Gholipour et al. *Studies in Medical Sciences* (2025) 36(4):57-65
https://doi.org/10.61882/umj.36.3.57
Studies is Medical Sciences

ORIGINAL RESEARCH ARTICLE

Performance Evaluation of Supervised Learning Algorithms in Modeling and Predicting Cardiovascular Disorders

Samad Gholipour[1], Parinaz Eskandarian[1], Hadi Lotfnezhad Afshar[2,3], Sara Mohammadi Kalashani[3]

Published: 25 December 2025

Abstract

Background Cardiovascular disorders remain a leading cause of morbidity and mortality worldwide, underscoring the need for accurate predictive models for early risk identification. Although supervised machine learning offers promising approaches for modeling complex clinical data, comparative evidence on the performance of different algorithms under standardized evaluation settings remains limited. Therefore, this study systematically compared widely used supervised learning classifiers using standardized preprocessing and clinically relevant performance metrics.

Methods The objective of this study was to comparatively evaluate the performance of five supervised learning algorithms for modeling and predicting cardiovascular disorders using a publicly available clinical dataset of 303 patients and 13 routinely collected clinical features. Following data preprocessing, the algorithms were implemented in RapidMiner Studio and evaluated using accuracy, sensitivity, specificity, precision, and F1-score based on an 80/20 train–test split with stratified five-fold cross-validation.

Results Among the evaluated models, the Random Forest algorithm demonstrated the highest overall performance, achieving an accuracy of 91.2%, sensitivity of 88.7%, specificity of 92.9%, precision of 89.4%, and F1-score of 89.0%. The Artificial Neural Network (accuracy 88.7%) and the Support Vector Machine (accuracy 86.9%) showed moderate performance, whereas the Decision Tree yielded the lowest accuracy (83.1%) and sensitivity (79.4%). This superior performance can be attributed to the ensemble structure of Random Forest, which reduces variance and overfitting while effectively capturing nonlinear relationships among clinical variables. Although the dataset size was relatively modest, it is widely used as a benchmark in cardiovascular prediction studies and is considered sufficient for comparative algorithm evaluation.

Conclusion Ensemble-based supervised learning models, particularly Random Forest, demonstrated more robust predictive performance than single classifiers for structured cardiovascular data. By evaluating sensitivity and specificity alongside accuracy, this study provides a transparent comparative baseline for cardiovascular disease prediction. Further validation using larger, independent clinical datasets is needed before these models can be considered for clinical implementation.

Keywords Cardiovascular Disorders, Machine Learning, Data Mining, Supervised Learning, Classification

✉ Samad Gholipour
gholipour.s@umsu.ac.ir

1. Department of Computer Science, School of Engineering, Afagh Higher Education Institute, Urmia, Iran
2. Health and Biomedical Informatics Research Center, Urmia university of medical Sciences, Urmia, Iran
3. Department of Health Information Technology, School of Allied Medical Sciences, Urmia University of Medical Sciences, Urmia, Iran

Received: 10 July 2025 / Revised: 01 October 2025 / Accepted: 21 November 2025

1 Introduction

The human heart, as the central organ responsible for pumping blood and delivering oxygen and nutrients to all cells of the body, is undoubtedly the vital artery of human existence. Its essential role in sustaining life underscores the critical importance of maintaining optimal heart health. However, the prevalence of cardiovascular disorders has risen significantly in recent years, emerging as one of the leading causes of mortality worldwide. Cardiovascular disorders, also referred to as cardiovascular diseases (CVDs), currently represent the primary cause of death globally. According to the World Health Organization, an estimated 17.9 million people died from CVDs in 2020. Over the past decades, cardiovascular disorders have remained the foremost cause of death worldwide.[1]

Cardiovascular diseases (CVDs) comprise a broad range of disorders affecting the heart and blood vessels, including arrhythmias, coronary artery disease, congestive heart failure, and ischemic heart disease. CVDs remain among the leading causes of morbidity and mortality worldwide, accounting for more deaths annually than any other disease category. Early identification of high-risk individuals and accurate prediction of cardiovascular outcomes can facilitate timely interventions, reduce preventable deaths, and improve patients' quality of life.[2,3]

In clinical practice, predicting CVD risk is challenging because cardiovascular outcomes are influenced by multiple interacting factors (e.g., age, sex, blood pressure, cholesterol, glucose status, ECG findings, and exercise-related indices). As a result, relying only on isolated risk factors may overlook complex relationships among variables. In this context, the use of computational approaches that can learn patterns from multivariate clinical datasets becomes increasingly valuable. As emphasized in recent work, early prediction and warning systems based on structured medical data can contribute to prevention and risk reduction strategies.[2] Consequently, traditional rule-based or single-factor risk assessment approaches may be insufficient for capturing complex nonlinear interactions among cardiovascular risk factors.[4,5]

In response to these challenges, the application of data mining and supervised machine learning introduces new dimensions to cardiovascular disorder prediction. Various data mining techniques have been employed to identify and extract useful information from clinical datasets with minimal user input and effort.[6] Over the past decade, researchers have explored different approaches to implementing data mining in healthcare to achieve accurate predictions of cardiovascular disorders. These techniques can support disease diagnosis and prevention through supervised machine learning algorithms that analyze raw data and uncover complex, hidden patterns related to disease occurrence. Such algorithms can identify interactions among dataset features and provide classification models that may improve predictive performance and decision-making in healthcare settings.[4]

Delivering high-quality and cost-effective clinical services remains a fundamental challenge for healthcare organizations. Achieving this goal requires accurate diagnosis, timely identification of high-risk individuals, and avoidance of misdiagnoses that may lead to unnecessary interventions or missed treatment opportunities.[3] Early detection of CVDs is also strongly linked to reduced healthcare costs and lower mortality rates. In this regard, data mining and classification algorithms can provide efficient and relatively low-cost solutions that align with the objectives of clinical research and evidence-based healthcare.[7,8]

A range of studies has investigated machine learning methods for cardiovascular disorder prediction using various datasets and modeling strategies. Dehkordi and Sajedi proposed a prescription-based predictive model using data mining techniques. They introduced an algorithm called Skating to enhance system accuracy and compared four classification algorithms: Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Skating under different labeling scenarios. Their findings indicated that Skating was the most accurate classifier, achieving an accuracy of 73.17%, which highlights both the feasibility of prediction and the need for more robust models in clinical-grade settings.[7]

In another study, standard datasets from the UCI repository (Cleveland and Hungary) were used to evaluate five classification algorithms: Random Forest (RF), Neural Networks, NB, regression-based classification, and Support Vector Machines (SVM).[9] Their results showed that regression-based methods yielded lower performance, while RF achieved the highest accuracy, reaching 98.136%. Similarly, Soni et al. employed DTs combined with a Genetic Algorithm and compared this approach with NB and clustering-based methods.[10] Their proposed system achieved an accuracy of 99.2%. Mansoor et al. evaluated Logistic Regression (LR) and RF for estimating cardiovascular risk and reported that LR slightly outperformed RF (89% vs. 88%) in their specific clinical context.[11]

Austin also compared tree-based approaches and highlighted the strong performance of conventional statistical models, such as LR, in predicting HD in certain settings.[4] However, reported performance values vary substantially across studies, even when similar benchmark datasets are used, highlighting the strong influence of preprocessing choices, validation strategies, and experimental design. Additional work has examined feature engineering and model selection. Le et al. applied

three classification methods to 58 features from the UCI dataset and reported that a linear SVM achieved an accuracy of 89.93%.[12]

Tarawneh and Embarak proposed a hybrid approach using 12 features and compared it with several algorithms, including KNN, DT, Artificial Neural Network (ANN), SVM, and NB.[13] Their hybrid method achieved 89.2% accuracy and outperformed the alternative models in that study. Chitra suggested a cascaded neural network architecture and compared it with SVM, reporting improved accuracy for the proposed approach (85% vs. 82%).[14] Latha and Jeeva combined majority voting with multiple classifiers and feature selection and achieved 85.48% accuracy, demonstrating the potential of ensemble strategies.[15] Mohan et al. proposed a hybrid model (HRFLM) using the Cleveland dataset and achieved 88.7% accuracy, reporting improvements over several baseline algorithms.[16]

Despite the extensive body of research on cardiovascular disorder prediction using data mining and machine learning techniques, several limitations remain evident in the existing literature. Many previous studies have focused primarily on reporting overall accuracy as the main performance indicator, while clinically critical metrics such as sensitivity and specificity, particularly the ability to correctly identify high-risk patients, have received comparatively less attention. Moreover, a considerable number of studies have relied on a single algorithm or narrowly focused modeling frameworks, which restricts comprehensive performance comparison across widely used supervised learning approaches. Recent research trends emphasize the need for systematic evaluation of multiple algorithms under uniform preprocessing and evaluation conditions, as well as the importance of model robustness and interpretability for clinical applicability.[2,5] As a result, it remains unclear which commonly used supervised learning algorithms provide the most balanced and clinically relevant performance when evaluated under identical preprocessing and validation conditions.

In addition, although ensemble-based methods such as RF and gradient-boosting approaches have demonstrated promising results in recent cardiovascular prediction studies, there is still no consensus regarding their comparative effectiveness against classical supervised learning algorithms when applied to standardized clinical datasets.[2] Recent works have highlighted the growing importance of explainable and reliable machine learning models in healthcare, especially in cardiovascular risk prediction, where model transparency and balanced performance play a critical role in clinical acceptance. These gaps indicate the necessity for a structured and comparative evaluation of commonly used supervised learning algorithms within a unified experimental framework.[5,8]

In light of the above methodological gaps and inconsistencies in prior work, the aim of this study is to systematically evaluate and compare the performance of widely used supervised learning algorithms, including DT, RF, SVM, KNN, and ANN, in modeling and predicting cardiovascular disorders using a standardized and well-established clinical dataset. This study focuses not only on overall accuracy but also on clinically relevant performance metrics such as sensitivity, specificity, and F1-score to ensure meaningful interpretation in medical contexts. By providing a comprehensive comparative analysis, the findings of this research aim to contribute to the selection of reliable and cost-effective predictive models that can support early diagnosis and decision-making processes in cardiovascular healthcare systems.

2 Methods

Study Design

The present study is a descriptive, analytical modeling study conducted to evaluate and compare the predictive performance of supervised machine learning algorithms for cardiovascular disorder prediction. Unlike interventional or implementation-based studies, the primary focus of this research is the analytical assessment of model behavior and performance using structured clinical data under a unified experimental framework. This design enables an objective comparison of different algorithms based on clinically relevant evaluation metrics rather than real-time clinical deployment.

The analytical process involved training and testing five supervised learning algorithms: DT, RF, SVM, KNN, and ANN on preprocessed clinical data.[9] These algorithms were selected due to their widespread use in cardiovascular prediction studies, representativeness of different learning paradigms (tree-based, distance-based, margin-based, and neural models), and their frequent application in comparable works reported in the literature.[15] The main research question addressed in this study was:

Which supervised learning algorithm demonstrates the most balanced and reliable performance in predicting cardiovascular disorders when evaluated using standardized clinical metrics? Specifically, we compared models using accuracy, sensitivity, specificity, precision, and F1-score, with emphasis on sensitivity due to the clinical cost of false negatives.

All modeling procedures were implemented using RapidMiner Studio, a widely used data mining and machine learning platform that enables reproducible analytical workflows and supports standardized evaluation of classification models.

Algorithm Selection Rationale

The selected models represent distinct and commonly used learning paradigms in clinical tabular prediction:

tree-based (DT, RF), margin-based (SVM), instance-based (KNN), and nonlinear function approximation (ANN). These algorithms are frequently reported in cardiovascular prediction literature and provide a meaningful baseline for comparative evaluation under consistent preprocessing and testing conditions. The selection was guided by their frequent use as baseline comparators in cardiovascular prediction studies.[17]

For SVM, a radial basis function (RBF) kernel was employed to capture nonlinear decision boundaries. The ANN model consisted of an input layer, two hidden layers with 64 and 32 neurons, and a sigmoid output layer for binary classification.

Data Collection

Data were obtained from a publicly available Kaggle repository that mirrors the Cleveland HD benchmark dataset. In this study, the dataset was obtained from the publicly available Kaggle repository, which aggregates cardiovascular datasets originally derived from clinical centers in the United States and Europe. These datasets are commonly used as benchmark data in cardiovascular prediction research and have been extensively cited in previous studies.[12,16]

The dataset was obtained from a Kaggle repository that mirrors the Cleveland HD benchmark. The Kaggle version was accessed on July 6, 2025; the original data-collection period is not consistently reported in the public copy.

The dataset includes 303 patient records with 13 clinically relevant features, such as age, sex, blood pressure, serum cholesterol, fasting blood glucose, electrocardiographic results, exercise-induced angina, and heart rate-related indicators. The target variable follows a binary classification scheme, indicating the presence or absence of cardiovascular disorders (Table 1).

Although the dataset size is relatively moderate, it has been widely adopted in comparative modeling studies and is considered sufficient for benchmarking and algorithm performance evaluation. The diversity of clinical attributes within the dataset supports meaningful comparative analysis across different supervised learning approaches.

Data Preprocessing

Data preprocessing is a critical step in ensuring the reliability and stability of machine learning models when applied to real-world clinical datasets. Raw medical data often contain missing values, noise, outliers, and heterogeneous feature scales, all of which can negatively affect model performance if left unaddressed.

Initially, the dataset was examined for missing values. If any variable exhibited excessive missingness, it was excluded; otherwise, remaining missing values were handled using statistical imputation methods, including mean substitution for numerical variables and mode substitution for categorical variables. Outliers were identified using interquartile range (IQR) and Z-score techniques, and extreme values were either normalized or adjusted based on their statistical deviation from the data distribution.

Categorical variables were transformed into numerical representations using One-Hot Encoding, allowing machine learning algorithms to process qualitative features without imposing artificial ordinal relationships. Since algorithms such as SVM, KNN, and ANN are sensitive to feature scale, Z-score normalization was applied to standardize numerical variables by centering them around a mean of zero with unit variance. This step ensured balanced feature contribution during model training.

Table 1 List of selected key features

Feature	Description	Field values
Age	Patient age	Numeric
Sex	Gender	1 = male, 0 = female
Cp	Chest pain type	1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic
Testbps	Resting blood pressure	Numeric
Chol	Serum cholesterol	Numeric
Fbs	Fasting blood sugar	1 = yes, 0 = no
Restecg	Resting electrocardiogram results	0 = normal, 1 = st-t abnormality, 2 = left ventricular hypertrophy
Thalach	Maximum recorded heart rate	Numeric
Exang	Exercise-induced angina	1 = yes, 0 = no
Oldpeak	ST depression induced by exercise relative to rest	Numeric
Slope	Slope of the peak exercise ST segment (indicates exercise capacity)	1 = upsloping, 2 = flat, 3 = downsloping
Ca	Number of major vessels colored by fluoroscopy	0–3
Thal	Thalassemia test result / Angiographic result	1 = normal, 2 = fixed defect, 3 = reversible defect

Model Training and Evaluation

Following preprocessing, the data were split into training (80%) and test (20%) sets to preserve an independent hold-out evaluation, while internal model development used stratified five-fold cross-validation (repeated five times) on the training set to improve stability and reduce split-dependence. This evaluation strategy is consistent with widely used accuracy-estimation and model-selection recommendations in the machine learning literature.[18] The final reported metrics (accuracy, sensitivity, specificity, precision, and F1-score) were computed on the held-out test set, with particular emphasis on sensitivity due to its clinical importance in minimizing false-negative predictions. Class distribution was examined; no resampling was applied unless the imbalance was substantial. Figure 1 shows the overall schematic of the algorithm modeling.

Table 2 Statistical characteristics of the dataset used

Feature	Min	Max	Mean
Age	29	77	54.37
Sex	0	1	0.68
Cp	0	3	0.97
Testbps	94	200	131.62
Chol	126	564	246.26
Fbs	0	1	0.15
Restecg	0	2	0.53
Thalach	71	202	149.65
Exang	0	1	0.33
Oldpeak	0	6.2	1.04
Slope	0	2	1.40
Ca	0	4	0.73
Thal	0	3	2.31

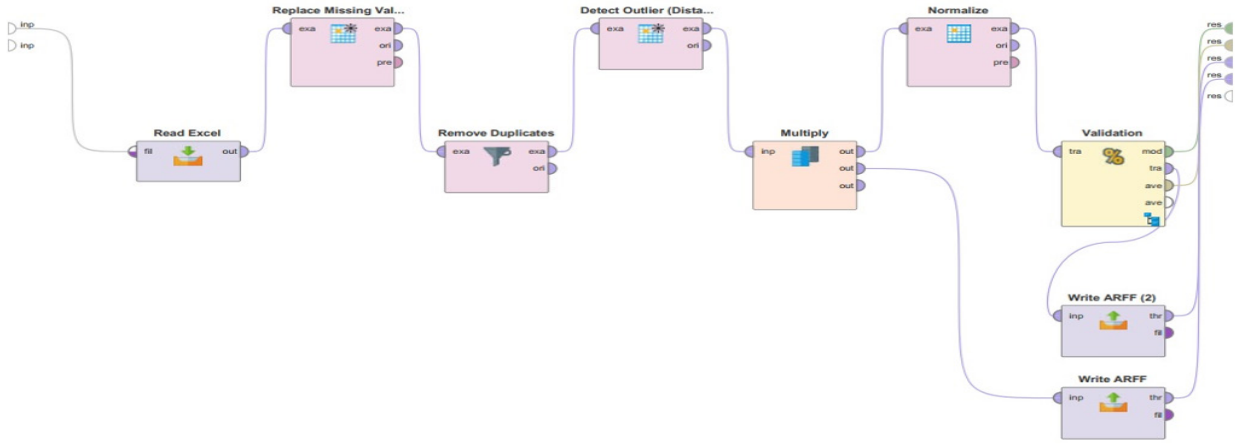

Figure 1 Overall schematic of the algorithm modeling

3 Results

Following the completion of model training and evaluation, this section presents the comparative results of the supervised learning algorithms applied to the cardiovascular dataset. The purpose of this section is to report the quantitative performance outcomes of each model based on predefined evaluation metrics, without interpretative analysis.

Dataset characteristics

Descriptive statistics of the dataset, including minimum, maximum, and mean values for each feature, are presented in Table 2.

Overall Performance of the Models

All supervised learning algorithms were evaluated using accuracy, sensitivity, specificity, precision, and F1-score metrics. The summarized performance results of the five models are presented in Table 3, providing a comprehensive comparison across all evaluation criteria.

Accuracy Comparison

Accuracy was used as an initial indicator of overall classification performance. As shown in Table 3, the RF model achieved the highest accuracy (91.2%), followed by the ANN with 88.7% and the SVM with 86.9%. The KNN classifier achieved an accuracy of 86.5%, while the DT model yielded the lowest accuracy at 83.1%.

Sensitivity and Specificity Analysis

Sensitivity and specificity were analyzed to assess the models' abilities to correctly identify patients with cardiovascular disorders and healthy individuals, respectively. The RF model demonstrated the highest sensitivity (88.7%) among the evaluated models. ANN and SVM followed with sensitivities of 85.6% and 85.3%, respectively. The DT model showed the lowest sensitivity (79.4%).

In terms of specificity, RF again achieved the highest value (92.9%), followed by SVM (91.1%) and ANN (89.1%). These results indicate variability in the models' abilities to distinguish between diseased and non-diseased cases.

Table 3 Summary of algorithm evaluation results

Algorithm	Accuracy	Sensitivity	Specificity	Precision	F1 Score
RF	91.2%	88.7%	92.9%	89.4%	89.0%
KNN (k = 5)	86.5%	83.2%	88.3%	84.1%	83.6%
ANN	88.7%	85.6%	89.1%	86.2%	85.9%
DT	83.1%	79.4%	85.6%	81.1%	80.2%
SVM	86.9%	85.3%	91.1%	87.4%	86.3%

F1-Score and Precision

The F1-score, which reflects the balance between precision and sensitivity, further differentiated model performance. RF achieved the highest F1-score (89.0%), followed by SVM (86.3%) and ANN (85.9%). KNN and DT demonstrated lower F1-scores of 83.6% and 80.2%, respectively. Precision values showed a similar trend, with RF yielding the highest precision (89.4%). Figure 2 illustrates the comparative performance of all evaluated algorithms across the selected metrics.

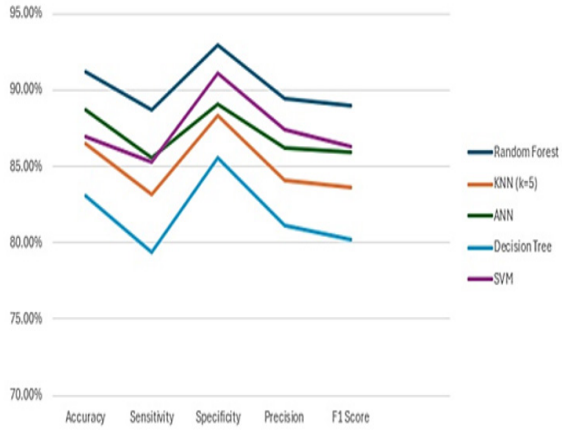

Figure 2 Comparative analysis of algorithms

4 Discussion

This study contributes to the existing literature on cardiovascular disorder prediction by providing a structured and comparative evaluation of widely used supervised learning algorithms under a unified preprocessing and validation framework. Unlike many previous studies that primarily emphasize overall accuracy or focus on a single modeling approach, this work systematically compares multiple classifiers using clinically relevant performance metrics, with particular attention to sensitivity and specificity. By using a well-established benchmark dataset and consistent evaluation procedures, the findings help clarify the relative strengths and limitations of commonly applied algorithms in cardiovascular risk prediction. The results highlight the practical value of ensemble-based models, particularly RF, in achieving balanced predictive performance, thereby providing methodological insights for future analytical and research-oriented applications in cardiovascular machine learning.

The present study evaluated the performance of several widely used supervised learning algorithms: DT, KNN, SVM, ANN, and RF for predicting cardiovascular disorders using a structured clinical dataset. Overall, ensemble-based and nonlinear models demonstrated higher quantitative performance across multiple metrics compared to single-tree and distance-based classifiers. The results demonstrated that ensemble-based and nonlinear models outperform simpler classifiers in this context, with the RF algorithm achieving the most balanced and robust performance across evaluation metrics.

From a clinical perspective, sensitivity is a critical metric, as failure to identify true positive cardiovascular cases may lead to delayed intervention and severe outcomes. Therefore, models were primarily compared based on their ability to minimize false negatives rather than accuracy alone.

A key reason RF performed best in our experiments is its ensemble nature: by aggregating multiple decorrelated DTs, it reduces variance, improves generalization, and remains robust to noise and moderate feature interactions. These properties are particularly valuable in small-to-medium clinical tabular datasets, where relationships among variables such as age, chest pain type, ST depression, and coronary vessel count are often nonlinear and interdependent. Unlike single DTs, which are prone to overfitting and instability, RF mitigates these limitations through bootstrap aggregation and randomized feature selection. In addition, RF does not require extensive feature engineering and provides intrinsic feature-importance measures, supporting interpretability, an essential requirement for clinical decision-support systems. The observed robustness of RF is particularly relevant for small-to-moderate clinical datasets, where nonlinear interactions and feature dependencies are common.

Our findings are consistent with prior evidence suggesting that ensemble and hybrid approaches can improve predictive stability and accuracy in cardiovascular risk modeling. For example, a hybrid RF–Linear Model approach has been reported to enhance prediction performance compared with standalone classifiers on structured heart-disease datasets.[16] Similarly, hybrid

and ensemble strategies have been shown to outperform individual algorithms such as KNN, DT, and NB in comparable cardiovascular prediction tasks.[13] Together, these studies support our observation that combining multiple learners or decision boundaries can lead to more stable decision-making in cardiovascular classification.

At the same time, the performance values reported in the literature vary substantially even when similar benchmark datasets are used, highlighting the importance of methodological design. A recent systematic study that examined multiple feature-selection families (filter, wrapper, and evolutionary approaches) on the Cleveland HD dataset reported that the best-performing configuration achieved a reported best performance of around mid-80%, depending on configuration, and that feature selection could reduce performance for some models, depending on the setup. In contrast, our RF model achieved an accuracy above 91% with a favorable balance between sensitivity and specificity. Such discrepancies may be explained by differences in preprocessing decisions, model-tuning depth, and the operationalization of train/test splitting and cross-validation. Importantly, the evidence suggests that preprocessing and evaluation protocols can meaningfully influence which algorithm appears "best," rather than classifier choice alone.[5]

Comparable conclusions arise when contrasting our results with other recent modeling studies. For instance, a machine-learning-based diagnostic model for HD has been reported with performance values around the mid-0.84 range (including accuracy and AUC depending on configuration). While these results are competitive, the higher performance observed in our study may reflect differences in dataset versions, class distributions, encoding and scaling pipelines, and validation design. This comparison reinforces that reported performance metrics in cardiovascular prediction studies can be strongly pipeline-dependent and should be interpreted cautiously unless preprocessing and validation steps are transparently described.[8]

Beyond comparing classical classifiers, recent research has increasingly emphasized training strategies that improve learning efficiency and evaluation reliability. For example, active-learning-based frameworks have been proposed to enhance model efficiency and robustness by intelligently selecting informative samples during model development. Although our study focused on classical supervised learners rather than active/adaptive learning, such evidence supports the broader conclusion that careful design of the learning process, data handling, validation, and interpretability can be as influential as the algorithm choice itself.[19]

Our findings also align with recent syntheses of machine-learning approaches for HD prediction, which highlight that tree-based ensembles and boosting-family methods often demonstrate strong performance on structured clinical features, especially when validation and feature handling are conducted rigorously. This supports our observation that models capable of capturing nonlinear relationships and feature interactions, such as RF, tend to perform well in cardiovascular risk classification tasks.[2] Feature-importance analysis consistently highlighted chest pain type, maximum heart rate (thalach), ST depression (oldpeak), and resting blood pressure as dominant predictors, aligning well with established clinical risk indicators.

Despite the promising results, several limitations should be acknowledged. The dataset used in this study contains 303 samples, which limits statistical power and may reduce generalizability. Small datasets can yield optimistic performance estimates, particularly if evaluation splits are favorable or if tuning inadvertently adapts to the evaluation protocol. Although internal cross-validation was employed to improve robustness, external validation on independent cohorts remains essential before clinical deployment. Future studies should also consider reporting confidence intervals and complementary metrics (e.g., calibration) to strengthen clinical interpretability.

In practical terms, RF appears to be a suitable candidate for early screening and decision-support prototypes due to its strong discrimination ability, relative robustness to noise, and interpretability through feature importance analysis. Nevertheless, real-world clinical adoption should be preceded by external validation, prospective testing, and safety evaluation to ensure reliable performance across diverse populations and care settings.

5 Conclusion

This study systematically compared five widely used supervised machine learning algorithms for predicting cardiovascular disorders using a standardized clinical benchmark dataset. Among the evaluated models, Random Forest demonstrated the most balanced overall performance, achieving superior accuracy while maintaining a favorable balance between sensitivity and specificity. These findings highlight the potential of ensemble-based methods as reliable analytical tools for cardiovascular risk prediction and provide a transparent comparative baseline for future research. However, the results are based on a moderate-sized public dataset with internal validation and should not be interpreted as evidence of clinical readiness. Further studies should validate these findings using larger, independent clinical cohorts and prospective evaluation before implementation in real-world healthcare settings.

Declarations

Acknowledgments
The authors sincerely thank the esteemed faculty members of the Computer Engineering Department at Afagh Higher Education Institute of Urmia and the faculty of Health Information Technology at Urmia University of Medical Sciences for their invaluable support and guidance throughout this research. Their expertise and dedication played a pivotal role in the successful completion of this study.

Artificial Intelligence Disclosure
The authors used ChatGPT (OpenAI) and Grammarly solely to improve the grammar, language, and overall clarity of the manuscript. These tools were not used to generate, analyze, or interpret the study data, develop the study design, or draw scientific conclusions. All scientific content, analyses, interpretations, and final manuscript revisions were performed by the authors, who take full responsibility for the accuracy and integrity of the work.

Authors' Contributions
Samad Gholipour contributed to data collection, supervision, and manuscript writing. Dr. Parinaz Eskandarian contributed to data analysis and manuscript writing. Dr. Hadi Lotfnezhad Afshar contributed to supervision, review, and editing. Sara Mohammadi Kalashani contributed to data analysis and editing.

Availability of Data and Materials
For this study, publicly available datasets and cardiovascular disorder-related data collected from Kaggle were used, which are accessible to the general public.

Conflict of Interest
The authors declare that they have no conflicts of interest in relation to the publication of this manuscript.

Consent for Publication
Not applicable.

Ethical Considerations
This study was conducted using a publicly available, de-identified dataset obtained from the University of California, Irvine (UCI) Machine Learning Repository. The dataset contains no personally identifiable information, and no direct involvement of human participants or access to confidential patient records occurred. Therefore, ethical approval and informed consent were not required for this study.

Funding
The authors declare that no funding or financial support was received for conducting this study or preparing this manuscript.

References

1. World Health Organization. Cardiovascular diseases [Internet]. Geneva: World Health Organization; 2020 [cited 2026 Jul 22]. Available from: https://www.who.int/health-topics/cardiovascular-diseases.
2. Islam MA, Majumder MZH, Miah MS, Jannaty S. Precision healthcare: A deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction. Comput Biol Med. 2024;176:108432. doi:10.1016/j.compbiomed.2024.108432.
3. Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Comput Biol Med. 2021;136:104672.doi:10.1016/j.compbiomed.2021.104672.
4. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. J Clin Epidemiol. 2013;66(4):398-407. doi:10.1016/j.jclinepi.2012.11.008.
5. Noroozi Z, Orooji A, Erfannia L. Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. Sci Rep. 2023;13(1):22588. https://doi.org/10.1038/s41598-023-49962-w.
6. Srinivas, K., B.K. Rani, and A. Govrdhan, Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2010. 2(2): 250–255.
7. Dehkordi SK, Sajedi H. Prediction of disease based on prescription using data mining methods. Health Technol. 2019;9(1):37-44. doi:10.1007/s12553-018-0246-2.
8. Al Bataineh A, Manacek S. MLP-PSO Hybrid Algorithm for Heart Disease Prediction. Journal of Personalized Medicine. 2022;12(8):1208. doi:10.3390/jpm12081208.
9. Jan M, Awan AA, Khalid MS, Nisar S. Ensemble approach for developing a smart heart disease prediction system using classification algorithms. Research Reports in Clinical Cardiology. 2018 ;9:33-45. doi:10.2147/RRCC.S172035.
10. Soni J, Ansari U, Soni S, Sharma D. Predictive data mining for medical diagnosis: An overview of heart disease prediction. International journal of computer applications. 2011;17(8):43-8.
11. Mansoor H, Elgendy IY, Segal R, Bavry AA, Bian J. Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: A machine learning approach. Heart & Lung. 2017;46(6):405-11. doi:10.1016/j.hrtlng.2017.09.003.
12. Le HM, Tran TD, Van Tran L. Automatic heart disease prediction using feature selection and data mining technique. Journal of Computer Science and Cybernetics. 2018;34(1):33-48.
13. Tarawneh M, Embarak O, editors. Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques. Advances in Internet, Data and Web Technologies; 2019 2019//; Cham: Springer International Publishing. doi:10.1007/978-3-030-12839-5_41.
14. Chitra R, Seenivasagam V. Heart disease prediction system using supervised learning classifier. Bonfring International Journal of Software Engineering and Soft Computing. 2013;3(1):1-7.

15. Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked. 2019;16:100203. doi:10.1016/j.imu.2019.100203.

16. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE access. 2019;7:81542-54.

17. Ramalingam V, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology. 2018;7(2.8):684-7.

18. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI); 1995 Aug 20–25; Montreal, Quebec, Canada. San Francisco (CA): Morgan Kaufmann Publishers; 1995: 1137-1143

19. El-Hasnony IM, Elzeki OM, Alshehri A, Salem H. Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction. Sensors [Internet]. 2022; 22(3):1184. doi: 10.3390/s22031184.