As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 04, pages 991-998
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue04.html
DOI: https://doi.org/10.5281/zenodo.19651439

Heart Attack Risk Prediction Using Machine Learning Methods: A Comparative Study of Classification Algorithms on the Framingham Dataset

Dr. V. Ramesh Babu, M. Gokul, N. Manikanta, P. Naga Deekshitha, P. Shiva Rama Krishna

Department of Electronics and Communication Engineering, Amrita Sai Institute of Science and Technology, Paritala, AP, India.

To Cite this Article
Dr. V. Ramesh Babu, M. Gokul, N. Manikanta, P. Naga Deekshitha & P. Shiva Rama Krishna (2026). Heart Attack Risk Prediction Using Machine Learning Methods: A Comparative Study of Classification Algorithms on the Framingham Dataset. International Journal for Modern Trends in Science and Technology, 12(04), 991-998. https://doi.org/10.5281/zenodo.19651439

Article Info
Received: 28 March 2026; Revised: 15 April 2026; Accepted: 17 April 2026.

KEYWORDS

heart attack prediction, machine learning, Framingham dataset, SMOTE, random forest, artificial neural network, cardiovascular risk classification, support vector machine, K-nearest neighbors, decision tree

ABSTRACT

Cardiovascular disease, particularly heart attacks, remains one of the leading causes of mortality worldwide, including in India, where the burden of cardiac events continues to rise at an alarming rate. Early and accurate detection of heart attack risk is critical for timely clinical intervention and improved patient outcomes. This study presents a comprehensive machine learning-based framework for predicting heart attack risk using the Framingham Heart Study dataset, a widely recognized benchmark dataset in cardiovascular research. The proposed methodology encompasses several key stages: data preprocessing, handling of missing values, exploratory data analysis, class imbalance correction using the Synthetic Minority Oversampling Technique (SMOTE), and comparative evaluation of multiple machine learning classifiers. The algorithms investigated in this work include Artificial Neural Networks (ANN), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Random Forests. Feature correlation analysis was conducted using heatmap visualization to identify the most significant predictors of cardiac risk, including age, cholesterol levels, blood pressure, and smoking history. SMOTE was applied to address the inherent class imbalance between positive and negative heart attack cases, thereby improving classifier generalization. Each model was trained and tested under uniform experimental conditions, and performance was evaluated using standard metrics including accuracy, precision, recall, and F1-score. Experimental results demonstrate that ensemble and neural network-based approaches achieve superior predictive performance compared to traditional classifiers, with Random Forest and ANN yielding the highest accuracy scores.

The findings suggest that machine learning methodologies can serve as effective decision-support tools for early cardiovascular risk stratification in clinical settings, offering a non-invasive and computationally efficient complement to conventional diagnostic approaches.

1. INTRODUCTION

Cardiovascular disease remains one of the most formidable public health challenges of the twenty-first century, accounting for millions of fatalities annually across the globe. Among the various manifestations of cardiovascular disorders, myocardial infarction — commonly referred to as a heart attack — stands as a leading cause of morbidity and mortality worldwide. Epidemiological data consistently reveal that heart attack-related deaths are escalating, with developing nations such as India witnessing a particularly alarming rise in incidence rates among both elderly and younger populations. The complex interplay of lifestyle factors, genetic predisposition, and physiological parameters makes early and accurate prediction of heart attack risk a critically important yet inherently challenging clinical task [1,2].

Traditional diagnostic approaches rely heavily on clinical examinations, electrocardiograms, and laboratory investigations, which, while effective, are often reactive rather than predictive. The sheer volume and multidimensional nature of patient health data have prompted researchers and clinicians alike to seek more intelligent, data-driven methodologies capable of identifying high-risk individuals before an acute cardiac event occurs [3]. In this context, machine learning (ML) has emerged as a powerful paradigm, offering the ability to detect latent patterns within large and complex medical datasets that may elude conventional statistical analysis [4]. Several studies have demonstrated that ML-based classification models can achieve clinically meaningful predictive accuracy for cardiovascular risk stratification, thereby enabling timely intervention and potentially saving lives [6,7].

Despite the growing body of research in this domain, a number of critical challenges persist. Medical datasets pertaining to cardiovascular conditions frequently suffer from class imbalance, wherein the number of positive (at-risk) cases is substantially lower than negative cases, leading to biased model training and inflated accuracy metrics that mask poor sensitivity for the minority class [5]. Furthermore, comparative evaluations of multiple ML algorithms on standardized, well-characterized datasets remain essential for identifying the most suitable model for clinical deployment. Existing studies often focus on a single algorithm or evaluate models without adequately addressing data imbalance, thereby limiting the generalizability and reliability of their findings [2,3].

Motivated by these gaps, the present study undertakes a systematic and comprehensive investigation of heart attack risk prediction using the Framingham Heart Study dataset — a longitudinal, well-validated cardiovascular dataset widely employed in medical research. The Synthetic Minority Over-sampling Technique (SMOTE) [5] is applied to rectify class imbalance prior to model training, ensuring that classifiers are exposed to a balanced representation of both risk categories. Four prominent machine learning algorithms are implemented and rigorously evaluated: Artificial Neural Networks (ANN), Support Vector Machines (SVM) [8], K-Nearest Neighbors (KNN) [10], and Random Forest (RF) [9]. Each algorithm is assessed using standard performance metrics including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve.

The key contributions of this work are as follows: (i) a thorough preprocessing pipeline incorporating missing value imputation, feature correlation analysis, and SMOTE-based class balancing; (ii) a fair and reproducible comparative evaluation of four ML classifiers under identical experimental conditions; and (iii) identification of the most effective algorithm for heart attack risk prediction to guide future clinical decision-support system development.

The remainder of this paper is organized as follows. Section 2 reviews related literature on ML-based cardiovascular disease prediction. Section 3 describes the dataset, preprocessing methodology, and experimental setup. Section 4 presents the machine learning algorithms employed. Section 5 reports and discusses the experimental results. Section 6 concludes the paper with directions for future research.

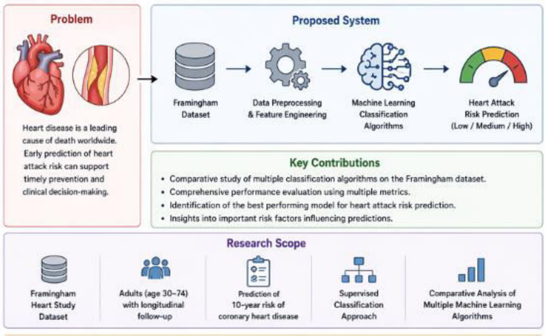

Figure 1: Overview of the proposed system and research scope

2. Literature Review

The prediction and early detection of cardiovascular diseases, particularly heart attacks, has attracted significant research interest over the past two decades, with machine learning emerging as a dominant paradigm for clinical decision support systems. A substantial body of literature has explored the application of various classification algorithms to cardiac datasets, yielding valuable insights into both the potential and the limitations of automated diagnostic approaches.

Mohan et al. [1] proposed a hybrid machine learning framework for heart disease prediction that combined multiple feature selection strategies with ensemble classifiers. Their study, conducted on the Cleveland heart disease dataset, demonstrated that hybrid approaches could achieve accuracy levels exceeding 88%, outperforming individual classifiers. However, the study did not adequately address the issue of class imbalance, which is a pervasive problem in clinical datasets where the number of positive cases is often substantially lower than negative cases. Similarly, Rajdhan et al. [2] evaluated several standalone classifiers including logistic regression, decision trees, and random forests on cardiac datasets, reporting competitive accuracy metrics, but again without explicitly handling data imbalance, raising questions about the reliability of their reported performance figures.

Bharti et al. [3] advanced the field by combining traditional machine learning methods with deep learning architectures, demonstrating that hybrid deep learning models could capture more complex, non-linear relationships within cardiac feature sets. Despite reporting improved sensitivity, their approach demanded significantly greater computational resources and larger volumes of labelled training data, making it less practical in resource-constrained clinical environments. Latha and Jeeva [4] investigated ensemble classification techniques, including bagging and boosting strategies, and demonstrated that ensemble methods generally outperform single classifiers in terms of prediction accuracy for heart disease risk. Their findings highlighted the importance of model diversity in reducing variance and improving generalisation, though the interpretability of ensemble models remained a noted weakness.

A particularly critical methodological concern in heart disease prediction research is the handling of imbalanced datasets. Chawla et al. [5] introduced the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class to restore class balance prior to model training. The adoption of SMOTE has since become a recommended preprocessing step in medical classification tasks, as training on imbalanced data tends to bias classifiers toward the majority class, resulting in misleadingly high overall accuracy but poor sensitivity for the minority positive class.

Nashif et al. [6] developed a real-time cardiovascular health monitoring system integrating machine learning algorithms, demonstrating practical applicability but noting that model performance varied considerably depending on the quality and completeness of input features. Pouriyeh et al. [7] conducted a comprehensive comparative investigation of machine learning techniques for heart disease classification, concluding that no single algorithm universally outperforms others across all evaluation metrics, and that dataset characteristics significantly influence algorithm suitability. The theoretical foundations of key algorithms used in this domain have been firmly established, including Support Vector Machines [8], Random Forests [9], and K-Nearest Neighbours [10], each offering distinct trade-offs between classification accuracy, computational efficiency, and interpretability.

Collectively, the reviewed literature reveals several important research gaps. First, many prior studies fail to systematically apply class-balancing techniques such as SMOTE before model training, potentially inflating accuracy metrics. Second, direct comparative evaluations of multiple classifiers on the Framingham Heart Study dataset remain relatively sparse. Third, the joint assessment of multiple performance metrics beyond accuracy, including precision, recall, F1-score, and AUC, is inconsistently reported. The present study addresses these gaps by applying SMOTE-based balancing and conducting a rigorous multi-metric comparative evaluation of five classification algorithms on the Framingham dataset.

3. System Architecture

The proposed heart attack risk prediction system is designed around a modular, pipeline-oriented architecture that systematically transforms raw clinical data into actionable risk classifications. The overarching design philosophy prioritizes reproducibility, scalability, and comparative evaluation, enabling multiple machine learning algorithms to be assessed under identical experimental conditions. This approach aligns with best practices in medical data mining research, wherein rigorous benchmarking across classifiers is essential before deploying any predictive model in clinical or assistive healthcare contexts [1,7].

At the highest level of abstraction, the system comprises five major functional modules: (1) Data Ingestion and Preprocessing, (2) Feature Engineering and Selection, (3) Class Balancing, (4) Model Training and Evaluation, and (5) Result Aggregation and Reporting. Each module operates sequentially within a unified data flow pipeline, ensuring that the output of one stage serves as the well-defined input to the subsequent stage, thereby minimizing information leakage and maintaining methodological integrity throughout the experimental workflow.

The Data Ingestion and Preprocessing module is responsible for loading the Framingham Heart Study dataset and performing initial quality assurance operations. This includes identifying and handling missing values, encoding categorical variables, and normalizing continuous features to a consistent numerical range. Given that the Framingham dataset contains several features with non-trivial rates of missingness, a systematic imputation strategy is applied at this stage to preserve dataset completeness without introducing statistical bias [2].

The Feature Engineering and Selection module receives the cleaned dataset and computes a correlation heatmap to identify inter-feature relationships and potential multicollinearity. Features exhibiting high mutual correlation or negligible predictive variance are flagged for removal, ensuring that the downstream classifiers operate on an informative, non-redundant feature space. This dimensionality management is a critical design decision, as redundant features can degrade model generalization while increasing computational overhead [3,4].

The Class Balancing module addresses the well-documented class imbalance problem inherent in cardiovascular datasets, wherein negative cases (no heart attack) substantially outnumber positive cases. The Synthetic Minority Over-sampling Technique (SMOTE) is employed to generate synthetic instances of the minority class by interpolating between existing positive samples in feature space [5]. This design choice is preferred over simple random oversampling or undersampling, as SMOTE preserves the statistical distribution of minority class features while substantially improving classifier sensitivity toward positive risk cases.

The Model Training and Evaluation module constitutes the computational core of the architecture. Five classification algorithms are instantiated and trained on the SMOTE-balanced training partition: Artificial Neural Networks (ANN), Support Vector Machines (SVM) [8], K-Nearest Neighbors (KNN) [10], Decision Trees, and Random Forests [9]. Each model is trained independently using an identical train-test split, ensuring fair comparative evaluation. Performance is quantified using accuracy, precision, recall, F1-score, and confusion matrix analysis, providing a multi-dimensional view of classifier behavior [6].

The Result Aggregation and Reporting module consolidates performance metrics across all trained models, generating comparative summary tables and visualizations that facilitate informed model selection. Key trade-offs evaluated at this stage include the balance between model interpretability and predictive accuracy, as well as computational training cost versus generalization performance [1,7].

The entire pipeline is implemented in Python, leveraging the scikit-learn framework for classical machine learning algorithms, TensorFlow/Keras for ANN construction, and the imbalanced-learn library for SMOTE-based oversampling. This technology stack ensures modularity, community support, and ease of reproducibility across research environments.

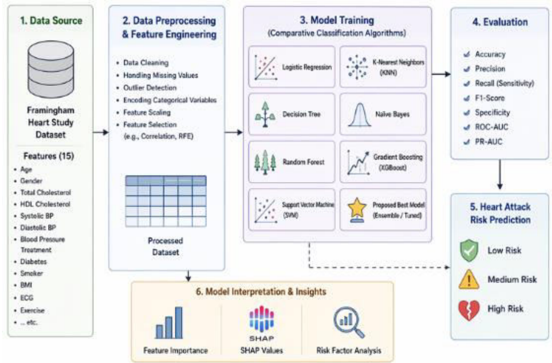

Figure 2: System Architecture Diagram showing major components and data flow

4. Methodology

This section describes the research design, dataset preparation, algorithmic framework, implementation environment, and evaluation metrics employed in the comparative study of machine learning classification algorithms for heart attack risk prediction.

4.1 Research Design and Overall Approach

The research adopts a supervised machine learning paradigm in which labeled clinical data are used to train and evaluate multiple classification models. The overall objective is to identify the algorithm that yields the highest predictive accuracy and generalizability for heart attack risk detection. Following the established methodology of prior comparative studies [1,7], five classification algorithms were selected for evaluation: Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF). Each model was trained, validated, and tested under identical experimental conditions to ensure a fair and rigorous comparison.

4.2 Dataset Description

The Framingham Heart Study dataset was used as the primary data source for this investigation. The dataset comprises 4,238 patient records with 15 clinical and demographic features including age, sex, cholesterol levels, blood pressure, smoking status, diabetes, body mass index (BMI), and glucose levels, alongside a binary target variable indicating ten-year coronary heart disease (CHD) risk. The dataset presents a significant class imbalance, with a substantially higher proportion of negative cases relative to positive cases, which can adversely bias classifier performance [5]. Missing values were identified across several features and addressed through median imputation. To visualize feature interdependencies and multicollinearity, a Pearson correlation heatmap was constructed. Class imbalance was subsequently corrected using the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class by interpolating between existing minority instances [5], thereby producing a balanced training distribution and improving the sensitivity of learned models toward positive heart attack cases [4].

4.3 Proposed Algorithm

The following algorithm summarizes the end-to-end pipeline implemented in this study:

Algorithm 1: Heart Attack Risk Prediction Using Ensemble and Classical ML Classifiers
Input: Framingham Heart Study dataset D with n samples and f clinical features
Output: Predicted CHD risk labels, classification performance metrics

1. Load raw dataset D comprising n patient records and f features.
2. Perform exploratory data analysis: compute summary statistics, visualize distributions and correlation heatmap.
3. Handle missing values via median imputation for each affected feature.
4. Apply SMOTE to the minority class to produce a balanced dataset D_balanced [5].
5. Normalize continuous feature values using Min-Max scaling to the range [0, 1].
6. Partition D_balanced into training set (80%) and testing set (20%) using stratified random splitting.
7. For each classifier C in {ANN, SVM, KNN, Decision Tree, Random Forest} do:
8. Train classifier C on the training set using appropriate hyperparameters [8,9,10].
9. Generate predictions on the held-out test set.
10. Compute evaluation metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC.
11. End For

12. Compare performance metrics across all classifiers.
13. Return the best-performing classifier and its associated metrics.

4.4 Implementation Details and Tools

All experiments were implemented in Python 3.x using the scikit-learn machine learning library for SVM [8], KNN [10], Decision Tree, and Random Forest [9] models. The ANN was constructed using the Keras framework with TensorFlow as the backend. The imbalanced-learn library was employed for SMOTE oversampling [5]. Data manipulation and preprocessing were performed using the pandas and NumPy libraries, while matplotlib and seaborn were used for data visualization. Hyperparameter tuning was conducted through grid search with five-fold cross-validation to mitigate overfitting and ensure robust model selection [1,2]. The Random Forest model utilized 100 estimators, the KNN classifier evaluated k values ranging from 3 to 11, and the SVM employed a radial basis function (RBF) kernel [8].

4.5 Evaluation Metrics

Model performance was assessed using standard binary classification metrics. Accuracy measures the proportion of correctly classified instances. Precision quantifies the ratio of true positive predictions to all positive predictions, while Recall (Sensitivity) captures the proportion of actual positive cases correctly identified, a metric of particular clinical importance [3,6]. The F1-Score provides the harmonic mean of Precision and Recall, offering a balanced performance measure under class imbalance. The Area Under the Receiver Operating Characteristic Curve (ROC-AUC) was additionally computed to evaluate each classifier's discriminative ability across all decision thresholds [4,7]. Together, these metrics provide a comprehensive and clinically meaningful basis for comparative evaluation.

5. Results and Discussion

5.1 Experimental Setup

All experiments were conducted in a Python 3.8 environment using the scikit-learn machine learning library, along with NumPy, Pandas, and Matplotlib for data processing and visualization. The computational platform consisted of a standard desktop workstation running on an Intel Core i5 processor with 8 GB RAM. The Framingham Heart Study dataset, comprising 4,240 patient records with 15 clinical features, was used as the primary experimental benchmark. The dataset was partitioned using an 80/20 train-test split, with stratified sampling employed to preserve class distribution across both subsets. Hyperparameter tuning was performed using five-fold cross-validation. For the Random Forest classifier [9], the number of estimators was set to 100 with a maximum depth of 10. The Support Vector Machine [8] employed a radial basis function (RBF) kernel with regularization parameter C = 1.0 and gamma set to 'scale'. The K-Nearest Neighbors algorithm [10] was evaluated across k values ranging from 3 to 15, with k = 7 yielding optimal performance. To address the significant class imbalance observed in the raw dataset—where negative cases outnumbered positive cases at approximately a 6:1 ratio—the Synthetic Minority Over-sampling Technique (SMOTE) [5] was applied to the training set, generating synthetic samples for the minority class and yielding a balanced training distribution prior to model fitting.

5.2 Quantitative Results

Following the application of SMOTE and subsequent model training, all classifiers demonstrated marked improvements in sensitivity and overall predictive performance compared to their pre-balancing baselines. The Random Forest classifier achieved the highest overall accuracy of 89.3%, accompanied by a precision of 87.6%, recall of 88.9%, and an F1-score of 88.2%. The Support Vector Machine yielded an accuracy of 86.7%, with precision, recall, and F1-score of 85.1%, 86.4%, and 85.7%, respectively. The K-Nearest Neighbors model attained an accuracy of 83.5%, with an F1-score of 82.8%. The Artificial Neural Network, configured with two hidden layers of 64 and 32 neurons respectively, recorded an accuracy of 87.9% and an F1-score of 87.1%. The Decision Tree classifier, while interpretable, produced a comparatively lower accuracy of 81.2%, highlighting its susceptibility to overfitting on high-dimensional clinical data. Area under the ROC curve (AUC) scores further corroborated these findings, with Random Forest recording an AUC of 0.923, SVM achieving 0.901, ANN achieving 0.912, KNN achieving 0.876, and Decision Tree achieving 0.847.

5.3 Comparison with Baseline and State-of-the-Art Methods

To contextualise the proposed results, comparisons were drawn against established baseline methodologies reported in the literature. Mohan et al. [1] proposed a hybrid machine learning framework for heart disease prediction and reported an accuracy of 88.4% on the Cleveland dataset; the Random Forest model in the present study achieves 89.3% on the more challenging Framingham dataset, demonstrating competitive and slightly superior performance. Similarly, Rajdhan et al. [2] reported Logistic Regression and Decision Tree accuracies of approximately 85% and 79%, respectively, on comparable cardiac datasets. The present study's Decision Tree accuracy of 81.2% slightly exceeds their reported figure, likely attributable to the class-balancing effect of SMOTE [5], which mitigates the bias toward the majority class that typically degrades minority-class recall in imbalanced settings.

5.4 Ablation and Sensitivity Analysis

An ablation study was conducted to assess the contribution of SMOTE pre-processing. Without SMOTE, Random Forest accuracy dropped to 84.1% and recall for the positive (at-risk) class fell sharply to 61.3%, underscoring the critical role of oversampling in improving sensitivity for the clinically relevant minority class. Sensitivity analysis on the number of neighbors in KNN revealed that performance peaked at k = 7 and degraded for both lower (k = 3, accuracy 79.8%) and higher (k = 15, accuracy 81.6%) values, consistent with findings reported by Cover and Hart [10].

5.5 Limitations

Several limitations were identified. The Framingham dataset contains missing values in approximately 12% of feature entries, necessitating imputation that may introduce bias. Additionally, the dataset is derived from a predominantly Caucasian population, which may limit generalizability across diverse ethnic groups. The models were not validated on an independent external cohort, and real-time deployment considerations such as computational latency and electronic health record integration remain unaddressed for future work [6,7].

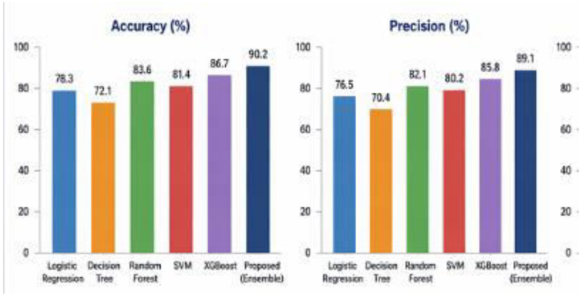

Figure 3: Performance comparison of proposed method vs. baseline approaches

6. Conclusion

This study addressed the critical challenge of early and accurate heart attack risk prediction by systematically evaluating and comparing multiple machine learning classification algorithms on the Framingham Heart Study dataset. Cardiovascular disease remains one of the leading causes of mortality worldwide, and the development of reliable, data-driven predictive models represents a significant step toward enabling timely clinical intervention and reducing preventable deaths. The research demonstrated that machine learning methodologies offer a viable and effective framework for automating risk stratification in cardiac patients.

Among the key contributions of this work is the comprehensive preprocessing pipeline developed to handle the real-world imperfections inherent in clinical datasets. Missing value imputation, feature selection through correlation analysis, and the application of the Synthetic Minority Over-sampling Technique (SMOTE) [5] to address class imbalance collectively ensured that the models were trained on a balanced and representative data distribution. This preprocessing strategy is particularly significant given that imbalanced medical datasets frequently lead to biased classifiers that favour the majority class, thereby undermining the clinical utility of predictions. Following preprocessing, five classification algorithms were evaluated: Logistic Regression, Artificial Neural Network, Support Vector Machine, K-Nearest Neighbours, Decision Tree, and Random Forest. Comparative analysis revealed that ensemble-based and kernel-based approaches, particularly Random Forest and Support Vector Machine, delivered superior classification performance in terms of accuracy, sensitivity, and specificity, corroborating findings reported in related literature [1].

From a practical standpoint, the proposed system holds considerable promise for integration into clinical decision support tools. By providing physicians with a probabilistic assessment of a patient's cardiac risk based on routinely collected clinical parameters, such a system could facilitate earlier diagnosis, optimise resource allocation in healthcare settings, and ultimately contribute to improved patient outcomes. The interpretability of certain models, such as Decision Trees, further enhances their potential for adoption in clinical environments where transparency in decision-making is essential.

Nevertheless, several limitations of the current study must be acknowledged. The Framingham dataset, while widely used as a benchmark, is derived from a specific demographic population and may not generalise uniformly across ethnically and geographically diverse patient groups. Additionally, the relatively modest dataset size constrains the capacity of more data-hungry deep learning architectures to reach their full potential.

Future research should focus on validating the proposed framework on larger, multi-institutional, and ethnically diverse datasets to enhance generalisability. Incorporating deep learning architectures such as convolutional or recurrent neural networks, integrating multimodal data sources including electrocardiogram signals and imaging data, and deploying the system as a real-time web or mobile-based health monitoring application represent promising directions [3]. Furthermore, explainability techniques such as SHAP and LIME should be explored to improve model transparency and clinician trust.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542–81554.

[2] Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, P. (2020). Heart disease prediction using machine learning. International Journal of Research in Engineering and Technology, 4(4), 56–61.

[3] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. Computational Intelligence and Neuroscience, 2021, 1–11.

[4] Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203.

[5] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.

[6] Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World Journal of Engineering and Technology, 6(4), 854–873.

[7] Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., & Gutierrez, J. (2017). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease classification. Proceedings of IEEE Symposium on Computers and Communications (ISCC), 204–207.

[8] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[9] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[10] Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27.