Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Advancing Cardiovascular Disease Prediction: An Interpretive Evaluation of Machine Learning and Deep Learning Models

Md Rahmathullah[1], Dr. S Nagakishore Bhavanam[2] and Dr. Vasujadevi Midasala[3]
[1] Research Scholar, Department of Electronics and Communications Engineering, University College of Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India
[2,3] Professor, Department of Computer Science and Engineering, Mangalayatan University Jabalpur, Madhya Pradesh, India
E-mail: rahmathullah@gmail.com

ARTICLE INFO

Received: 30 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

The heart is an important organ required for human beings. It is also the main organ of the cardiovascular system which pumps the blood to the body. The Cardiovascular disease (CVD) is the dominant cause of human death worldwide. Therefore, there is an urgent need to develop precise and practical predictive tools for understanding and diagnosing this disease in advance. Owing to advancements in Machine Learning (ML) and Deep Learning (DL), the accuracy of CVD prediction has significantly increased. Therefore, it offers a groundbreaking potential for early disease identification and provides individual treatment for patients. In this study, the prominent models used in ML such as Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), and K-nearest neighbors (KNN), and important DL models in DL such as artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNN), were implemented, compared, and critically analysed. By comparing and analysing different existing models, it is proposed that, by using multimodal data and hybrid models, the accuracy can be increased to the next highest benchmark. The paper concludes that the RF and DT models performed extraordinarily well with an accuracy of 98.5% for the dataset used in the present study.

Keywords: Artificial intelligence, machine learning, deep learning, cardiovascular disease, hybrid models, multimodal data, evaluation metrics, heart disease.

1. Introduction

Cardiovascular disease (CVD) is the dominant cause of human being deaths worldwide [1, 2], including in India [3], causing 17-18 million deaths roughly every year, representing about 32% of the annual worldwide fatalities. Therefore, there is an immediate and urgent need for accurate and early prediction of CVD. It also reduces health care costs and resources. Traditional diagnostic methods for diagnosing the disease have been used to determine CVD conditions. The lack of availability of shared datasets and the limited computational power are other reasons for this.

Currently, the availability of larger datasets and increased computational power combined with advancements in the Machine Learning (ML) and the Deep Learning (DL) techniques have revolutionized the predictive landscape. Accurate results can be obtained by enabling models to uncover complex nonlinear patterns in both simple and multimodal datasets. With the advent of various ML and DL models, it is challenging to determine the model that can be used to accurately predict CVD. In this study, a comparative analysis was performed for all of the most common ML and DL models, thus providing a path for future research and clinical implementation. The results were analyzed using the following criteria – Accuracy (A) , Receiver Operating Characteristics Area Under Curve (ROC-AUC), Precision (P), Recall (R), F1-Score (F1), and Confusion Matrix (CM).

2. Methods

The significant increase in mortality due to CVD emphasizes the importance of early detection to allow early guidance and management to reduce morbidity and mortality. Owing to further breakthroughs in Artificial Intelligence, ML and DL are now considered superior resources in the Artificial Intelligence field for heart disease

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

prediction. These methods allow for the analysis of enormous datasets, identification of key complicated patterns, and assistance from healthcare providers for diagnosing at risk patients.

2.1 Machine Learning Techniques for Heart Disease Prediction

Machine Learning (ML) approaches are widely used to predict CVD risk. Risk is predicted by considering various health parameters that are typically available in the form of a dataset. Free datasets are available on Kaggle for this study. Commonly used ML models include logistic regression, decision trees, random forests, support vector machines, and K-nearest neighbors.

2.1.1 Logistic Regression

Logistic Regression (LR) is the simplest and most common model used for predicting the heart disease. Instead, it is more useful for binary predictions, like presence or absence of CVD, given predictor factor variables. It assumes that the probability of a specific outcome can be modelled by a logistic function; therefore, it is interpretable and computationally efficient.

The research done by Zulkiflee et al. (2021) used the LR to assess the probability of heart disease. They applied three techniques: Binary Logistic Regression (BLR), BLR paired with the Least Quartile Difference (LQD) method, and BLR paired with the Median Absolute Deviation (MAD) method. The dataset included 271 patients with 12 characteristics. The MAD model delivered the best result (86.6%. Nevertheless, LR is a popular medical diagnostic method owing to its transparency and ease of implementation. [4]

According to Anshori et al. (2022), LR is used to know the heart disease, based on patient medical records. Iteration 14 using the LR model yielded a prediction accuracy of 81.35% (0.02 s). [5]

In another study by Kavya et al. (2023), where LR method was used, leveraging a dataset from Kaggle with over 4,000 records and 15 attributes and achieving an accuracy of 88%. [6]

Mathematically, the LR model calculates the probability $P(Y = 1 \mid X)$ of the existence of heart related disease given a set of predictor variables. The probability was modelled using a sigmoid function, as shown in Equation 1.

$$P(Y = 1 \mid X) = \frac{1}{1+e^{-z}} \qquad (1)$$

Where z represents a linear relationship of variables used to predict, as shown in Equation 2.

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \qquad (2)$$

The parameters of the model are calculated using the Maximum Likelihood Estimation (MLE) approach, which maximizes the likelihood function:

$$L(\beta) = \prod_{j=1}^{m} P\,(Y_j \mid X_j)^{Y_j}(1 - P(Y_j \mid X_j))^{(1-j)} \qquad (3)$$

where m represents the total number of observations.

Although LR has benefits, it also has limitations, specifically, its inability to manage nonlinear relationships or relationships between variables. Its predictive performance can be improved using polynomial features and interaction terms L1 (lasso) and L2 (ridge) regression.

2.1.2 Decision Trees and Random Forests

Decsion Trees (DT) is another popular model used to predict heart disease. The decision tree builds decision rules using a hierarchical partitioned approach for the dataset based on feature values to make class predictions. Such a structure provides DT with high interpretability and performance. However, Random Forests (RF), which are an ensemble of DT, are known to have better performance and accuracy than a single decision tree because they reduce the risk of overfitting.

Madhumita Pal et al. (2021) employed the RF algorithm to diagnose cardiac illness.. The dataset was obtained from Kaggle and contained 303 samples and 14 parameters, and it was written in Python Jupyter Notebook. The RF

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

algorithm achieved accuracy, sensitivity, and specificity of 86.904%, 90.558%, and 82.688%. Among several methods, RF performed the best in classifying heart diseases (diagnosis rate: 93.3%) . [7]

The study done by Kellen et al. (2023) proposed an enhanced CVD prediction model using the RF algorithm, which outperforms other models such as K-nearest neighbors (KNN), Support Vector Machines (SVM), and LR. The model attained an accuracy of 99%, which was significantly higher than that of any other model. The suggested model's excellent accuracy makes it an attractive option for early identification and avoidance of CVD. [8]

Mathematically, a decision tree recursively splits the dataset into subsets using criteria such as Gini impurity or entropy. The impurity measure for node t is given by Equation 4.

$$Gini(t) = 1 - \sum_{i=1}^{c} p_i^2 \qquad (4)$$

where p_i represents the proportion of instances belonging to class . Alternatively, entropy can be used as shown in Equation 5.

$$Entropy(t) = -\sum_{i=1}^{c} p_i \log_2 p_i \qquad (5)$$

A split was chosen to minimize the weighted sum of impurities for the child nodes, calculated using Equation6.

$$Gain = Impurity_{parent} - \sum_j \frac{|N_j|}{|N|} Impurity_{child_j} \quad (6)$$

where $|N_j|$ represent the number of occurrences in the child node and $|N|$ is the total number of occurrences in the parent node.

RF improves predictive accuracy and robustness by reducing overfitting. In RF, multiple DT are trained on bootstrapped samples of the dataset and predictions are aggregated through majority voting for classification, as shown in Equation 7.

$$\hat{Y} = \mathrm{mode}(Y_1, Y_2, \dots, Y_T) \qquad (7)$$

where Y_t represents the prediction of an individual decision tree in the ensemble.

DT offers simplicity and interpretability, whereas RF enhance the predictive performance by aggregating multiple models. Therefore, RF is the preferred choice for medical diagnosis.

2.1.3 Support Vector Machines

Another powerful classification model for CVD is the support vector machine (SVM). SVM are more useful when the data are not linearly separable, because they can convert input features into a space with higher dimensions in which a separate hyperplane can be found.

Mathematically, SVM computes the perfect hyperplane by solving the next optimization as given in Equation 8.

$$\min_{w,b} \frac{1}{2} || w ||^2 \qquad (8)$$

subject to Equation 9.

$$y_i(w \cdot x_i + b) \geq 1, \forall i \qquad (9)$$

where w is the weight.vector, b is the bias. term, and y_i i is the class label. For non linearly separatable data, SVMs use a kernel. function $K(x_i, x_j)$ to project data into a greater dimensional space, enabling better separation.

Linear, polynomial, radial basis function (RBF), and sigmoid kernels are among the most commonly utilized. The regularization parameter and kernel coefficient have a major impact on the model's performance.

Vijayashree et al. (2018) suggested a machine learning framework for feature selection in heart disease classification, leveraging an upgraded Particle Swarm Optimization (PSO) algorithm coupled with an SVM classifier. By varying these parameters, an accuracy of 88.22% was achieved. [9]

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The efficacy of Naive Bayes, KNN, and SVM algorithms in categorizing patients with heart disease was compared in the work done by Damayunita et al. (2022). Focusing on elements such as BMI, physical health, and sleep duration, the dataset comprised 18 characteristics from 320,000 medical records. At 92%, the SVM algorithm attained the best accuracy, followed by KNN (91%), and Naive Bayes (88%). This study recommends SVM as the most reliable method for predicting heart diseases. [10]

SVMs provide robust performance in heart disease classification, particularly when feature selection and kernel methods are applied optimally.

2.1.4 K-Nearest Neighbors

The K-nearest. neighbor (KNN) algorithm is alternately widely used approach for heart disease prediction. KNN classifies a target point depending on the most classes of its nearest neighbors, making it a simple yet effective method for many classification problems.

Mathematically, given a query instance x_q , the KNN algorithm finds the distance between x_q and all trained instances using a distance value, such as the Euclidean. distance, as shown in Equation 10.

$$d(x_i, x_q) = \sqrt{\sum_{j=1}^{n} (x_{i,j} - x_{q,j})^2} \qquad (10)$$

where $x_{i,j}$ represents the value of feature j for training instance i , and $x_{q,j}$ is the corresponding value for the query instance. The algorithm selects the K closest neighbors and assigns the most frequent class to x_q as shown in Equation 11.

$$\hat{Y} = \text{mode}(Y_1, Y_2, \ldots, Y_K) \qquad (11)$$

Assegie et al. (2021) used the KNN algorithm to propose a heart disease prediction model. He used the Kaggle dataset, which contains 1025 observations with almost equal numbers of heart disease positive and heart disease negative instances. In the evaluation phase, the model registered a 91.99% accuracy. [11]

Shorewala et al. (2021) investigated the early diagnosis of coronary heart disease (CHD) using multiple ML approaches, including KNN, Binary Logistic.Classification, and Naïve.Bayes, with a focus on ensemble methods like bagging, boosting, and stacking. The study used the CVD Dataset with 70,000 records. The bagged models enhanced the accuracy by 1.96%. The boosted models showed an average 73.4% accuracy where as the stacked model with different ensemble classifiers and models achieved the highest accuracy of 75.1%. [12]

The paper presented by Li et al. (2024) introduced a combination of RF and K-nearest neighbor (KNN) heart disease detection models. By combining these models, they achieved an added accuracy of 10.3% compared to the original KNN model. [13]

2.2 Deep Learning Approaches for Heart Disease Prediction

Deep Learning (DL) is considered as the subset of ML. The capability of DL in learning complex patterns from large datasets without manual feature extraction has attracted significant attention in heart disease prediction. These methods are particularly useful for large, high dimensional datasets.

2.2.1 Artificial Neural Networks

Artificial Neural Networks (ANN) form the basis of DL. ANNs are widely used to predict CVD. ANNs are constructed to resemble the human being's brain by using numerous complexly networked neuronal layers. Neurons add nonlinearity to the model by processing the weighted inputs and applying an activation function.

Mathematically, a neuron processes the input X by using a weighted sum, as shown in Equation 12.

$$z = \sum_{i=1}^{n} w_i \, x_i + b \qquad (12)$$

where w_i are the weights, x_i are the features input, and b , the bias term The output is processed using an activation function that can be a sigmoid or Rectified Linear Unit (ReLU) function., as given in Equation 13.

$$A(z) = \frac{1}{1+e^{-z}} \ \ or \ \ A(z) = max\,(0, z) \qquad (13)$$

Talukdar et al. (2023) studied the application of ANN to detect the cardiovascular illnesses. This study analyzed data from hospitals in Assam, India. The method based on ANNs using the Multi Layer Perceptron ANN with the back propagation algorithm has achieved 81% accuracy.[14]

Raniya et al. (2023) constructed DL diagnosis system using ANN to know the heart disease in advance. The accuracy achieved using ANN was 93.44%, beating classic ML models such as SVM by 7.5%. [15]

Mathematically, the backpropagation algorithm, which is a crucial part of ANN training, updates the weights using the gradient method to minimize the loss function, as shown in Equation 14.

$$\Delta w = -\eta \frac{\partial L}{\partial w} \qquad (14)$$

where η is the learning rate and L is the loss function, which is commonly chosen as the binary cross entropy for classification tasks as shown in Equation 15.

$$L(\, y, \hat{y} \,) = -\sum_{k=1}^{m} y_k \, log \; \hat{y}_k + (1 - y_k) log\,(1 - \hat{y}_k) \qquad (15)$$

2.2.2 Convolutional Neural Networks

Although Convolutional Neural Networks (CNN) are typically employed for image processing, they have also been effectively utilized for structured data, including heart disease prediction. CNNs use convolutional layers to find the hierarchical patterns provided by the input data, thereby applying trainable filters to identify the spatial dependencies in the dataset.

Mathematically, a convolution operation is defined, as shown in Equation 16.

$$S(i, j) = \sum_p \sum_q X(i - p, j - q) * Y(p, q) \qquad (16)$$

where X is the feature matrix of input, Y is the convolutional kernel, and $S(i, j)$ is the output feature map.

Chibueze et al. (2024) recommended a model based on CNN for the identifying the heart disease. Using Magnetic Resonance Imaging (MRI) data from Kaggle and a local hospital, the model obtained a 94% accuracy level. With four convolutional layers, the CNN design proved adept in classifying the MRI data of the heart. A cross valuation of ten fold validated the higher effectiveness of the model, with an 94.13% accuracy. [16]

Sajja et al. (2020) advised a similar model to predict CVD accurately. The model was evaluated against conventional ML techniques including LR, KNN, Naive Bayes, and SVM. Using the UCI ML Cleveland dataset, the model outperformed all the other techniques, with a prediction accuracy of 94.78%. [17]

Mathematically, CNNs use backpropagation to update weights, minimizing the loss function through the techniques for optimization like stochastic gradient descent (SGD) or the Adam optimizer:

$$\theta = \theta - \eta \nabla L(\theta) \qquad (17)$$

where η is the learning rate and $L(\theta)$ is the loss function.

2.2.3 Recurrent Neural Networks

In general, Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) networks process data which is sequential by maintaining hidden states over time. Unlike conventional models, LSTMs captures long term dependencies in data which is series in time using memory cells and gating mechanisms.

Omankwu et al. (2023) used an RNN and Gated Recurrent Units (GRU) to envisage the heart disease. The accuracy of was 98.6876%. By integrating the Synthetic Minority Oversampling Technique for balancing the data and applying the Adam optimizer for enhanced learning rates, the model solves the constraints of the current models, such as poor accuracy and data imbalance. [18]

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Alkhodari et al. (2021) used bidirectional long short term memory (BiLSTM) networks along with CNN to identify valvular heart disease (VHD) using phonocardiogram (PCG) recording. Using preprocessed data from 1,000 PCG recordings, wavelet smoothing, and z-score normalization, the model was trained using a cross valuation system of ten fold. The accuracy:, sensitivity:, and specificity: were 99.32%, 98.30%, and 99.58%, respectively. The CNN-BiLSTM model taken as whole performed best. [19]

Mathematically, an LSTM cell is defined as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \qquad (18)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \qquad (19)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \qquad (20)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh\,(W_c x_t + U_c h_{t-1} + b_c) \quad (21)$$

$$h_t = o_t \odot tanh\,(c_t) \qquad (22)$$

where i_t, f_t, o_t and are the input, forget, and output gates, respectively, and represents the cell state.

3. Results and Discussion

3.1 Description of the used dataset:

A freely available heart disease dataset from Kaggle [20] was used in this study. The size of the dataset is 1025 rows with 14 columns. The 13 columns contain independent attributes, and the last column is the target attribute. Table 1 gives a brief summary of the traits.

Column Number	Name	Information	Minimum value	Maximum Value
1	age	age in years	29	77
2	sex	(1 = male; 0 = female)	0	1
3	cp	chest pain type	0	3
4	trestbps=	resting blood pressure (in mm Hg on admission to the hospital)	94	200
5	chol	serum cholestoral in mg/dl	126	564
6	fbs	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)	0	1
7	restecg	resting electrocardiographic results	0	2
8	thalach	maximum heart rate achieved	71	202
9	exang	exercise induced angina (1 = yes; 0 = no)	0	1
10	oldpeak	ST depression induced by exercise relative to rest	0	6.2
11	slope	the slope of the peak exercise ST segment	0	2
12	ca	number of major vessels (0-3) colored by flourosopy	0	3
13	thal	Thalassemia= 1=Normal; 2 = fixed defect; 3 = reversable defect	1	3
14	target	Presence of disease =1, Absence of disease = 0	0	1

Table 1. *List of attributes in the dataset*

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The categorical attributes are shown in the Figure 1

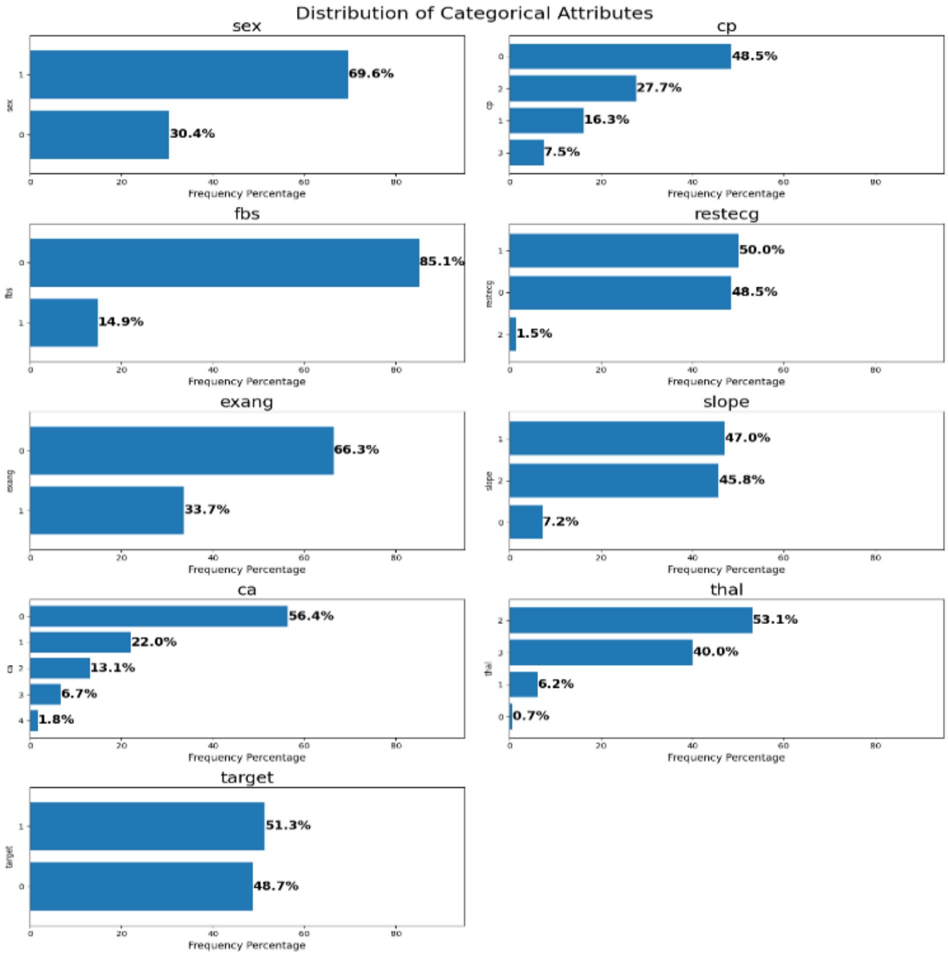

Figure 1. *Distribution of attributes*

The attribute importance is defined as the extent to which a attribute affects a model's predictions. The attribute and its relative medical importance is shown in Figure 2.

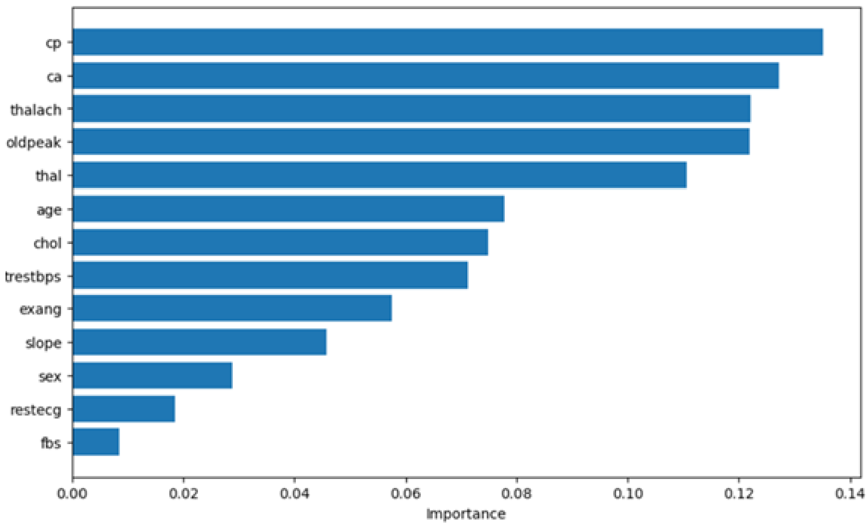

Figure 2. *Attribute Versus Importance*

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.2 Evaluation Metrics

The following metrics were evaluated to test the ML and DL models.

3.2.1 Accuracy or Classification Accuracy

The accuracy (ACC) was used to evaluate the performance of the model. This is calculated based on the number of correct predictions (NCP) divided by the total input samples taken (IS), as shown in Equation 23.

$$ACC = \frac{NCP}{IS} \qquad (23)$$

3.2.2 Receiver operating characteristics – Area Under Curve

Receiver operating characteristic (ROC) curves were used to determine the performance of the model to determine whether it was good at distinguishing between positive and negative outcomes. It plots the true positive rates (TPR) against the false positive rates (FPR) at different thresholds. The area under the ROC curve, Area Under Curve (AUC) measures the performance of the ML algorithms. The perfect model had an area under the curve of 1.

3.2.3 Precision

The model's performance is measured using Precision (PR). It expresses the number of positive predictions made by the model are actually positive. This is given by Equation 24.

$$PR = \frac{TP}{TP + FP} \qquad (24)$$

Where TP = $True\ Positive\ Predictions$ and FP = $False\ positive\ predicitons$

3.2.3 Recall

Recall (RC) is the given by the Equation (25)

$$RC = \frac{TP}{TP + FN} \qquad (25)$$

Where FN = $False\ Negative\ predicitons$

Lower Recall and higher precision yielded good accuracy.

3.2.4 F1-score

F1-score (F1) was used to tell about the model accuracy. This is the harmonic mean of PC and RC. Its value is between 0 and 1. This is given by Equation 26.

$$F1 = 2 \times \left(\frac{1}{\frac{1}{Precision} + \frac{1}{Recall}}\right) \qquad (26)$$

3.2.5 Confusion Matrix

The confusion matrix (CM) is an NXN matrix that helps assess the efficacy of the model. This stipulates a comparison between the estimated and the actual values for a given dataset. Figure 3 gives the basic structure of the 2X2 confusion matrix.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Positive cases predicted correctly as positive	Negative cases but predicted wrongly as positive
Positive cases but predicted wrongly as Negative	Negative cases predicted correctly as Negative

Figure 3. *Confusion Matrix*

3.3 Experiments and Results

In this paper, we focus on using the same dataset for predicting heart disease using the most commonly used ML and DL models and compare all of them collectively.

3.3.1 Using Machine Learning Techniques

Figure 4 illustrates the flowdiagram of CVD prediction using the ML models.

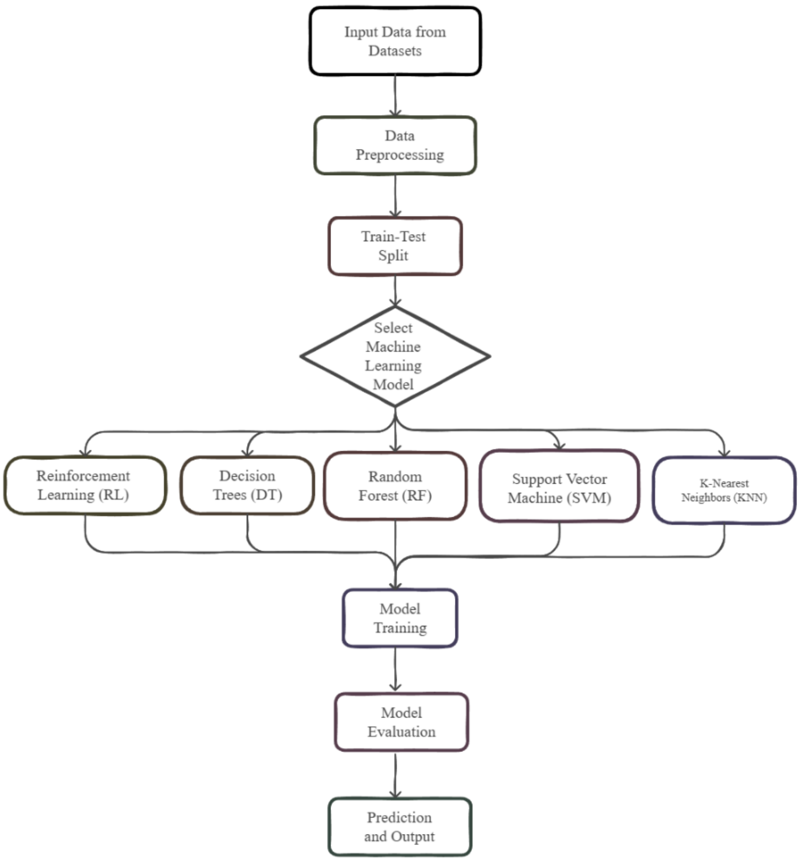

Figure 4. *Flow Diagram for predicting Heart Disease using ML*

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The dataset was divided into 80% to train the model and 20% to test the models. LR, DT, RF, SVM, and KNN models were implemented on Python on the Kaggle notebook, and the following results were obtained:

Screenshots of the results of various models are presented below. A comparative analysis is presented in the later part of this section.

```
Model: Logistic Regression
Accuracy: 0.7951219512195122
ROC-AUC: 0.8787359604035789
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.72      0.78       102
           1       0.76      0.87      0.81       103

    accuracy                           0.80       205
   macro avg       0.80      0.79      0.79       205
weighted avg       0.80      0.80      0.79       205
```

Figure 5. *Output of Logistic Regression*

From Figure 5, it can be observed that LR had an ACC of 79.5% and an ROC-AUC of 0.879. the PR and RC for both classes (0 and 1, respectively) were satisfactory. This can be considered reasonable. It can be noted that the model can still be optimized to improve the results.

```
Model: Decision Tree
Accuracy: 0.9853658536585366
ROC-AUC: 0.9854368932038835
Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       102
           1       1.00      0.97      0.99       103

    accuracy                           0.99       205
   macro avg       0.99      0.99      0.99       205
weighted avg       0.99      0.99      0.99       205
```

Figure 6. *Output of Decision Tree*

From Figure 6, it can be observed that the Decision Tree performed well and achieved an ACC of 98.5% and an ROC-AUC of 0.985. The PR and RC for both classes (0 and 1) were excellent, and the F1 score was close to 1, indicating minimal misclassifications. The hyperparameters were tuned using default. Therefore, it exhibits a higher performance.

```
Model: Random Forest
Accuracy: 0.9853658536585366
ROC-AUC: 1.0
Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       102
           1       1.00      0.97      0.99       103

    accuracy                           0.99       205
   macro avg       0.99      0.99      0.99       205
weighted avg       0.99      0.99      0.99       205
```

Figure 7. *Output of Random Forest*

It is clear from Figure 7, that RF performed extraordinarily, achieved an ACC of 98.5%, similar to the DT, and scored an ROC-AUC of 1. The slight increase in the ROC-AUC over the previous DT suggests that the ensemble method used in RF improved the ability of the model to correctly classify the data. The PR and RC for both classes (0 and 1) were similar to those of the DT, and the F1 score was close to 1. We can conclude that the RF and DT models perform similarly.

```
Model: SVM
Accuracy: 0.8878048780487805
ROC-AUC: 0.9631639063392347
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.83      0.88       102
           1       0.85      0.94      0.89       103

    accuracy                           0.89       205
   macro avg       0.89      0.89      0.89       205
weighted avg       0.89      0.89      0.89       205
```

Figure 8. *Output of SVM*

From Figure 8, it can be observed that the SVM achieved an ACC of 88.7% and an ROC-AUC of 0.963. The PR and RC for both classes (0 and 1, respectively) were satisfactory. The F1 score is close to 0.88. We can conclude that the SVM performs better than LR and KNN. The SVM is more suitable when linear separability exists.

```
Model: KNN
Accuracy: 0.8341463414634146
ROC-AUC: 0.9485532076908434
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.77      0.82       102
           1       0.80      0.89      0.84       103

    accuracy                           0.83       205
   macro avg       0.84      0.83      0.83       205
weighted avg       0.84      0.83      0.83       205
```

Figure 9. *Output of KNN*

From Figure 9, it can be observed that KNN achieved a lower ACC of 83.41% and an ROC-AUC of 0.949. The PR and RC for both classes (0 and 1, respectively) were satisfactory. The F1 score is close to 0.84. We can conclude that KNN underperforms in terms of accuracy because its sensitivity depends on the number of neighbors.

3.3.2 Using Deep Learning Techniques

A flow diagram for predicting heart disease using DL Techniques is given in Figure 10.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

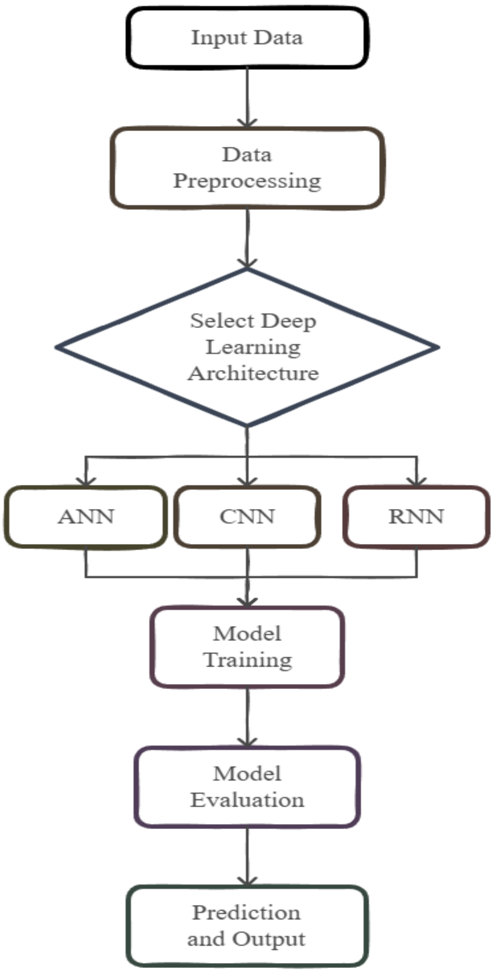

Figure 10. *Flow diagram of Deep Learning*

Even though some DL models work best with images such as CNNs and Sequential data such as RNNs, in this paper, the same dataset was used for comparing and analyzing with the ML models.

3.3.2.1 Using Artificial Neural Networks

The Artificial Neural Networks (ANN) model used in this study was a 64 layered Neural network. The Rectified Linear Unit (RELU) was the activation function used for learning deeper patterns. The output function used was a sigmoid function to perform binary classification in terms of the presence or absence of CVD.

```
Results for ANN:
Accuracy: 0.8341463414634146
ROC-AUC: 0.833904435560632
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.78      0.82       102
           1       0.81      0.88      0.84       103

    accuracy                           0.83       205
   macro avg       0.84      0.83      0.83       205
weighted avg       0.84      0.83      0.83       205
```

Figure 11. *Output of ANN*

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

It can be noticed from Figure 13 that the ANN has a decent ACC of 83.41% and an ROC-AUC of 0.834. The PR and RC for both classes (0 and 1, respectively) were satisfactory. The F1 score is close to 0.84. It can be observed that its performance is generally lower than that of the DT and RF methods. We conclude that the lower performance of ANN is due to inadequate network architecture, or the model requires more data for training itself.

```
Results for CNN:
Accuracy: 0.775609756097561
ROC-AUC: 0.7753188654102419
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.72      0.76       102
           1       0.75      0.83      0.79       103

    accuracy                           0.78       205
   macro avg       0.78      0.78      0.77       205
weighted avg       0.78      0.78      0.77       205
```

Figure 12. *Output of CNN*

Figure 12 gives the alarming observation that the CNN showed a lower ACC of 77.56% and an ROC-AUC of 0.775. The PR and RC values for both classes (0,1 respectively) were average. The F1 score is close to 0.76. It can be observed that its performance is generally lower than that of the previous models. We conclude that the lower performance of CNN is because its architecture is well suited for image based datasets. It is not intended to be used for structured or tabular datasets, as in this paper. The CNN model could be trained using a extensive dataset to train and obtain better results.

```
Results for RNN:
Accuracy: 0.751219512195122
ROC-AUC: 0.7512373881591472
Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.75      0.75       102
           1       0.75      0.75      0.75       103

    accuracy                           0.75       205
   macro avg       0.75      0.75      0.75       205
weighted avg       0.75      0.75      0.75       205
```

Figure 13. *Output of RNN*

Figure 13 also gives the observation that the RNN has scored the lowest ACC of 75.12% and an ROC-AUC of 0.751. The PR and RC values for both classes (0 and 1) were just average. The F1 score was 0.75. It can be observed that its performance is the lowest compared to those of all previous models. We conclude that the lower performance of the RNN is because the architecture of the CNN is well suited for the sequential dataset. It could not be used for structured or tabular datasets in this experiment. The model also requires additional training and fine tuning.

3.3.3 Discussion of the Results with Graphical Analysis

3.3.3.1 Accuracy Bar Graph:

The Bar Graph shown in Figure 14 clearly compares all the ML and DL models in one figure, and it can be perceived that the DT and RF models achieved superiority in terms of accuracy.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

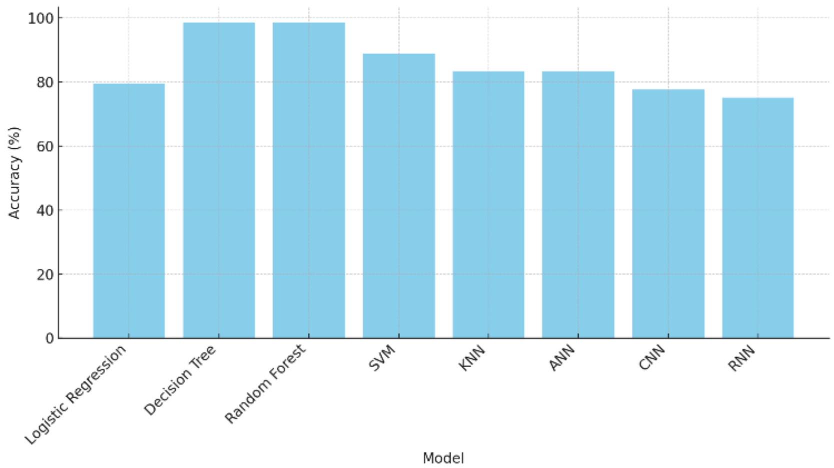

Figure 14. *Accuracy Comparison of ML and DL Models*

3.3.3.2 ROC-AUC Line Plot:

The Line Plot shown in Figure 15 clearly depicts all the ML and DL models in one figure, and it can be noticed that with respect to the ROC-AUC, the DT and RF models achieved superiority evaluated to all the other models.

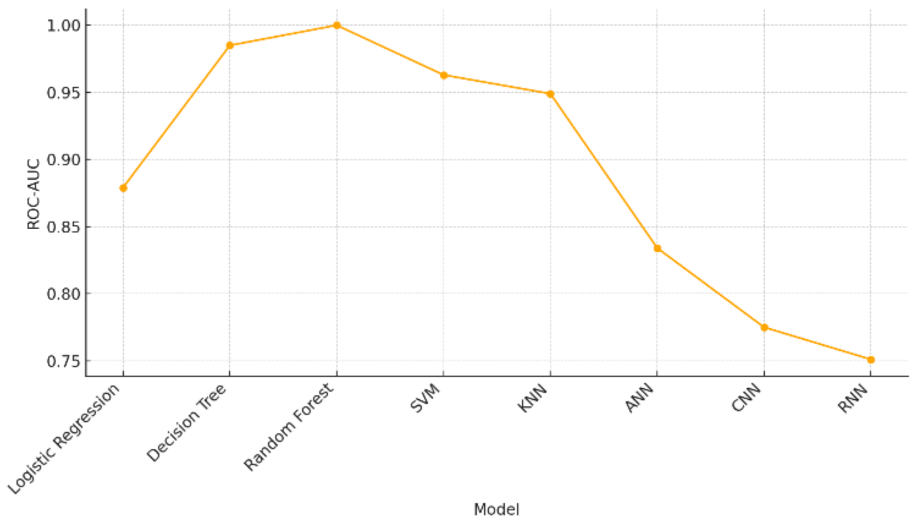

Figure 15. *ROC-AUC Comparison of ML and DL Models*

3.3.3.3 F1- Score bar Graph:

The F1-Score Bar Graph shown in Figure 16 is used to compare all the ML and DL models, and it can be witnessed that the DT and RF models have achieved superiority in terms of the F1-Score.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

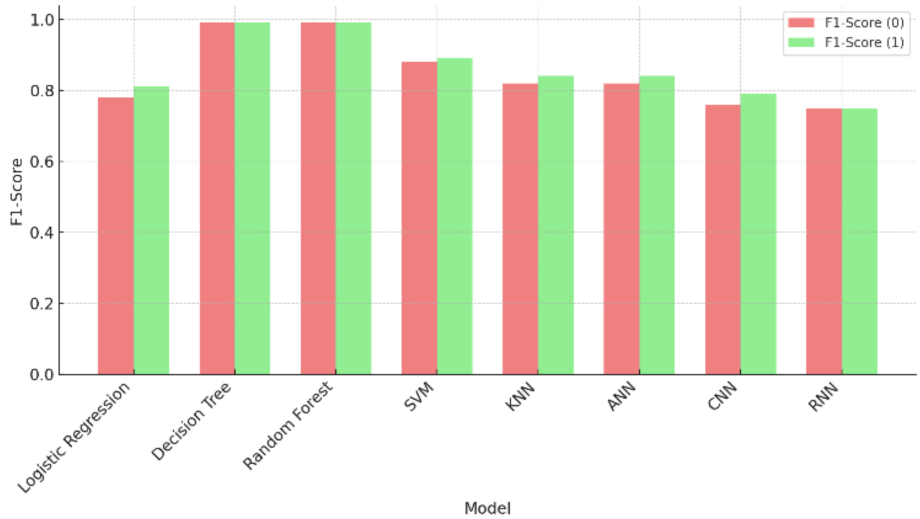

***Figure 16.** F1-Score Comparison of ML and DL Models*

3.3.3.4 Confusion Matrix:

Figure 17 displays the confusion matrix for all models. From the confusion matrix, we can see that the Decision Tree and RF models have the best performance among all models, with the highest accuracy (98.44%), perfect precision (1.00), and excellent recall (96.96%).

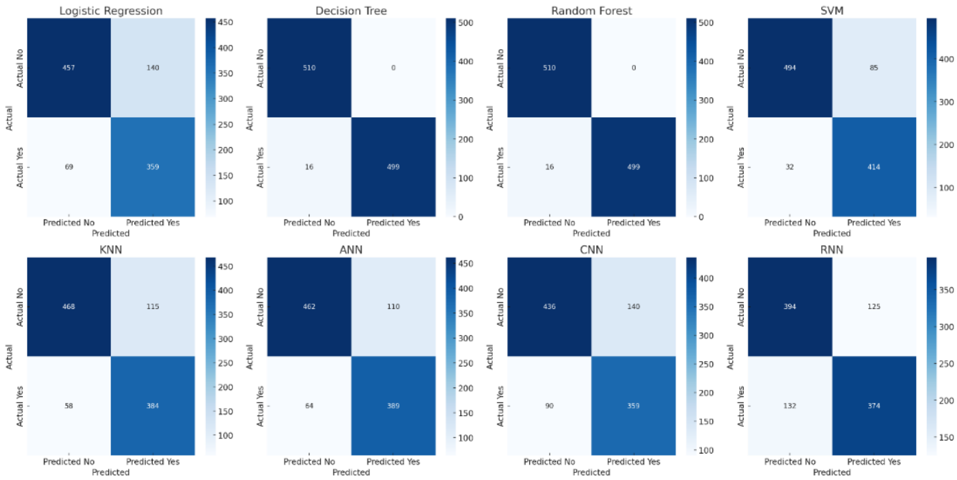

***Figure 17.** Confusion mtrix*

3.3.3.5 Summary Table:

Table 2 provides a summary of all evaluation metrics for all the models discussed above.

Metric	LR	DT	RF	SVM	KNN	ANN	CNN	RNN
ACC	79.51	98.54	98.54	86.78	83.41	83.41	77.56	75.12
ROC-AUC	0.879	0.985	0.985	0.931	0.93	0.94	0.78	0.71
PR (Class 0)	0.85	0.97	0.97	0.89	0.93	0.93	0.82	0.83
RC (Class 0)	0.72	1	1	0.83	0.87	0.91	0.92	0.88
PR (Class 1)	0.76	1	1	0.83	0.87	0.86	0.73	0.78
RC (Class 1)	0.87	0.97	0.97	0.94	0.88	0.88	0.68	0.62
F1 (Class 0)	0.78	1	1	0.86	0.9	0.9	0.87	0.83
F1 (Class 1)	0.81	0.97	0.97	0.88	0.87	0.84	0.7	0.75

***Table 2.** Comparison of all models together*

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4 Future Directions in Heart Disease Predictions: Hybrid Models and Multimodal Approaches

There is much scope for developing the prediction of CVD using Hybrid Models (HM) that are made by combining the ML and DL together.

Mohammead et al. (2023) suggested a HM of ML for predicting the heart disease. The model combines a continuous wavelet transform (CWT) with a CNN to analyze electrocardiogram data and achieves an ACC of 97.2% in early heart disease identification. [21]

A Hybrid RF with Linear Model (HRFLM) was introduced by Mohan et al. (2019) to improve the ACC of heart disease prediction. The ACC of 88.4% was recorded.[22]

Shiwlani et al. (2024) discussed the integration of multimodal data using DL methods to diagnose heart disease. Different data modalities, such as clinical data, electrocardiograms, and imaging modalities, have been combined to enhance the diagnosis of heart disease. [23]

Future research can combine multimodal (MM) data, demographic data, ECG signals, and imaging data, and using hybrid models combining traditional ML and DL techniques will guide to substantial improvements in the prediction ACC. These HM use the strength of traditional ML and DL models for interpretability while harnessing DL capabilities to capture complex nonlinear relationships.

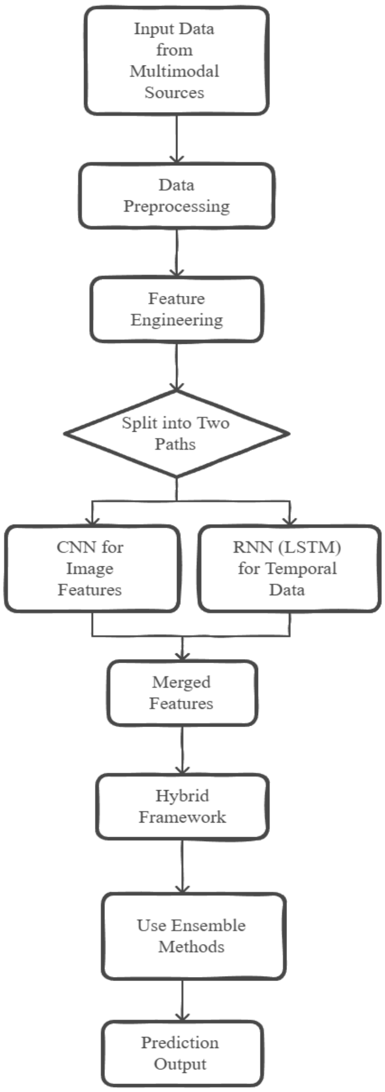

Figure 18. *Flow diagram for the proposed Multimodal Hybrid Model diagram*

Figure 18 shows the flow diagram for using the multimodal hybrid model to improve the early identification of the heart disease.

Future research on MM HM learning should explore optimizing architectures and feature fusion techniques to improve the accuracy of heart disease prediction further.

5. Conclusion

After comparing and evaluating a diverse set of ML and DL models for predicting the heart disease, it is clear that classical models such as DT and RF perform better in terms of ACC, ROC-AUC, and F1. These models have the capability to be used with very high accuracy (98.5%) and high PR and RC and are consequently highly effective at discriminating between cases of heart disease. SVM and LR show moderate performance with fair ROC-AUC and balanced PR-RC but are nowhere near the ensemble methods. The DL models, ANN, CNN, and RNN show poorer performance overall, which is likely a result of the increased susceptibility to the small size and level of feature abstraction in the dataset in which these models may underfit. Subsequently, simpler models were more robust. The CM also showed that models such as DT and RF had fewer misclassifications, making them more appropriate choices for practical heart disease prediction tasks. In conclusion, ensemble approaches such as RF and DT are the best choices for heart disease identification in terms of both prediction accuracy and model interpretation, which are crucial in the field of human health care.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not for profit sectors.

Conflict of interest

The authors have no conflicts of interest to declare.

Data access statement

The Kaggle dataset is used in this paper, which is freely accessible. It can be accessed from:
www.kaggle.com/datasets/johnsmith88/heart-disease-dataset/data

Ethics statement

This study did not involve any subjects that required ethical approval.

Information consent statement

This work did not involve any objects that required informed consent from any one.

Clinical trials

This study did not involve human clinical trials.

Author contribution

Md Rahmathullah: Concept, experimentation, analysis, editing; Dr. S Nagakishore Bhavanam: Reviewing and Supervision; Dr. Vasujadevi Midasala: Writing

References

[1] McAloon, C. J., Boylan, L. M., Hamborg, T., Stallard, N., Osman, F., Lim, P. B., & Hayat, S. A. (2016). The changing face of cardiovascular disease 2000–2012: An analysis of the world health organisation global health estimates data. International journal of cardiology, 224, 256-264. https://doi.org/10.1016/j.ijcard.2016.09.026

[2] McFarlane, S. I., Jean-Louis, G., Zizi, F., Whaley-Connell, A. T., Ogedegbe, O., Makaryus, A. N., & Maraj, I. (2012). Hypertension in the high-cardiovascular-risk populations. International journal of hypertension, 2011, 746369. doi:10.4061/2011/746369

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

[3] Jan, B., Dar, M. I., Choudhary, B., Basist, P., Khan, R., & Alhalmi, A. (2024). Cardiovascular diseases among Indian older adults: A comprehensive review. Cardiovascular Therapeutics, 2024(1), 6894693. https://doi.org/10.1155/2024/6894693
[4] Zulkiflee, N. F., & Rusiman, M. S. (2021). Heart Disease Prediction Using Logistic Regression. Enhanced Knowledge in Sciences and Technology, 1(2), 177-184.
[5] Anshori, M., & Haris, M. S. (2022). Predicting heart disease using logistic regression. Knowledge Engineering and Data Science (KEDS), 5(2), 188-196.
[6] Kavya, S. M., PrathanyaSree, C., Deepasindhu, M., Nowshika, B., & Shijitha, R. (2023). Heart Disease Prediction Using Logistic Regression. Journal of Coastal Life Medicine, 11, 573-579.
[7] Pal, M., & Parija, S. (2021, March). Prediction of heart diseases using random forest. In Journal of Physics: Conference Series (Vol. 1817, No. 1, p. 012009). IOP Publishing. doi:10.1088/1742-6596/1817/1/012009
[8] Sumwiza, K., Twizere, C., Rushingabigwi, G., Bakunzibake, P., & Bamurigire, P. (2023). Enhanced cardiovascular disease prediction model using random forest algorithm. Informatics in Medicine Unlocked, 41, 101316. https://doi.org/10.1016/j.imu.2023.101316
[9] Vijayashree, J., & Sultana, H. P. (2018). A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. Programming and Computer Software, 44, 388-397. https://doi.org/10.1134/S0361768818060129
[10] Damayunita, A., Fuadi, R. S., & Juliane, C. (2022). Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients. Jurnal Online Informatika, 7(2), 219-225. https://doi.org/10.15575/join.v7i2.919
[11] Assegie, T. A. (2021). Heart disease prediction model with k-nearest neighbor algorithm. International Journal of Informatics and Communication Technology (IJ-ICT), 10(3), 225. http://doi.org/10.11591/ijict.v10i3.pp225-230
[12] Shorewala, V. (2021). Early detection of coronary heart disease using ensemble techniques. Informatics in Medicine Unlocked, 26, 100655. https://doi.org/10.1016/j.imu.2021.100655
[13] Li, B., & Bao, H. (2024, January). A Heart Disease Detection Model Based on Random Forest and KNN. In Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering (pp. 656-659). https://doi.org/10.1145/3672758.3672867
[14] Talukdar, J., & Singh, T. P. (2023). Early prediction of cardiovascular disease using artificial neural network. Paladyn, Journal of Behavioral Robotics, 14(1), 20220107. https://doi.org/10.1515/pjbr-2022-0107
[15] Sarra, R. R., Dinar, A. M., & Mohammed, M. A. (2023). Enhanced accuracy for heart disease prediction using artificial neural network. Indonesian Journal of Electrical Engineering and Computer Science, 29(1), 375-383. http://doi.org/10.11591/ijeecs.v29.i1.pp375-383
[16] Chibueze, K. I., Didiugwu, A. F., Ezeji, N. G., & Ugwu, N. V. (2024). A CNN based model for heart disease detection. Scientia Africana, 23(3), 429-442. https://doi.org/10.4314/sa.v23i3.38
[17] Sajja, T. K., & Kalluri, H. K. (2020). A Deep Learning Method for Prediction of Cardiovascular Disease Using Convolutional Neural Network. Rev. d'Intelligence Artif., 34(5), 601-606. https://doi.org/10.18280/ria.340510
[18] Omankwu, O. C., & Ubah, V. I. (2023). Hybrid deep learning model for heart disease prediction using recurrent neural network (RNN). NIPES-Journal of Science and Technology Research, 5(2). https://doi.org/10.5281/zenodo.8014330
[19] Alkhodari, M., & Fraiwan, L. (2021). Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. Computer Methods and Programs in Biomedicine, 200, 105940. https://doi.org/10.1016/j.cmpb.2021.105940
[20] Lapp, D. (2019, June 6). Heart disease dataset. Kaggle. http://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset/data
[21] Mohammad, F., & Al-Ahmadi, S. (2023). WT-CNN: a hybrid machine learning model for heart disease prediction. Mathematics, 11(22), 4681. https://doi.org/10.3390/math11224681
[22] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, 81542-81554. https://doi.org/10.1109/ACCESS.2019.2923707

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[23] Shiwlani, A., Ahmad, A., Umar, M., Dharejo, N., Tahir, A., & Shiwlani, S. (2024). Analysis of multi-modal data through deep learning techniques to diagnose CVDs: A review. International Journal, 11(1), 402-420. https://doi.org/10.15379/ijmst.v11i1.3659