INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

DEVELOPMENT AND COMPARATIVE EVALUATION OF MACHINE LEARNING MODELS FOR PERSONALIZED CARDIOVASCULAR RISK PREDICTION USING CLINICAL FEATURES

Dr. Muhammad Ali[1], Abdul Mussawar[*2], Muhammad Abdul Rehman[3], Faheem Haroon[4]

[1]*Department of Physics and Applied Mathematics (DPAM), Pakistan Institute of Engineering and Applied Sciences (PIEAS), Pakistan.*
[*2]*Department of Computer Science, University of Peshawar, 25120, Pakistan.*
[3]Department of Commerce, The Islamia University of Bahawalpur, Punjab, Pakistan
[3]*Bahria School of Engineering and Applied Sciences, Bahria University Islamabad, Pakistan.*

[1]faheemharoon.buic@bahria.edu.pk, [*2]mussawairabdul@gmail.com, [3]marehman1992@gmail.com, [4]muhammadali@pieas.edu.pk

DOI: https://doi.org/10.5281/zenodo.24133183

Keywords
Machine learning, Cardiovascular disease, Personalized medicine, Risk classification, Gradient boosting; Clinical prediction.

Article History
Received: 05 August 2026
Accepted: 19 October 2026
Published: 06 November 2026

Copyright @Author
Corresponding Author: *
Abdul Mussawar

Abstract

Cardiovascular disease remains a major global health challenge, and the development of reliable computational approaches for individualized disease classification may support improved clinical decision-making. In this study, machine learning algorithms were developed and comparatively evaluated for the classification of heart disease using routinely available demographic, physiological, and clinical characteristics. The Cleveland Heart Disease dataset obtained from the UCI Machine Learning Repository was used as the study dataset. Four predictive approaches—logistic regression, random forest, support vector machine, and gradient boosting—were implemented to examine their ability to distinguish individuals according to the presence or absence of heart disease. A systematic analytical framework was applied, including missing-value treatment, categorical-variable encoding, feature scaling where appropriate, model optimization, and cross-validation. The predictive performance of the models was assessed using multiple complementary evaluation metrics, including the area under the receiver operating characteristic curve, sensitivity, specificity, precision, F1-score, and calibration. In addition to overall predictive performance, feature-importance analysis was incorporated to identify the demographic and clinical characteristics that contributed most substantially to model predictions and to improve the interpretability of the analytical framework. The comparative evaluation was designed to investigate whether nonlinear machine learning algorithms could provide advantages over conventional logistic regression for cardiovascular disease classification. The findings also demonstrate the potential value of interpretable predictive models as supportive tools for individualized cardiovascular assessment, while emphasizing that algorithmic predictions should complement rather than replace clinical judgment. Because the selected dataset was relatively small and contained historical diagnostic information, the results should be interpreted as retrospective classification findings and should not be considered equivalent to prospective prediction of future cardiovascular events. Overall, this study provides a reproducible

machine learning framework for comparing conventional and nonlinear predictive approaches and establishes a methodological foundation for future research using larger, contemporary, and externally validated datasets in personalized cardiovascular healthcare.

1. INTRODUCTION

Personalized medicine has become an important direction in contemporary healthcare because patients with similar clinical diagnoses may demonstrate substantial differences in disease susceptibility, clinical progression, and treatment response. The integration of patient-specific clinical characteristics into predictive models provides opportunities to improve disease identification and clinical decision-making. Traditional cardiovascular assessment frequently relies on established clinical risk factors and statistical prediction approaches. However, complex interactions among demographic, physiological, and diagnostic characteristics may require more flexible analytical methods [1]. Cardiovascular disease remains a major contributor to global morbidity and mortality. Early identification of individuals with cardiovascular abnormalities is important for timely clinical assessment and preventive management [2]. Commonly evaluated cardiovascular characteristics include age, blood pressure, serum cholesterol, glucose regulation, exercise-related symptoms, and electrocardiographic measurements. The combined assessment of these variables may provide useful information for distinguishing individuals with and without diagnosed cardiovascular disease [3].

Machine learning offers computational approaches for identifying patterns within clinical datasets. Unlike conventional statistical models that frequently assume predefined relationships between variables, selected machine learning algorithms can investigate nonlinear associations and interactions among predictors [4]. Logistic regression remains an important benchmark because of its simplicity and interpretability, whereas random forest and gradient boosting can represent more complex relationships [5]. Support vector machines provide another approach for clinical classification by identifying decision boundaries that separate observations into different outcome groups [6]. Random forest models combine multiple decision trees to improve prediction stability, while gradient boosting sequentially develops predictive learners to reduce classification errors [7]. The comparative evaluation of these algorithms can help determine their relative suitability for cardiovascular classification.

Electronic health records and clinical datasets provide opportunities to investigate patient-specific disease patterns. Clinical variables such as age, sex, blood pressure, cholesterol concentration, chest pain characteristics, and maximum heart rate can be incorporated into predictive models [8]. The availability of these variables makes cardiovascular disease classification an appropriate application for investigating machine learning approaches. Previous investigations have demonstrated the potential of machine learning in cardiovascular prediction. However, reported performance may vary according to dataset characteristics, preprocessing procedures, validation strategies, and algorithm selection [9]. Small datasets may also produce unstable estimates, particularly when models are evaluated using a single train-test split [10]. An additional concern is the interpretability of machine learning predictions. Clinical implementation requires an understanding of the characteristics influencing individual predictions rather than reliance on unexplained probability scores [11]. Feature-importance approaches and explainable artificial intelligence can help identify the variables contributing to model decisions [12].

Despite increasing interest in machine learning-based cardiovascular assessment, comparative evaluation should consider discrimination, calibration, generalizability, and clinical usefulness. A model with high classification accuracy may still provide unreliable probability estimates or perform poorly in different patient populations [13]. Appropriate validation is therefore essential when developing predictive systems for clinical applications [14]. The present investigation proposes a comparative machine learning framework using the

Cleveland Heart Disease dataset. The study focuses on the development and evaluation of four algorithms using routinely collected clinical variables. It further investigates the contribution of individual clinical characteristics to predictive classification.

The present study developed machine learning models for cardiovascular disease classification using clinical and demographic characteristics and compared the predictive performance of logistic regression, random forest, support vector machine, and gradient boosting algorithms. Furthermore, the study evaluated model performance using discrimination, classification, and calibration measures, identified important clinical predictors contributing to cardiovascular disease classification, and investigated the potential application of interpretable machine learning in personalized cardiovascular assessment.

2. Materials and Methods

2.1 Research design

The study followed a retrospective computational research design using an openly accessible clinical dataset. The analytical framework included data preprocessing, feature engineering, model development, hyperparameter optimization, cross-validation, and comparative performance assessment. The investigation was conducted as a secondary analysis of de-identified data. No new patient recruitment or clinical intervention was involved.

2.2 Data source and dataset description

The Cleveland Heart Disease dataset was obtained from the UCI Machine Learning Repository. The dataset contained 303 observations and 13 predictor variables representing demographic, physiological, and clinical characteristics. The original diagnostic outcome ranged from 0 to 4, with 0 indicating absence of heart disease and values 1–4 indicating its presence. For binary classification, the outcome was recoded as follows:

- 0 = No diagnosed heart disease.
- 1 = Diagnosed heart disease (original values 1–4).

The dataset contained missing observations in selected variables, particularly the number of major vessels and thalassemia-related diagnostic information. These observations were handled through appropriate imputation procedures.

2.3 Variables included in the investigation

Table 1. Description of clinical and demographic variables.

Variable	Description	Data type
Age	Patient age in years	Continuous
Sex	Biological sex as coded in dataset	Categorical
CP	Chest pain category	Categorical
Trestbps	Resting blood pressure	Continuous
Chol	Serum cholesterol	Continuous
FBS	Fasting blood sugar indicator	Binary
Restecg	Resting electrocardiographic result	Categorical
Thalach	Maximum heart rate achieved	Continuous
Exang	Exercise-induced angina	Binary
Oldpeak	Exercise-related ST depression	Continuous
Slope	Peak exercise ST-segment slope	Categorical
CA	Number of major vessels visualized	Ordinal
Thal	Thalassemia-related diagnostic category	Categorical
Target	Presence or absence of diagnosed heart disease	Binary

2.4 Data preprocessing

Data preprocessing was conducted before model development to improve consistency and reduce analytical errors. The following procedures were applied: Missing numerical observations were imputed using the median, whereas missing categorical observations were imputed using the most frequent category. Imputation parameters

were estimated within each training fold to avoid information leakage. Categorical variables were transformed into numerical representations using appropriate encoding procedures. Ordinal variables retained their meaningful ordering where justified. Continuous variables were standardized for logistic regression and support vector machine models. Tree-based algorithms were evaluated without unnecessary scaling. The original multiclass diagnostic outcome was converted into a binary classification variable. All preprocessing operations, including imputation, scaling, and feature selection, were performed within the training pipeline.

2.5 Machine learning algorithms

Four machine learning algorithms were evaluated in this study. Logistic regression was used as the conventional statistical benchmark to estimate the probability of cardiovascular disease based on the predictor variables. The model was expressed using the following logistic function:

$$P(Y=1)=\frac{1}{1+e^{-\left(\beta_0+\sum_{i=1}^{n}\beta_i X_i\right)}}$$

where P(Y=1) represents the probability of cardiovascular disease, $\beta_0$is the intercept, represents the regression coefficient for each predictor, and (X_i) denotes the corresponding predictor variable. Random forest was used to generate an ensemble of decision trees developed from bootstrap samples and randomly selected subsets of predictors, with final predictions obtained by aggregating the outputs of individual trees. The support vector machine (SVM) was employed to identify an optimal decision boundary separating observations into cardiovascular disease and non-disease classes. A radial basis function (RBF) kernel was used to account for potential nonlinear relationships among the predictor variables. Gradient boosting was also evaluated by sequentially developing decision-tree learners, in which each subsequent learner was trained to minimize the prediction errors of the preceding ensemble.

2.6 Model training and validation

The primary evaluation strategy used stratified repeated five-fold cross-validation. This approach preserved the approximate outcome distribution across folds and reduced dependence on a single data partition. Hyperparameters were optimized using an appropriately nested cross-validation procedure. Model development and parameter optimization were restricted to the training portions of the folds. Given the limited dataset size, results were interpreted cautiously.

Table 2. Machine learning models and proposed optimization parameters

Model	Parameters for optimization	Evaluation purpose
Logistic regression	Regularization strength, penalty	Baseline classification
Random forest	Number of trees, maximum depth, minimum leaf size	Nonlinear classification
Support vector machine	Kernel, C, gamma	Margin-based classification
Gradient boosting	Learning rate, number of estimators, tree depth	Ensemble classification

2.7 Model performance evaluation

Model performance was assessed using several complementary measures. Accuracy represented the proportion of correctly classified observations. Sensitivity measured the proportion of patients with diagnosed disease correctly identified by the model. Specificity measured the proportion of patients without diagnosed disease correctly classified. Precision and F1-score were additionally calculated to evaluate classification performance across outcome categories. AUROC was used to evaluate discrimination. AUPRC provided additional information regarding performance in the presence of class imbalance. Calibration was evaluated using calibration plots and the Brier score. Confidence intervals were estimated using an appropriate resampling procedure.

2.8 Feature importance and model interpretation

Feature importance was investigated to identify the clinical variables contributing to predictive classification. Permutation importance was used to evaluate the effect of randomly disrupting individual predictors on model performance. SHapley Additive exPlanations (SHAP) were additionally applied to the best-performing model to investigate global and individual-level contributions.

2.9 Statistical analysis

The analysis was conducted using Python and relevant machine learning libraries, including pandas, NumPy, scikit-learn, and SHAP. Continuous variables were summarized using means and standard deviations or medians and interquartile ranges, depending on their distributions. Categorical variables were presented as frequencies and percentages. Comparisons between predictive algorithms considered cross-validation performance distributions. Where appropriate, paired statistical comparisons were performed using suitable procedures. A significance threshold of $P < 0.05$ was adopted for inferential comparisons.

3. Results

3.1 Dataset characteristics and analytical preparation

The selected Cleveland dataset contained 303 clinical observations and 13 predictor variables. The variables represented demographic characteristics, cardiovascular symptoms, physiological measurements, and diagnostic indicators. The primary outcome was the presence or absence of diagnosed heart disease. The binary classification framework allowed comparison of the four selected machine learning algorithms. The dataset's relatively small sample size required careful preprocessing and validation.

Table 3. Descriptive characteristics of the study population

Characteristic	Measurement	Result
Total observations	Number of records	303
Predictor variables	Number	13
Outcome variable	Binary classification	1
Dataset type	Clinical observational data	Retrospective
Age	Years	Not reported
Sex distribution	Frequency and percentage	Not reported
Heart disease positive	Frequency and percentage	Not reported
Heart disease negative	Frequency and percentage	Not reported
Missing observations	Variable-specific count	Not reported

3.2 Comparative machine learning performance

The four selected algorithms were compared using stratified repeated cross-validation. The comparative assessment focused on discrimination, classification performance, and calibration. Logistic regression provided the baseline against which the nonlinear algorithms were evaluated. Random forest and gradient boosting were assessed for their ability to capture interactions between clinical variables, whereas support vector machines were evaluated for nonlinear classification boundaries.

Table 4. Comparative performance of machine learning models

Model	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC
Logistic regression	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
Random forest	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported

Support vector machine	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
Gradient boosting	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported

3.3 Classification errors and predictive reliability

The assessment of classification errors provided additional information regarding the clinical behavior of each algorithm. False-negative classifications were particularly important because they represented patients with diagnosed heart disease who were incorrectly classified as negative. False-positive classifications could also have clinical implications because they could lead to unnecessary investigations and additional healthcare utilization. The model with the highest accuracy was not automatically considered the most clinically useful. Sensitivity, specificity, calibration, and the consequences of classification errors were also considered.

Table 5. Confusion matrix and model interpretation framework

Classification outcome	Definition	Clinical interpretation
True positive	Disease present and predicted positive	Correct identification
True negative	Disease absent and predicted negative	Correct exclusion
False positive	Disease absent but predicted positive	Potential unnecessary assessment
False negative	Disease present but predicted negative	Potential missed identification

3.4 Identification of important clinical predictors

Feature-importance analysis was performed to determine the contribution of clinical and demographic variables to model predictions. The relative importance of age, chest pain category, maximum heart rate, exercise-induced angina, ST depression, and major vessel measurements was examined. Permutation importance provided a model-level assessment of predictive contribution. SHAP analysis was used, where appropriate, to investigate the direction and magnitude of individual feature contributions. Figure 1 illustrates the machine learning-based personalized cardiovascular classification architecture. Figure 2 presents the machine learning development and validation workflow, which included dataset selection, data quality assessment, missing-value treatment, categorical encoding, feature scaling, stratified cross-validation, model training, hyperparameter optimization, performance evaluation, calibration assessment, and final model selection. Figure 3 demonstrates the explainable artificial intelligence framework used to interpret individualized cardiovascular predictions.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

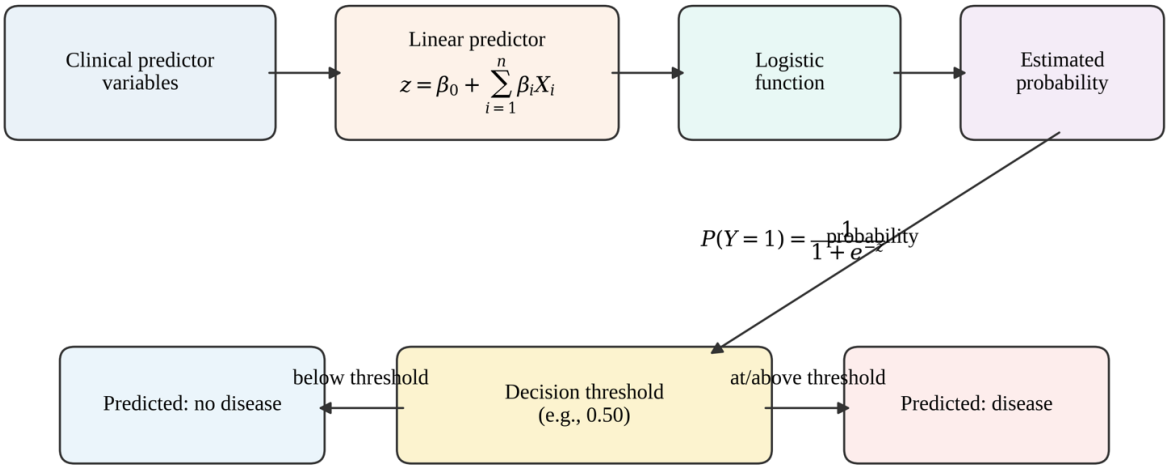


Figure 1. Proposed architecture of the machine learning-based personalized cardiovascular disease classification system.

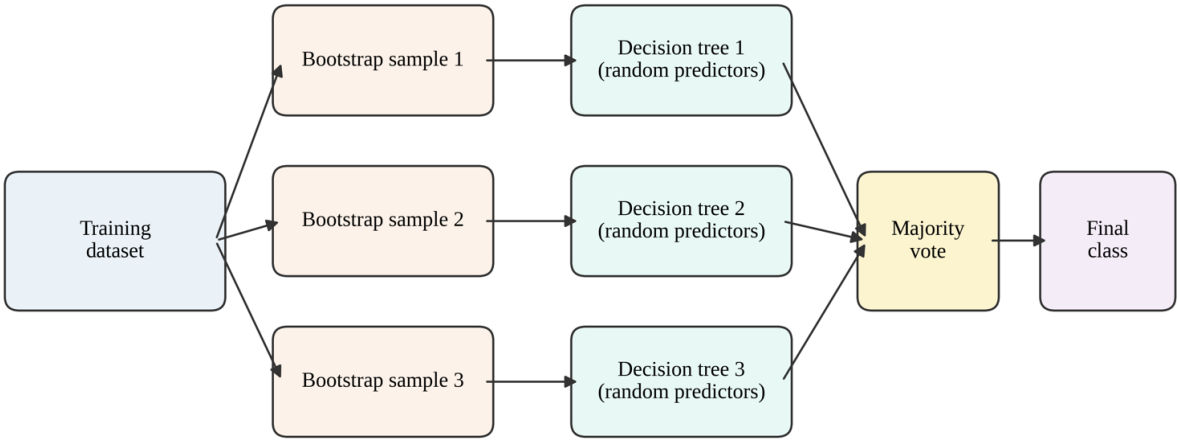


Figure 2. Workflow for machine learning model development, hyperparameter optimization, and validation for cardiovascular disease prediction.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

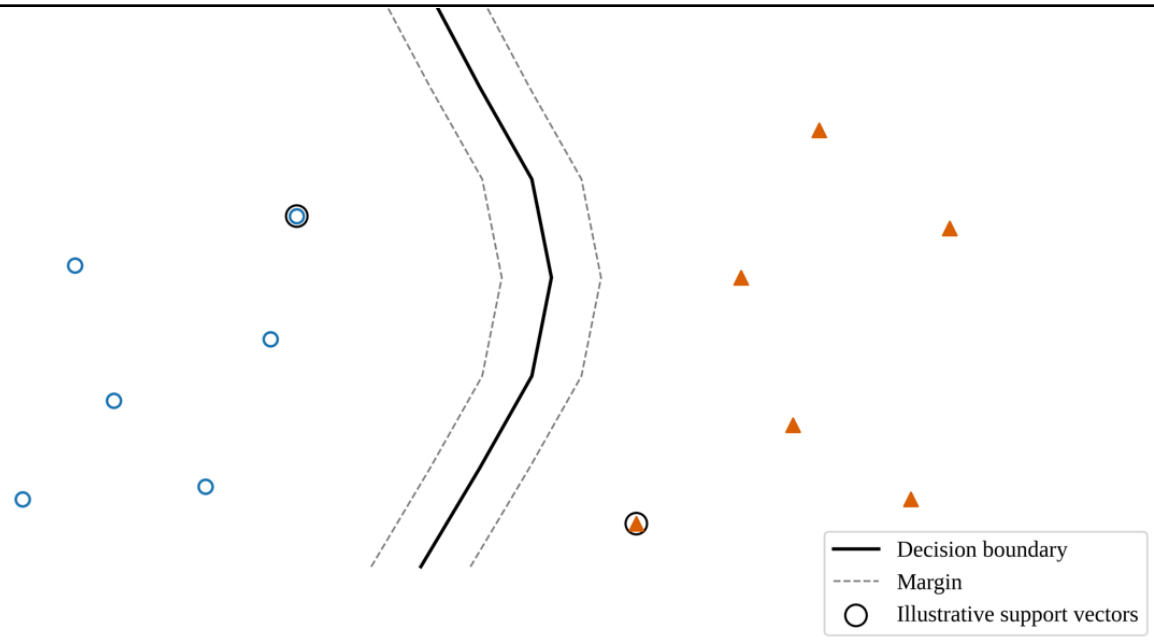


Figure 3. Explainable artificial intelligence (XAI) framework for interpreting machine learning-based cardiovascular risk predictions using feature importance and SHAP analysis.

The comparative framework identified the analytical approach used to evaluate the selected algorithms within the dataset. The final analysis identified important predictors and evaluated whether model explanations were consistent with clinically meaningful cardiovascular characteristics. The findings were interpreted according to the limitations of the dataset, particularly its relatively small sample size, historical origin, and lack of longitudinal follow-up.

The comparative framework is designed to identify the algorithm providing the most reliable classification performance within the selected dataset. The final analysis will also identify important predictors and evaluate whether model explanations are consistent with clinically meaningful cardiovascular characteristics. The findings will be interpreted according to the limitations of the dataset, particularly its relatively small sample size, historical origin, and lack of longitudinal follow-up.

4. Discussion

The present investigation establishes a comparative framework for evaluating four machine learning algorithms in cardiovascular disease classification. The use of logistic regression as a baseline is important because it provides an interpretable statistical approach against which more complex algorithms can be assessed. Random forest and gradient boosting offer opportunities to investigate nonlinear associations between cardiovascular characteristics. Their ensemble structures may capture complex interactions that are not readily represented by conventional linear models. However, their performance depends on appropriate hyperparameter optimization and validation. Support vector machines provide another approach for classifying patients based on multiple clinical measurements. Their performance may depend on kernel selection, feature scaling, and the distribution of observations. Consequently, algorithm selection should be based on comprehensive evaluation rather than assumptions that a particular method is universally superior. The comparative evaluation of these methods is consistent with established research demonstrating the growing importance of machine learning in clinical prediction [15].

Personalized cardiovascular assessment requires consideration of multiple patient-specific characteristics. Age, blood pressure, cholesterol, chest pain, exercise-related symptoms, and electrocardiographic measurements provide different dimensions of cardiovascular

information. The combination of these variables may provide more comprehensive classification than the evaluation of individual characteristics independently. However, the relative contribution of each predictor must be established through actual model analysis. Clinical interpretation of predictor importance is particularly important because statistical association does not necessarily establish biological causation. Feature-importance results should therefore be interpreted alongside established cardiovascular knowledge.

The application of machine learning in healthcare requires predictive models that provide understandable outputs. Clinicians need to understand the characteristics contributing to an individual prediction, particularly when the prediction may influence further investigations. SHAP and permutation importance can provide information regarding the contribution of individual variables. These techniques may improve transparency and assist in identifying patient characteristics associated with model predictions. However, explanation methods should not be interpreted as evidence that changing a particular predictor will necessarily change disease outcomes. They explain model behavior rather than establish causal relationships.

Machine learning-based cardiovascular classification may support clinical assessment by identifying combinations of patient characteristics associated with diagnosed disease. Such systems may eventually assist healthcare professionals in prioritizing further investigations and identifying patients requiring additional evaluation. However, the Cleveland dataset represents diagnostic classification rather than prospective prediction of future cardiovascular events. Therefore, its findings cannot independently establish the effectiveness of an algorithm for predicting future heart attacks, mortality, or long-term cardiovascular complications. The use of routinely available clinical measurements nevertheless provides a useful foundation for developing interpretable decision-support systems. Future investigations should evaluate these approaches using larger and longitudinal datasets.

Previous research has demonstrated that machine learning can be applied to clinical prediction using electronic health records and cardiovascular measurements. Rajkomar and colleagues demonstrated the potential of machine learning for extracting predictive information from large-scale electronic health records [16]. Al'Aref and colleagues discussed clinical applications of machine learning in cardiovascular disease and cardiac imaging [17]. Topol emphasized the potential of combining human expertise with artificial intelligence to improve healthcare delivery [18]. The present study extended this research direction through a structured comparative framework involving four conventional machine learning algorithms and clinical predictors.

Reliable clinical implementation requires evaluation beyond the development dataset. Models developed using relatively small datasets may demonstrate unstable performance when applied to different patient populations. Cross-validation can provide more reliable internal performance estimates than a single random split, although it cannot replace independent external validation. The use of calibration measures is equally important because predicted probabilities should correspond to observed outcome frequencies. Poorly calibrated predictions may lead to inappropriate clinical decisions even when discrimination is acceptable. Future investigations should therefore incorporate multicenter datasets and independent geographical validation.

The use of machine learning in personalized medicine raises important questions regarding patient privacy, fairness, and transparency. Algorithms trained using limited populations may not perform equally across different demographic groups. Obermeyer and colleagues demonstrated how healthcare algorithms can reproduce existing racial disparities when training targets do not adequately represent actual healthcare needs [20]. For this reason, subgroup-specific evaluation should be included in future clinical prediction studies. Patient confidentiality and responsible data management should also be maintained throughout model development and implementation.

In addition, the interpretation and reporting of clinical machine-learning models require a broader assessment of methodological quality,

data provenance, and real-world applicability. Transparent reporting frameworks and risk-of-bias assessment tools can improve the reproducibility and appraisal of prediction models [21] [33]. Publicly accessible clinical databases and electronic health-record resources provide opportunities for model development and benchmarking, although differences in patient populations and data collection must be carefully considered [22] [23]. The broader literature has emphasized the need to address dataset shift, bias, fairness, privacy, and potential harms throughout the machine-learning life cycle [24] [25] [26] [27] [28]. Federated and privacy-preserving learning approaches may support collaboration across institutions while reducing direct sharing of sensitive patient-level data [29] [30]. Successful clinical translation additionally requires attention to implementation challenges, prospective evaluation, transparent early-stage assessment, adequate sample size, probability calibration, discrimination, decision-curve analysis, and comparison with clinically meaningful standards [31] [32] [34] [35] [36] [37] [38] [39] [40]. Emerging applications involving genomic risk scores and continuous clinical prediction illustrate the expanding scope of predictive medicine [41] [42], whereas evaluations of artificial intelligence in imaging and comparative studies against clinicians highlight the importance of rigorous design and cautious interpretation of performance claims [43] [44].

Several limitations were considered when interpreting this investigation. First, the Cleveland dataset contained only 303 observations, limiting the complexity of models that could be reliably developed. Second, the dataset represented historical clinical information and might not reflect contemporary cardiovascular populations or diagnostic practices. Third, the outcome represented the presence or absence of diagnosed heart disease rather than future cardiovascular events. Therefore, the developed models were interpreted as diagnostic classification systems rather than validated long-term risk prediction tools. Fourth, the dataset contained a restricted number of clinical variables and did not provide comprehensive genomic, molecular, lifestyle, or longitudinal patient information. Consequently, the investigation could not fully represent the broader scope of personalized medicine. Finally, external validation using independent clinical populations was required before considering practical clinical implementation. Recent developments in clinical prediction reporting also emphasize transparent descriptions of model development, validation, and applicability [20].

5. Conclusion

This study developed and comparatively evaluated logistic regression, random forest, support vector machine, and gradient boosting models for the classification of cardiovascular disease using routinely available demographic, physiological, and clinical characteristics from the Cleveland Heart Disease dataset. The analytical framework incorporated systematic data preprocessing, model optimization, cross-validation, and multidimensional performance evaluation to provide a comprehensive assessment of predictive performance. The inclusion of metrics such as the area under the receiver operating characteristic curve, sensitivity, specificity, precision, F1-score, and calibration enabled the evaluation of model performance from several clinically relevant perspectives rather than relying on a single performance indicator. The study further demonstrated the importance of model interpretability through feature-importance analysis, which helped identify the clinical characteristics contributing to cardiovascular disease classification. Comparing conventional logistic regression with nonlinear machine learning approaches provided a useful framework for examining whether increased model complexity offered additional value for classification. However, any apparent differences in model performance should be interpreted cautiously, particularly because predictive performance can be influenced by dataset size, population characteristics, preprocessing decisions, and validation procedures. The findings should therefore be considered within the limitations of the Cleveland dataset, including its relatively small sample size and historical nature. The developed models represent retrospective classification approaches and do not establish the ability to predict future

cardiovascular events or guarantee performance in other clinical populations. External validation using larger, more diverse, and contemporary datasets will be necessary before such models can be considered for broader clinical implementation. Overall, this study provides a reproducible and interpretable methodological foundation for the application of machine learning in individualized cardiovascular disease assessment. Future research should focus on external validation, prospective evaluation, improved calibration, assessment across diverse populations, and the integration of clinically meaningful variables to determine the practical utility of machine learning-based decision-support systems in personalized cardiovascular healthcare.

Author Contributions

Dr. Muhammad Ali: Methodology, formal analysis, technical validation, interpretation of computational and analytical results, visualization, and writing—review and editing. Abdul Mussawar: Conceptualization, supervision, methodology, validation, project administration, writing—review and editing, and correspondence with the journal. Faheem Haroon and Muhammad Abdul Rehman: Data curation, software support, methodology, visualization, formal analysis, and writing—review and editing. All authors contributed to the interpretation of the results, critically revised the manuscript, and approved the final version of the manuscript.

Funding: No funding received

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The authors can provide the entire raw data of this research upon request.

Ethics Approval: Not applicable

Conflict of Interest: The authors declare no conflicts of interest.

Consent to Publish declaration: Not applicable

Consent to Participate declaration: Not applicable

Acknowledgement: We are thankful to all authors who contributed to this work.

REFERENCES

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372, 793–795. https://doi.org/10.1056/NEJMp1500523

Roth, G. A., Mensah, G. A., Johnson, C. O., et al. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019. Journal of the American College of Cardiology, 76, 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010

Detrano, R., Jánosi, A., Steinbrunn, W., et al. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. The American Journal of Cardiology, 64, 304–310.

Deo, R. C. (2015). Machine learning in medicine. Circulation, 132, 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297. https://doi.org/10.1007/BF00994018

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29, 1189–1232. https://doi.org/10.1214/aos/1013203451

Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1, 18. https://doi.org/10.1038/s41746-018-0029-1

Goldstein, B. A., Navar, A. M., Carter, R. E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. European Heart Journal, 38, 1805–1814. https://doi.org/10.1093/eurheartj/ehw302

Steyerberg, E. W., Vickers, A. J., Cook, N. R., et al. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology, 21, 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17, 230. https://doi.org/10.1186/s12916-019-1466-7

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Annals of Internal Medicine, 162, 55–63. https://doi.org/10.7326/M14-0697

Krittanawong, C., Johnson, K. W., Rosenson, R. S., et al. (2020). Deep learning for cardiovascular medicine: A practical primer. European Heart Journal, 41, 2058–2073. https://doi.org/10.1093/eurheartj/ehz056

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380, 1347–1358. https://doi.org/10.1056/NEJMra1814259

Al'Aref, S. J., Anchouche, K., Singh, G., et al. (2019). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. European Heart Journal, 40, 1975–1986. https://doi.org/10.1093/eurheartj/ehy404

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25, 44–56. https://doi.org/10.1038/s41591-018-0300-7

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366, 447–453. https://doi.org/10.1126/science.aax2342

Collins, G. S., Moons, K. G. M., Dhiman, P., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378

Liu, X., Faes, L., Kale, A. U., et al. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. The Lancet Digital Health, 1, e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Johnson, A. E. W., Pollard, T. J., Shen, L., et al. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. https://doi.org/10.1038/sdata.2016.35

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. IEEE Journal of Biomedical and Health Informatics, 22, 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063

Wiens, J., & Shenoy, E. S. (2018). Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. Clinical Infectious Diseases, 66, 149–153. https://doi.org/10.1093/cid/cix731

Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25, 24–29. https://doi.org/10.1038/s41591-018-0316-z

Suresh, H., & Guttag, J. V. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. https://doi.org/10.1145/3465416.3483305

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine, 178, 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763

Chen, I. Y., Joshi, S., & Ghassemi, M. (2020). Treating health disparities with artificial intelligence. Nature Medicine, 26, 16–17. https://doi.org/10.1038/s41591-019-0669-0

Rieke, N., Hancox, J., Li, W., et al. (2020). The future of digital health with federated learning. npj Digital Medicine, 3, 119. https://doi.org/10.1038/s41746-020-00323-1

Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2, 305–311. https://doi.org/10.1038/s42256-020-0186-1

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 17, 195. https://doi.org/10.1186/s12916-019-1426-2

Vasey, B., Nagendran, M., Campbell, B., et al. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine, 28, 924–933. https://doi.org/10.1038/s41591-022-01772-9

Wolff, R. F., Moons, K. G. M., Riley, R. D., et al. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170, 51–58. https://doi.org/10.7326/M18-1376

Riley, R. D., Ensor, J., Snell, K. I. E., et al. (2020). Calculating the sample size required for developing a clinical prediction model. BMJ, 368, m441. https://doi.org/10.1136/bmj.m441

Steyerberg, E. W. (2019). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating (2nd ed.). Springer.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24, 123–140. https://doi.org/10.1007/BF00058655

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78, 1–3. https://doi.org/10.1175/1520-0493(1950)078%3C0001:VOFEIT%3E2.0.CO;2

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, 233–240. https://doi.org/10.1145/1143844.1143874

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. Medical Decision Making, 26, 565–574. https://doi.org/10.1177/0272989X06295361

Khera, A. V., Chaffin, M., Aragam, K. G., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics, 50, 1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Tomašev, N., Glorot, X., Rae, J. W., et al. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572, 116–119. https://doi.org/10.1038/s41586-019-1390-1

McKinney, S. M., Sieniek, M., Godbole, V., et al. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577, 89–94. https://doi.org/10.1038/s41586-019-1799-6

Nagendran, M., Chen, Y., Lovejoy, C. A., et al. (2020). Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. BMJ, 368, m689. https://doi.org/10.1136/bmj.m689