Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/150606

Machine Learning Framework for Early Coronary Heart Disease Prediction

Ahmed A.F Osman[1], *P. S. Metkewar*[2], *Sonakshi Ruhela*[3], *Aafaq A. Rather*[4,*], *Sachin Naik*[2], *Basant Tiwari*[2], *Anuja Bokhare*[5], *Maryam Nasser Almusallam*[1], *Ala Abdullah*[1], *Osman Elwasila Elshiekh Idress*[6], *Ali Salem Bin Sama*[6], *and Abdalla Ahmed*[7]

[1] Applied College, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia
[2] Department of Computer Science and Applications, Dr Vishwanath Karad MIT World Peace University, Pune, India
[3] Department of Liberal Arts, Faculty of Psychological Science, Rochester Institute of Technology, Dubai, UAE
[4] Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune, India
[5] Dr D. Y. Patil School of Science & Technology, Dr D. Y. Patil Vidyapeeth, Pune - 411018, Maharashtra, India
[6] Department of Management Information Systems (MIS), College of Business Administration, King Faisal University, Al-Ahsa, Eastern Province, Kingdom of Saudi Arabia
[7] Department of Mathematics and Statistics, College of Science, King Faisal University, KSA

Received: 2 Sep. 2026, Revised: 7 Oct. 2026, Accepted: 16 Oct. 2026
Published online: 1 Nov. 2026

Abstract: Heart disease remains one of the leading causes of mortality worldwide, highlighting the need for early and accurate risk prediction to support timely intervention and preventive healthcare. The primary objective of this study is to develop and evaluate a machine learning-based framework for the early prediction of coronary heart disease (CHD), particularly before the onset of clinical symptoms. To achieve this objective, clinical and lifestyle-related data are analyzed using three classification algorithms: k-nearest neighbors (KNN), decision tree (DT), and random forest (RF). The predictive performance of these models is systematically evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and specificity. The experimental results indicate that the random forest classifier achieves the best overall performance among the evaluated models, attaining an accuracy of 80.88%. The comparative findings demonstrate the effectiveness of ensemble-based learning in identifying individuals at potential risk of coronary heart disease. Furthermore, the results suggest that machine learning techniques can provide a scalable and cost-effective approach to preliminary cardiovascular risk assessment. In conclusion, the proposed framework demonstrates the potential of data-driven predictive models to support early risk identification, facilitate informed clinical decision-making, and strengthen preventive healthcare strategies for coronary heart disease management.

Keywords: Comparative analysis, K-Nearest Neighbor (KNN), decision tree, random forest, heart disease prediction, machine learning algorithms

1 Introduction

This paper focused on following parameters:

- –Usage of Predictive model to anticipate the likelihood of Coronary Illness.
- –Machine learning enables cardiologists to select the optimal treatment option for individual patients.
- –Usage of machine learning algorithms — K-Nearest Neighbors (KNN), Decision Trees, and Random Forests to perform comparative analysis.
- –Performance parameters, such as accuracy, precision, recall, F1–Score, specificity, and support to evaluate model performance.
- –Random Forest gave the best result among the tested algorithms achieving the highest overall accuracy and reliability

One of the most major issues in global health, cardiac problems cause millions of deaths yearly. Usually tough to forecast until symptoms appear—leading to serious problems or even fatality—are these conditions: strokes, heart attacks, and

* Corresponding author e-mail: afadol@kfu.edu.sa; aafaq7741@gmail.com

© 2026 NSP
Natural Sciences Publishing Cor.

other cardiovascular problems. Improving patient results, lowering the demands on healthcare systems, and ultimately saving lives all depend on early detection of heart disease. Though valuable, conventional ways of diagnosing and forecasting heart disease — clinical exams, blood tests, and imaging techniques including Electrocardiograms (ECGs) and Computational Tomography (CT) scans—have their constraints. These diagnostic treatments are burdensome, costly, and often inaccessible in resource-limited settings. Moreover, they usually diagnose heart issues only after symptoms have started to show, so by the time they are found out the disease may already be in a late stage. As reported by the World Health Organization (2024), heart-related illnesses are responsible for roughly 17.9 million fatalities each year, which constitutes 32% of total deaths worldwide [32]. Timely and cost-effective diagnosis continues to be a major obstacle, especially in nations with lower and middle incomes.

Given these constraints, Machine Learning (ML) has arisen as an appealing method for forecasting heart disease [1],[3],[5],[6]. Since they can process great quantities of information and recognize patterns beyond what humans could, machine learning methods can give early on warnings of possible cardiovascular problems before clinical symptoms develop. By considering several types of data sources — demographic data (age, gender), lifestyle variables (diet, physical activity), and clinical parameters (blood pressure, cholesterol levels)—these algorithms forecast the possibility of heart disease [4], [7]. Machine learning is being increasingly seen as a useful weapon for enhancing medical decision making in light of the rising amount of healthcare data.

Recent years have seen several investigations of how machine learning algorithms can be used to forecast heart disease. For example, research has shown the effective application of techniques such as K-Nearest Neighbors (KNN), Random Forest [2],[5],[6],[10] .

1.1 Cardiac Problems and Global Health

- –Cardiac diseases, including strokes, heart attacks, and other cardiovascular issues, cause millions of deaths annually [32].
- –These conditions are often difficult to predict until symptoms appear, leading to serious complications or even fatality.
- –Early detection of heart disease is crucial for improving patient outcomes, reducing healthcare burdens, and saving lives.

Traditional methods (e.g., clinical exams, blood tests, ECGs, CT scans) are effective but have limitations including Expensive and time-consuming, not always accessible in low-resource environments and usually detect heart disease after symptoms appear, often at an advanced stage.

1.2 Machine Learning in Heart Disease Prediction:

ML has emerged as a promising solution for early heart disease prediction. ML can process large volumes of data and recognize complex patterns that humans may miss, providing early warnings before clinical symptoms develop. ML algorithms can predict heart disease using: Demographic data (e.g., age, gender), Lifestyle factors (e.g., diet, physical activity), Clinical parameters (e.g., blood pressure, cholesterol levels) and ML is increasingly seen as a tool to enhance medical decision-making with the growing volume of healthcare data.

1.3 Recent Research on Machine Learning for Heart Disease

Various studies have explored machine learning for heart disease prediction, particularly using algorithms such as:

- –KNN: A non-parametric algorithm that classifies based on proximity to ‘k’ nearest neighbours.
- –Random Forest: An ensemble of decision trees that improves accuracy by reducing overfitting and capturing complex relationships.
- –Decision Trees: Provide transparent decision-making processes that are easy to interpret and understand in a clinical setting.

1.4 Challenges in Machine Learning for Heart Disease

Many studies have focused on individual algorithms without comprehensive comparisons across datasets [7], [9]. Issues include: Lack of standardization in data preprocessing, feature selection, and evaluation metrics, small sample sizes and limited datasets, affecting the generalizability of findings and There is a need for robust comparisons to identify the best-performing algorithms for predicting heart disease.

© 2026 NSP
Natural Sciences Publishing Cor.

1.5 Our Study Objectives

Main Goal: Conduct a comprehensive comparison of KNN, Random Forest, and Decision Trees using a large, publicly available heart disease dataset. Evaluate performance based on key metrics like accuracy, precision, and recall to determine the best predictive algorithm. Provide valuable insights into the strengths and limitations of these algorithms for heart disease prediction.

1.6 Potential of Machine Learning in Healthcare

Medical datasets are complex and require sophisticated analytical methods to uncover hidden patterns.

Machine learning can provide solutions by handling large amounts of diverse data, including risk factors like age, blood pressure, cholesterol, and lifestyle choices.

This study aims to explore how these algorithms can enhance decision-making in cardiology and improve patient outcomes.

1.7 Conclusion and Motivation

This research is inspired by the worldwide shift towards data-informed cardiology, in which machine learning models are progressively utilized to enhance early diagnosis and assist in clinical decision-making [1], [3], [7]. By comparing the performance of KNN, Random Forest, and Decision Trees, the research aims to provide clear guidance on the most effective algorithms for heart disease prediction. The findings will contribute to the body of research on machine learning in healthcare and help clinicians make informed decisions when selecting predictive models. The ultimate goal is to develop accurate, efficient, and accessible tools for early heart disease detection, leading to better health outcomes for patients worldwide.

In Section II, literature review has been incorporated. Methodology given in Section III. Section IV covers with results and discussion, also in Section V conclusion and future scope given subsequently.

2 Literature Review

The author here focused on the Synthetic Minority Oversampling Technique (SMOTE) to handle the imbalance issue to detect patient status. They used ML classifier and (SMOTE) together and this helped their model to achieve an accuracy of 99.2% and 98.52% in heart disease detection. Haung et al. (2022) provided the coronary artery calcification (CAC) score for predicting heart disease using ML based risk prediction model. In this model the integration of imaging and clinical data used to enhance the diagnostic accuracy. This study has been carried out using several algorithms including Random Forest, XGBoost and Support Vector Machine. XGBoost provides the most effective performance in terms of accuracy and robustness. The finding of this research states that the CAC score is a key determinant of CHD risk and which demonstrates that the machine learning models outperforms traditional evaluation tools such as Framingham Risk Score. In this, the integration of imaging biomarkers with the patient health data, this approach provides the precise and individualized CHD risk assessment. The limitation of this study includes the lack of large-scale validation and limited interpretability for the implementation in clinical applications [1].

Velmurugan et al. (2023) [2] proposed an IoT enabled ML model that integrates the real-time dynamic data aquired through sensors. The proposed model consists of ML classification techniques which focuses on practical system implementation aspects such as sensor interfacing, pre-processing of data and edge-to cloud data flow which demonstrates that integration of continuous physiological signals with well known risk features can increase the performance of early detection in constrained settings using the work which is primarily focuses on proof-of-concepts with limited external validation and leaves questions about the long term data quality, generalizability across the populations as well as in the field of clinical workflow integration.

Forrest et al.(2023, The Lancet)[3] propose a validate a machine learning based biomarker for coronary heart disease using longitudinal cohort data, focus on rigorous derivation and external validation across independent population size , this study strength lies in its large-scale , longitudinal design and rigorous evaluation of the predictive calculus beyond the traditional risk factors, presenting a potentially translatable marker for stratified screening, while this challenge remain on clinical implementation , evaluation of assessment of cost-effectiveness in routine care, interpretability of complex models.

Muhammad et al. (2021) [4] proposed a coronary heart disease prediction model using comparative study of multiple classification families and feature extraction techiques. This work provides a experimental comparative that provides the

© 2026 NSP
Natural Sciences Publishing Cor.

importance of data preprocessing, feature selection and model creation in CAD prediction. This model has the, imitation of data size and heterogeneity typical of retrospective datasets which highlights a need for the more standardized benchmark and prospective validation techniques.

Rahim, A., Rasheed (2021)[5] proposed a new framework known as MaLCaDD (Machine Learning Cardiovascular Disease Diagnosis) that aims to predict cardiovascular disease with exceptional precision. This framework addresses the values which are missing by using mean replacement and tackles data imbalance through Synthetic Minority Over Sampling Technique (SMOTE). Feature Importance techniques are then used for effective feature selection. The prediction process employs an ensemble of Logistic Regression and KNN classifiers to achieve heightened accuracy.

Hassan et al(2022,sensors)[6] propose a multi classification techniques to estimate several machine learning classifiers for predicting presence of coronary heart disease and report comparative performance across algorithms and feature sets; this study advances applied direction on classifier choice and preprocessing pipelines for CHD detection, yet it is limited by dataset representativeness and the continuing challenge of translating classifier accuracy into actionable clinical decision support with clear explainability.

A Comprehensive review of ML and DL for Heart Disease Risk (HDR) prediction has been presented by Karna et al. (2025) [7] which includes advancements in algorithms, datasets, evaluation metrics. The study has identified research gaps related to dataset bias, lack of standard reporting, and limited interpretability. This review contributes for future research such as multi-center validation, faireness-aware modelling and interoperability between predictive system and clinical workflows.

Li J, Aguirre AD (2025)[8] This study introduces a fresh approach to prize ECG-related characteristics and prognosticates ECG types in real-time, within a bit of an alternate. The models are trained using a comprehensive dataset of nearly 40,000 ECGs, rigorously labeled by expert cardiologists from different hospitals and countries. These models can identify 7 types of ECG signals, including Normal, AF, Tachycardia, Bradycardia, Arrhythmia, Other, or Noisy.

Yang et al. (2025, Physics of Fluids) [9] provides the proposed a real-time prediction of coronary heart blood-flow pressure from CT angiography using deep learning approach. This system provides a fast substitute for computational fluid dynamics which enable the nearest immediate physiological assessment from the imaging, which contributes lies in accelerating hemodynamic interfaces from the anatomical scanned imaging, this study requires the rigorous validation against the invasive measurement and sensitivity which analyses for imaging artifacts before the clinical deployment.

This paper focuses on enhancing the discovery of heart failure using patient health data and machine learning ways. To achieve this, the authors have employed nine machine learning algorithms and compared their performance. Also, they have introduced a new system called Principal Component Heart Failure (PCHF) point engineering, opting for crucial features to optimize performance. Through this, they drafted a new dataset grounded on eight pivotal features. Rigorous trials were conducted to estimate colorful algorithms, pressing their proposed decision tree system as the top pantomime with an emotional 100% delicacy score by Panda et al. [10].

This paper introduces a new approach, the Recursion Enhanced Random Forest with Improved Linear Model (RFRF-ILM), for heart complaint discovery. Its end is to identify crucial features using machine learning ways to enhance cardiovascular complaint vaticination. The model combines different point combinations and established bracket styles, performing in bettered vaticination perfection. This study identifies factors contributing to cardiovascular complaints, exploring their relationship using the Internet of Medical Effects (IoMT) platform for data analysis. The findings punctuate the correlation between aging and the frequency of coronary roadway complaints, alongside the significance of high blood pressure. With accuracy, stability, and F1–Score ratios of 96.6%,96.8%, and 96.7% independently, this exploration emphasizes the significance of visionary measures against heart complaints by Kaur et al. [11].

A deep learning method for CAD detection and segmentation in ultrasound imaging using gated spatial CNN which is adaptive in nature has been proposed by Singh et.al (2025) [12] This study provides the interpretation of ultrasound by refining the boundary delineation and localizing the lesion.

The deep learning algorithms provides the promising results to automatically localize and interpret the coronary heart calcification on oncology PET and CT scans, this study has been proposed by wu et al. (2025) [13] . The limitation of this study includes the false positives, standardization across the scanner types and integration into the oncology workflows which require the further research.

A survey based on AI methods which is used for cardiovascular prognosis, which summarizes the different trends in the algorithms and provides the focus on data and implementation related issues proposed by Ramathilagam et al. (2024, ICNWC) [14]. This study has limited practical performance evaluations and requires the synthetic benchmarking.

Li et al. (2024) [15] proposed the study on prediction of coronary heart disease using hybrid deep learning model. This study uses multiple neural network frameworks which takes the heterogeneous attributs from the images. The constraints of this study is diverse training data, which limits the generalizability of the proposed model.

Mansoor et al. [16] and authors in [33] propose a method which combines the different machine learning and deep learning technique for the CAD prediction. This study implies the feature fusion and having improved optimistic architecture. This study provides the promising results but it requires the external validation and evaluation of the clinical utility.

© 2026 NSP
Natural Sciences Publishing Cor.

EI-Sofanyu (2024, IEEE Access) [17] proposed and compared the multiple data classification techniques for prediction of heart disease. This study emphasizes on the comparing the performance across the standard datasets, tuning of the clusters, and preprocessing techniques. The limitations of this study includes the limited dataset, and the model needs to be tested for clinically.

Hammoud et al. [18] conducted a comparative analysis of seven ML models, including Logistic Regression, Support Vector Classifier, K-Nearest Neighbor, Random Forest, Decision Tree, Naïve Bayes, and Gradient Boosting Classifier. They utilized the JUH Heart Dataset, comprising 486 patient records, and found that the Random Forest model achieved an accuracy of 94.3%, highlighting its effectiveness in heart disease classification. Banday et al. [19] introduced a hybrid approach combining quantum computing with classical ML algorithms to predict CHD risk. This method demonstrated superior performance over traditional models, offering a promising direction for future research in medical diagnostics. Karaki et al. [20] applied various ML techniques to predict CHD, with the Random Forest model, when combined with oversampling techniques, achieving an accuracy of 84%. Their study in Table 1, underscores the potential of ensemble methods in enhancing predictive performance. Although previous studies have demonstrated the potential of machine learning techniques for predicting coronary heart disease, several challenges remain in identifying the most suitable classification model for reliable and accurate prediction. Existing approaches have explored different algorithms, but their predictive performance varies depending on the dataset, selected features, and evaluation criteria. Furthermore, there is a need for a systematic comparative assessment of different machine learning classifiers using consistent evaluation measures to determine their relative strengths and limitations. In particular, comparing K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF) models can provide valuable insights into their effectiveness in coronary heart disease prediction. Addressing this research gap, the present study aims to evaluate these three classification techniques using clinical and lifestyle-related attributes and assess their performance through accuracy, precision, recall, F1-score, and specificity. The central objective is to identify the model that provides the most effective overall predictive performance and examine its potential for supporting early risk assessment and preventive healthcare decision-making. Recent studies have further highlighted the importance of machine learning and advanced statistical techniques in medical prediction and diagnosis. Khande et al. [34] conducted a comparative analysis of machine learning models for early heart disease diagnosis, demonstrating the relevance of evaluating different classification techniques for improving cardiovascular risk prediction. Their findings provide a direct foundation for further comparative investigations of machine learning algorithms in heart disease diagnosis. In another study, Khande et al. [35] developed an enhanced prediction framework for chronic kidney disease using the XGBoost machine learning algorithm, illustrating the applicability of advanced ensemble learning techniques in medical prediction. Additionally, Isyaku et al. [36] investigated the Area-Biased Burhan Distribution, focusing on its structural properties, parametric inference, and applications to medical data. Their work highlights the complementary role of statistical distribution modelling in analysing medical datasets and supporting data-driven healthcare research. Collectively, these studies demonstrate the growing contribution of machine learning and statistical modelling to medical diagnosis and prediction. Motivated by these developments, the present study focuses specifically on comparing K-Nearest Neighbors, Decision Tree, and Random Forest algorithms to identify an effective approach for early coronary heart disease prediction.

Table 1: Summary of Literature Review

Reference	Contribution	Methodology	Limitations	Future Scope
Velmurugan et al. (2023)	Proposed IoT based ML model for heart disease prediction. This study uses real-time acquisition of data	IoT sensors, cloud integration and ML classification techniques.	Clinical Validation requires and small dataset	This model can be extended to large scale with IoT implementation and observing real time data.
Forrest et al. (2023)	Proposed ML based biomarker for CAD and performance is evaluated on two longitudinal sections.	ML on sectional cohort dataset and external validation of CAD risk.	Small interpretation and complex in model evaluation.	Clinical evaluation and proposing cost effective model.
Muhammad et al. (2021)	Comparative study of ML models.	Multiple classification techniques, feature engineering and selection with comparative study.	Limited dataset.	Model needs to be tested over large dataset.
Hassan et al. (2022)	Detection of Heart disease using multiple machine learning classifiers.	Use of clinical data with model evaluation using RF, KNN and SVM.	Dataset is not balanced.	Needs clinical validation.

Continued on next page

© 2026 NSP
Natural Sciences Publishing Cor.

Table 1 – *Continued from previous page*

Reference	Contribution	Methodology	Limitations	Future Scope
Karna et al. (2025)	Compedious review on ML and DL model.	LR analysis and systematic review.	Lagging in experimental validation.	Use of standardised dataset for the validation.
Yang et al. (2025)	Use of deep learning algorithms, real time prediction of coronary blood flow.	DL frameworks and angiography data. This model estimates the blood flow.	Validation of the model require.	Combining of cardiac imaging and clinical examination.
Singh et al. (2025)	Proposed SCNN for ultrasound CAD images.	Segmentation and gated CNN	Small dataset.	Validation on diverse dataset images.
Wu et al. (2025)	CAC localization with deep learning.	DL , CAC Oncological scans.	Risks for false positive and heterogeneity	Various Scanner Validations.
Ramathilagam et al. (2024)	Use of AI Techniques.	Comparative study and CVD prediction..	Less practical testing.	Benchmarking of dataset and AI Techniques.
Li et al. (2024)	Hybrid deep learning model.	Feature fusion , CNN and LSTM.	Diverse datasets	External Validation
Mansoor et al. (2024)	Use of ML and DL techniques, hybrid classification model.	Feature fusion techniques.	Small dataset	System needs to be tested on real time data.
El-Sofany (2024)	Comparative study of classification algorithms.	DT, SVM and RF models , standardized dataset.	Small dataset.	Needs to apply AI approach.
Huang et al. (2022)	Use of CAC Scores using ML models.	Use of XG Boost algorithm	No external validation	AI integration
Trigka & Dritsas (2023)	Temporal modeling	Modelling of risk factors.	Less data and its complexity	Use of RNN and LSTM
Ma et al. (2024)	CHD prediction in diabetic patients	ML algorithm	Limited dataset	Needs to be tested on generalized dataset.
Jindal et al. (2021)	Use of ML algorithms.	Compared various ML learning algorithms SVM RF.	Small dataset	Need to use real time data.
Rubini et al. (2021)	Use of various ML algorithsm.	Standard dataset and performance evaluation	Limiting in clinical dataset validation.	Needs to use hybrid features.
Nayeem et al. (2022)	Use of ML algorithms.	Standardized dataset and ML classifiers	Limited dataset	Needs to apply on real time hospital data.
Baghdadi et al. (2023)	Comprehensive LR	Systematic review	Performance evaluation	Clinical examinations
Katarya & Meena (2021)	Comparative Study of ML algorithm.	Systematic comparison	Imbalanced dataset	Common benchmarks
Yuan et al. (2023)	CAC score,DL on echocardiograms	CAC, DL Model	CT based CAC.	Integration with echo software.
Riyaz et al. (2021)	Review of ML models.	Meta review	Limited dataset	Use of transformer models
Ramesh & Lakshmanna (2024)	Fuzzy based hybrid DL	Integration CNN and fuzzy	Complexity of model	Clinical validation

3 Methodology

The Proposed methodology's implementation begins with downloading a dataset from Kaggle.com. The data cleaning and preprocessing are done before splitting the dataset. After cleaning and normalizing, the data is visualized. The data visualization helped us to observe the trends and the relationship of attributes in the dataset taken.

The dataset is also resolved into test and train data. The model is trained by applying algorithms to the training dataset, and the model is tested as shown in Table 2.

© 2026 NSP
Natural Sciences Publishing Cor.

Table 2: Clinical Features of Dataset and Their Description

Features	Description
Age	Instance age in years
Sex	Instance gender
Chest Pain	0=No pain; (1)=Mild pain; (2)=Moderate pain; (3)=Severe pain; (4)=Unbearable pain
BP	Instance Blood pressure; 120/80=Normal
Cholesterol	Serum Cholesterol
FBS over 120	Fasting Blood sugar
EKG result	Electrocardiogram results
Max HR	Maximum Heart rate
Exercise angina	Exercise-induced angina
ST depression	ST depression induced by exercise relative to rest
Slope of ST	The slope of the peak exercise ST segment
Number of vessels fluro	Number of significant vessels (0-3) colored by fluoroscopy
Thallium	3=Normal; 6=Fixed defect; 7=reversible defect

The dataset in table 2 contains both numerical and categorical variables. Numerical variables, such as age, blood pressure, cholesterol, maximum heart rate, and ST depression, were treated as quantitative features. Categorical variables, including sex, chest-pain type, fasting blood sugar, electrocardiogram results, exercise-induced angina, slope of ST, and thallium results, were represented using their corresponding numerical category codes. For example, chest-pain categories were coded from 0 to 4, while thallium test results were represented by the codes 3, 6, and 7. Binary categorical variables, such as sex and exercise-induced angina, were represented using numerical codes. These representations enabled the categorical information to be incorporated into the machine learning algorithms. Data cleaning, preprocessing, and normalization were performed before model training and evaluation.

3.1 Dataset

Kaggle is a stage run by Google that assists with work in information shrewdness and machine proficiency. It has incalculable objection-related datasets. For scholarly purposes, these unreservedly open datasets are utilized. In this model, the dataset of heart grumblings is utilized. With 207 lines and 14 sections. The dataset incorporates both information that is useful and credits that aren't valuable. Thus, in pre-handling the helpful information is named and information cleaning is finished to eliminate the invalid qualities.

3.2 Preprocessing

This pre-handling position is the significant stage; significant information is found from the dataset of heart objections. This stage is compulsory because the crude information isn't trustworthy and untreated, so pre-handling is performed for additional ways of delivering prepared crude information. The characteristic's qualities are regularized and changed over into mathematical structure.

3.3 Data Clean

The nature of information has a fundamental influence and is the most definitively portrayed thing to be. For this investigation, information cleaning has bettered the nature of our dataset. Information drawing is important as it eliminates needless or irrelevant properties of information from the dataset. This step of the model will make the dataset more exact. In this piece of the methodology, the Invalid (NaN) values are eliminated from the dataset to make it more helpful as these qualities drop the efficiency of the calculation. At the information cleaning stage, the dataset is likewise regularized to not have any nebulosity after the drawing.

3.4 Data Visualization

The visualizations enable us to uncover significant trends and patterns present in the dataset. For this case, authors can observe the distribution of periods, the frequency of different types of casket pain, and the relationship between maximum

© 2026 NSP
Natural Sciences Publishing Cor.

heart rate and exercise-convinced angina. This visual perceptivity gives a precious environment for making informed opinions during model training and evaluation.

The data representation in Fig. 1 shows us a gender-wise representation of people having heart problems. From the representation, the authors can conclude that several males are prone to maximum heart problems compared to the females. The Pie Chart shows that 67.8% of males and 32.2% of females have a heart problem

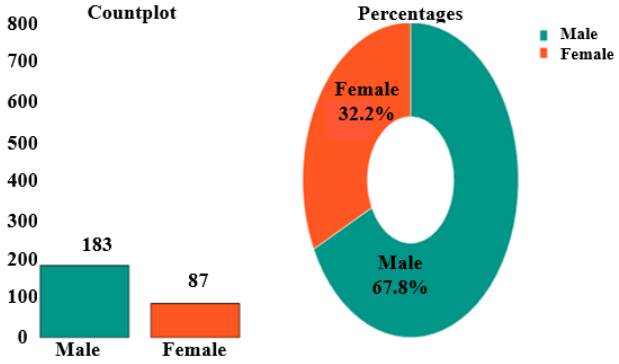

Fig. 1: Gender-wise distribution of the people having heart problems.

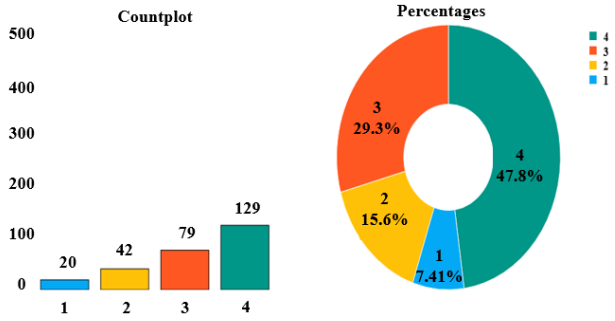

Fig. 2: Chest pain type distribution of the people having heart problems.

Fig. 2 Represents the chest pain type of the dataset. Here 1=Mild pain and 4=Unbearable pain caused by heart problems which describes the person might be suffering from heart problems. This also represents that 7.4% of the people have mild pain in their chest similarly 15.6% of people have moderate pain, 29.3% of people have Severe pain, and 47.8% have unbearable pain during a heart attack.

The above Fig. 3 Represents the Fasting Blood Sugar Distribution among the people; it also represents that 230 people have high blood sugar which can be a cause of heart problems.

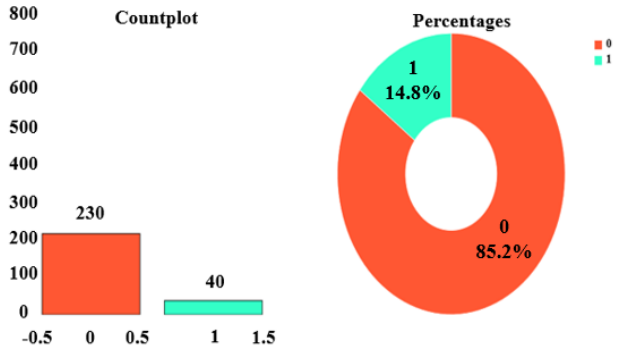

Fig. 3: Fasting blood sugar distribution.

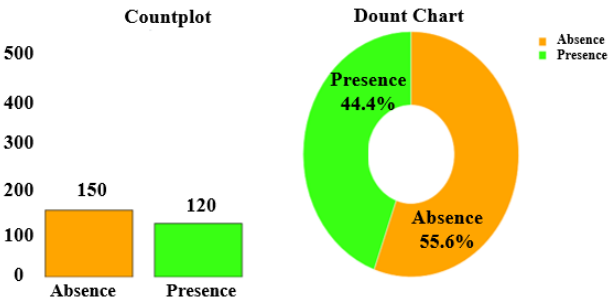

Fig. 4: Heart disease distribution among the people in the dataset.

The above Fig. 4 Represents the Heart Disease Distribution in which authors can visualize that 44% or 120 of the people do not have heart disease and similarly 55.6% or 150 of the people have heart disease.

The above Fig. 5 depicts the Resting ECG distribution which represents the Normal ECG count of people 137 in the dataset Left ventricular hypertrophy (LVH) has a count of 131 and ST has a count of 2.

The above Fig. 6 represents the presence of Exercise Angina in the people. This also represents that 181 people do not have Exercise Angina and 89 people have exercise angina.

Fig. 7, shows the relationship between gender (' male and female') and age. It uses a violin plot to display the distribution of periods for each gender. also, individual data points are represented as strip plots over the violin plots, and a point plot displays the average age for each gender.

© 2026 NSP
Natural Sciences Publishing Cor.

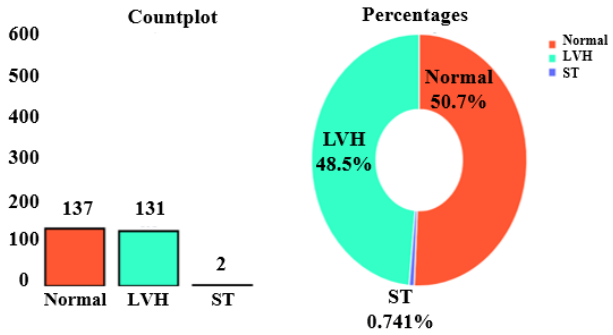

Fig. 5: Resting ECG Distribution of the people having heart problems.

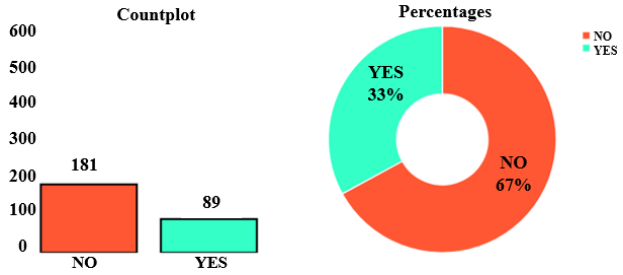

Fig. 6: Exercise angina distribution of the people.



Fig. 7: Violin Plot between Age and sex of the people having heart problems.

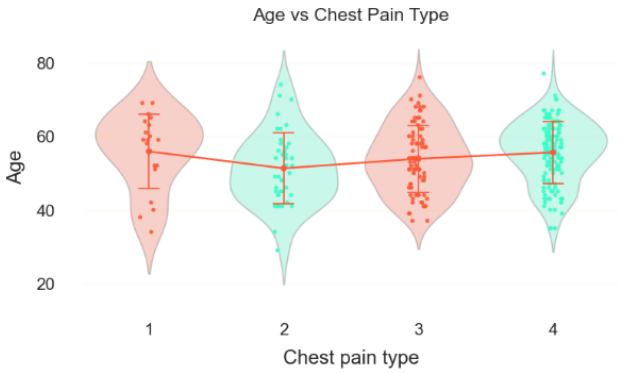

Fig. 8: Violin Plot between Age and chest pain type of the people.

Fig. 8, depicts the relationship between different types of casket pain and age. also, the first action uses a violin plot to show the distribution of periods for each casket pain type. Strip plots and point plots are also employed to give fresh perceptivity.

The third subplot explores the relationship between fasting blood sugar situations ('Over 120' or 'Not Over 120') and age. Again, violin plots, strip plots, and point plots are used to fantasize the data as shown in Fig. 9.

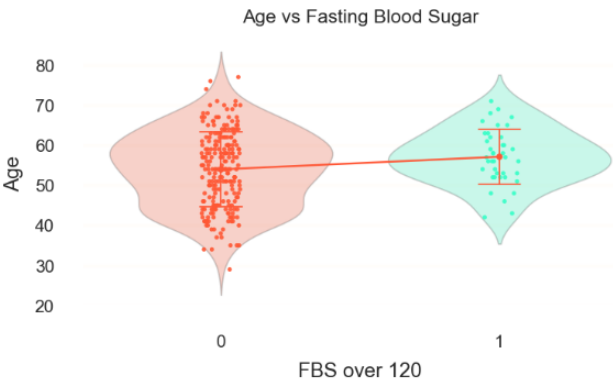

Fig. 9: Violin plot between age and fasting blood sugar.

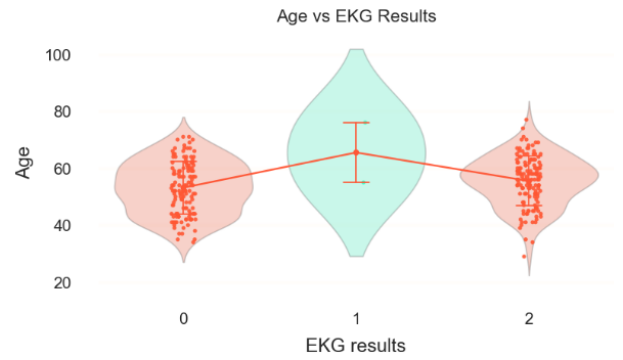

Fig. 10: Violin Plot between Age and EKG results of the people having heart problems.

© 2026 NSP
Natural Sciences Publishing Cor.

Fig. 10, illustrates the relationship between EKG results ('Normal', 'Abnormal,' or 'Other') and age. The violin plot, strip plot, and point plot styles are employed as in the former plots.

Fig. 11, shows how the presence or absence of exercise – angina ('Yes' or 'No') relates to age. The visualization ways used are harmonious with the former plots.

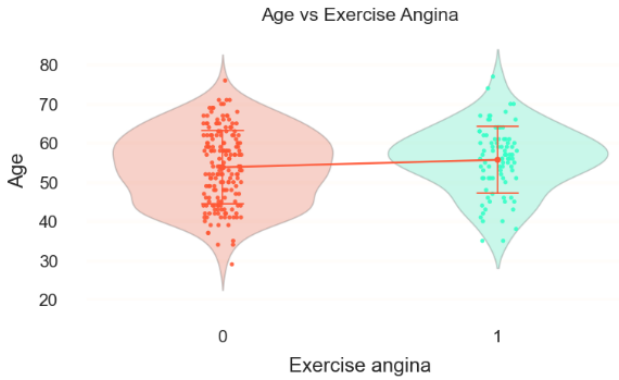

Fig. 11: Violin plot between age and exercise angina of the people having heart problems.

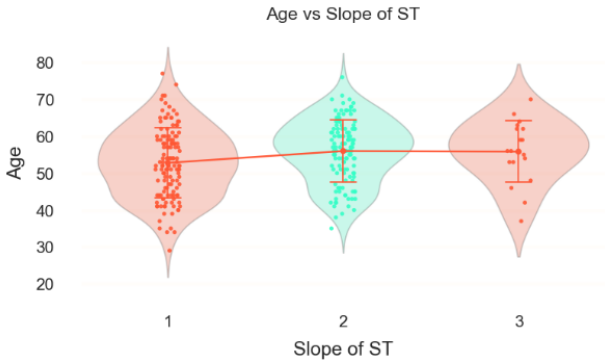

Fig. 12: Violin plot between age and slope ST of the people having heart problems.

Fig. 12, examines the relationship between the pitch of the peak exercise ST member ('Up,' 'Flat,' or 'Down') and age. The chosen visualization styles help portray the distribution and trends effectively. A violin plot is used to fantasize the distribution of a nonstop variable (Age in this case) across different orders of a categorical variable. In your case, authors are creating violin plots for different categorical variables from the heart complaint dataset sex, chest pain type slope ST, Exercise Agnia.). Each violin plot is divided into sections, one for each order, and the range of each section represents the viscosity of data points in that order. The Age values are shown on the y- axis and sex, chest pain type slope ST, Exercise Agnia, etc. are shown on the X-axis. All the plots represent aged between 35 to 70 are more prone to heart problems.

The above Fig. 13 represents the number of people having Cholesterol, which also represent the number of male and female having high cholesterol. Almost 50 people are suffering from high cholesterol.

The correlation matrix heatmap in Fig. 14 visualizes how different attributes in the dataset are related to one another. Each square cell in the heatmap has a number that indicates the relationship between two qualities. The colours in these cells indicate whether the association is favourable or negative. Shades of red, for example, indicate a good association, whereas shades of blue indicate a negative one. The intensity of the hue indicates the strength of the link. There's a value inside each cell that informs us of the actual strength of the connection, rounded to two decimal places. This integer might range between –1 and 1. When it is 1, it indicates that there is a perfect positive relationship when one thing rises, the other rises as well. When the value is –1, there is a completely negative relationship when one thing rises, the other falls. A 0 indicates that there is no straight-line link between the two items. The heatmap's diagonal line from top-left to bottom-right is filled with 1s. This line symbolizes the traits being compared to themselves, making it ineffective. Because the association of a property to itself tells us nothing new, it's frequently made faint or deleted in the heatmap.

New, it's frequently made faint or deleted in the heatmap.

3.5 Training and Testing

ML calculation depends on information for information and issues. The information is known as a preparation dataset, and information is checked for information to verify model execution. The Heart dataset is settled into information bases of trains and assessments. By applying calculations on the preparation dataset to learn, the models are produced and assessed utilizing the testing dataset. Irregular signals and data sources are utilized in the testing system to substantiate whether the model is working accurately. Calculations have a fundamental impact on the development of machine education and information shrewdness models. These calculations can incorporate managed calculations, semi-directed calculations, unaided calculations, and supporting calculations. These sorts are partitioned into various calculations, and the accompanying three calculations utilized in this investigation paper are Arbitrary Woods, Choice Tree, and K-Closest Neighbor, every Directed Calculation.

© 2026 NSP
Natural Sciences Publishing Cor.

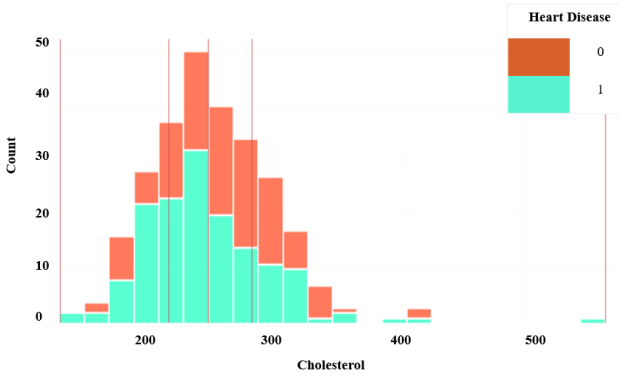

Fig. 13: Histogram of cholesterol distribution among the people.

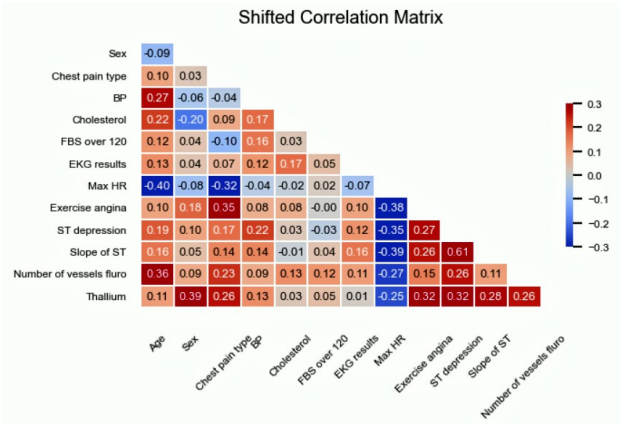

Fig. 14: Represents the Correlation between all the attributes used in the model.

3.6 Random Forest

The random forest AI calculation is tree-based. Various choice trees are framed utilizing the qualities of the dataset and the mean of the trees are anticipated. This forms the choice trees and finds the outcome considering the choice made. Random Forest utilizes bootstrap totalling/sacking for tree learning. Take an information of X = x1, x2, x3, x4 . . . , xn with reaction table being Y = y1, y2, y3, y4 . . . , yn [4], it repeats bootstrap amassing from b = 1 to B. The inconspicuous examples x' are made by taking the normal expectations from practically all the part trees on x': (1) The standard deviation for the vulnerability of forecast of those choice trees is 1.

3.7 Decision Tree

One of the most efficient supervised learning techniques is decision tree for both regression & classification applications. Each these internal nodes denotes a test on an attribute, each branch a test result & each leaf node resulting in a tree structure that resembles flowchart. By continually splitting training the training data based on the values of attributes, a stopping criterion is reached, like a maximum depth of the tree or the smallest number of samples required to split a node. The Gini impurity or entropy metric, which measures the degree of randomness or impurity in the subsets, is used by the Decision Tree algorithm to determine the best feature to use in dividing the data. locating the building that maximizes information gain or impurity reduction following the split is the objective.

3.8 K-Nearest Neighbour

Consider a group of friends who have diverse interests in topics like sports, movies, and cuisine. You now want to determine whether a new item you discovered is something you'd want. KNN steps up in this situation. It's like seeking advice from your pals. You examine their preferred items to determine if the new item is similar to those they enjoy.

Find Neighbours: KNN searches your friends for those who share the same interests as you. These acquaintances resemble your "neighbors".

Ask Your Peers Say you're unsure if you'll enjoy a new film. KNN looks for pals who enjoy the same types of movies. You will likely enjoy the movie if many of your movie-loving pals do.

The majority wins when KNN tallies your movie-loving friends' votes. KNN says you're likely to appreciate the movie if more of your friends liked it than didn't. How many friends (neighbors) you want to ask is indicated by the letter "k" in KNN. If you choose a tiny k, only a few close friends will be contacted. If you choose a high number, you will poll more friends—including some who might not be very similar to you.

KNN is essentially like getting recommendations from friends who share your interests. You're probably going to like something if many of them do. It's a pleasant way to base decisions on the viewpoints of individuals close to you.

© 2026 NSP
Natural Sciences Publishing Cor.

The basic formula used to find the value of K in this algorithm is decided by Euclidean Formula as shown in Eq. (1). The formula is given by,

Euclidean Distance between points A1 and B2:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

Fig. 15, represents the confusion matrix, is a tool used in the machine-learning algorithm that is used to understand the performance of the proposed models. In this, after the prediction is over, we compare them with actual labels which is True Value. The table is usually divided into 4 sections namely True Positive, True Negative, False Positive, and False Negative.

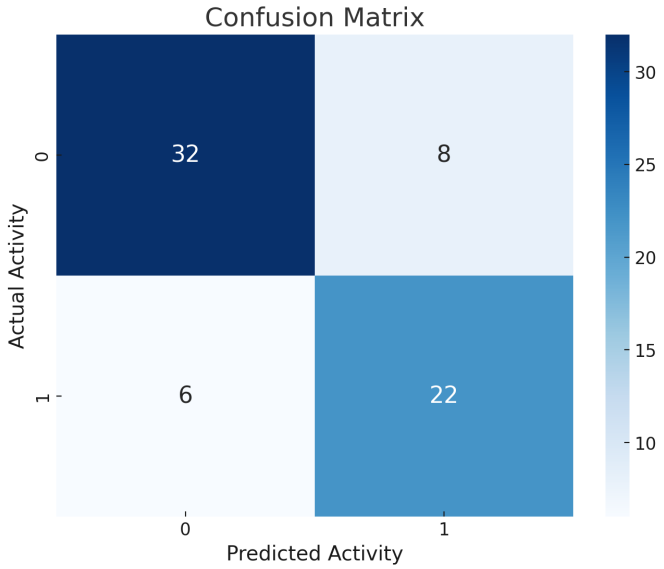

Fig. 15: Confusion matrix representing true positive, true negative, False positive, and False negative.

4 Results and Discussion

The superior performance of the Random Forest (RF) classifier compared with K-Nearest Neighbors (KNN) and Decision Tree (DT) can be attributed to its ensemble learning mechanism, which combines multiple decision trees to generate a more stable and reliable prediction. Unlike a single Decision Tree, which may be sensitive to variations in the training data and prone to overfitting, Random Forest uses bootstrap sampling and random feature selection to reduce variance and improve generalization. Furthermore, KNN relies on distance-based calculations and may be affected by differences in feature scales, irrelevant attributes, and the distribution of observations within the dataset. In contrast, Random Forest can effectively capture complex nonlinear relationships and interactions among clinical and lifestyle-related variables without relying on distance calculations. These characteristics may explain its comparatively better predictive performance in the present study, where Random Forest achieved an accuracy of 80.88%, compared with 79.4% for KNN and 69.1% for Decision Tree. The findings suggest that the ensemble-based structure of Random Forest is advantageous for the dataset considered in this investigation.

Fundamental patient characteristics including age, gender, blood pressure, and heart rate lie at the center of this forecast model. Machine learning techniques are proving to be useful aids in assisting cardiologists to decide on the best therapy for every patient. Most significantly, machine learning models have consistently accurate predictions based on the data provided as shown in Table 3.

© 2026 NSP
Natural Sciences Publishing Cor.

Table 3: Comparison of the Three Different Algorithms Based On Different Parameters

Algorithm	Accuracy	Precision	Specificity	Support	F1–Score	Recall
Decision Tree	0.691	0.74	0.72	40	0.73	0.72
Random Forest	0.808	0.81	0.88	40	0.84	0.88
KNN	0.794	0.84	0.80	40	0.82	0.80

Within the study, three distinct methodologies KNN- supervised Machine learning algorithm, Decision Tree algorithm, and Random Forest — were subordinated to rigorous comparison across multiple parameters including F1–Score, Accuracy, Precision, Recall, particularity, and Support. This scrupulous analysis aimed to identify the approach that demonstrated superior efficacity in terms of performance and results. The findings underscore the paramount importance of judicious algorithm selection and lay a robust foundation for future inquiries geared towards refining predictive models for bolstered heart disease management and prevention. Notably, the research disclosed that the Random Forest algorithm emerged as the frontrunner among the three, exhibiting a commendable accuracy rate of 80.88%. From the confusion matrix in Fig. 16 we can find the following metrics:

Accuracy score is a way to evaluate model performance by comparing predictions made after executing the algorithm with test data. For a prediction to match the real data, a value between 0 and 1 is generated based on the ratio of the complete predicted value. To determine the forecast's accuracy:

–TP = Both the prediction and the actual are positive (normal).

–FP denotes that the prediction is positive (normal) while the real is negative (abnormal).

–FN = The prediction is negative (abnormal), but the actual is positive (normal).

–TN = Both the prediction and the actual are negative (abnormal)

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{2}$$

Further,

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{3}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{4}$$

$$\mathrm{F1} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FN}+\mathrm{FP}} \tag{5}$$

Precision, recall, and F1–Score are the other evaluation criteria for the proposed model. Precision (P) in Eq. (3) is defined as the proportion of correctly categorized positive instances in relation to the total number of positive examples. Recall (R) in Eq. (4) measures how well positive cases are predicted. The weighted average of the Precision and Recall scores is the F1–Score (F1) as in Eq. (5).

Fig. 17 represents the error rate vs. value of K that should be taken. The results from these studies highlight that Random Forest is the most consistently effective algorithm for predicting thyroid diseases, achieving high accuracy, precision, recall, and F1–Scores across various datasets. Feature selection and preprocessing steps significantly contribute to improving performance. However, ensemble methods offer the highest levels of accuracy and are worth considering for future studies in this domain. Although most studies demonstrate strong performance as in Table 4, some models still need refinement, particularly in balancing precision and recall.

© 2026 NSP
Natural Sciences Publishing Cor.

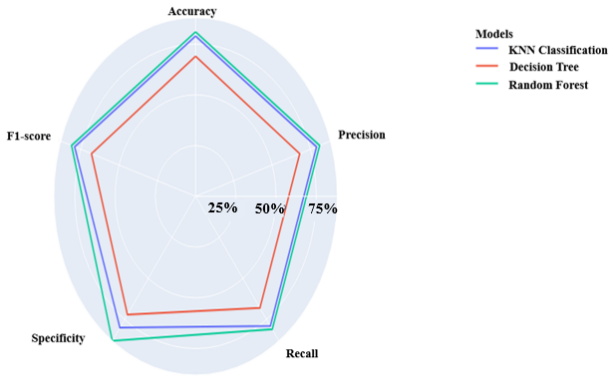

Fig. 16: Visual representation of the proposed model based on comparison metrics.

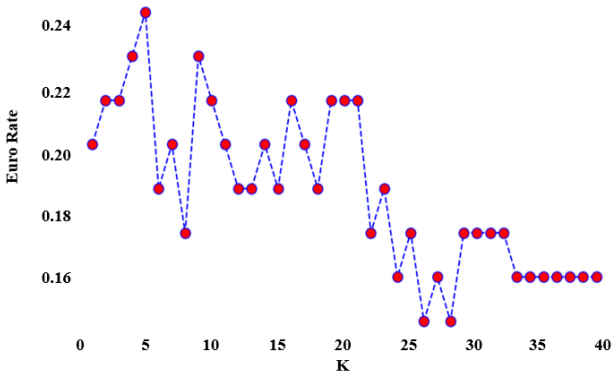

Fig. 17: Represents the Error rate vs. value of K that should be taken.

Table 4: General Comparison Based on Recent Studies

Metric	General Trend in Literature	Observed Results (Current Study)
Accuracy	Random Forest models generally outperform Decision Trees and KNN, often exceeding 80% accuracy.	Consistent with literature — Random Forest achieved 80.8% accuracy, the highest among all models.
Precision	KNN may yield higher precision with optimal k-value tuning; Random Forest also performs strongly.	KNN attained the highest precision (0.84), aligning with reported trends.
Specificity	Ensemble-based classifiers such as Random Forest typically show superior specificity due to reduced variance.	Random Forest demonstrated the highest specificity (0.88), corroborating prior findings.
F1–Score	Random Forest often achieves a balanced trade-off between precision and recall, resulting in higher F1-scores.	Random Forest produced the highest F1-score (0.84), confirming its robustness.
Recall	Random Forest models generally exhibit improved recall, minimizing false negatives.	The Random Forest model achieved the top recall (0.88), in agreement with existing studies.

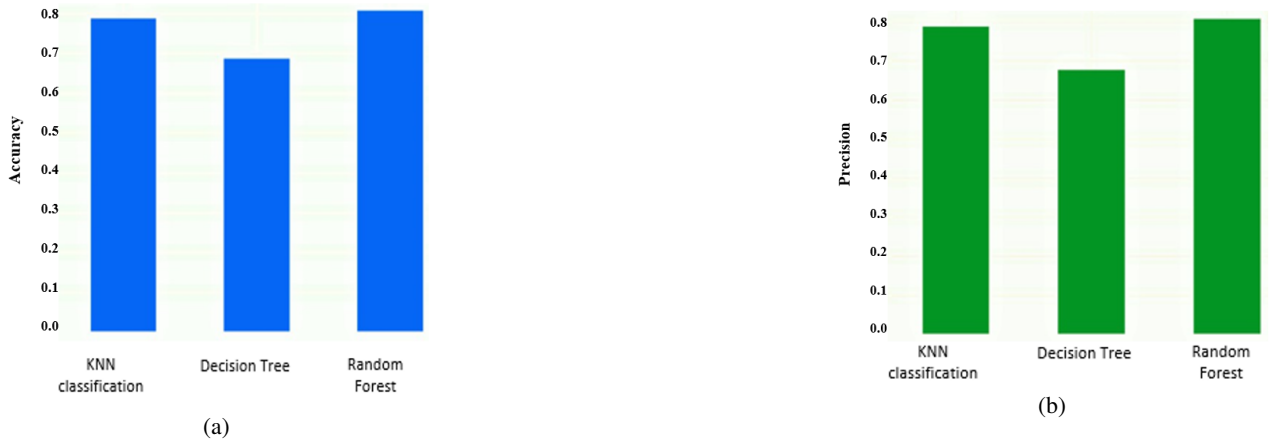

Fig. 18: Bar graph representing algorithms used: (a) Accuracy of ML algorithms (b) Precision of ML algorithms.

© 2026 NSP
Natural Sciences Publishing Cor.

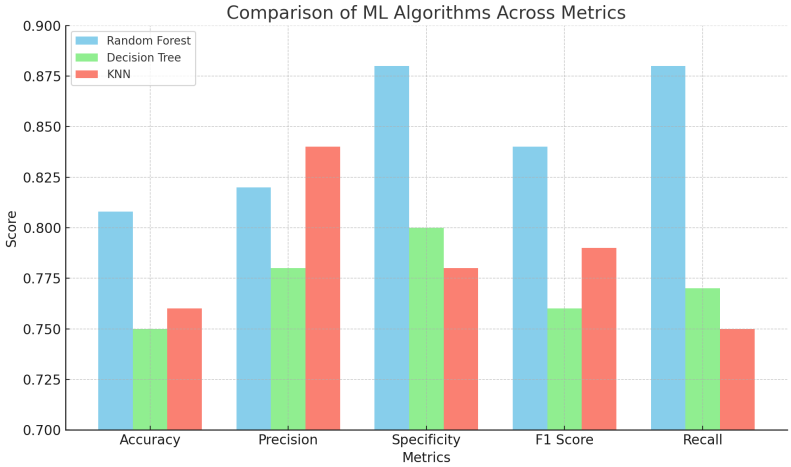

Fig. 19: Bar chart representing comparison of ML algorithms across metrics.

Here above Fig. 18 and Fig. 19 present the bar chart visualizing the comparison of Random Forest, Decision Tree, and KNN across key evaluation metrics.

5 Conclusion And Future Scope

This study investigated the application of machine learning techniques for the early prediction of coronary heart disease (CHD), with the primary objective of identifying an effective classification model for cardiovascular risk assessment. Three machine learning algorithms, namely K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF), were comparatively evaluated using clinical and lifestyle-related attributes. Their predictive performance was assessed through standard evaluation metrics, including accuracy, precision, recall, F1-score, and specificity. The experimental findings revealed that the Random Forest algorithm achieved the highest accuracy of 80.88%, outperforming KNN (79.4%) and Decision Tree (69.1%). These findings demonstrate the potential of machine learning, particularly ensemble-based classification techniques, in supporting early identification of individuals at risk of coronary heart disease. Furthermore, the comparative analysis highlights the importance of selecting an appropriate classification algorithm to achieve reliable predictive performance.

Despite these promising findings, the study has certain limitations. The predictive performance of the proposed framework is dependent on the quality, size, and representativeness of the dataset used for analysis. The evaluation was restricted to three classification algorithms, and the obtained accuracy indicates that further improvements are required before the framework can be considered for practical clinical applications. Moreover, variations in demographic characteristics, clinical conditions, and lifestyle factors may influence the generalizability of the predictive models across different patient populations. Therefore, the proposed framework should be considered a supportive risk-assessment approach rather than a replacement for professional medical diagnosis.

Future research may focus on incorporating larger and more diverse datasets containing additional clinical, demographic, genetic, and lifestyle-related attributes to improve predictive accuracy and generalizability. Furthermore, the investigation can be extended by incorporating other machine learning and ensemble learning techniques, such as Support Vector Machines (SVM), Logistic Regression, Gradient Boosting, and XGBoost. Advanced feature-selection techniques, hyperparameter optimization, and explainable artificial intelligence (XAI) methods may also be explored to improve model performance, interpretability, and clinical reliability. Finally, validating the proposed framework using independent datasets and real-world clinical settings would provide further evidence of its practical applicability. Overall, this study provides a foundation for developing more accurate, interpretable, and data-driven machine learning frameworks to support early coronary heart disease risk prediction and preventive healthcare.

VI. Declarations

Ethical Approval: No ethical approval is required. Data Availability: Data will be made available on reasonable request. Conflict of Interest: There is no conflict of interest among all the authors. Consent to Participate declaration: not applicable.

© 2026 NSP
Natural Sciences Publishing Cor.

Consent to Publish declaration: not applicable
Clinical trial number: not applicable
Author's Contribution:
All the authors have equally contributed.
Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [KFU265346].

References

[1] Huang, Y., Ren, Y., Yang, H., Ding, Y., Liu, Y., Yang, Y., ... & Zhang, Y. (2022). Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment. Computers in biology and medicine, 151, 106297.
[2] Velmurugan, A. K., Padmanaban, K., Kumar, A. S., Azath, H., & Subbiah, M. (2023, January). Machine learning IoT based framework for analysing heart disease prediction. In AIP Conference Proceedings (Vol. 2523, No. 1, p. 020038). AIP Publishing LLC.
[3] Forrest, I. S., Petrazzini, B. O., Duffy, Á., Park, J. K., Marquez-Luna, C., Jordan, D. M., ... & Do, R. (2023). Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. The Lancet, 401(10372), 215-225.
[4] Muhammad, L. J., Al-Shourbaji, I., Haruna, A. A., Mohammed, I. A., Ahmad, A., & Jibrin, M. B. (2021). Machine learning predictive models for coronary artery disease. SN Computer Science, 2(5), 350.
[5] Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. IEEE access, 9, 106575-106588.
[6] Hassan, C. A. U., Iqbal, J., Irfan, R., Hussain, S., Algarni, A. D., Bukhari, S. S. H., ... & Ullah, S. S. (2022). Effectively predicting the presence of coronary heart disease using machine learning classifiers. Sensors, 22(19), 7227.
[7] Karna, V. V. R., Karna, V. R., Janamala, V., Devana, V. K. R., Ch, V. R. S., & Tummala, A. B. (2025). A comprehensive review on heart disease risk prediction using machine learning and deep learning algorithms. Archives of Computational Methods in Engineering, 32(3), 1763-1795.
[8] Li J, Aguirre AD, Moura V, Jin J, Liu C, Zhong L, Sun C, Clifford G, Westover MB, Hong S. An Electrocardiogram Foundation Model Built on over 10 Million Recordings. NEJM AI. 2025 Jul;2(7):10.1056/aioa2401033. doi: 10.1056/aioa2401033. Epub 2025 Jun 26. PMID: 40771651; PMCID: PMC12327759.
[9] Yang, Y., Li, B., Wen, C., Fan, L., Li, T., Feng, Y., ... & Liu, Y. (2025). Deep learning-based real-time prediction of coronary artery blood flow pressure from computed tomography angiography. Physics of Fluids, 37(2).
[10] Panda, R., Majhi, R., & Mohapatra, P. (2022). Heart failure detection using principal component-based feature engineering and machine learning techniques. Biomedical Signal Processing and Control, 78, 103901. https://doi.org/10.1016/j.bspc.2022.103901
[11] Guo, C., Zhang, J., Liu, Y., Xie, Y., Han, Z., & Yu, J. (2020). Recursion enhanced random forest with an improved linear model (RFRF-ILM) for heart disease detection on the Internet of Medical Things platform. IEEE Access, 8, 59247–59256. https://doi.org/10.1109/ACCESS.2020.2981159
[12] Singh, A., Nagabhooshanam, N., Kumar, R., Verma, R., Mohanasundaram, S., Manjith, R., & Rajaram, A. (2025). Deep learning based coronary artery disease detection and segmentation using ultrasound imaging with adaptive gated SCNN models. Biomedical Signal Processing and Control, 105, 107637.
[13] Wu, K. C., Hsieh, T. C., Hsu, Z. K., Chang, C. J., Yeh, Y. C., Lu, L. S., ... & Kao, C. H. (2025). Application of deep learning in automated localization and interpretation of coronary artery calcification in oncological PET/CT scans. The International Journal of Cardiovascular Imaging, 41(3), 453-465.
[14] Ramathilagam, A., Kaviya, B., Thiyagarajan, M., & Amutha, S. (2024, April). Artificial intelligence methods for the prognosis of cardiovascular disease. In 2024 2nd International Conference on Networking and Communications (ICNWC) (pp. 1-7). IEEE.
[15] Li, F., Chen, Y., & Xu, H. (2024). Coronary heart disease prediction based on hybrid deep learning. Review of Scientific Instruments, 95(1).
[16] Mansoor, C. M. M., Chettri, S. K., & Naleer, H. M. M. (2024). Development of an efficient novel method for coronary artery disease prediction using machine learning and deep learning techniques. Technology and Health Care, 32(6), 4545-4569.
[17] El-Sofany, H. F. (2024). Predicting heart diseases using machine learning and different data classification techniques. IEEE Access.
[18] A. Hammoud, A. Karaki, R. Tafreshi, S. Abdulla, and M. Wahid, "Coronary heart disease prediction: a comparative study of machine learning algorithms," Journal of Advances in Information Technology, vol. 15, no. 1, pp. 27–32, 2024.
[19] M. Banday, S. Zafar, P. Agarwal, and M. A. Alam, "Early detection of coronary heart disease using hybrid quantum machine learning approach," arXiv preprint arXiv:2409.10932, 2024.
[20] J. A. Karaki, P. Ilono, S. Baweja, J. Naghiyev, R. S. Yadav, and M. A. Z. Khan, "Predicting coronary heart disease using a suite of machine learning models," arXiv preprint arXiv:2409.14231, 2024.
[21] Huang, Y., Ren, Y., Yang, H., Ding, Y., Liu, Y., Yang, Y., ... & Zhang, Y. (2022). Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment. Computers in biology and medicine, 151, 106297.

© 2026 NSP
Natural Sciences Publishing Cor.

[22] Trigka, M., & Dritsas, E. (2023). Long-term coronary artery disease risk prediction with machine learning models. Sensors, 23(3), 1193.
[23] Ma, C. Y., Luo, Y. M., Zhang, T. Y., Hao, Y. D., Xie, X. Q., Liu, X. W., ... & Lin, H. (2024). Predicting coronary heart disease in Chinese diabetics using machine learning. Computers in Biology and Medicine, 169, 107952.
[24] Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing.
[25] Rubini, P. E., Subasini, C. A., Katharine, A. V., Kumaresan, V., Kumar, S. G., & Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. Annals of the Romanian society for cell biology, 25(2), 904-912.
[26] Nayeem, M. J. N., Rana, S., & Islam, M. R. (2022). Prediction of heart disease using machine learning algorithms. European Journal of Artificial Intelligence and Machine Learning, 1(3), 22-26.
[27] Baghdadi, N. A., Farghaly Abdelaliem, S. M., Malki, A., Gad, I., Ewis, A., & Atlam, E. (2023). Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. Journal of Big Data, 10(1), 144.
[28] Katarya, R., & Meena, S. K. (2021). Machine learning techniques for heart disease prediction: a comparative study and analysis. Health and Technology, 11(1), 87-97.
[29] Yuan, N., Kwan, A. C., Duffy, G., Theurer, J., Chen, J. H., Nieman, K., ... & Ouyang, D. (2023). Prediction of coronary artery calcium using deep learning of echocardiograms. Journal of the American Society of Echocardiography, 36(5), 474-481.
[30] Riyaz, L., Butt, M. A., Zaman, M., & Ayob, O. (2021, August). Heart disease prediction using machine learning techniques: a quantitative review. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3 (pp. 81-94). Singapore: Springer Singapore.
[31] Ramesh, B., & Lakshmanna, K. (2024). A novel early detection and prevention of coronary heart disease framework using hybrid deep learning model and neural fuzzy inference system. IEEe Access, 12, 26683-26695.
[32] World Health Organization, Cardiovascular diseases (CVDs), Geneva, Switzerland, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[33] Saravanan, S., Dar, S. A., Rather, A. A., Qayoom, D., & Ali, I. (2025). Deep Learning Models for Intrusion Detection Systems in MANETs: A Comparative analysis. Decision Making Advances, 3(1), 96–110. https://doi.org/10.31181/dma31202556
[34] Rajeshree Khande, Walid Ayadi, Nikita Bhandhari, Yasser Farhat, P. S. Metkewar, Sharvari R. Shukla, Aafaq A. Rather, Mushtaq A. Lone (2025), Comparative Analysis Of Machine Learning Models for Early Heart Disease Diagnosis, International Journal of Statistics in Medical Research, Vol. 14 (1), pp 590-600, DOI: https://doi.org/10.6000/1929-6029.2025.14.56
[35] Rajeshree Khande, Nrashant Singh, Sachin Naik, Satvik Bodke, Aniket Gaikwad, P.S. Metkewar, Raeesa Bashir, Aafaq A. Rather (2026), Enhanced Prediction of Chronic Kidney Disease Using Xgboost Machine Learning Model, International Journal of Statistics in Medical Research, Vol. 15 (1), pp 109- 120 DOI: https://doi.org/10.6000/1929-6029.2026.15.10
[36] Ahmad Isyaku, Aafaq A. Rather, Ahmed A.F. Osman, Maryam Nasser Almusallam (2026), The Area-Biased Burhan Distribution: Structural Properties, Parametric Inference and Application to Medical Data; International Journal of Statistics in Medical Research, Vol. 15 (3), pp 378-389; DOI: https://doi.org/10.6000/1929-6029.2026.15.33

© 2026 NSP
Natural Sciences Publishing Cor.