RESEARCH



Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance



M Darshan Teja¹ and G Mokesh Rayalu^{1*}

Abstract

Cardiovascular disease is the leading cause of mortality globally, necessitating precise and prompt predictive instruments to enhance patient outcomes. In recent years, machine learning methodologies have demonstrated significant potential in enhancing the precision and efficacy of health-related predictions, especially in the identification of heart disease. The dataset used in this study came from the UC Irvine Machine Learning Repository and included data from Cleveland, Switzerland, Hungary, Long Beach, and Statlog. We selected seven of the 1,190 cases, each with 12 attributes, for analysis. We used different machine learning models, like Random Forest, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, Gradient Boosting, AdaBoost, XGBoost, and Bagged Trees, to check performance using accuracy, precision, recall, F1-score, and ROC-AUC. K-fold cross-validation (K=10, K=5) was conducted to guarantee the robustness and generalizability of these models. Random Forest exhibited remarkable stability, attaining 94% accuracy with K=10 and 92% with K=5, whereas XGBoost had a minor decrease during cross-validation (90% for K=10, 89% for K=5). KNN demonstrated possible overfitting, evidenced by a notable decline in accuracy (71% for K=10, 72% for K=5). XGBoost and Bagged Trees achieved the highest accuracy of 93%, followed by Random Forest and KNN at 91%. Furthermore, Random Forest and Bagged Trees exhibited the highest ROC-AUC values at 95%, and XGBoost demonstrated a ROC-AUC of 94%. The results demonstrate the effectiveness of ensemble methods in predicting cardiac diseases, along with the potential for future advancement through the incorporation of hybrid models and advanced survival analysis techniques.

Keywords Heart diseases, Machine learning models, Confusion matrix, ROC-AUC, Metrics

Introduction

Cardiovascular disorders (CVDs) are the primary cause of mortality worldwide, resulting in 17.9 million deaths per year. Heart attacks and strokes are the leading causes, with one-third occurring prematurely in people under 70 [1]. Behavioral risk factors include unhealthy diet,

G Mokesh Rayalu

mokesh.g@vit.ac.in

¹Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, India



inactivity, tobacco use, alcohol, and air pollution. Heart failure, stroke, and heart attacks are more likely in people with intermediate risk factors. In Cleveland, unhealthy diets, smoking, alcohol, fat, and tobacco use contribute to a 50–60% mortality rate [2]. In Hungary, cardiovascular diseases are among the most common causes of death, with 60,000 deaths annually [3]. In Switzerland, cardiovascular disease and cancer were the most common killers in 2022 [4]. Comprehensive datasets from Cleveland, Hungary, Switzerland, Long Beach, and Statlog demonstrate the global burden of cardiac disease and

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

^{*}Correspondence:

continuous attempts to study and treat these disorders through modern medical research and data analysis [5].

Machine learning (ML) is an effective method for analysing complicated information in a variety of industries, including healthcare. Machine learning (ML) approaches use training algorithms to find patterns and make predictions based on data without explicit programming [6]. A wide array of ML models, including Logistic Regression, Random Forest, Support Vector Machines, K-Nearest Neighbours, Gradient Boosting Machines, Neural Networks, XGBoost, Model Averaged Neural Networks, Flexible Discriminant Analysis, Conditional Inference Trees, Bagged Trees, Naive Bayes, Multivariate Adaptive Regression Splines, Boosted Generalised Linear Models, Random Trees, and Bayesian Generalised Linear Models, have been applied to various datasets to extract [7].

Machine learning techniques are critical for increasing the accuracy and efficiency of identifying cardiac disease. By analysing massive datasets, ML models can forecast the onset of cardiac disease, identify high-risk individuals, and recommend the best treatments. Performance criteria like as accuracy, recall, precision, and ROC-AUC are utilised to assess these models. High accuracy reflects the model's capacity to correctly classify examples; recall evaluates the model's ability to identify all relevant cases; precision evaluates the number of genuine positive findings; and ROC-AUC offers an overall measure of performance across all classification thresholds.

Through the application of cutting-edge machine learning methods, this research aims to provide a thorough examination of cardiac disease prognosis. The literature on heart disease prediction is reviewed in the Review of literature section, with an emphasis on earlier approaches and their results. The technique used for this study is described in the Methodology section, along with the steps involved in gathering data and the features that were selected. The dataset, together with its sources, features are all described in Data set. The measurements and descriptions of the machine learning models used are described in the section of Methods. The study and findings, which highlight the various models' performances, are presented in Analysis and Results. In Discussion section, the results are compared with earlier studies, highlighting the improvements achieved in this investigation, makes recommendations for further development and finally we concludes in the last section.

Review of literature

Numerous academic contributions have investigated cardiac disease detection using data mining and machine learning approaches, drawing on datasets from Cleveland and other nations. Researchers have used a variety of methodologies, including ensemble

Chandrasekhar et al. (2023) employed machine learning to increase cardiovascular disease forecasting accuracy. They employed six approaches: random forest, K-nearest neighbour, logistic regression, Naïve Bayes, gradient boosting, and AdaBoost classifier. The Cleveland dataset produced the greatest results, with logistic regression reaching 90.16% accuracy. AdaBoost scored 90% accuracy on the IEEE Dataport dataset. A softly selected ensemble classifier integrating all six techniques improved performance to 93.44% and 95%, outperforming logistic regression and AdaBoost classifiers [8]. Perumal et al. (2020) They focus evaluated three classifiers: LR, SVM, and KNN, focusing on their performance metrics. LR and SVM outperformed the KNN classifier in terms of accuracy, sensitivity, specificity, and MCC. The LR classifier had higher sensitivity, specificity, accuracy, MCC, and error rate, while the SVM had lower performance [9].

Gao et al. (2021) created a system for predicting heart illness that employs ensemble approaches such as boosting and bagging, as well as PCA and Linear Discriminant Analysis are two examples of feature extraction methods. They utilised the Cleveland heart disease dataset to evaluate the performance of five classifiers. The bagging ensemble learning technique, when paired with Decision Trees and PCA feature extraction, produced the greatest results, with an amazing accuracy of 98.6%. This demonstrates the efficacy of integrating ensemble approaches with feature extraction algorithms for heart disease prediction [10]. kahramanli et al. (2008) The suggested medical categorization system was tested using the Pima Indian diabetes and Cleveland heart disease datasets. The performance measures employed were accuracy, sensitivity, and specificity. The technique obtained 84.24% and 86.8% accuracy for the heart disease datasets, respectively, which were among the best when compared to prior research and UCI websites [11].

Hassan D et al. (2023) suggest a novel technique for heart disease prediction based on a pre-trained DNN, PCA, and logistic regression. The Cleveland dataset had been used to assess the effectiveness of the suggested strategy. Experimental findings demonstrated that the suggested technique worked well on both training and testing data, with 91.79% and 93.33% accuracy rates, respectively [12]. Samuel et al. (2016) novel decision support system, which combines ANN and Fuzzy AHP approaches, was evaluated on 297 heart failure patients using an online clinical dataset. The system attained an average prediction accuracy of 91.10%, which is 4.40% greater than the standard ANN technique. It also outperformed seven earlier approaches, which ranged in accuracy from 57.85 to 89.01% [13].

Rahman et al. (2018) investigated the Cleveland heart disease dataset and used a variety of approaches to predict heart disease. They used Naive Bayes with a threshold of 0.5, resulting in an accuracy of 88.35%, and with a threshold of 0.4, they increased their accuracy to 89.32%. Logistic regression was also used, resulting in an accuracy of 84.47%. Furthermore, a Neural Network technique achieved a noteworthy accuracy of 90.2%. The suggested ensemble approach, RIHDPS, was the study's most important contribution, with a maximum accuracy of 91.26%. This demonstrates how ensemble approaches improve forecast accuracy for heart disease [14].

For instance, B Shi et al. (2022) suggested a JASMA-SVM framework that uses clinical vitamin D and thyroid data to accurately predict recurrent spontaneous abortion. X Fei et al. (2020) added sparse learning to MEKLM to improve the use of neuroimaging to diagnose Alzheimer's and Parkinson's diseases. In the field of computer vision, L. Zang et al. (2021) created a deep active learning framework for scene classification that is based on biological inspiration. It uses object-based segmentation and multimodal feature fusion to get good results on remote sensing datasets. Additionally, Chen, M. R., Zeng, G. Q., & Lu, K. D. (2019) presented the MaOPEO-HM algorithm, a novel many-objective optimization approach with adaptive hybrid mutation, which outperformed several state-of-the-art evolutionary algorithms on benchmark problems [31].

Table 1 summarizes the work of many authors on heart disease data, including research that used the same dataset, as well as their findings. However, in our study, we focused on a subset of features are sex, chest pain type, fasting blood sugar, resting ECG, exercise angina, ST slope, and target. Unlike prior research, which failed to find ROC-AUC, our study contains this critical statistic. By adding ROC-AUC, we can provide a more thorough evaluation of the model's performance, filling a gap in the literature and establishing a more rigorous framework for heart disease prediction. This inclusion of ROC-AUC guarantees that the models' diagnostic capacity is properly evaluated, increasing the dependability of the results.

Methodology

This study used data from the UCI Repository, which included five separate heart disease datasets: Cleveland, Switzerland, Hungarian, Long Beach VA, and Statlog. The merged dataset comprised of 1190 records and 12 characteristics. For the purpose of this study, only seven relevant characteristics were chosen for investigation. The dataset was divided into training and testing sets with an 8:2 ratio. Following the data splitting, 15 machine learning models were applied to the dataset. The performance of these models was measured using a variety of measures, including accuracy, precision, F1 score, recall, confusion matrix, and ROC-AUC. This extensive

 Table 1
 An overview of cutting-edge strategies for predicting cardiovascular illnesses. Individual classifiers

R. No	Author	Year	Data Set	Methods	Results
					(Accuracy)
[7]	Hassan et al.	2024	Cardiovascular diseases dataset	XGboost, Random Forest	85.23, 86.36
[15]	Almazroi et al.	2023	Cleveland, Hungarien, Switzerland, Long Beanch HDD	Deep Learning	83
[12]	Hassan D et al.,	2023	Cleveland HDD	DNN+PCA+LR	91.79–93.33
[9]	Perumal et al.;	2020	Coronary heart diseases from Cleveland	LR, SVM, KNN	85,87,69
[13]	Samuel et al.;	2016	Heart Failure Rate	ANN, Fuzzy AHP	91.10, 89.01
[10]	Gao et al.;	2021	Heart Diseases	PCA+DT	98.6
[14]	Rahman et al.;	2018	Cleveland heart diseases data set	Navie Bayes (0.4 and 0.5), LR, Neural Network, RIHDPS	89.32, 88.35, 84.47, 90.21, 91.26
[11]	Kahramanli et al.;	2008	Pima Indian Diabetes and Clever data set	ANN, Fuzzy Neural Network	84.24, 86.8
[8]	Chandrasekhar, N., & Ped- dakrishna, S.	2023	Clever HDD sets and IEEE Dataport data set	LR, AdaBoost	90.16,90
[16]	Singh, A., & Kumar, R.	2020	UCI repository dataset	SVM, DT, Linear Regression	83, 79, 78
[17]	Maji, S., & Arora, S.	2019	UCI respository	DT, ANN	78,77
[18]	Senthilkumar et al.	2019	UCI respository heart diseases dataset	HRLM	88.7
[19]	Amin et al.	2019	Heart diseases data set	VOTE(NB + LR)	87.4
[20]	Chun-an cheng & Hung- wen chiu	2017	UCI laboratory heart diseases data	Hybrid model for NN, DT, SVM, Naïve bayes	86.8
[21]	Reddy, G. T., & Khare, N.	2017	Cleveland, Hungarian and Switzerland datasets	Radial Basis Network Link Network	78
[22]	Khateeb, N., & Usman, M.	2017	HDD	KNN	80

examination enabled a detailed comparison of the models' performance in predicting cardiac disease as overall process is shown in the Fig. 1.

Data set

The heart disease dataset utilised in this work was acquired from the UC Irvine Machine Learning Repository and Kaggle, and it included 12 features and 1,190 entries. The collection is made up of data from several sites, including Cleveland (303 records), Hungary (294 recordings), Switzerland (123 records), Long Beach, Va (200 records), and Stalog (270 records). Combining

all of these datasets yielded a comprehensive dataset of 1,190 records. For our research, we chose the seven most important variables from the original 12, with an emphasis on those that make a major contribution to heart disease prediction. This meticulous selection is intended to improve the performance and accuracy of the various machine learning models used in this study.

We selected seven key variables for analysis from Table 2, which contained dataset details: sex, chest pain kind, fasting blood sugar, resting ECG, exercise angina, ST slope, and Target. These characteristics were chosen based on their relation to heart disease prognosis. Using



S.No	Variable Name	Description and range	Feature Type
1.	Age	Age 30 to 90	Integer
2.	Sex	0-female, 1-male	Binary
3.	Chest-pain	Chest pain type 1 to 4	Nominal
4.	Rest-bp	Resting Blood Pressure 90 to 180	Integer
5.	Serum-chol	Serum Cholestrol mg/dl [126,564]	Integer
б.	Fastingbloodsugar (FBS)	FBS above 120 mg/dl [1=true, 0=false]	Binary
7.	Electrocardiographic (ECG)	Resting electrocardiographic results	Nominal
8.	Max-heart-rate	Maximum heart rate achieved 90 to 180	Integer
9.	Angina	exercise induced angina [0=no, 1=yes]	Binary
10.	Old peak	oldpeak = ST depression induced by exercise relative to rest	Integer
11.	Slope	the slope of the peak exercise ST segment	Normal
12.	Target	0-No, 1-Yes	Binary

Table 2 Data set description

these chosen characteristics, multiple machine learning algorithms were applied to the dataset to predict heart disease outcomes. This concentrated strategy is intended to improve the predicted accuracy and reliability of the models by focusing on the most useful factors.

Methods Metrics

Confusion matrix:

A confusion matrix is a tabular format that evaluates the effectiveness of a classification model. It juxtaposes the anticipated classifications with the actual classifications, yielding insights about the model's performance efficacy. The matrix comprises four fundamental elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

$$CM = \left[\begin{array}{cc} TP & FN \\ FP & TN \end{array} \right]$$

Accuracy:

Accuracy is the fraction of accurately predicted cases (including true positives and true negatives) among all instances.

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

Precision:

Precision is also known as positivity. The predictive value is the proportion of real positive forecasts among all positive ones.

$$Precision = \frac{TP}{TP + FP}$$
(2)

Recall:

Recall, also known as responsiveness or true positive rate, is a metric that indicates the proportion of real positive forecasts across all positives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1Score:

The score of F1 is the harmonic average of precision and recall, resulting in a balance of the two measures [23].

$$F1 \ score = 2 \ X \frac{Precision \ X \ Recall}{Precision + Recall}$$
(4)

ROC-AUC (Receiver Operating Characteristic- Area Under the curve)

The Recall (TPR) and False Positive Rate (FPR) are contrasted on the ROC curve. The classifier's performance is summarised at every level of classification by the AUC (Area Under the Curve) [24].

$$AUC = \int_{0}^{1} TPRd \left(FPR\right) \tag{5}$$

Models

A statistical model called logistic regression uses binary classification to forecast if an input is part of a particular class or not. The logistic function is employed to translate the output into probability values [20].

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$$
(6)

Random forest:

Logistic regression:

As part of a group learning technique called random forest, many decision trees are built during training and the mean estimate for regression of each tree or the mode of the categories for classification are returned.

$$\widehat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x)$$
(7)

Where $\hat{f}(x)$ is the aggregated prediction from B trees [22].

SVR (Support Vector Regression).

SVR is a support vector machine that solves regression issues. It seeks a function that deviates by no more than e from the actual goal values.

$$f(x) = \sum_{i=1}^{n} (\alpha_{i} - \alpha_{i}^{*}) K(x_{i}, x) + b \qquad (8)$$

Where $K(x_i, x)$ is the Kernel function.

KNN (K-nearest neighbors)

KNN is a slow, non-parametric learning method that may be applied to regression and classification. The average of the values of the KNN is the result of regression.

$$\widehat{f}(x) = \frac{1}{k} \sum_{i \in N_k(x)} y_i \tag{9}$$

Where $N_k(x)$ are the k-nearest neighbors of x.

GBM (gradient boosting machine)

Gradient Boosting Machine is an ensemble learning approach that constructs a model stage by stage using weak learners such as decision trees and optimises a loss function using gradient descent.

$$F_{m}(x) = F_{m-1}(x) + vh_{m}(x)$$
(10)

Where $h_m(x)$ is the base learner, v is the learning rate. Neural network:

A neural network is a set of algorithms that attempt to recognise correlations in data using an approach comparable to the human brain. It consists of layers of nodes, or neurons, that are linked in a network [23, 26].

$$a^{(l)} = g\left(W^{(l)}a^{(l-1)}\right) + b^{(l)} \tag{11}$$

Where g is the activation function, $W^{(l)}$ and $b^{(l)}$ are the weights and biases of layer l.

XGBoost

XGBoost (Extreme Gradient Boosting) is an optimised gradient boosting system that use decision trees and is intended for speed and performance.

$$F_{m}(x) = F_{m-1}(x) + \gamma_{m}h_{m}(x)$$
(12)

Where γ_m is the weight of the mth tree, and $h_m(x)$ is the m-th tree's prediction.

Model averaged neural network (MANN)

To increase generalisation and prevent overfitting, model-averaged neural networks are trained using several neural networks and then averaged to produce predictions. This ensemble technique combines the strengths of individual models [21].

$$\widehat{y} = \frac{1}{M} \sum_{m=1}^{M} \widehat{y}^{(m)}$$
(13)

Flexible discriminant analysis (FDA)

Flexible discriminant analysis extends linear discriminant analysis by including non-linear correlations through the use of basis expansions and non-parametric regression strategies.

$$g_{k}(x) = \sum_{j=1}^{p} \emptyset_{j}(x) \beta_{jk}$$
(14)

Where $\emptyset_{j}(x)$ are basis functions and β_{jk} are coefficients for class k.

Conditional inference tree (CIT)

Conditional inference trees are a sort of decision tree that selects splits using hypothesis testing, therefore avoiding the overfitting and selection bias found in regular decision trees.

The splitting criterion is based on the test statistic:

Test statistic =
$$max_j\left(\frac{|T_j|}{\sigma_j}\right)$$
 (15)

Where T_j is the test statistic for predictor j and σ_j is its standard error.

Bagged tree (BT)

Bagged Trees are an ensemble approach that creates numerous decision trees using distinct bootstrap samples of training data and then aggregates their results. This helps to minimise variation and increase model stability.

$$\widehat{f}_{b}(x) = \frac{1}{T} \sum_{t=1}^{T} f_{t}(x)$$
(16)

Where T is the number of bootstrap samples and $f_t(x)$ is the prediction from the t-th tree.

Navie bayes (NB)

Navie bayes is a probabilistic classifier based on the Bayes theorem, which assumes predictor independence. It determines the posterior probability of each class based on the supplied characteristics [25].

$$P(C_{k}|x) = \frac{P(C_{k}) \prod_{i=1}^{n} P(x_{i}|C_{k})}{P(x)}$$
(17)

Where C_k is the class, x_i are the input features, and P(x) is the marginal probability of x.

Multivariate adaptive regression splines (MARS)

MARS is a regression approach that models relationships by fitting piecewise linear regressions known as splines, which automatically pick and combine predictor variables to increase prediction accuracy.

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x)$$
 (18)

Where $h_m(x)$ are basis functions and β_m are coefficients.

Boosted generalized linear model (BGGLM)

Boosted The generalised linear model combines the flexibility of boosting methods with the interpretability of generalised linear models, iteratively refining the model through residual fitting [27].

$$F_m(x) = F_{m-1}(x) + v.GLM_m(x)$$
 (19)

Where v is the learning rate, and $GLM_m(x)$ is the m-th GLM fit to the residuals.

Bayesian generalized linear model (BGLM)

The Bayesian Generalised Linear Model (BGLM) builds on generalised linear models by including Bayesian inference, which allows for the integration of previous information and probabilistic interpretations of model parameters.

$$p(\beta | y) \propto p(y|\beta) p(\beta)$$
(20)

Where $p(y|\beta)$ is the likelihood and $p(\beta)$ is the prior distribution.

Analysis and results

Feature selection

The heart disease dataset's correlation matrix reveals significant relationships between various variables in Fig. 2. Age is moderately correlated with resting blood pressure and exercise-induced angina, suggesting older individuals may have higher blood pressure and angina during exercise. Sex is positively correlated with the target variable, indicating a higher likelihood of heart disease in males. Chest pain type is strongly correlated with exerciseinduced angina and the target variable, underscoring its importance in heart disease diagnosis. Max heart rate has a significant negative correlation with age and the target variable, suggesting higher heart rates are associated with younger age and lower heart disease prevalence. Exercise-induced angina and ST slope have strong positive correlations with the target variable, and old peak is positively correlated with ST slope and the target variable.

Many research studies have shown that different kinds of chest pain, like angina, atypical angina, and non-anginal pain, are strong signs that someone might get heart disease. Researchers have derived their findings from these studies [32, 33]. The Framingham Heart Study came to the conclusion that elevated levels of blood sugar in the fasting state are associated with an increased risk of cardiovascular diseases [34, 35]. This was one of the findings of the study. An important diagnostic tool for figuring out if someone has had a myocardial infarction or ischemia is finding any changes in the data from an ECG [36]. In the presence of ischemic heart disease, there is a strong association between the occurrence of exercise-induced angina and the risk of developing the condition. When attempting to diagnose coronary artery disease, the slope of the ST segment at maximal activity is a highly essential component to consider [37]. We chose seven features or variables to use machine learning analysis on based on the results of correlation and another research paper.

Evaluation metrics for machine learning models

In this work, we want to predict heart disease by analyzing a dataset that originally included 12 Features or variables. To enhance the functionality of the model, we carefully chose the seven most relevant variables. Applying these variables to different machine learning models produces substantial outcomes. Logistic regression has an accuracy of 0.83, a precision of 0.84, a recall of 0.78, and F1 score of 0.81. Random Forest performs admirably, with an accuracy of 0.91, a precision of 0.92, a recall of 0.90, and an F1 score of 0.91. SVM regression produces comparable results as logistic regression, whereas KNN matches Random Forest's performance with an accuracy of 0.91. GBM has an accuracy of 0.87, with balanced precision and recall scores. Neural networks and modelaveraged neural networks both score an accuracy of 0.85.

XGBoost has the best accuracy of 0.93, precision of 0.92, recall of 0.92, and F1 score of 0.92. Flexible Discriminant Analysis (FDA) and MARS both have an accuracy of 0.84, with minor differences in precision and recall. The Conditional Inference Tree (CIT) model has a lower accuracy of 0.81, while other metrics are marginally lower than those of the FDA and MARS. The Bagged Tree model is comparable to XGBoost's strong performance, with an accuracy of 0.93 and continuously good precision and recall. Naive Bayes obtains an acceptable accuracy of 0.85.

Bayesian Generalised Linear Models (BGLM) and Boosted Generalised Linear Models (BGGLM) perform similarly, with accuracies of 0.83 and 0.835, respectively as results shown in Table 3. This investigation compares the efficacy of several machine learning models in predicting cardiac disease, demonstrating the



Fig. 2 The findings of correlation across all characteristics are shown as a heatmap

value of choosing essential factors to improve model performance.

ROC-AUC

The comparison of multiple machine learning models on the heart disease dataset using ROC-AUC (Receiver Operating Characteristic-Area Under the Curve) metrics indicates considerable disparities in predicting ability mention in Table 4; Fig. 3 (where x-axis represents Specificity and y-axis represents Sensitivity). Notably, ensemble approaches such as Random Forest and Bagged Tree had the highest ROC-AUC values of 0.95, showing a higher capacity to identify between individuals with and without heart disease. XGBoost also scored admirably, with a ROC-AUC of 0.94, indicating its capability to handle complicated data structures and capture intricate patterns within the dataset. The Gradient Boosting Machine (GBM) had a ROC-AUC of 0.92, whereas models such as SVM regression, KNN, Neural Network, MANN, FDA, CIT, and Naïve Bayes consistently had ROC-AUC values of 0.91, showing their dependability in prediction tasks.

Traditional models, such as Logistic Regression, MARS, BGLM, and BGGLM, performed marginally worse, with ROC-AUC values of 0.90. While these models still have decent accuracy, the results show that ensemble methods and advanced boosting techniques outperform them in terms of predictive power. These findings highlight the significance of model selection and optimisation in heart disease prediction, implying that combining powerful ensemble algorithms might greatly improve diagnostic accuracy. Furthermore, the higher performance

Table 3 Individual classifier results were evaluated

S.No	Models	Accuracy	Precision	Recall	F1Score
1	Logistic Regression	0.83	0.84	0.78	0.81
2	Random Forest	0.91	0.92	0.90	0.91
3	SVM Regression	0.83	0.84	0.79	0.82
4	KNN	0.91	0.91	0.91	0.91
5	GBM	0.87	0.87	0.86	0.86
6	Neural Network	0.84	0.85	0.80	0.82
7	XGBoost	0.93	0.92	0.92	0.92
8	MANN	0.84	0.86	0.79	0.82
9	FDA	0.84	0.85	0.82	0.83
10	CIT	0.81	0.78	0.82	0.80
11	Bagged tree	0.93	0.93	0.92	0.93
12	Navi Bayes	0.85	0.87	0.80	0.83
13	MARS	0.84	0.85	0.82	0.83
14	BGLM	0.83	0.84	0.80	0.82
15	BGGLM	0.83	0.84	0.79	0.82

Table 4 ROC-AUC values for machine learning models

S.No	Models	AUC
1.	Logistic Regression	0.90
2.	Random Forest	0.95
3.	SVM regression	0.91
4.	KNN	0.91
5.	GBM	0.92
6.	Neural Network	0.91
7.	XGBoost	0.94
8.	MANN	0.91
9.	FDA	0.91
10.	CIT	0.91
11.	Bagged Tree	0.95
12.	Navie Bayes	0.91
13.	MARS	0.90
14.	BGLM	0.90
15.	BGGLM	0.90

of models such as Random Forest and Bagged Tree, together with the promising results of XGBoost, suggests that utilising ensemble learning approaches may lead to more reliable and accurate cardiac disease prediction systems, ultimately improving patient outcomes.

Confusion matrix

Overall, Random Forest, XGBoost, and Bagged Tree models outperform in predicting heart disease, with the lowest misclassification rates and best accuracy in confusion matrix. KNN, GBM, and Neural Networks all perform well, making them reliable prediction models. Models such as Logistic Regression, SVM, FDA, MANN, and Naive Bayes perform well but with greater error rates, whereas CIT, BGLM, and BGGLM have lesser accuracy and may require more optimisation for better performance from Table 5.

Cross-Validation

This study uses a combined dataset from several sources to show how well different machine learning models predict cardiac disease. We significantly improved predicted performance by applying fifteen different models and concentrating on seven key features. The maximum accuracy was 93% for XGBoost and Bagged Trees and 91% for Random Forest and KNN. Cross-validation results showed that XGBoost performance slightly decreased while Random Forest performance staved high (94% for K = 10 and 92% for K = 5). This added to the idea that the model was stable. ROC-AUC values showed that Random Forest and Bagged Trees got 95%, XGBoost got 94%, and GBM got 92%. This showed that the models were more reliable. Notably, under cross-validation, KNN's accuracy dramatically decreased, suggesting possible overfitting. Conventional models like Naïve Bayes, SVM Regression, and Logistic Regression performed moderately yet consistently. The results demonstrate in the Table 6, to evaluate models for more than just accuracy because cross-validation offers a reliable assessment of their ability to generalize. Taking everything into account, our results surpass those of previous studies, showcasing the potential of advanced machine learning techniques for precise clinical cardiac disease forecasting.

Discussion

Table 1 summarizes the work of many writers on heart disease datasets, describing different approaches and their accuracies. Almazraoj et al. [15] achieved 83% accuracy with deep learning, whereas Singh et al. achieved 83% accuracy with SVM, 79% with decision trees, and 78% with linear regression. Maji et al. [17] discovered that decision trees and artificial neural networks (ANN) had accuracies of 78% and 77%, respectively, whereas Reddy et al. [21] reached 78% using RBNN. In our investigation, we employed a subset of characteristics and sophisticated machine learning algorithms to get much improved accuracy. XGBoost and Bagged Trees obtained 93% accuracy, while Random Forest and KNN got 91%. In addition, we computed ROC-AUC values, which were not addressed in earlier investigations. Our findings suggest that Random Forest and Bagged Trees have ROC-AUCs of 95, XGBoost has 94, GBM has 92, and the remaining models vary between 90 and 91, as shown in Table 4. This extensive examination indicates the higher performance and diagnostic capacity of our chosen models when compared to earlier efforts.

Limitations

 Lack of Comprehensive Feature Evaluation: Many earlier studies used all accessible characteristics without rigorously examining their impact on model



Fig. 3 ROC-AUC plot for different Machine learning models

performance. Our study identifies seven significant variables based on domain expertise and assesses their influence with machine learning models.

- Limited Comparison of Machine Learning Models: Before we looked at ensemble methods, most study only looked at a few models, like logistic regression, decision trees, or neural networks. To find the best method, our study compares 15 models, such as advanced ensemble methods like XGBoost and Bagged trees.
- Absence of ROC-AUC Analysis: Some studies said they were accurate, but they didn't give ROC-AUC scores, which are a more accurate way to measure how well a classification system works. This study fills in the blanks by looking at the ROC-AUC values of all the models, which gives a strong assessment of how well they can predict.
- Lack of Model Validation Techniques: A number of earlier studies may have overfitted because they failed to use appropriate validation procedures.

K-fold cross-validation is used in our study to guarantee accurate and broadly applicable findings.

• Limited Model Interpretability and Deployment consideration: Some studies didn't look at how the results would work in the real world, like how to make the computer work faster or how to make it easier to use models. Our work talks about the tradeoffs between model complexity and performance, which can be used in real life.

Future directions

- Exploring Survival Analysis Models: In the future, researchers could look into Gaussian Process
 Regression and other methods for survival analysis to figure out how heart diseases will get worse over time.
- Hybrid Model Approaches: It's possible that combining several learning models (like XGBoost, Random Forest, and Bagged Trees) through

Table 5 Confusion matrix for different models

S.No	Models	Confusion Matrix
1	Logistic Regression	$\begin{bmatrix} 88 & 16 \\ 24 & 100 \end{bmatrix}$
2	Random forest	$\begin{bmatrix} 24 & 109 \\ 101 & 8 \end{bmatrix}$
2	C) (AA De grossie e	
3	SVIVI Regression	89 16
4	KNN	
		9 115
5	GBM	97 14
		15 111
6	Neural Network	$\begin{bmatrix} 90 & 15 \\ 10 & 15 \end{bmatrix}$
7	VCD	
/	XGBOOST	
8	ΜΔΝΝ	
0		
9	FDA	92 16
		20 109
10	CIT	92 25
		20 100
11	Bagged Tree	
10	New Provide	
12	Navi Bayes	90 11
13	MARS	
15	MANS	92 10 20 100
14	BGI M	
15	BGGLM	89 16
		23 109

Table 6	K-fold Cross-Validation accuracy for machine learning
models	

S.No	Models	Accuracy	K=10	K=5
			Accuracy	Accuracy
1	Logistic Regression	0.83	0.84	0.84
2	Random Forest	0.91	0.94	0.92
3	SVM Regression	0.83	0.86	0.85
4	KNN	0.91	0.71	0.72
5	GBM	0.87	0.89	0.86
6	Neural Network	0.84	0.85	0.84
7	XGBoost	0.93	0.90	0.89
8	MANN	0.84	0.85	0.84
9	FDA	0.84	0.85	0.84
10	CIT	0.81	0.83	0.82
11	Bagged tree	0.93	0.91	0.92
12	Navi Bayes	0.85	0.84	0.83
13	MARS	0.84	0.85	0.84
14	BGLM	0.83	0.84	0.83
15	BGGLM	0.83	0.84	0.84

ensemble or meta-analysis methods could make predictions more accurate overall.

- Real-time clinical Application: Incorporating the created models into a clinical decision support system could assist clinicians in the early diagnosis of cardiac disease and enhance patient outcomes.
- Validation on larger, Real-world Datasets: In the future, bigger, real-world datasets from hospitals or electronic health records (EHRs) should be used

to test the model's generalizability and practical applicability.

Conclusion

This work offers a thorough assessment of machine learning models for predicting cardiac disease, utilizing a consolidated dataset from many sources. We employed fifteen distinct models, including XGBoost, Random Forest, Bagged Trees, and conventional classifiers such as Logistic Regression and Naïve Bayes, by picking seven essential features. XGBoost and Bagged Trees attained the highest accuracy of 93%, closely followed by Random Forest and KNN at 91%. To guarantee model robustness, we utilized k-fold cross-validation (K=10, K=5), wherein Random Forest showed enhanced stability (94% for K = 10, 92% for K = 5), while XGBoost displayed a marginal decrease. The ROC-AUC scores showed that the models were even more reliable. Random Forest and Bagged Trees got 95%, XGBoost got 94%, and GBM got 92%. Furthermore, tests such as precision, recall, and F1-score proved that these models worked, with Bagged Trees and Random Forest showing a strong balance between sensitivity and specificity. The confusion matrix elucidated misclassification tendencies, facilitating model enhancement. KNN demonstrated a propensity for overfitting, as its accuracy markedly declined during crossvalidation. These results show that ensemble models are better at predicting heart disease and stress how important cross-validation is for generalizability. Our method is better than what has been done before, which shows that modern machine learning techniques can help doctors make accurate and reliable decisions.

Acknowledgements

We would like to convey my sincere gratitude to the Vellore Institute of Technology, Vellore management authority and School of Advance of Technology for giving us with all the assistance and resources.

Author contributions

Darshan Teja M: Data collection, conceptualization, manuscript drafting, literature review, methodology, data analysis. Mokesh Rayalu G: Supervision, project administration, Conceptualization, critical manuscript revision.

Funding

Open access funding provided by Vellore Institute of Technology. Not applicable.

Data availability

The data is secondary, obtained from the UCI Machine Learning Repository. https://archive.ics.uci.edu/.

Declarations

Ethics approval and consent to participate

This research did not involve experiments on live vertebrates, higher invertebrates, or human participants. Therefore, ethical approval and informed consent were not required.

Competing interests

The authors declare no competing interests.

Content for publication

Not applicable.

Received: 20 January 2025 / Accepted: 4 March 2025 Published online: 22 March 2025

References

- World Health Organization. Cardiovascular diseases. (2021). https://www.who .int/health-topics/cardiovascular-diseases#tab=tab=1
- Cleveland Clinic H, Disease. (2023). https://my.clevelandclinic.org/health/dise ases/24129-heart-disease
- Boruzs K, et al. High inequalities associated with socioeconomic deprivation in cardiovascular disease burden and antihypertensive medication in Hungary. Front Pharmacol. 2018;9:839.
- Swissinfo. Cardiovascular disease and cancer were most common killers in Switzerland in 2022,(2024) https://www.swissinfo.ch/eng/society/cardiovasc ular-disease-and-cancer-were-most-common-killers-in-switzerland-in-2022/4 9048308
- Dua D, Graff C. UCI Machine Learning Repository: Dataset on Heart Disease. University of California, Irvine, (2019).https://archive.ics.uci.edu/dataset/45/hea rt+disease
- 6. Pillai R, Oza P, Sharma P. Review of machine learning techniques in health care. *Proceedings of ICRIC 2019: Recent Innovations in Computing*. Springer International Publishing, 2020.
- Hassan M, Mehedi, et al. Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. Comput Electr Eng. 2024;115:109130.
- Chandrasekhar, Nadikatla. Samineni peddakrishna.enhancing heart disease prediction accuracy through machine learning techniques and optimization. Processes. 2023;11(4):1210.
- Perumal R, Kaladevi AC. Early prediction of coronary heart disease from Cleveland dataset using machine learning techniques. Int J Adv Sci Technol. 2020;29(06):4225–34.
- 10. Gao X-Y et al. Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. *Complexity* 2021.1 (2021): 6663455.
- 11. Kahramanli H, Allahverdi N. Design of a hybrid system for the diabetes and heart diseases. Expert Syst Appl. 2008;35(1–2):82–9.
- Hassan D, Hussein HI, Masoud M. Hassan. Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. Biomed Signal Process Control. 2023;79:104019.
- Samuel O, Williams, et al. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Syst Appl. 2017;68:163–72.
- Rahman M, Jamil-Ur et al. Ensemble of multiple models for robust intelligent heart disease prediction system. 2018 4th international conference on electrical engineering and information & communication technology (ICEEICT). IEEE, 2018.
- 15. Almazroi A, Ali, et al. A clinical decision support system for heart disease prediction using deep learning. IEEE Access. 2023;11:61646–59.
- Singh A. and Rakesh Kumar. Heart disease prediction using machine learning algorithms. 2020 international conference on electrical and electronics engineering (ICE3). IEEE, 2020.
- Maji S, Arora S. Decision tree algorithms for prediction of heart disease. Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017. Springer Singapore, 2019.
- Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542–54.
- 19. Cheng C. A., Chiu H. W. (2017, July). An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide

database. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2566–2569). IEEE.

- Amin M. S., Chiam YK, Varathan KD. Identification of significant features and data mining techniques in predicting heart disease. Telematics Inform. 2019;36:82–93.
- Reddy G, Thippa, Khare N. An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model. J Circuits Syst Computers. 2017;26(04):1750061.
- 22. Khateeb N, Usman M. Efficient heart disease prediction system using K-nearest neighbor classification technique. *Proceedings of the international conference on big data and internet of thing.* 2017.
- 23. Hastie T, et al. The elements of statistical learning: data mining, inference, and prediction. Volume 2. New York: springer; 2009.
- Bishop CM. Pattern recognition and machine learning by Christopher M. Bishop. LLC: Springer Science + Business Media; 2006.
- James G, et al. An introduction to statistical learning. Volume 112. New York: springer; 2013.
- 26. Aggarwal CC. Neural networks and deep learning. Volume 10. Cham: springer; 2018. p. 978.
- 27. Murphy KP. Machine learning: a probabilistic perspective. MIT Press; 2012.
- Shi B, et al. Prediction of recurrent spontaneous abortion using evolutionary machine learning with joint self-adaptive Sime mould algorithm. Comput Biol Med. 2022;148:105885.
- Fei X, et al. Projective parameter transfer based sparse multiple empirical kernel learning machine for diagnosis of brain disease. Neurocomputing. 2020;413:271–83.
- Zhang L, et al. Bioinspired scene classification by deep active learning with remote sensing applications. IEEE Trans Cybernetics. 2021;52(7):5682–94.
- Chen M-R, Zeng G-Q, Kang-Di, Lu. A many-objective population extremal optimization algorithm with an adaptive hybrid mutation operation. Inf Sci. 2019;498:62–90.
- 32. Iacovelli JJ, et al. Statin administration improves vascular function in heart failure with preserved ejection fraction. J Appl Physiol. 2024;136(4):877–88.
- 33. Alpenglow JK et al. α -Adrenergic regulation of skeletal muscle blood flow during exercise in patients with heart failure with preserved ejection fraction. J Physiol (2024).
- 34. Aponte Ribero, Valerie et al. Diabetes and the risk of cardiovascular events and all-cause mortality among older adults: an individual participant data analysis of five prospective studies. Eur J Clin Invest (2024): e14340.
- Wu H, et al. Life's essential 8 and risks of cardiovascular morbidity and mortality among individuals with type 2 diabetes: A cohort study. Diabetes Metabolic Syndrome: Clin Res Reviews. 2024;18(6):103066.
- Koyun E, et al. Electrocardiographic prognostic marker in pulmonary arterial hypertension: RS time. Arquivos Brasileiros De Cardiologia. 2024;121:e20240083.
- 37. Levine GN et al. 2015 ACC/AHA/SCAI focused update on primary percutaneous coronary intervention for patients with ST-elevation myocardial infarction: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention and the 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. *Circulation* 133.11 (2016): 1135–1147.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.