

ORIGINAL RESEARCH ARTICLE

An ensemble-driven machine learning framework for robust heart disease prediction using dimensionality reduction and comparative evaluation

Achyut Mitra[†], Radha Krishna Jana^{†*}, Bidisha Bhabani, and Tanaya Das

Department of Computer Science and Engineering, JIS University, Kolkata, West Bengal, India

Abstract

Coronary heart disease is a major cardiovascular condition characterized by reduced coronary blood flow, often due to atherosclerotic narrowing of the coronary arteries. Early and prompt identification of these abnormalities is critical to minimizing healthcare costs and improving patient survival rates. Despite impressive progress in predictive modeling, many existing machine learning models suffer from overfitting, in which they perform well on training or test data but fail to generalize to larger and more heterogeneous patient populations. This limitation diminishes their clinical validity and applicability in medical diagnosis. To address these issues, the proposed study presents an ensemble-based machine learning approach for the early prediction of heart disease. The proposed system incorporates several supervised learning algorithms to increase predictive accuracy and generalization performance. Decision tree, support vector machine, random forest, and classifiers trained on principal component analysis-transformed features were compared. The framework involves systematic data preprocessing, dimensionality reduction, and class imbalance management to ensure methodological robustness and consistency of diagnostic outcomes. Experimental results demonstrate that the ensemble-based approach achieves superior predictive performance while minimizing data bias. These findings highlight the potential of machine learning-based diagnostic systems to support clinical decision-making, improve healthcare resource allocation, and enhance patient outcomes by enabling timely and accurate cardiovascular risk identification.

[†]These authors contributed equally to this work.

*Corresponding author:

Radha Krishna Jana
(radhakrishna.jana@jisuniversity.ac.in)

Citation: Mitra A, Jana RK, Bhabani B, Das T. An ensemble-driven machine learning framework for robust heart disease prediction using dimensionality reduction and comparative evaluation. *Brain & Heart*. 2026;4(3):025440069. doi: 10.36922/BH025440069

Received: October 31, 2025

Revised: January 12, 2026

Accepted: February 11, 2026

Published online: May 8, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Coronary arteria dysfunction; Systolic diastolic pressure; Ensemble forest technique

1. Introduction

Heart failure, coronary artery disease, and cardiac rhythm disorders remain among the leading causes of death worldwide. The World Heart Federation estimates that there are currently about 20.5 million cardiovascular disease-related deaths every year, and this number is expected to increase significantly in the next few decades. Early diagnosis and classification of cardiac conditions are thus of utmost importance, as early clinical intervention has been shown to significantly impact therapeutic outcomes and mortality rates.^{1,2}

Timely diagnosis of cardiac anomalies enables prevention, reduces the risk of progression, and lowers the likelihood of serious complications.^{3,4} Such proactive steps will enable clinicians to customize treatment programs for individual patients, deal with modifiable risk factors, and finally enhance prognosis. Moreover, early diagnosis reduces the economic burden on health care systems, and the intervention's benefits include lowering the long-term cost of treatment and the social costs associated with cardiovascular morbidity and mortality.^{5,6} Historically, the diagnosis of the heart has been largely dependent on the skills and clinical judgment of the physician.^{7,8} Although these conventional methods are quite effective in most instances, they are subjective, time-consuming, and require specialized training. The complexity and heterogeneity of cardiovascular pathologies are further compounded by the nature of the problem, which can lead to misclassification even by professionals.^{9,10}

This study aims to address these challenges by developing and comparatively evaluating machine learning (ML)-based systems for heart disease prediction using the Framingham dataset, informed by a review of relevant prior studies.¹¹ Particular attention is given to methodological frameworks, data attributes, performance functions, and existing constraints in the formation of reliable predictive frameworks. The Framingham dataset was used as a benchmark dataset for model training and evaluation.^{12,13}

The literature is reviewed with respect to supervised learning in cardiac diagnostics, including protocols for data collection, preprocessing of incomplete records and outliers, feature selection, and resolution of class imbalance via augmentation and resampling.¹⁴ Lastly, cross-validation and algorithm selection schemes are used to evaluate the performance of models in order to ascertain the strength of diagnostics and generalizability.¹⁵⁻¹⁷

1.1. Motivation

Early and accurate diagnosis of heart disease remains a major global healthcare challenge because cardiovascular disorders are the major cause of morbidity and mortality. Cardiac symptoms are highly heterogeneous and difficult to interpret, underscoring the need for sophisticated computational assistance in clinical decisions. ML systems that leverage many clinical predictors provide fast, practical diagnostic information. This paper comprises six sections: introduction, literature review, dataset description, methodology pipeline, experimental analysis, and conclusion. It presents a leakage-free, structured framework for benchmarking the Framingham dataset,

with greater emphasis on reproducibility, methodological rigor, and disciplined evaluation, rather than on presenting a new algorithm.

2. Literature survey

A hybrid diagnostic model was suggested, as the method focuses on feature selection to optimize classification performance. Rajendran *et al.*¹⁸ proposed an entropy-based feature engineering pipeline with a dual-classifier ensemble comprising logistic regression and naive Bayes. Ahsan *et al.*¹⁹ addressed the issue of imbalanced datasets. Their analysis demonstrated the significance of identifying methodological trends, technological gaps, and the potential for current research in ML-based cardiac disease classification. Behera *et al.*²⁰ employed several ML classifiers to identify individuals in high-risk groups using clinical variables such as blood pressure, diabetes status, lipid levels, body mass index, tobacco use, and family history. Ghasemieh *et al.*²¹ developed an extreme gradient boosting-based model to forecast emergency hospital readmission in cardiac patients. The model utilized behavioral characteristics and a large confidential dataset obtained from the Massachusetts Institute of Technology Laboratory of Computational Physiology, and contributed a novel classification architecture adapted for acute readmission cases.¹⁹ Emmons-Bell *et al.*²² recommended a convolutional neural network (ConvNet) model trained on phonocardiogram signals using the hybrid constant-Q transform (HCQT). The model employed mel-frequency cepstral coefficients and a five-layer regularized ConvNet to extract acoustic features through the constant-Q transform, variable-Q transform, and HCQT. Their work was reported as the first to apply HCQT to phonocardiogram signal classification. The study evaluates metrics and statistical tests for ML, highlighting the importance of proper evaluation methods. It explains that relying only on accuracy can be misleading, especially for imbalanced datasets. Various performance metrics, including precision, recall, F1-score, and receiver operating characteristic-area under the curve (ROC-AUC), have been discussed, highlighting their specific roles in model assessment. The paper emphasizes that the choice of metric should depend on the application and error impact. It also explores statistical tests for comparing models and warns against their misuse, which may lead to false conclusions. The study recommends using cross-validation techniques to improve reliability. Additionally, it suggests reporting confidence intervals to improve the interpretation of results. Combining appropriate metrics with proper statistical validation yields more robust, trustworthy ML outcomes.²³

3. Proposed model

The Framingham Heart Study dataset was used to develop and evaluate the models. The first step is an in-depth data collection, including patient demographics, clinical signs, and diagnostic test results. After collection, the dataset undergoes preprocessing protocols that include data cleansing and normalization, as well as gap filling to make the data consistent and reliable. Figure 1 shows the entire ML life cycle used in this investigation, including data ingestion, preprocessing, model construction, and evaluation.

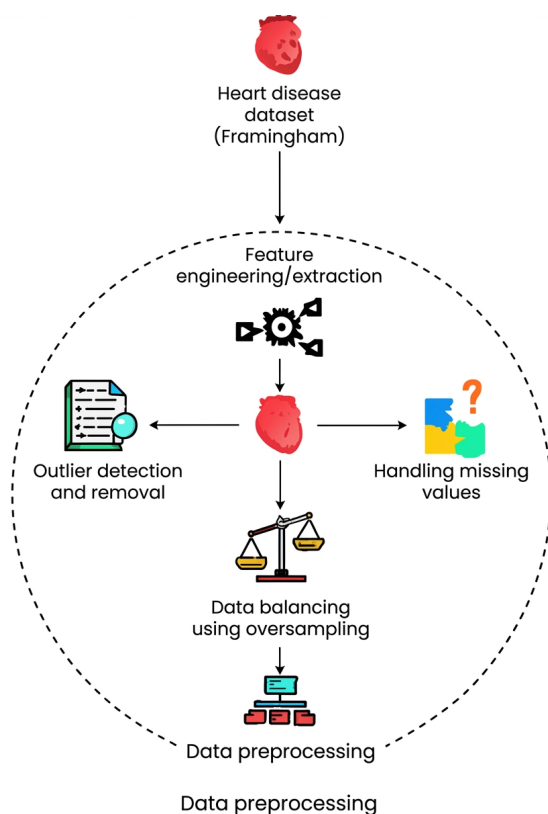


Figure 1. Workflow architecture for predictive model development

3.1. Dataset description

The Framingham dataset is a commonly used resource for assessing long-term cardiovascular risk. It consists of 4,240 patient records across 16 demographic, lifestyle, and clinical variables, including systolic and diastolic blood pressure. The target attribute, 10-year coronary heart disease, is binary and indicates whether a coronary event occurred within the subsequent 10 years. Since the dataset contains missing values and class imbalance, preprocessing steps such as imputation and oversampling are needed to preserve data integrity and ensure reliable model training.

The following features render the dataset appropriate for testing forecasting ML methods.

4. Methodology

This section presents the conceptualization of the protocol undertaken in the evaluation process. One predictive model is a fully automated classifier used to identify individuals with a high likelihood of developing coronary heart disease. Class imbalance was addressed using random oversampling. Although the synthetic minority over-sampling technique was also a possible approach, it was not used to avoid introducing artificial bias into clinical variables. This option maintains the balanced dataset's interpretability and clinical plausibility. The dimensionality reduction method employed was principal component analysis (PCA), which was used solely to improve computational efficiency. PCA was not applied as a classifier but as a preprocessing step before model training.

4.1. Data collection

The input to the model is the Framingham Heart Study data, which will be used to develop the models. It consists of 4,240 individual records, 15 predictive attributes, and a binary target variable. The information encompasses a wide range of clinical and demographic attributes of cardiovascular risk assessment. However, there exist gaps in the data and non-contributory variables in the raw data that should be preprocessed with care prior to model training.

4.2. Data preprocessing

Preprocessing is a critical step that transforms raw clinical data into a standardized format suitable for ML applications. The phase involves data cleaning, standardization, and missing-value imputation to address the inconsistencies, noise, and missing data. Feature selection is performed to obtain informative predictors, and random oversampling is conducted.

The correlation heatmap in Figure 2 provides a visual representation of the strength and directional linear relationships between pairs of features, with color intensity indicating the correlation coefficient. Exploratory analysis was also used to identify potential outliers that could adversely affect model performance.

While PCA significantly reduces dimensionality and improves computational efficiency, it may obscure the direct clinical interpretation of statistics based on the original variables (such as systolic blood pressure, diabetes status, or smoking habits) of the data. This trade-off between efficiency and interpretability is particularly critical in

clinical settings. To address this limitation, interpretability algorithms such as SHapley Additive exPlanations (SHAP) can be used to quantify each feature's contribution to the model's outputs. SHAP is comparatively less complex and preserves clinically relevant insights that are important to clinicians.²⁴⁻²⁶

4.3. Model training

In machine learning, structured, preprocessed data are fed into algorithms, which, in turn, identify patterns to learn predictive relationships. The training data consist of a number of input features and corresponding output labels, allowing the model to learn associations and decision boundaries. This supervised learning paradigm is used to develop accurate classifiers for predicting unseen data.

4.4. Model construction

A number of supervised learning procedures are used to create predictive models for cardiac disease in this study.

4.4.1. Decision tree classifier

The decision tree classifier operates by recursively splitting the data based on conditional rules, and is represented as a tree structure in which internal nodes represent decision

criteria, and leaf nodes represent class labels.

The most commonly used variants are classification and regression trees and C4.5, which have been widely applied in medical diagnostics. These models are favored because they present classification outcomes in terminal nodes, while intermediate branches represent attributes that influence classification. Using the decision tree algorithm on the Framingham dataset, the model identifies optimal split points, with leaf nodes assigned predictive labels based on learned patterns.

As illustrated in Figure 3, the decision tree derived from the Framingham data outlines the decision-making pathway for classifying individuals based on characteristics such as smoking habits, diabetes, and systolic blood pressure.

For all classifiers, hyperparameters were explicitly tuned to ensure methodological transparency. The decision tree was implemented using the classification and regression trees algorithm with a maximum depth of 10 and Gini impurity as the splitting criterion. The random forest classifier was constructed with 500 trees, each trained on bootstrapped samples, and a maximum feature selection of $\sqrt{n}$. The support vector machine (SVM) employed

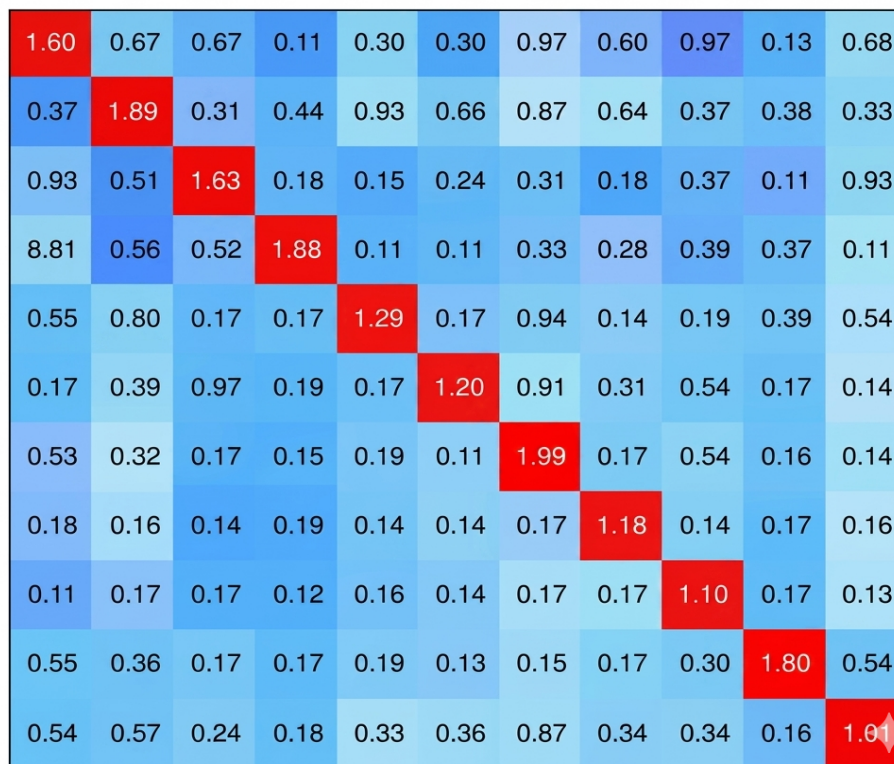


Figure 2. Variable interdependence map for Framingham clinical features

a radial basis function kernel with $C = 1.0$ and $\gamma = 0.01$. For dimensionality reduction, PCA retained components explaining 95% of the variance. Model evaluation was performed using a stratified 10fold crossvalidation scheme with an 80:20 train–test split to ensure generalizability and minimize overfitting.

4.4.2. Support vector machines

The SVM is a supervised learning algorithm used for both binary and multi-class classification. When applied to heart disease prediction, SVM maps patient information into a higher-dimensional feature space, where an optimal separating hyperplane is constructed to distinguish between patients at risk of developing coronary disease and healthy patients. SVMs can handle both linearly and non-linearly separable data by using kernel functions (polynomial, radial basis function, and sigmoid kernels). The algorithm optimizes the distance between support vectors, enabling better generalization, minimizing classification error in the decision, and improving diagnostic reliability for heterogeneous clinical data.

4.4.3. Random forest

Random Forest is a robust ensemble classification algorithm that constructs a sequence of choice trees to provide powerful and accurate estimations. Each tree is trained on a random subset of the data and features, thereby enhancing diversity and reducing overfitting. For classification tasks, the final output is determined by majority voting across all trees, whereas in regression tasks, the average of the outputs is used. It is an effective and stable stochastic approach, particularly suitable for complex, high-dimensional data. Random forests remove the biases of individual models by combining the results of thousands of weak learners to produce a more accurate model that is highly applicable in medical diagnosis, including heart disease prediction. The ensemble technique (random forest) is illustrated in Figure 4.

5. Performance evaluation and diagnostic validation

The model testing process is a critical component of the machine learning pipeline and is used to measure the predictive strength and generalization ability of a trained classifier on unseen data. This evaluation ensures the

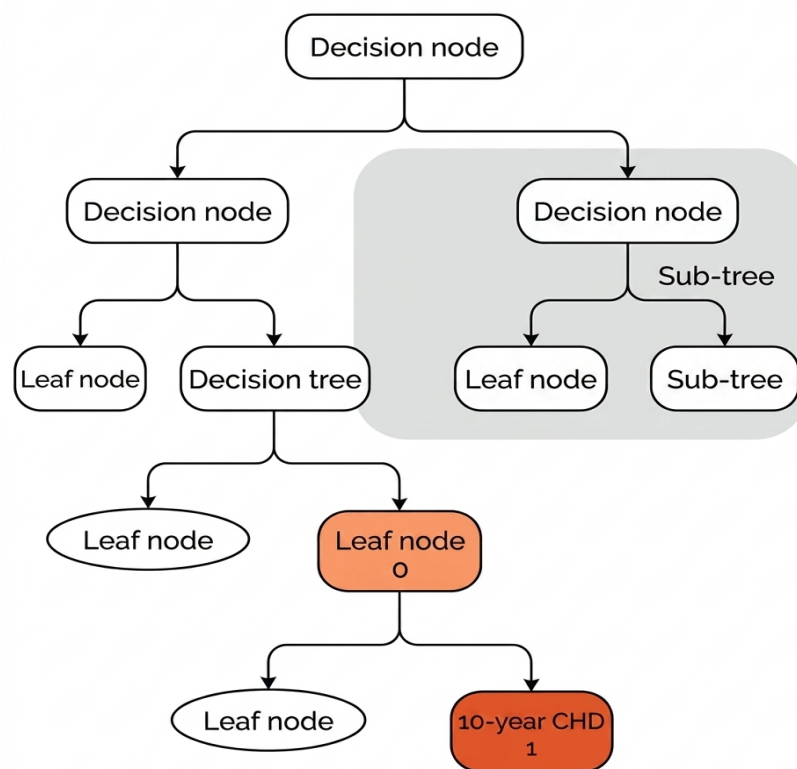


Figure 3. Structural Layout of the decision tree classifier
Abbreviation: CHD: Coronary heart disease.

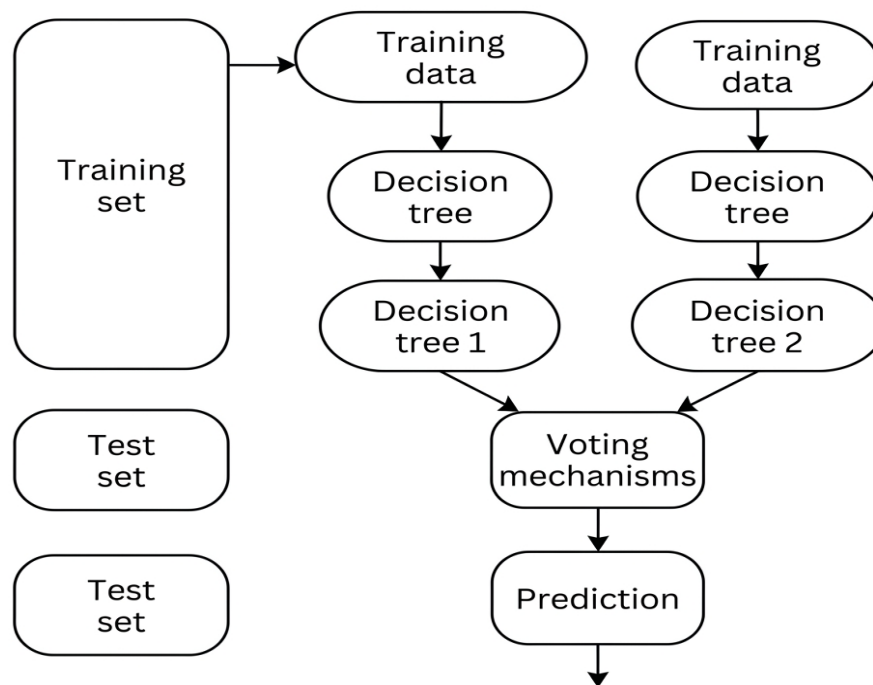


Figure 4. Ensemble tree structure of the random forest classifier

reliability of the model when applied in real-world clinical settings, as it reflects diagnostic performance across a broad spectrum of patient profiles.

5.1. Model evaluation

It is necessary to evaluate the performance of the heart disease prediction model to determine its applicability in clinical practice and the reliability of its diagnostic output. Figure 5 provides a graphical representation of true positives, true negatives, false positives, and false negatives, which constitutes the confusion matrix of the selected classifier.

5.2. Performance appraisal and metrics

Effective and clinically viable ML models for cardiac diagnostics require rigorous performance evaluation. This enables comparison among different algorithms and supports informed selection of models for implementation in healthcare systems. Standard evaluation metrics are used to assess the efficiency of the classification.

Table 1 indicates that the model performed well across all relevant metrics, with precision indicating that 84% of predicted positives were correct, and recall indicating that 84% of actual positives were correctly captured. The F1-score balances these two measures at 0.84. The weighted average accounts for differences in class sizes, whereas the

macro average treats all classes equally. Both values being similar suggests that the model performs consistently across classes, indicating balanced predictive performance.

Table 1. Performance metrics for a decision tree classifier

Metric type	Precision	Recall	F1-score	Support
Weighted average	0.84	0.84	0.84	1,531
Macro average	0.84	0.84	0.84	1,531
Accuracy	–	–	0.84	1,531

Table 2 shows that the random forest classifier achieved 97% performance across all evaluation metrics, with precision indicating that 97% of predicted positives were accurate, and recall indicating that 97% of actual positives were accurately identified. The F1-score balances these two measures at 0.97. The weighted average accounts for class size differences, whereas the macro average treats all classes equally. The similarity of these values, along with the high accuracy, indicates consistent model performance across classes.

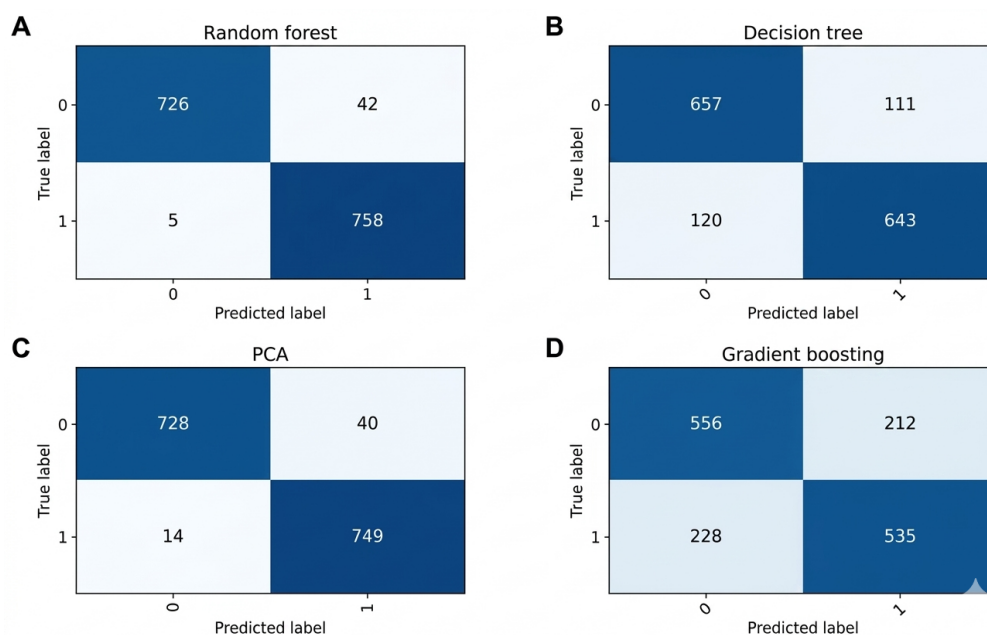


Figure 5. Comparative confusion matrices of classification models. (A) Random Forest. (B) Decision tree. (C) Principal component analysis (PCA). (D) Gradient boosting.

Table 2. Performance metrics for a random forest classifier

Metric type	Precision	Recall	F1-score	Support
Weighted average	0.97	0.97	0.97	1,531
Macro average	0.97	0.97	0.97	1,531
Accuracy	–	–	0.97	1,531

Table 3 shows that the PCA-based classifier achieved an overall accuracy of 80%. Precision indicates that 80% of predicted positives were correct, while recall indicates that 80% of actual positives were correctly identified. The F1-score of 0.79 balances these two measures. The weighted average reflects the class distribution, and the macro average treats all classes equally. The similarity between these metrics suggests relatively balanced performance across classes, although lower than that of the other models.

Table 4 suggests that the SVM classifier achieved the highest performance of 68% across all evaluation metrics, with precision indicating that 68% of predicted positives were accurately classified, and the recall indicating that 68% of actual positives were correctly identified. The

F1-score, balancing the two measures at 0.68, and the weighted average similarly reflect class distribution, while the macro average treats all classes equally.

Table 3. Performance metrics for principal component analysis-preprocessed classifier

Metric type	Precision	Recall	F1-score	Support
Weighted average	0.8	0.8	0.79	1,531
Macro average	0.8	0.8	0.79	1,531
Accuracy	–	–	0.8	1,531

Table 4. Performance metrics for support vector machine

Metric type	Precision	Recall	F1-score	Support
Weighted average	0.68	0.68	0.68	1,531
Macro average	0.68	0.68	0.68	1,531
Accuracy	–	–	0.68	1,531

The best performance and highest accuracy among all the classifiers discussed was achieved by the random forest classifier. The accuracy and recall rates of all algorithms are shown in Figure 6, and it should also be noted that there is a trade-off among models, which is an important factor in model selection.

5.2.1. Model selection

Model selection should be carried out systematically using cross-validation, hyperparameter optimization, and ensemble methods. K-fold cross-validation is applied to reduce overfitting and provide evidence of model generalizability across different subsets of the data. In addition, comparisons with baseline models and assessments of computational efficiency are required to

select the most appropriate classifier.

To ensure fairness and reliability in the evaluation process, it is critical to include preprocessing steps such as missing-value imputation, class balancing, and bias mitigation. A large and representative dataset is also necessary to reduce bias in performance estimation and improve model robustness.

Finally, the performance of machine learning models for heart disease detection should be carefully evaluated using metrics such as accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC-ROC (Figure 7 provides a graphical representation of the capability of ML models to predict heart disease).

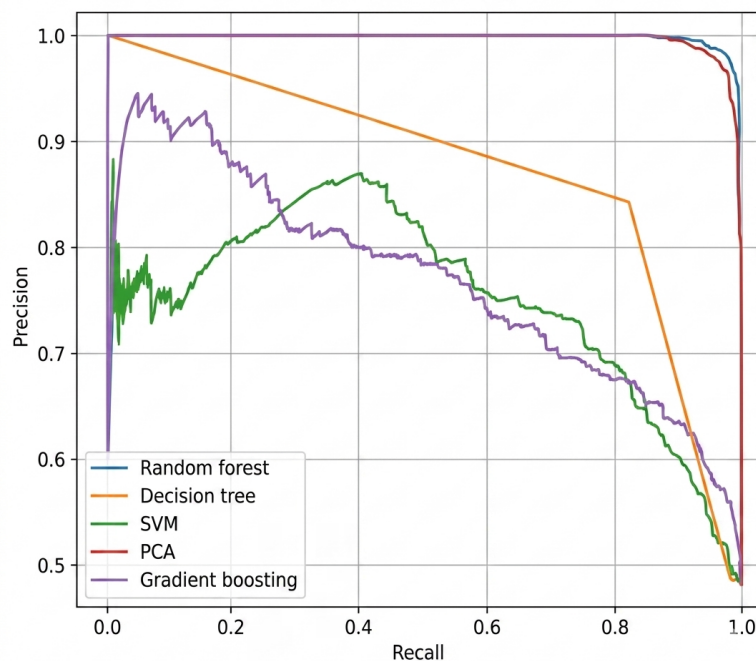


Figure 6. Comparative precision–recall profiles of classification algorithms
Abbreviations: PCA: Principal component analysis; SVM: Support vector machine.

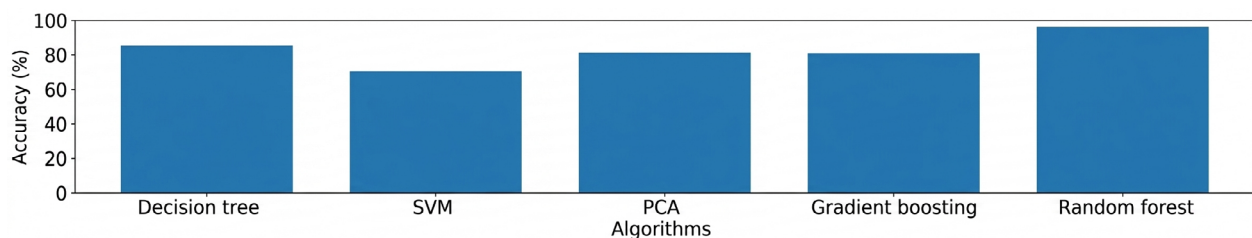


Figure 7. Benchmarking machine learning classifiers for heart disease prediction
Abbreviations: PCA: Principal component analysis; SVM: Support vector machine.

In addition to accuracy metrics, clinical interpretability was considered to ensure practical usefulness. SHAP analysis identified the most significant predictors, including systolic blood pressure, diabetes, and smoking habits; thus, the computational results were consistent with known cardiovascular risk factors. This level of interpretability helps bridge the knowledge gap between statistical performance and clinical decision-making, enabling clinicians to better understand why a specific prediction was made.

6. Conclusion

This paper provides a comprehensive discussion of ML tools used for the early identification and prediction of heart disease. The results highlight a conceptual shift toward data-driven models in improving diagnostic accuracy, optimizing clinical workflows, and enabling personalized treatment approaches. The comparative performance of various classifiers has been systematically evaluated using rigorous assessment metrics.

Combining cross-validation procedures with model selection strategies strengthens the predictive models and enhances their ability to generalize across different patient populations. The proposed learning system has been shown to outperform several existing heart disease detection methods and many commonly used ML algorithms. It is worth noting that the dimensionality reduction improves training efficiency and computational performance without degrading classification accuracy.

The experimental results confirm that the optimized system can serve as a reliable clinical decision-support tool, assisting in the timely identification of cardiac risk and improving patient prognosis. The study provides a valuable resource for practitioners, researchers, and healthcare policymakers aiming to apply ML in cardiovascular diagnostics. Further development and implementation of such intelligent systems have the potential to reduce mortality rates, improve clinical accuracy, and enhance the overall effectiveness of healthcare provision. The proposed evaluation pipeline shows competitive performance across a range of classifiers; however, it is not significantly superior to external benchmarks when evaluated beyond the Framingham dataset.

A key limitation of this study is its reliance on a single dataset, namely the Framingham dataset. External datasets, such as the Cleveland Heart Disease dataset or Medical Information Mart for Intensive Care III, should be used for independent validation to ensure generalizability across populations. Future work should focus on multi-dataset validation and on integrating explainability frameworks to improve clinical interpretability and acceptance. This

work is novel in that it is leakage-free, carefully controls class imbalance without artificial enhancement, and systematically compares both ensemble and baseline classifiers within a reproducible and clinically relevant diagnostic framework.

Acknowledgments

We would like to thank our Vice-Chancellor and the Head of the Department for their support of this work.

Funding

None.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Radha Krishna Jana

Formal analysis: Achyut Mitra

Investigation: Radha Krishna Jana

Methodology: Achyut Mitra

Writing—original draft: Bidisha Bhabani

Writing—review & editing: Tanaya Das

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The dataset used in this study is available on Kaggle (<https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset>).

References

1. Chen LH, Hsiao HD. Feature selection to diagnose a business crisis using a GA-based support vector machine: an empirical study. *Expert Syst Appl.* 2008;35(3):1145-1155. doi: 10.1016/j.eswa.2007.08.010.
2. Temitayo F, Stephen O, Abimbola A. Hybrid GA-SVM for efficient feature selection in e-mail classification. *Comput Eng Intell Syst.* 2012;3(3):17-28.
3. Khammassi C, Krichen S. A GA-LR wrapper approach for feature selection in network intrusion detection. *Comput Secur.* 2017;70:255-277. doi: 10.1016/j.cose.2017.06.005.
4. De Hert M, Detraux J, Vancampfort D. The intriguing relationship between coronary heart disease and mental disorders. *Dialogues Clin Neurosci.* 2018;20(1):31-40.

- doi: 10.31887/DCNS.2018.20.1/mdehert
5. Pathan MS, Nag A, Pathan MM, Dev S. Analyzing the impact of feature selection on the accuracy of heart disease prediction. *Health Anal.* 2022;2:100060.
doi: 10.1016/j.health.2022.100060.
 6. El-Shafiey MG, Hagag A, El-Dahshan ESA, Ismail MA. A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. *Multimed Tools Appl.* 2022;81(13):18155-18179.
doi: 10.1007/s11042-022-12425-x.
 7. Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. *Soft Comput.* 2018;22(2):387-408.
doi: 10.1007/s00500-016-2474-6.
 8. Wang J, Rao C, Goh M, Xiao X. Risk assessment of coronary heart disease based on cloud-random forest. *Artif Intell Rev.* 2022;56(1):203-232.
doi: 10.1007/s10462-022-10170-z.
 9. Vijayashree J, Sultana HP. A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. *Program Comput Softw.* 2018;44(6):388-397.
doi: 10.1134/S0361768818060129.
 10. Tao Z, Huiling L, Wenwen W, Xia Y. GA-SVM-based feature selection and parameter optimization in hospitalization expense modelling. *Appl Soft Comput.* 2019;75:323-332.
doi: 10.1016/j.asoc.2018.11.001.
 11. Huang CL, Wang CJ. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst Appl.* 2006;31(2):231-240.
doi: 10.1016/j.eswa.2005.09.024.
 12. Katarya R, Srinivas P. Predicting heart disease at early stages using machine learning: a survey. In: *Proceedings of the IEEE International Conference Electronics Sustainable Communication Systems (ICESC), July 2-4, 2020, Coimbatore, India*; 2020:302-305.
doi: 10.1109/ICESC48915.2020.9155586.
 13. Nandipati SCR, Chen XY. Polycystic ovarian syndrome (PCOS) classification and feature selection by machine learning techniques. *Appl Math Comput Intell.* 2020;9:65-74.
 14. Coffey S, Roberts-Thomson R, Brown A, et al. Global epidemiology of valvular heart disease. *Nat Rev Cardiol.* 2021;18(12):853-864.
doi: 10.1038/s41569-021-00570-z.
 15. Zhou Y, Zhu X, Cui H, et al. The role of the VEGF family in coronary heart disease. *Front Cardiovasc Med.* 2021;8:738325.
doi: 10.3389/fcvm.2021.738325.
 16. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart Association. *Circulation.* 2022;145(8).
doi: 10.1161/cir.0000000000001052.
 17. Liu A, Diller GP, Moons P, et al. Changing epidemiology of congenital heart disease: effect on outcomes and quality of care in adults. *Nat Rev Cardiol.* 2023;20(2):126-137.
doi: 10.1038/s41569-022-00749-y.
 18. Rajendran R, Karthi A. Heart disease prediction using entropy-based feature engineering and ensembling of machine learning classifiers. *Expert Syst Appl.* 2022;207:117882.
doi: 10.1016/j.eswa.2022.117882.
 19. Ahsan MM, Siddique Z. Machine learning-based heart disease diagnosis: a systematic literature review. *Artif Intell Med.* 2022;128:102289.
doi: 10.1016/j.artmed.2022.102289.
 20. Behera MP, Sarangi A, Mishra D, Sarangi SK. A hybrid machine learning algorithm for heart and liver disease prediction using modified particle swarm optimization with support vector machine. *Procedia Comput Sci.* 2023;218:818-827.
doi: 10.1016/j.procs.2023.01.062.
 21. Ghasemieh A, Lloyed A, Bahrami P, et al. A novel machine learning model with stacking ensemble learner for predicting emergency readmission of heart-disease patients. *Decis Anal J.* 2023;7:100242.
doi: 10.1016/j.dajour.2023.100242.
 22. Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. *Heart.* 2022;108(17):1351-1360.
doi: 10.1136/heartjnl-2021-320131.
 23. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep.* 2024;14(1).
doi: 10.1038/s41598-024-56706-x.
 24. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. *Algorithms.* 2023;16(2):88.
doi: 10.3390/a16020088.
 25. Abdulsalam G, Meshoul S, Shaiba H. Explainable heart disease prediction using ensemble-quantum machine learning approach. *Intell Autom Soft Comput.* 2023;36(1):761-779.
doi: 10.32604/iasc.2023.032262.
 26. Wong TT, Yeh PY. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng.* 2020;32(8):1586-1594.
doi: 10.1109/TKDE.2019.2912815.