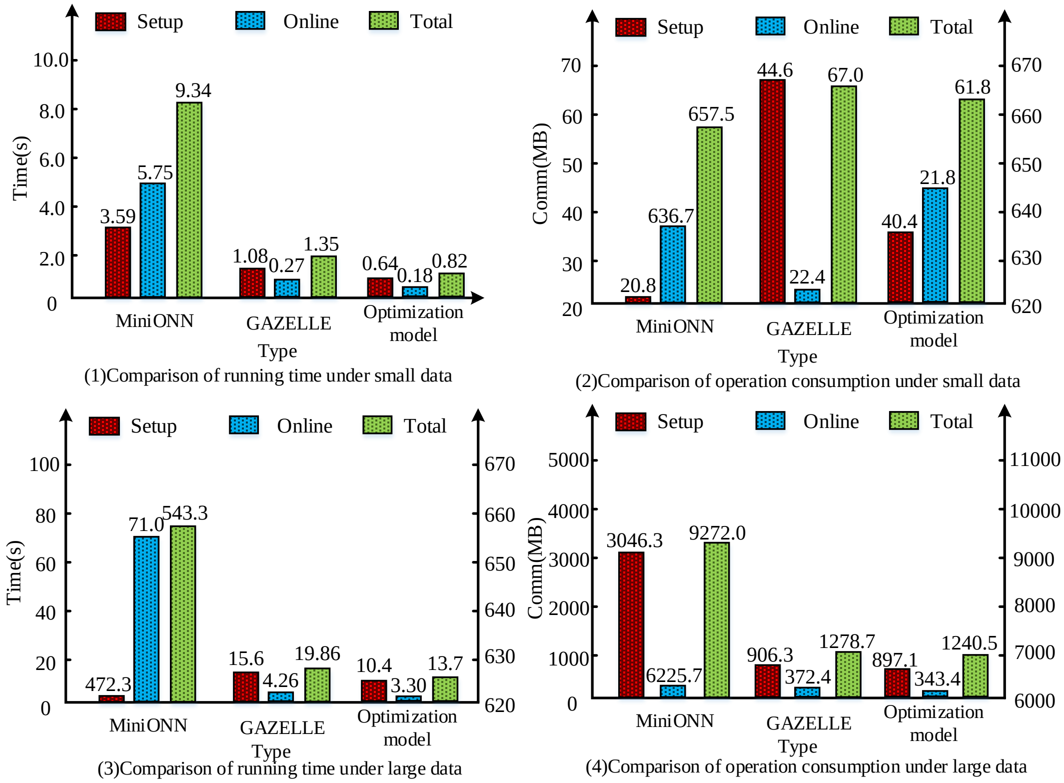


Fig. 8. Data set performance test

Fig. 8 compares and analyzes the models from the perspectives of running time and running cost. The results show that the running time of the model optimized in this study is 0.64ms under the small data set, 0.18ms online, and 0.82ms overall. The overall running time is less than 9.34ms of the minion model and 1.35ms of the gazelle model, and the running time in the setting stage and online stage is smaller; In terms of running cost, the model optimized in this study sets the running cost to 40.4 MB under the small data set, the online running cost to 21.8 MB, and the overall running cost to 61.8 MB. The overall running cost is less than 657.5 MB for the minion model and 67.0 Mb for the gazelle model, and the running cost is less in the setting stage and online stage; Under the large data set, the set running time of the model optimized in this study is 10.4ms, the online running time is 3.30ms, and the overall running time is 13.7ms. The overall running time is less than 543.3ms of the minion model and 19.86ms of the gazelle model, and shows significant advantages in the running time of the setting stage and online stage; In terms of running cost, the model optimized in this study sets the running cost as 897.1 MB under large data set, 343.4 MB on-line running cost and 1240.5 MB overall running cost. The overall running cost is less than 9272.0 MB for the minionn model and 1278.7 MB for the gazelle model, and the running cost in the setting stage and online stage is smaller. The analysis of the encryption and decryption efficiency of the model is shown in Fig. 9.

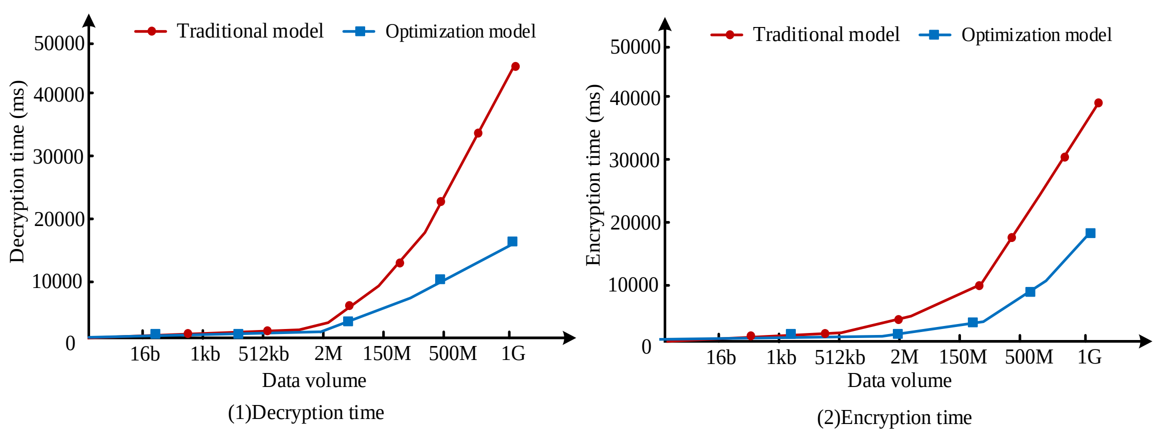


Fig. 9. Analysis of encryption and decryption efficiency

It can be seen that the encryption time and decryption time of the traditional model and the model optimized in this study show a positive proportional trend with the increase of the data scale. In terms of encryption time, taking the 512KB data scale as the starting point, the encryption time of the traditional model and the model designed in this study has increased significantly, but the encryption time of the model optimized in this study on each data scale node is relatively less, and the highest time point is no more than 20000ms; The decryption time of each node of the traditional research model is not less than 2m, but the decryption time of each node of the traditional design model is not less than 2m. It can be seen that the optimized models have higher operation efficiency in the process of encryption and decryption.

B. *Analysis of Output Encryption Effect of argmax Algorithm*

In the cloud platform storage and transmission encryption analysis part, this research first analyzes the running time and transmission efficiency of optimizing the traditional module of argmax. In the analysis process, the research will take the category parameters and precision values as the main distinguishing dimensions. The correlation analysis of category parameters is shown in Fig. 10.

It can be seen that the operation time and operation consumption data of the optimized argmax output encryption module under the three category parameters of 10, 100, and 1000 are 0. It can be seen that the optimized argmax output encryption module does not need to operate in the setting stage. In the online operation stage, when the category parameter is 10, the operation time of the system is 0.21ms; When the category parameter rises to 100, the running time of the system is 2.01ms; When the category parameter rises to 1000, the running time of the system is 20.01ms; It can be seen that the running time of the optimized argmax output encryption module shows an increasing trend with the increase of category parameters. In terms of system operation consumption, when the category parameter is 10, the system operation consumption data is 0.59 Mb; When the category parameter rises to 100, the consumption data of the system is 5.92 Mb; When the category parameter rises to 1000, the consumption data of the system is 59.21 Mb; It can be seen that the amount of data transmitted by the optimized argmax output encryption module shows an increasing trend with the growth of category parameters. It can also be seen from the growth law of the two indicators that with the increase of category parameters in the order of 10 times, the running time and the amount of transmitted data of the system also increase in the order of 10 times. This indicates that the operation efficiency and loss of the system are relatively stable and will not exceed the expected additional loss with the increase in operation demand, which proves that the overall system is still stable in the face of the excessive operation. As different precision operations will lead to different system operation costs, if the system has a large difference in operation costs between high precision and low precision, the overall system operation stability will be insufficient, which will easily lead to higher performance consumption and resource waste. Therefore, the research will analyze the system costs under different precision. The specific accuracy analysis results are shown in Fig. 11.

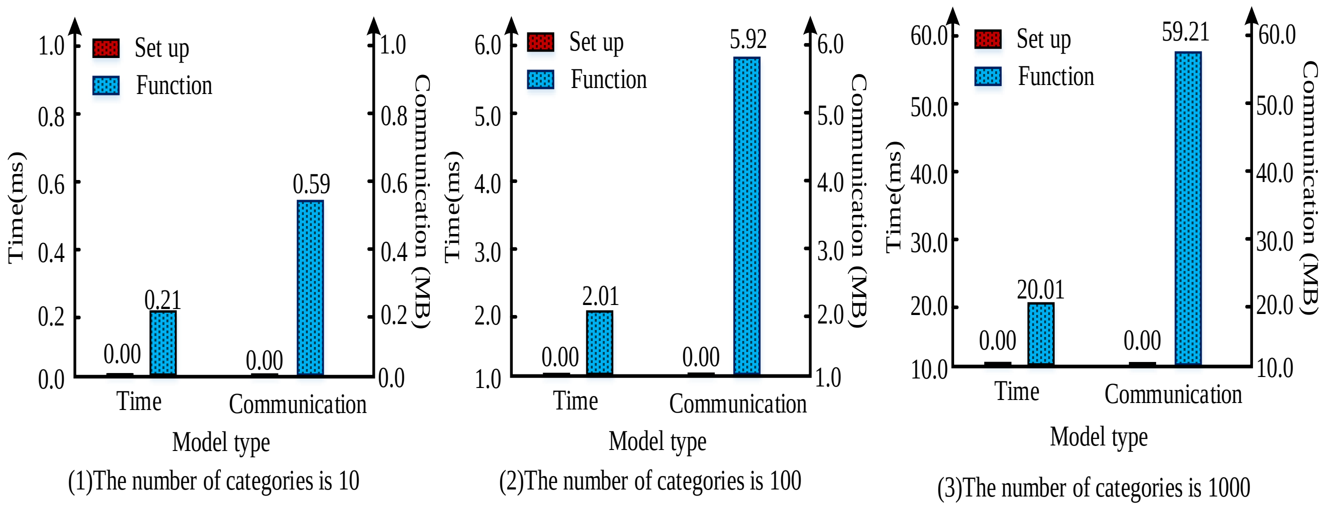


Fig. 10. Category parameter analysis

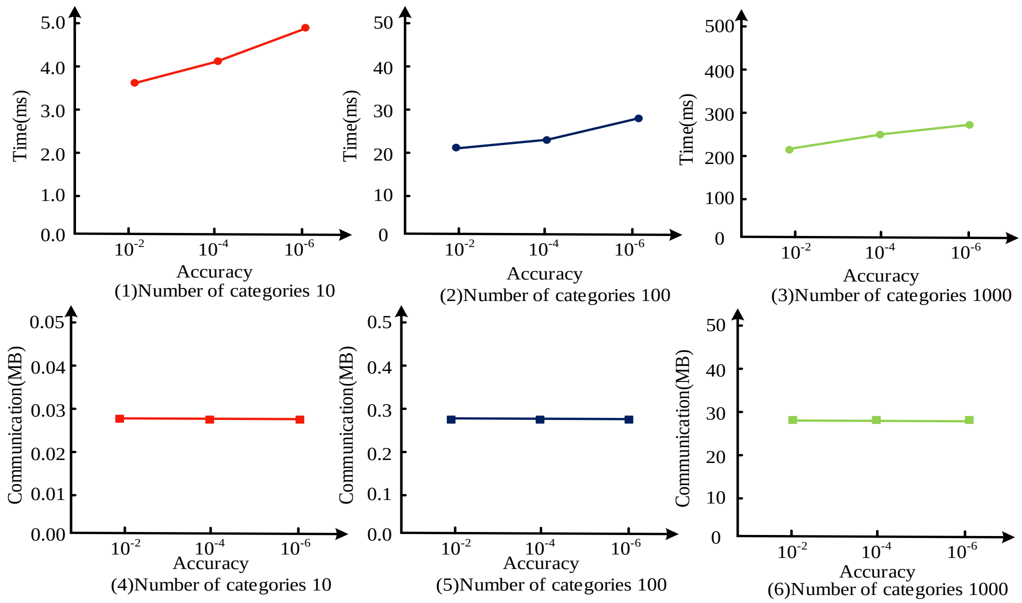


Fig. 11. Accuracy analysis

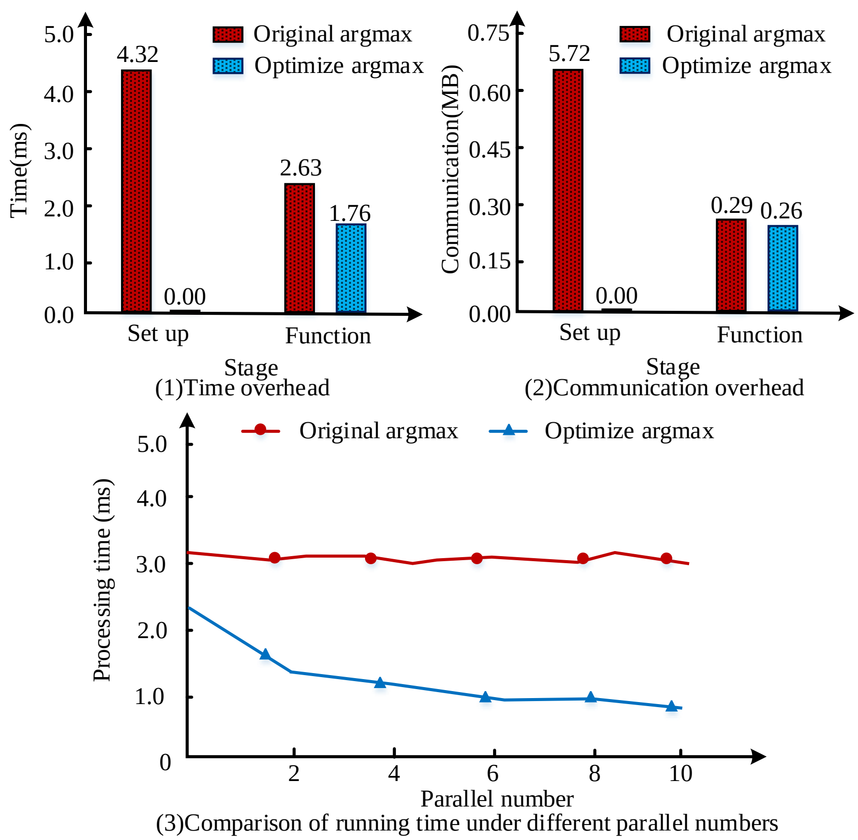


Fig. 12. Performance comparison of overall transmission module

Fig. 11 analyzes the running time and running consumption of 10, 100, and 1000 parameters under the accuracy conditions of 10-2, 10-4, and 10-6. The results show that when the category parameter is 10, the running time of the system increases gradually with the increase of accuracy, the overall range is between 3.0 and 5.0, while the running consumption does not fluctuate, and remains at 0.028 under all accuracy; When the category parameter is 100, the running time of the system also shows a gradually increasing trend with the increase of accuracy, the overall range is between 0.2 and 0.3, while the running consumption does not fluctuate and remains at 0.28 under all accuracy; When the category parameter is 1000, the running time of the system still shows a gradually increasing trend with the increase of accuracy, the overall range is between 20 and 30, while the running consumption does not fluctuate and remains at 28 under all accuracy. It can be seen that the setting and online operation efficiency of the system increase with the increase of accuracy, but the overall operation cost does not change with the change of accuracy but maintains the overall consistency, and the overall operation resource consumption is relatively stable. The comparative analysis of the overall system performance is shown in Fig. 12.

Fig. 12 compares the optimized argmax output encryption module with the traditional argmax output encryption module. The results show that the optimized argmax output encryption module in this study will not produce running time and running consumption in the system setting stage, which is because the optimized argmax output encryption module does not need to be calculated in the setting stage. From the perspective of online operation, the running time and running loss of the optimized argmax output encryption module are relatively small. The running time of the traditional argmax output encryption module is 2.63ms, while the running time of the optimized argmax output encryption module is 1.76ms, the running loss of the traditional argmax output encryption module is 0.29mb, and the running time of the optimized argmax output encryption module is 0.26mb. In addition, from the perspective of system processing time, with the increase of parallel number, the overall system processing time of the optimized argmax output encryption module is less than that of the traditional argmax output encryption module, and with the increase of parallel number, the system processing time of the optimized argmax output encryption module shows a certain downward trend. It can be seen that the optimized argmax output encryption module has stronger encryption processing performance.

V. Conclusion

To solve the budget problems and technical difficulties caused by the localized encrypted storage of oral English teaching data, this study takes the cloud platform as the main encrypted storage platform of teaching resources, optimizes the convolutional neural network from the perspective of ciphertext encryption and decryption, and takes it as the main model of cloud platform data storage encryption. On this basis, it studies the encryption improvement of argmax algorithm, makes it more efficient in storing encrypted data, and finally tests the computing performance of the main part and transmission part of the encrypted storage model of the cloud platform. The results show that the overall running time of the encryption model is 0.82 MS under small data sets; the overall running cost is 61.8 MB, and the overall running time is 13.7ms under large data sets; the overall running cost is 1240.5 MB, which is less than other comparable models. The longest consumption time of encryption and decryption of the research and design encryption model is no more than 20000 MS, which is significantly less than that of the traditional model. In terms of encrypted data transmission, the running time of the optimized argmax output encryption module is 1.76ms and the running loss is 0.26 MB, which is less than the traditional encryption model. It can be seen that the encryption model designed in the research has higher security and higher encryption and decryption efficiency in the encrypted storage and transmission of oral English education resources. At the same time, the cloud platform itself also has strong realizability and economic advantages, which is more suitable to help the school carry out the encrypted storage of oral English education resources and improve the security and access efficiency of educational resources.

References

[1] J. Deng, W. Yang, Q. W. Li, et al. "Research and Application of Fire Power Cloud Platform," Procedia Engineering, vol. 211, pp. 911-916, 2018.

[2] X. Feng, F. Yan, X. Y. Liu, et al. "Development of IoT Cloud Platform Based Intelligent Raising System for Rice Seedlings," Wireless Personal Communications, vol. 122, no. 2, pp. 1695-1707, 2021.

[3] R. S. Kumar, L. Parthiban, "Privacy preservation in big data with encrypted cloud data storage using walrus," International Journal of Pure and Applied Mathematics, vol. 119, no. 15, pp. 1833-1842, 2018.

[4] M. Uphoff, M. Wander, T. Weis, et al. "SecureCloud: An Encrypted, Scalable Storage for Cloud Forensics," pp.1934-1941, 2018.

[5] M. Li, G. Wang, S. Liu, et al. "Multi-keyword Fuzzy Search over Encrypted Cloud Storage Data." Procedia Computer Science, vol. 187, no. 2, pp. 365-370, 2021.

[6] K. Xue, W. Chen, W. Li, et al. "Combining Data Owner-Side and Cloud-Side Access Control for Encrypted Cloud Storage," IEEE Transactions on Information Forensics and Security, pp. 2062-2074, 2018.

[7] G. K. Kumar, E. A. Reddy, B. Mamatha, et al. "Access Policy's Over Encrypted Cloud Storage for Secure Deduplication," International Journal of Engineering & Technology, vol. 7, no. 3, pp. 27-31, 2018.

[8] E. Rao, "Enhanced Effective and Privacy Preserving Multi Keyword Search over Encrypted Data in Cloud Storage Using Blowfish Algorithm," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 2, pp. 2845-2853, 2021.

[9] H. Farhadi, M. Mokhtarzade, H. Ebadi, et al. "Rapid and automatic burned area detection using sentinel-2 time-series images in google earth engine cloud platform: a case study over the Andika and Behbahan Regions, Iran," Environmental Monitoring and Assessment, vol. 194, no. 5, pp. 1-19, 2022.

[10] S. Liu, Y. Dai, Z. Cai, et al. "Construction of Double-Precision Wisdom Teaching Framework Based on Blockchain Technology in Cloud Platform," IEEE Access, vol. 9, pp. 11823-11834, 2021.

[11] P. Liu, J. Wang, Z. Wang, et al. "Cloud Platform-Oriented Electrical Vehicle Abnormal Battery Cell Detection and Pack Consistency Evaluation with Big Data: Devising an Early-Warning System for Latent Risks," IEEE Industry Applications Magazine, no. 99, pp. 2-13, 2021.

[12] S. Choi, J. Woo, H. P. Yang, et al. "User-Friendly Method of Digital Twin Application based on Cloud Platform for Smart Manufacturing," Transactions of the Korean Society of Mechanical Engineers A, vol. 45, no. 2, pp. 175-184, 2021.

[13] Q. Yan. "Design of Teaching Video Resource Management System in Colleges and Universities based on Microtechnology," Security and Communication Networks, no. 4, pp. 1-11, 2021.

[14] O. Kolesnyk, IP. Bubeník, J. Apek. "Cloud platform for learning factories," Transportation Research Procedia, vol. 55, no. 23, pp. 561-567, 2021.

Denoising Method of Interior Design Image based on Median Filtering Algorithm

Tao Li
School of Design, Chongqing Industry Polytechnic College, Chongqing, China

***Abstract*—Interior design image generation process is prone to the interference of many factors, resulting in the interior design image denoising effect decreases, denoising time increases, so the interior design image denoising method based on median filtering algorithm is proposed. The architecture of interior design image collection is set up, including video signal conversion module, compression coding module, programmable logic chip module and power module. The interior design image collection is realized by using sensors to collect interior design related video information and converting video signals. Based on the results of image acquisition, the median filtering algorithm based on rough set theory is used to realize the denoising of interior design images. Experimental results show that the denoising effect of the proposed method is better, the average signal-to-noise ratio of interior design images is 54.6dB, and the denoising time is always lower than 0.3s, which can be widely used in practice.**

Keywords—Median filtering algorithm; interior design; image denoising; image acquisition architecture; the rough set

I. INTRODUCTION

Interior design is a comprehensive discipline that studies the artificial environment. It is “a rational creative activity that takes science as the functional basis and art as the form of expression in order to shape an indoor environment where spiritual wealth and material wealth coexist” [1]. With the application of computer image digital interior decoration design technology, the design drawing is no longer an abstract design drawing, but can fully and truly present the design effect and improve the design quality. Before the formal start of interior design, designers should make full preparations. First of all, the designer should build a complete interior design layout in his mind. Based on this layout, the framework should be gradually enriched by collecting data and understanding customer needs [2]. According to the information provided by the customer, the decoration design concept map is sorted out and used as the design framework, which is constantly modified and optimized to form an interior design image. Therefore, the rise of computer image digitization technology has provided great help for interior design. This technology makes designers say goodbye to the traditional manual drawing method and improves the efficiency of drawing work [3]. On the other hand, each detail of interior space can be expressed in a dynamic way, which improves the expressiveness of design drawings. Therefore, the digital work of design drawings has gradually become the mainstream direction of the development of the industry, with high market potential. However, due to the influence of many factors in the process of interior design digitalization, interior design images include noise, resulting in the deterioration of interior design quality, so it is urgent to study a method of interior design image denoising.

Research for indoor design class image denoising is not much, much more to the rest of the research achievements in the field of migration to the interior design in the image denoising to reference [4], for example on at night, such as indoor under low illumination scene image sensor collected video images are widespread low signal-to-noise ratio and poor visual effect, In order to improve the de-noising effect, a video de-noising method based on time-space frequency(TSF) domain set is proposed. First, a digital gain low illumination image noise model is established, and then, according to the theory of the original video before and after two adjacent frames, finish the image registration based on the optical flow method, get the displacement map, in time-space domain to high frequency image denoising and reconstruction, and then realize the infinite impulse response combined with subsequent noise image processing, at the same time in the airspace for noise reduction, the output after a frame denoising image. However, this method has low signal-to-noise ratio and poor image quality for interior design images, so the practical application effect is not good. In reference [5], an image denoising method based on deep learning is proposed in order to make the noise image clear and obvious, as the image will be polluted by noise due to imperfect equipment and lose the details of the original image. The structure of image denoising network based on convolutional neural network is constructed, and the back propagation algorithm in convolutional neural network is optimized, so as to accelerate the training speed of the model and improve the denoising effect effectively. In order to solve the problem of poor image denoising and rendering in interior design system and consider data rendering, a research method of interior design system combining digital processing technology and image processing technology is proposed, which transmits the image after data processing to the logic layer, by digital processing module using the new threshold function of image denoising algorithm of energy distribution into consideration, Image denoising is realized by multiplying the denoising factors with the same wavelet coefficients. In the image processing module, the bidirectional reflection distribution function model is used to simulate the diffuse reflection of objects, and the Torrance-Sparrow microplane model is used to simulate the specular reflection of objects to achieve image rendering. The logic layer transfers the processed interior design image to the application layer, and displays the interior design result through PC or browser interface, so as to improve the image denoising effect [6]. However, this method is too complicated and has the problem of long denoising time for interior design images.

In order to solve the problems of the above methods, this paper proposes a denoising method of interior design image based on median filtering algorithm, and verifies the effectiveness of the proposed method through experiments.

II. Standard Image Acquisition

A. *Interior Design Image Collection*

In order to improve the quality of interior design class image acquisition, this paper designed image acquisition architecture interior design class, including video signal conversion module, compression module, a programmable logic chip module, power supply module, mainly through the use of sensors to collect interior design related video information, and transformed the video signal processing, image acquisition in interior design class.

1) Video signal conversion module: DMD sensor is used to collect interior design related video information, and the video signal is converted. The conversion of video signals is realized by a dedicated video processing chip [7]. This paper uses special video processing chip SAA7113, is a powerful and simple operation of the bit video input processing chip, the chip adopts technology, through the bus and or microcontroller connected to form an application system. It contains two analog processing channels, which can realize video signal source selection, anti-aliasing filtering, conversion, automatic box position, automatic gain control, clock generation, multi-standard decoding, brightness contrast saturation control and multi-standard data decoding.

After the SAA7113 is turned on, the chip does not immediately collect analog video signals for A/D conversion processing and output digital signals. It must be initialized by the front-end processor through the serial bus to set up its internal register before it can operate normally. LLC is the output of line locking clock, which is twice the frequency of pixel clock, and 27MHz is used to synchronize the whole interior design image acquisition architecture. One CYCLE of LLC generates one byte of image data, and the rising edge of LLC can be used as the write signal of frame storage after being inverted in CPLD [8-9]. The function of RTS0 and RTS1 is determined by programming the internal register. RTS0 and RTS1 are thresholds for different routers. The interface is set RTS0 as the horizontal output reference signal line effective signal, and RTS1 as the vertical output reference signal and parity field signal. RTS0: high level means a row of valid pixels, low level means field blanking signal. The rising edge of RTS0 represents the beginning of the effective singular field image, and it is also used to represent the beginning signal of the frame image. The falling edge of RTS1 indicates the start of output of a valid even field image. SAA7113 and its peripheral circuits are shown in Fig. 1.

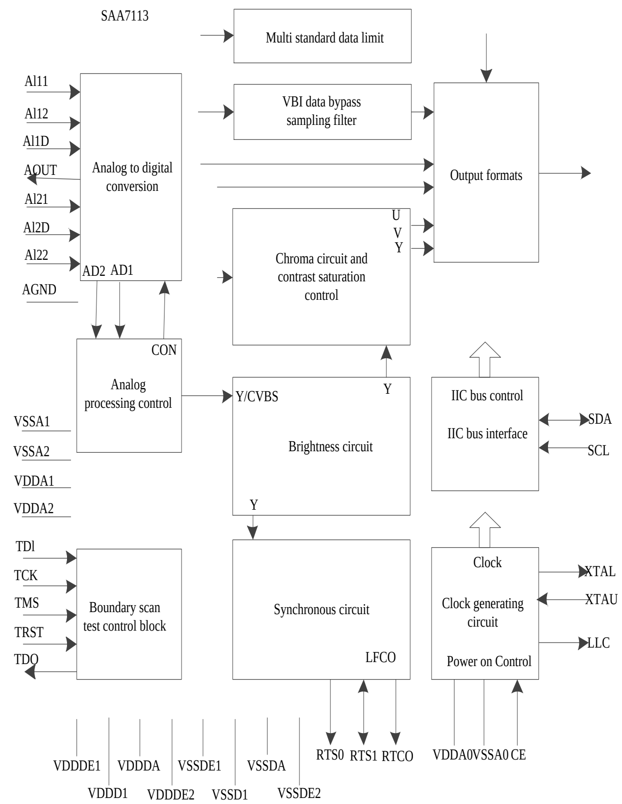


Fig. 1. SAA7113 and its peripheral circuit

2) Compression coding module: The JPEG compression and coding part adopts THE JPEG special codec chip ZR36060, which can realize the image compression and decoding function. In this paper, only its compression and coding function is used [10-11]. ZR36060 is specially designed for computer video acquisition and editing applications, which can easily realize real-time compression and decompression of video signals. During compression, ZR36060 receives digital video signal and encodes it as JPEG stream period; during decompression, it receives a JPEG stream and decodes it into digital video output. The internal structure of ZR36060 is shown in Fig. 2.

VCLK and VCLK×2 are the external clock input. External clock input is connected to the machine terminal by using external oscillation pulse. VCLK×2 is the video clock. For CCIR digital video format, it is connected to the 27MHz clock. VCLK is half the frequency of VCLK×2 and requires synchronization with VCLK×2. SATRT and FRAME are external control signals. The interface of ZR36060 can be divided into video interface, host interface and code interface. The video signal is input and output by the video terminal, the control word of the chip is written and read by the host interface, and the JPEG code is input and output by the code in the activation mode.

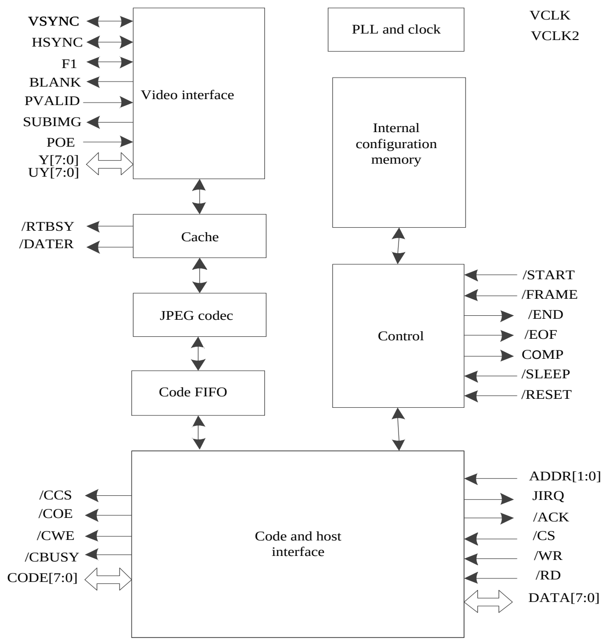


Fig. 2. ZR36060 internal structure

3) Programmable logic chip module: The microcontroller AT89S52 is the transaction control core of this paper [12]. It will enter the working state after power-on and will run continuously. Therefore, the reset circuit of the microcontroller part adopts power-on reset. Interior design image acquisition architecture does not have high requirements for the running speed of the microcontroller. The clock circuit of the microcontroller is composed of 12mb external crystal and grounding capacitor. Reset circuit and clock circuit are shown in Fig. 3.

The communication between the microcontroller and GPRS network is realized through the universal asynchronous serial interface (UART), and its level standard is RS-232, which is a serial physical interface standard developed by the American Electronics Industry Association. The RS-232 has the fastest transmission speed and the maximum transmission distance of 30m. The serial port adopts MAX232 level conversion chip and DB9 interface. The interface circuit is shown in Fig. 4.

4) Power module: The front-end circuit of interior design image acquisition architecture includes digital part and analog part. The interface level of each chip consist of 5 volts and 3.3 volts. To reduce the interference of digital signals with analog circuits, the digital and analog power circuits need to be separated and connected to a point on the circuit board. Considering the power consumption of image acquisition, a three-terminal voltage stabilizer power supply of 5V to 3.3V is utilized. The circuit of the power supply is shown in Fig. 5.

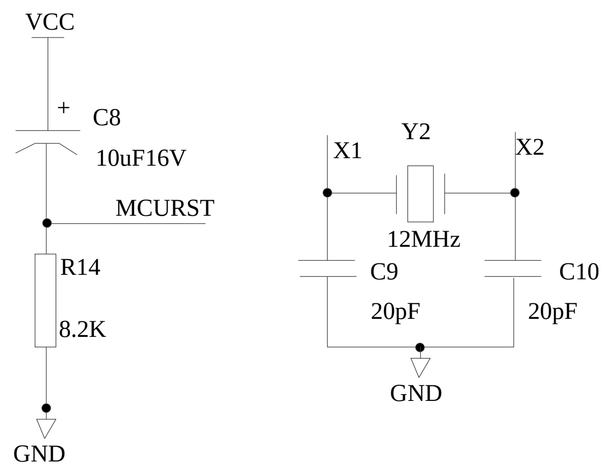


Fig. 3. Reset and crystal oscillator circuit

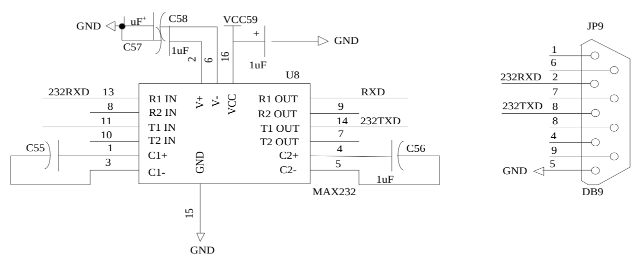


Fig. 4. UART interface circuit

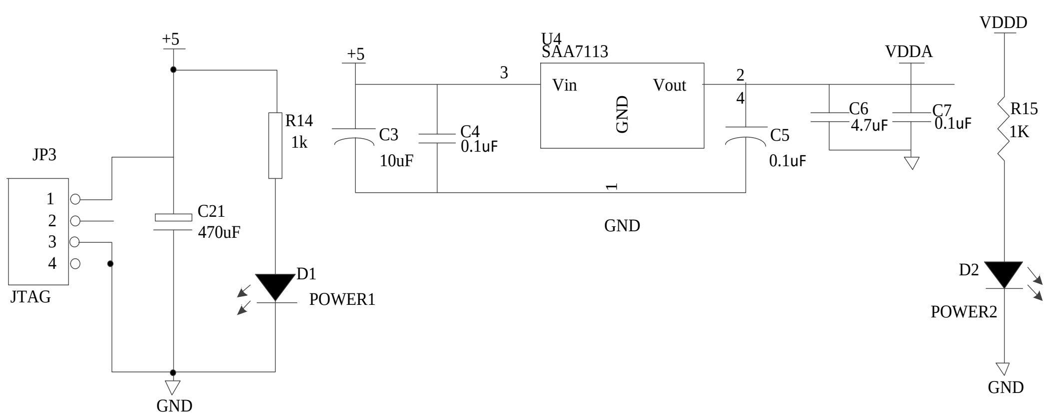


Fig. 5. Power supply circuit

III. Interior Design Image Denoising Method

A. Image Denoising based on Median Filtering Algorithm

As a representative of nonlinear filtering methods, standard median filtering technology can protect image edges while reducing noise, but it only considers the sorting information of data in the filtering window, not the timing information of data, so it will produce edge jitter in image processing. It is possible to remove important image details such as points, corners, and thin lines [13-14]. Considering the problems existing in the standard median filtering, national and international scholars have improved it from many aspects, and produced many improved median filtering algorithms, which make full use of the sorting and timing information of data and improve the filtering quality. Therefore, this paper applies it to interior design image denoising process to improve the effect of denoising.

The basic idea of median filtering is to replace the value of a point in interior design images or digital sequences with the median value of each point in a neighborhood of the point [15]. The median is defined as follows:

For a group of numbers $x_1, x_2, ..., x_n$, which are expressed as $x_{t1}, x_{t2}, ..., x_{tn}$ in order of size, then:

$$y = Med\{x_1, x_2, ..., x_n\} = \begin{cases} x_{t[(n+1)/2]}, n \in 2n+1 \\ \left[x_{t[(n)/2]} + x_{t(n/2+1)}\right]/2, n \in 2n \end{cases} \quad (1)$$

y is called the median of sequence $x_1, x_2, ..., x_n$.

A neighborhood of a particular size or shape of a point is called a window. In the one-dimensional case, the median filter is a sliding window with an odd number of pixels. The value of the center pixel of the window is replaced by the median value of each pixel in the window [16-17] for filtering. Suppose $\{x_{i,j}, (i,j) \in I^2\}$ represents the gray value of each pixel of interior design images, then the two-dimensional median filtering with a filtering window of A can be defined as:

$$y_{i,j} = Med\{x_{i,j}\} = Med\{x_{i+r,j+r}, (r,s) \in A(i, j \in I^2)\} \quad (2)$$

Median filtering also has the following main features:

1) The output of median filter is related to the density distribution of noise, and it has obvious effect on the interference of narrow pulses with small pulse width and long distance.

2) Through the observation of the overall test, it is found that the spectrum characteristics of the median filter have little fluctuation and the mean value is relatively stable. Therefore, it can be considered that the spectrum of the signal after the median filter is basically unchanged [18].

3) Some specific input signals can remain unchanged after median filtering, such as monotonically decreasing or monotonically increasing sequences.

For interior design image processing, standard median filtering mainly has the following deficiencies:

1) From the perspective of signal estimation theory, because the standard median filter is a non-parametric estimation, it does not make full use of the statistical knowledge of the observation model available in practice during processing, so it is quite blind and conservative.

2) Although median filtering has a good suppression effect on long trailing probability distribution noise, such as impulse noise, it has a less ideal suppression effect on short trailing distribution noise and medium trailing distribution noise, such as uniform distribution noise and Gaussian distribution noise.

3) When the number of interference samples in the filtering window is larger than half of the window length, median filtering has no effect. In this case, increasing the size of the filtering window can improve the noise filtering ability, but will destroy the image details.

4) After the median filtering of the image, some thin lines or sharp edges and corners may be eliminated, which destroy the geometric structure of the image.

5) Edge jitter will occur after median filtering for nonconstant signals disturbed by pulse. Even in low noise area, edge displacement still exists.

In order to further improve the filtering effect of standard median filter, rough set theory is used. For noisy interior design images, the effect and quality of early processing often determine the quality of subsequent processing (such as feature extraction, image segmentation, etc.) [19-20]. The complexity and strong correlation of image information cause incompleteness and uncertainty in image processing. Rough set theory has been successfully applied to image processing, among which image filtering is a key research direction in the future. Its basic idea is to treat the information expressed in each image as a knowledge system, and to detect and process the noise pixels in the image by using the approximate set concept of rough set. In this paper, a median filtering algorithm based on rough set theory is proposed to denoise interior design images.

An interior design image can be regarded as a knowledge system. Image I and equivalence relation R constitute an approximate space of interior design image, which is represented by $K = (I, R)$. For a filtering window, an approximate space is formed by the pixels in the window(W) and equivalence relation R, and is represented by $K = (W, R)$.

Rough set theory includes conditional and decision attribute. The conditional attribute set $C = \{c_1, c_2\}$, where c_1 is the extreme value attribute and c_2 is the region attribute of the pixel (hereinafter referred to as the region attribute).

Let the extreme value attribute $c_1 = \{0,1\}$, where 0 represents that the current pixel is the extreme point, that is, its gray value is greater than or equal to the maximum or less than or equal to the minimum gray value of the neighborhood pixel, and 1 represents that the current pixel is not the extreme point. Generally, it is assumed that the gray image to be processed consists of flat region and detail region. If the gray value of the current pixel is close to that of most neighboring pixels, it indicates that the pixel is located in a flat region. Otherwise, it indicates that it is in the detail area. Let the region attribute be $c_2=\{0,1\}$, where 0 means that the current pixel is in the flat region and 1 means that the current pixel is in the detail region. According to the equivalent concept of the irresolvable relation, draw the molecular diagram, classified by attribute C.

1) Drawing the molecular diagram according to c_1: Let x represent image pixel, and the equivalence relation R_{c_1} is defined as the following:

If the gray values of the two pixels are greater than or equal to the maximum or less than or equal to the minimum gray values of the neighboring pixels, then the two pixels are related, that is, they belong to the equivalence class and can be described by the following formula:

$$R_{c_1}(x)=\left\{x \middle| f_{ij} \geq \max(f_w) \quad or \quad f_{ij} \leq \min(f_w)\right\} \tag{3}$$

Where, f_{ij} is the gray value of the corresponding pixel, $\max(f_w)$ represents the maximum gray value of pixel in neighborhood x, and $\min(f_w)$ represents the minimum gray value of pixel in neighborhood x. $R_{c_1}(x)$ represents the set composed of all pixels with extreme gray value, and the complement of R_{c_1} represents the set composed of all pixels with non-extreme gray value.

2) Drawing the molecular diagram according to c_1: R_{c_2} is defined as a small difference between the gray value of a pixel and its adjacent pixels. $R_{c_2}(x)$ represents the set of pixels in all plane areas, and R_{c_2} is complement represents a group of pixels in all detail areas. The difference of gray values is measured by introducing fuzzy membership degree. For a certain pixel point x, the following fuzzy membership function is defined as its degree of belonging to surrounding points:

$$u_z(x)=\left[1+\frac{\left(f_{ij}-f_z\right)^2}{f_z+\lambda}\right]^{-1} \tag{4}$$

Where, z represents the point in the neighborhood of x, f_z represents the gray value of the neighborhood pixel of pixel x, and λ is a constant, which is the square of the mean gray value of the image. According to the definition of membership function, the larger $u_z(x)$ is, the more similar x and z are, and the more likely x and z are to be classified into the same category. Therefore, we can use fuzzy membership function to judge the similarity of gray values between pixels.

The flatness pixel area to be processed for the current pixel () is determined based on the above analysis, can choose the current pixel as elements to be processed, calculate the fuzzy membership degree, on the other a pixel the sum of membership degree, the greater the means of each pixel gray value are more close to, on the basis of pixel gray similarity to measure area, To judge the flatness of the region. Therefore, for any point in the image, the following conditions can be used to determine whether it is located in the detail region or flat region: If the sum of the membership degree of the point to be processed relative to all neighborhood points is greater than or equal to a threshold value, the point is considered to be located in the flat region; otherwise, it is considered to be located in the detail region.

To sum up, it can be concluded that:

$$R_{c_2}(x)=\left\{x \middle| \sum_{x\in Z} u_z(x) \geq T\right\} \tag{5}$$

Where, W represents the selected window centered on the current pixel, that is, a neighborhood of the current pixel, and T represents the preset threshold.

After combining the subgraphs, we can get:

$$W=R_{c_1}(x)-R_{c_2}(x) \tag{6}$$

The pixel in A is median filtered, while the gray values of the other pixels remain unchanged.

In the interior design noisy image, the pixel at position (i,j) is represented by x, and its gray value is f_z. A_{ij} is the filtering window for median filtering of pixel point x, the neighborhood pixel of x is represented by z, and the gray value of pixel z is set as A_{ij}. T is the preset threshold.

The denoising process of interior design images based on median filtering algorithm is as follows:

Step 1: Determine the gray maximum value $\max(f_w)$ and minimum value $\min(f_w)$ of the neighborhood pixel of pixel x in the filtering window A_{ij};

Step 2: Judge the extreme value attribute of pixel point x. If $\min(f_w) \leq f_{ij} \leq \max(f_w)$, When the pixel point x is a signal point, directly output f_{ij}, and execute step 6; Otherwise, when the pixel x is an extreme point, the fuzzy membership function is used to further judge its regional attributes.

Step 3: Calculate the fuzzy membership of pixel x with respect to its neighbor pixel z according to the fuzzy membership function $u_z(x)$, and sum it.

Step 4: Compare the summation result $\sum_{x\in Z} u_z(x)$ with threshold T. If $\sum_{x\in Z} u_z(x) \geq T$, pixel point x is located in the flat area and is the signal point. Output f_{ij} directly and go to step 6. If $\sum_{x\in Z} u_z(x) < T$, pixel x is a noise point, and median filtering is performed on it.

Step 5: According to the standard median filtering principle, replace the gray value f_{ij} of pixel x with the gray median of pixel in the filtering window;

Step 6: If all pixels in the image are processed, complete the process; otherwise, go to the first step to continue to process with the next pixel, so as to achieve interior design image denoising.

IV. Experimental Design

In order to verify the effectiveness of the interior design image denoising method based on median filtering algorithm designed in this paper, experimental tests were carried out. The proposed test parameters are shown in Table I.

In order to verify the application effect of different methods, LYX's method, HCX's method, LYF's method and the method designed in this paper are selected as experimental methods. By selecting multiple interior design scenes, collecting interior design images with sensors, and taking the collected data as experimental sample data, by comparing the denoising effect of interior design images of different methods. The signal-to-noise ratio of interior design images and the denoising time of interior design images are used to verify the practical application effects of different methods.

1) Comparison of denoising effects of interior design images: The indoor design image denoising effects of LYX's method, HCX's method, LYF's method and proposed method in this study are compared in 50 experimental scenes. The comparison results are shown in Fig. 6-7.

TABLE I. Experimental Environment Parameters

Runtime environment	Parameter	Explanation
Hardware environment	CPU	Intel(R) Core (TM)i5-9400
	Frequency	2.90GHz
	RAM	16.0GB
Software environment	Operating system	Windows 10
	Version	18362.1082 pro
	Digits	64bit
	Analog software language	APDL
	Simulation software	Matlab 7.0

By analyzing the results of Fig. 6 and Fig. 7, it can be seen that the image denoising effect of HCX's method is the worst of the four methods, and the image denoising effect of LYX's method and LYF's method is better than that of HCX's method. Compared with the three experimental comparison methods, the image denoising effect of this method is better.

(a) Reference [4] method

(b) Reference [5] method

(c) Reference [6] method

(d) Paper method

Fig. 6. Scene 1 image denoising effect

(a) LYX's method

(b) HCX's method

(c) LYF's method

(d) Proposed method

Fig. 7. Scene 2 image denoising effect

2) Comparison of signal-to-noise ratio of interior design images: The signal-to-noise ratio of interior design images of LYX's method, HCX's method, LYF's method and this proposed method are compared. The results are shown in Table II.

By analyzing the data in Table II, it can be seen that the maximum signal-to-noise ratio of interior design images in LYX's method is 39.7dB, the minimum value is 32.7dB, and the average value is 35.9dB; The maximum signal-to-noise ratio of interior design image in HCX's method is 28.9dB, the minimum value is 23.1dB, and the average value is 26.4dB; The maximum signal-to-noise ratio of interior design image in LYF's method is 39.6dB, the minimum value is 29.7dB, and the average value is 33.1dB; Compared with these methods, the maximum signal-to-noise ratio of interior design image in this method is 58.4dB, the minimum value is 50.6dB, and the average value is 54.6dB, which shows that the signal-to-noise ratio of interior design image processed by this method is higher, the image is clearer, and the practical application effect is better.

TABLE II. SIGNAL TO NOISE RATIO OF INTERIOR DESIGN IMAGES

Number of experiments	Image signal-to-noise ratio/dB			
	LYX's method	HCX's method	LYF's method	Paper method
10	35.1	25.6	30.6	50.6
20	34.2	28.9	31.2	51.4
30	36.9	27.6	35.6	52.3
40	39.7	24.7	30.7	56.4
50	36.4	25.6	35.4	58.4
60	35.6	28.7	29.8	53.2
70	34.1	23.1	29.7	56.7
80	32.7	25.8	32.8	58.2
90	38.5	26.7	35.7	53.6
100	35.7	27.5	39.6	54.7
average value	35.9	26.4	33.1	54.6

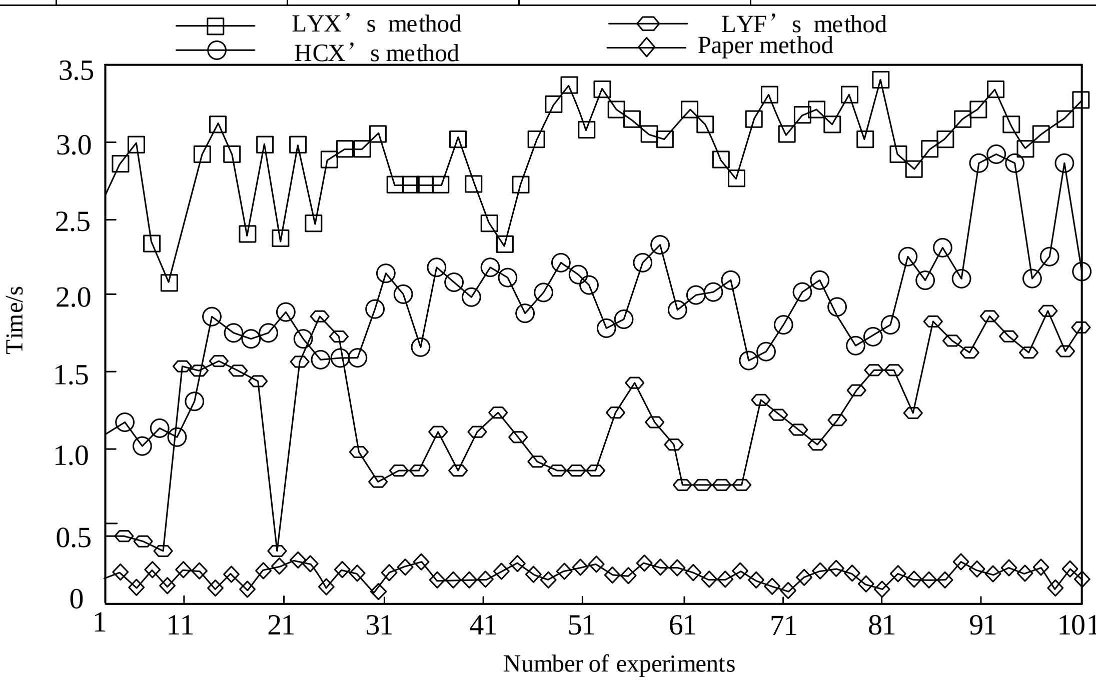


Fig. 8. Comparison of denoising time of interior design images

3) Comparison of denoising time of interior design images: The denoising time of interior design images of LYX's method, HCX's method, LYF's method and proposed method in this study are compared. The results are shown in Fig. 8.

By analyzing the data in Fig. 8, it can be seen that the denoising time of interior design image of LYX's method is between 2.1s and 3.4s, the denoising time of interior design image of HCX's method is between 1.0s and 2.9s, the denoising time of interior design image of LYF's method is between 0.3s and 1.8s, The denoising time of this method for interior design images is always less than 0.3s, which shows that the denoising time of this method for interior design images is shorter and more effective.

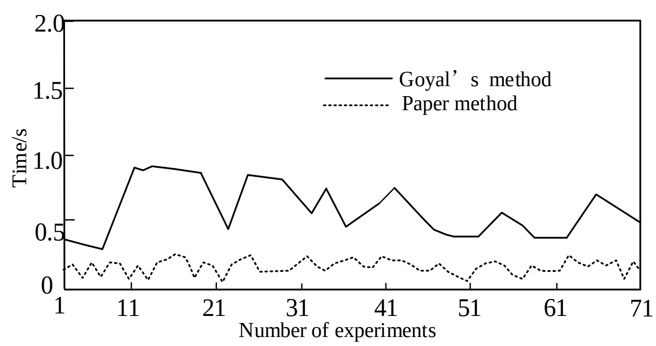


Fig. 9. Comparison of denoising time of interior design images

4) Comparison of image denoising time with the latest method: The method proposed in this paper and Goyal's method are used to denoise the interior design image. The results are shown in Fig. 9.

By analyzing the data in Fig. 8, it can be seen that the denoising time of Goyal's method for the design image is between 0.35s and 0.98s, and the overall time required is still greater than that of the research method. This shows that the method has shorter and more effective noise reduction time for interior design images.

V. Conclusion

With the development of computer technology, the functions provided by digital interior design software have become more and more abundant. The application of relevant software can completely replace the traditional manual design method. Digital room design software can quickly and conveniently complete sketch design. In short, it is to process the two-dimensional plane image through computer software, so that it can be presented in a digital way, so that people can understand the content of the image in a more intuitive way. However, in the process of interior design image generation, interior images are easily interfered by various external factors, resulting in the rise of image noise. Therefore, this paper proposes an interior design image denoising method based on median filter algorithm. The experimental results show that the image denoising effect of this method is better. The average signal-to-noise ratio of interior design image is 54.6dB, and the denoising time of interior design image is always less than 0.3s. It can realize accurate and rapid denoising of interior design image and improve image quality.

References

[1] N. Takahashi, T. Sakamoto, H. Shoji, et al. "Estimation of impression of store interior design based on color features extracted using object segmentation," International Symposium on Affective Science and Engineering, vol. 21, no. 1, pp.1-4, 2021.

[2] S. Vanhee, B. Vandevoort, F. Floré, et al. "Beyond Distinction-Based Narratives, Interior Design's Educational History as a Knowledge Base," Journal of Interior Design, vol. 46, no. 4, pp. 13-25, 2021.

[3] Ly, Z. B. Yi, X. A. Xian, et al. "Interior design of hierarchical micro/nanostructures for enhancing energy storage ability of polyanilines through frozen interfacial polymerization," Electrochimica Acta, vol. 386, no. 1, pp. 138448-138457, 2021.

[4] L. Y. Xian, M. Kang, "Spatiotemporal frequency domain video denoising algorithm in low illumination scene," Modern Computer, vol. 17, no. 1, pp.128-134, 2021.

[5] H. Chaoxiu, "Research on image denoising based on depth learning," Journal of Sichuan Vocational and Technical College, vol. 32, no. 1, pp. 151-158, 2022.

[6] L. Y. Fan, "Research on interior design system of digital processing technology and image processing technology," Modern Electronic Technology, vol. 44, no. 9, pp. 38-42, 2021.

[7] G. Chen, W. U. Guiyan, "Design of real-time video processing system based on ZYNQ chip," Modern Electronics Technique, vol. 42, no. 6, pp. 76-78, 2019.

[8] M. Zhilevski, V. Hristov, "Design of an automated railway crossing system with Verilog language in CPLD" IOP Conference Series Materials Science and Engineering, vol. 878, no. 1, pp. 012046-012057, 2020.

[9] X. Chen, S. H. Huang, X.H. Zhang, "Fast turn off control strategy of cascaded multilevel PWM signal based on FPGA/CPLD," Micromotor, vol. 54, no.1, pp. 39-44, 2021.

[10] D. Cea-Dominguez, J. C. Moure, J. Bartrina-Rapesta, et al. "GPU-Oriented Architecture for an End-to-End Image/Video Codec Based on JPEG2000," IEEE Access, vol. 8, no. 1, pp. 68474-68487, 2020.

[11] T. Brahimi, F. Khelifi, A. Kacha, "An efficient JPEG-2000 based multimodal compression scheme," Multimedia Tools and Applications, vol. 12, no. 6, pp. 1-11, 2021.

[12] G. Nurhayata, I. W. Sutaya, K. U. Ariawan, "The development of prepaid water meters based on AT89S52 Microcontroller," Journal of Physics: Conference Series, vol. 1810, no. 1, pp. 012004-012012, 2021.

[13] G. X. Tan, Y. C. Wang, D. D. Fu, et al. "SBW signal processing and control strategy based on median filter," Computer Simulation, vol. 38, no. 7, pp. 134-13, 20218.

[14] Y. Zhuang, Q. Liu, C. Qiu, et al, "A Handwritten Chinese Character Recognition based on Convolutional Neural Network and Median Filtering," Journal of Physics Conference Series, vol. 1820, no. 1, pp. 012162-012173, 2021.

[15] S. Agarwal, K. H. Jung, "Median filtering forensics based on optimum thresholding for low-resolution compressed images," Multimedia Tools and Applications, vol. 81, no. 5, pp. 7047-7062, 2022.

[16] Mursal, H. Ibrahim, "Median Filtering Using First-Order and Second-Order Neighborhood Pixels to Reduce Fixed Value Impulse Noise from Grayscale Digital Images," Electronics, vol. 9, no. 12, pp. 2034-2045, 2020.

[17] G. Wu, W. Zheng, "The Local Median Filtering Method for Correcting the Laser Return Intensity Information from Discrete Airborne Laser Scanning Data," Remote Sensing, vol. 12, no. 10, pp. 1681-1692, 2020.

[18] J. Zhang, Y. Liao, X. Zhu, et al. "A Deep Learning Approach in the Discrete Cosine Transform Domain to Median Filtering Forensics," IEEE Signal Processing Letters, vol. 27, no. 99, pp. 276-280, 2020.

[19] T.Wang, M. Zhou, "Integrating rough set theory with customer satisfaction to construct a novel approach for mining product design rules," Journal of Intelligent and Fuzzy Systems, pp. 1-25, 2021.

[20] C. F. Chien, H. J. Wu, "Integrated circuit probe card troubleshooting based on rough set theory for advanced quality control and an empirical study," Journal of Intelligent Manufacturing, pp. 1-13, 2022.

Application of CAD Aided Intelligent Technology in Landscape Design

Juan Du
Department of Architectural Engineering, Shijiazhuang University of Applied Technology, Shijiazhuang, China

***Abstract*—The current landscape design methods ignore the depth rendering of scene elements, resulting in low spatial utilization of landscape plant diversity index and landscape spatial pattern. Therefore, this study explores the application of CAD in landscape design. AutoCAD aided intelligent technology is adopted to display the scene in multiple directions and from all angles with terrain design, planning design and planting design as the main contents. Using 3D graphics engine to render landscape elements. On this basis, the spatial coordination planning model of plant landscape is established. The color attribute of landscape space staggered pattern is added to 3D visual reconstruction model by image library function, and the CAD intelligent technology is applied in landscape design. The results shows, the method scored higher in graphic refresh rate, visual brightness and visual contrast, a higher plant diversity coefficient in multiple iterations, and a higher spatial utilization ratio of the landscape pattern than the other two design methods, and the spatial utilization ratio of the landscape pattern of the proposed method is higher than that of the reference method.**

***Keywords*—Landscape design; CAD intelligent technology; engine rendering; image library function; 3D visual reconstruction; digital design; landscape architecture**

I. Introduction

Landscape design requires consideration of relevant theories, based on landscape planning and the integrated use of scientific, technology and art means to create and protect the outdoor environment. The design work needs to be both aesthetically pleasing and coherent [1]. These landscape elements provide a lot of material for creating a high-quality urban spatial environment, but to form a unique urban landscape, it is necessary to organize various landscape elements systematically and combine them with the layout to make them form a complete and harmonious landscape system and orderly spatial form. Regardless of the scale of landscape design, it is expected to achieve a step-by-step effect in space organization. Therefore, it is very important for designers to express their works in an all-round, real and even dynamic way. Current landscape design process mostly uses AutoCAD to draw the 2D construction drawings [2].

Contemporary society with the development of information technology, production methods from the past empirical to knowledge-based, scientific and information direction transition, as the human habitat of the landscape because of the impact of computer network technology, multimedia technology, electronic communication technology is showing unprecedented changes. The landscape design is also deeply engraved with the brand of information technology. Multimedia technology, computer network technology and other information technology are becoming an indispensable part of contemporary landscape design, and they are also an important driving force of landscape design changes. Through the use of information technology and from the information expression of the landscape, people feel the comprehensive impression of the landscape's efficiency and smoothness [3]. As the most widely used CAD software at present, it has rich drawing and editing functions, and provides convenient operation under friendly user interface, so it has a good reputation in the design industry. In particular, AutoCAD software can also provide various editing interfaces and tools, and strengthen the function of 3D modeling and image processing, supporting ActiveXAutomation program design interface.

The study [4] puts forward the reconstruction method of landscape spatial staggered pattern based on the coupling mechanism of pattern and process, analyzes the connotation and extension between landscape spatial pattern and ecological development process, considers that human activities will interfere with and control the reconstruction of landscape spatial pattern, and reveals the external performance and internal mechanism of landscape spatial pattern from the aspects of function and connectivity of landscape spatial pattern, Finally, the steps of landscape spatial pattern optimization are summarized to realize the optimization of landscape spatial pattern. The research [5] proposed a landscape spatial staggered pattern reconstruction method based on remote sensing image, analyzed the landscape spatial pattern characteristics of northern rural areas by using remote sensing image technology, reflected the characteristic changes of northern rural spatial pattern, and calculated the forest land area of northern rural areas by using the patch density and type of landscape spatial pattern, experiments show that proposed method can improve optimization effect of landscape spatial pattern.

The existing methods ignore the depth rendering of scene elements, which leads to the low spatial utilization of landscape plant diversity index and landscape spatial pattern. Therefore, this study explores the application of CAD in landscape design. Therefore the study uses CAD as the development platform.

II. Application of CAD Intelligent Technology in Landscape Design

A. CAD Software in Landscape Image Processing

AutoCAD (hereinafter referred to as "CAD") [6]: developed by Autodesk, USA, is the core software of choice for landscape design. It has powerful 2D graphics editing function and powerful 3D drawing and solid modeling function. It assisted intelligent technology to generate vector graph, its main feature is to generate a variety of editable straight lines, circles, text and size. It has many other editing features, such as copy, pruning, mirroring, and rotation. Mainly used for the drawing of landscape plans, elevations, sections, construction drawings, so that the main line of landscape graphics. Before making landscape, it is better to collect and set up various elements of the library system, which is of great significance to the completion of image processing quickly and beautifully. Different software needs different atlas types, according to the mountains, rocks, plants, architecture, sculpture, sketching and other related factors, you can establish different atlas system. In addition, thinning classification can improve the efficiency of graphics processing.

CAD has become an important technology for landscape designers, and CAD technology is widely used in the landscape design process because of its convenience, creativity and visual impact. It has become a mainstream technology for landscape design due to its easy modification and fast and beautiful drawing. At present, the methods applied to landscape design are mainly divided into two categories: one is the application of GIS technology; the other is the application of traditional image processing software. The former is mostly used in the planning and design of scenic spots, which can generate real topography, landscape and background, and manage and analyze geographic data. And the latter is more widely used in garden planning, elevation drawings, section drawings and rendering, which can draw more beautiful effects. The software can be divided into several modules of modeling, editing, terrain design, planning design, planting design, sprinkler design, data statistics, construction drawing production, rendering and animation production, and the overall structure of the software is shown in Fig. 1.

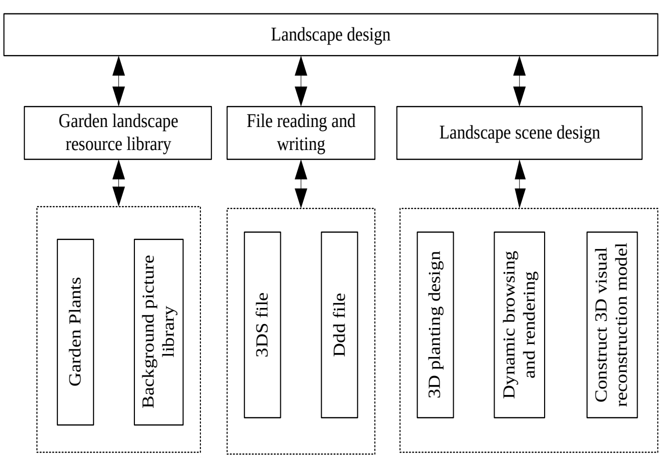


Fig. 1. Application of CAD in landscape architecture design

B. Scene Multi Feature 3D Rendering

In the digital technology of garden design, the system through three-dimensional animation, three-dimensional rendering, virtual scene design, the design intention is well expressed, so that the design technology of garden landscape from the original two-dimensional construction drawing stage gradually upgrade, really reach the stage of three-dimensional simulation, so that the efficiency of garden design greatly improved, in China has been applied as a commercial software, its development prospects are very promising [7-8]. 3D graphics engine is a set of functional modules that allow building full-fledged graphics applications with minimal effort. The visualizer has the ability to extend the architecture, so developers can create their own object classes, inherit them from existing objects, and thus place their own properties and rules in them. Classes can be set up as needed until the objects in the scene are rendered. 3D graphics engine creates graphics applications that use a geometric representation of the scene, i.e. lines, surfaces, edges and other objects drawn in the active window. The direct interaction of the user with the graphics application has some kind of representation of the scene display. This can complicate the use of efficient methods if the view does not have a hierarchical structure, but is represented, for example, by a simple linear array of objects. Using hierarchical scene views has several advantages. In this case, scene objects have very similar functionality and can be grouped into individual groups and then combined into higher-level groups. The advantages that platform has over 2D design drawings are shown in Table I.

Platform can be used to build the texture and transparency of water body, but it cannot reflect the refraction and reflection of the dynamic water body under the illumination of light. Static water body needs to be built by platform software first, and then the refraction and reflection of water body can be realized by animated walkthrough software, so that the water body model has strong vividness and authenticity [9, 10]. There are a large number of tree models in landscape design. If the number of trees is too large, the data and rendering consumption will be too large. TreeEngine engine can be used to load and render tree models in landscape. The tree model rendering consumption can be effectively reduced by adding the branch and leaf information of the tree model to the hierarchical detail model and building the leaf information with the Billboard model. The TreeEngine engine effectively draws and renders parameterized tree models, and includes all the functions of the OSG scene. Load the created tree model into the rendering process of the created 3D terrain scene model, as shown in Fig. 2.

TABLE I. COMPARISON WITH PREVIOUS METHODS

/	2D design drawings	Platform
Models	2D	3D
Similarity	Possible detail errors	Realistic restoration
Reuse	Inheritance is strong and conducive to innovation	Inheritance is not possible

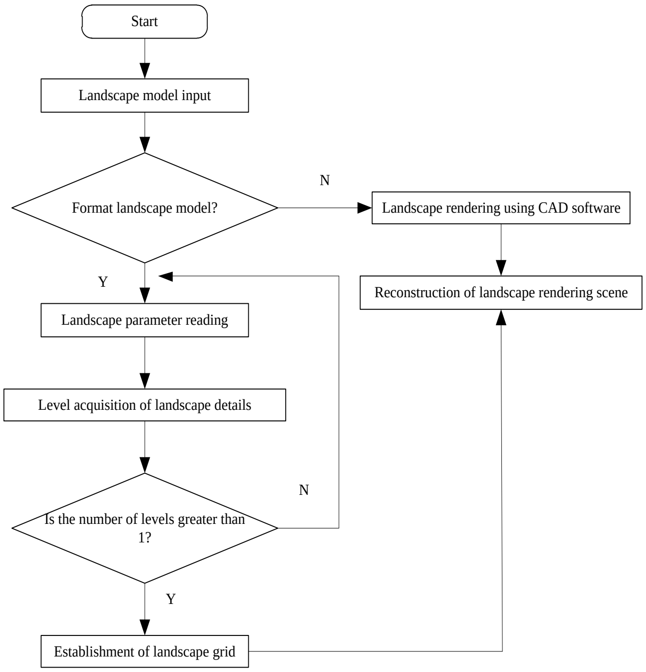


Fig. 2. Tree model rendering flow chart

The tree model is read into the TreeEngine engine, the tree format is parameterized, and the tree multi-level detail is obtained. When the multi-level detail level is more than one, the tree trunk and the whole texture and mesh are established.

C. Spatial Coordination Planning Model of Plant Landscape

According to the analysis results of landscape pattern, the following formula is used to extract the spatial information characteristics of ecological plant landscape in green city:

$$Q(x,y) = Q^n(x,y) + \Re \tag{1}$$

In the formula, $Q(x,y)$ represents the feature information existing in the landscape group space, $Q^{(n)}(x,y)$ represents the feature information existing in the landscape group space in block n, $\Re$ represents the feature extraction coefficient, and (x,y) represents the position of the landscape group in the whole ecological plant landscape space.

Combining the information of decision variables, the relationship between landscape spatial planning and different decision variables was determined. The utility function is used to solve the problem of ecological plant landscape spatial planning from the perspective of ecological plant landscape spatial coordination planning. Through the above analysis, the following formula is used to describe the spatial planning of green urban ecological plant landscape:

$$\max R = \sum_{i=1}^{I} \Im_i \cdot Q(x,y) \cdot \delta \tag{2}$$

Formula (2) satisfies the following formula:

$$\Im_i = Q_i(x,y), i = 1, 2, \cdots \tag{3}$$

In the formula, I represents the number of planning objectives, $\Im_i$ represents the importance of i corresponding to ecological plant landscape management objectives, δ represents the rationality coefficient of spatial planning, and q_i represents the value of i corresponding to the objective variable.

Utility functions generally consist of sub-efficacy functions (Fig. 3), restrictive efficacy functions (Fig. 4), evenly increasing efficacy functions (Fig. 5) and evenly decreasing efficacy functions (Fig. 6).

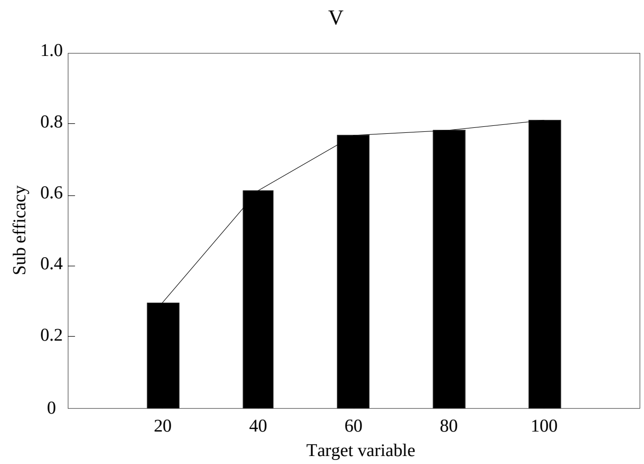


Fig. 3. Sub efficacy function

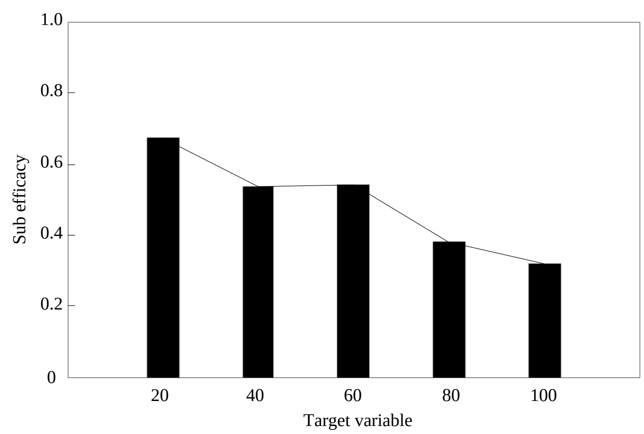


Fig. 4. Restricted efficacy function

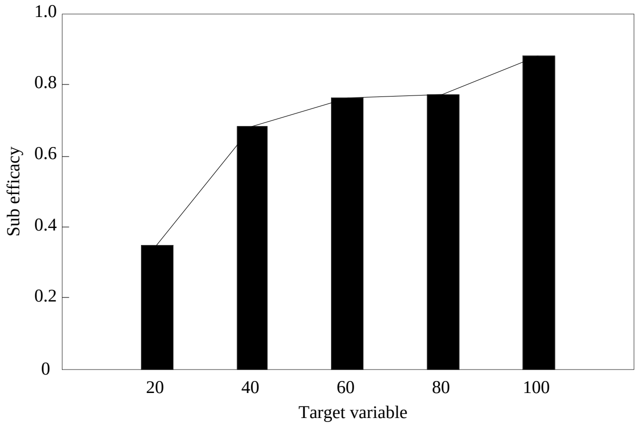


Fig. 5. Uniformly increasing efficacy function

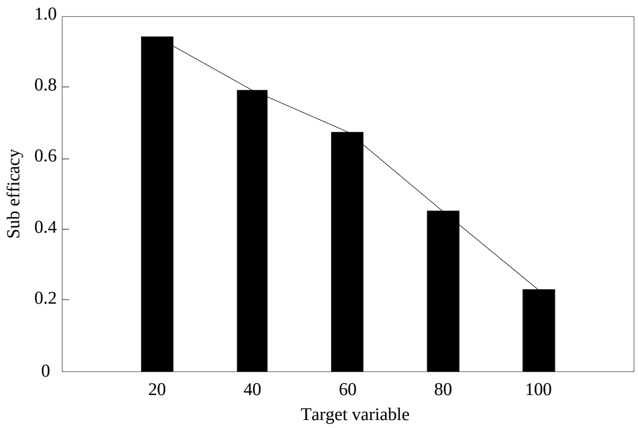


Fig. 6. Efficacy function of uniform decline

Taking green urban ecological plant landscape planning as an object, a green urban ecological plant landscape coordination and planning model is established by maximizing the utility function combined with basic model and utility function theory:

$$U = \frac{\aleph_1 + \aleph_2 + \aleph_3 + \aleph_4}{\delta \cdot \sum_{i=1}^{I} \Im_i} \tag{4}$$

Where, $\aleph_1$, $\aleph_2$, $\aleph_3$ and $\aleph_4$ represent the weight of sub objectives.

D. 3D Planting Design

a) Description of planting method: 3D graphics engine can provide a lot of drawing tools, such as grass point filling, garden corridor, water barge, slate road, gravel road, hedge line, forest edge line, etc., and can also be used to mark plants, positioning line, elevation, garden road, angle, radius and area, etc. The system also has similar to AutoCad fill pattern function and rich content of the two-dimensional library, and the corresponding effect can be printed. The system also has similar functions to AutoCad and a rich library of 2D drawings, and can print out the corresponding effect drawings. In general, the system is used for parameterization, definition of plots, road planning, generation of steps, fences and walls, ponds, planters, planters, ramps, curbs, and other landscape models and design of mountain roads. The design function also provides a rich plant database and pictures, and records the various plant habits, providing a variety of planting methods such as mixed planting, piece planting, column planting, and solitary planting, as well as modifying the number and properties of some already planted plants, and expanding the planting on flat surfaces to undulating terrain. The expression of plantation plan should meet the standard of landscape design. It should be shown according to plantation legend or plantation point and plant contour. In order to achieve the unity of 2D and 3D planting design, the planar planting plan is displayed in the top view and the plantation plan is displayed in the perspective view.

b) Methods of representation for plants: In landscape construction, many plants may be used. If a 3D model is used to represent plants, then a plant will use a large number of patches, which back use a larger system space and even cause the software to crash. Based on the above reasons, the system provides another plant representation, using platform bulletin board, using pictures to simulate the plant. Bulletin board is a very useful technology, it can be a simple way to achieve a lot of special effects, but it is most attractive to achieve these effects with very low system resources. So for each plant entity, it only needs to store its location and the index information of the picture, instead of the storage cost of the 3D model, so it saves the system space greatly. In the display effect, the system will collect nearly a thousand kinds of plants in the south and north of the picture have done the hollow processing, using platform expansion function to achieve the hollow texture display effect, the results more realistic.

E. Dynamic Browsing and Rendering Technology

After the creation of a 3D model, after assigning materials to entities and setting up light sources and viewpoints (cameras), dynamic browsing can be conducted at any time, 3D realistic rendering or animation can be made, and the rendering effects and texture can be reflected by the rendering effect pictures. As can be seen, the difficulty of building a three-dimensional model is that it is difficult to achieve a large volume and a large amount of data for the implementation of the model to browse. The landscape design software uses a more cost-effective display card and its mature interface language for acceleration. At the same time, data processing is greatly reduced by smoothing the special 3D mesh. Landscape design software rationally uses the triangle belt and triangle fan mode provided by platform to reduce the repetition of records on nodes and reduce number of nodes representing stereotyped objects [11-12].

In addition, the use of a specific 3D mesh surface smoothing processing also greatly reduces the data processing, in practical applications played a significant effect. Since the platform requires the map to be a power of two in size. Therefore, it will be sorted when it is passed to the platform. Considering the characteristics of landscape design, we cannot use a simple cutting method, but a combination of color compression, color combination and softening technology to achieve the desired results.

Through reading transparent channel information of 32-bit color bitmap and combining with application of platform template, the hollowed-out texture display technology is realized. The tree image is mapped by hollowing out, and the tree image size is set to power of two to improve the display accuracy. In the aspect of realistic rendering and rendering algorithm, by comparing the depth buffer, ray tracing and radiance, the improved radiance algorithm is used to solve the problem of low rendering speed. The light module in the platform can realize the whole scene, and can display various effects such as daytime and nighttime scenes. In a daytime environment, ambient lighting consists of “parallel light”, “light” and “sky light” by way of simulated sunlight. The camera is equivalent to another point and determines the final rendering type. The landscape design software can drag and modify the position of the camera, display the position and angle of the camera in real time, and display the observation range of the camera in the camera view to achieve a WYSIWYG effect. Meanwhile, the software also supports six types of light sources, including point light source, cone light source, parallel light source, column light source, surface light source and sunlight source. Each light source can be interactively set and modified.

F. Construct 3D Visual Reconstruction Model

Based on the index of landscape spatial staggered pattern, the characteristic points of landscape spatial staggered pattern are mapped into the 3D visual reconstruction model, and the 3D visual reconstruction model is established [13-14]. If there are A points in the 3D visual reconstruction model, the space coordinate of the scanning base station of texture mapping is (X,Y,Z), and the a point coordinate of the staggered pattern of A points and landscape space is E, the mapping relationship between (x,y) points and a points of the staggered pattern of landscape space in the 3D visual reconstruction model is as follows:

$$\begin{cases} x = h_1(X,Y) \\ y = h_2(X,Y) \end{cases} \tag{5}$$

$$\begin{cases} x = \sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{ij} X^i Y^j \\ y = \sum_{i=0}^{n} \sum_{j=0}^{n-i} b_{ij} X^i Y^j \end{cases} \tag{6}$$

Because of the error of data transmission and the noise in the feature data of landscape pattern, the coordinates of X and Y axes are different, but the coordinates of Z axes are the same. In order to make the mapping result more accurate, the threshold σ is introduced, and the A_i is assumed to be a point in the 3D visual reconstruction model. The threshold σ is used to find A‘s adjacent point $\dot{A}_i$ in the A_i-point range. By expanding formula (6), the mapping model of A and $\dot{A}$ from 3D visual reconstruction model to point a can be obtained:

$$\begin{cases} h_1(X,Y) = x = A_{00} + A_{10}X + A_{01}Y + A_{20}X^2 + A_{11}XY + A_{02}Y^2 \\ h_2(X,Y) = y = B_{00} + B_{10}X + B_{01}Y + B_{20}X^2 + B_{11}XY + B_{02}Y^2 \end{cases} \tag{7}$$

$$\begin{cases} \dot{h}_1(\dot{X}_i,\dot{Y}_i) = C_{00} + C_{10}\dot{X} + C_{01}\dot{Y} + C_{20}\dot{X}^2 + C_{11}\dot{X}\dot{Y} + C_{02}\dot{Y}^2 \\ \dot{h}_2(\dot{X}_i,\dot{Y}_i) = D_{00} + D_{10}\dot{X} + D_{01}\dot{Y} + D_{20}\dot{X}^2 + D_{11}\dot{X}\dot{Y} + D_{02}\dot{Y}^2 \end{cases} \tag{8}$$

In the formula, A represents the landscape autocorrelation coefficient; B represents landscape heterogeneity coefficient; C represents the basic characteristics of landscape pattern; D represents the global characteristics of landscape pattern [15].

Using formula (7) and formula (8), it can be obtained that the mapping model includes 24 positions. When the number of 3D visual reconstruction model and its corresponding landscape spatial staggered pattern is greater than 12, the 3D visual reconstruction models of $h_1(X,Y)$, $h_2(X,Y)$, $\dot{h}_1(\dot{X}_i,\dot{Y}_i)$ and $\dot{h}_2(\dot{X}_i,\dot{Y}_i)$ can be calculated according to the principle of least square method.

Substituting $\dot{A}_i(\dot{X}_i,\dot{Y}_i,\dot{Z}_i)$ into formula (7), respectively obtain that the coordinates of $\dot{a}_i$ are $(\dot{x}_i,\dot{y}_i)$ and the distance between $\dot{a}_i$ and a is d_1; Substitute $A(X,Y,Z)$ into formula (8) and obtain that the coordinate of a' is (x', y') and the distance between a' and a is d_2.

Comparing the size of d_2 and d_1, the size of the sum is compared and the corresponding polynomial with smaller distance is selected to express the mapping relationship of the 3D visual reconstruction model.

III. Experimental Design of Performance Test of the Proposed Method

A. Presentation of Landscape Design Effect

Select Visual Studio 2021 as the development platform, using VC+ language as the development language to develop this landscape design software. The landscape design of a residential quarter in a certain city is selected as the verification object. The residential quarter is located in the southeast of the city and is a newly developed ecological residential quarter. The total area of the residential quarter is 38759m^2, the total construction area is 16748m^2, and the residential quarter includes eight residential buildings, including villas, multi-storey areas and high-rise areas. The landscape design requires that the green area within the residential quarter shall be more than 40%.

As can be seen in Fig. 7, the 3D model built using the platform can contain all the architectural information needed in the relevant landscape project, and the 3D terrain scene model and the actual building can be presented in a 1:1 scale, which makes the design drawings more intuitive and easy to modify.

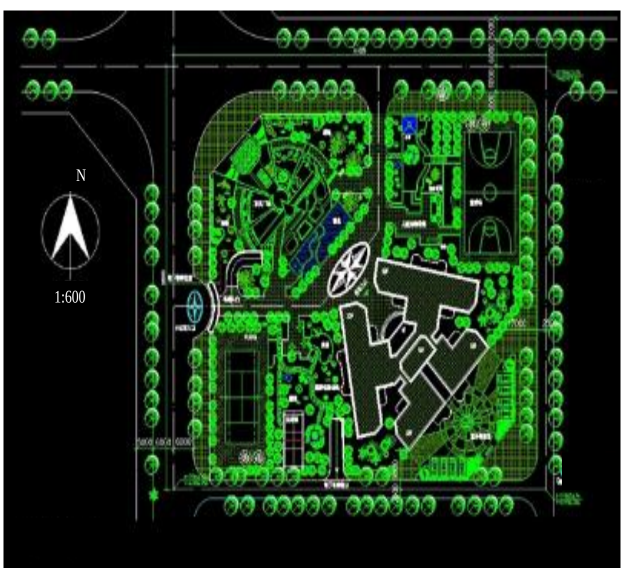


Fig. 7. Reconstruction of 3D terrain scene model

Fig. 8. Final effect of landscape design

The final effect drawing of landscape design for this residential area is shown in Fig. 8.

The simulation results of Fig. 8 show that the proposed method can effectively reduce the labor intensity of landscape designers and shorten the design cycle. Compared with the 2D landscape design, the 3D terrain scene model is more accurate. Because the 3D terrain scene model is 1:1, the accuracy of landscape design is improved effectively. The method of this paper can be used to design landscape in 3D scene model.

Analytic Hierarchy Process is selected to evaluate the effect of landscape design in this paper. Analytic Hierarchy Process is a quantitative and qualitative multi-objective decision analysis method. The results of AHP evaluation of the landscape designed by this method are shown in Table I.

TABLE II. Landscape Design Results of this Method

Evaluating indicator	Weight	Evaluation results / score
Illumination	0.08237	94
Visual contrast	0.09263	98
Resolving power	0.08356	94
Visual brightness	0.06657	91
Shadow processing	0.16849	98
Graphics refresh rate	0.26478	99

As can be seen from Table II, the AHP method can be used to evaluate the final effect of the landscape design using the 10 indicators listed in the table to rate each indicator, and the average of each score sought is 93.7. The result shows that the method has high design effectiveness.

B. Application Performance Test of the Proposed Method

In order to the overall effectiveness of the proposed method, it is necessary to test the analysis method of green urban ecological plant landscape spatial coordination planning model. This test is completed in Simulink platform. Set the plant diversity coefficient in the interval [0,1]. The larger the plant diversity coefficient, the more categories of plants in the landscape space, and the better the diversity. The proposed method, the landscape spatial design method based on pattern coupling mechanism proposed in study [4] and the landscape spatial design method based on remote sensing image proposed in research [5] are tested respectively.

The plant diversity coefficients of different methods are shown in Fig. 9.

By analyzing the data in Fig. 9, it can be seen that the plant diversity coefficient obtained by the proposed method in multiple iterations is more than 0.8, and the plant diversity coefficient obtained by the landscape spatial design method based on pattern coupling mechanism proposed in study [4] fluctuates around 0.5. The plant diversity coefficient obtained by research [5] method in multiple iterations fluctuates around 0.5. Comparing the test results of different methods, it can be seen that the plant diversity coefficient of the proposed method is high in multiple iterations.

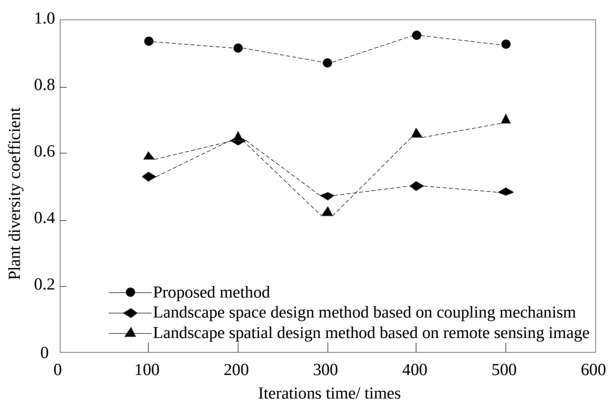


Fig. 9. Plant diversity coefficients of different methods

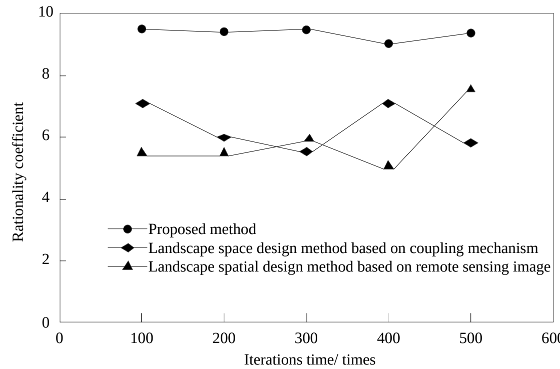


Fig. 10. Rationality coefficients of different methods

The above methods are further tested through the rationality coefficient of green urban ecological plant landscape spatial planning. The rationality coefficients of different methods are shown in Fig. 10.

According to the data in Fig. 10, when the proposed method is used to design the ecological plant landscape space of green city, the rationality coefficient is more than 8, When the landscape space design method based on pattern coupling mechanism proposed in study [4] and the landscape space design method based on remote sensing image proposed in research [5] are used to plan the ecological plant landscape space of green city, the rationality coefficients obtained are within the interval [4, 8]. The comparison shows that the rationality coefficients of the proposed methods in multiple iterations are higher than those in references. Because the proposed method uses 3D graphics engine to render garden landscape scene elements. Based on this, the spatial coordination planning model of plant landscape is constructed. Color attribute of landscape spatial staggered pattern is added to 3D visual reconstruction model through image library function, which improves the planning rationality of the proposed method.

The landscape spatial design method based on pattern coupling mechanism proposed in study [4], the landscape spatial design method based on remote sensing image proposed in research [5] and the proposed method are used to calculate the utilization rate of landscape spatial pattern. The calculation formula is as follows:

$$\eta = \frac{S_0}{S} \quad (9)$$

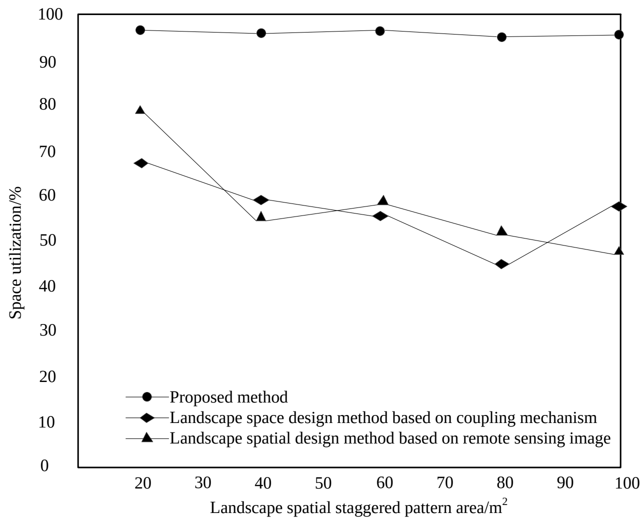


Fig. 11. Comparison results of landscape spatial pattern utilization

In the formula, S_0 represents the area of the utilized landscape spatial pattern; S represents the area of reconstructed landscape spatial pattern. The results are shown in Fig. 11.

As can be seen from the results in Fig. 11, the spatial utilization rate of landscape spatial pattern is significantly higher than that of the other two design methods, indicating that the design method has advantages in spatial utilization rate and can effectively improve the spatial utilization rate of landscape spatial pattern.

IV. CONCLUSION

This study explores the application of CAD in landscape design. It has powerful 2D graphics editing function and powerful 3D drawing and solid modeling function. In this paper, platform technology is used to generate landscape vector map. Topographic design, planning design and planting design are main contents. Landscape can be dynamically browsed and displayed. After reconstructing terrain and environment information of landscape to be designed by binocular stereo vision camera, landscape elements are rendered by 3D graphics engine. Based on this, coordination and planning model of plant landscape space is constructed by maximizing utility function combined with basic model and utility function theory. The color attribute of landscape space staggered pattern is added to the 3D visual reconstruction model by using image library function, and the platform intelligent technology is applied in landscape design. The results show that the design evaluation index score of the proposed method is higher, the plant diversity coefficient is higher, and the spatial utilization ratio of the landscape spatial pattern is higher.

REFERENCES

[1] Dezio, “Agritourism and slow lines: hybrid practices for a landscape design model to support agriculture in mountain regions,” Vermont as read from an Italian perspective. Ciudades, vol 20, no. 24, pp. 79-98, 2021.

[2] Y. Wei, “Environmental monitoring and landscape design of green city based on remote sensing image and improved neural network,” Environmental Technology & Innovation, vol. 23, no. 1–2, pp.101718, 2021.

[3] S. Radley, “Humphry Repton: Landscape Design in an Age of Revolution: by Tom Williamson, London, Reaktion Books, 2020, 312 pp.136 illustrations, 114 in colour, 35.00 (hbk), ISBN: 9781789142990. Landscapes,1-2,” 2021.

[4] S. L Kang, B. R. Yue, “The Method of Landscape Spatial Pattern Optimization Based on Coupling Mechanism of Patterns and Processes,” Chinese Landscape Architecture, vol.33 no.3, pp.50-55,2017.

[5] Han, “Study on Spatial Pattern Characteristics of Rural Landscape in Northern China Based on Remote Sensing Images,” Bulletin of Science and Technology, vol. 34, no. 11, pp. 176-179, 2018.

[6] P. Shan, W. Sun, “Auxiliary use and detail optimization of computer VR technology in landscape design,” Arabian Journal of Geosciences, vol. 14, no. 9, pp.1-14, 2021.

[7] P. Shan, W. Sun, “Research on landscape design system based on 3D virtual reality and image processing technology,” Ecological Informatics, no. 9, pp.101287, 2021.

[8] S. Yang, J. Yang, “Application Prospect of CAD-SketchUp-PS Integrated Software Technology in Landscape Planning and Design,” Computer-Aided Design and Applications, vol. 18, no. S3, pp. 153-163, 2020.

[9] M. Zhang, X. Deng, “Color Effect of Landscape Architecture Design under Computer Aided Collaborative Design System”. Computer-Aided Design and Applications, vol.19, no. S3, pp.13-22, 2021.

[10] M. Khakzand, K. Aghabozorgi, “Developing a Creative Landscape Design Process Based on the Interaction of Architecture and Landscape,” The International Journal of Design Education, vol.15, no.2, pp.133-151, 2021.

[11] Kot, D. Chan, T. Chung, H, “TsuiImage Rendering Techniques in Postmortem Computed Tomography: Evaluation of Biological Health and Profile in Stranded Cetaceans,” Journal of Visualized Experiments, vol. 163, no.163, pp. 61701, 2020.

[12] P. Paudyal, F. Battisti, P. L. Callet, J. Gutiérrez, “Perceptual Quality of Light Field Images and Impact of Visualization Techniques,” IEEE Transactions on Broadcasting, pp. 1-14, 2020.

[13] H. Huang, X. Yang, S. Y. Tan, Peng, “Fast Reconstruction of 3D Point Cloud Model Using Visual SLAM on Embedded UAV Development Platform,” Remote Sensing, vol. 12, no.12, pp.3308, 2020.

[14] M. Zachariou, N. Dimitriou, O. “Arandjelovi, Visual Reconstruction of Ancient Coins Using Cycle-Consistent Generative Adversarial Networks,” Sci, vol. 2, no. 3, pp. 52, 2020.

[15] Y. B. Wang, T. T. Li, “Simulation of Urban Landscape Pattern Evolution and Dynamic Characteristic Change Prediction,” Computer Simulation, vol.38, no. 04, pp. 417-42, 20211.

Filtering and Enhancement Method of Ancient Architectural Decoration Image based on Neural Network

Yanan Wang
Architectural Engineering Institute, Sanmenxia Polytechnic, Sanmenxia, China

***Abstract*—Due to poor ambient light or uneven lighting, the old decoration image acquisition methods are easy to cause the image blur. To solve this problem, this paper proposes a neural network-based filtering enhancement method for ancient architectural decoration images, which preserves image details by enhancing contrast, smoothing noise reduction and edge sharpening. Based on the convolutional neural network which is composed of encoder, decoder and layer hop connection, the residual network and hole convolution are introduced, and the hole U-Net neural network is constructed to fuse the pixel feature blocks of different levels. This method enhanced the image contrast according to the gray level and frequency histogram, and aiming at the gray value of the pixel to be processed in the image. And the middle value of the gray value of the neighborhood pixel is used to filter the noise of the ancient building decoration image. The paper also analyzes the joint strength of beams and columns in ancient buildings, and calculates the elastic constants of beams and columns and the stress at the joint of them, considering the image texture characteristics of the wood in ancient buildings with the mortise and tenon connection of beams and columns. Experimental results show that the proposed method has good noise suppression performance, can effectively obtain image detail features, and significantly improve the subjective visual effect of ancient architectural decoration images.**

Keywords*—Neural network; decorative images of ancient buildings; filter enhancement method; encoder; decoder; pixel gray value

I. Introduction

Ancient buildings are one of the most important cultural heritages left over by the history of various countries, marking the national civilization, the pride of various countries and the wealth of all mankind [1-2]. The comprehensive protection of ancient buildings is conducive to the promotion of national traditional culture and can benefit future generations of mankind. On the vast land, there are many distinctive ancient villages and residential buildings, as well as many magnificent royal ancient buildings, which can be said to be countless. These buildings integrate ancient folk culture and architectural art, and are the witness of history and the crystallization of culture [3-4]. Its existence not only leaves exquisite evidence for the development of architectural history and human civilization history, but also provides diversified development reference for contemporary architectural art [5]. With the substantial improvement of people's living standards and cultural level, tourism has developed very rapidly around the world, in which the visit of ancient buildings has become more and more popular, and more places of interest have begun to be open to the public and for the public to visit. More and more developers began to explore the cultural connotation and historical value of ancient buildings, followed by the birth of a series of antique buildings, trying to restore the ancient architectural style in terms of layout and building materials. In recent years, people pay more and more attention to the protection of ancient buildings [6], especially China, which has a long history and many ancient buildings. But after years of erosion, most China ancient buildings have been weathered, corroded and even man-made damage to varying degrees. How to reconstruct the damaged ancient buildings and ensure the high degree of restoration after reconstruction has become the main research problem in the field of ancient building restoration [7]. Image filtering and enhancement processing is an enhancement strategy to improve the decorative features of ancient buildings. It can improve the overall image quality from the aspects of pixel definition, contrast and brightness. However, in daily life, because of the photographer's photography level and surrounding environment, the initial image quality is usually poor. Therefore, it is imperative to design a reasonable and complete image filtering and enhancement processing method.

The body of the study consists of six parts. In Section III, the convolutional neural network is introduced and modified into U-Net neural network, which establishes the basic framework for the proposed method. In Section IV, the beams and columns of ancient buildings are introduced, and their properties such as handover strength, elastic constant and connection stress are calculated. The characteristics of the application objects of the proposed method are investigated to make the method more suitable for this kind of image. In Section V, the image filtering enhancement method based on neural network is introduced in detail. The validity of the proposed method is verified by experimental analysis in the Section VI. Conclusion and prospect are described in Section VII.

II. Related Works

There are many research types in the field of filtering and enhancement of ancient architectural images. The study [8] proposes a quantitative evaluation of the architectural style of Agora, an ancient Greek city in Athens based on image correlation and fractal dimension analysis. Agoras is the center of Western European life and can be called the starting point of western European civilization. In previous studies, image processing technology was used to restore the initial shape of buildings and cities. Then, each building in the three-dimensional model of the square is built around the formation of the square. In this study, the 3D model of temple focuses on temple buildings and corridor extracts the facade of each building, quantitatively evaluates the similarity by performing fractal dimension analysis and image analysis, and considers the correlation of the two analysis results. The research [9] pointed out the evolution of Vladivostok's spatial structure and architectural image, and the problems and contradictions existing in the modern stage of the development of Far East maritime cities, which is the result of insufficient consideration of the basic principles of its formation and development. The solution to these problems is to analyze their spatial structure and architectural image from the perspective of evolution. This method can reveal all the existing shortcomings and contradictions, and formulate the main direction, principles and means of urban environmental construction, so as to improve and humanize the urban environment. From the perspective of evolutionary method and complex analysis method proposed by the author, the development of Vladivostok urban spatial structure and architectural image is considered. In the latest research, the performance of machine learning has been further optimized. The authors [10] proposed a new NA-DE hybrid neural network algorithm. It incorporates NNA and differential evolution algorithms. In the experiment, the controller optimized by this algorithm has better performance. In addition, it shows better robustness to the uncertainty of parameters in the system. The literature [11] proposes a trainable network framework based on attention mechanism. The framework combines model training, feature extraction and band selection. The experimental results show that the model constructed based on this framework has higher regression R^2 value and classification accuracy, and the original spectral space information can be used to effectively select the band while training the model.

Although the above research has made some progress, because only the feature information is enhanced and the enhanced information is not matched, the visual effect of image filtering and enhancement is poor, and it is easy to lose the local detail information of the image. Therefore, this study proposed a neural network-based filtering and enhancement method for ancient architectural decoration images. In this method, residual network and hole convolution are introduced to build hole U-Net neural network, which can fuse pixel feature blocks of different levels, enhance the contrast of decorative images of ancient buildings, and filter the noise of decorative images of ancient buildings.

III. Neural Network

Using cavity convolution with different convolution rates, a neural network [12] is constructed to obtain more and more detailed image information of ancient architectural decoration, provide as perfect a visual image as possible for image filtering and enhancement, and further improve the image enhancement effect.

The neural network composed of encoder, decoder and layer hopping connection is shown in Fig. 1. The encoder is similar to the feature extraction layer, while the decoder mainly carries out an operation similar to deconvolution. In the process of decoding, image features extracted during encoding are also added. Jump joins can bridge the semantic gap between the encoder and the decoder, and it can also recover fine-grained details of the target object. The image is first down-sampled by the encoder to obtain the high-level semantic feature map, and then up-sampled by the decoder to restore the feature map to the resolution of the original image. Each time the decoder is upsampled, the jump connection will fuse the feature maps corresponding to the same resolution in the decoder and the encoder in a splicing way to help the decoder recover the details of the target better.

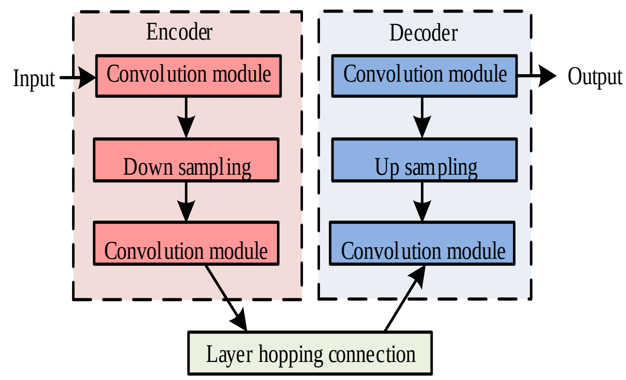


Fig. 1. Neural network architecture

As can be seen from Fig. 1, the encoder and decoder have symmetry, and the number of ancient building decoration image channels and resolution and other parameters correspond one by one. They are respectively composed of two convolution modules. Each convolution module contains two convolution stages, and its functions are to extract the deep features of ancient building decoration image and repair the details of ancient building decoration image; The function of layer hopping connection is to associate the image feature blocks between codec and decoder, and further optimize the image restoration effect of ancient architectural decoration of decoder by using the deep features extracted by encoder feature blocks.

The process of using neural network filtering to enhance the image of ancient architectural decoration is described as follows:

1) Based on the gray level and single channel initial ancient building decoration image, the encoder uses the maximum pool of two steps for down sampling processing, so that the ancient building decoration image channel is multiplied by 2 and the resolution is divided by 2;

2) After the decoder uses the same step deconvolution of 2*2 convolution kernel for up sampling, it also performs the same processing on the ancient building decoration image and resolution. Finally, the number of compression channels is the number of types to be output, and the segmentation result of the ancient building decoration image is obtained.

3) Layer hopping connection integrates the semantic information and orientation information contained in different levels of features through splicing, so that the number of feature channels is one fourth of the input, and accurately segment the ancient building decoration image.

The semantic information and orientation information of ancient architectural decoration image have certain randomness, so the deep and shallow feature weights change constantly with the image content; Moreover, the decoration image of ancient buildings will change with the passage of time, which increases the probability of problems such as blurred boundary, uneven illumination and high noise in the visual image [13], which brings great difficulty to the subsequent processing and application of the image. Therefore, in order to balance the feature weights of different levels of neural network and improve the image quality of ancient architectural decoration, the residual network and cavity convolution are integrated into the neural network to construct the cavity neural network, so as to remove the network gradient and fully extract the multi-scale features of the image.

Assuming that based on the neural network with convolution layers of n, the following formula is used to derive the gradient of the network whose first layer weight is w_1, and the calculation formula is:

$$\frac{\partial Q}{\partial e_1} = \frac{\partial Q}{\partial q_n} \cdot \ldots \cdot \frac{\partial q_3}{\partial q_2} \cdot \frac{\partial q_2}{\partial q_1} \cdot \frac{\partial q_1}{\partial e_1} \tag{1}$$

In formula (1): Q represents the loss function, $Q(y, f(x)) = |y - f(x)|$. $q_i\left(i = 1, 2, \ldots, n\right)$ represents the activation function output with the number of layers of i, and ∂ represents the sign of partial derivative. When the partial derivative of $\frac{\partial Q}{\partial q_n}$ satisfies $\frac{\partial Q}{\partial q_n} < 1$, the more network layers, the faster the network gradient decreases, and infinitely approaches 0.

According to the correlation between the output value of the simultaneous loss function and the number of network layers, the residual network structure composed of convolution layer and curve is used to remove the network gradient. The input items reach the network output through the curve and convolution layer respectively, so the definition formula of residual network structure is as follows:

$$W_a = D\left(n\right) + n \times Q \tag{2}$$

In formula (2): W_a represents the network output part, $D\left(n\right)$ represents the residual network part, and n represents the input item or direct mapping structure.

Neural network hole convolution is to increase the convolution rate index in the convolution core of convolution layer, expand the field of view of convolution layer and prevent the loss of down sampling information. Based on the 3*3 cavity convolution with the expansion rate of 1, when the expansion rate increases to two, insert a zero between the two adjacent points of the convolution core to make it 7*7; when the expansion rate increases to four, insert three 0 between the two adjacent points of the convolution core to make it 15*15, and so on. Thus, a positive correlation between the field of view of cavity convolution and the expansion rate is established.

After the convolution of residual network and hole, the neural network is improved into hole U-Net neural network, as shown in Fig. 2.

According to Fig. 2, the operation process of U-NET is shaped like “U”, which can be divided into two parts: code and decode, including input and output layer, convolution layer and pooling layer, but not full connection layer. Through the network structure, pixel feature blocks of different levels in the decoration images of ancient buildings are fully integrated. Published in 2015, U-Net is a variant of FCN. Its original intention is to solve the problem of biomedical images. Due to its good effect, it has been widely applied in various directions of semantic segmentation, such as satellite image segmentation, industrial defect detection, etc [14]. Its structure is simple but effective, and it can recognize multi-scale features of image features. Therefore, it is selected in this study to expand the neural network.

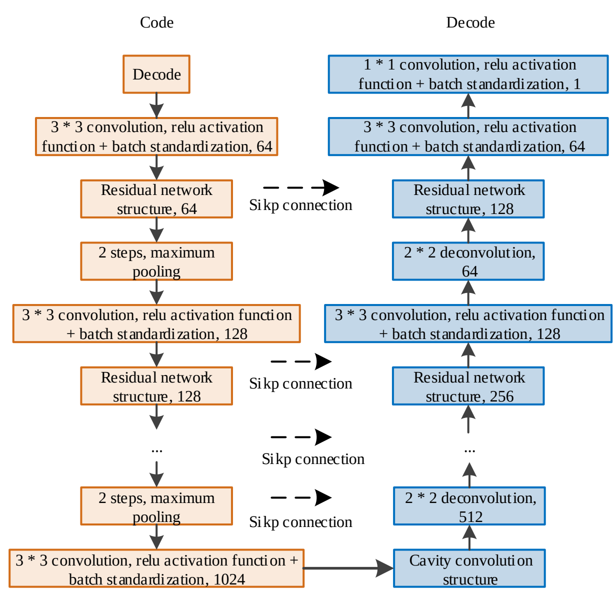


Fig. 2. U-Net neural network architecture

IV. Basic Characteristics of Beams and Columns in Ancient Buildings

A. Orthotropic

In general, wood is constructed according to cells with different shapes, sizes and arrangement order. This construction characteristic also determines the anisotropy of wood. The anisotropy of wood elasticity is obtained by applying the principle of orthogonal symmetry to wood. When wood materials in ancient buildings are cut from different directions, they will show inconsistent texture characteristics. Specifically, they can be divided into four characteristics: along grain direction, chord direction, radial direction and tangent direction. The longitudinal cutting of wood refers to the direction parallel to the wood fiber in the process of actual operation, that is, the cutting direction of the trunk from the bottom line, which becomes the direction along the grain; Chord direction and radial direction refer to the direction perpendicular to wood fiber, which can be called transverse grain direction; In general, after the wood is cut, the center of the trunk will show an annual ring like a concentric circle, while the radial cutting and chord cutting are perpendicular to the annual ring. The only difference is that the radial direction passes through the tangent direction of the tree center, and the chord direction does not pass through the tangent direction.

B. Mechanical Properties of Beams and Columns in Ancient Buildings

1) Flexural strength: Since the bending strength of wood is higher than the strength under external force, in general, the external force strength and bending strength along the grain of wood can be improved respectively, and on this basis, it conforms to the following formula:

$$W_{sd} = \left(3\frac{\sigma_l}{\sigma_y} - 1\right) / \left(\frac{\sigma_l}{\sigma_y} + 1\right) \quad (3)$$

In formula (3), σ_l and σ_y represent the external force strength and bending strength of wood samples respectively.

When the wood in ancient buildings is damaged by bending, the wood fiber under external force on the cross section of the material will be unstable and wrinkled. With the strength of cyclic load, the position of fiber instability will develop to the neutral axis of the section, and finally the wood will break.

2) Bearing strength: In ancient buildings, wood bearing pressure refers to the performance of transmitting load on the cross section of the intersection of beams and columns when they intersect into nodes. The intersection point of force is called bearing stress on this cross section, and the corresponding resistance is called bearing strength. However, because the actual intersecting section is unstable, the bearing capacity along the grain of wood is lower than that along the grain.

When the cross grain of the whole surface is under pressure, the appropriate maximum proportion will be selected as the measurement standard of the bearing strength. The bearing pressure of part of the cross grain is consistent with that of the whole surface. However, due to the bearing pressure of the cross grain on the length of some wood, it is necessary to consider the improvement of the strength caused by the fiber action of the wood on both sides of the bearing pressure point.

In general, the elastic characteristic constants of wood include: elastic modulus E_i, Poisson's ratio μ_{ij} and shear elastic modulus G_{ij}. The elastic modulus of longitudinal L, radial R and chord T of wood cutting is the elastic modulus in the main direction. Assuming $E_L = E_1$, $E_R = E_2$ and $E_T = E_3$. E_L 、 E_R and E_T represents the elastic modulus of wood cut longitudinally, radially and vertically, respectively. And it is determined that there is only one normal stress in the main direction, the ratio of normal stress to linear strain in the corresponding direction can be written as:

$$E_i = \sigma_i / \varepsilon_i \qquad \left(i = L, R, T \quad \text{or } i = 1, 2, 3\right) \quad (4)$$

According to the ratio of formula (4), Poisson's ratio μ_{ij} is the normal stress, and σ_i acts alone in the L direction; when the corresponding bearing stress is zero, the opposite number of the ratio of stress in L direction to strain in R direction is:

$$\begin{aligned} &\mu_{i,j} = -\varepsilon_j / \varepsilon_i \\ &\left(i = L, R, T \quad \text{or } i = 1, 2, 3,\ j = 1, 2, ..., 3\right) \end{aligned} \quad (5)$$

According to formula (5), $G_{L,R}$, G_{RT} and G_{TL} can be described as shear elastic modulus in bearing plane LR, RT and TL respectively.

C. Characteristic Analysis of Beam Column Joints

1) Characteristic points of mortise and tenon structure: Mortise and tenon connection is a semi-rigid connection between rigid connection and hinge connection. Because most ancient building structures adopt straight mortise connection, the straight mortise is mainly analyzed, and the finite element numerical simulation analysis of mortise and tenon connection of building structure is carried out [15]. According to the unique characteristics of the building structure in this technology, the virtual spring element will be used to simulate the semi-rigid structure. In practical application, this spring element has no substantial mass and element.

2) Analysis of tenon joint intersection: In order to calculate and solve the stress effectively and simply, assuming that the shear capacity of wood along the x axis and y axis is the same, and the bending and torsion capacity in different directions is the same, the simplified $k_x = k_1$, $k_y = k_z = k_2$ and $k_{\theta x} = k_{\theta y} = k_{\theta z} = k_3$ can be obtained. Only these three unknown variables need to be solved, and there are:

$$k_x = \frac{F_x}{\Delta_x} \qquad (6)$$

According to formula (6), similarly, the load F_z treatment along the z axis can be carried out on the model to obtain the shrinkage strength of the mortise joint along the z axis:

$$k_z = \frac{F_z}{\left|\Delta_z^4 - \Delta_z^5\right|} \qquad (7)$$

In formula (7), Δ_z^4 and Δ_z^5 are respectively described as the average value of all nodes on section A_4 and section A_5 moving along the z axis.

According to the calculation result of formula (7), the torque M_y rotating around the y axis is applied to the surface A_3 of the mortise joint, and the state is determined after multiple iterative calculations, the moving distance of i node in the upper surface A_1 along the x axis is solved, where Δ_z^4 and Δ_z^5 are expressed as the average value of multiple nodes on FF and A_4 moving along the A_5 axis respectively. According to the simplified expression of formula (7), the value of stress in each direction of mortise and tenon joint can be obtained.

V. Filtering and Enhancement of Ancient Architectural Decoration Image based on Neural Network

A. Filtering and Enhancement of Ancient Architectural Decoration Image

Using the hollow U-Net neural network, after fusing the different levels of features in the ancient building decoration image, the image is transformed into a mode more convenient for automatic analysis and visual observation through the stages of enhancing contrast, smoothing and noise reduction and edge sharpening, so as to realize the filtering enhancement of the ancient building decoration image and improve the visual effect.

1) Contrast enhancement: The pixel gray variable θ of the ancient building decoration image is any integer between 0 and 255. When processing the image, the gray variable is normalized to the interval of 0 to 1. If the gray variable is continuous, the initial gray distribution of the image can be described by using the probability density function $S(\theta)$. The decoration image of ancient buildings is usually discrete, and the discrete gray level can be expressed by Z_x. The corresponding gray level distribution is A_{qw}, and the calculation formula is:

$$A_{qw} = S(\theta) \times Z_x \times R_{as} \times V_{av} \qquad (8)$$

In formula (7): R_{as} represents the total number of image pixels, and V_{av} represents the number of gray pixels. Randomly select a normalized gray variable θ and convert the gray value using the following expression:

$$H_{dz} = M(\theta) \times S_{ab} \times A_{qw} \qquad (9)$$

In formula (8): H_{dz} represents the new gray value after conversion, $M(\theta)$ represents the monotonically increasing conversion function, and S_{ab} represents the probability of occurrence of the converted gray value.

Most ancient architectural decoration images have fixed shapes, so histogram regularization processing needs to be carried out on the basis of equalization processing to further enhance the image contrast [16-17]. For the converted gray value function H_{dz}, the equalization result is T_k. After replacing the inverse gray u with uniform gray s, the calculation formula for contrast enhancement of ancient building decoration image is as follows:

$$D_{bd} = H_{dz} \times T_k \times T(\theta) \times u \times s \qquad (10)$$

If there is $T(\theta) = T^{-1}\left(\frac{1}{\theta}\right)$, the histogram regularization completes the contrast enhancement of ancient architectural decoration image.

2) Smooth noise reduction: In the process of collecting, transmitting and quantifying the ancient building decoration image, the noise of fuzzy image information will be formed. Therefore, the median filter method is used to remove the noise of the ancient building decoration image on the premise that the image edge contour and line can be well preserved.

For the gray value of the pixel to be processed in the image, the intermediate value of the gray value of the pixel in its neighborhood is used to complete the replacement. The implementation form is as follows:

$$Z_{JZ} = med\left\{M(\alpha_1, \beta_1), M(\alpha_2, \beta_2), \ldots, M(\alpha_n, \beta_n)\right\} \qquad (11)$$

In formula (10): Z_{JZ} represents the intermediate value of the neighborhood gray of the pixel to be processed, med represents the median value, $M(\alpha_n, \beta_n)$ represents the pixel whose coordinate is (α_n, β_n), α represents any integer, and β represents the number of rows and columns of the pixel.

3) Edge sharpening: A gradient method is designed to sharpen the pixels with significant changes in image edge and gray level. Based on the field theory, set a vector field as $t = t(\alpha, \beta, \gamma)$, and obtain the vector field gradient according to the extreme value of directional derivative at any point, as shown below:

$$A_{grad}(t) = \frac{\partial O}{\partial \alpha} a + \frac{\partial O}{\partial \beta} b + \frac{\partial O}{\partial \gamma} c \qquad (12)$$

In formula (11): a, b and c represent arbitrary constants.

Given the pixel $O(\alpha, \beta, \gamma)$, the following vector formula is used to define the corresponding gradient of the coordinate (α, β, γ):

$$B_{grad}\left[O(\alpha, \beta, \gamma)\right]^{\mathrm{T}} = \begin{bmatrix} \frac{\partial O}{\partial \alpha} \\ \frac{\partial O}{\partial \beta} \\ \frac{\partial O}{\partial \gamma} \end{bmatrix} \qquad (13)$$

The amplitude of the vector is $G\left[O(\alpha, \beta, \gamma)\right]$, and the following absolute value operation method is used for approximate calculation:

$$\begin{aligned} G\left[O(\alpha, \beta, \gamma)\right] &= \left|O(\alpha, \beta, \gamma) - O(\alpha + 1, \beta, \gamma)\right| \\ &+ \left|O(\alpha, \beta, \gamma) - O(\alpha, \beta + 1, \gamma)\right| \end{aligned} \qquad (14)$$

In order to simplify the complexity of edge sharpening, the obtained gradient $G\left[O(\alpha, \beta, \gamma)\right]$ is used to describe the pixel $O(\alpha, \beta, \gamma)$ and obtain the gradient image.

B. *Filtering and Enhancement of Ancient Architectural Decoration Image*

Before realizing the filtering and enhancement of ancient building decoration image, obtain the ancient building decoration image data through tilt photography technology and preprocess it. Obtain the ancient building decoration image data through three-dimensional laser scanning system and preprocess it [18]; Then, based on the preprocessed tilt image data and point cloud data, ICP algorithm is used to register the tilt image data and point cloud data; Finally, according to the registration results, the filtering and enhancement of ancient architectural decoration image is realized by using CGA rules.

1) Image preprocessing of ancient architectural decoration: The acquisition of ancient building decoration image needs to determine the UAV aerial photographing area, route and time according to the target ancient building and UAV elevation information, set the UAV flight take-off and landing points, and obtain the ancient building decoration image by controlling the UAV. In tilt photography, the vertical ground shooting method is adopted, also known as central projection. In order to ensure the accurate expression of the target ancient architecture, the central projection image is converted into the map projection mode of Orthophoto projection, and the conversion formula is as follows:

$$\begin{cases} x = -j\,\dfrac{a_1\left(X - X_g\right) + b_1\left(Y - Y_g\right) + c_1\left(Z - Z_g\right)}{a_3\left(X - X_g\right) + b_3\left(Y - Y_g\right) + c_3\left(Z - Z_g\right)} \\ y = -j\,\dfrac{a_2\left(X - X_g\right) + b_2\left(Y - Y_g\right) + c_2\left(Z - Z_g\right)}{a_3\left(X - X_g\right) + b_3\left(Y - Y_g\right) + c_3\left(Z - Z_g\right)} \end{cases} \qquad (15)$$

In formula (14), j represents focal length, a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2 and c_3 represent the parameters of external orientation element matrix, (X, Y, Z) represents point cloud data coordinate information, g represents ranging observation value, and (X_g, Y_g, Z_g) represents the coordinate information of a point of the target building in the ground coordinate system.

The obtained target ancient architecture decoration image has multiple perspectives, such as left perspective, right perspective, front perspective, rear perspective and positive perspective. The inclined image of each perspective contains rich geographic information of the target ancient architecture. The definition of the target ancient architectural decoration image plays an important role in image filtering and enhancement. Tilt images from different perspectives will produce large overlapping areas, resulting in some deviation between tilt image data and real data, which will affect the accuracy of filtering and enhancement of ancient architectural decoration images. Therefore, it is necessary to preprocess the target ancient building decoration image data to improve the filtering and enhancement accuracy.

Tilt image preprocessing is mainly divided into two stages, as follows:

Stage 1: Configure relevant information, such as sensor pixel information and parameter information, so as to reduce the displacement deviation during filtering and enhancement of ancient building decoration image;

Stage 2: uniform light and color processing, so as to ensure the filtering of ancient architectural decoration image and enhance the consistency of hue. The hierarchical diagram of point cloud data of ancient building decoration image preprocessing is shown in Fig. 3.

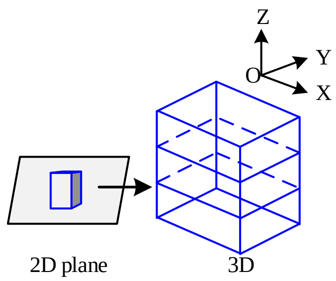
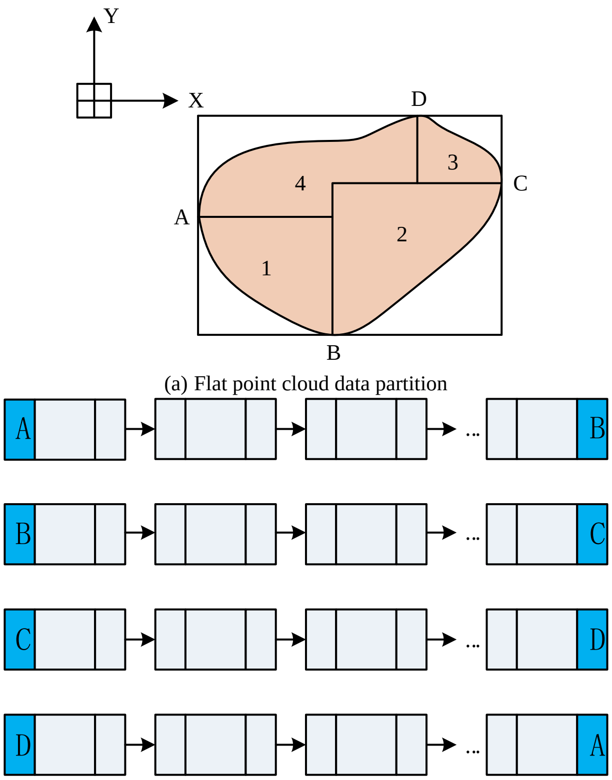


Fig. 3. Hierarchical diagram of point cloud data of ancient building decoration image preprocessing



(a) Flat point cloud data partition

(b) Flat point cloud linked list

Fig. 4. Image preprocessing point cloud data partition storage graph

As shown in Fig. 3, the image preprocessing point cloud data in each sub cube is still stored hierarchically with the Z axis coordinate value from small to large. The maximum and minimum values of coordinates X and Y are used as the point cloud data segmentation points, the plane data is divided into four parts and stored as a point cloud data linked list. The above point cloud data preprocessing process is shown in Fig. 4.

According to Fig. 4, calculate the cube space containing all point cloud data, and the calculation formula is:

$$\begin{cases} L_{abc} = roundup\left(X_{\max} - X_{\min}\right) \\ W_{abc} = roundup\left(Y_{\max} - Y_{\min}\right) \\ H_{abc} = roundup\left(Z_{\max} - Z_{\min}\right) \end{cases} \quad (16)$$

In formula (15), L_{abc} represents the length of the cube and W_{abc} represents the width of the cube; H_{abc} represents the height of the cube; $roundup(\cdot)$ represents the upward rounding function; $\left(X_{\max}, Y_{\max}, Z_{\max}\right)$ represents the maximum coordinate of cube space; $\left(X_{\min}, Y_{\min}, Z_{\min}\right)$ represents the minimum coordinate of cube space.

According to the above calculation results, the point cloud data of ancient buildings are simplified, and the specific steps are as follows:

Step 1: Combine the point cloud data to segment points, and cluster and sort the point cloud data of each layer;

Step 2: Set an appropriate reduction step according to the filtering and enhancement requirements of ancient architectural decoration images;

Step 3: Based on the point cloud data reduction standard, retain the point cloud data that meets the reduction standard, and delete the point cloud data that does not meet the reduction standard;

Step 4: If the point cloud data is in the layer, return to step 1; If the point cloud data is not in the layer, proceed to step 5;

Step 5: If the point cloud data is in the cube, update the point cloud data segmentation points; if the point cloud data is not in the cube, the point cloud data reduction is completed and the reduced point cloud data is output.

2) Filtering and enhancement of ancient architectural decoration image: According to the above preprocessing results of ancient building decoration image, the filtering and enhancement of ancient building decoration image is realized under the neural network by applying CGA rules. The specific process is as follows:

CGA rule is a computer language that can automatically build models and has the characteristics of target description and visual observation. The basic idea of the rule is to replace a shape symbol with a certain number of new shapes. By inserting extruded taper and scatter, rendering texture and so on, the decorative details of ancient buildings can be modeled [19]. Nowadays, CGA rules have been integrated into the CE platform, and the rule programming has been completed in collaboration with C++ language to realize the filtering and enhancement of ancient architectural decoration images with numbering programs. The CGA rule commands are shown in Table I.

Based on the registration results of the ancient building decoration image and the point cloud data, execute the CGA rule command shown in Fig. 1, and finally complete the filtering and enhancement of the ancient building decoration image through the conversion operation. The expression is:

$$V_{fg} = \left(X_g, Y_g, Z_g\right) \times R_s + H_t \times Y_r \quad (17)$$

TABLE I. CGA RULE COMMAND TABLE

Operation Category	Function Name	Command
Addition operation	I(insert)	Insert
	Extrude taper	Pull up
	RoofGable;roof Hip;Roof Pyramid;roof Shed	Roof lifting
	Scatter	Scatter
Subtraction operation	InnerRect	Inner rectangle
	Shape L;shape U;shape O	Deformation
	Convexity	De convexity
	Setback	Back off
	Offset	Deviation
	Comp split	Decompose
Conversion operation	Mirror;mirrorScope；Reverse Normals	Image
	Texture	Mapping
	T(scope translate) translateUV	Move
	AlignScope ToAxes center	Alignment
	R(scope rotate) rotateUV	Rotate
	S(scope size) sizeUV	Zoom
	Set	Set up
	Color	To color

In equation (16), R_s represents the image scaling coefficient, H_t represents the image motion vector, and Y_r represents the rotation coefficient. Through the above conversion operation, the research on the filtering and enhancement method of ancient building decoration image under neural network is realized, which not only provides a means to solve the contradiction between opening and protection of ancient building decoration image, but also provides accurate and complete data support for ancient building management and repair.

VI. EXPERIMENTAL ANALYSIS

In order to verify the performance of the filtering and enhancement method of ancient architectural decoration image based on neural network in practical application, an experiment is designed. The specific experimental process is as follows:

A. Selection of Experimental Objects

A Hutong in Beijing, China is selected as the experimental object of ancient architectural decoration image, which is a quadrangle in the Ming and Qing Dynasties. The South and east of the ancient building are qilinbei Hutong and scissors Lane Hutong respectively. Opposite is Fuxue primary school. It can be seen that the experimental object has a strong flavor of old Beijing life. The experimental object covers an area of about 15000 square meters and is now a key unit of cultural relics protection in Beijing. With the change of history, the owner of the experimental object has changed for several generations, but the shape and architectural layout of the main house have not changed significantly, and the original appearance of ancient buildings in the Ming and Qing Dynasties is basically retained. The schematic diagram of experimental objects of various ancient architectural decoration images is shown in Fig. 5.

Fig. 5. Schematic diagram of experimental object

B. Experimental Instrument Setting

The experiment is carried on Matlab 7.1 platform and runs on the microcomputer platform configured with Genuine Intel(R) CPUT1600 1.66 GHZ and 1GB RAM. The effects of the decorative image filtering and enhancement methods of this method, literature [8] method、literature [9] method and literature [10] method are compared. In the experiment, the UAV camera and three-dimensional laser scanner are mainly used to obtain the information data of the experimental object. In order to ensure the accuracy of the obtained information, the parameters of the experimental instrument are reasonably set, as shown in Table II and Table III. In addition, the superparameter of the neural network as shown in Table IV is set, so that the maximum iteration of the network is 1000 times, and the training will stop when the set accuracy is reached.

TABLE II. CAMERA PARAMETER SETTING

Camera Parameters	Numerical Value	Image Parameters	Numerical Value
Camera model	ILCE-QX1	Image size	5462 * 3632 pixels
Time of exposure	1/1000 秒	Ground resolution	5cm
Aperture value	F/5.6	Horizontal resolution	350dpi
Focal length	16mm	Vertical resolution	350dpi
Maximum Aperture	2.96875	Image overlap	80%-90%
ISO speed	ISO-200	Image bit depth	24

TABLE III. PARAMETER SETTING OF 3D LASER SCANNER

Parameter Name	Numerical Value	Company
Scanning principle	Pulse	-
Field of view range	360×270	Degree
Ranging range	1-350	m
Angle measurement accuracy	±60	mrad
Single point ranging accuracy	±6(50)	mm
Maximum scan rate	5000	Spot/s

There are about 30 courtyards and 60 houses, including rockeries, shrubs, gardens, etc.

TABLE IV. HYPERPARAMETER SETTING OF CONVOLUTIONAL NEURAL NETWORK

Parameter Name	Numerical value
Learning rate	0.001
Batch_size	128
Iterations	1000
Step size	1

C. Result Analysis

The signal-to-noise ratio is used to measure the noise suppression performance of the four methods in the process of image enhancement. Generally, the higher the signal-to-noise ratio, the smaller the noise contained in the ancient building decoration image, and the better the image quality. The calculation process of signal-to-noise ratio is as follows:

$$X_{SNR} = \sum_{i,j}^{n} x'^{2}_{ij} / \sum_{i,j}^{n} y'_{ij} - x'^{2}_{ij} \tag{18}$$

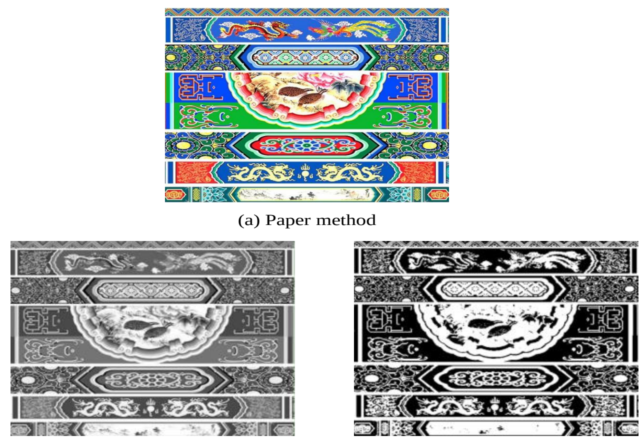
Different Gaussian white noises are introduced into the image, and four methods are used for enhancement respectively. The enhanced images' signal-to-noise ratio is compared. The results are shown in Table V.

TABLE V. COMPARISON RESULTS OF SIGNAL-TO-NOISE RATIO OF THREE METHODS

White Gaussian Noise	Reference [8] Method	Reference [9] Method	Reference [10] Method	Paper Method
1	11.15	11.44	13.12	**14.52**
2	10.20	10.43	12.20	**12.42**
3	9.73	9.91	11.33	**11.89**
4	7.82	8.99	9.96	**10.92**
5	7.57	7.82	9.69	**10.45**
6	7.18	7.54	9.10	**9.99**
7	7.11	7.43	8.84	**9.85**
8	7.07	7.24	8.52	**9.75**
9	6.92	7.04	8.13	**8.99**
10	6.76	6.94	8.21	**8.77**

It can be seen from Table V that under the same Gaussian white noise, the signal-to-noise ratio of the method in this paper is greater than that of the methods in literature [8], [9] and literature [10], which can effectively suppress the adverse impact of noise on the image. This is because this method uses the median filter method to remove the noise of ancient architectural decoration image on the premise of better preserving the image edge contour and lines. For the gray value of the pixel to be processed in the image, the intermediate value of the gray value of the adjacent pixel is used to complete the replacement, which is helpful to suppress the adverse effect of noise on the image to a certain extent.

Fig. 6 is a comparison diagram of ancient building decoration image after filtering and enhancement processing under three methods.



(a) Paper method

(b) Reference [8] method

(c) Reference [9] method

Fig. 6. Image filtering and enhancement processing effect under three methods

It can be seen from Fig. 6 that the images processed by the methods of literature [8] and literature [9] can only get gray images, and the processing of image target details is relatively rough, resulting in blurred image edge feature extraction. However, the filtering enhancement processing effect of ancient building decoration image under this method is very close to the original image, and the color is consistent, Visually, the image obtained by this method is clearer than that obtained by the literature method. This method can show more detailed information in the image, greatly improve the image contrast, enhance the processed image content, and has strong practicability and robustness. Mainly because most of the ancient architectural decoration images in this method have fixed shapes, histogram regularization processing is carried out on the basis of equalization processing to further enhance the image contrast.

VII. CONCLUSION AND PROSPECT

The filtering and enhancement method of ancient architectural decoration image based on neural network can effectively suppress the adverse impact of noise on the image, show more detailed information in the image, greatly improve the image contrast and enhance the processed image content. It has strong practicability and robustness. Although the work has achieved preliminary results, there are still many deficiencies, and a large number of technical problems need to be further studied and solved. Although this work has achieved preliminary results, it only has a good adaptability to the decorative images of Chinese ancient buildings, and cannot be well adapted to other types and styles of decorative images. In the following work, other types of decorative images of ancient buildings should be studied, and the information acquisition problem of some colored images that are not covered and exposed to light and wind and rain for a long time will be studied.

REFERENCES

[1] S. Khokhlova, and A. O. Nagler, “The marfa kurgan in the Stavropol territory: an example of an ancient architectural structure,” Archaeology Ethnology and Anthropology of Eurasia, vol. 48, no. 2, pp. 38-48, 2020.

[2] Campiani, R. L. Stuardo, N. Lercari, “The mausoleum architectural project: reinterpreting Palenque’s temple of the inscriptions through 3d data-driven architectural analysis,” Ancient Mesoamerica, vol. 3, no. 3, pp. 1-16, 2021.

[3] A. Paray, “Composite culture as reflected in ancient art and architecture of Kashmir,” International Journal of Science and Research (IJSR), vol. 10, no. 2, pp. 181-186, 2021.

[4] J. M. Mateus, “Construction history and the history of construction cultures: between architecture and engineering in Portugal,” Buildings, vol. 10, no. 4, pp. 1-10, 2020.

[5] T. Z. Eraliyevna, G. E. Shavkatbekovna, “History, concept and origin of architectural art,” ACADEMICIA an International Multidisciplinary Research Journal, vol. 11, no. 6, pp. 714-716, 2021.

[6] I. Shokry, M. M. Khalifa, S. N. Abdel, “Toward maintaining the identification and architectural heritage to develop the historical areas in ain al-sira,” Resourceedings, vol. 2, no. 2, pp. 237-342, 2020.

[7] G. A. Michel, E. Magda, Z. Saied, “Investigating the impact of employing nanotechnology in heritage building restoration (egytian engineers syndicate),” Journal of Environmental Science, vol. 50, no. 3, pp. 39-66, 2021.

[8] H. Ohuchi, T. Kimura, S. Zong, S. Kanai, T. Kuroiwa, “Quantitative evaluation of architectural style using image correlation and fractal dimension analysis in agora of ancient greek city athens,” Advances in Science and Technology, vol. 103, no. 8, pp. 37-45, 2020.

[9] E. Erysheva, V. Moor, “Evolution of vladivostok’ spatial structure and architectural image. IOP Conference Series Materials Science and Engineering,” vol. 753, no. 3, pp. 022081, 2020.

[10] B.Aoa, A. Ys, A. Hz, BS A，LW A, “Observer-based interval type-2 fuzzy PID controller for PEMFC air feeding system using novel hybrid neural network algorithm-differential evolution optimizer,”Alexandria Engineering Journal,vol.61, no.9, pp.7353-7375, 2022.

[11] Z. Zheng, Y. Liu, M. He, D. Chen, L. Sun, F. Zhu, ”Effective band selection of hyperspectral image by an attention mechanism-based convolutional network，”RSC Adv, vol. 12, no.14, pp.8750-8759, 2022.

[12] G. Chrysos, S. Moschoglou, G. Bouritsas, J. Deng, S. P. Zafeiriou, “Deep polynomial neural networks,” IEEE Transactions on Software Engineering, vol. 2, no. 99, pp. 1-10, 2021.

[13] P. Negretti, G. Bianconi, G. Cannata, G. Catillo, R. Steri, R. Barrasso, G. Bozzo, “Visual image analysis for a new classification method of bovine carcasses according to eu legislation criteria,” Meat Science, vol. 183, no. 1, pp. 108654, 2021.

[14] Y. Ren, X. Li, Xu H,Yang X,”Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images”,IEEE Geoscience and Remote Sensing Letters, vol.19, pp.2-5, 2021.

[15] G. Wu, Y. Zhong, Y. Gong, H.Ren, “Application of Modern Wood Product Glulam in Timber Frame with Tenon-Mortise Joints and Its Structural Behavior”, Journal of Renewable Materials, vol.7, no.5, pp.451-461, 2019.

[16] W. G. Feather, H. Lim, M. Knezevic, “A numerical study into element type and mesh resolution for crystal plasticity finite element modeling of explicit grain structures,” Computational Mechanics, vol. 67, no. 1, pp. 1-23, 2020.

[17] M. Alkovsk, E. Müller, M. Meffert, N. Firman, D. Gerthsen, “Comparison of segmentation algorithms for fib-sem tomography of porous polymers: importance of image contrast for machine learning segmentation,” Materials Characterization, vol. 171, no. 1, pp. 110806, 2020.

[18] S. Gassenmaier, J. Herrmann, D. Nickel, S. Kannengiesser, S. Afat, F. Seith, A. E. Othman, “Image quality improvement of dynamic contrast-enhanced gradient echo magnetic resonance imaging by iterative denoising and edge enhancement,” Investigative Radiology, publish ahead of print, vol. 56, no. 7, pp. 465-470, 2021.

[19] Y. Chen, Z. G.Yang, “Identification Algorithm of Building Sign Information in Natural Scene,” Computer Simulation, vol. 38. no. 8, pp. 450-454, 2021.

A Neural Network-Based Algorithm for Weak Signal Enhancement in Low Illumination Images

Dawei Yin[1*], Jianwei Li[2]
General Affairs Office, Laiwu Vocational and Technical College, Laiwu, China[1]
Department of Information Engineering, Laiwu Vocational and Technical College, Laiwu, China[2]

***Abstract*—There is noise interference in low-illumination images, which makes it difficult to extract weak signals. For this reason, this paper proposes a low-illumination image weak signal enhancement algorithm based on neural network. Multi-scale normalization is performed on low-light images, and multi-scale Retinex is used to enhance weak signals in low-light images. On this basis, the GAN artificial neural network is used to detect the weak signal of the weak signal in the image, the normalization of the weak signal of the low-illumination image is completed based on the residual network, the self-encoding parameters of the depth residual are generated, and the weak signal enhancement result of the low-illumination image is output. The experimental results show that the method in this paper has better enhancement effect on low-illumination images and better image denoising effect. When the scale value is large, the low-contrast area of the low-illumination image has a better enhancement effect. The saturated area of the low-light image has a better enhancement effect.**

Keywords—Artificial neural network; GAN neural network; low-light image; weak signal enhancement

I. Introduction

With the advancement and development of science and technology, the methods of image acquisition are becoming more and more abundant, and people's requirements for image quality are getting higher and higher. However, the process of image acquisition will be affected by many factors. Under special lighting conditions, optical imaging equipment may cause uneven exposure of the obtained image, loss of scene details, and unclear recognition of weak and small targets due to uneven lighting. Since the dynamic range of the shooting equipment is limited, if you just adjust the exposure rate of the equipment, it still cannot solve the problem of overexposure or oversaturation in some areas. Under the conditions of lack of light and low visibility such as at night and dusk, the image collected by the image acquisition device not only has dark areas, but also the brightness and contrast of the image will be seriously reduced, making it difficult to distinguish the details of the image or even unable to see any details. The identification and judgment of images and the extraction of information in project management and control have a certain impact. Therefore, it is the current focus in the field of image processing to study and enhance the detail features of low-light images and ultimately improve the overall quality of the image [1].

Image weak signal enhancement is to perform specific processing on a given image, purposefully emphasizing the overall or local characteristics of the image, making the original unclear image clear or emphasizing some interesting features, and expanding the difference between the characteristics of different objects in the image. To meet some special application requirements [2], the difference between them can be suppressed, and the features that are not of interest can be suppressed. Low-illumination images are generally images collected in scenes with insufficient ambient light. The gray levels of such images are concentrated in a lower gray-scale range, and the details of the images are not obvious. In addition, the image will also contain a lot of noise, which seriously affects the image quality. To make the low-illumination image have a better display effect or meet the input requirements of other image processing algorithms, it is necessary to carry out enhancement and noise reduction processing to emphasize the useful information in the image and effectively suppress the interfering information such as noise [2].

Study [3] proposes a low-light image adaptive enhancement algorithm based on maximum difference map decision. First, the concept of a maximum difference map is proposed, and the initial illumination component is roughly estimated by the maximum difference map; then, an alternate guided filtering method is proposed. The algorithm uses alternating guided filtering to correct the initial illumination components to achieve accurate estimation of the illumination components; finally, an adaptive gamma transform for image brightness is designed, which can adaptively adjust the gamma transform parameters according to the acquired illumination components to enhance the image, also eliminates the effects of uneven lighting. Research [4] proposes a recurrent image enhancement network based on generative adversarial networks (GAN). An unsupervised learning method is introduced to estimate the original illumination of low-light images by reducing the loss of cycle consistency and adversarial loss. The image enhancement model formula is used to enhance the brightness of the images collected under insufficient illumination. Finally, the synthetic low-light image data set and the real natural low-light image data set are evaluated qualitatively and quantitatively. Liu Y's team proposed a GAN model that can perceive the lighting details. This model can fine-tune the image lighting effect in the case of high noise and loss of texture details, and use residual dense encoding and decoding strategy to suppress noise. The performance results show that the model is better in the ultra-low light image data set [5]. Jung E's team proposed a multi-frame GAN concept that can enhance stereo vision image sequences under low light conditions. This method is based on the reversible generation of the confrontation

*Corresponding Author

network, which can transfer the features of bright lighting scenes to the bottom lighting scenes. The research results show that this method has great advantages in visual mileage [6]. Ma Y et al., focusing on the field of medical images, proposed a GAN model that can add terms to the circular structure and illumination constraints of medical images. This model can learn the overall and local details of images. The research results show that this method is better than traditional methods [7]. The above methods are all effective to a certain extent, but it is difficult to ensure the details of the image while realizing the enhancement of the low-illumination image, and the visual perception of the enhanced image still needs to be improved. To this end, this paper proposes a low-light image weak signal enhancement algorithm based on a neural network.

II. Research on Weak Signal Enhancement of Low Illumination Image

A. Multi-Scale Normalization of Low-Light Images

For low-light images, normalization can reduce the probability of image distortion during the enhancement process and improve the ability to retain details. Through the unified management of image features, the misjudgment rate of target information is reduced and the enhancement efficiency is improved. However, in most cases, the original low-light image contains a large number of isolated discrete pixels, and there is no spatial correlation between these pixels, but the pixels follow the Gaussian distribution, indicating that the data of the original image has a certain influence of noise. To this end, the image is denoised first and then normalized [8].

In this paper, the wavelet transform method is used to denoise the data points in the multi-frequency domain in the image, which helps to deepen the pixel signal characteristics in the image, and can ensure the processing effect of signal noise even in the state of different resolutions. The key to wavelet transformation lies in the selection of the transformed threshold. There are two commonly used thresholds: one is the hard threshold, which is closer to the actual situation and retains the edge information of the image better, but the noise reduction effect for Gaussian white noise is poor. The second is the soft threshold, which has a certain continuity, which can improve the low-light enhancement performance and make the visual expression more natural. In contrast, the latter is more suitable for low-light image processing [9].

By establishing a unified Gaussian frequency division noise reduction model, the original low-light image is used as the input value of the model. In the model, the transform value of each wavelet packet frequency band depends on the frequency band value of the image, and the complex wavelet packet is decomposed. It is assumed that X is the initial value of the image, P is the noise, and Z is the observation image with noise. The expression formula is as follows:

$$Z = X + P \quad (1)$$

If each noise in the image follows the Gaussian distribution of its frequency band, its mean value is 0, and its variance value is σ_k^2. The variance value can be calculated according to statistical values [10].

After determining the frequency band parameters of the image, use the multi-scale method to quickly estimate the noise reduction model, and then use the soft threshold to correct the estimated value ξ to obtain the actual output value e^2. The correction formula is:

$$\begin{aligned} e^2 &= |\xi|^2 / 4 - \sigma_k^2 & |\xi| \geq 4\sigma_k^2 \\ e^2 &= 0 & |\xi| < 4\sigma_k^2 \end{aligned} \quad (2)$$

From the corrected output value, derive the noise value in all frequency bands in the image:

$$\xi = e^2 / \left(e^2 + \sigma_k^2\right) \quad (3)$$

Finally, the estimated value is restored to the original decomposition domain by wavelet packet transform to realize image noise reduction.

Calculate the multiple scale values of the denoised image, and divide the size threshold of the image normalization process according to the scale value. The algorithm framework is as follows.

Given a low-light image of size $p \times q$, where q is the sampling factor, M is the scaling factor of the image, and $m = 1, 2, \cdots, M$. Pixel sampling according to $\frac{p}{q} \times \frac{q}{p}$, pixel point $i \in C$.

The constraint function is:

$$C_{s-1,s}\left(i, j\right) = \frac{1}{|N_i|} \qquad \forall j \in N_i \quad (4)$$

In the formula, N_i is the neighborhood of image sampling point i.

Calculate the refraction value W_s^c of the image passing through the light, the radius is r, and the multi-scale value of the image is calculated by $\left(W_s^c, C_{s-1,s}\right)$:

$$W = \begin{bmatrix} W_1^c & & \\ & \ddots & \\ & & W_s^c \end{bmatrix} \quad (5)$$

$$C=\begin{bmatrix} C_{1□2}-I_2 & & \\ & \ddots & \\ & & C_{s-1,s}-I_s \end{bmatrix} \quad (6)$$

In the formula, W represents the multi-scale weight. Calculate the image projection value Q:

$$Q=I-AC^T\left(CA^{-1}C^T\right)^{-1}CA \quad (7)$$

In the formula, A represents a multi-scale normalized matrix:

$$A(i,i)=\sum_{i,j=1}^{n}W(i,j) \quad (8)$$

The low-light image is substituted into the matrix, and the final normalized result can be obtained by multiple iterative calculations.

B. Weak Signal Enhancement of Low-Light Images based on Multi-Scale Retinex

1) Low-light sub-image segmentation based on LIP model: In image processing, general arithmetic operations are not suitable for some actual image processing work, and the result obtained by directly adding (or multiplying) two images has a certain gap with the human visual effect, and will produce “hyper-interval value” problem, the LIP model provides a new arithmetic construct that defines new vector operations such as addition, subtraction, multiplication, etc. The gray values of the images applied with this model are all in the (0, M) interval, so as to avoid the problem of exceeding the interval value, which is also consistent with the saturation characteristics of the human visual system [11].

In order to make the enhancement effect of the proposed method more obvious, when the low-illumination image is denoised, it is decomposed in two dimensions according to the background intensity and gradient information.

Let $I(x,y)$ be the background intensity, which is obtained by calculating the weighted mean of the pixels in the field, as shown in the following formula:

$$I(x,y)=m\otimes n\left(\overset{m}{2}\otimes\sum_{L}f(x,y)\right) \quad (9)$$

In the formula, the set of neighborhood pixels in the four directions of the pixel to be processed is L; the weights are m and n; the set of domain pixels on the diagonal of the pixel to be processed is L'.

If the gradient $G(x,y)$ of the pixel value of the low-light image is used as the information transition rate, the maximum difference value of the low-light image pixel needs to be defined. The calculation method is as follows:

$$Id=\left[\max(f(x,y))\Theta\min(f(x,y))\right] \quad (10)$$

Let I_i be the threshold of the background intensity; G_i be the threshold of the gradient, and use the following formula to divide the low-light image into regions:

$$\begin{cases} I_1=a*Id;I_2=b*Id;I_3=c*Id \\ GG_1=0.01\beta\max(GG(x,y)I(x,y)) \\ GG_2=GG_1I_2 \\ GG_3=GG_1I_3 \end{cases} \quad (11)$$

Pixels in saturated regions of low-light images satisfy the following equation:

$$\begin{cases} I(x,y)\geq I_3 \\ \begin{cases} GG(x,y) \\ I(x,y) \end{cases} \geq GG_3 \end{cases} \quad (12)$$

Pixels in the de Vries region of a low-light image satisfy the following equation:

$$\begin{cases} I_2(x,y)\geq I(x,y)\geq I_1 \\ \begin{cases} GG(x,y) \\ I(x,y) \end{cases} \geq GG_2 \end{cases} \quad (13)$$

Pixels in the Weber region of a low-light image satisfy the following equation:

$$\begin{cases} I_3(x,y)\geq I(x,y)\geq I_2 \\ \begin{cases} GG(x,y) \\ I(x,y) \end{cases} \geq GG_1 \end{cases} \quad (14)$$

The proposed method decomposes the low-illumination image into several sub-images and then merges the remaining pixels into the low-contrast area uniformly, which completes the division of each area of the low-illumination image and realizes the purpose of image enhancement for different sub-images [12].

2) Local multi-scale Retinex algorithm: The Retinex theory believes that an image can be divided into incident components and reflection components. First, the low-illumination image is segmented by the area division method of the above LIP model. Then, according to the illumination characteristics of each sub-image after segmentation, Retinex of different scales σ is used to highlight the advantages of Gaussian functions of different scales are enhanced, and the specific process of the method is shown in Fig. 1.

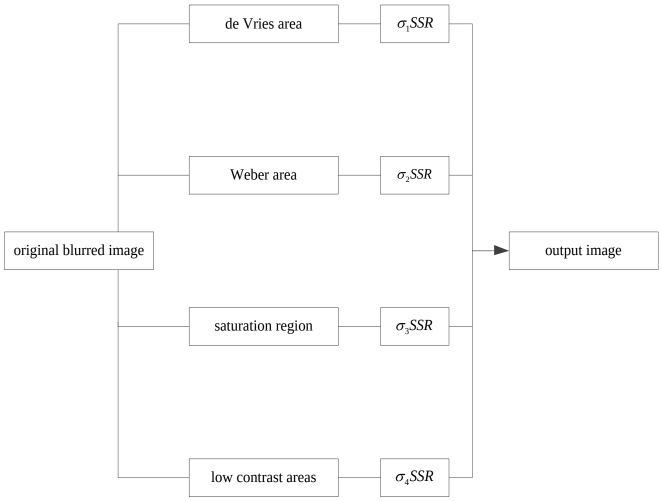


Fig. 1. Local multi-scale Retinex algorithm process

The pixels in the low-contrast area are filtered using Gaussian template $F_1(x,y)$ with scale σ_1.

First, the incident component of the image in this area is estimated, and then the reflection component is discarded to complete the enhancement of the low-contrast area in the low-illumination image, as shown in formulas (15) and (16):

$$F_1(x,y)=2\frac{1}{\pi}\sigma_1 \exp(-\frac{x^2+y^2}{2\sigma_1^2}) \tag{15}$$

$$R_1(x,y)=\lg S_1(x,y)-\lg\left[S_1(x,y)*'F_1(x,y)\right] \tag{16}$$

In the formula, the Gaussian function with a scale of σ_1 is $F_1(x,y)$; the pixel in the low-contrast area is $S_1(x,y)$, and the processing result of the low-contrast area in the original image is $R_1(x,y)$.

Use Gaussian filters of different scales to estimate the incident components of the remaining de Vries area, Weber area, and saturation area, and complete the enhancement of each area according to the above calculation method. The calculation method is as shown in formulas (17) and (18). Show:

$$F_K(x,y)=2\frac{1}{\pi}\sigma_K \exp(-\frac{x^2+y^2}{2\sigma_K^2}) \tag{17}$$

$$R_K(x,y)=\lg S_K(x,y)-\lg\left[S_K(x,y)*F_K(x,y)\right] \tag{18}$$

Through the above calculation, three sub-images $R_2(x,y)$、$R_3(x,y)$、$R_4(x,y)$ are obtained in the same way for the other three regions, and the sub-images $R_1(x,y)$、$R_2(x,y)$、$R_3(x,y)$、$R_4(x,y)$ are combined together to obtain the final enhancement effect.

Because the pixel contrast in the low-contrast area of the low-illumination image is low, choosing a relatively small σ -value can better highlight the details of the image and achieve an enhanced effect. Among them, σ_1、σ_2、σ_3、σ_4 are different scales selected in the four regions decomposed in the low-illumination image by using the Gaussian function.

The de Vries region represents the low-light region of the low-light image, the Weber region represents the medium-light region of the low-light image, and the saturated region represents the high-light region of the low-light image. Therefore, the scale can be selected according to the rule of $\sigma_2<\sigma_3<\sigma_4$. At the same time, through the improved multi-scale Retinex algorithm, the proposed method satisfies the enhancement requirements for different specific pixel regions in low-illumination images, so that low-illumination images can obtain better color fidelity and detail enhancement effects.

III. Improved Algorithm for Weak Signal Enhancement of Low-Light Images based on Neural Network

A. *Weak Signal Detection of Weak Signal in Image based on Artificial Neural Network*

Before designing the low-light image weak signal enhancement algorithm, extract the weak signal of the image, apply the mathematical model and other techniques to plan the image edge data into a set of multi-dimensional data structure [13], and then use the random process of the data and the feedback in the artificial neural network. The mechanism divides the weak signal of the image into functional regions and non-functional regions [14], and its noise function expression is as follows:

$$g(y)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{(y-\mu)}{2}\right) \tag{19}$$

In the formula, y represents the gray value of the noisy edge image, and μ represents the expected value of the image edge noise on the gray value. The formation of Rayleigh noise is mainly caused by other elements in the edge data collection environment during the image generation process, and the noise is relatively high. The complex data edge environment, and then the formation of noise, the function representation is as follows:

$$g(y)=\frac{2}{b}(y-a)\exp\left[-\frac{(y-a)^2}{b}\right], y\geq a \tag{20}$$

In the formula, a represents the high probability complex factor existing in the weak signal data environment of the image, and b represents the correlation between the weak signal data of the image and the original data of the image edge. In order to observe the influence of different noises on the image edge more intuitively, this paper extracts the noise data of the original image edge, and then the Matlab software is used for model simulation processing. The obtained weak signal detection processing results of different types of images are as follows:

The application of artificial neural network technology to extract image weak signal edge noise data requires the application of various types of algorithms and filtering techniques [15-17]. As the feedback object and data extraction library of artificial neural network, the filtering algorithm formula used in this paper is as follows

$$f(x,y)=\frac{1}{M\times N}\sum_{M,N\in R}f(m,n) \tag{21}$$

In the formula, M and N respectively represent the average noise wave value of different types in the image weak signal edge data, which can produce a certain degree of defense against the image edge noise data, and complete the image edge noise without losing the original image clarity. Data detail processing and extraction will not affect the post-processing of images such as missing data.

After filtering and classifying the relevant data in the weak signal of the image by applying filtering technology, two parts of noise coefficients are generated, which respectively exist in the image edge database with different noise frequencies. Initialize the data information at the edge of weak signal. The specific change expression is as follows:

$$w_{j,k}^{n}=\begin{cases} w_{j,k} & \left|w_{j,k}\right|\geq\lambda \\ 0 & \left|w_{j,k}\right|<\lambda \end{cases} \tag{22}$$

In the formula, λ represents the image edge noise threshold, and $w_{j,k}$ represents the noise coefficient in the neural network. To accurately extract the noise coefficient, the data connection state in the artificial neural network has been in a continuous mode, resulting in obvious threshold division and easy to make other noise factors. Reshape the details of the weak signal edge of the image.

B. Create a Normalized Processing Structure for Weak Signals in Low-Light Images

GAN technology is actually a low-light image weak signal confrontation processing program, which mainly processes the weak signals of low-light images one by one according to the corresponding steps. First, we need to prepare a low-light image weak signal discriminator and a generator. The discriminator mainly makes a corresponding judgment on whether the weak signal of the low-illumination image belongs to the real data distribution, and calculates the respective probability values of the data input and output. The generator captures relevant data and indirectly forms a similar data distribution. Calculate the objective function of the GAN. Calculated as follows:

$$G\left\{F(a,b)\right\}=E\left[Lg(1+\sqrt{5})\right] \tag{23}$$

In the formula, G represents the value of the objective function, F represents the set range of the function, a represents the maximum value of the function, b represents the minimum value of the function, and E represents the corresponding separation function of the GAN objective function. Through calculation, its objective function is obtained. Then, a model is generated using the objective function and balanced optimization is performed on it. In a specific low-light image weak signal processing environment, use SRGAN to build a GAN network architecture processing structure, as shown in Fig. 2:

Fig. 2 shows the related process of the low-light image weak signal processing structure of the created GAN network. After that, a normalized processing relationship is established between the weak signal loss function of the low-illumination image obtained in the model and the adversarial function, as shown in the following formula:

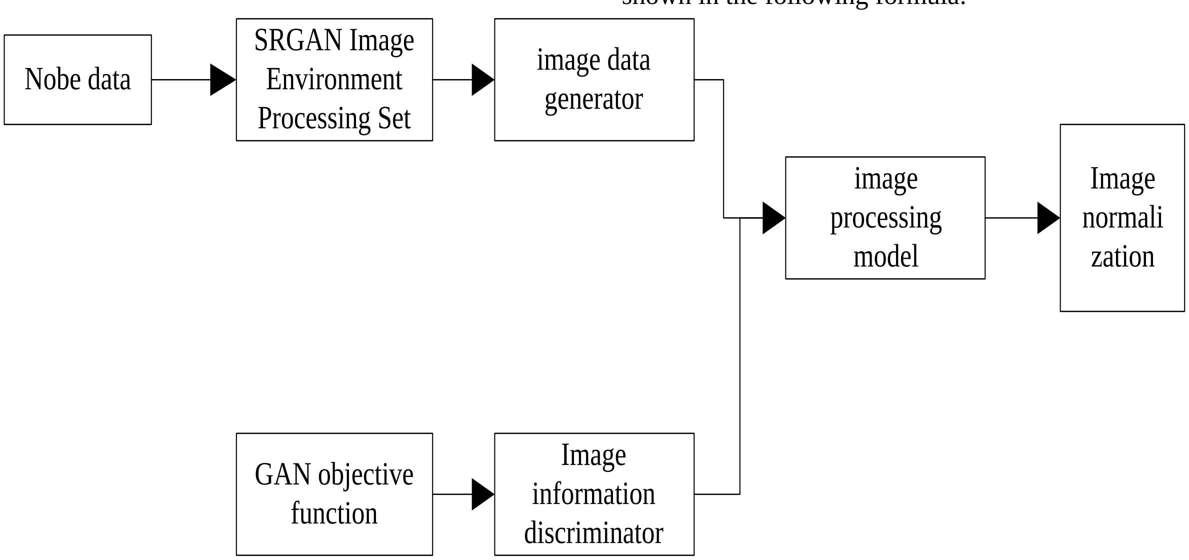


Fig. 2. GAN network low-light image weak signal processing architecture

$$Y = B - C + r \quad (24)$$

In the formula, Y represents the relationship between the loss function and the adversarial function, B represents the directional processing distance, C represents the target processing distance, and r represents the error distance. After the processing relationship is established, the weak signal of the low-illumination image is imported into the processing system, and the generator is used to first obtain the data information corresponding to the weak signal of the low-illumination image, and then a convolution layer of the weak signal of the low-illumination image is established to improve its resolution. Subsequently, the convolutional layer is activated using the ReLU activation layer in the discriminator, and then 10 residual network modules are constructed through the processing steps, and these modules are skip-connected. At this time, the connection surfaces of each layer are superimposed on each other and continue to be superimposed under the action of the convolution layer. Such a low-light image weak signal processing mode can prevent the low-level low-light image weak signal features from being generated. loss, and expand the size, length, width, and area of weak signals in low-light images. In general, it can be expanded to about four times the weak signal of the original low-light image, and finally, a picture with four times the resolution is obtained. Then, the discriminator is used again to superimpose the weak signal of the low-illumination image processed above by using the processing path again to superimpose the connection surface. This time, multiple convolution structures need to be repeated, and their lengths are separated into two types, one and two, using different processing types used to process low-light images, weak signals, and pictures in different formats. Subsequently, the image scale was changed to 1/16 in the original image processing system. Adjust the number of channels to 612, activate the discrimination result, output the result in another system dimension, and set the sigmoid function to perform the final processing on the weak signal of the low-light image, to use the related technology of GAN to create a low-light image. Normalized processing structure for weak signals.

C. Low-Light Image Weak Signal Enhancement Algorithm based on Residual Network

After the normalization of the weak signal of the low-illumination image is completed, next, a low-illumination residual network is constructed to enhance the weak signal of the low-illumination image. First, the auto-encoding parameters of the depth residual are generated. Auto coding is a neural network control value that uses a backpropagation algorithm to process data information. It compresses the weak signal of a low-light image to an editable size, hides its spatial representation, and then outputs it through the representation. Next, the low-light processing coefficient of the residual processing model is established in a low-light processing environment. The calculation formula is as follows:

$$X = (\alpha + \beta) - \frac{1}{l} \quad (25)$$

In the formula, X represents the low-light processing coefficient of the model, represents the processing distance, represents the straight-line processing distance of the model, and one represents the number of inertial protons. Through the above calculation, the low-light processing coefficient is obtained. On this basis, next, the total iteration value of the discrimination of the weak signal of the low-light image is calculated. The calculation formula is as follows:

$$K = \frac{\Delta i}{\Delta j} - \frac{1}{v} \quad (26)$$

In the formula, K represents the total iterative value of discrimination for weak signal processing of low-illumination images, i represents the total distance of calculation iterations, j represents the number of iterations, and v represents the objective function. After obtaining the iterative value of the weak signal processing of the low-illumination image, next, the processing system is used to establish an enhancement relationship between the weak signal of the original low-illumination image and the weak signal of the target low-illumination image to enhance its versatility.

The cyclic residual neural network is used to enhance the weak signal of the low-light image. First, optical color analysis needs to be performed on the low-light image that needs to be processed, and the color that cannot be edited is recorded. Then, use the residual neural network technology to amplify the weak signal of the low-light image to a size that can be processed, separate the color replacement factors in the low-light image, and adjust the color of the low-light image to a single mode, and turn on the intelligent color in the system processing. Adjustment function, a kind of original low-light image weak signal, restores the color and increases the control parameters of its saturation so that the low-light image weak signal overall is more three-dimensional. Then, set the residual coefficients that satisfy the weak signal enhancement processing of low-illumination images in the system, add auxiliary graphics processing programs, and change the processing mode of the system to multiple-simultaneous loop processing. Then, in the low-light and low-light image processing environment, the corresponding low-light image weak signals and photos are enhanced at various levels, to complete the low-light image weak signal enhancement processing by the cyclic residual neural network.

IV. Experimental Analysis

A. Experimental Conditions

To verify the effectiveness of the low-light image weak signal enhancement algorithm based on a neural network, a simulation comparison experiment was carried out. Select the images in the Joint Video Stitching and Stabilization from Moving Camer (http://www.liushuaicheng.org/TIP/VideoStitching2016/index.html) dataset, and the schematic diagram of the low-light image is shown in Fig. 3.

Fig. 3. Schematic diagram of low-light image

TABLE I. GAN NEURAL NETWORK PARAMETERS

Index	Parameter
The number of network layers	3
The number of iterations	94
The weight of the first layer	0.5
The offset term	0.7
The learning rate	0.001

The parameters of the GAN neural network are set as shown in Table I.

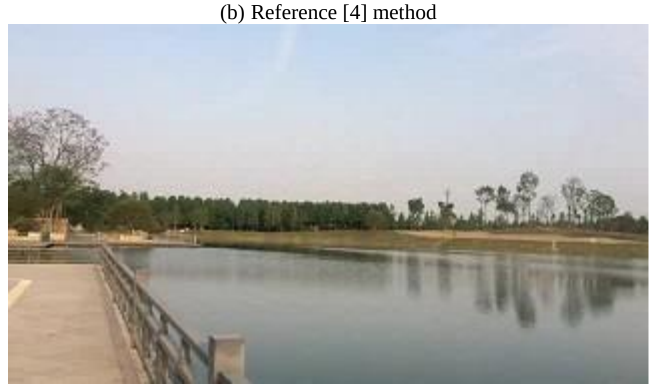
Under the above experimental conditions, the method of studies [3] and [4] are used as experimental comparison methods to test the effectiveness of the low-light image weak signal enhancement algorithm.

B. Analysis of Experimental Results

Fig. 4 is a low-illumination image enhancement effect diagram obtained by applying the method of study [3], the method of research [4] and the method of this paper to the low-illumination image enhancement operation.

(a) Reference [3] method

(b) Reference [4] method



(c) This paper method

Fig. 4. Low-light image enhancement effect diagram

As can be seen from the above figure, the original low-illumination image is dark in brightness and the blurring phenomenon is serious. The enhancement effect of the method in studies [3] and [4] is poor, while the low-illumination image enhanced by the method in this paper has low brightness and sharpness. It shows that the application of this method can realize the enhancement of low-light images, and the enhancement effect is better, which can better meet the needs of practical work.

Clarity and information entropy are two important indicators that reflect the enhancement effect of weak signals in low-light images. Table II shows the values of sharpness and information entropy at different scales obtained by performing sub-image enhancement operations on saturated areas of weak signals and low-contrast areas of low-illumination images using the method in this paper.

TABLE III. CLARITY AND INFORMATION ENTROPY AT DIFFERENT SCALES

Scale value	low contrast area		saturation area	
	sharpness	information entropy	sharpness	information entropy
0.6	20.52	6.21	23.63	8.63
1.2	21.52	6.59	22.52	8.52
1.8	22.03	7.03	22.02	8.01
2.4	22.96	7.42	21.96	7.56
3.0	23.65	7.86	21.56	6.98
3.6	24.52	7.99	20.32	6.23
4.2	25.63	8.09	19.56	5.56
4.8	26.66	8.33	18.96	4.98
5.4	27.33	8.86	18.23	4.88
6.0	27.86	9.01	17.63	4.35

It can be seen from Table II that with the increase of the scale value, the sharpness and information entropy value of the low-contrast area of the low-illumination image both show an increasing trend. When the value is 0.6, it is increased by 34%; the sharpness and information entropy of the saturated area of the low-light image shows a decreasing trend with the increase of the scale. When the scale value increases from 0.6 to 6.0, the sharpness of the image decreases by 25%. Description: When the value of the scale is large, the low-contrast area of the low-light image has a better enhancement effect, and when the value of the scale is small, the saturated area of the low-light image has a better enhancement effect. Therefore, when applying the method in this paper to perform image enhancement on low-illumination images, large-scale and small-scale combined operations should be performed in the low-contrast and saturated regions of the low-illumination image to achieve better image enhancement effects.

Fig. 5 shows the sharpness curves under different signal-to-noise ratios obtained by applying the method in this paper, the method in study [3] and [4] to perform weak signal multi-scale enhancement operations on low-illumination images respectively. In Fig. 5, with the increase of the signal-to-noise ratio, the sharpness curves of the three methods all show an upward trend, but the sharpness curves of the method in this paper are higher than those of the methods in researches [3] and [4] from the beginning, and with the increase of the signal-to-noise ratio, the sharpness curve obtained by applying the method in this paper to perform the image enhancement operation on the low-illumination image has an obvious upward trend. Increased by 50% and 65%. Prove: Applying the method in this paper to perform image enhancement operations on low-light images can obtain higher image clarity, and the advantages of low-light image enhancement are more obvious.

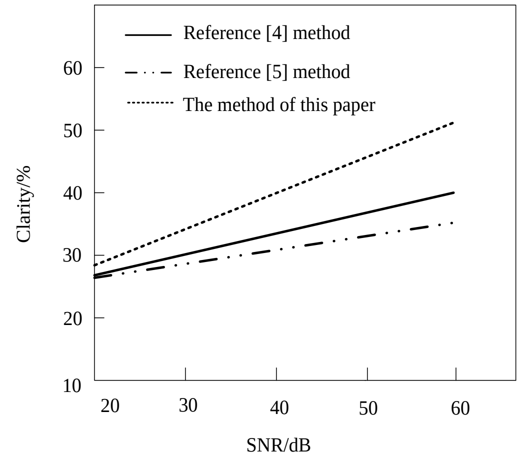


Fig. 5. Comparison of sharpness curves

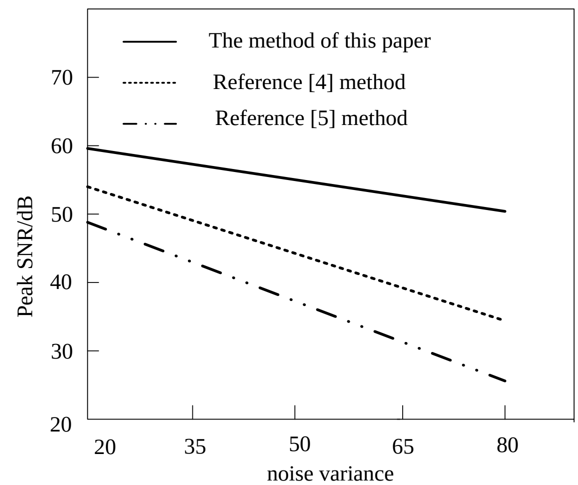


Fig. 6. Peak signal-to-noise ratio curves of different methods

In order to verify the advantages of the method in this paper in denoising low-illumination images, the peak signal-to-noise ratio curves under different noise variances obtained by applying three methods to perform denoising operations on low-illumination images are shown in Fig. 6.

Analysis of Fig. 6 shows that the peak signal-to-noise ratio curves obtained by applying the three methods to perform denoising operations on low-illumination images all gradually decrease with the increase of noise variance, but the peak signal-to-noise ratio curves obtained by applying the method decrease from one to one. In the beginning, it is higher than the methods of [3] and [4], and with the increase of noise variance, the downward trend of the curve is very insignificant. Experiments show that the method in this paper is used to perform denoising operation on low-illumination images, and the denoising effect is better than the methods in [3] and [4], and it has more advantages in denoising low-illumination images.

V. Conclusion

This paper proposes a low-light image weak signal enhancement algorithm based on a neural network. The experimental results show that the method in this paper has a better enhancement effect on low-light images. When the scale value is large, the low-contrast area of the low-light image has a better enhancement effect. When the scale value is small, the saturation area of the low-light image has a better effect. It has a better enhancement effect, and the image-denoising effect of this method is better.

References

[1] Z. Zhang, "Similarity Retrieval Simulation of Low Illumination Image Features Based on Pentatree Decomposition," Computer Simulation, vol. 38, no.6, pp. 460-463,474, 2021.

[2] J. Wang, Y. Liu, D. Zou, Z. Ren, S. Yan, "Controlling the chain orientation and crystal form of poly(9,9-dioctylfluorene) films for low-threshold light-pumped lasers. Macromolecules, vol. 54, no. 9, pp. 4342-4350, 2021.

[3] R. Wang, Y. Xueting, Z. Zhou, Z. Geng, "Adaptive enhancement algorithm of low illumination image based on maximum difference image decision," Journal of Computer Applications, vol. 40, no. 4, pp. 1164-1170, 2020.

[4] L. Huang, Y. Shaozhen, "Research on low-light image enhancement algorithm based on GAN," Journal of Fuzhou University (Natural Science Edition), vol. 48, no. 5, pp. 551-557, 2020.

[5] Y. Liu, Z. Wang, Y. Zeng, et al. "PD-GAN: Perceptual-Details GAN for extremely noisy low light image enhancement," ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, PP. 1840-1844, 2021.

[6] E. Jung, N. Yang, D. Cremers, "Multi-frame GAN: Image enhancement for stereo visual odometry in low light," Conference on Robot Learning. PMLR, pp. 651-660, 2020.

[7] Y. Ma, Y. Liu, J. Cheng, et al. "Cycle structure and illumination constrained GAN for medical image enhancement," International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp. 667-677, 2020.

[8] S. Tang, F. Yu, "Construction and verification of retinal vessel segmentation algorithm for color fundus image under bp neural network model," The Journal of Supercomputing, vol. 77, no.1, pp. 1-15, 2021.

[9] W. Zhang, F. Wang, N. Li, "Prediction model of carbon-containing pellet reduction metallization ratio using neural network and genetic algorithm," ISIJ International, vol.61, no.6, pp. 1915-1926, 2021.

[10] K. Srinivas, A. K. Bhandari, "Low light image enhancement with adaptive sigmoid transfer function," IET Image Processing, vol. 14, no. 4, pp. 668-678, 2020.

[11] Paul, T. Sutradhar, P. Bhattacharya, S. P. Maity, "Adaptive clip-limit-based bi-histogram equalization algorithm for infrared image enhancement," Applied Optics, vol. 59, no. 28, pp. 9032-9041, 2020.

[12] P. Li, K. T. Lo, "Survey on jpeg compatible joint image compression and encryption algorithms," IET Signal Processing, 2020, vol. 14, no. 8, pp. 475-488.

[13] I. Bondarenko, E. I. Suvorova, "Method of contrast enhancement and background correction in electron diffraction patterns of polycrystalline materials," Crystallography Reports, vol. 66, no. 4, pp.594-600, 2021.

[14] M. Ghafoor, S. A. Tariq, I. A. Taj, N. M. Jafri, T. Zia, "Robust palmprint identification using efficient enhancement and two-stage matching technique," IET Image Processing, vol. 14, no. 11, pp. 2333-2342, 2020.

[15] H. Danesh, K. Maghooli, A. Dehghani, R. Kafieh, "Synthetic oct data in challenging conditions: three-dimensional oct and presence of abnormalities," Medical & Biological Engineering & Computing, vol. 60, no. 1, pp.189-203, 2021.

[16] J. A. Fessler, "Optimization methods for magnetic resonance image reconstruction: key models and optimization algorithms," IEEE Signal Processing Magazine, vol.37, no.1, pp.33-40, 2020.

[17] L. I. Yang, Y. Kubota, M. Okamoto, S. Shiba, T. Ohno, "Determination of deformable image registration algorithms for accumulating dose in carbon-ion radiotherapy for pancreatic cancer," Anticancer Research, vol.41, no.2, pp.835-843, 2021.

Digital Intelligent Management Platform for High-Rise Building Construction Based on BIM Technology

Rui Deng[1], Chun'e Li[*2]
School of Building Management, Chongqing Metropolitan College of Science and Technology, Chongqing, China[1, 2]

***Abstract*—In this study, the digital intelligent management platform of high-rise building construction based on BIM technology is used for real-time monitoring and management of construction progress and quality. In the data acquisition and processing layer, construction site data is obtained through RFID technology. After processing such as cleaning and integration, it is input to the BIM model layer to dynamically generate various real-time BIM models, and these real-time BIM model information is input to the application layer to query, monitor and correct the construction progress and quality. The results are presented by the display layer. The actual application results show that the real-time BIM models generated by the platform have clear details and can realize the query, monitoring and correction functions for the construction progress of high-rise buildings, and effectively correct the construction progress according to the construction progress monitoring query results to achieve the unification with the planned progress. It can effectively realize the visual measurement of the size of each component in construction and monitor the construction quality in real time.**

Keywords*—BIM technology; high-rise building construction; digitization; intelligent management; BIM model; RFID technology

I. Introduction

With more and more complex high-rise construction and longer construction cycle in recent years, it makes the control of construction progress and quality more and more difficult. Therefore, the collaborative management of construction progress and quality of high-rise buildings has become a very important research topic in the field of engineering construction [1]. Many studies have been done by scholars at home and abroad for this purpose. At present, most of the studies are exploring the working principles of management and platform for comprehensive monitoring of progress and quality implementation in the construction phase. However, such management platforms lack specific responses in case of changes in project construction and are not effective in ensuring that high-rise buildings proceed as planned [2, 3]. Therefore, there is a need to create a platform that can adapt to this change.

J. Zhang et al. (2021) study the project critical chain schedule monitoring method based on project resource segmentation, and determine the early warning state value by multi-objective fuzzy comprehensive evaluation method, solve the overdue minimization objective function, and finally use BPNN to predict the next phase of the project so as to deal with the problems in the project implementation process in a timely manner. In order to realize the monitoring and management of project progress, this monitoring method of construction project progress can effectively predict the project risk situation and has good effect on reducing the overdue completion rate and overcapacity rate, but it does not implement monitoring and management of construction quality of the construction project [4]; Zhou, C et al. (2020) proposed that the digital construction management platform in EPC mode is mainly from EPC mode starting from the EPC model to solve the key problems in the digital construction of large enterprises, so as to analyze the necessity of applying the EPC model in digital construction. Although this platform can realize the basic digital management in construction, it lacks the function of real-time correction and adjustment according to the plan in the actual construction process [5]; Ma, G et al. (2019) proposed an engineering construction schedule optimization model based on improved genetic algorithm, constructed the overlapping strategy decision model of engineering schedule, combined with the optimization algorithm to output the overlapping strategy with precise overlapping rate, and obtained the overlapping strategy in meeting the premise of schedule requirements, the optimal schedule plan is obtained. This model can achieve the optimal management of project construction schedule, but does not have the ability to control the project construction quality [6]. In recent decades, information simulation technology has been widely used in the construction industry, especially the Building Information Modeling (BIM) technology. This technology is considered as an essential tool for construction project management. All project participants can share and exchange information with each other with the help of BIM technology platform, which is also an upgrade of CAD technology. In a real sense, it realizes industrialization, refinement, digitization, informatization, and intelligence in the construction industry [7, 8]. It not only reflects the geometric information of building entities, but also shows a large amount of building information such as structure type, material type, manufacturer, construction quality, etc. during the whole life cycle of a construction project. It not only provides a unified data and information source for the collaborative management of all project participants, but also promotes the reform process of building construction progress and quality management [9]. Obtaining 3D models through BIM technology can enhance the interactivity of manual models [10]. However, due to the large volume and high investment in the construction of high-rise

*Corresponding Author

buildings, the wide range of specialties and long period, the impact on the surrounding environment and complex construction organization and other challenges. Therefore, in the construction process of high-rise buildings, we need to adopt BIM technology to solve the problems of construction progress and quality management in real time. For the pre-processing of the data, RFID technology is required. This technology, also known as radio frequency identification, is a communication technology [11]. This technology allows the identification of specific targets and the reading and writing of relevant data through wireless communication without the need to establish mechanical or optical contact between the identification system and the specific target.

Based on the above analysis, this paper innovatively proposes a digital intelligent management platform for high-rise building construction based on BIM technology. The digitalization of construction progress and quality in the construction process of complex high-rise buildings is visually and intelligently monitored and managed, which improves the comprehensive management level in the construction process of high-rise buildings and ensures the construction progress and quality of high-rise buildings. Firstly, RFID technology is used for data collection, and then the data is input into the BIM model proposed in this study.

II. Digital Intelligent Management Platform for High-Rise Building Construction

A. *Overall Architecture Design of the Platform*

Architecture design of the digital intelligent management platform for high-rise building construction based on BIM technology. The key objective of the platform is to realize the visual digital intelligent management of the construction progress and quality of high-rise buildings, as well as to ensure the construction quality and progress. The overall architecture design of the platform is shown in Fig. 1.

The intelligent management platform is mainly composed of data acquisition and processing layer, BIM model layer, application layer and display layer. The data of high-rise building's construction site is collected through the data acquisition and processing layer. After processing the collected data, the BIM model layer uses the processed high-rise building's construction site data to build terrain BIM model, high-rise building's BIM model, high-rise building's construction information BIM model, construction machinery and equipment's BIM model and the 4D BIM model of high-rise building construction. Various BIM models are introduced into the application layer to prepare the progress plan and progress report of high-rise building construction, monitor and correct the construction quality and progress in time, simulate and optimize the construction process, and present various application results in the application layer to the construction manager in real time through the display layer, so as to complete the digital intelligent management of high-rise building construction progress and quality.

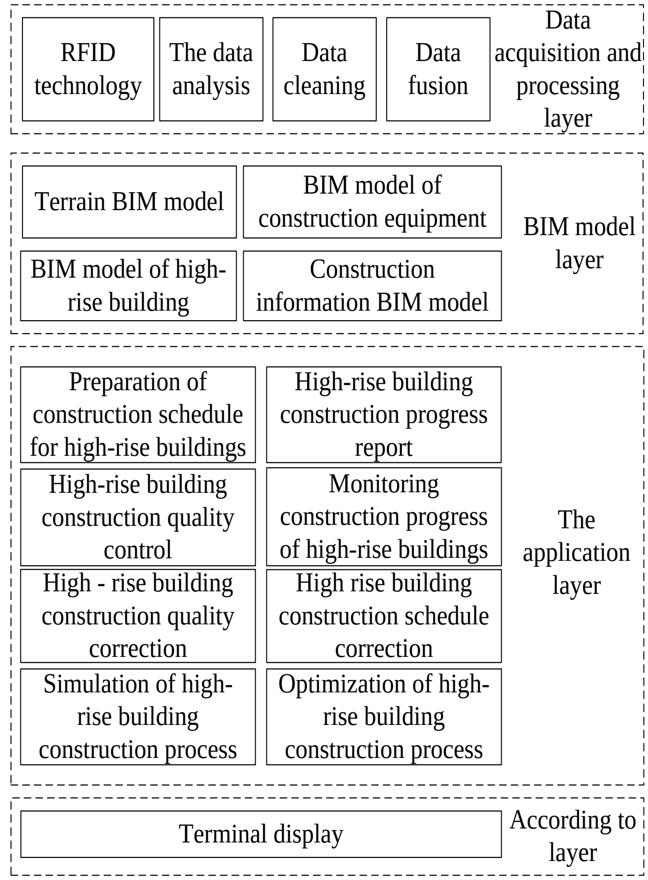


Fig. 1. Overall architecture diagram of digital intelligent management platform for high-rise building construction based on BIM technology

B. *Design of Data Acquisition and Processing Layer in Digital Intelligent Management Platform for High-Rise Building Construction*

In the digital intelligent management platform for high-rise building construction designed in this paper, the data acquisition and processing layer mainly realizes the collection and processing of building and mechanical equipment data on the high-rise building construction site. Therefore, this layer can be divided into two parts: data acquisition and data processing. The key contents of the two parts are as follows:

1) Data collection part: mainly using RFID technology to collect construction site data of high-rise buildings in real time. It provides strong data support for the realization of digital intelligent management of the building. In terms of applicability, RFID technology has technical advantages in the field of building management. Its superior non-contact reading and writing function can be used to quickly identify information such as high-rise building construction [12]. In this study, RFID technology is used to quickly identify and collect information from each instrument and equipment in high-rise buildings and construction, and to complete the storage and transmission of the collected information. Its structure is shown in Fig. 2.

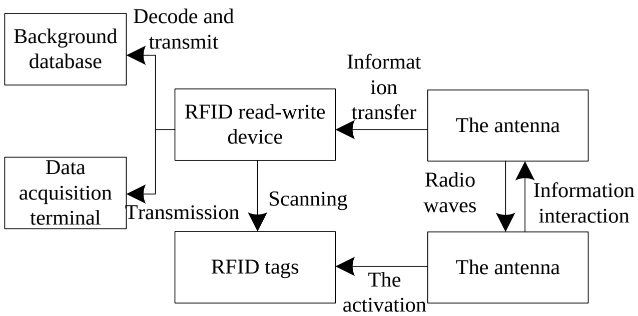


Fig. 2. Structure diagram of data acquisition part of RFID technology

The realization process of RFID technology in the data acquisition part is as follows: i) identification and information acquisition of buildings and mechanical equipment: the RFID technology in the data acquisition part is mainly composed of RFID tag, RFID reader and antenna. In the on-site construction process, RFID tag is generally placed in mechanical equipment, and RFID reader and writer are placed in each floor of buildings; when the RFID reader scans the RFID tag, the antenna of the RFID reader transmits radio waves, and the antenna of the RFID tag receives the transmitted radio waves, activates the RFID tag, and collects the existing building and mechanical equipment and other information, such as building and mechanical equipment ID; for mechanical equipment with motion range on the construction site, RFID can not only collect its construction information, but also realize the accurate positioning of mechanical equipment according to the collected information. ii) Information storage and transmission of buildings and mechanical equipment: after the RFID reader obtains the actual construction information of buildings and mechanical equipment on the construction site in real time, it can write the information into the RFID tag, decode it and transmit it to the RFID background database for storage by GPS or wireless network, or directly to the data acquisition terminal.

2) Data processing part: the data processing part is the intermediate filter layer for effective information transmission from the data acquisition part to the BIM model layer. The amount of information collected by RFID technology in the data acquisition part is huge and miscellaneous. It must be further processed by data analysis, data cleaning and data integration, so as to provide efficient data information flow for the management of high-rise building construction. Since a large amount of complex data will be generated during the construction of high-rise buildings, NoSQL, Sybase, Oracle and other databases can be selected for data storage of the information of buildings and mechanical equipment on the construction site obtained by RFID technology in the data acquisition part [13]; At the same time, due to the non-standard data form, data conflict, data duplication and other situations in the obtained information, in order to effectively manage the construction progress and quality of high-rise buildings in the digital intelligent management platform, it is necessary to analyze and clean such complex and diverse information data. Through the statistical analysis and network analysis of such information data, the classification and standardized processing of such information can be realized [14]; The cleaned and re selected data cannot be directly applied to the intelligent management platform. It is necessary to sort out such data forms through data integration, so that the BIM model layer in the intelligent management platform can directly extract and use these data, and use these data to obtain various corresponding BIM models, so as to lay a foundation for the application and implementation of the management platform.

C. *Design of BIM Model Layer in Digital Intelligent Management Platform for High-Rise Building Construction*

1) Structure Design of BIM Model Layer

Based on its own characteristics, if the traditional construction management mode is adopted, the management efficiency is not only low, but also easy to cause the loss of information and other problems. In this paper, the BIM model layer is designed by using BIM integration technology in the digital intelligent management platform of high-rise building construction, which can effectively solve the above problems. According to the digital intelligent management requirements of high-rise building construction, the BIM model layer structure of the digital intelligent management platform for high-rise building construction is designed. This layer uses different software to build three-dimensional terrain's BIM model, high-rise building's BIM model, construction machinery and equipment's BIM model, high-rise building's construction information BIM model and the 4D BIM model of high-rise building construction. Through the application layer of the digital intelligent management platform of high-rise building construction, the integration and application of this BIM model are realized, so as to realize the digital visual intelligent management of high-rise building's construction progress and quality. The structural design of BIM model layer is shown in Fig. 3.

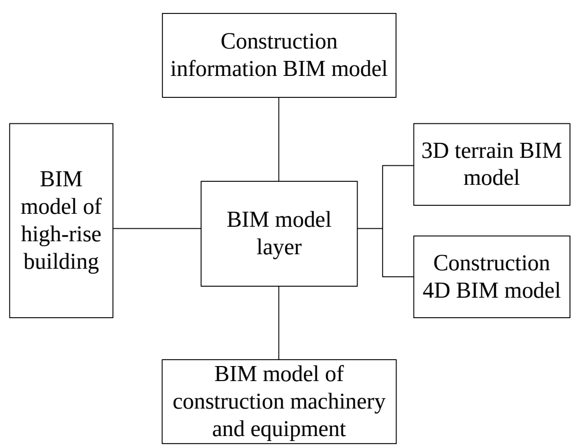


Fig. 3. Structure design drawing of BIM model layer

Various BIM models in the BIM model layer are described as follows:

a) Three-dimensional terrain's BIM model: this model is a three-dimensional real scene model, which can well reflect the terrain and geomorphic characteristics of the high-rise building's construction area. It is intuitive, vivid and realistic. The model is composed of terrain mapping data and real image. After collecting the terrain image of the area where the high-rise building construction is located, the three-dimensional digital model of surrounding buildings, terrain and geomorphic vector data and other information, it is imported into BIM software for processing, to obtain the three-dimensional BIM model of terrain and landform.

b) BIM model of high-rise building: in order to maximize the coincidence between the construction simulation and the actual construction process, taking the construction process of high-rise building as the goal and based on the construction scheme plan, the construction project of high-rise building is divided into different construction areas for modeling. The model is the basis of realizing digital intelligent management of high-rise building construction. It integrates physical and functional characteristics, including the spatial location, shape structure and material properties of high-rise building's construction projects. BIM model construction of high-rise building mainly includes two parts: terrain surface solid model and high-rise building's partition solid model. Whether it is terrain surface solid model or high-rise building's partition solid model conforming to partition, the multi-level of detail accuracy of components generated by BIM component assembler directly determines the accuracy of partition solid model of high-rise building construction [15-17]. BIM software editor can be used to improve the overall accuracy of generated components. The specific process is shown in Fig. 4.

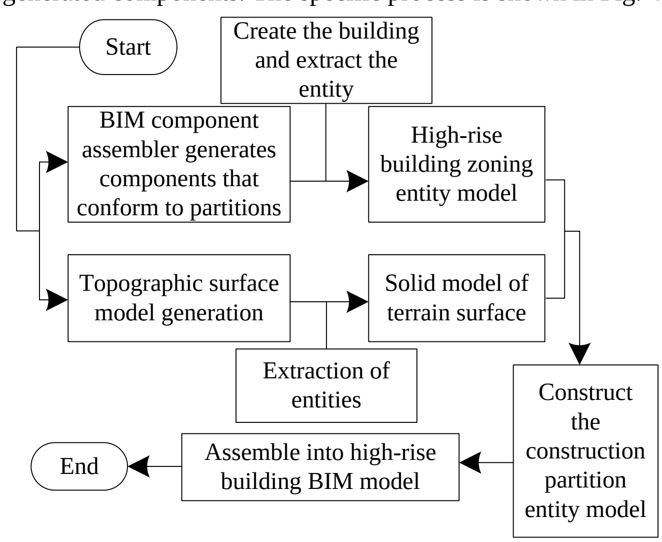


Fig. 4. BIM model assembly flow chart of high-rise building

c) BIM model of high-rise building's construction information: the BIM model of high-rise building's construction information based on BIM technology mainly includes construction progress and construction quality information, which has better visibility. This model can be created with the help of Microsoft Project (MSP), and the bar chart of on-site construction progress can be prepared according to the created model.

d) BIM model of construction machinery and equipment: the information data of various construction machinery and equipment on the high-rise building's construction site obtained and processed by the data acquisition and processing layer in the platform is used to import into the BIM software to generate the real-time BIM model of construction machinery and equipment in the operation of the construction site.

e) 4D BIM model of high-rise building construction: the 4D BIM model of high-rise building construction is the integration of BIM model and high-rise building's BIM model of high-rise building's construction information.

2) Design of Dynamic Modeling Process in BIM Model Layer

The accuracy of traditional manual modeling methods cannot be guaranteed, and it is time-consuming and laborious. It costs a lot in practical application. Due to many uncontrollable factors in the intermediate process of automatically creating BIM model, it is easy to miss or deviate from the model. Therefore, in the digital intelligent management platform of high-rise building construction, the dynamic modeling process of BIM model layer combines time series and planning model to realize the dynamic generation of various BIM models. In the design stage before the construction of high-rise building project, the pre generated BIM model corresponding to the construction plan of high-rise building project [18] is the planning model. According to this planning model, various BIM models can be dynamically created, which can make full use of the planning model to complete the verification of the real-time creation model, ensure the accuracy and integrity of high-rise building construction, and reduce the complexity of automatic modeling.

The dynamic modeling method used in the BIM model layer in this paper is to obtain the spatial position and orientation of each model component in the actual high-rise building's construction project by using the processed on-site construction data on the basis of the data acquisition and processing layer, and change the planning model to obtain various actual BIM models according to this kind of actual information. This way of creating the actual BIM model based on the planning model can not only ensure the accuracy of the BIM model, but also reduce the complexity and workload. For the convenience of description, U_a is defined as the actual BIM model generated by the BIM model layer, and U_b is defined as the plan model. Then the relationship between the two is:

$$U_a = \delta\left(U_b\right) \quad (1)$$

Where, δ represents the model change function, which aims to transform the planned model components according to the information of the actual high-rise building's construction site, such as rotation, clipping, spatial position movement and so on. The planning model U_b can be described by a series of elements C_b and the relationship H_b between elements, that is:

$$U_b = \left\langle C_b, H_b \right\rangle \quad (2)$$

Therefore, only a certain transformation of C_b and H_b is needed to generate U_a. The transformation equation used here is:

$$\max{}^l \left[fit\left(B, U_b^l\right) \right] \rightarrow \tilde{U}_b \sim U_a \quad (3)$$

Where, B represents the characteristics of the obtained construction site data; U_b^l represents the BIM planning model modified by l iterations; $fit\left(B, U_b^l\right)$ represents a function for calculating the similarity between B and U_b^l; $\tilde{U}_b$ represents an approximate solution of U_a and the optimal solution of fit function. This dynamic modeling method of real-time generating BIM model based on planning model not only has high modeling efficiency, but also has high precision, and can reconstruct the covered part. The dynamic modeling process of BIM model layer combining time series and planning model is shown in Fig. 5.

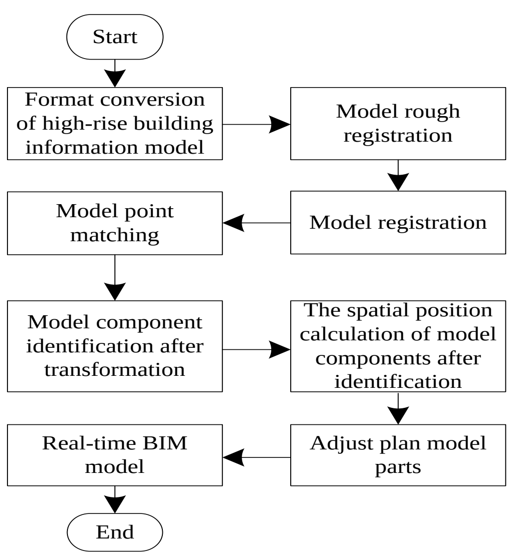


Fig. 5. Dynamic modeling process diagram of BIM model layer combined with time series and planning model

The detailed steps of dynamic modeling process of BIM model layer combining time series and planning model are as follows:

1) Format conversion of high-rise building's information model: in order to obtain the detailed information in the BIM model, it is necessary to convert the irregular geometry in the BIM model into triangles, that is, convert the parametric BIM model into STL data format [19]. Relatively complex planes or surfaces can be represented in detail by triangles.

2) Rough model registration: rough registration of the converted model is carried out based on the spherical coordinate system. The converted model is an entity to be pseudo scanned in the coordinate system. In the matching process, it is necessary to select the datum points of high-rise building's construction engineering facilities to obtain more than three point pairs, and then rough registration is carried out.

3) Fine model registration: for each scanning point in the data generated after laser scanning of the solid parts of the converted model, the model points of the corresponding parts can be found in the planning model after mapping, to record the included angle in the horizontal direction and the inclination angle in the vertical direction when the laser scanner scans the solid parts of the converted model. These two angles can determine the direction of the ray for the pseudo scanning of the converted model; Then, by calculating the intersection and distance between the triangular surface of a part in the planning model and the previously determined ray, the obtained intersection is the model point matching the scanning point; The vertices and corresponding intersections of STL triangles in the converted model need to be represented by spherical coordinate system, so the horizontal included angle and vertical inclination of the boundary of each triangle element can be identified.

4) Model point matching: in the construction of high-rise buildings, each model point of each component can be matched with the corresponding scanning point. The previous coarse registration and fine registration ensure that each model point has the same angle with the actual scanning point. Therefore, it only needs to compare the distance ΔD between the model point and the scanning point to determine the relationship between this distance and the preset threshold ΔD_{max}. If $\Delta D \leq \Delta D_{max}$, it is considered that the two-point pairing is successful. Therefore, the definition of D2 directly determines whether the automatic matching can be successful. ΔD_{max} is defined as:

$$\Delta D_{\max} = \varepsilon_1 + \overline{\varepsilon}_2 \quad (4)$$

Where, ε_1 represents the acceptable construction error in the actual high-rise building construction; $\overline{\varepsilon}_2$ represents the average value of registration error.

5) Recognition of model components after conversion: for the point matching method, as long as a certain number of point pairs are judged to match, it can be considered that the components are recognized, but its stability depends on the number of point pairs sought [20]. Here, a method with higher stability is combined to judge whether the object is recognized more effectively. The process of the method is as follows: i) define parameter A_e as the surface area of the model, that is, the sum of the areas formed by the equal division of the distance between adjacent discrimination points. ii) Define the model's surface area threshold S_e as:

$$S_e = n \tan\alpha \tan\beta \, \mathrm{dist}^2_{\max} \quad (5)$$

Where, n represents the minimum number of model recognition point pairs; $\tan\alpha$ and $\tan\beta$ represent the tangent functions of the horizontal included angle and vertical inclination of the model components respectively; $\mathrm{dist}^2_{\max}$ represents the highest distance function of the distance between adjacent discrimination points. The accuracy of component identification is guaranteed by setting an appropriate threshold S_e. When the threshold S_e is less than or equal to A_e, the component can be considered to be successfully identified.

6) Spatial position calculation of model components after identification: after the components of the converted model are successfully identified, the position and orientation of such components need to be calculated. After the third step of fine registration, the converted model will be paired with the real-time high-rise building's state data points, so that the spatial position information such as the position and orientation of the components that have been successfully identified can be calculated. According to the spatial position information of such actual components, various real-time BIM models in high-rise building construction can be obtained by adjusting the parts of the plan model, such as rotation and clipping.

D. Construction Progress and Quality Monitoring Process of Application Layer in Digital Intelligent Management Platform for High-Rise Building Construction

In this paper, the application layer of the digital intelligent management platform for high-rise building construction mainly realizes the functions of construction progress planning, report generation, monitoring, correction, construction quality monitoring, and the optimization and simulation of the whole construction process, etc. Among them, construction progress monitoring and construction quality monitoring are particularly keys. Therefore, the design is implemented for the monitoring process of high-rise building's construction progress and quality. Fig. 6 shows the details.

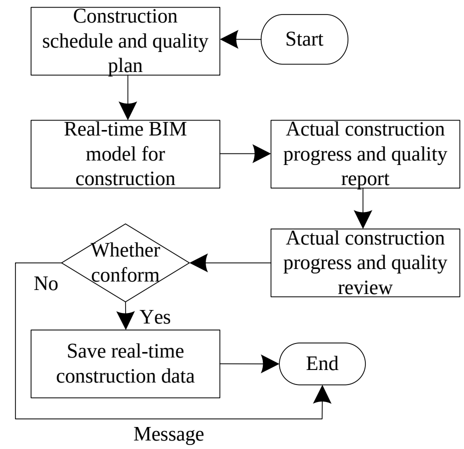


Fig. 6. Monitoring process diagram of construction progress and quality

Construction progress and quality monitoring is based on the construction plan. The BIM model layer is used to import various BIM models in the actual construction of high-rise buildings in the application layer, so as to visually and intelligently supervise whether the actual construction progress and quality of high-rise building projects under construction are consistent with the construction plan, so as to achieve the purpose of digital intelligent management of high-rise building's construction progress and quality. The management of high-rise building's construction progress and quality realizes the input of time nodes and construction tasks of high-rise building's project construction. In the management process, the generated construction progress chart and construction quality model chart can be reviewed to observe whether they meet the construction requirements, and the construction progress and quality can be checked and recorded in real time.

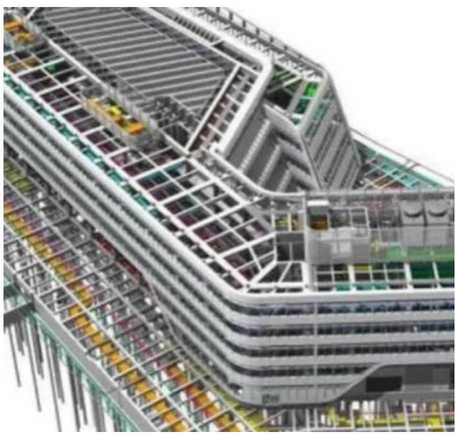
In the digital intelligent management platform of high-rise building construction, the construction progress and quality report is generated according to different time periods. Through the comparative analysis of the current construction progress and quality and the planned construction progress and quality, the report is formed and output through the display layer to realize the supervision and reminder of high-rise building's construction progress and quality; At the same time, it can query the completion of high-rise building's construction tasks according to the time, content, cycle and other keywords of construction tasks, and adjust the construction progress and quality of high-rise buildings in real time to avoid the waste of time, manpower and economic cost. So far, the design of digital intelligent management platform for high-rise building construction based on BIM technology has been completed.

III. Analysis of Experimental Results

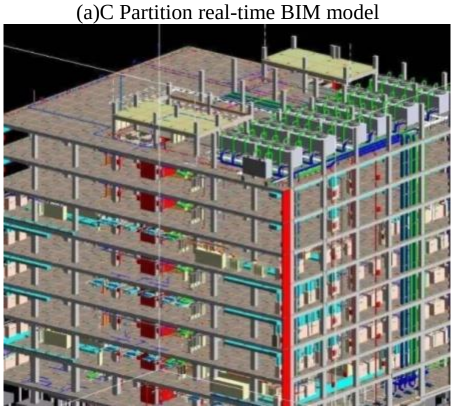
Taking a high-rise building project under construction in an area as an example, this platform is applied to the construction management of the high-rise building project to test the practical application effect of this platform. The high-rise building project is divided into 10 partitions, each partition is numbered a ~ j. Digital intelligent management is implemented for the construction progress and construction quality of each partition through the platform of this paper. The specific inspection process and results are analyzed as follows:

A. Construction Progress Management Process and Result Analysis

Firstly, two partitions c and f are randomly selected from the 10 partitions of the experimental high-rise building project, and the real-time BIM models of the two selected partitions are obtained through the platform in this paper. The presentation results are shown in Fig. 7.



(a)C Partition real-time BIM model



(b)F Partition real-time BIM model

Fig. 7. Real-time BIM model of each partition of experimental high-rise building presented by this platform

It can be seen from Fig. 7 that the real-time BIM model of each partition of the experimental high-rise building presented by the platform in this paper can show the internal information and layered information of the building in detail, and the visualization effect is clear, which is convenient for users to view and monitor the construction progress in real time, so as to realize the digital intelligent management of high-rise building construction.

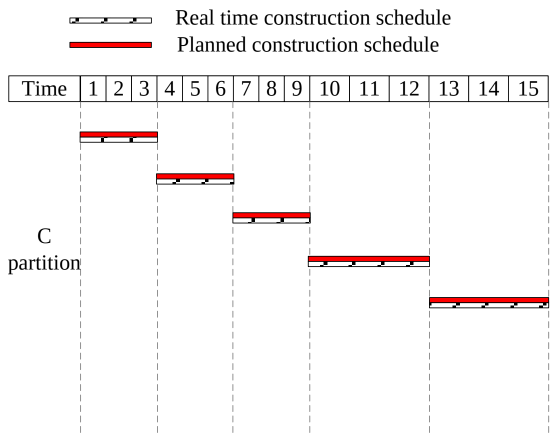
The real-time construction progress of partitions c and f of the experimental high-rise building is obtained through the platform in this paper. The deviation between the obtained real-time construction progress and the planned construction progress is compared and analyzed to realize the real-time monitoring of the construction progress. The bar chart of deviation comparison and analysis is shown in Fig. 8.

According to the analysis of Fig. 8, the real-time construction progress of each partition of the experimental high-rise building can be obtained through the platform. After comparing the implementation deviation between the real-time construction progress of each partition obtained by the platform and the planned construction progress, it is found that the real-time construction progress of partition c of the experimental high-rise building is basically consistent with the planned construction progress, and there is almost no deviation; The real-time construction progress of partition f is different from the planned construction progress, and the deviation between the two is obvious. The actual construction progress of the partition is one day later than the planned construction progress. Therefore, the platform in this paper can monitor the actual construction progress of experimental high-rise buildings in real time, so as to provide guarantee for correcting the construction progress in time and ensuring the completion of the overall construction on schedule.

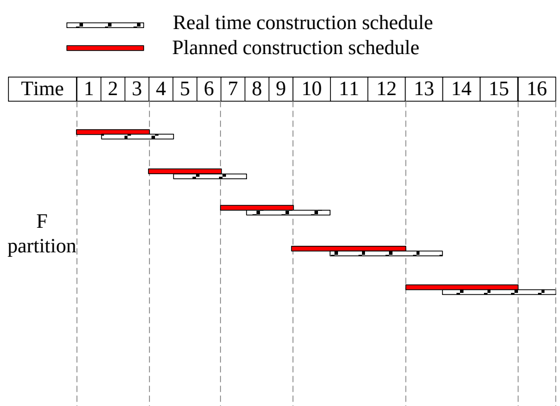
Through the platform in this paper, the real-time construction progress of all partitions of the experimental high-rise building is obtained according to the above methods, and compared with the implementation deviation of the planned construction progress of each partition respectively, the obtained data are stored in the background database of the platform in this paper, and the number of each partition is used to query the planned and actual construction progress data of each partition to complete the construction progress query task. The query results are shown in Table I.

TABLE I. REAL-TIME QUERY RESULTS OF CONSTRUCTION PROGRESS OF THE PLATFORM IN THIS PAPER

Partition number	Plan time	The actual time	Schedule analysis	Delays
a	March 10, 2021	March 15, 2021	The required progress has been reached, but the progress is delayed	Delay 5 days
b	March 13, 2021	March 12, 2021	Arrive ahead of schedule	Without delay
c	March 8, 2021	March 8, 2021	According to the original schedule to reach the specified progress	Without delay
d	March 15, 2021	March 18, 2021	The required progress has been reached, but the progress is delayed	Delay 3 days
e	March 20, 2021	March 24, 2021	The required progress has been reached, but the progress is delayed	Delay 4 days
f	March 6, 2021	March 7, 2021	The required progress has been reached, but the progress is delayed	Delay 1 day
g	March 5, 2021	March 5, 2021	According to the original schedule to reach the specified progress	Without delay
h	March 17, 2021	March 18, 2021	The required progress has been reached, but the progress is delayed	Delay 1 day
i	March 1, 2021	March 3, 2021	The required progress has been reached, but the progress is delayed	Delay 2 days
j	March 25, 2021	March 31, 2021	The required progress has been reached, but the progress is delayed	Delay 6 days



(a)Comparison between real-time construction progress and planned construction progress in c partition



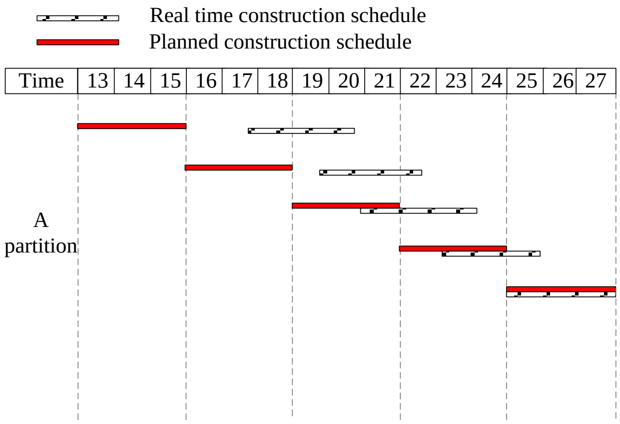
(b)Comparison between real-time construction progress and planned construction progress in f partition

Fig. 8. Horizontal chart of deviation between real-time construction progress and planned construction progress in each partition

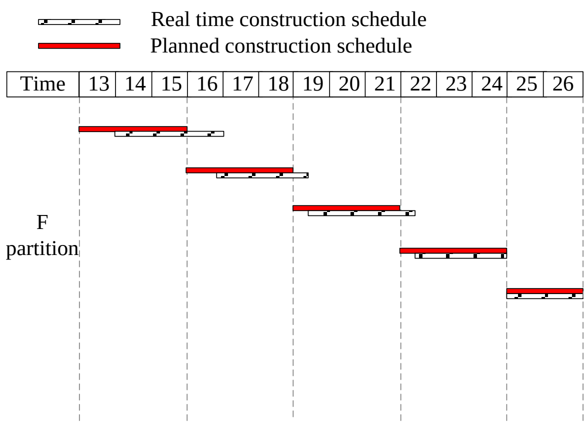
It can be seen from Table I that the platform in this paper can query the actual construction progress of each partition of high-rise building in real time. From the query results, it is known that the actual construction progress of three partitions b, c and g of the experimental high-rise building project is not delayed compared with the planned construction progress, and the construction progress of the other seven partitions is delayed to varying degrees.

Based on the above query results, under the management of the platform in this paper, the construction progress correction is implemented for the seven partitions with construction progress delay. Taking three partitions a, f and j as examples, the bar chart of the real-time construction progress and planned construction progress of each partition after the correction of this platform management is presented to test the correction effect of construction progress management of this platform. The construction progress bar chart of each partition is shown in Fig. 9.

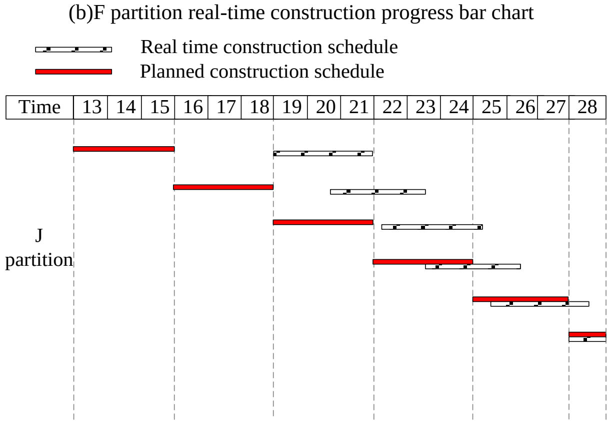
As can be seen from Fig. 9, after the construction progress correction is implemented under the management of the platform in this paper, the actual construction progress and planned construction progress of each partition with construction progress delay in the experimental high-rise building project can be gradually unified. The construction progress management effect of the platform in this paper is remarkable, which can provide effective guarantee for the construction period of high-rise buildings. The data processing efficiency and progress error of this method, digital twin (DT) technology and EPC mode digital building management platform in construction progress are compared and analyzed. The results are shown in Table II.



(a)A partition real-time construction progress bar chart



(b)F partition real-time construction progress bar chart



(c)J partition real-time construction progress bar chart

Fig. 9. Horizontal chart of real-time construction progress and planned construction progress in each partition after correction of platform management in this paper

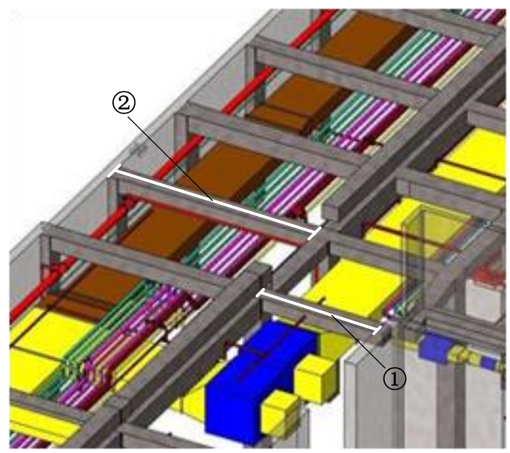
TABLE II. COMPARATIVE ANALYSIS TABLE

Comparison Items	BIM	DT	EPC
Data processing efficiency	91.89%	78.93%	79.97%
Progress advancement error	0.156	0.342	0.459

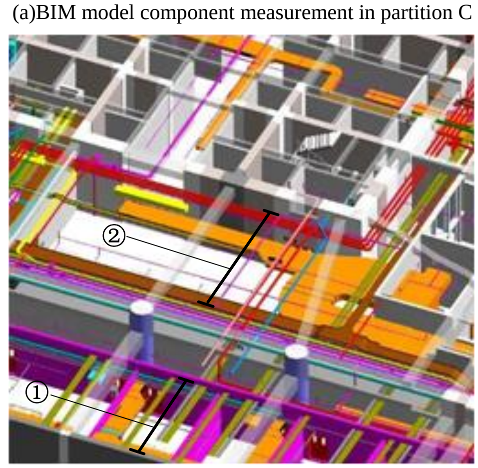
It can be seen from Table II that the data processing efficiency of BIM model is 91.89%, and the progress error is 0.156, which is better than the processing progress of digital twin (DT) technology and digital building management platform under EPC mode. The construction progress management of BIM technology platform has a remarkable effect, which can provide an effective guarantee for the construction period of high-rise buildings.

B. *Construction Quality Management Process and Result Analysis*

Based on the platform in this paper, the parts of the real-time BIM model of the two partitions c and f of the experimental high-rise building are obtained. Using these parts, we can measure whether the size of the internal components in the construction of each partition of the experimental high-rise building meets the construction requirements in real time, so as to achieve the purpose of real-time visual monitoring of the high-rise buildings' construction quality. Taking the random components in the real-time BIM model of partitions c and f obtained by the platform in this paper as an example, the real-time measurement is shown in Fig. 10.



(a)BIM model component measurement in partition C



(b)BIM model component measurement in partition F

Fig. 10. Real-time measurement of BIM model components in each partition of experimental high-rise building obtained by the platform in this paper

As can be seen from Fig. 10, the internal details for the BIM model components of each partition of the experimental high-rise building obtained by the platform in this paper are clear, which can be used to realize the real-time visual measurement of the dimensions of each component in the component, so as to achieve the purpose of real-time monitoring the quality of each component in the construction of high-rise building.

IV. Discussion

This paper mainly analyzes the practical application effect of the platform in this paper from two aspects: the first aspect is the management effect of this platform on the construction progress of high-rise buildings; The second aspect is the management effect of the platform on the construction quality of high-rise buildings. Among them, the management effect of this platform on the construction progress of high-rise buildings is mainly tested through the actual construction progress monitoring, query and correction. The test results show that this platform can realize the functions of real-time progress monitoring, query and correction in the actual construction of high-rise buildings, and has a significant effect on the comprehensive management of construction progress; The management effect of this platform on the construction quality of high-rise buildings is only tested by the quality monitoring in the actual construction. The test results show that this platform can realize the real-time monitoring function of each component quality in the actual construction of high-rise buildings, but it is not verified for other functions of construction quality. This part needs to be further carried out in the future research, in order to improve the verification of the comprehensive management effect of high-rise building construction quality.

V. Conclusion

With the development of science and technology, infrastructure construction is developing rapidly, and also showing the development trend of scale and complexity. In order to ensure that building construction can be put into use properly, it is necessary to ensure that the project can be completed on time as planned. Therefore, this paper proposes a digital intelligent management platform for high-rise building construction based on BIM technology to achieve comprehensive management of progress and quality during high-rise building construction by integrating the processes of data collection and processing, dynamic generation of BIM models, and construction progress and quality monitoring. The application effect of the platform was tested through practical engineering applications. The results show that the data processing efficiency of the BIM model proposed in the study is 91.89%, and the progress error is significantly lower than that of the digital construction management platform model under digital twin (DT) technology and EPC mode, and the application effect in construction progress monitoring, query, correction and construction quality monitoring is significant, which can realize the digital, visualization and intelligent management of high-rise building construction. In the future, we will continue to study the energy consumption and cost management in high-rise construction and further expand the practical application of this platform. This research can track the construction progress and monitor the construction quality in real time, which has significant effect compared with other technologies.

Reference

[1] M. Kopsida & I. Brilakis, “Real-Time Volume-to-Plane Comparison for Mixed Reality–Based Progress Monitoring,” Journal of Computing in Civil Engineering, vol.34, pp. 04020016, 2020.

[2] Z. Wang, Q. Zhang, B. Yang, T. Wu, K. Lei, B. Zhang, “Vision-Based Framework for Automatic Progress Monitoring of Precast Walls by Using Surveillance Videos during the Construction Phase,” Journal of Computing in Civil Engineering, vol.35, 2020.

[3] Y.C. Lee, M. Scarpiniti & A. Uncini, “Advanced Sound Classifiers and Performance Analyses for Accurate Audio-Based Construction Project Monitoring,” Journal of Computing in Civil Engineering, vol.34, pp.04020030, 2020.

[4] J. Zhang, & K. Li, “Critical Chain Schedule Monitoring Based on Project Resource Segmentation.” Management Review, vol. 33, pp. 224-231, 2021.

[5] C. Zhou, X. Zhang, F. Peng, Z. Guo, Y. Zhang & S. Xue, “Study on Application of Digital Construction Management Platform Under EPC Mode,” Construction Technology, vol.49, pp. 135-141, 2020.

[6] G. Ma, & L. Zhang, “Construction Project Schedule Optimization Based on Improved Genetic Algorithm,” Journal of Civil Engineering and Management, vol. 36, pp. 1-6, 2019.

[7] K. Asadi, H. Ramshankar, M. Noghabaei & K. Han, “Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction,” Journal of Computing in Civil Engineering, vol. 3, pp.04019031.1-04019031.15, 2019.

[8] X. Ma, & D. Liu, “3D Reconstruction Simulation of BIM Building Based on Perspective Augmented Reality,” Computer Simulation, vol.37, pp. 229-233, 2020.

[9] M. Xiao, & X. Zhou, “Safety Management Method of Coastal Buildings against Typhoon Based on BIM Technology,” Journal of Coastal Research, vol.103, pp. (sp1):539, 2020.

[10] X. Hua, “Application of BIM Technology in Port Breakwater Construction,” Journal of Coastal Research. vol. 103, pp.(sp1):548, 2020.

[11] W. Zhu, X. Meng, X. Peng, J. Cao & M. Raynal, “Collisions are Preferred: RFID-based Stocktaking with a High Missing Rate,” IEEE Transactions on Mobile Computing, pp. 1, 2019.

[12] X. Liu, J. Yin, J. Liu, S. Zhang & B. Xiao, “Time Efficient Tag Searching in Large-scale RFID Systems: A Compact Exclusive Validation Method,” IEEE Transactions on Mobile Computing, pp.99, 2020.

[13] A.V. Yamaev, M.V. Chukalina, D.P. Nikolaev, A.V. Sheshkus & A.I. Chulichkov, “Neural Network for Data Preprocessing in Computed Tomography,” Automation and Remote Control, vol. 82, no. 10, pp.1752-1762, 2021.

[14] H. Zhang, M. Huang, J. Yang & W. Sun, “A Data Preprocessing Method for Automatic Modulation Classification Based on CNN,” IEEE Communications Letters, pp.99, 2020.

[15] H. Lai, X. Deng & T. Chang, “BIM-Based Platform for Collaborative Building Design and Project Management,” Journal of Computing in Civil Engineering, vol. 33, no.3, pp.05019001.1-05019001.15, 2019.

[16] L. Yang & W. Hu, “Life Cycle Cost Management of Marine Engineering Project: A BIM Technology Based Study,” Journal of Coastal Research, vol. 107, pp. (sp1):85, 2020.

[17] M. Valinejadshoubi, A. Bagchi & O. Moselhi, “Development of a BIM-Based Data Management System for Structural Health Monitoring with Application to Modular Buildings: Case Study,” Journal of Computing in Civil Engineering, vol.33, no.3, pp. 05019003.1-05019003.16, 2019.

[18] Y. Weng, N. Mohamed, B. Lee, N. Gan, M. Li & M. Tan, “Extracting BIM Information for Lattice Toolpath Planning in Digital Concrete Printing with Developed Dynamo Script: A Case Study,” Journal of Computing in Civil Engineering, vol. 35, no.3, 05021001, 2021.

[19] J. Ninic, H.G. Bui, C. Koch & G. Meschke, "Computationally Efficient Simulation in Urban Mechanized Tunneling Based on Multilevel BIM Models," Journal of Computing in Civil Engineering, vol. 33, no. 3, pp. 04019007.1-04019007.17, 2019.

[20] M. Li, "Application of Building Information Model (BIM) in the Design of Marine Architectural Structures," Journal of Coastal Research, vol. 94, pp. (sp1):117, 2019.

Research on Key Technologies of Smart City Building Interior Decoration Construction based on In-Depth Learning

Li Zhang[1], Aimin Qin[2*]
Information Technology and Creative Design College, Qingyuan Polytechnic, Qingyuan, China[1]
College of Architecture and Civil Engineering, West Anhui University, Lu'an, China[2]

***Abstract*—The intelligentization of building interior decoration construction is of great significance to the construction of smart city, and robot automation has brought an opportunity for this. Robot self-decoration is the development trend in the future. One of the key issues involved, is the self-planning of mobile path. In this regard, the research adopts the proximal policy optimization algorithms (PPO) to improve the self-planning path ability of the decoration robot. For the information of lidar and robot status, the Full Connect Neural Network (FCNN) is used to process it. In addition, the reward function and the corresponding Credit Assignment Problem (CAP) model are designed, to accelerate the learning process of path planning. Aiming at the dynamic uncertainty in the actual environment, the adaptive loss function is used to build an auxiliary model to predict the environmental change. The simulation results show that the research and design strategy significantly improves the learning efficiency and path planning success rate of the decoration robot, and shows good adaptability to the dynamic environment, which has important reference significance for the practical application of the decoration robot.**

***Keywords*—Interior decoration; path planning; deep reinforcement learning; reward function; credit allocation**

I. Introduction

Smart city is the universal vision of urban development in various countries. It uses modern technologies such as the Internet, big data and the Internet of things to solve urban governance problems, involving many fields such as architecture, transportation, medical treatment and smart home [1]. Intelligent robot is indispensable in the comprehensive environment of smart city [2]. In smart city architecture, the building scale is gradually expanding, but the number of labor is decreasing. Building construction automation has become an inevitable trend to improve production efficiency and quality. Robot technology provides new conditions and foundation for this [3]. On the whole, the development of construction robots is still at an early stage. The main robots that have been put into use at home and abroad are wall construction robots, decoration robots, facade maintenance robots, 3D printing robots, etc. [4-6]. The decoration robot has realized the plastering, painting, door and window installation and other processes of interior decoration, which has promoted the automation of interior decoration. The decoration robot mainly includes sensor, positioning system, control system and other parts. The sensor combined with the positioning system enables the decoration robot to realize the movement path planning, positioning, obstacle avoidance and other actions. Traditional robots use simultaneous localization and mapping (SLAM) to construct prior knowledge to avoid obstacles [7]. Therefore, the degree of intelligence is not high, and automatic identification cannot be carried out. Faced with the uncertainty of the actual indoor environment, the adaptability to the dynamic environment is not high, and the ability of path planning and obstacle avoidance needs to be improved. Therefore, a path planning strategy based on PPO deep reinforcement learning algorithm is proposed, which enables the decoration robot to obtain external information through the sensor system, make path selection decisions through autonomous learning, and improve the intelligent degree of the robot. Through the auxiliary model based on adaptive loss function, the adaptive degree of the robot is improved, providing a new way for the practical application of the indoor decoration robot.

II. Related Work

Interior decoration is an important part of urban construction, and its construction automation is an important direction of future development. In this regard, robot technology is a key technology, in which the research on robot path planning is an important issue and research hotspot. Using voxel space modeling and vector field path optimization, Min JK et al. proposed a new method for planning the robot's three-dimensional obstacle avoidance path to realize the robot's three-dimensional obstacle free path planning in the factory environment [8]. Guo X et al. proposed a robot intelligent assembly path planning scheme based on the improved Q-learning algorithm by adding a dynamic reward function to accelerate the convergence speed and design a trap avoidance scheme. Experiments show that the six joint robot UR10 has a good three-dimensional path optimization ability under this algorithm [9]. Zhang S et al. used information entropy to describe the population diversity of ant colony algorithm, thus improved an adaptive ant colony algorithm and improved the optimization ability of the algorithm. Through evaluation and significance test, it is proved that the algorithm is feasible in mobile robot path planning and has good performance [10]. Zhao W et al. proposed a heuristic search path planning method to solve the defects of traditional path planning methods in high demand environments. The experimental results show that more robots in narrow and long channels can run in a coordinated and

*Corresponding Author

orderly manner and have better path planning ability [11]. Jing et al. designed an intelligent interactive system for home robots, so that the robot has a perception and positioning system, and can realize autonomous navigation and obstacle avoidance. Experiments show that the robot using this system can better adapt to the family dynamic environment, and the human-computer interaction effect is good [12]. Islam M R et al. reset and introduced a new basic operator of chemical reaction optimization algorithm, and proposed an improved meta heuristic path planning algorithm. The experimental results show that the algorithm has good performance and can effectively improve the robot's path optimization ability in complex tasks [13].

Sharma K and Doriya R use the intelligent distance measurement method to improve the path planning ability of multi robot systems when performing tasks in large warehouses. This method has achieved good results in the coordinated operation and path optimization of multi robot systems [14]. Kang J G et al. designed a rerouting method using trigonometric inequality to improve the Rapid Exploring Random Tree (RRT). The simulation results show that, compared with RRT and RRT connection algorithm, the planning ability and path optimization ability of this algorithm are improved [15]. Zhang Z et al. proposed a path planning method based on the improved A* algorithm according to the special kinematic characteristics of the spherical motion robot, which improved the search efficiency of the robot and could find the optimal path in a short time [16]. Zhang TW et al. proposed an improved firefly algorithm (FA) combined with genetic algorithm (GA) to solve the defect of local optimal solution of FA algorithm. The experimental results show that the performance and accuracy of the improved algorithm are improved, and good results can be achieved when applied to robot path planning [17]. Chi W et al. proposed a heuristic path search method by using the feature extraction algorithm of Generalized Voronoi Diagram (GVD) to improve the path planning ability of fast robots in obstacle environments. The results show that this method is highly efficient, the results of one-time feature extraction can be reused, and the robot can efficiently search the obstacle free path [18]. Yang Y et al., based on the global and local levels, used the two-level path planning method to enable the robot to effectively avoid obstacles in a multi obstacle mobile environment. The results show that the actual application effect is good [19]. Xu T et al. perceived the shape of obstacles and planned the obstacle avoidance path of the robot arm under specific circumstances by improving the artificial potential field method. The results show that the method is highly adaptable to the environment and can effectively avoid obstacles [20]. In spherical tank inspection, Li J et al. planned the path of their inspection robot by improving Fleury algorithm, and recognized the welding line through depth learning network. The results show that the application effect of the path planning method is good [21].

From many researches on robot path planning, it can be seen that the current common path planning methods mainly include graph planning algorithm, spatial sampling, deep reinforcement learning and so on. The deep reinforcement learning method combines the deep learning and reinforcement learning, so that the robot can realize the autonomous learning of path planning in the interaction with the external environment. Compared with the traditional methods, it has obvious advantages, and is more suitable for the path planning of indoor construction of decoration robots. However, previous studies have mostly used the existing algorithm framework to build models, neglecting the design of the algorithm itself, especially the shaping of the reward function, and the quality of the reward function is directly related to the effect of robot learning. The novelty of this paper is to optimize the algorithm through the construction of reward function, aiming to improve the ability of robot autonomous planning. In this regard, this research focuses on the shaping and use of reward function, and deeply discusses the deep reinforcement learning method of robot path planning, hoping to propose an effective autonomous planning learning scheme.

III. Design of Self Planning Learning Scheme for Mobile Path of Decoration Robot

A. *Learning Method of PPO Algorithm and FCNN Neural Network*

With the development of smart cities, intelligent interior decoration has become a new direction of interior decoration development. Robots replace manual workers to carry out relevant decoration operations. During this period, the robot path planning problem needs to be solved. In this paper, PPO algorithm is used as the basic algorithm of path planning learning for decoration robot. This algorithm is a gradient optimization algorithm based on policy garden and off policy. It combines the advantages of strategy and value function, and performs well in tasks with continuous control and continuous scenarios. Generally, a reinforcement learning process can be described as a Markov Decision Process (MDP) [22]. This process can be expressed as a four tuple $M=(S,A,R,P)$, S is the environment state set, A is the action set, R is the reward function, and P is the state transition function. In the interactive environment, at each time step t, the agent's status is s_t and the action is a_t. It gets a reward r_{t+1} and moves to the next state s_{t+1}. When the cumulative reward reaches the maximum G_t, as shown in equation (1), the agent enters the termination state.

$$G_t=\sum_{n=0}^{T-t-1}\gamma^n R_{t+n+1} \tag{1}$$

In equation (1), $\gamma\in[0,1]$ is the attenuation coefficient. PPO algorithm is based on Actor-Critic algorithm. The Actor-Critic algorithm combines two reinforcement learning methods, value based and policy based, and is divided into two different network structures, Actor and Critic [23]. The idea of the algorithm is to use Actor network to generate actions to interact with the environment, and Critic network to evaluate actions. In the strategy π_φ, the state estimation of Critic output is shown in equation (2):

$$Q_\rho(s,a)\approx Q^{\pi_\varphi}(s,a) \tag{2}$$

In equation (2), ρ is the parameters of Actor network, φ is the parameters of Critic network. Therefore, as a Policy Gradient (PG), the calculation of Actor-Critic algorithm is shown in formula (3):

$$\nabla_{\varphi} J(\varphi) \approx E_{\pi_{\varphi}}\left[\nabla_{\varphi} \log \pi_{\varphi}(s,a) Q_{\rho}(s,a)\right] \tag{3}$$

The Policy Gradient has a high square difference, which can be solved by subtracting a Baseline [12]. During actual operation, you can select a Baseline function and add an Advantage function in the form of $A^{\pi_{\varphi}}(s,a) = Q^{\pi_{\varphi}}(s,a) - B^{\pi_{\varphi}}(s)$, so as to obtain a new calculation formula of Actor-Critic algorithm, as shown in (4):

$$\nabla_{\varphi} J(\varphi) \approx E_{\pi_{\varphi}}\left[\nabla_{\varphi} \log \pi_{\varphi}(s,a) A^{\pi_{\varphi}}(s,a)\right] \tag{4}$$

According to formula (4), the Actor-Critic algorithm adopts online update, that is, it needs to collect data again after completing an update, which is inefficient. In this regard, PPO algorithm introduces Importance Sampling into the existing framework, so that samples can be reused and offline updates can be realized.In this method, when it is impossible to obtain the expected $e(x)$ of a variable x subject to a continuous random distribution m, a distribution n with close values is set, and the expected $f(x)$ of the distribution is calculated to obtain the expected score distribution m, as shown in formula (5):

$$\begin{aligned} E_{x\sim m}\left[e(x)\right] &= \int f(x) m(x) dx = \int f(x) \frac{m(x)}{n(x)} n(x) dx \\ &= E_{x\sim m}\left[e(x)\frac{m(x)}{n(x)}\right] \end{aligned} \tag{5}$$

Therefore, the Actor-Critic gradient calculation formula of Importance Sampling is introduced, that is, the gradient calculation formula of PPO algorithm is shown in (6):

$$J^{\bar{\varphi}}(\varphi) \approx E_{(s_t,a_t)-\pi_{\varphi}}\left[\frac{m_{\varphi}(a_t \mid s_t)}{m_{\bar{\varphi}}(a_t \mid s_t)} A^{\bar{\varphi}}(s_t,a_t)\right] \tag{6}$$

The path planning training flow chart of PPO algorithm is shown in Fig. 1.

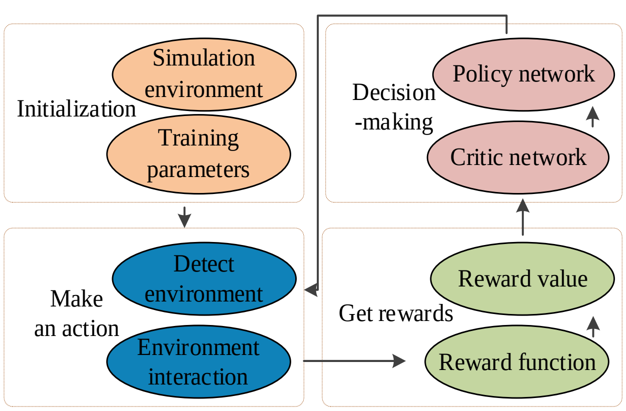


Fig. 1. PPO training process

FCNN neural network is used to process the state information of laser radar and robot. The algorithm has good fitting effect on nonlinear data. The full connection layer is converted into a convolution layer, which can be classified at the pixel level. However, it is not precise enough in the sampling process and is insensitive to some details. FCNN consists of input layer, hidden layer and output layer. The basic unit is neurons. There is no correlation between neurons in each layer, but all neurons in the other layer are associated. Information propagates unidirectionally from input layer to input layer. Its structure is shown in Fig. 2.

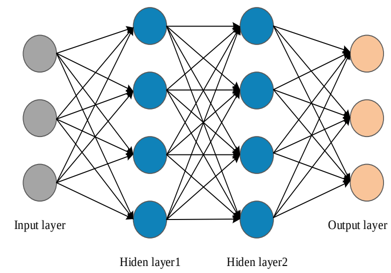


Fig. 2. Structure of FCNN

The propagation mode of FCNN can be expressed as formula (7):

$$\begin{cases} L_i = K_{i-1} H_i + C_i \\ K_i = F_i(L_i) \end{cases} \tag{7}$$

In equation (7), L_i represents the linear input of layer, K_i represents the output after layer activation, H_i represents the weight matrix from layer to layer, C_i represents the offset vector of layer, and F_i represents the activation function of layer. In deep learning, a reverse algorithm is usually used to make FCNN propagate from the output layer to the input layer to update the neuron weight matrix and bias vector. The principle is to bring the data set $I = \{(x^p, y^p)\}_{p=1}^{P}$ into FCNN training so that the output layer get $\hat{y}^p$ as shown in formula (8):

$$\hat{y}^p = FCN\left[\left(x^p, \varphi\right)\right] \tag{8}$$

In Eq. (8), $FCN[\ ]$ represents a FCNN neural network and φ represents the training parameters of the FCNN network. After we get $\hat{y}^p$, we can use the cost function to calculate $Loss(y^p, \hat{y}^p)$, and the gradient of the network can be obtained according to the partial derivative and learning rate σ of the loss function to the weight and offset. Finally, the inverse algorithm updates the weight H and offset C of the network, as shown in equation (9):

$$\begin{cases} H - \sigma \dfrac{\partial Loss(H,C)}{\partial H} \Rightarrow H \\ C - \sigma \dfrac{\partial Loss(H,C)}{\partial C} \Rightarrow C \end{cases} \tag{9}$$

B. Design and Optimization of Reward Function

In the process of deep reinforcement learning, the learning goal is to achieve the maximum cumulative reward, so the setting of reward function is directly related to the realization effect of learning goals [24]. The research aims at the problem of autonomous planning of the robot's indoor moving path, which involves two key abilities, namely, the ability of path optimization and the ability of obstacle avoidance. Therefore, two potential energy reward functions can be created respectively, and an initial reward function can be combined to get a better training effect. The initial reward function $R(s,a,\vec{s})$ is shown in equation (10):

$$R(s,a,\vec{s}) = \begin{cases} 40 & if\ \delta_d < 0.25m \\ -20 & if\ \delta_0 < 0.2m \\ -0.15 & p(every\ step) \end{cases} \quad (10)$$

In Eq. (10), $\delta_d < 0.25m$ means that the robot successfully reaches the target within 0.25m, successfully completes the task, and the reward value is 60, and starts the next round of training; $\delta_0 < 0.2m$ means that if the robot reaches within 0.2m of the obstacle and the task fails, the reward value is -30 and the next round of training starts; p is punish, motivating robots to perform training tasks. For the path optimization capability, it involves the distance from the end of the task and the angle of the travel direction. First, set the potential energy function F_r for the target distance as shown in equation (11):

$$\begin{cases} F_r(s,a,\vec{s}) = \delta_r(\vec{s}) - \delta_r(s) \\ \delta_r = F_\varepsilon(r_{ax} - r_{tx}, r_{ay} - r_{ty}) \\ F_\varepsilon = \sqrt{(\nabla rx)^2 + (\nabla ry)^2} \end{cases} \quad (11)$$

In Eq. (11), r_{ax}, r_{ay} indicates the position of the target position on the x-axis and y-axis; r_{tx}, r_{ty} indicates the position of the trolley in the x-axis and y-axis; δ_r represents the distance between the robot and the target; ∇rx, ∇ry represents the difference between the target position and the trolley position in the x-axis and y-axis; δ_r represents Euclidean distance; s indicates the current status; a indicates the action taken in the current state; $\vec{s}$ indicates the next state to move to. Set the potential energy function F_a for the robot travel direction angle δ_g as shown in equation (12):

$$\begin{cases} F_g(s,a,\vec{s}) = \delta_g(\vec{s}) - \delta_g(s) \\ \delta_g = \begin{cases} \pi/4 & if\ g \le \pi/4 \\ abs(g) & if\ g \le \pi/4 \end{cases} \end{cases} \quad (12)$$

In Eq. (12), g represents the orientation angle between the robot and the target position, and $abs(g)$ represents the absolute value. Within the range of 2 meters in diameter centered on the robot, the distance δ_d between the obstacle and the robot sets the potential energy reward function F_d as shown in Eq. (13):

$$\begin{cases} F_d(s,a,\vec{s}) = \delta_d(\vec{s}) - \delta_d(s) \\ \delta_d = \begin{cases} otr & if\ otr \le 1 \\ 1 & if\ otr \ge 1 \end{cases} \end{cases} \quad (13)$$

In Eq. (13), otr is the distance between the current robot and the obstacle. Adding the three potential energy reward functions set to the initial reward function $R(s,a,\vec{s})$ is the new reinforcement learning reward function, as shown in Eq. (14):

$$\dot{R}(s,a,\vec{s}) = R(s,a,\vec{s}) + F_r + F_g + F_d \quad (14)$$

In the training scheme adopted in the study, every time the robot reaches the target point or fails to avoid obstacles, it will re-enter the new training process, and each round will obtain a training path t_s including reward sequence r_s, as shown in Eq. (15):

$$t_s = s_1, a_1, r_1, s_2, a_2, r_2, \ldots, s_n, a_n, r_n \quad (15)$$

This training process is carried out in rounds until the cumulative reward reaches the maximum G_t. The global training in this way involves a credit allocation problem [25]. Generally speaking, in the whole process of robot path selection, the final selection step is relatively more Critic. However, as the number of training steps increases, the discount coefficient decreases exponentially. Therefore, the correlation between the reward return performance of the previous steps and the number of subsequent steps is very low, so that the variance of the final cumulative return is too high. This has a great impact on the update of the strategy gradient, and the sample utilization and training efficiency are therefore low. In order to solve this problem, we can build a credit allocation model, reallocate the reward value at the end of the decision sequence, and obtain the cumulative return value with smaller variance, so as to make the gradient update of reinforcement learning algorithm more stable and accelerate the training speed. The training structure of adding credit allocation model is shown in Fig. 3.

In Fig. 3, S_{ld} is the lidar information, S_s is the self-status information, $V_\eta(s)$ is the status output value of the critical network, $\pi_\iota(a)$ is the action distribution of the actor network, and $V_\varphi(s)$ is the status output value after credit allocation.After a complete path planning training, the robot can obtain three different reward values, namely, the positive reward value 60 for completing the training, the negative reward value -30 for collision obstacles, and the common reward value [-2.0,2.0] for reaching the maximum number of training steps. Each reward is shown in Fig. 4.

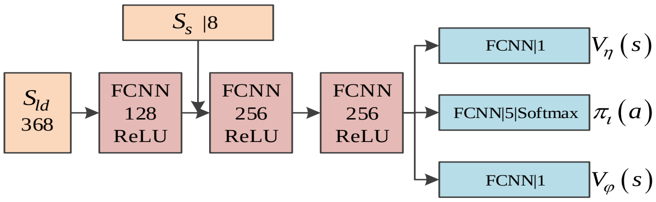
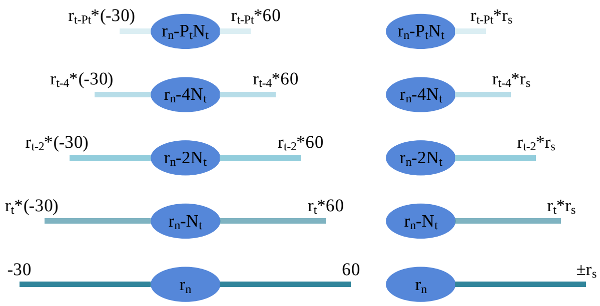


Fig. 3. Training model after adding credit allocation



（a）Positive and negative rewards （b）Normal rewards

Fig. 4. Credit allocation models

The core of the problem to be solved is to allocate the reward impact of the later stage to the reward of the previous stage. First, the whole sequence needs to be segmented. The operation requirements are shown in formula (16):

$$P_t = round\left(S_t / N_t\right) \tag{16}$$

In Eq. (16), S_t is the reward sequence length, N_t is the credit allocation indicator, that is, the sequence length divided, N_t is the number of divided sequences, and $round(\ )$ refers to rounding. On the basis of the new segmented sequence, the reward of the later segment of the original sequence is redistributed in the direction from the back to the front and the discount coefficient γ is taken as the criterion, so as to update the reward sequence r_s and recalculate the cumulative reward G_t. Then, the advantage functions AF_{old} and AF_{new} of the original sequence and the updated sequence are introduced respectively, and their values are the respective maximum cumulative reward G_t and the state output value of the Critic network. The two advantage functions are multiplied by the coefficients κ and λ respectively, so obtain the final form AF of the reward function, as shown in formula (17):

$$AF = \kappa AF_{old} + \lambda AF_{new} \tag{17}$$

In addition, in order to make the credit allocation perceived by the algorithm, a CACritic network is introduced to output additional credit allocation status. Finally, the loss functions of CACritic network, Critic network and Actor network are calculated by PPO algorithm respectively, so as to complete the shaping of the reward function.

C. Auxiliary Model Construction of Adaptive Loss Function

In the static environment, because the elements of the environment are fixed, the “State-Action-Reward” of machine training is fully responsive. As long as the time and strategy required for training are met, the robot can tend to repeat training on a successful path or finally retain the training state related to all environmental elements, so it is easy to obtain the optimal path and achieve the maximum reward. However, in the real decoration scene, the environmental state will change constantly, and new structures and elements will be generated. Taking the same action at different times of the same path may result in different reward situations, which requires more adaptation to costs and has a negative impact on machine learning. The change of this environment is accidental. Therefore, the study assumes that the uncertain s-a-r condition obeys the Gaussian distribution, and then constructs an auxiliary model to predict this distribution. In the training process, the uncertainty of s-a-r is transformed into the uncertainty of obtaining rewards to affect the robot's decision-making. Therefore, the core of the auxiliary model is to connect two FCNN on the basis of the training model to predict the mean and variance of s-a-r distribution, so as to represent the mean of the reward and measure the uncertainty of the reward, and then use the adaptive regression loss function to compare the predicted value with the actual reward value. The reward uncertainty perceived by the robot with the help of the auxiliary model can be used as a reference for the next path planning decision-making, so as to improve the adaptability of the robot and speed up the speed of completing the learning task. The training structure with auxiliary model is shown in Fig. 5.

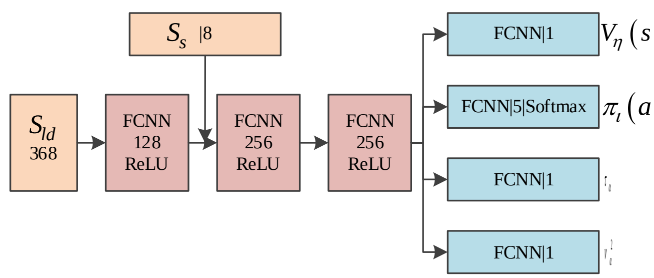


Fig. 5. Training model after adding credit allocation auxiliary task

In Fig. 5, τ_α is the s-a-r distribution mean predicted by the auxiliary model, and v_α^2 is the distribution variance.The specific method to build the model is to set the S-A-R data set of the dynamic environment, which follows the Gaussian distribution, described as $Trend(s,a,r)$,as shown in Eq. (18):

$$Trend\left(s,a,r\right) \in N\left(\tau_\alpha, \upsilon_\alpha^2\right) \tag{18}$$

In Eq. (18), τ_α is value of distribution, υ_α^2 is variance of distribution, used to measure the uncertainty of reward. FCNN is used to predict the mean and variance of S-A-R distribution. The prediction results are compared with the actual reward situation for regression training, and then the adaptive loss function W_α is used to calculate the difference between the two, as shown in equation (19):

$$W_\alpha = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{2\upsilon\left(s_j\right)^2} \left\| h\left(s_j\right) - r_j \right\|^2 + \frac{1}{2}\log \upsilon\left(s_j\right)^2 \tag{19}$$

In Eq. (19), s_j represents the state of the path, r_j is the reward obtained in this state, $h(s_j)$ is the predicted mean value obtained by the auxiliary model, $\upsilon(s_j)^2$ is the variance obtained by the prediction model, and $\log\upsilon(s_j)^2$ is the constraint added to prevent the gradient turbulence caused by too high variance. Finally, the loss functions of CACritic network, Critic network and Actor network are calculated by PPO algorithm respectively, so as to complete the addition of auxiliary model.

IV. Effect Analysis of Self Planning Learning Strategy for Mobile Path of Decoration Robot

A. *Simulation Experiment Setup*

The common robot operating system (ROS) is used as the programming framework of the robot running program, and the simulation environment is built in combination with the Gazebo simulation platform. Gazebo can perform simulation independently of ROS, or install ROS related function packs to build an environment model as a node of ROS. In addition, the turtlebot3 of the turtlebot series robots is selected as the simulation robot. It has high support for ROS, strong modularity and flexibility, and has a large number of relevant resources in the ROS community, which brings great convenience to the research. The ROS simulation training framework thus established is shown in Fig. 6.

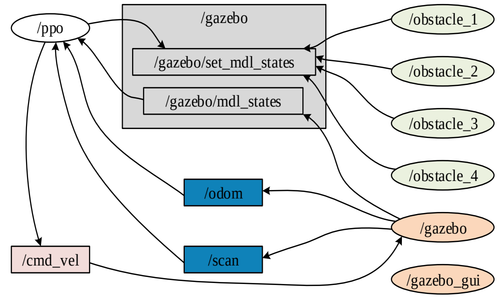


Fig. 6. Framework of ROS simulation training

In Fig. 6, /ppo refers to the algorithm used in the experiment/ cmd_ Vel is the moving speed command sent to the robot/ gazebo and /gazebo_ GI refers to the Gazebo simulation program/ gazebo/set_ mdl_ States and /gazebo/mel_ States means setting obstacles and robot states/ Odom and /scan refer to mobile state sensor and laser ranging sensor/ obstacle_ 1~/obstacle_ N means moving obstacles.Firstly, the experiment simulation environment is built by using Gazebo software. The research is aimed at the indoor movement of the decoration robot. Therefore, considering the actual situation and the needs of experimental training, two kinds of square enclosed spaces including static obstacles and dynamic obstacles are built. The plan is shown in Fig. 7.

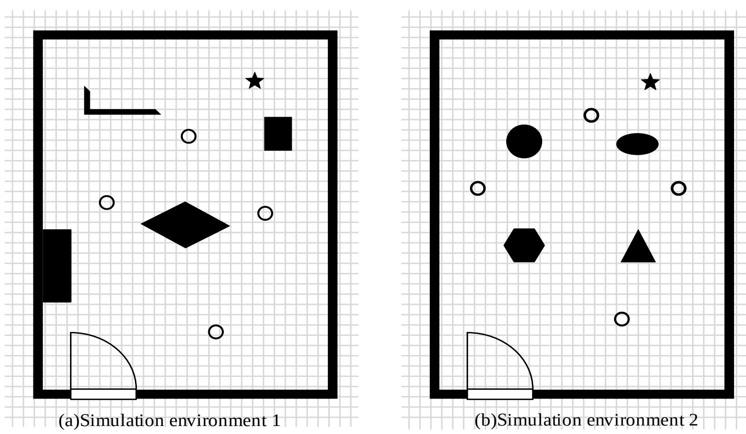
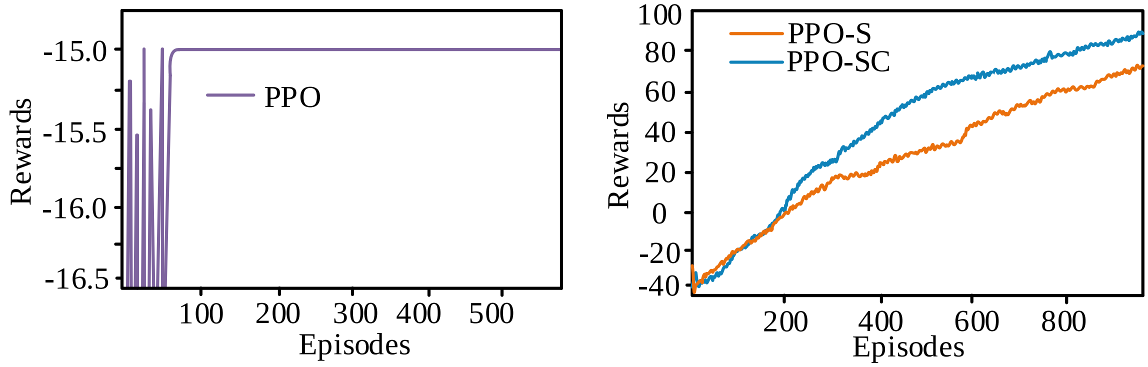


Fig. 7. Gazebo simulation environment

In Fig. 7, the closed block represents the static obstacle, the circular graph represents the dynamic obstacle, and the five pointed star represents the target position. Secondly, the linear velocity and angular velocity at different levels are set for the action parameters of the robot; For the running environment of the experimental program, the operating system is Ubuntu20.04 LTS, the processor is Intel i5 12600K, the graphics card is NVIDIA RTX2070Ti, python version 3.9, and pytoch version 1.9.1; The machine status includes real-time distance to the target, travel direction angle, travel speed information and completed action information; The ranging system adopts light detection and ranging (LIDAR).

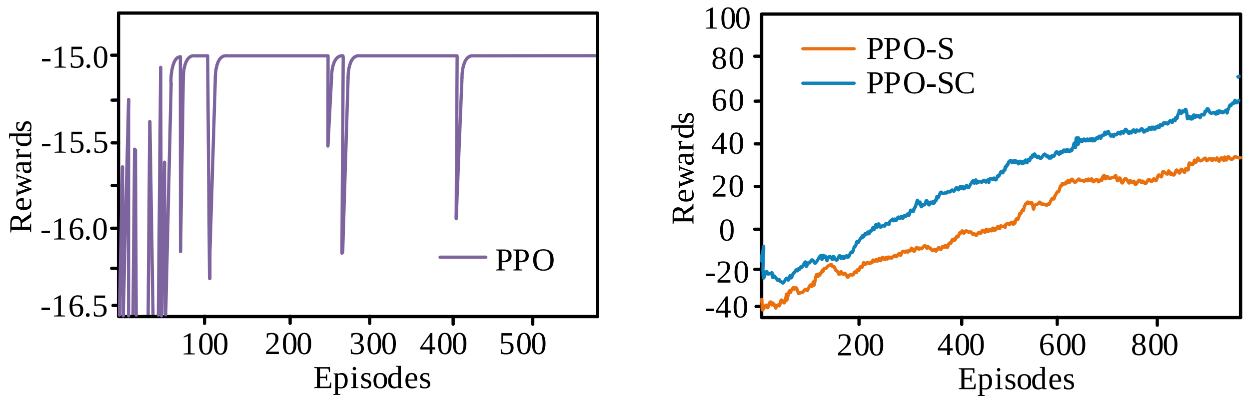
B. *Path Planning Ability Test and Performance Analysis of Decoration Robot*

Firstly, the training effect of path planning for decoration robot is compared between the initial reward function training model PPO and the improved reward function training model. The improved reward function training model is divided into improved reward function PPO-S and added credit allocation model PPO-SC. In static simulation environment 1, the path planning training effect of each model is shown in Fig. 8.



(a)Reward effect of PPO (b)Reward effect of PPO-S & PPO-SC

Fig. 8. Simulation environment 1(Static state)



(a)Reward effect of PPO (b)Reward effect of PPO-S & PPO-SC

Fig. 9. Simulation environment 1(Dynamic state)

The results show that the PPO model robot has more collisions with its random exploration route, and gets some negative rewards. Finally, in order to avoid collision, it stays in place, and the reward remains at -15, and the task fails. PPO cannot guide the robot to complete the training task. In the PPO-S and PPO-SC models, the robot can complete the learning of path planning, and the PPO-SC model has faster rising speed and higher success rate. Therefore, in static environment 1, the new reward function and credit allocation model improve the robot's path planning ability. In dynamic environment 1, the training effect of each model is shown in Fig. 9.

The results show that compared with the static environment, the robot collision situation of PPO model increases and gets more negative rewards. In the simulation environment, the robot will still collide with the dynamic obstacles when it is stationary. In PPO-S and PPO-SC models, the training efficiency of the robot is significantly lower than that of the static model, and the stability is poor, but the training effect of PPO-SC is still better. To sum up, PPO-S with potential energy function can effectively guide the robot to achieve the goal and complete the training task, and the learning ability is improved and the training effect is better after credit distribution and reward sequence redistribution. However, the dynamic environment adaptability is poor and the training effect is average. In order to verify the effect of auxiliary tasks on improving the adaptability to dynamic environment, the training effect of PPO-S model and PPO-SA model combined with auxiliary tasks is compared. The results are shown in Fig. 10.

The results show that in both static and dynamic environments, PPO-SA model has better reward performance than PPO-S model. The reward rises faster and the learning efficiency is higher. In the dynamic environment, the robot training effect under PPO-S model is significantly worse than that in the static environment, the number of collisions has increased significantly, the training effect is poor, and the reward curve fluctuates greatly. In the PPO-SA model, through early learning, the auxiliary model in the middle and late stages can effectively predict the reward distribution, so that the robot can avoid obstacles in advance in route selection, the training effect has been significantly improved, and the reward curve is relatively stable. In order to further verify the effect of the research and design strategy in robot path planning, the deep deterministic policy gradient (DDPG) algorithm, A-star algorithm and RRT algorithm are used to design the robot indoor path planning modell. Conduct training in static environment 1, and compare the percentage of the difference between the completed path and the optimal path. The results are shown in Fig. 11.

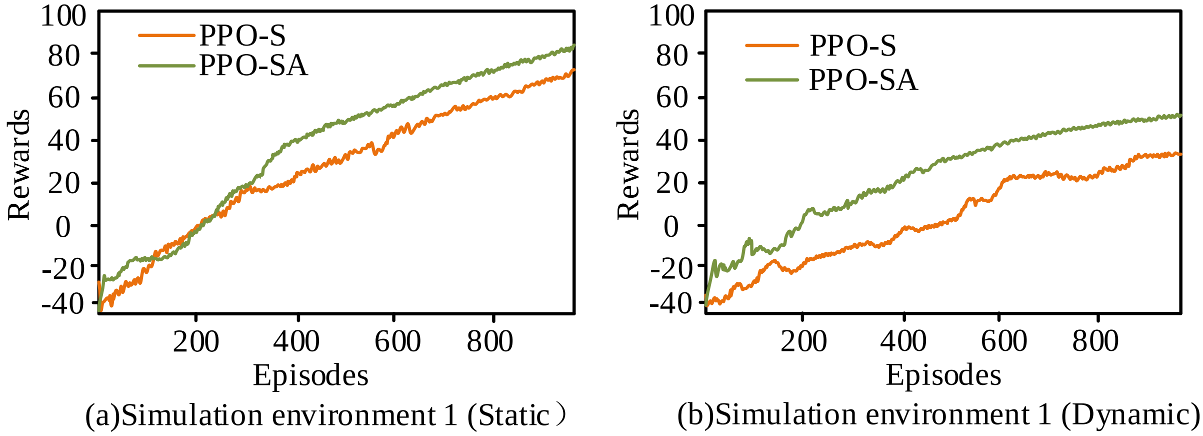


Fig. 10. Reward effect of PPO-S & amp; PPO-SA

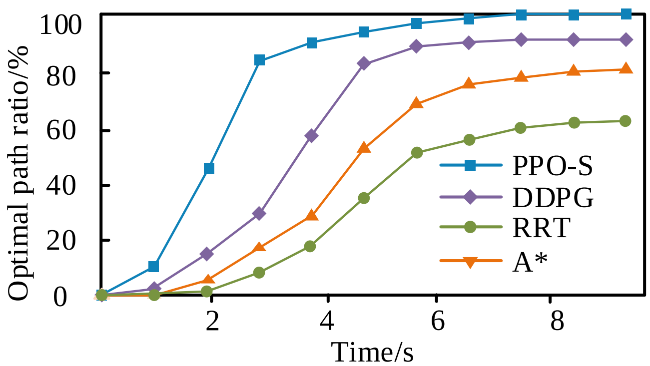


Fig. 11. Ratio of path difference in simulation environment1

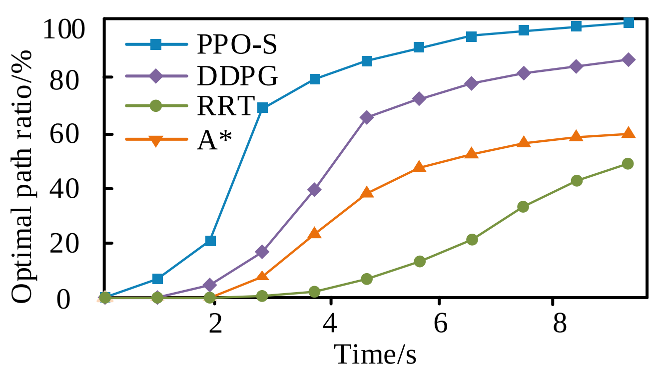


Fig. 12. Ratio of path difference in simulation environment2

It can be seen from the figure that in the static simulation environment 1, the path optimization ability of the PPO-S model studied and designed is better, the curve rises fastest, and the time to find the optimal path is the shortest. Conduct training in the static simulation environment 2, and compare the proportion of the difference between the completed path and the optimal path. The results are shown in Fig. 12.

It can be seen from the figure that in the static simulation environment 2, the path optimization ability of the PPO-S model studied and designed is better, the curve rises fastest, and the time to find the optimal path is the shortest. Each model is used to train robots in dynamic environment 1 and dynamic environment 2. Compared with PPO-SA model, the results show that the designed training model has achieved good training effect, while other algorithms used for comparison are not applicable in dynamic environment. It shows that the traditional common path planning algorithms are usually placed in static environment, and it is difficult to deal with dynamic environment without adding improvement measures. In the static simulation environment 2, 50 simulation experiments were conducted, and the statistics of the success times and success rates of different models are shown in Fig. 13.

TABLE I. SUCCESS TIMES AND SUCCESS RATE OF THE MODE

Model	Success times (times)	Success rate (%)
PPO-S	47	94
DDPG	40	80
RRT	21	42
A*	33	66

In Table I, from the success times and success rates of different models, PPO-S model has the highest success rate of 94%, significantly higher than other models, while DDPG model takes the second place, with a success rate of 80%. The success rate of A * model is the lowest, with only 21 successful times. It can be seen that the adaptability of PPO-S model is relatively optimal.

V. CONCLUSION

In order to promote the application of decoration robot in automatic decoration construction, it is a key problem to solve the autonomous planning of mobile path. In this paper, FCNN neural network is introduced into PPO algorithm and applied to interior decoration path planning. Aiming at the two core problems of path planning, a potential reward function is established. The reward function is redesigned based on the initial reward function. In order to solve the problem that the variance of cumulative reward returns is too large due to the difference between the front and rear reward values, a credit allocation model is constructed. By predicting the distribution of rewards through the auxiliary task model, the robot can predict the distribution of dynamic obstacles through training, and avoid obstacles in advance. The simulation results show that compared with the training model of the initial reward function, the learning efficiency of the robot is significantly improved and there are fewer negative rewards in the training of the new reward function after adding credit allocation; In a dynamic environment, the reward curve is more stable, the learning efficiency is higher, and the adaptability is better. Compared with other common path planning algorithms, PPO-S model has better performance in path planning, with the highest success rate of 94%, 14% higher than DDPG. Therefore, the research and design of deep reinforcement learning algorithm for indoor mobile path planning of decoration robot is feasible, and the training effect is also very good, which has important practical significance. Applying this scheme to the actual environment is the next research direction.

ACKNOWLEDGMENT

The research is supported by: 2018 Guangdong Young Innovative Talents Project “Research on the application of soft decoration materials in interior decoration design from the perspective of Internet” (Project No.: 2018gwqncx141).

REFERENCE

[1] Y. Chen, D. Han. “Water quality monitoring in smart city: A pilot project,” Automation in Construction, vol. 89, pp. 307-316, 2018.

[2] H. Liu, Y. Deng, D. Guo, et al. “An Interactive Perception Method for Warehouse Automation in Smart Cities.,” IEEE Transactions on Industrial Informatics, PP(99):1-1, 2020.

[3] A. J. Wit, L. Vasey, V. Parlac, et al. “Artificial intelligence and robotics in architecture: Autonomy, agency, and indeterminacy.” International Journal of Architectural Computing, vol. 16, no. 4, pp. 245-247, 2018.

[4] S. Seriani, A. Cortellessa, S. Belfio, et al. “Automatic path-planning algorithm for realistic decorative robotic painting.” Automation in Construction, vol. 56(aug.), pp. 67-75, 2015.

[5] Y. S. Lee, S. H. Kim, M. S.Gil , et al. “The study on the integrated control system for curtain wall building faade cleaning robot,” Automation in Construction, vol. 94(OCT.), pp. 39-46, 2018.

[6] T. Kim, H. Lim, K. Cho, “Conceptual robot design for the automated layout of building structures by integrating QFD and TRIZ.” The International Journal of Advanced Manufacturing Technology, vol. 120, no. 3, pp. 1793-1804, 2022.

[7] H. Chen, H. Huang, Y. Qin, et al. “Vision and laser fused SLAM in indoor environments with multi-robot system.” Assembly Automation, vol. 39, no. 2, pp. 297-307, 2019.

[8] J. K. Min, P. Kang, “Generation of a 3D robot path for collision avoidance of a workpiece based on voxel and vector field.” Journal of Mechanical Science and Technology, vol. 36, no. 1, pp. 385-394, 2022.

[9] X. Guo, G. Peng, Y. Meng, “A modified Q-learning algorithm for robot path planning in a digital twin assembly system.” The International Journal of Advanced Manufacturing Technology, vol. 119, no. 5-6, pp. 3951-3961, 2022.

[10] S. Zhang, J. Pu, Y. Si, “An Adaptive Improved Ant Colony System Based on Population Information Entropy for Path Planning of Mobile Robot.” IEEE Access, PP(99):1-1, 2021.

[11] W. Zhao, R. Lin, S. Dong, et al. “Dynamic node allocation based multirobot path planning.” IEEE Access, PP(99):1-1, 2021.

[12] Q. Li, J. Wu, et al. “Autonomous Tracking Control for Four-Wheel Independent Steering Robot Based on Improved Pure Pursuit.” Journal of Beijing Institute of Technology, vol. 29, no.106(04), pp. 35-42, 2020.

[13] M. R. Islam, P. Protik, S. Das, et al. “Mobile robot path planning with obstacle avoidance using chemical reaction optimization.” Soft Computing, pp. 1-28, 2021.

[14] K. Sharma, R. Doriya, “Coordination of multi-robot path planning for warehouse application using smart approach for identifying destinations.” Intelligent Service Robotics, vol. 14, no.2, pp. 313-325, 2021.

[15] J. G. Kang, D. W. Lim, Y. S. Choi, et al. “Improved RRT-Connect Algorithm Based on Triangular Inequality for Robot Path Planning.” Sensors, vol. 21, no. 2, pp. 1-34, 2021.

[16] Z. Zhang, Y.Wan, Y. Wang, et al. “Improved hybrid A* path planning method for spherical mobile robot based on pendulum.” International Journal of Advanced Robotic Systems, vol. 18, no. 1, pp. 671-680, 2021.

[17] T. W. Zhang, G. H. Xu, X. S. Zhan, et al. “A new hybrid algorithm for path planning of mobile robot.” The Journal of Supercomputing,: pp. 4158–4181, 2021.

[18] W. Chi, J. Wang, Z. Ding, et al. “A Reusable Generalized Voronoi Diagram Based Feature Tree for Fast Robot Motion Planning in Trapped Environments.” IEEE Sensors Journal, PP(99):1-1, 2021.

[19] Y. Yang, Z. Lin, M. Yue, et al. “Path planning of mobile robot with PSO-based APF and fuzzy-based DWA subject to moving obstacles.” Transactions of the Institute of Measurement and Control, vol. 44, no. 1, pp. 121-132, 2022.

[20] T. Xu, H. Zhou, S. Tan, et al. “Mechanical arm obstacle avoidance path planning based on improved artificial potential field method.” Industrial Robot,vol. 49, no. 2, pp. 271-279, 2022.

[21] J. Li, S. Jin, C. Wang, et al. “Weld line recognition and path planning with spherical tank inspection robots.” Journal of Field Robotics, vol. 39, no. 2, pp. 131-152, 2022.

[22] Q.V. Do, I.Koo, “Actor-critic deep learning for efficient user association and bandwidth allocation in dense mobile networks with green base stations.” Wireless Networks, vol. 25, no. 5, pp. 5057-5068, 2019.

[23] J. Guo, *L. I.* Mengtian, L. I. Quewei, et al. “GradNet: Unsupervised Deep Screened Poisson Reconstruction for Gradient-Domain Rendering.” ACM Transactions on Graphics, vol. 38, no. 6, pp. 223.1-223.13, 2019.

[24] X. Ning, H. Wang, et al. “Multi-Level Policy and Reward-Based Deep Reinforcement Learning Framework for Image Captioning.” IEEE Transactions on Multimedia, vol. 22, no. 5, pp. 1372-1383, 2019.

[25] B. A. Richards, T. P. Lillicrap, “Dendritic solutions to the credit assignment problem.” Current opinion in neurobiology, vol. 54, pp. 28-36, 2018.

Research on Face Recognition Technology of Subway Automatic Ticketing System based on Neural Network and Deep Learning

Shuang Wu[1], Xin Lin[2], Tong Yao[3*]
Department of Transportation Engineering, Yangzhou Polytechnic Institute, Yangzhou, China[1, 2]
School of Economics and Management, Sichuan Tourism University, Chengdu, China[3]

***Abstract*—Face recognition technology is the core technology of the subway ticketing system, which is related to the efficiency of people's ticket purchase. In order to improve people's experience of taking public transport, it is necessary to improve the performance of face recognition technology. In this study, the Back Propagation (BP) algorithm is used to optimize the parameters of the SoftMax classifier of the convolutional neural network, and the branch structure is added to the structure of the SphereFace-36 convolutional neural network to extract the local features of the face. Based on the improved neural network, the face recognition system of the subway automatic ticketing system is established. The results show that the area under the ROC curve is the highest for validation and identification of the optimization model; The recognition accuracy of the optimized model in different data sets is 1.0%, 0.7%, 1.1%, 0.9% and 0.6% higher than that of SphereFace-36 respectively, and its specificity is higher than that of SphereFace-36, with the maximum difference of 9%; The average accuracy of global feature extraction and recognition of the optimized network model is 83.01%. In the simulation experiment, the optimized model can accurately recognize facial features, which has high practical value and can be applied to the automatic ticketing system.**

Keywords*—Automatic ticketing system; BP; CNN; deep learning; face recognition; SphereFac； SoftMax classifier

I. Introduction

Face authentication does not require physical contact with other systems, has a high false acceptance rate and false rejection rate, and can minimize errors [1]. Because of these unique characteristics of face recognition systems, it is considered one of the effective methods of data protection. Face recognition technology is widely used in many fields such as security, medical care, education, etc., and face recognition is required in various social places, which is very important for security, transportation, etc. [2]. Face recognition is mainly divided into appearance recognition and feature recognition. By extracting facial features, it is judged whether it matches the face information in the image or video stream, to identify the identity [3]. However, in the real environment, face recognition technology will be affected by posture, gender, lighting, and other conditions, so it is necessary to improve and optimize the method to improve face recognition's accuracy [4]. Changes in head posture will bring challenges to face recognition technology, and an unconstrained environment will increase the difficulty of face recognition in the field of human-computer interaction. Researchers designed a tool to collect key pose samples from face datasets and then identify changing head poses through cascaded convolutional neural networks [5]. With the continuous development of society, people have higher and higher requirements for travel quality. Face recognition technology can help people save time cost of transportation. In the subway automatic ticketing system, fast ticket purchases can be realized by scanning the face. The application of face recognition technology has significantly improved the service efficiency and quality of the subway ticketing system. Due to factors such as the new crown epidemic, wearing masks for self-protection in public places is necessary, which makes the face recognition technology of the automatic ticketing system of the subway more difficult. As an important factor, the occlusion problem can affect the accuracy of facial landmark detection in face recognition, and the occlusion of the face affects the application of face recognition [6]. To improve the face recognition technology of the subway automatic ticketing system, the neural network was optimized in this experiment, and the accuracy of the face recognition system was improved through deep learning technology, hoping to meet people's travel needs. To improve the face recognition technology of the subway ticketing system, the neural network was optimized in this experiment. The research content of this paper is mainly to optimize the parameters of the SoftMax classifier of the convolutional neural network using the Error Back Propagation (BP) algorithm, and add a branch structure on the basis of the structure of the SphereFace-36 convolutional neural network to extract the local features of the face. Then, based on the improved neural network, the face recognition system of the subway ticketing system is established. Finally, the performance of the optimized neural network and face recognition system is simulated and analyzed. It is hoped that the accuracy of face recognition system can be improved through deep learning technology to meet people's travel needs.

II. Related Work

In promoting the development of machine learning and deep learning, the application of CNN plays an important part and can reduce the complexity of feedback nerves. With in-depth research of scientific research, CNN has been continuously optimized and improved, and it has been effectively applied in various fields such as medicine, autonomous driving, and face recognition. In the technical research of heartbeat detection in the medical field, Chandra B

*Corresponding Author.

et al. applied a convolutional neural network for multimodal data analysis, and CNN can effectively improve the robustness and accuracy of heartbeat detection [7]. In disease classification, the bilateral CNN method can effectively distinguish dystrophic scoliosis from non-dystrophic scoliosis. This method has good classification accuracy, has strong application value, and provides a great deal for clinical diagnosis. Good technical support [8]. O Elharrouss and other scholars cascade the CNN technology, perform pose estimation, face segmentation, and recognition through three CNN models, and realize face recognition from the left, front and rear directions of the human face. The model training results show that the new model has high robustness [9]. As one of the ways of people’s daily communication, image data shows explosive growth with network technology’s continuous development. Machine vision is based on image recognition. CNN can perform feature extraction and recognition of image information, thereby extracting and analyzing different images. However, when recognizing face images, CNN may lose some image feature information during propagation and computation, and the multi-level information fusion method can further recover the information lost in CNN propagation and computation [10]. SoftMax classifier is the main technology of convolutional neural network to realize feature classification. The enhanced algorithm based on SoftMax classifier can effectively improve the accuracy and processing time of face recognition [11]. BP algorithm can be used to train artificial neural network to realize face recognition process. To improve the performance of face recognition system, artificial neural network is combined with the famous meta heuristic optimization algorithm, namely harmony search algorithm (HSA) [12]. Genetic algorithm, particle swarm optimization algorithm, etc. are heuristic algorithms, which can effectively solve the combination problems by examples. In recent years, researchers have obtained meta heuristic algorithm by combining random algorithm with local search algorithm. This algorithm can explore the solution space more thoroughly, so as to obtain a better solution. For example, Dwarf Mongoose Optimization Algorithm [13], the arithmetic optimization algorithm [14], Reptile Search Algorithm (RSA) [15], and other optimization algorithms are all improvements to heuristic algorithms, which can be independent of problems. Although meta heuristic algorithm has certain advantages, its intrinsic parameters still need to be adjusted and further studied in order to make the technology better used for problem solving.

Led by the rapid development of computer science and other fields, face recognition technology has become increasingly sophisticated and perfected. In various industries and fields, face recognition technology has been well applied. For example, based on face recognition technology such as neural networks and UAV, the online attendance system for students in the education industry is designed, and the research on the estimation of long-distance absolute distance and target tracking of face recognition technology is designed [16-17]. Some literature has conducted in-depth discussions on the lateralization of face recognition. The study believes that although the left hemisphere has achieved language evolution and is also the realization area of facial recognition, there is no evidence that the right hemisphere does not correlate with face recognition [18]. Image feature extraction is a key technology for face image classification in face recognition systems. Nejrs S et al. applied two-dimensional discrete Fourier transform technology to face image classification. The performance difference comparison method based on the correlation coefficient is effective. The classification accuracy of the target image reaches more than 90%, which proves that the classification result of the new model method is accurate [19]. Face recognition technology includes the recognition of facial expressions, and some scholars use machine learning algorithms to optimize face recognition technology. By using the gradient direction histogram to collect facial emotion data, and then applying the fast learning network algorithm to classify the extracted features for facial emotion recognition, the recognition of people’s facial expressions and emotions is realized [20]. CNN is an efficient and accurate technical method in face recognition method and has high robustness [21]. To improve the distinguishing degree of feature extraction in face recognition technology, researchers use the optimized loss function to train the existing neural network and use the multi-action form supervision force to supplement the origin loss, which promotes the feature learning of face recognition [22]. Face recognition technology is easily affected by factors such as illumination and environment, which reduces the accuracy of face recognition. Chen C et al. proposed to use the fuzzy discriminant analysis method to establish a face recognition system, which can reduce the influence of the environment on the face recognition technology to a certain extent. This method calculates the fuzzy membership degree of the sample, calculates the fuzzy mean value, and then establishes a new model to achieve face feature extraction [23]. Some scholars have established a face beauty prediction model with a simple structure and suitable for embedding in small devices. The model uses CNN technology to extract facial features and expands the face database to predict facial beauty [24].

Facial sign detection is an important step in face recognition technology. Under the influence of the COVID-19, masks have become a necessity for people's daily travel, and mask occlusion has brought challenges to facial sign detection. Based on the above research, it can be seen that CNN technology has a good application effect in face recognition technology. At the same time, in order to improve the efficiency of CNN face recognition, other methods can be introduced to optimize the parameters. In this experiment, BP neural network is introduced to optimize the weight matrix and bias parameter of SoftMax classifier in CNN algorithm, and branch structure is added to the structure of SphereFace-36 convolution neural network to extract local features of face. It is hoped that the optimization and improvement of the method can improve the efficiency of the subway ticketing system and provide convenience for people to purchase tickets.

III. FACE RECOGNITION TECHNOLOGY OF SUBWAY AUTOMATIC TICKETING BASED ON NEURAL NETWORK AND DEEP LEARNING

A. *Face Recognition Technology based on Neural Network and Deep Learning*

Due to the impact of the epidemic, people are required to wear masks when choosing public transportation, which puts forward higher requirements for the face recognition technology of the subway ticketing system. The face recognition system needs to achieve high accuracy of face recognition when ticket buyers wear masks and other facial masks. In the existing CNN face recognition systems, the extraction of individual facial sign features can improve the accuracy of face recognition and better classify features [25-28]. CNN is a feedforward neural network model, and convolutional neural networks can be used to detect targets in face recognition technology. CNN can repeatedly cut the target image through different sliding windows and then perform feature extraction and recognition. The CNN network layer includes three network layers: convolutional layer, downsampling layer, and fully connected layer. The weight matrix *W* and bias term *b* parameters in the convolutional layer need to be trained and learned, and formula (1) is the calculation formula of convolution.

$$y_{mm} = f\left(\sum_{j=0}^{J-1}\sum_{i=0}^{I-1} x_{m+i,n+j} W_{ij} + b\right) \quad (1)$$

i in formula (1) represents the *ith* parameter, *jth* parameter is represented by *j*, and the samples' total number is represented by *m*. The subsampling layer can use a nonlinear downsampling method to reduce the amount of input data, thereby the feature map's scale is reduced without affecting the feature extraction of the target. If the size of uniform sampling is $S_1 \times S_2$, then the convolution formula (2) can be obtained.

$$y_{mm} = \frac{1}{S_1 S_2}\sum_{j=0}^{S_2-1}\sum_{i=0}^{S_1-1} x_{m\times S_1+i,\, n\times S_2+j} \quad (2)$$

As a CNN classifier, the fully connected layer can be implemented by convolution operations in practical applications, and is widely used in deep learning network models, but there is also the problem of parameter redundancy. When performing multi-classification, the SoftMax multi-classification layer is added to the convolutional neural network. Assume that the training set sample $T = \{(x^{(1)}, y^{(1)}),...,(x^{(m)}, y^{(m)})\}$, which $x^{(i)}$ represents the sample *i*, $y^{(i)}$ is the label of the sample. For the training set *T*, the SoftMax classifier can determine the category of the samples in it. In the case of rule restrictions, the probability $P(y=j|x),(j=1,....k)$ of converting to the *kth* sample is calculated by the formula, and then the sample category is calculated for face recognition, and the calculation formula of SoftMax can be obtained, see formula (3).

$$f(x^{(i)}|\theta) \begin{bmatrix} p(y^{(i)}=1|x^{(i)},\theta) \\ p(y^{(i)}=2|x^{(i)},\theta) \\ \vdots \\ p(y^{(i)}=k|x^{(i)},\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (3)$$

$\theta = [\theta_1^T \theta_2^T ... \theta_k^T]$ representation learning parameters in Eq. (3). The network model is continuously trained in the data set *T*, and the optimal one is obtained through iterative calculation and fitting θ. Formula (4) is the loss function used in the training.

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^{k} e^{\theta_j^T x^{(i)}}}\right] \quad (4)$$

In the formula (3), $1\{y=j\}$ means that its value is 1 in case of $y=j$, and 0 in case of $y \neq j$. By calculating network prediction and real value's error, the loss function's value can be obtained and the optimization process of the neural network can make the loss function minimize. The direction of signal transmission is forward, and the propagation direction of the error is opposite to it, which are BP's characteristics. The parameters' reverse adjustment in the network layer can be achieved by BP. Fig. 1 shows the BP network's structure.

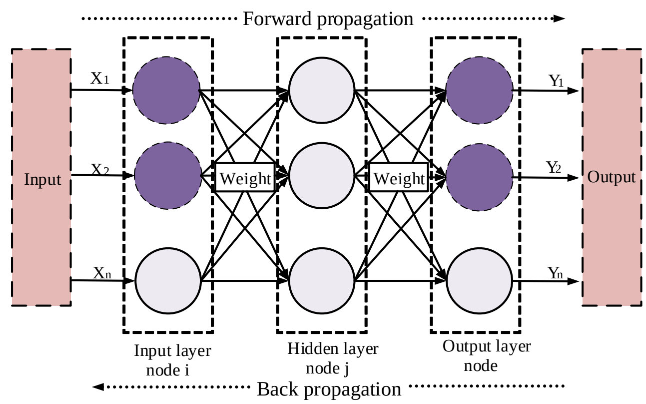


Fig. 1. Structure diagram of BP neural network

Through the continuous adjustment of the network weights in the BP neural network, the final network output close to the expected output can be obtained, that is, the error between the network prediction result and the given label is continuously reduced. Assuming that there are a total of *n* samples with a given label in the training set $\{(x_1, y_1),(x_2, y_2),....,(x_n, y_n)\}$, the neural network's training method is gradient descent. The variance loss function of the model is shown in formula (5) for the samples (x_i, y_i) in the training set.

$$L(w,b;x_i,y_i) = \frac{1}{2}\left\| h_{w,b}(x_i) - y_i \right\|^2 \quad (5)$$

In formula (5), *w* represents the weight and *b* represents the bias. The variance loss function can calculate the error, which can be used to evaluate the convergence situation. The loss function of *n* samples can be derived by formula (6).

$$\begin{aligned} L(w,b) &= \left[\frac{1}{n}\sum_{i=1}^{n} L(w,b;x_i,y_i)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_l+1}(w_{ji}^{(l)})^2 \\ &= \left[\frac{1}{n}\sum_{i=1}^{n}\left\|h_{w,b}(x_i)-y_i\right\|^2\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_l+1}(w_{ji}^{(l)})^2 \end{aligned} \tag{6}$$

In formula (6), *i* represents the *ith* parameter, *j* represents the network layer unit of the previous layer, *l* represents the *lth* layer of the neural network, and *S* represents the variance. The parameters λ are introduced to adjust the loss value of the variance loss function and the proportion of the attenuation term. In the process of BP neural network training, it is necessary to continuously fine-tune the size of parameters *w* and *b*. The derivation formula of parameter *w is shown* in formulas (7) and (8).

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha\frac{\partial}{\partial w_{ij}^{(l)}}L(w,b) \tag{7}$$

In formula (7), α denotes the learning rate, and the step size of gradient descent can be controlled by adjusting the learning rate. The derivation formula of parameter *b* is shown in formula (8).

$$b_i^{(l)} = w_{ij}^{(l)} - \alpha\frac{\partial}{\partial b_i^{(l)}}L(w,b) \tag{8}$$

Through formulas (7) and (8), the partial derivatives of all sample parameters *w* and *b* can be deduced, and the formula for partial derivatives of all sample parameters *w* is shown in formula (9).

$$\frac{\partial}{\partial w_{ij}^{(l)}}L(w,b) = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial w_{ij}^{(l)}}L(w,b;x_i,y_i)\right] + \lambda w_{ij}^{(l)} \tag{9}$$

The partial derivative formula of all sample parameters *b* is shown in formula (10).

$$\frac{\partial}{\partial b_{ij}^{(l)}}L(w,b) = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial b_{ij}^{(l)}}L(w,b;x_i,y_i)\right] \tag{10}$$

In the BP neural network, according to the input sample signal of forward propagation, the two parameters of weight and bias are randomly initialized, and the loss value is calculated. The partial derivatives of the optimization function to the neuron weights in the neural network of each layer can be obtained in the process of backpropagation, to obtain the error signals of all neurons, as the basis for modifying the parameters *w* and *b*. Through the construction of a multi-layer neural network, the basic neural network can be continuously developed to form deep learning. In the field of image recognition, underlying features of the image can be detected by deep learning in the low-level neural network layer, then integrate and abstractly combine to generate the corresponding high-dimensional image features, and finally realize the classification of high-dimensional features at the top layer. When the deep neural network propagates forward, the corrected weight value and bias value of the neural network can be obtained by calculating the error, and then the weight and bias item are corrected by backpropagation.

B. Face Recognition Technology for Subway Automatic Ticketing based on Neural Network and Deep Learning

The application of face recognition technology in many scenarios is limited by many conditions, and the clarity of photos is one of the limiting factors. In the actual subway automatic ticketing face recognition scene, the face recognition technology will reduce the accuracy of face recognition due to the influence of masks, glasses, hats, and other obstructions, expressions, light, and other factors. To make the robustness of the face recognition technology for subway ticket vending improve, CNN was optimized in this experiment. A branch structure was added to the structure of SphereFace-36 to get the face's local features. Fig. 2 shows the network structure after the fusion of local features and global features. The resolution of facial details can be improved after local and global features are fused.

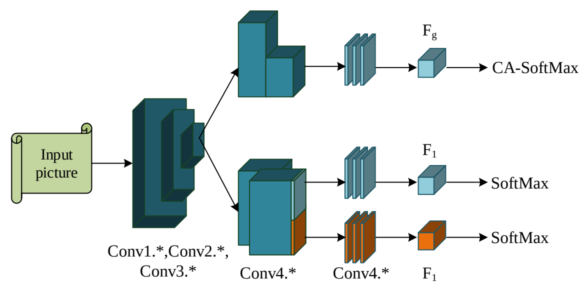


Fig. 2. Structure diagram of a network

One branch in the network structure diagram is the SphereFace-36, which can get the face's global features, and the other branch divides the feature image into upper and lower parts, which can get the face detailed features. The loss functions applied in these two branch network diagrams are CA-SoftMax (Center Angular SoftMax Loss) and SoftMax Loss respectively. In the original SoftMax Loss, it needs to be satisfied $w_1^T x > w_2^T x$, ie $\|w_1\|\|x\|\cos(\theta_1) > \|w_2\|\|x\|\cos(\theta_2)$. To achieve the effect of increasing the distance between classes, the variable *m* was added in this experiment to strictly limit the decision boundary of the classification, see formula (11).

$$\|w_1\|\|x\|\cos(m\theta_1) > \|w_2\|\|x\|\cos(\theta_2) \tag{11}$$

In the range of [0, π], since the *cos* function decreases monotonically, the larger the value of variable *m* is. Under the condition that and *x* are unchanged, when the classification is correct, the smaller the angle of θ_1, the smaller the angle between the sample feature vector and w_1, increase the distance between classes. The separability of the feature angle is a key part of Angular SoftMax Loss, and weight normalization can improve the separability of samples. After the parameter *w* is normalized by the weight, the value is 1, that is $\|w_i\| = 1$, the bias term *b* is 0, and formula (12) is the loss function.

$$L = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|x_i\|\phi(\theta_{yi})}}{e^{\|x_i\|\phi(\theta_{yi})} + \sum_{j\neq y_i} e^{\|x_j\|\cos(\theta_j)}}\right) \quad (12)$$

$$\phi(\theta_{yi}) = (-1)^k \cos(m\theta_{yi}) - 2k, \theta_{yi} \in \left[\frac{k}{m}\pi, \frac{(k+1)}{m}\pi\right], k \in [0, m-1], m \geq 1$$

In Angular SoftMax Loss, the expression effect of facial features can be improved, but the convergence speed of the function is slow during training, and it is easy to fall into a local minimum. The experiment updated the weights in Angular SoftMax Loss, that is, the sample center, named CA-SoftMax Loss. In formula (13), at time t+1, the expected weight value is the mean value of all sample features and their weight values that conform to $y_i = j$ at time t.

$$w_j^{t+1} = \frac{w_j^t + \sum_{i=1}^m \delta(y_i = j)\cdot(x_i)}{1 + \sum_{i=1}^m \delta(y_i = j)} \quad (13)$$

The mean value of Euclidean distance is approximately equal to the mean value of angular distance after weight normalization. The updated value of the sample center can be obtained by formulas (14) and (15).

$$w_j^{t+1} - w_j^t = \frac{\sum_{i=1}^m \delta(y_i = j)\cdot(x_i) - \delta(y_i = j)\cdot w_j^t}{1+\delta(y_i = j)} = \frac{\sum_{i=1}^m \delta(y_i = j)\cdot(x_i - w_j^t)}{1+\delta(y_i = j)} \quad (14)$$

$$\Delta w_j = w_j^{t+1} - w_j^t = \frac{\sum_{i=1}^m \delta(y_i = j)\cdot(w_j^t - x_i)}{1+\delta(y_i = j)} \quad (15)$$

To control the step size of the sample center update, a parameter is introduced during the update α, and it is set to 0.5, which achieves the purpose of preventing the jitter of the class center, see formula (16).

$$w_j^{t+1^*} = w_j^t - \alpha\Delta w_j \quad (16)$$

Based on neural network and deep learning, the face recognition system of subway automatic ticketing was designed and developed in the experiment. The process flow of the face recognition system is shown in Fig. 3. The overall framework of the system includes four modules: video image acquisition, face detection, normalization, and face recognition. The system needs to perform face registration first, and then extract the face features, and then save the acquired information and features into the database. When performing face recognition, it should extract the face information and features in the video and make it compared with the database's information and features to determine whether they belong to the same person.

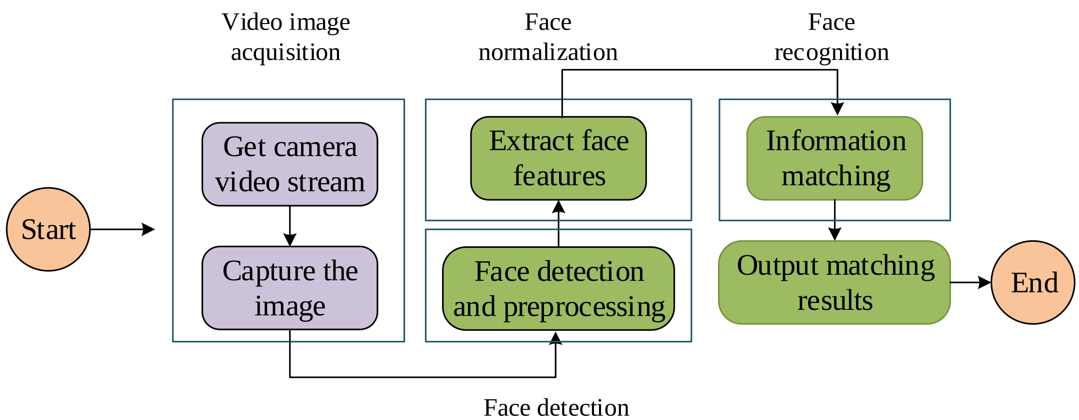
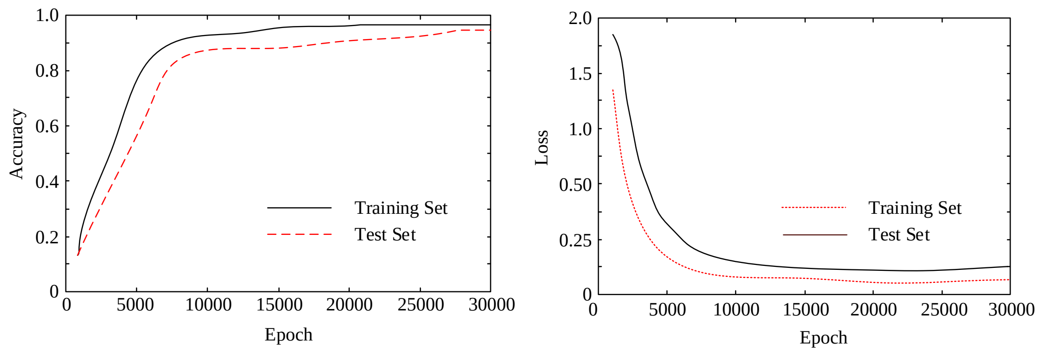


Fig. 3. Flow chart of face recognition

IV. Simulation Analysis of Face Recognition System For Subway Automatic Ticketing

To improve the classification performance of the improved CNN, the improved method is trained in the BLUFR general data set, and the training is completed in the training set and the test set respectively. It can be seen from Fig. 4 that the accuracy rate of the optimized CNN model tends to be stable when it is trained to about 8000 times. The final accuracy rate in the training set is 98.5%, and the final accuracy rate in the test set is 97.6%. The loss function value of the optimized CNN model tends to be stable when it is trained to about 5000 times. The loss function value in the training set is 0.09, and the loss function value in the test set is 0.18. It shows that the optimized CNN model can get higher accuracy and lower loss function value after training, and the algorithm has better performance.

In the process of neural network training, each weight value needs to be assigned in advance, that is, weight initialization. The method of weight assignment is very important to the convergence speed of the neural network model and the performance of the model. In this study, the original Angular SoftMax Loss has been updated with weights, and the form of the original function has not been changed. The weights can be updated faster and more efficiently after the weights are updated. In the data set WebFace, the loss function in the face recognition system is compared, and SphereFace-36 is selected as the test neural network, the parameter settings are consistent, and the initial learning rate is 0.01. The difference between the optimized CA-SoftMax Loss and the original loss function is shown in Fig. 5.



(a) Accuracy (b) Loss

Fig. 4. Performance verification diagram of improved CNN model

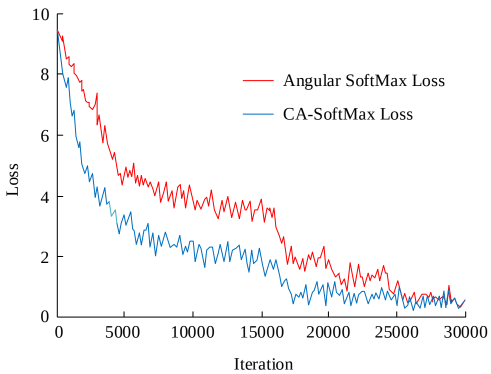
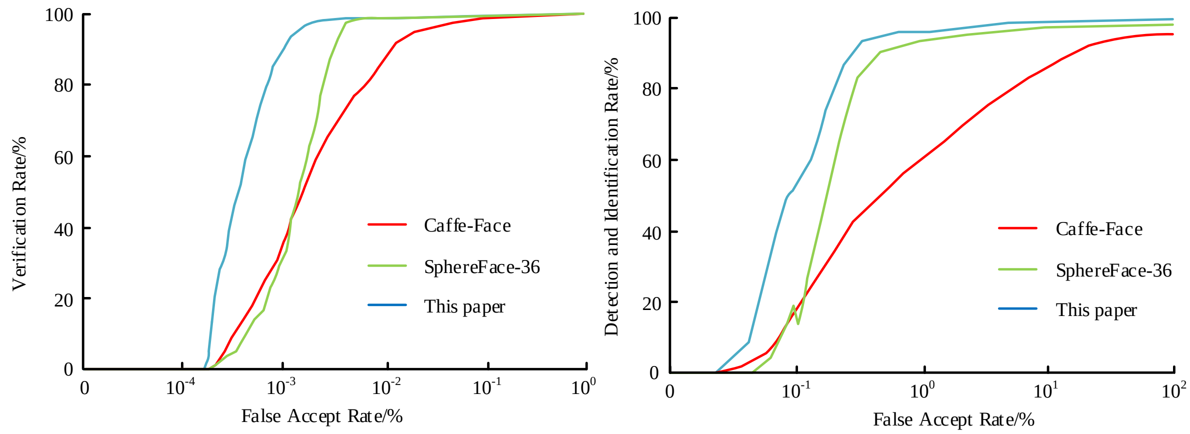


Fig. 5. Loss function curve

From Fig. 5, the initial Loss value of the optimized CA-SoftMax Loss and Angular SoftMax Loss is about 9.2. With the iterations number and pictures number increased, the Loss value decreases rapidly. The optimized CA-SoftMax Loss converges faster than the original Angular SoftMax Loss loss function, and the Loss value does not change or increase abnormally due to network gradient dissipation or gradient explosion during the descent process. In the experiment, we chose to test the model accuracy in the BLUFR general data set, LFW general data set, AgeDB-30 general data set, CFP-FP general data set and YTF general data set, and compared SphereFace-36, Caffe-Face, and optimization. The verification rate and recognition rate of the post-branch network structure are shown in Fig. 6.

Fig. 6 (a) and Fig. 6 (b) are the validation ROC curve and the identification ROC curve of the validation model, respectively. The area under the optimized branch network structure verification ROC curve and identification ROC curve is higher than that of the SphereFace-36 and Caffe-Face models, indicating that the optimized network structure proposed in this experiment has a better Good face recognition effect. In BLUFR general dataset, LFW general dataset, AgeDB-30 general dataset, CFP-FP general dataset and YTF general dataset, SphereFace-36, Center Loss, Angular SoftMax Loss, Caffe-Face. In this experiment, the optimization model is used to compare the algorithm recognition performance, and the results are shown in Table I.

The "-" in Table I indicates that the algorithm has not been tested for recognition performance in the corresponding dataset. It can be seen from the results in different datasets that the accuracy of the basic SphereFace-36 algorithm is lower than that of the optimized network model. In the BLUFR dataset, LFW dataset, AgeDB-30 dataset, CFP-FP dataset, and YTF dataset, the recognition accuracy of the optimized model is improved by 1.0%, 0.7%, 1.1%, 0.9%, and 0.6%, respectively. The evaluation indicators of the performance comparison in the experiment selected the true rate, false positive rate, prediction accuracy rate, and average accuracy rate, respectively in the BLUFR general data set, LFW general data set, AgeDB-30 general data set, CFP-FP general data set Compared with the YTF general data set. Table II shows the test results of SphereFace-36 and the optimized model in this experiment.



(a) Verification of ROC curve

(b) Detection and Identification of ROC curve

Fig. 6. ROC curve

TABLE I. ALGORITHM PERFORMANCE CONTRAST

Method	BLUFR	LFW	AgeDB	CFP	YTF
SphereFace-36	98.70%	93.25%	94.33%	92.01%	95.17%
Center Loss	98.31%	93.01%	-	92.57%	93.62%
Angular SoftMax Loss	98.12%	-	93.58%	-	94.83%
Caffe-Face	99.13%	92.87%	94.27%	92.33%	95.10%
This paper	99.70%	93.95%	95.43%	93.11%	95.77%

TABLE II. INDEX COMPARISON RESULTS

Data set	Method	True positive (%)	False positive (%)	Prediction accuracy (%)	Average accuracy (%)
BLUFR	SphereFace-36	84.84	9.79	90.59	87.87
	This paper	93.93	4.04	96.85	94.94
LFW	SphereFace-36	83.14	9.60	88.78	86.11
	This paper	92.05	3.95	94.92	93.04
AgeDB-3 0	SphereFace-36	78.98	9.12	84.34	81.80
	This paper	87.44	3.76	90.17	88.38
CFP-FP	SphereFace-36	86.88	10.03	92.78	89.98
	This paper	96.19	4.13	99.19	97.22
YTF	SphereFace-36	84.84	9.79	90.59	87.87
	This paper	93.93	4.04	96.85	94.94

From Table II, in BLUFR, LFW, AgeDB-30, CFP-FP, and YTF general data set, the true rate of optimizing the branch network structure is 93.93%, 92.05%, 87.44%, 96.19%, 93.93%. In the BLUFR general dataset, the LFW general dataset, the AgeDB-30 general dataset, the CFP-FP general dataset, and the YTF general dataset, the false positive rates of the optimized branch network structure are 4.04%, 3.95%, 3.76%, and 4.13%, 4.04%, respectively. In BLUFR, LFW, AgeDB-30, CFP-FP, and YTF general data set, the prediction accuracy of optimized branch network structure is 96.85%, 94.92%, 90.17%, 99.19%, 96.85%, respectively. In the BLUFR general dataset, the LFW general dataset, the AgeDB-30 general dataset, the CFP-FP general dataset, and the YTF general dataset, the average accuracies of optimizing the branch network structure are 94.94%, 93.04%, 88.38%, 97.22%, and 94.94%, respectively. In each general dataset, the true rate, prediction accuracy, and optimized network model's average accuracy are higher than those of the basic SphereFace-36; in each general dataset, the false positive rate of the optimized branch network is lower than that of the basic SphereFace-36. In Fig. 7, the sensitivity and specificity of the basic SphereFace-36 and optimized network models are tested on the BLUFR general dataset, the LFW general dataset, the AgeDB-30 general dataset, the CFP-FP general dataset, and the YTF general dataset.

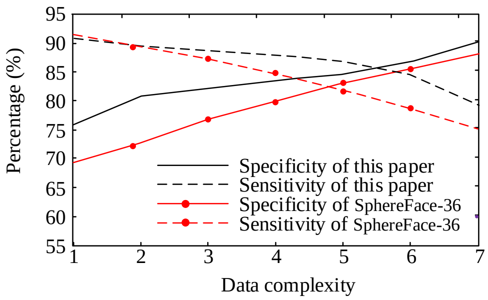


Fig. 7. Comparison results of sensitivity and specificity

Fig. 7 describes the comparison results of specificity and sensitivity between the basic SphereFace-36 and the optimized network model. In terms of specificity, the specificity of the optimized network model is higher than that of the basic SphereFace-36, and the difference can be up to 9 %. When the complexity of the data is low, the specificity of the optimized network model is significantly higher than that of the basic SphereFace-36; as the complexity of the data increases, the specificity gap between the basic SphereFace-36 and the optimized network model gradually narrows. In the sensitivity comparison of the methods, when the data complexity is low, the difference between the basic SphereFace-36 and the optimized network model is not significant, but as the data complexity increases, the optimized network model's sensitivity declines significantly less than the basic one. Fig. 8 shows the global feature recognition and extraction effects of different methods. Two sample images are chosen to optimize network model and the basic SphereFace-36 global feature's performance comparison.

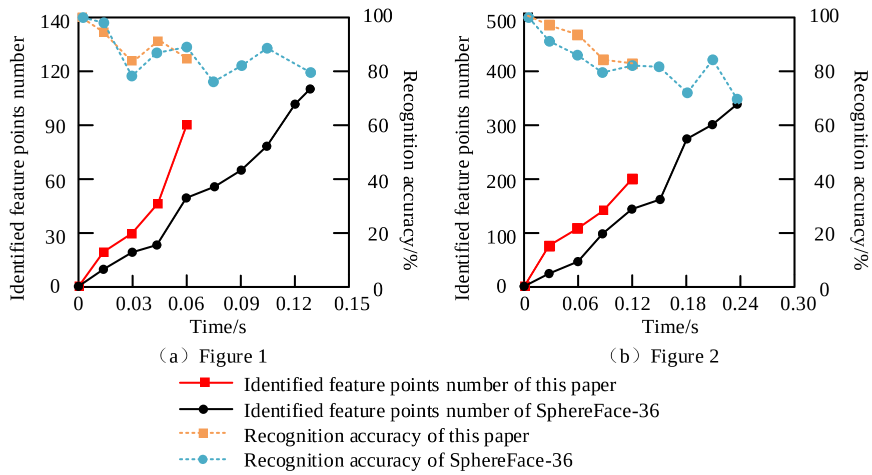


Fig. 8. Test results of global feature recognition and extraction

From Fig. 8, in the global feature extraction and recognition test of the two sample images, the optimized network model proposed in this experiment recognizes the number of feature points and the accuracy rate is higher than the basic SphereFace-36. Among them, the average accuracy rate of the optimized method is 83.01%, and basic SphereFace-36's average accuracy rate is 67.98%. The above results can show that the optimized network model has higher performance in global feature extraction and recognition of face images performance.

V. Conclusion

In the results of this research on face recognition technology, the area under the ROC curve for verification and recognition of the optimized model is the highest; The recognition accuracy of the optimized model in different data sets is 1.0%, 0.7%, 1.1%, 0.9% and 0.6% higher than that of SphereFace-36 respectively, and its specificity is higher than that of SphereFace-36, with the maximum difference of 9%. The average accuracy of global feature extraction and recognition of the optimized network model is 83.01%. The results show that the optimized method can accurately recognize facial features, and has high practical value, which can be applied to the automatic ticketing system. In a word, the improved method has good performance and plays a positive role in the establishment of face recognition system. Although the face recognition technology in this study has a good recognition effect, the running time of the algorithm is relatively long in the experimental process, which affects the running speed of the algorithm. In the later research, it is necessary to further improve the optimization of parameters. In addition, in actual application scenarios, face recognition may be affected by environmental factors, resulting in the inability to recognize low resolution faces. Therefore, in subsequent research, specific analysis of problems in the actual scene is required to improve the performance of face recognition technology.

References

[1] V. Sharma, A. M. Joshi, "VLSI Implementation of Reliable and Secure Face Recognition System," Wireless Personal Communications, vol. 122, no. 4, pp. 3485 - 3497, 2021.

[2] M. Kumar, S. Gupta, "2D - human face recognition using SIFT and SURF descriptors of face's feature regions." The Visual Computer, vol. 37, no 11, pp. 447 - 456, 2021.

[3] Y. Kortli, M. Jridi, A.A. Falou, M. Atri. "Face Recognition Systems: A Surve." Sensors (Basel, Switzerland), vol. 20, no. 2, pp. 1 - 34, 2020.

[4] F. Zhang, T. Zhang, Q. Mao, C. Xu. "A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment."

IEEE Transactions on Image Processing, vol. 29, no. 1, pp. 6574 - 6589, 2020.

[5] S. F. Baltanas, J. R. Ruizsarmiento, J. Gonzalezjimenez, "Improving the Head Pose Variation Problem in Face Recognition for Mobile Robots." Sensors, vol. 21, no. 2, pp. 659 - 676, 2021.

[6] M. Sadiq, D. Shi, J. Liang, "A robust occlusion - adaptive attention - based deep network for facial landmark detection." Applied Intelligence, vol. 52, no. 8, pp. 9320 - 9333, 2022.

[7] B. Chandra, S. Sastry, C.Jana, S. Robust, "Heartbeat Detection From Multimodal Data via CNN - Based Generalizable Information Fusion." IEEE Transactions on Biomedical Engineering, vol. 66, no. 3, pp. 710 - 717, 2019.

[8] H. Zhong, W. Yimu, Q. Xiaodong, Y. Rui, Q. Yong, H. Kelei, Z. Zezhang. "Classification of neurofibromatosis-related dystrophic or nondystrophic scoliosis based on image features using Bilateral CNN." Medical Physics, vol. 48, no. 4, pp. 1571 - 1583, 2021.

[9] O. Elharrouss, N. Almaadeed, S. Al - Maadeed, F. Khelifi. "Pose - invariant face recognition with multitask cascade networks." Neural Computing and Applications, vol. 34, no. 8, pp. 6039 - 6052, 2022.

[10] G. Lou, H. Shi, "Face image recognition based on convolutional neural network." China Communications, vol. 17, no. 2, pp. 117 - 124, 2020.

[11] R. Silwal, A. Alsadoon, P. Prasad, O.H. Alsadoon, A. Al-Qaraghuli. "A novel deep learning system for facial feature extraction by fusing CNN and MB-LBP and using enhanced loss function." Multimedia Tools and Applications, vol. 79. no. 41/42, pp. 31027-31047, 2020.

[12] M. M. Hussein, A. H. Mutlag, H. Shareef, "An Improved Artificial Neural Network Design for Face Recognition utilizing Harmony Search Algorithm." IOP Conference Series Materials Science and Engineering, vol. 745, pp. 012038, 2020.

[13] J. O. Agushaka, A. E. Ezugwu, L. Abualigah, "Dwarf Mongoose Optimization Algorithm." Computer Methods in Applied Mechanics and Engineering, vol. 391. no. 1, pp. 1-38, 2022.

[14] L. Abualigah, A. Diabat, S. Mirjalili, et al. "The Arithmetic Optimization Algorithm." Computer Methods in Applied Mechanics and Engineering, vol. 376. no.1, pp. 1-38, 2021.

[15] K. H. Almotairi, L. Abualigah, "Improved reptile search algorithm with novel mean transition mechanism for constrained industrial engineering problems." Neural Computing and Applications, vol. 34. no.20, pp.17257-17277, 2022.

[16] A.S. Bist, W. Febriani, C. Lukita, S. Kosasi, U. "Rahardja. Design of Face Recognition AttendX for Recording Student Attendance Data Based on Artificial Intelligence Technology." Solid State Technology, vol. 63, no. 2s, pp. 4505 - 4518, 2020.

[17] Y.H. Pu, P.S. Chiu, Y.S. Tsai, M.T. Liu, Y.Z. Hsieh, S.S. Lin. "Aerial face recognition and absolute distance estimation using drone and deep learning." The Journal of Supercomputing, vol. 78, no. 4, pp. 5285 - 5305, 2022.

[18] B. Rossion, A. Lochy, "Is human face recognition lateralized to the right hemisphere due to neural competition with left - lateralized visual word recognition?" A critical review. Brain Structure and Function, vol. 227, no. 2, pp. 599 - 629, 2021.

[19] S. Nejrs, M. S. Al – An, "Face Image Classification Based on Feature Extraction. Solid State Technology," vol. 63. no. 6, pp. 13515 – 13526, 2020.

[20] M.R.M. Alsemawi, M.H. Mutar, E.H. Ahmed, H.O. Hanoosh, F.T. Al - Dhief. "Emotions Recognition from Human Facial Images Using Machine Learning Technique." Solid State Technology, vol. 63, no. 5, pp. 8749 - 8761, 2020.

[21] H. B. Fredj, S. Bouguezzi, C. Souani, "Face recognition in unconstrained environment with CNN." The Visual Computer, vol. 37, no. 2, pp. 217 - 226, 2021.

[22] Z. Sun, L. Tian, Q. Du, et al. "Feature learning via multi - action forms supervising force for face recognition." Neural Computing and Applications, vol. 34, no. 6, pp. 4425 - 4436, 2022.

[23] C. Chen, X. Zhou, "Collaborative representation - based fuzzy discriminant analysis for Face recognition." The Visual Computer, vol. 38, no. 4, pp. 1383 - 1393, 2022.

[24] J. Gan, K. Jiang, H. Tan, G. He. "Facial Beauty Prediction Based on Lighted Deep Convolution Neural Network with Feature Extraction Strengthened." Chinese Journal of Electronics, vol. 29, no. 2, pp. 312 - 321, 2020.

[25] C. Han, S. Shan, M. Kan, S. Wu, X. Chen. "Personalized Convolution for Face Recognition." International Journal of Computer Vision, vol. 130, no. 2, pp. 344 - 362, 2022.

[26] Rizwan S A, Ghadi Y Y, Jalal A, Kim K. "Automated Facial Expression Recognition and Age Estimation Using Deep Learning." Computers, Materials, & Continua, vol. 6, pp. 5235-5252, 2022.

[27] Chen L, Li M, Lai X, Hirota K Pedrycz W. "CNN-based Broad Learning with Efficient Incremental Reconstruction Model for Facial Emotion Recognition." IFAC-PapersOnLine, vol. 53. no.2, pp.10236 - 10241, 2020.

[28] Mao L, Yan Y, Xue J H, Wang H. "Deep Multi-Task Multi-Label CNN for Effective Facial Attribute Classification." IEEE transactions on affective computing, 2022, vol. 13. no.2, pp. 818-828.

Research on Improved Xgboost Algorithm for Big Data Analysis of e-Commerce Customer Churn

Li Li
Department of Economics & Management, Sichuan TOP IT Vocational Institute, Chengdu, China

***Abstract*—With the increasing cost of acquiring new users for e-commerce enterprises, it has become an important task for e-commerce enterprises to actively carry out customer churn management. Therefore, based on the distributed gradient enhancement library algorithm (XGBoost), this research proposes a big data analysis study on e-commerce customer churn. First, it conducts an evaluation analysis on e-commerce customer segmentation and combines the random forest algorithm (RF) to build an RF XGBoost prediction model based on customer churn. Finally, it verifies the performance of the prediction model. The results show that the area under receiver operating characteristic curve (AUC) value, prediction accuracy, recall rate, and F1 value of the RF-XGBoost model are significantly better than those of the RF, XGBoost, and ID3 decision trees to build an e-commerce customer churn prediction model; The average output error of RF-XGBoost model is 0.42, and the average output error is relatively good, indicating that the model proposed in this study has a smaller error and higher accuracy. It can make a general assessment of the customer churn of e-commerce enterprises, and then provide data support for the customer maintenance work of e-commerce enterprises. It is helpful to analyze the relevant factors affecting customer churn, to Equationte targeted customer service programs, thus improving the economic benefits of e-commerce enterprises.**

Keywords*—*E-commerce; customer churn; random Forest; XGBoost; big data

I. Introduction

At present, domestic e-commerce enterprises and websites are also gradually increasing, and people are getting used to buying goods and services online [1]. Most traditional industries have built their e-commerce platforms, and the opportunities for customers to choose products are gradually increasing [2]. For e-commerce companies, they must face fierce competition for customer resources from other companies, and companies that fail to compete will face elimination from the market. Therefore, finding a suiTABLE method to more accurately predict the purchase behavior and churn of users, and taking effective measures to reduce the loss rate has become a hot topic in related industries. In the absence of new customer resources, some e-commerce companies began to focus on old customers, and used data mining-related technologies to analyze the shopping behavior of old customers to mine useful data. By formulating marketing plans based on these data, enterprise can avoid the loss of old customers and increase more economic benefits [3]. Some scholars have pointed out that enterprises will accumulate massive amounts of data in the daily operation process, from which valuable information can be mined and customers who are about to be lost can be accurately predicted [4]. For e-commerce companies, when they find customers who are about to be lost, taking corresponding retention measures can effectively reduce the loss rate of corporate customers and make customers active on the company's e-commerce platform [5]. Therefore, this research takes e-commerce enterprise customers as the research object, analyzes the relevant behavioral data recorded by the background when customers purchase online, subdivides them according to customer value, combines the distributed gradient enhancement library algorithm (XGBoost) with the random forest algorithm (RF), and constructs a big data analysis model for e-commerce customer loss.

II. Related Work

Janabi S and his team proposed an integrated system, in recent years, under the global economic downturn under the new crown epidemic; all walks of life are competing for customer resources. Therefore, the analysis of big data for enterprise customer churn has attracted people's attention. Janabi S and his team proposed an integrated system to help telecom companies achieve intelligent prediction of customer churn, use training data to build an intelligent predictor of customer churn, and use a genetic algorithm (GA) algorithm to group customers for decision-making. The integrated system is superior to the system built by traditional methods and has high accuracy [6]. Li W et al. used feature extraction technology to analyze lost customers and proposed a customer loss prediction method. The online test shows that the proposed customer churn prediction method can effectively predict customer churn, which is superior to the traditional prediction technology [7]. Gu Y et al. studied the effect of the decision tree algorithm in customer churn prediction. Each performance index of this algorithm is superior to other traditional algorithms. After applying this method, it can help enterprises retain old customers and avoid the loss of old customers [8]. Jiao G and his team predicted telecommunication customer churn and compared four prediction models on a telecommunication dataset, and found that the model built by XGBoost and RF algorithm has the best prediction ability among many algorithms [9]. Eria K's studio studied the relevant characteristics of customer data, and on this basis proposed a customer churn prediction method, that reduces the data according to missing values and irrelevant variables, which helps reduce churn analysis. The computational cost is incurred by teachers using big data in churn prediction and analysis [10].

Sharma T and other researchers propose an improved XGBoost algorithm as a model to address customer churn in the telecom industry. Compared with non churning customers, the previous model paid more attention to prediction accuracy, and the proposed model correctly classified all churning customers, having better performance [11]. Swetha P et al. designed a data characteristic function model to judge and predict the loss rate of enterprise customers. In the process of research, the loss function is constructed and combined with the XG_ Boost method. The proposed model is helpful to identify correct and wrong classification examples of South Asia's global mobile communication system service providers [12]. Scholars such as Naser AM built a customer churn prediction model to determine whether each customer is a churned customer and get more opinions about serving customers and use the extreme gradient to boost the classification model of "XGBoost", which is added to the original network by computing A centrality measure for the new attribute of the dataset to evaluate model performance. Experiments show that the proposed clustering-based churn detection method combining social influence and web content significantly improves the prediction accuracy of telecom datasets [13].

Although a large number of studies have focused on enterprise customer churn, there are still few studies on e-commerce enterprise customer churn. And the prediction accuracy needs to be improved, and the application of the XGBoost loss prediction model is not flexible enough. Therefore, based on the improved XGBoost algorithm, this study proposes a prediction method for e-commerce customer churn, to contribute to the development of e-commerce enterprises.

III. E-Commerce Customer Segmentation Evaluation and Customer Churn Prediction Model Construction

A. *E-commerce Customer Value Segmentation Evaluation*

The "28" law proposed by the Italian economist Pareto shows that in any group of things, the most important part is only a small part, accounting for about 20%, and the remaining 80%, although the majority, is secondary. Its core content is that 80% of the results in life almost all come from 20% of activities. According to the famous "28" law, nearly 80% of the benefits of an enterprise are brought by 20% of the main customers [14]. Therefore, in the current situation that resources are getting smaller and smaller, e-commerce companies need to further subdivide customers, and at the same time customize different service plans for different customers, to avoid the loss of customers. Based on the traditional customer behavior analysis model (RFM), this study analyzes the customer value evaluation indicators of e-commerce companies [15]. The period indicator (T) of the first online activity and the last online activity in the observation period is introduced through the RFM model, and the RFM model (RFMT) based on the time factor is constructed. It mainly includes four key indicators in the observation period, namely the time interval (R) between the last time the customer had consumption behavior and the observation point of analysis, the total consumption times (F), the total consumption amount (M) and T .

The customer segmentation process is shown in Fig. 1. The first step is to build an index system, mainly using expert scoring to evaluate four key indicators; the second step is to evaluate customer value (V), by calculating the normalized index weight value used to obtain the corresponding V value; the third step is to classify the customer groups. The V value is analyzed through K-means, customer types are divided, and the indicators of each type of center point are analyzed to get the category labels of various customers, and set the value threshold, so that the enterprise can further identify the customers with special value.

According to the actual situation of the research, and according to the method in the literature, a single-layer indicator system for e-commerce customer segmentation is proposed, the target layer is "customer value evaluation", and the indicator layer is four key indicators [16]. Then sort the four key indicators proposed, calculate the weight of each indicator, and realize the calculation of customer value V .

According to the single-level indicator system of e-commerce customer segmentation, the four key indicators are compared with each other in pairs and graded. From the comparison results, the feature discriminant matrix can be constructed $X = (x_{ij})_{n \times n}$, as shown in Eq. (1).

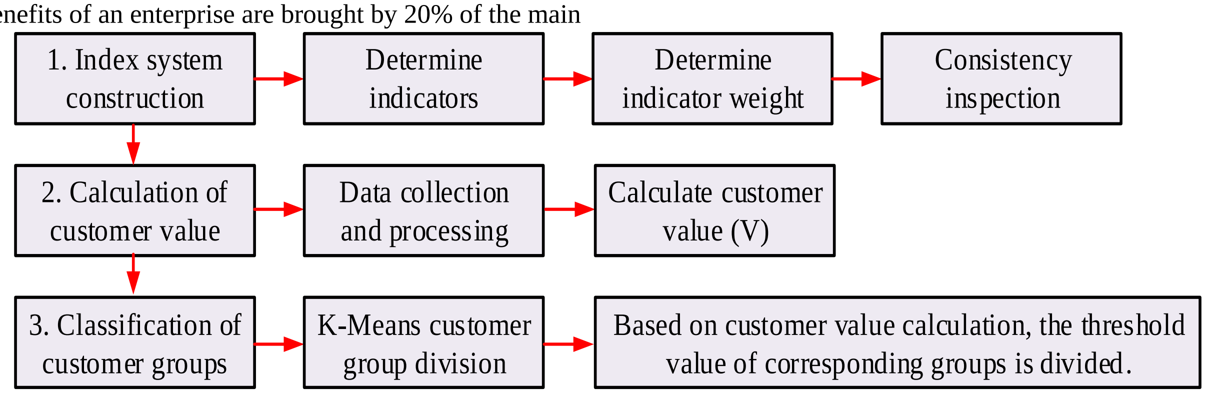


Fig. 1. Customer segmentation model framework

$$X = \begin{vmatrix} x_{ij} & X_1 & X_2 & \dots & X_n \\ X_1 & x_{11} & x_{12} & \dots & x_{1n} \\ X_2 & x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ X_n & x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix} \quad (1)$$

It can be seen from Eq. (1) that x_{ij} the contrast between the $x_{ij} = 1/x_{ji}$ representation feature i and the feature exists at the same time j. According to the characteristics of the matrix in Eq. (1), to obtain the weight of each index, it is necessary to use the expert scoring method to score X the eigenvectors. Then, the consistency test is performed by the largest eigenroot $\lambda_{\max}$ pair X, as shown in Eq. (2).

$$CI = (\lambda_{\max} - n)/(n-1) \quad (2)$$

To be able to better evaluate X the consistency, this study introduces an index X related to RI. Use the CR value to judge X whether it passes the test, if $CR > 0.1$ it fails the test, otherwise it passes the test, see Eq. (3).

$$CR = CI / RI \ (2) \quad (3)$$

In the classification of customer groups, the scores of each customer in the RFMT model are compared with the average of the indicators, and those higher and lower than the average are marked as "+" and "-" respectively. Based on this, customers are divided into 16 sub-categories, and then through expert analysis, the e-commerce customers are finally divided into four groups, namely core customers, important customers, important development customers, and general customers. After the customer is preprocessed, the customer's score on the four indicators will be obtained, and the weight value of each indicator will be obtained by using the expert scoring method. Finally, the customer's V value will be calculated by the Eq. (4). Among them, W represents the feature vector obtained by expert scoring.

$$V = W_R \times \frac{1}{R} + W_F \times F + W_M \times M + W_T \times T \quad (4)$$

B. Improved XGBoost Prediction Model

First of all, we need to measure the data characteristics of customers. This research is implemented through random forests (RF). When the RF algorithm randomly selects the training subset applied to the decision tree algorithm, the unselected data are out-of-bag data [17]. The error rate of the out-of-bag data evaluation model is named E1; then random interference terms are added to the out-of-bag data, x and the evaluation model error is $E2$. If $E2 > E1$, then the importance of feature x (I) is higher, and I, the higher the value, the more important the feature is. Assuming that the number of training decision trees generated by the RF algorithm is k, then I the calculation Eq. is shown in Eq. (5).

$$I = \frac{\sum(E1 - E2)}{k} \quad (5)$$

The introduction of the feature importance measure is convenient for the RF algorithm to select a better feature set [18]. In the process of screening important features through the RF algorithm, a tree-building process will be carried out at the same time. This process mainly includes four processes: one is I to calculate the value of the I features in the data set by Eq. (5). Sort by size, and output the I values of all features; the third is to select the previous x_i feature in the output to regroup into multiple feature sets, input each combined feature set into the RF model, and then calculate the out-of-bag error rate of each feature set, among them $i = 1, 2, 3...$, $x_1 < x_2 < x_3....$; Fourthly, all out-of-bag error rates are obtained by comparison, and the feature set whose out-of-bag error rate does not change continuously is selected as the optimal feature set, so as to facilitate subsequent empirical analysis and research.

The XGBoost algorithm belongs to a classic ensemble algorithm, which implements part of the generalized linear machine learning algorithm based on the boosted decision tree (GBDT) [19]. XGBoost model can use CPU multithreading to realize parallel operation and obtain the best parameters. See Eq. (6) for the calculation of its objective function.

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (6)$$

In Eq. (6), θ is the model parameter and $L(\theta)$ is the loss function. The higher the value, the higher the model accuracy. However, the model is easy to regard noise as a learning sample, which leads to overfitting. Overfitting refers to the phenomenon that the model predicts the known data well, but the prediction effect on the unknown data is poor, which will reduce the robustness of the model.

The regularization term ($\Omega(\theta)$) can improve the generalization ability of the model and the complexity of the evaluation model. Let the number of ensemble trees in the XGBoost model be k.

$$y_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \rho \quad (7)$$

In Eq. (7), f_k is a base classifier, K and ρ respectively represents the number and space of the base classifier. If y_i represents the class i mark, the calculation of the objective function is shown in Equation (8).

$$Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (8)$$

The XGBoost model is a boosted tree model with a special form, and its essence is an additive model. To solve the objective function of the model, it is necessary to obtain each tree f_k. However, all trees f_k cannot be obtained at one time, so it is necessary to train according to the model that has been iterated several $t-1$ times to obtain the first t tree. After iteration t times, the loss function of the model is expanded by second-order Taylor expansion, and then the constant term is removed, and a regular term is added to constrain the loss function, to control the complexity of the model. Based on the above content, t the objective function of the first iteration can be obtained, as shown in the Eq. (9).

$$Obj^{(t)} \approx \sum_{i=1}^{n}\left[g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right]+\Omega(f_t) \quad (9)$$

For the objective function, g_i and h_i are their first and second derivatives respectively. After minimizing the objective function, the model can achieve the best classification effect, thus improving the prediction accuracy [20]. The loss function when not optimized is drawn in Fig. 2. From the analysis of the curve gradients in the graph, it is found that the gradients of the first two types of curves are consistent, indicating that the losses of the two types of errors are regarded as the same, and the positive examples are misclassified more than the negative ones after optimization, which is consistent with the actual situation.

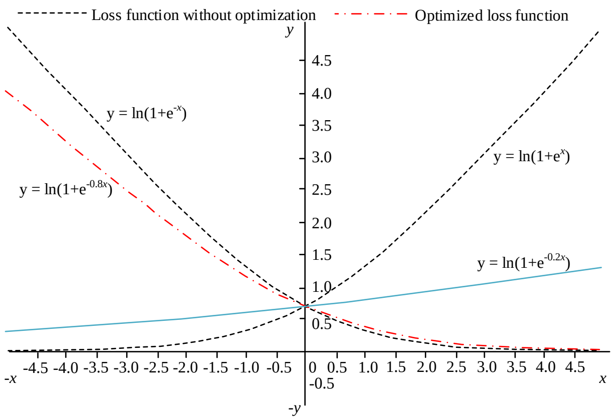


Fig. 2. Comparison diagram of loss function before and after optimization

Due to the serious loss of e-commerce customers, this study judges the loss of customers by studying the basic data and behavioral data of customers and uses the RF feature selection method combined with the XGBoost algorithm to build a prediction model, to better handle the relationship between e-commerce companies and customers. This research combines RF and XGBoost to build a prediction model, and its process is shown in Fig. 3. In classifier 1, if the customer is not lost, the output is 0; if the customer is lost, the output is 1. In classifier 2, the customers who have not lost the e-commerce company are predicted. If there is a possibility of losing customers, the output result is 1; otherwise, the output is 0.

Firstly, the model should be trained through the data test set, and each parameter in XGBoost should be tuned to obtain the optimal parameters of the algorithm, to make the model prediction effect in the later stage the best. In addition, each tree structure in the algorithm needs to be made simpler. For the detailed process of training the objective function, see Eq. (10).

$$\begin{cases} \hat{y}_l^{(0)} = 0 \\ \hat{y}_l^{(1)} = f_1(x_i) = \hat{y}_l^{(0)} + f_1(x_i) \\ \hat{y}_l^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_l^{(1)} + f_2(x_i) \\ \quad\quad \cdots\cdots \\ \hat{y}_l^{(t)} = \sum_{k=1}^{t} f_t(x_i) = \hat{y}_l^{(t-1)} + f_t(x_i) \end{cases} \quad (10)$$

The process of training the model and obtaining the optimal parameters proposed in this study mainly includes four steps: first, obtaining a new decision tree, second, deriving the loss function, third, obtaining the predictive value of the sample, and fourth, returning to the first step until the result of parameter optimization is optimal.

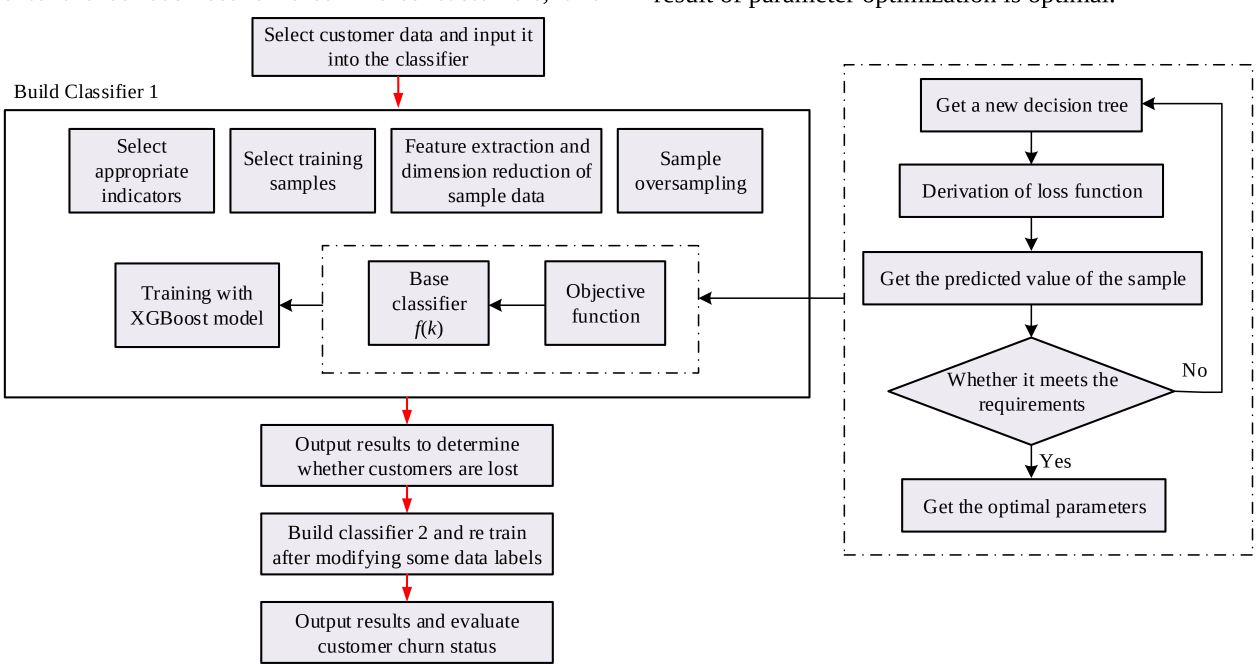


Fig. 3. Customer churn prediction model proposed in this study

IV. Model Performance Analysis

This study selects the data of an e-commerce platform from April 1, 2020, to April 1, 2022, and analyzes the performance of the model constructed in this study. Take the data from April 1, 2020, to February 19, 2022, as the training set, and the data from the rest of the dates as the test set. The data mainly includes the original sample skeleton data, advertising information attribute description, user information data, and user behavior log data. After the cleaning and conversion of the selected data with missing values and abnormal values, we selected 121949 effective users, a total of 4488519 user behavior data for customer churn prediction research. Based on the above customer segmentation results, Table I shows the churn rate of each type of customer.

The results in Table I show that corresponding to the characteristics of each type of customer after customer segmentation, the order of customer churn rate from high to low is: general customers, important development customers, important customers, and core customers. Moreover, the churn rate of general customers exceeds 90%, which is much higher than that of core customers, indicating that general customers may just consume on the platform by accident, and can be identified as temporary customers, which is very easy to lose. The core customer churn rate is the lowest, indicating that such customers have the highest loyalty to the company and stronger customer stickiness. The difference between the churn rate of important customers and core customers is 6.46%, and the difference between the churn rate of important development customers and important customers is 4.06%, indicating that the difference in churn rate between them is relatively stable, but the churn rate of general customers is more important The churn rate of development customers has increased by 43.47%, which makes managers see the fundamental difference between general customers and the other three types of customers, and reminds managers to give up properly and invest more energy in this part of customers.

Build two prediction models (Model A and Model B) based on RF-XGBoost. Among them, the training of model A is mainly through the pre-processed data; Model B uses the original data set for training. See Fig. 4 for the training iteration process. For the accuracy rate and loss curve, the curve of model A is relatively stable, while model B has large fluctuations. It shows that data preprocessing can improve the stability of the model and prediction accuracy.

TABLE I. CUSTOMER CHURN RATE OF CUSTOMER SEGMENTS

/	Customer volume	The number of customers lost	Loss rate
Core customers	23975	8983	37.47%
Key customers	24316	10683	43.93%
Important development customers	26812	12865	47.99%
General customers	46705	42723	91.46%

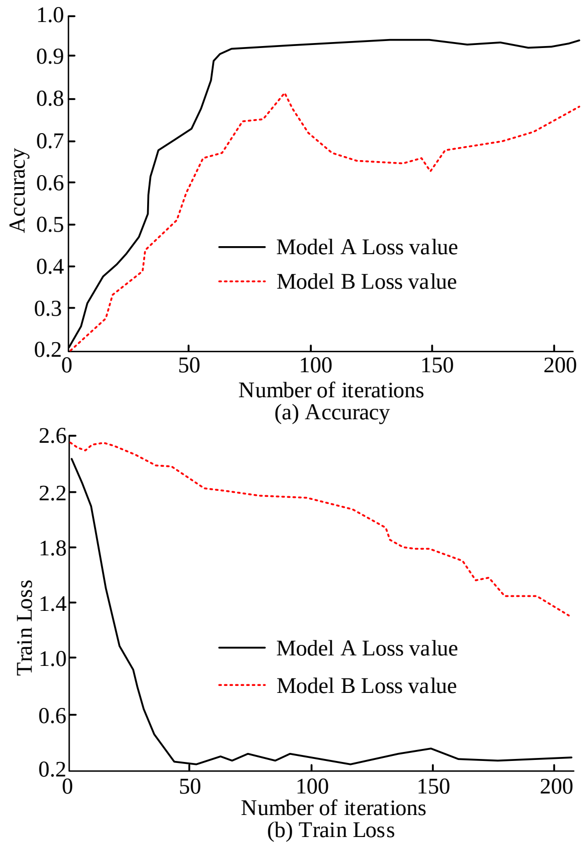


Fig. 4. Data preprocessing effect

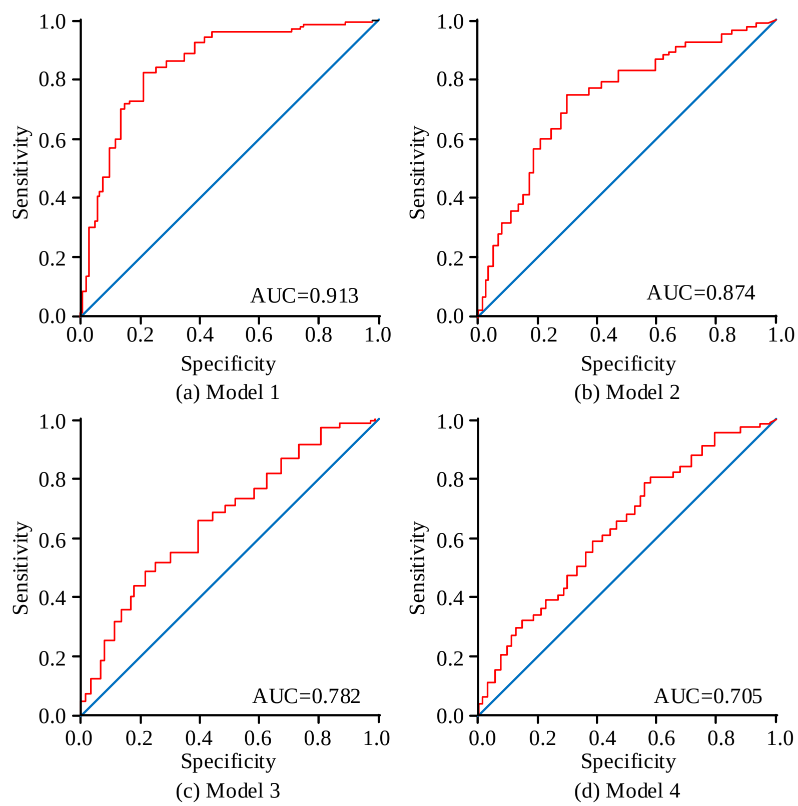


Fig. 5. ROC curve of four models

The prediction models are built based on RF-XGBoost, RF, XGBoost, and ID3 decision trees, which are represented as models 1 to 4 respectively. The performance of the built multiple models is evaluated by the receiver operating characteristic curve (ROC), as shown in Fig. 5. The results showed that model 1 had the largest AUC value, significantly exceeding the other three models. It shows that the performance of the model proposed in this study is better than several traditional prediction models, so it has high practicality and can provide accurate and scientific data support for enterprise customer churn management.

The training results of each model are compared as shown in Fig. 6. The number of iterations required for Model 1 to reach the best accuracy is 45, which is significantly less than the other three models. After iteration to the best accuracy, the error is 0.08, which is the best among all models. Therefore, the proposed model has better convergence and training effects, which is better than the other three models.

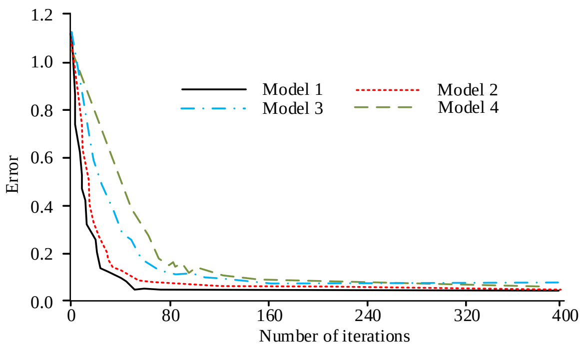


Fig. 6. Training results of each model

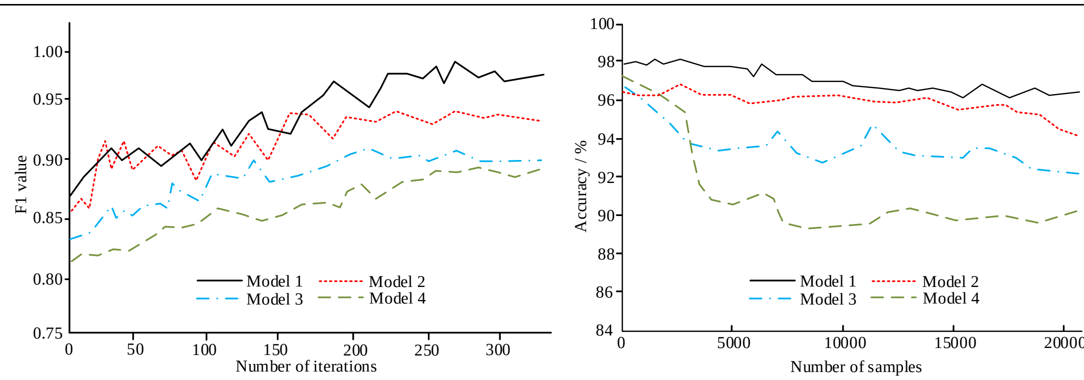
Test sets are used to test the performance of each model, as shown in Table II. For the accuracy rate, Model 1 has the highest prediction accuracy, reaching 92.15%. For the recall rate, Model 1 has the highest recall rate, reaching 0.93. For the F1 value, model 1 is the highest, reaching 0.91. In addition, the prediction time of model 1 is the shortest, 12.7s.

Further, we analyze the changes of F1 indicators and prediction accuracy of the four models in the iterative process, and the results are shown in Fig. 7. In Fig. 7 (a), after 300 iterations of model 1, the value of F1 reaches the maximum, and the value of F1 in the whole iteration process also remains optimal. In all models, this study proposes that the value of F1 in the model is optimal. In Fig. 7 (b), the accuracy of all models will decrease with the increase of test data. On the whole, no matter whether the number of samples increases or decreases, the prediction accuracy of model 1 is the best. It is said that the model proposed in this study has excellent prediction accuracy, which can accurately predict customer churn based on user data, thus guiding e-commerce enterprises to Equationte customer maintenance plans.

Divide the selected test set data into 25 pieces, test the four models respectively, and compare the average error between various models and the real output results. After 25 tests on four models, see Fig. 8. The average output error of model 1 is 0.42, which is the lowest among all comparison models, suggesting that the model proposed in this study has a smaller error and higher accuracy.

TABLE II. PERFORMANCE TEST RESULTS OF FOUR MODELS

Model	Index			
	Accuracy / %	Recall / %	F1	Detection time / s
1	92.15	0.93	0.91	12.7
2	85.34	0.89	0.84	33.5
3	86.19	0.87	0.82	48.3
4	83.48	0.80	0.76	36.9



(a) F1

(b) Accuracy

Fig. 7. F1 value and accuracy change curve of four models

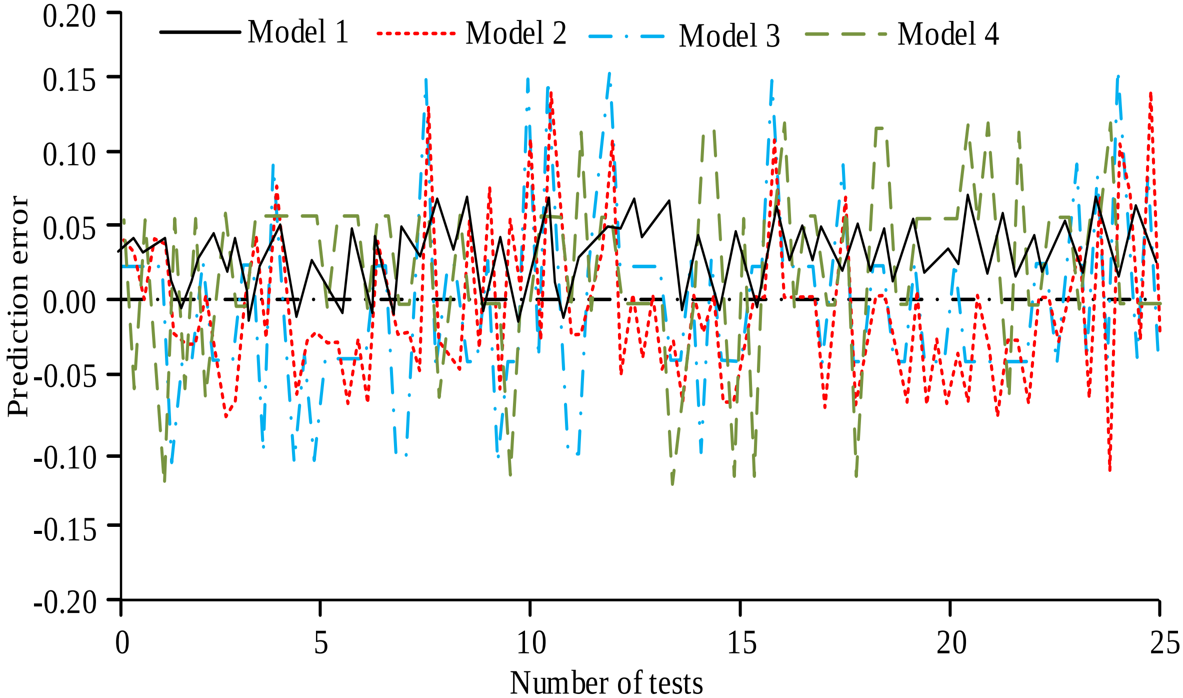


Fig. 8. Comparison of output errors of various models

The prediction model proposed in this study has excellent performance, including prediction accuracy, stability, and convergence. It can make a rough assessment of customer churn, thus providing data support for e-commerce enterprises' customer maintenance work, helping to analyze the relevant factors affecting customer churn, to Equationte targeted customer service programs, thus improving the enterprise's revenue.

In view of the importance of the model proposed in this study to other models in the application process, the ROC curve is selected for comparative study of model performance. It can be seen from the Fig. 9 that the AUC area of the optimized model proposed in this experiment is larger than that of the non-optimized model, which can significantly improve the accuracy of the method.

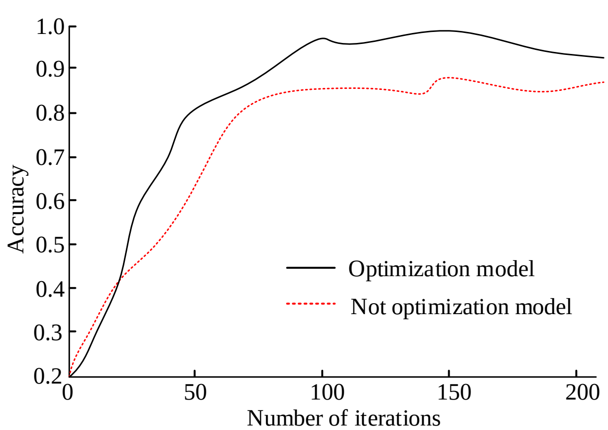


Fig. 9. Performance comparison between different models

V. Conclusion

For e-commerce enterprises, they must face fierce competition for customer resources from other enterprises. While enterprises that fail in the competition will be eliminated from the market. This research takes e-commerce enterprise customers as the research object, analyzes the relevant behavioral data recorded by the background when customers purchase online, subdivides them according to customer value, combines the distributed gradient enhancement library algorithm (XGBoost) with the random forest algorithm (RF), and constructs a big data analysis model for e-commerce customer churn. The results show that the AUC value of model 1 is the largest, which shows that the performance of the model proposed in this study is significantly better than several traditional prediction models, so it has high practicality and can provide accurate and scientific data support for enterprise customer churn management. The number of iterations required for Model 1 to reach the best accuracy is 45, which is significantly less than the other three models. After iteration to the best accuracy, the error is 0.08, which is the best among all models. Therefore, the proposed model has better convergence and better training effect. For the accuracy rate, Model 1 has the highest prediction accuracy, reaching 92.15%; for the recall rate, Model 1 has the highest recall rate, reaching 0.93; For F1 value, model 1 is the highest, reaching 0.91; In addition, the prediction time of model 1 is the shortest, 12.7s. After 300 iterations of model 1, the value of F1 reaches the maximum, and the value of F1 in the whole iteration process also remains optimal. In all models, this study proposes that the value of F1 in the model is optimal. The prediction accuracy of model 1 is the best, indicating that the model proposed in this study has excellent prediction accuracy. Since the data obtained in this study are preliminary data of e-commerce customers, deeper customer behavior data such as evaluation, sharing, return and exchange cannot be obtained, which leads to the too one-sided prediction of customer behavior. We will try to get more customer behavior data.

REFERENCES

[1] B. Senthilnayaki, M. Swetha, D. Nivedha, “Customer Churn Prediction.” IARJSET, Vol. 8. No. 6, pp. 527-531, 2021.

[2] A. Amin, S. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain, K. Huang, “Customer churn prediction in the telecommunication sector using a rough set approach.” Neurocomputing, Vol. 237, NO. (MAY10), pp. 242-254, 2017.

[3] R. Yu, X. An, B. Jin, J. Shi, O.A. Move, Y. Liu, “Particle classification optimization-based BP network for telecommunication customer churn prediction.” Neural Computing and Applications, Vol. 29, No. 3, pp. 707-720, 2018.

[4] I. Alshourbaji, N. Helian, Y. Sun, M. Alhameed, “Anovel HEOMGA Approach for Class Imbalance Problem in the Application of Customer Churn Prediction.” SN Computer Science, Vol. 2, No. 6, pp.1-12, 2021.

[5] A. Sk, B. Hl. “Customer Churn Prediction in Influencer Commerce: An Application of Decision Trees.” Procedia Computer Science, Vol. 199, pp. 1332-1339, 2022.

[6] S. Janabi, F. Razaq, “Intelligent Big Data Analysis to Design Smart Predictor for Customer Churn in Telecommunication Industry,” Springer, Cham. Springer, Cham, Vol. 53, pp. 4659-4676, 2018.

[7] W. Li, C. Zhou, “Customer churn prediction in telecom using big data analytics.” IOP Conference Series: Materials Science and Engineering, Vol. 768, No. 5, pp. 052070, 2020.

[8] Y. Gu, T. D. Palaoag, J. Cruz, “Comparison of Main Algorithms in Big Data Analysis of Telecom Customer Retention.” IOP Conference Series: Materials Science and Engineering, Vol. 1077, No. 1, pp. 012045, 2021.

[9] G. Jiao, H. Xu, “Analysis and Comparison of Forecasting Algorithms for Telecom Customer Churn.” Journal of Physics: Conference Series, Vol. 1881, No. 3, pp. 032061 (6pp), 2021.

[10] K. Eria, B. P. Marikannan, “Significance-Based Feature Extraction for Customer Churn Prediction Data in the Telecom Sector.” Journal of Computational and Theoretical Nanoscience, Vol. 16, No. 8, pp. 3428-3431, 2019.

[11] T. Sharma, P. Gupta, V. Nigam, M. Goel, “Customer Churn Prediction in Telecommunications Using Gradient Boosted Trees.” Vol. 1059, pp. 235–246, 2020.

[12] P. Swetha, B. Dayananda, “Improvised_XgBoost Machine learning Algorithm for Customer Churn Prediction.” EAI Endorsed Transactions on Energy Web, Vol. 7, No. 30, pp. 164854, 2018.

[13] A. M. Naser, Al-Shamery E S. Churners Prediction Based on Mining the Content of Social Network Taxonomy. International Journal of Recent Technology and Engineering, 8(Issue-2S10):341-351, 2020.

[14] A. Vignesh, T. Y. Selvan, G. G. Krishnan, A.N. Sasikumar, A.V.D. Kumar. “Efficient Student Profession Prediction Using XGBoost Algorithm // International Conference on Emerging Current Trends in Computing and Expert Technology.” Springer, Cham, Vol. 35, pp. 140-148, 2020.

[15] P. Lalwani, M. K. Mishra, J. S. Chadha, P. Sethi, “Customer churn prediction system: a machine learning approach.” Computing, Vol. 104, No. 2, pp. 271-294, 2021.

[16] D. Liu, X. Zhang, Y. Shi, H. Li, “Prediction of Railway Freight Customer Churn Based on Deep Forest.” Vol. 12837, pp. 479-489,2021.

[17] M. Panjasuchat, Y. Limpiyakorn, “Applying Reinforcement Learning for Customer Churn Prediction.” Journal of Physics: Conference Series Vol. 1619, NO. 1, pp. 012016 (5pp), 2020.

[18] H. Wei, Q. T. Zeng, “Research on sales Forecast based on XGBoost-LSTM algorithm Model.” Journal of Physics: Conference Series, Vol. 1754, No. 1, pp. 012191 (6pp), 2021.

[19] M. Li, X. Fu, D. Li, “Diabetes Prediction Based on XGBoost Algorithm.” IOP Conference Series: Materials Science and Engineering, Vol. 768, No. 7, pp. 072093 (7pp), 2020.

[20] Z. Liu, Q. Kong, L. Yang, “Power consumption prediction with K-nearest-neighbours and XGBoost algorithm.” International Journal of Wireless and Mobile Computing, Vol. 15, No. 4, pp. 374-381, 2018.