Iraqi Journal for Computer Science and Mathematics

Manuscript 1239

Improving Heart Attack Prediction Accuracy Performance Using Machine Learning and Deep Learning Algorithms

Mosleh Hmoud Al-Adhaileh

Mohammed Ibrahim Ahmed Al-mashhadani

Eidah M Alzahrani

Theyazn H.H. Aldhyani

IRAQI JOURNAL FOR COMPUTER SCIENCE AND MATHEMATICS 2025;6:30–43

ORIGINAL STUDY

Improving Heart Attack Prediction Accuracy Performance Using Machine Learning and Deep Learning Algorithms

Mosleh Hmoud Al-Adhaileh [a], **Mohammed Ibrahim Ahmed Al-mashhadani** [b], **Eidah M Alzahrani** [c], **Theyazn H.H. Aldhyani** [d,*]

[a] Deanship of E-Learning and Distance Education and Information Technology, King Faisal University, P.O. Box 4000, Al-Ahsa 31982, Saudi Arabia
[b] AL-Iraqia University, College of Education, Computer Department, Baghdad, Iraq
[c] Computer Science Department Al-Baha University, Saudi Arabia
[d] Applied College in Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia

ABSTRACT

Accurate classification of cardiovascular diseases (CVDs) is of utmost importance for cardiologists to provide appropriate treatments. Diagnosing and predicting cardiovascular conditions are crucial medical responsibilities in this context. The healthcare sector is increasingly utilizing deep learning (DL) and machine learning (ML) algorithms due to their ability to identify patterns in data. Diagnosticians may reduce the number of misdiagnoses by using DL and ML techniques for the categorization of cardiovascular disease incidence. To reduce the mortality linked to CVDs, this research offers a unique model that properly predicts and classifies these problems. This research presents approaches such as deep learning, random forest (RF), support vector machines (SVM), and K-nearest neighbors for predicting heart disease arrest. We implemented the suggested model using 303 real-world instances from Kaggle. In the testing stage, the KNN model's accuracy was 92%. By comparison, the accuracy of the DL model was 87%. The RF model's accuracy was 84%. The results indicate that the KNN algorithm outperforms other algorithms in terms of accuracy. We compared the study's results with a variety of existing systems.

Keywords: Cardiovascular, Deep learning, Machine learning, E-health, Heart disease

1. Introduction

According to data from the World Health Organization (WHO), heart disease is a major global health risk. There are multiple factors that can contribute to CVDs, such as hypertension, obesity, elevated cholesterol levels, diabetes, and arrhythmias [1, 2]. Most patients succumb to cardiovascular disease due to an insufficient diagnosis during the early stages. Thus, it is crucial to employ effective disease classification and prediction algorithms to understand disease prognosis.

CVDs are responsible for many deaths worldwide, making up approximately 70% of all fatalities. Risk factors for heart disease in developed nations often include poor dietary choices, tobacco use, excessive sugar intake, and obesity [5, 6]. Moreover, there has been a rise in the occurrence of chronic diseases in low- and middle-income nations [7]. Between 2010 and 2015, experts estimated that cardiovascular disorders would have a global economic impact of around USD 3.7 trillion [8, 9]. Approximately 25–30% of the annual medical expenditures borne by firms may be attributable to personnel suffering

Received 11 October 2024; revised 28 December 2024; accepted 31 December 2024.
Available online 12 April 2025

* Corresponding author.
E-mail addresses: madaileh@kfu.edu.sa (M. H. Al-Adhaileh), mohamed672005@yahoo.com (M. I. A. Al-mashhadani), Em.Alzahrani@bu.edu.sa (E. M Alzahrani), taldhyani@kfu.edu.sa (T. H.H. Aldhyani).

https://doi.org/10.52866/2788-7421.1239

from cardiovascular diseases [10]. Hence, the timely identification of cardiovascular disease is crucial to minimize its effects on individuals and institutions, both in terms of health and finances. The World Health Organization predicts that the number of deaths caused by CVDs will increase to 23.6 million by 2030. Cardiovascular disease and cerebrovascular events are the main causes of these deaths [11]. To address the issue of high mortality rates and the economic impact on society, it is crucial to utilize data mining and ML techniques to accurately predict and classify the probability of developing cardiovascular diseases.

Through the use of medical research, a great number of risk factors that are related to heart disease have been identified. These include inactivity, poor dietary choices, obesity, and bad lifestyle choices. The risk of developing heart disease may be increased by several variables, such as hypertension, smoking, stress, diabetes, higher blood pressure, and high cholesterol levels [12–17]. The complex and diverse nature of heart disease makes it more important to develop effective methods of prediction and prevention to lessen the impact it may have. Researchers have successfully analyzed datasets and uncovered significant insights across many domains, including healthcare, using data mining and ML approaches [18]. ML algorithms can effectively sift through mountains of medical data in search of patterns that will allow for more precise disease detection, diagnosis, and prediction [19–21]. Though several research have investigated sickness prediction in the past, it is still difficult to reliably forecast cardiac disease [22]. To fill this informational void, this research introduces a model for cardiac disease prediction that integrates different ML tactics, hyperparameter optimization approaches, and preprocessing techniques. The model's stated goal is to evaluate the classification accuracy of different ML algorithms to make accurate predictions about heart disease. The purpose of this research is to examine the benefits and drawbacks of using modern ML methods to analyze heart diseases. Following that, we analyzed the data and used seven ML/DL prediction models to improve CV and vascular disease diagnosis. This study incorporates DL, RF, KNN, and SVM models. The following noteworthy research results are presented in this article.

This study examines and utilizes well-known ML algorithms to analyze the Cleveland heart disease datasets, assessing performance classification metrics. The proposed research seeks to enhance the precision of ML and DL methods. Comparing the suggested research with other articles in the field revealed an impressive performance of 92%. The proposed idea aims to answer the following main research question:

What ways might DL and ML models enhance the accuracy and generalizability of cardiovascular disease prediction?

2. Related works

This section presents an assessment of ML models specifically developed for predicting CVDs, along with an extensive examination of the most recent research in this field. This section provides an overview of the advancements, methodologies, and significant findings in the application of ML models for predicting cardiovascular disease.

The assessment of various data from multiple sources is conducted using cutting-edge ML models [23]. The choice of an ML model may greatly affect CVD prediction accuracy and reliability [2, 4]. Given the vast amount of health and medical data available and the rapid advancements in computational power, ML models have become essential tools in various applications [25]. We employ ML models to categorize patients into distinct risk groups, thereby facilitating focused therapies. These models can assist in the identification of patients with a high susceptibility to cardiovascular disease and facilitate the development of personalized treatment strategies. Healthcare practitioners are incorporating ML models into clinical workflows to offer immediate decision help in real time [26]. We will examine various machine-learning algorithms frequently used for predicting cardiovascular illness in this section.

Common ML models include Logistic Regression [27], Decision Trees (DT) [28–30], RF approach [31], SVM [32], KNN, AdaBoost [33], Naïve Bayes [34], and CNN, also known as DL [35]. The study introduced a model that sought to pinpoint the most effective machine learning technique for precisely forecasting early-stage cardiovascular diseases. According to the findings, the RF approach demonstrated the highest accuracy rate of 95.4% in effectively classifying cardiovascular disease [36]. In [37], researchers used a neural network and fuzzy logic to categorize CVDs. The system was 87.4% accurate.The study [38] introduced a diagnosis method for cardiovascular illness. The method utilized a set of artificial neural networks. This technique was incorporated into the statistical measuring tool Enterprise Miner. In reference [39], a machine learning-based method was developed to diagnose cardiovascular illness. The system utilized an Artificial Neural Network (ANN) algorithm in conjunction with a feature selection approach, resulting in noteworthy outcomes.

The study described in [40] introduced an advanced medical detection system created to identify cardiovascular disorders. This system employed various predictive ML models, such as Naïve Bayes, DT, and ANN. Researchers conducted a study that developed a three-phase methodology to forecast the occurrence of CVDs in individuals with angina using ANN. The approach exhibited a precision rate of 88.89%.

Shah et al. [42] developed a CVD prediction ML model. This study used data from the UCI repository's CVDs dataset, which included 303 occurrences and 17 attributes. The authors used naive Bayes, DT, RF, and KNN supervised classification techniques. Drod et al. [43] employed machine learning to identify the primary risk factors for cardiovascular disease (CVD) in individuals with metabolic-associated fatty liver disease (MAFLD). We used principal component analysis (PCA) to identify the most relevant characteristics, and then utilized machine learning (ML) techniques to analyze the data and identify those who were highly likely to have cardiovascular disease (CVD). Alotalibi [44] conducted a study where he tested the ability of machine learning to predict cardiac failure. The study constructed prediction models utilizing data from the Cleveland Clinic Foundation and machine learning techniques like decision tree, logistic regression, RF, naive Bayes, and SVM. This research illuminates ML's heart failure prediction capabilities. Hasan and Bao [45] investigated the best CVDs prediction feature selection method. Their investigation included numerous algorithms. Start with filter, wrapper, and embedding, the three most frequent feature selection techniques. We compared RF, SVM classifier, KNN, naive Bayes, and XGBoost to find the best predictive analytics method.

These gaps include issues like managing unbalanced datasets, the restricted interpretability of model conclusions, and the incapacity of some standard approaches to elucidate complicated, non-linear connections within the data. Smaller datasets recognize models like KNN and SVM for their simplicity and efficacy, while RF offers resilience through ensemble learning. DL proficiently identifies complex patterns and connections throughout extensive datasets.

3. Methodology

In this research, we have developed the methodology to predict and classify cardiovascular disease using advanced ML and DL approaches, which have the potential to deliver substantial advantages to healthcare professionals and patients. In order to achieve this objective, we used a range of ML and DL methodologies on a dataset, documenting our discoveries in this research paper. Fig. 1 illustrates the architecture of the suggested system. To optimize the performance of the ML and DL algorithms, we have used feature engineering and normalization techniques.

3.1. Heart disease dataset

This study utilized the Kaggle database to obtain the heart disease dataset. The dataset comprises two classes, 0 denoting no disease and 1 denoting disease, with both columns displaying numeric values. Fig. 2 shows the features of the dataset.

In the "Sex" attribute, males represent 1, while females represent 0. There are four types of chest pain included in the 'cp' (chest pain type) attribute. There are two categories of fasting blood sugar levels for the 'fbs' attribute. The 'restecg' attribute categorizes resting electrocardiograms into three groups, while the 'exang' attribute differentiates exercise into two classes, specifically associated with exercise angina. Moreover, the 'slope' attribute (ST slope) divides into three distinct categories. The remaining four attributes, 'trestbps' (resting blood pressure), 'chol' (cholesterol), 'age', and 'oldpeak', consist of numerical values. Fig. 3 depicts the classes present in the dataset.

3.2. Preprocessing steps

In this study, we initially normalize the dataset's features using StandardScaler. This method standardizes the attributes by subtracting the mean and scaling them to have a variance of one. This ensures that all features have an equal contribution to the training process of the model. We have normalized the heart disease data and then divided the dataset into training and testing sets. We divide the dataset into two subsets: we use 80% for training the model and the remaining 20% for evaluating its performance. The random state parameter ensures the reproducibility of the split, maintaining the same data partition across multiple model runs. Fig. 4 presents the features of the dataset after scaling.

3.3. Machine learning

3.3.1. Random forest (RF) classifier

The RF Classifier is a ML algorithm that generates numerous DTs during training and selects the class by aggregating the predictions of these trees by majority voting. This approach capitalizes on the advantages of several DTs to enhance the overall efficiency and

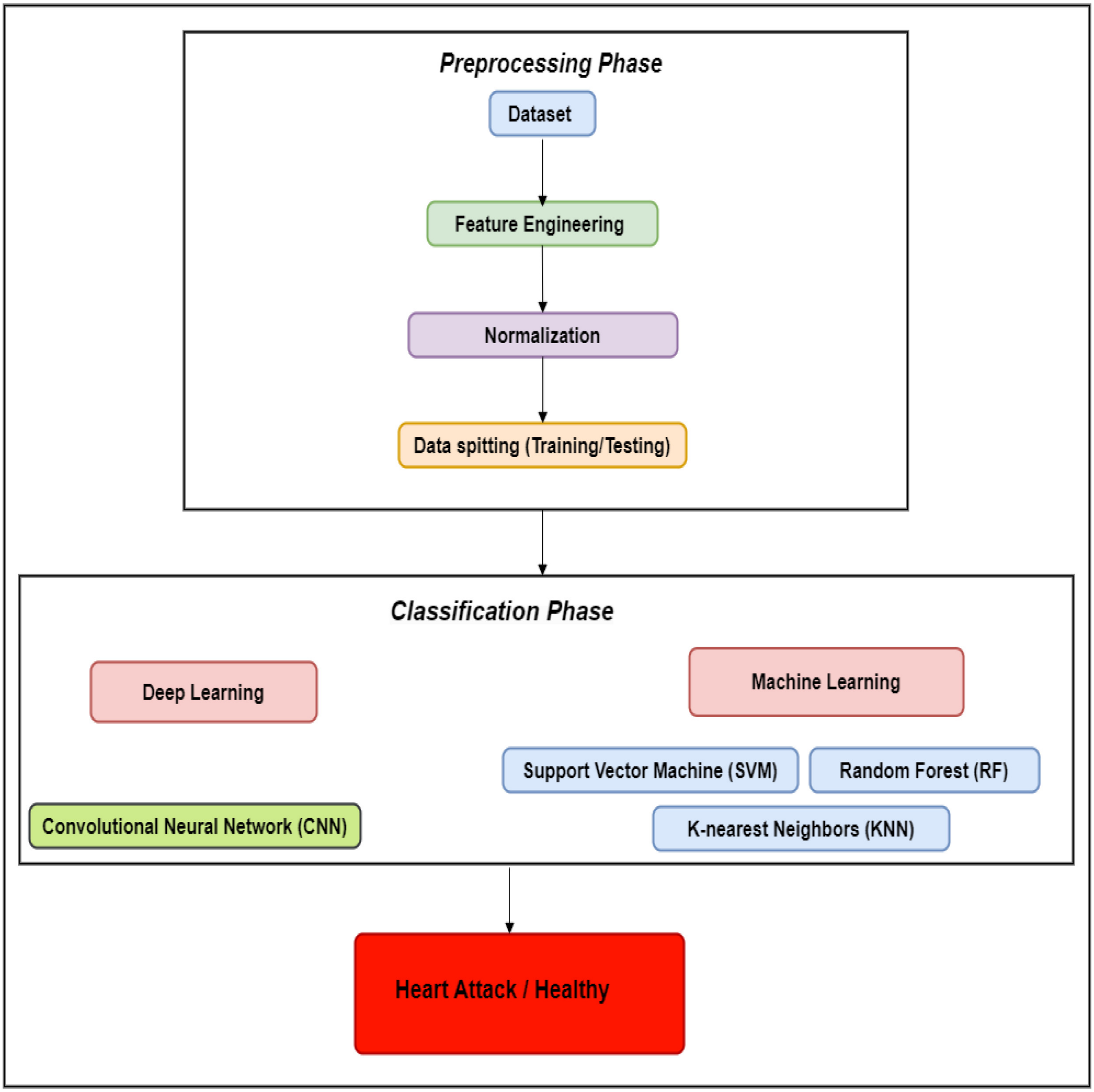

Fig. 1. Framework of the proposed.

resilience of the model. By calculating the mean of the outcomes from multiple trees, it mitigates the problem of overfitting in comparison to using a single decision tree. Additionally, it is capable of effectively processing huge datasets that have a high number of dimensions.

Every tree in the forest undergoes training using a randomly selected subset of the data. This approach creates diversity and enhances the ability of the model to generalize. The ultimate forecast is determined by combining the forecasts from each individual tree, typically through majority vote in classification tasks. This methodology facilitates the identification and analysis of intricate patterns and connections within the dataset.

There are a total of 100 trees in the RF Classifier. The system generates probabilities for each class, subsequently calculating performance indicators like the ROC curve. This model is renowned for its exceptional precision and adeptness in properly managing unbalanced datasets. Fig. 5 displays the RF structure in this research.

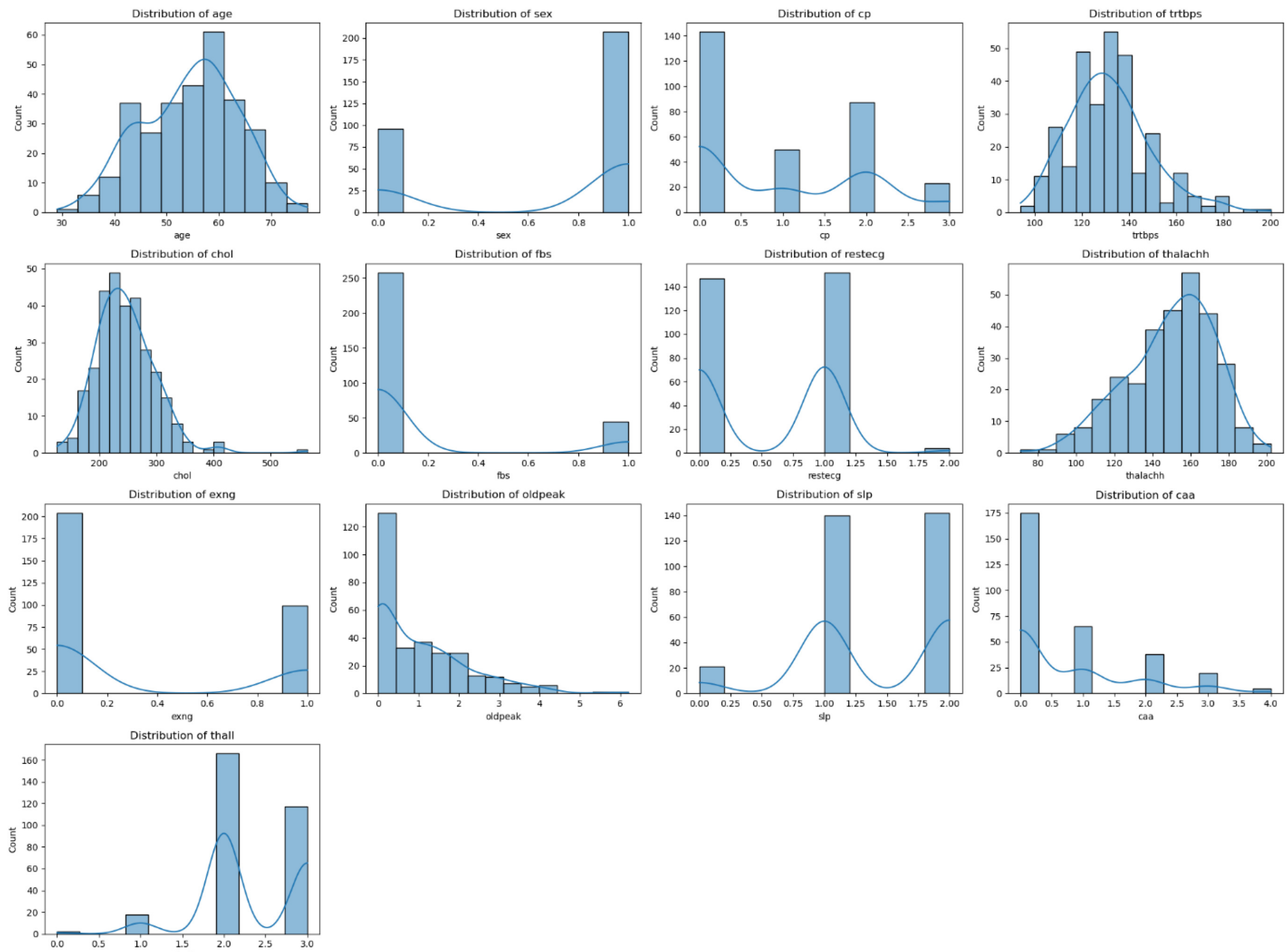

Fig. 2. Features of dataset.

3.3.2. K-Nearest neighbours (KNN)

The K-Nearest Neighbours (KNN) Classifier is a straightforward technique for instance-based learning that categorises data points by determining the majority class among their neighbours. The algorithm operates by storing the complete training dataset and generating predictions by evaluating the similarity between the test instance and its closest neighbours in the feature space.

KNN utilises a distance metric, usually the Euclidean distance, to assess the proximity between data points. The selection of k, the number of neighbours, is a critical factor in determining the effectiveness of the model. Smaller values of k result in a more responsive model, while bigger values lead to a more generalized classification.

The KNN model in this research utilises a set of 5 neighbouring data points to generate predictions. The evaluation process assesses the quality of the model by comparing the predicted labels with the actual values. It also creates probabilities that indicate the likelihood of class membership. These probabilities are valuable for calculating metrics like the ROC curve and AUC. Fig. 6 shows the structure of the KNN model in this proposed system.

3.3.3. Support vector machine (SVM)

The SVM classifier is an effective ML technique that seeks to identify the best hyperplane that maximises the separation between distinct classes. A hyperplane is a border that distinguishes data points belonging to one class from those belonging to another. The SVM approach is highly efficient in dealing with datasets that have a large number of dimensions. They are particularly useful for classification problems where a straight line cannot easily separate the classes. The SVM approach can employ various kernels to convert the input space into a space with more dimensions, enabling a linear separation. We frequently employ the radial basis function (RBF) kernel to address non-linear situations.

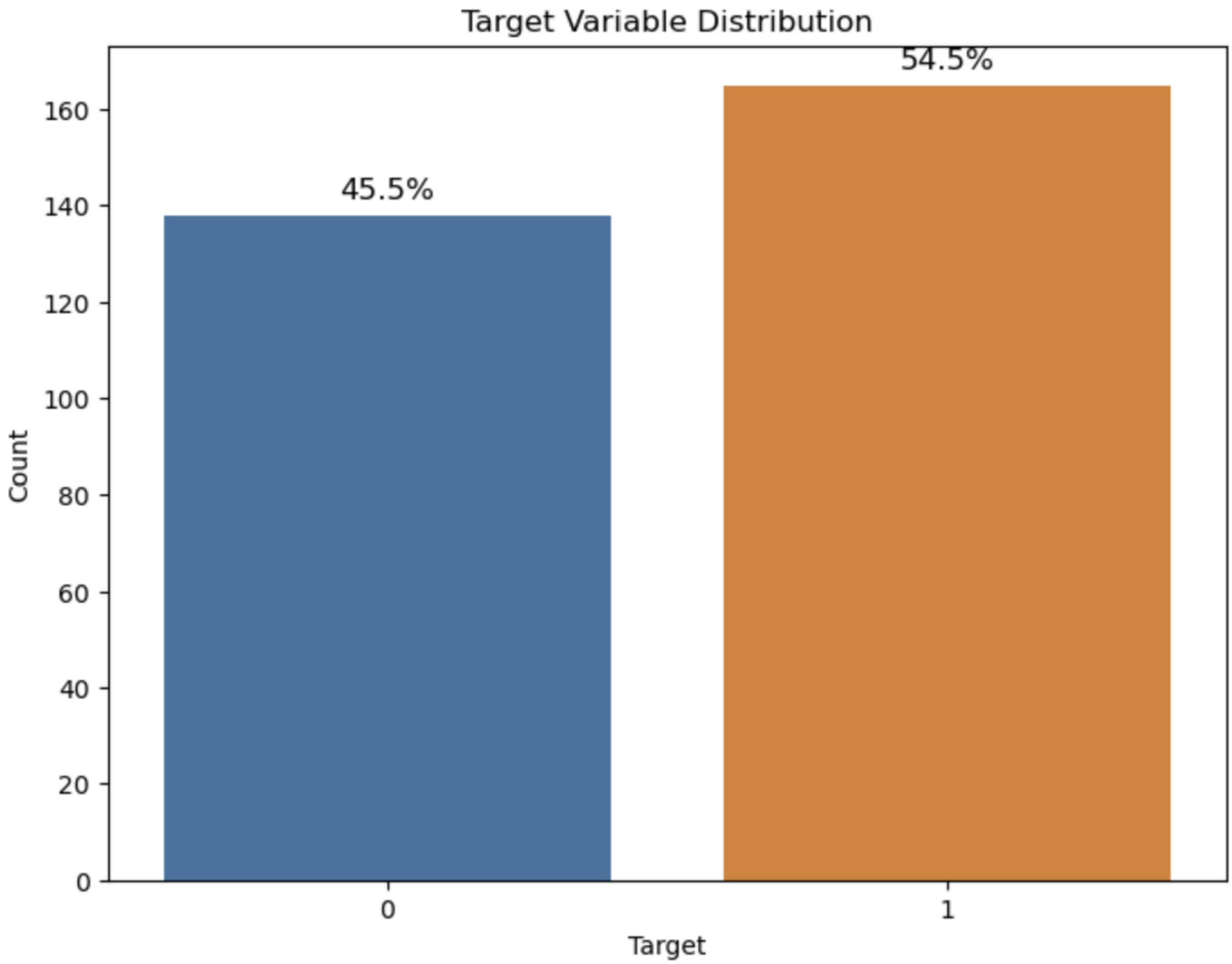

Fig. 3. Classes of dataset.

The provided code configures the SVM model to facilitate the use of probability estimates. This enables the creation of class probabilities and simplifies the performance measurement process using the ROC curve. The SVM algorithm is known for its resilience in handling diverse data types and its ability to achieve exceptional accuracy. Fig. 7 displays the structure of SVM.

3.4. Deep learning model

This research effort illustrates the process of constructing and training a deep learning model using TensorFlow and Keras [49]. The model is constructed using the Sequential API, which enables a direct and uncomplicated arrangement of layers. The process begins with a Dense layer consisting of 64 neurones and ReLU activation. Next, a Dropout layer with a 50% dropout rate intervenes to reduce overfitting. Afterwards, there are two more dense layers, with 32 and 16 neurones, respectively, both employing ReLU activation. The last layer consists of a dense layer that contains only one neurone and uses a sigmoid activation function. This configuration is well-suited for problems involving binary classification. Once the model architecture is defined, the code proceeds to compile the model using the Adam optimiser and the binary cross-entropy loss function. The Adam optimiser has been selected because of its flexible learning rate capabilities, which often lead to accelerated convergence. The model has been trained with a batch size of 32 for 50 epochs, utilizing the training data for heart disorders. Another technique used to evaluate DL performance on unseen data during training is a validation split, which accounts for 20% of the training data. The fit method produces a history object that includes metrics for each epoch. These metrics can be utilised for additional analysis or visualisation of the heart diseases at the training process. Fig. 8 presents the structure of the DL model.

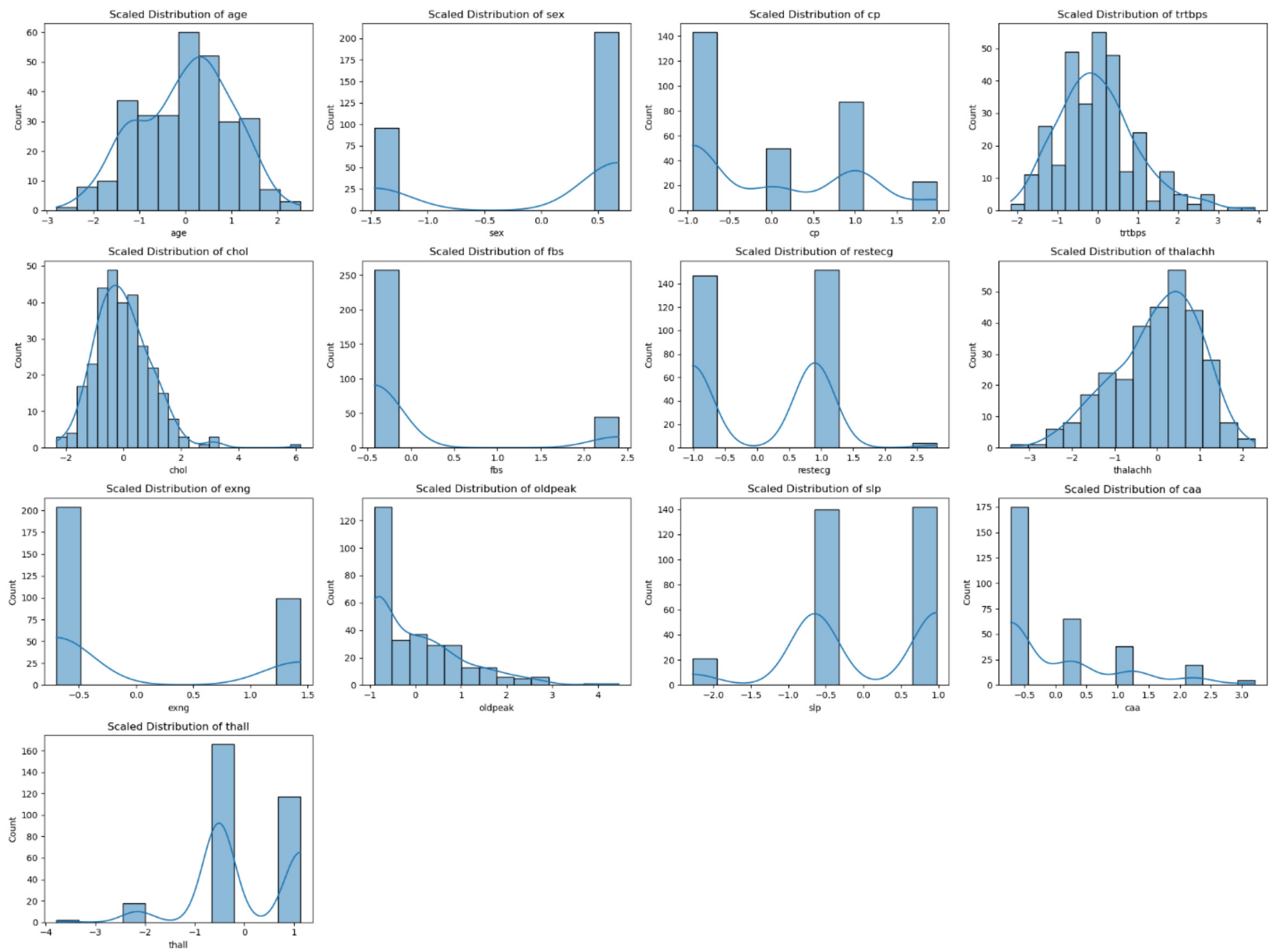

Fig. 4. Scaling features.

4. Experiment

This section showcases the empirical findings of DL and ML algorithms in the classification and prediction of heart attacks using a standardised dataset. We include a description of the experimental setting, a definition of assessment measures, and a detailed presentation of classification results for each unique model.

4.1. Environment setup

The experimental results of our investigation were obtained utilizing a PC equipped with a Core i7 processor and 8 GB of RAM. Our models were created using the TensorFlow framework. The hardware and software configurations were crucial for the effective training and evaluation of ML and DL models.

4.2. Evaluation metrics

Assessing the efficacy of DL models is essential for comprehending their efficiency. Various metrics are employed for this objective. The evaluation measures provide an alternative perspective on the model's advantages and disadvantages.

4.2.1. Confusion matrix

The evaluation of a binary classification model is dependent on the use of a confusion matrix. It quantifies the level of accuracy with which the model predicts outcomes on the test dataset. The matrix is divided into four distinct categories: True Positives (TP) refer to the cases of heart attack that are accurately predicted by the model. False Positives (FP) refer to the instances where the model incorrectly classified negative cases (normal) as positive (heart attack). True Negatives (TN) refer to the instances where the model accurately predicted negative outcomes. False Negatives (FN) are instances that the model mistakenly classified as negative (normal) when they were actually positive (heart attack).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (1)$$

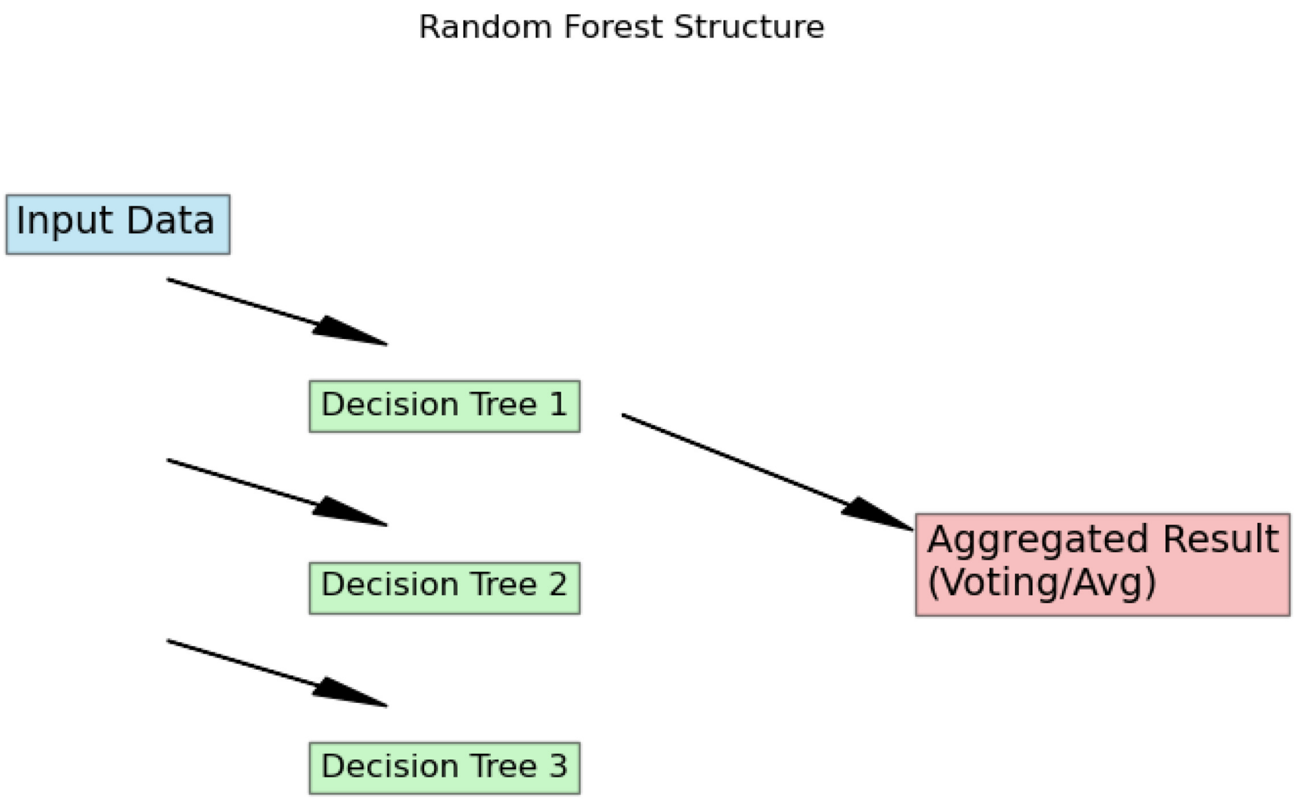

Fig. 5. RF structure.

Table 1. Results of RF approach.

	Precision%	Recall%	F1-score%	Support	Accuracy%
Normal	83	83	83	29	
Disease	84	84	84	32	84
Macro average	84	84	84		

$$F1\text{-}score = 2 * \frac{precision \times Recall}{precision + Recall} \times 100\% \tag{2}$$

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ positives} \times 100\% \tag{3}$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Negatives} \times 100\% \tag{4}$$

4.3. Results of machine learning models

The RF model achieved an overall accuracy of 84% in heart attack classification, as shown in Table 1. The model attained an 84% recall rate and an 84% precision rate for heart attacks, accurately detecting the majority of them while generating a minimal number of false positives.

Table 2. Results of KNN.

	Precision%	Recall%	F1-score%	Support	Accuracy%
Normal	90	93	92	29	
Disease	94	91	92	32	92
Macro average	92	92	92	61	

Table 3. Results of SVM.

	Precision%	Recall%	F1-score%	Support	Accuracy%
Normal	84	90	87	29	
Disease	90	84	87	32	87
Macro average	87	87	87	61	

Table 2 displays the overall accuracy performance of the KNN algorithms based on the applied metrics. The KNN algorithm uses the default hyperparameter values. The KNN classifier had a higher accuracy rate of 92%. The model attained a peak accuracy, recall, and F1-score, all of which reached 92%. The KNN algorithm has shown improved performance in identifying the illness class (1) compared to the normal class.

Table 3 presents the application of SVM classifiers on the heart disease dataset to detect the presence

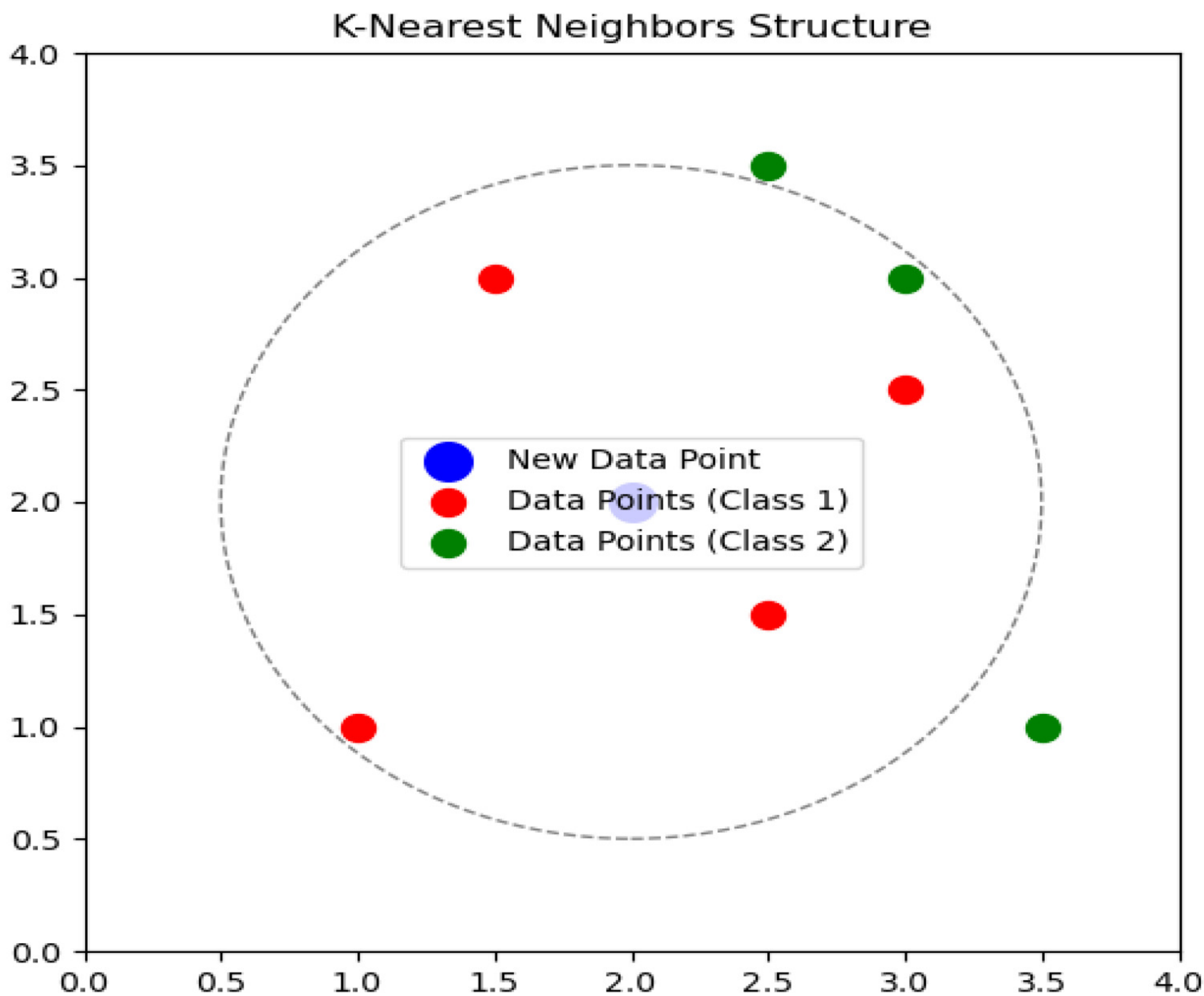

Fig. 6. KNN structure.

of CVDs following adjustments to hyperparameters. The results indicate that the SVM algorithm achieved an accuracy rate of 87%, along with notable recall, precision, and F1 scores, all of which were also 87%. The SVM approach has shown significant efficacy in disease classification (1).

Table 4. Results of deep learning approach.

	Precision%	Recall%	F1-score%	Support	Accuracy%
Normal	86	86	86	29	
Disease	88	88	88	32	87
Macro average	87	87	87	61	

4.4. Results of deep learning model

A deep learning model tailored to binary classification tasks is used by the system. This model's goal is to analyze many parameters in order to predict the probability of acquiring heart disease. The model is designed with many fully connected layers that use ReLU activations. To prevent overfitting, there are also dropout layers. A batch size of 32 was used throughout the model's 50 epochs of training on the dataset. Next, it was tested on an independent set of data to see how well it performed. Multiple methods are used to condense the results. With metrics like accuracy, recall, and F1-scores broken down by class, the classification report shows how well the model can differentiate between good and bad examples. Table 4 displays the outcomes of the DL model. The DL model attained an accuracy rate of 87%.

The model's training and validation results are depicted using two essential plots: the accuracy plot and the loss plot, which is presented in Fig. 9. The plots

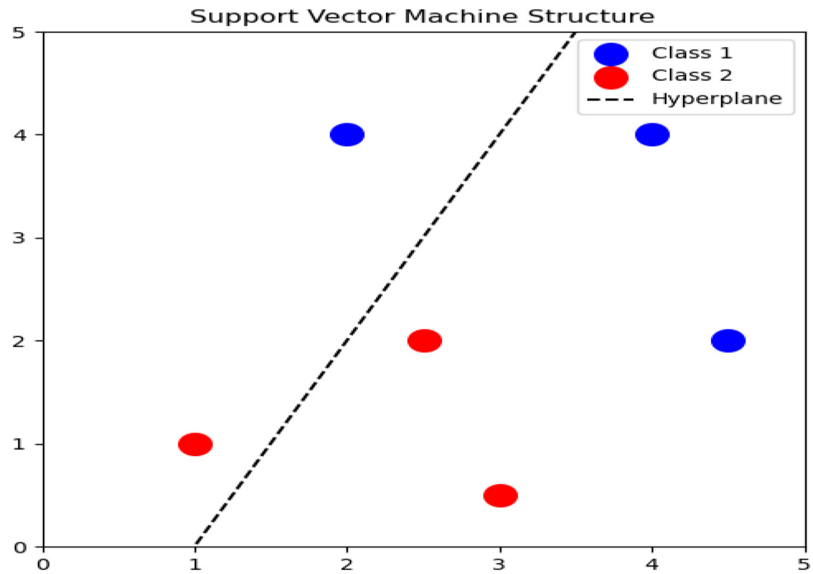

Fig. 7. SVM structure.

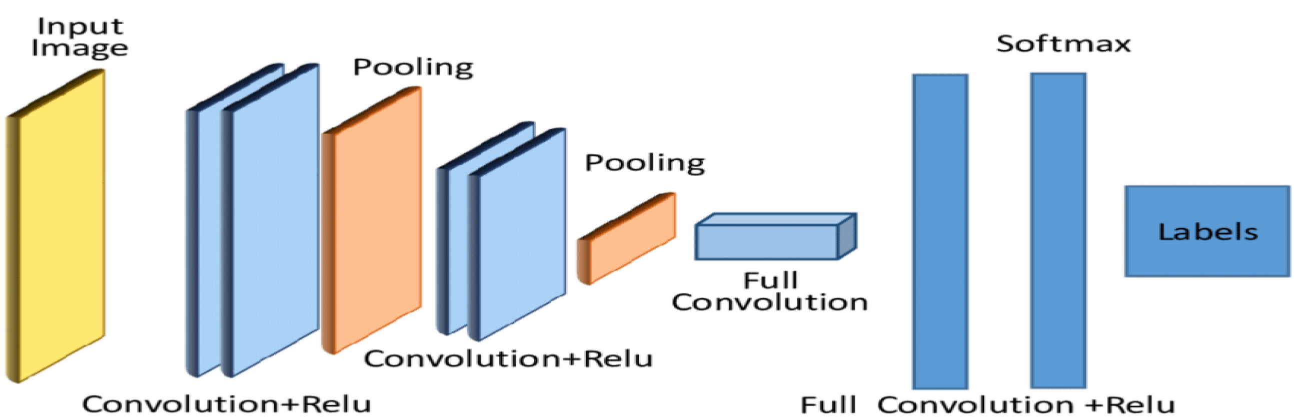

Fig. 8. Deep learning model.

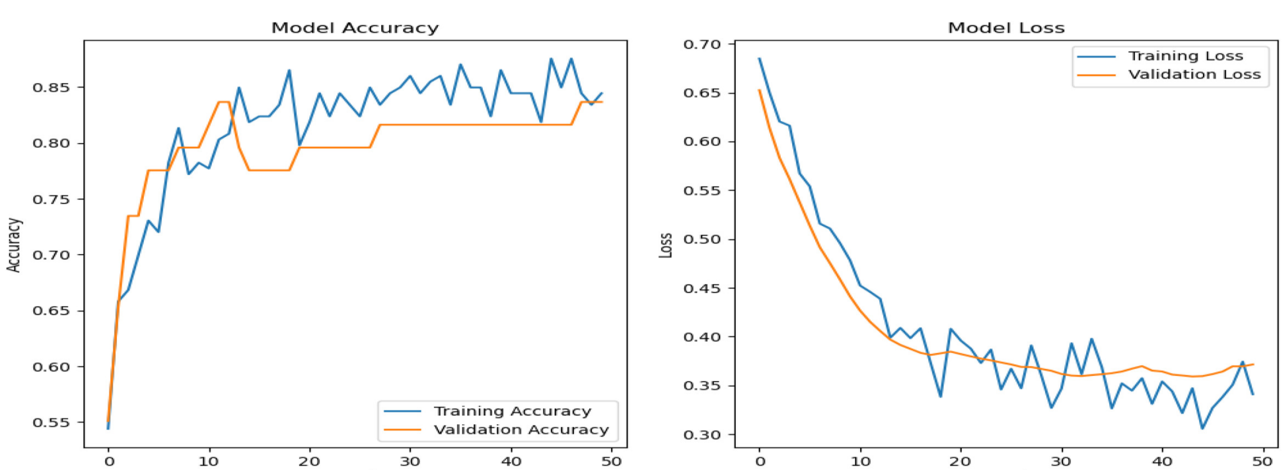

Fig. 9. Performance of DL model.

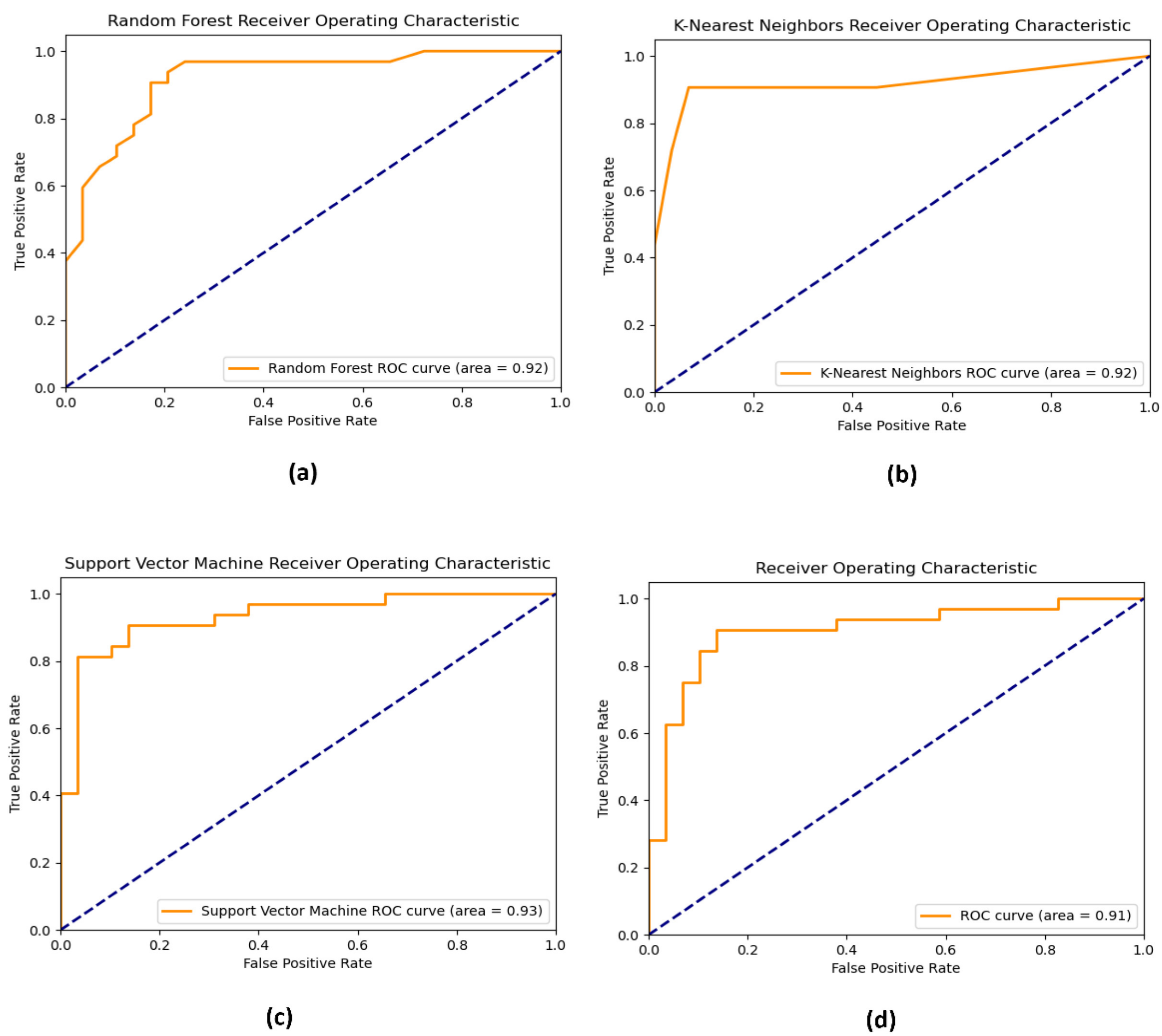

Fig. 10. ROC of ML and DL.

are derived from the training history of the model, which records metrics over epochs. The accuracy plot displays the model's precision on both the training and validation datasets throughout the epochs. The y-axis represents accuracy as a percentage, while the x-axis refers to the 50 epochs. It is noted that the model attained above 87% in training and 87% in the testing phase. The accuracy loss reaches 0.35 during the validation phase.

The Receiver Operating Characteristic (ROC) plot illustrates the model's capacity to differentiate between positive and negative classes in a visual manner. The plot illustrates the relationship between the TP and the FP, with the TP represented on the y-axis and the FP on the x-axis. The ROC curve of each model is shown in a dark orange colour, illustrating the relationship between sensitivity TP and specificity (1 - FP) as the decision threshold varies. The graphic features a dashed navy line that runs diagonally, symbolising the line of no discrimination. This line indicates the point at which the model's predictions would be indistinguishable from random guessing. Fig. 10 shows the ROC of heart disease system prediction.

The ROC curve for the RF and KNN model was achieved at 92%. The ROC curve of the Support SVM model may differ according to the selected kernel, but it typically indicates the model's ability to establish a clear separation between different classes. The SVM model achieved (ROC = 93), and this high performance is compared with different proposed systems. The deep learning model achieved 91%.

Table 5. Comparison of heart attacks system with existing models.

Ref.	Model	Acc %
[46]	Nonlinear Regression	89
[47]	SVM	89
[48]	SVM	88
Proposed KNN		92

The model we provided was compared to other current research articles on the same dataset, as displayed in Table 5. The recommended KNN algorithm demonstrated superior performance compared to previous methods, with an accuracy rate of 92%.

5. Conclusions

This study aimed to categorize cardiac disorders using models and real-world data. We conducted an analysis of the dataset, which consisted of individuals who had been diagnosed with cardiac conditions. Our approach involved utilizing machine learning (ML) and deep learning (DL) techniques to forecast the likelihood of disease occurrences. The dataset was preprocessed by normalising the heart disease data to improve the accuracy of the ML and DL models. The dataset was divided into 80% for training and 20% for testing, using a random state of 42. To predict heart disease, we assessed four DL and ML algorithms in this study: RF, KNN, SVM, and deep learning, in addition to a number of hyperparameter optimization strategies. Kaggle datasets were utilised to evaluate these different ML and DL models. Our studies' findings showed that the algorithms' performance was greatly impacted by the hyperparameter optimisation strategies. The outcomes demonstrate that all DL and ML models had higher accuracy rates. The KNN model trained on the dataset produced the best classification accuracy (92%) out of all the models tested. For the KNN algorithm, the corresponding values for Sensitivity, Specificity, Accuracy, and AUC were 92%, 92%, 92%, and 92%, respectively. The ROC metric (93%), however, gave the SVM algorithm an excellent performance grade. Researchers and practitioners working on heart disease-related tasks may find this model useful. Furthermore, a limited dataset is a small one, increasing the likelihood of yielding biased or less accurate predictions. Recognizing these limits and exploring viable methods, such as data augmentation, might enhance the system's reliability. Using advanced DL models to enhance system performance is the primary focus of future development.

Dataset: https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset.

Funding statement

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU250007).

Ethical approval

Not applicable.

Conflict of interest

The authors declare no conflict of interest.

Data availability

Dataset link: https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-predictiondataset.
Access date: 12/10/2024
Name of repository: Kaggle

Author contribution

MHA and THHA: Conceptualization, Software, Writing - original draft, Writing - review & editing. MIA and EM: Data curation, Methodology, Writing - original draft, Writing - review & editing. THHA: Data curation, Supervision, Writing - original draft, Writing - review & editing. MHA: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. MIA: Formal analysis, Project administration, Writing - original draft, Writing - review & editing. EM: Project administration, Validation, Writing - original draft, Writing - review & editing. MHA: Funding acquisition, Resources, Writing - original draft, Writing - review & editing. EM: Resources, Visualization, Writing - original draft, Writing - review & editing.

References

1. World Health Statistics. Cardiovascular Diseases, Key Facts. 2021. Available online: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 2024).
2. R. P. Choudhury and N. Akbar, "Beyond diabetes: A relationship between cardiovascular outcomes and glycaemic index," *Cardiovasc. Res.*, vol. 117, pp. E97–E98, 2021.
3. C. Estes, Q. M. Anstee, M. T. Arias-Loste, H. Bantel, S. Bellentani, J. Caballeria, M. Colombo, A. Craxi, J. Crespo, C. P. Day *et al.*, "Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and

United States for the period 2016–2030," *J. Hepatol.*, vol. 69, pp. 896–904, 2018.
4. K. Drożdż, K. Nabrdalik, H. Kwiendacz, M. Hendel, A. Olejarz, A. Tomasik, W. Bartman, J. Nalepa, J. Gumprecht, and G. Y. H. Lip, "Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach," *Cardiovasc. Diabetol.*, vol. 21, pp. 240, 2022.
5. H. S. N. Murthy and M. Meenakshi, "Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease," In *Proceedings of the International Conference on Circuits, Communication, Control and Computing*, Bangalore, India, 21–22 November 2014, pp. 329–332.
6. E. J. Benjamin, P. Muntner, A. Alonso, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, S. R. Das *et al.*, "Heart disease and stroke statistics—2019 update: A report from the American Heart Association," *Circulation*, vol. 139, pp. e56–e528, 2019.
7. V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Inform. Med. Unlocked*, vol. 26, p. 100655, 2021.
8. D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. de Ferranti, J.-P. Després, H. J. Fullerton, V. J. Howard *et al.*, "Heart disease and stroke statistics—2015 update: A report from the American Heart Association," *Circulation*, vol. 131, pp. e29–e322, 2015.
9. J. Maiga and G. G. Hungilo, Pranowo, "Comparison of machine learning models in prediction of cardiovascular disease using health record data," In *Proceedings of the 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia, 24–25 October 2019, pp. 45–48.
10. J. Li, A. Loerbroks, H. Bosma, and P. Angerer, "Work stress and cardiovascular disease: A life course perspective," *J. Occup. Health*, vol. 58, pp. 216–219, 2016.
11. Purushottam, K. Saxena, and R. Sharma, "Efficient heart disease prediction system," *Procedia Comput. Sci.*, vol. 85, pp. 962–969, 2016.
12. J. Nahar, T. Imam, K. S. Tickle, and Y. P. Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," *Expert Syst. Appl.*, vol. 40, pp. 96–104, 2013.
13. J. E. Dalen, J. S. Alpert, R. J. Goldberg, and R. S. Weinstein, "The epidemic of the 20th century: Coronary heart disease," *Am. J. Med.*, vol. 127, pp. 807–812, 2014.
14. S. Kenchaiah, J. Narula, and R. S. Vasan, "Risk factors for heart failure," *Med. Clin.*, vol. 88, pp. 1145–1172, 2004.
15. B. Dahlöf, "Cardiovascular disease risk factors: Epidemiology and risk assessment," *Am. J. Cardiol.*, vol. 105, pp. 3A–9A, 2010.
16. G. Tarkanyi, A. Tenyi, R. Hollos, P. J. Kalmar, and L. Szapary, "Optimization of large vessel occlusion detection in acute ischemic stroke using machine learning methods," *Life*, vol. 12, p. 230, 2022.
17. M. Sayadi, V. Varadarajan, F. Sadoughi, S. Chopannejad, and M. Langarizadeh, "A machine learning model for detection of coronary artery disease using noninvasive clinical parameters," *Life*, vol. 12, p. 1933, 2022.
18. N. Jothi and W. Husain, "Data mining in healthcare—A review," *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015.
19. D. Waigi, D. S. Choudhary, D. P. Fulzele, and D. Mishra, "Predicting the risk of heart disease using advanced machine learning approach," *Eur. J. Mol. Clin. Med.*, vol. 7, pp. 1638–1645, 2020.
20. T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "Predictive analysis of heart diseases with machine learning approaches," *Malays. J. Comput. Sci.*, pp. 132–148, 2022.
21. S. Haseena, S. K. Priya, S. Saroja, R. Madavan, M. Muhibbullah, and U. Subramaniam, "Moth-Flame optimization for early prediction of heart diseases," *Comput. Math. Methods Med.*, vol. 1, p. 9178302, 2022.
22. V. V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: A survey," *Int. J. Eng. Technol.*, vol. 7, pp. 684–687, 2018.
23. S. Kalamkar and G.M. A, "Clinical data fusion and machine learning techniques for smart healthcare," In *Proceedings of the 2020 International Conference on Industry 4.0 Technology (I4Tech)*, Pune, India, 13–15 February 2020, pp. 211–216.
24. M. M. Ahsan, M. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies*, vol. 9, p. 52, 2021.
25. T. M. Ghazal, M. K. Hasan, M. T. Alshurideh, H. M. Alzoubi, M. Ahmad, S. S. Akbar, B. Al Kurdi, and I. A. Akour, "IoT for smart cities: Machine learning approaches in smart healthcare—A review," *Future Internet*, vol. 13, p. 218, 2021.
26. M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, p. 541, 2022.
27. T. Addison and A. Bhadrashetty, "Advanced Predictive model for heart disease in clinical decision support systems", *EDRAAK*, vol. 2023, pp. 11–15, Feb. 2023, doi: 10.70470/EDRAAK/2023/003.
28. A. U. Haq, J. Li, J. Khan, M. H. Memon, S. Parveen, M. F. Raji, W. Akbar, T. Ahmad, S. Ullah, L. Shoista *et al.*, "Identifying the predictive capability of machine learning classifiers for designing heart disease detection system," In *Proceedings of the 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, Chengdu*, China, 14–15 December 2019, pp. 130–138.
29. I. K. A. Enriko, M. Suryanegara, and D. Gunawan, "Heart disease prediction system using k-Nearest neighbor algorithm with simplified patient's health parameters," *J. Telecommun. Electron. Comput. Eng. (JTEC)*, vol. 8, pp. 59–65, 2016.
30. P. Palimkar, R. N. Shaw, and A. Ghosh, "Machine learning technique to prognosis diabetes disease: Random forest classifier approach," In *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021* Springer, Singapore, 2022, pp. 219–244.
31. A. U. Haq, J. Li, M. H. Memon, M. H. Memon, J. Khan, S. M. Marium, "Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection," In *Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Bombay, India, 29–31 March 2019, pp. 1–4.
32. P. Cunningham and S. J. Delany, "k-Nearest neighbour classifiers—A Tutorial," *ACM Comput. Surv. (CSUR)*, vol. 54, pp. 1–25, 2021.
33. R. E. Schapire, "Explaining adaboost," In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, Springer, Berlin/Heidelberg, Germany, 2013, pp. 37–52.
34. M. Burhanuddin, "Epigenetic modifications as predictive biomarkers for cardiovascular disease: A longitudinal study on DNA methylation patterns and disease risk", *SHIFAA*, vol. 2023, pp. 59–65, Jun. 2023, doi: 10.70470/SHIFAA/2023/007.
35. F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, and K.S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Inf. Fusion*, vol. 63, pp. 208–222, 2020.
36. M. A. Kadhim and A. M. Radhi, "Heart disease classification using optimized Machine learning algorithms," *Iraqi J. Comput. Sci. Math.*, vol. 4, pp. 31–42, 2023.

37. H. Kahramanli and N. Allahverdi, “Design of a hybrid system for the diabetes and heart diseases,” *Expert Syst. Appl.*, vol. 35, pp. 82–89, 2008.
38. R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” *Expert Syst. Appl.*, vol. 36, pp. 7675–7680, 2009.
39. M. A. Jabbar, B. Deekshatulu, and P. Chandra, “Classification of heart disease using artificial neural network and feature subset selection,” *Glob. J. Comput. Sci. Technol. Neural Artif. Intell.*, vol. 13, pp. 4–8, 2013.
40. S. Sa, “Intelligent heart disease prediction system using data mining techniques,” *Int. J. Healthc. Biomed. Res.*, vol. 1, pp. 94–101, 2013.
41. E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, “Heart diseases diagnosis using neural networks arbitration,” *Int. J. Intell. Syst. Appl.*, vol. 7, p. 72, 2015.
42. D. Shah, S. Patel, and S. K. Bharti, “Heart disease prediction using machine learning techniques,” *SN Comput. Sci.*, vol. 1, p. 345, 2020.
43. K. Drożdż, K. Nabrdalik, H. Kwiendacz, M. Hendel, A. Olejarz, A. Tomasik, W. Bartman, J. Nalepa, J. Gumprecht, and G.Y.H. Lip, “Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach,” *Cardiovasc. Diabetol.*, vol. 21, p. 240, 2022.
44. F. S. Alotaibi, “Implementation of machine learning model to predict heart failure disease,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, pp. 261–268, 2019.
45. N. Hasan and Y. Bao, “Comparing different feature selection algorithms for cardiovascular disease prediction,” *Health Technol.*, vol. 11, pp. 49–62, 2020.
46. A. K. Dubey, K. Choudhary, and R. Sharma, “Predicting heart disease based on influential features with machine learning,” *Intell. Autom. Soft Comput.*, vol. 30, pp. 929–943, 2021.
47. R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, “Enhanced heart disease prediction based on machine learning and χ^2 statistical optimal feature selection model,” *Designs*, vol. 6, p. 87, 2022.
48. A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, “A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms,” *Mob. Inf. Syst.*, vol. 2018, p. 3860146.
49. Maad M. Mijwil, Ioannis Adamopoulos, and Pramila Pudasaini, “Machine learning helps in quickly diagnosing cases of ‘New Corona’,” *Mesopotamian Journal of Artificial Intelligence in Healthcare*, vol. 2024, 2024, pp. 16–19, https://doi.org/10.58496/MJAIH/2024/003.