VER
Vascular & Endovascular Review

Explainable Machine Learning Framework for Early Heart Disease Detection Using SMOTE and SHAP

Mahfuz Islam Khan Jabed[1], Muhammad Imran[2], Abdul Ali Khan[3], Mohiuddin Mehedi[4], Ashraful Islam[5], Rahat Pervez[6]

[1] Department of Information Technology, Washington University of Science & Technology, Virginia, USA
Email: mahfuzislamkhanjabed@gmail.com
ORCID iD: https://orcid.org/0009-0001-2141-2894
[2] Department of Information Technology, Washington University of Science & Technology, Virginia, USA
Email: imranm.student@wust.edu
ORCID iD: https://orcid.org/0009-0001-2069-0868
[3] Department of Biology, New Mexico Highlands University, Las Vegas, New Mexico, USA
Email: abdulalik5111@gmail.com
ORCID iD: https://orcid.org/0009-0002-0118-3497
[4] Department of Information Technology, Washington University of Science & Technology, Virginia, USA
Email: momehedi.student@wust.edu
ORCID iD: https://orcid.org/0009-0007-4228-6253
[5] Department of Information Technology, Washington University of Science & Technology, Virginia, USA
Email: ashralam.student@wust.edu
ORCID iD: https://orcid.org/0009-0001-8067-7331
[6] Department of Cyber Security, Bay Atlantic University, Washington D.C, USA
Email: Rahat335@gmail.com
ORCID iD: https://orcid.org/0009-0000-8194-1520

Corresponding Author: Mahfuz Islam Khan Jabed
Email: mahfuzislamkhanjabed@gmail.com

ABSTRACT

Cardiovascular diseases (CVDs) remain the leading cause of death globally, creating a persistent need for early screening tools that are both accurate and clinically interpretable.1 This study presents an end-to-end, explainable machine learning framework for binary heart disease prediction using a structured clinical dataset derived from the UCI Heart Disease benchmark, where published experiments commonly use a 14-variable subset and focus on distinguishing presence versus absence of disease.2 To strengthen reliability and reduce bias, the pipeline integrates stratified train–test splitting, feature scaling for scale-sensitive learners, and Synthetic Minority Over-sampling Technique (SMOTE) to address potential class imbalance in the training split.3 Multiple models are compared, including Logistic Regression, Naïve Bayes, KNN, SVM, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Extra Trees, XGBoost, and MLP. Performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC, with 5-fold cross-validation to estimate generalization stability. On the held-out test set, Extra Trees achieved the highest ROC-AUC (90.80%), while SVM obtained the highest accuracy (83.61%). Cross-validation ranked Random Forest (mean ROC-AUC ≈ 90.06%) and AdaBoost (≈ 89.95%) as top performers, and GridSearchCV further optimized Extra Trees to a best cross-validated ROC-AUC of 0.912. Finally, explainability is provided through SHAP, which attributes predictions to clinically meaningful features, supporting transparent decision support rather than black-box output.4

KEYWORDS: Heart Disease Prediction, Machine Learning, Explainable AI, SHAP, SMOTE, Classification Models, Healthcare Analytics, Cross-Validation, Hyperparameter Tuning, Clinical Decision Support.

How to Cite: Mahfuz Islam Khan Jabed, Muhammad Imran, Abdul Ali Khan, Mohiuddin Mehedi, Ashraful Islam, Rahat Pervez , (2026) Explainable Machine Learning Framework for Early Heart Disease Detection Using SMOTE and SHAP, Vascular and Endovascular Review, Vol.9, No.1, 316-324

INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for an estimated 19.8 million deaths in 2022 approximately 32% of all global deaths underscoring the urgent public-health imperative for early identification of high-risk individuals [1]. In the United States alone, heart disease is responsible for approximately one in every five deaths and imposes an annual economic burden exceeding $239 billion in combined healthcare expenditures and lost productivity [5]. Beyond the United States, the global prevalence of major cardiovascular risk factors including hypertension, diabetes mellitus, dyslipidemia, obesity, and physical inactivity continues to rise in both high-income and low-to-middle-income countries, amplifying the demand for scalable, cost-effective early screening tools capable of functioning in resource-constrained clinical environments [10].

Traditional cardiovascular risk scoring systems such as the Framingham Risk Score, SCORE2, and the ACC/AHA Pooled Cohort Equations have long served as clinical decision aids; however, they rely on a fixed set of variables and linear statistical assumptions that may not adequately capture the complex, nonlinear interactions among risk factors present in heterogeneous patient populations [11]. Machine learning (ML) methods have emerged as a compelling alternative, offering the capacity to learn high-dimensional, nonlinear decision boundaries from structured clinical data without requiring explicit parametric assumptions. Studies have consistently demonstrated that ensemble methods such as Random Forests, Gradient Boosting, and XGBoost as well as support vector machines and deep neural architectures can outperform traditional risk scores on benchmark cardiovascular datasets when properly validated [8].

Despite this promise, the widespread clinical adoption of ML-based cardiac risk models remains limited. A primary barrier is interpretability: clinicians must understand why a model predicts elevated risk in order to meaningfully integrate that output into diagnostics, triage decisions, and patient communication. Without interpretability, even high-accuracy models risk clinician distrust, regulatory non-compliance, and patient harm if their reasoning is opaque or misaligned with established medical knowledge. This interpretability barrier is repeatedly identified as a central obstacle to translating AI-based heart disease prediction into real-world clinical practice [6]. A second challenge is class imbalance: clinical datasets frequently contain fewer confirmed disease-positive cases than negative ones, causing standard classifiers to be biased toward the majority class and to systematically underdetect the very patients most in need of intervention [3].

Explainable AI (XAI) has emerged as a critical paradigm to address the interpretability limitation by producing human-readable evidence alongside model predictions. Among XAI techniques, SHAP (SHapley Additive explanations) is grounded in cooperative game theory's Shapley values and provides both global feature importance rankings and patient-level attribution scores that quantify each feature's marginal contribution to a specific prediction [4]. LIME (Local Interpretable Model-agnostic Explanations) complements SHAP by fitting locally faithful surrogate models around individual predictions, enabling clinicians to interrogate whether the model's reasoning for a specific patient aligns with established clinical knowledge [7]. Guleria et al. demonstrated that integrating XAI frameworks into cardiovascular classification pipelines significantly improves clinician acceptance and model trustworthiness compared to black-box approaches [12]. In high-stakes clinical domains such as cardiology, this transparency is not merely desirable but essential: predictions must be defendable to patients, auditable by regulatory bodies, and actionable within existing clinical workflows. Emerging regulatory frameworks including the EU AI Act and FDA guidance on Software as a Medical Device (SaMD) increasingly mandate that AI systems deployed in healthcare settings provide explainable outputs, cementing XAI as a foundational requirement rather than an optional enhancement.

Against this backdrop, this paper contributes an explainable, reproducible, and evaluation-rigorous ML pipeline for early heart disease detection. Specifically, the pipeline: (i) benchmarks eleven diverse ML models spanning linear, probabilistic, kernel-based, neural, and ensemble learners on a standardized UCI Heart Disease dataset to provide a comprehensive comparative baseline; (ii) applies SMOTE exclusively within the training split for class imbalance mitigation, preventing data leakage into evaluation and preserving the integrity of held-out test results; (iii) validates predictive stability through stratified 5-fold cross-validation with ROC-AUC scoring to capture fold-level variance in addition to point estimates; (iv) optimizes the highest-performing ensemble model using GridSearchCV-based hyperparameter tuning; and (v) generates SHAP-based global and instance-level explanations that align predictive drivers with established clinical risk factors, supporting transparent and auditable decision support. Together, these contributions address a reproducibility and interpretability gap that persists in the cardiovascular ML literature, particularly for benchmark-scale studies where methodological rigor is often sacrificed for scale or complexity.

RELATED WORK

The intersection of machine learning and cardiovascular medicine has experienced rapid growth over the past decade, driven by expanding availability of electronic health records, improvements in computational infrastructure, and growing recognition that traditional risk stratification tools have ceiling-level performance on complex patient populations. Recent scholarship increasingly emphasizes that predictive accuracy alone is insufficient for real-world cardiovascular deployment; transparency, robustness, and clinical usability are equally necessary conditions for adoption [8]. Comprehensive reviews document progress across multiple data modalities structured EHR data, cardiac imaging (CT angiography, MRI, echocardiography), wearable biosignals (ECG patches, photoplethysmography), and privacy-preserving federated learning architectures designed for multi-institutional data sharing while consistently noting persistent challenges: lack of prospective clinical validation, dataset heterogeneity, and the absence of standardized evaluation protocols that would permit meaningful cross-study comparisons [8]. Jakkani and Singh further highlight that while deep learning models achieve impressive performance on imaging tasks, structured clinical data pipelines incorporating classical and ensemble ML methods remain the most practically deployable in settings where imaging data is unavailable or cost-prohibitive [8].

Several studies have specifically demonstrated the value of combining strong classification performance with post-hoc explainability tools. Bilal et al. integrated machine learning with SHAP and LIME on a large-scale Kaggle cardiovascular dataset containing 308,737 patient records, reporting 91.94% accuracy with an 8.06% miss rate, and illustrated how XAI

explanations can foster clinician trust when operating at population scale [9]. However, the large feature space and proprietary data characteristics of that study limit its direct reproducibility for researchers using standard benchmark datasets. Salah and Srinivas similarly applied an explainable ML framework to long-term CVD risk prediction in adolescents, finding that SHAP-identified predictors were highly consistent with established pediatric cardiology risk factors, reinforcing the clinical face validity of SHAP-driven interpretation across diverse demographic subgroups [10]. Guleria et al. proposed an XAI framework for cardiovascular disease classification using ensemble techniques on structured clinical data, demonstrating that combining SHAP with well-tuned ensemble classifiers consistently outperformed single-model baselines while producing clinically coherent feature attributions [12]. Abdulsalam et al. explored an ensemble-quantum ML approach for heart disease prediction with explainability, achieving competitive classification results while highlighting that model diversity within ensemble strategies is critical for robust minority-class detection [13].

Class imbalance handling has been identified as a central methodological concern in cardiovascular prediction pipelines. Talukder et al. proposed XAI-HD, a hybrid ML/deep learning framework with extensive preprocessing and multiple balancing strategies including SMOTE variants, reporting error-rate reductions of 20–25% across multiple datasets validated by Wilcoxon signed-rank statistical testing [6]. Their work establishes two core themes that motivate the present study: first, that imbalance handling materially affects both fairness and sensitivity toward disease-positive minority cases; and second, that explainability is not a secondary concern but an integral requirement for clinically acceptable AI. Chawla et al.'s foundational SMOTE algorithm which synthesizes minority-class examples through linear interpolation between k-nearest neighbors remains the most widely adopted oversampling technique for structured clinical data, having been empirically validated across dozens of medical classification tasks [3].

The UCI Heart Disease benchmark dataset has served as a canonical testbed for cardiovascular ML research since its introduction, and numerous studies have leveraged it to evaluate and compare classification algorithms under controlled conditions [2]. Laftah and Al-Saedi applied explainable ensemble learning on UCI-derived heart disease data, demonstrating that stacking and voting classifiers augmented with SHAP explanations achieve a superior interpretability–performance trade-off compared to single models, and that feature attributions consistently identify chest pain type, maximum heart rate, and number of major vessels as dominant predictors across ensemble configurations [14]. Patro and Padhy extended this line of work to a remote health monitoring context, combining ML and deep learning within an XAI framework and reporting that SHAP-guided feature selection improved model generalization while reducing computational over head a finding with direct implications for deployment in telehealth and primary care screening settings [11]. Adalarasu et al. broadened the XAI scope to multiple cardiovascular disease subtypes, showing that SHAP explanations produced consistent, clinically plausible feature rankings across different disease categories, reinforcing the generalizability of SHAP as an explanation mechanism beyond binary classification [15].

Despite these advances, a practical gap persists for researchers and practitioners in resource-constrained or early-stage clinical settings: a lightweight, fully reproducible, model-comparative framework that uses a well-established benchmark dataset, validates generalization stability through cross-validation, addresses class imbalance rigorously through training-only SMOTE application, and provides clear SHAP-based interpretability all without requiring large-scale EHR infrastructure, proprietary datasets, or specialized deep learning hardware. Many existing studies either sacrifice reproducibility for scale, or achieve interpretability at the expense of systematic multi-model benchmarking. The present study addresses this gap by unifying rigorous evaluation protocols, ensemble optimization via GridSearchCV, and SHAP explanations within a single coherent pipeline applied to the UCI Heart Disease benchmark, providing a transparent and replicable baseline that future clinical-scale studies can directly build upon.

METHODOLOGY

3.1 Dataset and Problem Formulation

The task is formulated as binary classification predicting heart disease presence (1) versus absence (0). The dataset is derived from the UCI Heart Disease benchmark, which contains 76 raw attributes, while most published experiments use a standardized subset of 14 variables and reduce the outcome to presence vs. absence.[2] The implementation uses 13 predictor features (e.g., age, sex, chest pain type, resting BP, cholesterol, fasting blood sugar, resting ECG, maximum heart rate, exercise-induced angina, ST depression/oldpeak, slope, number of major vessels, thalassemia status) and one binary target variable.

Recommended table placement: Include a “Dataset Variables” table here (Table 1) listing each feature, type (continuous/categorical encoded), and clinical meaning. This improves reviewer perception of clinical grounding.

3.2 Data Preparation and Train–Test Split

Data preprocessing includes consistency checks, missing-value inspection, and preparation of features/target. Because accurate estimation requires unbiased evaluation, the dataset is partitioned using a stratified 80/20 train–test split to preserve class proportions across splits.

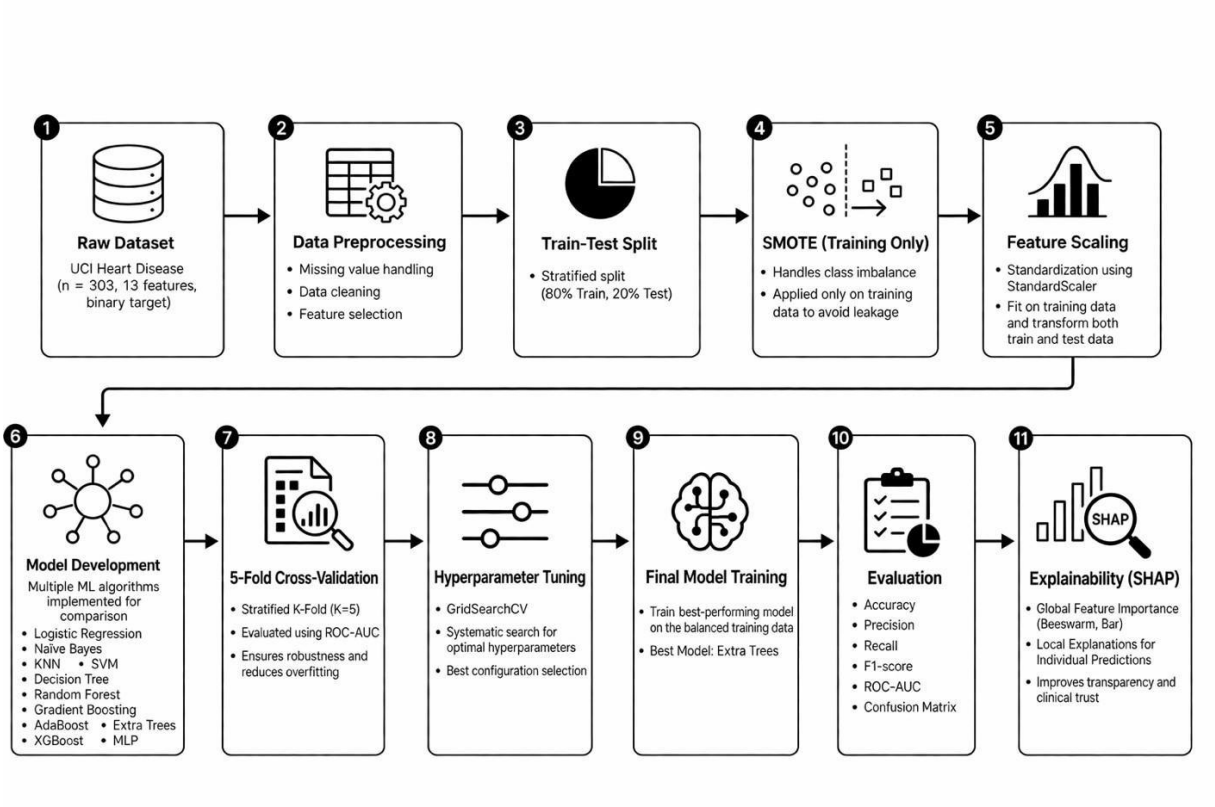

Figure 1. End-to-end machine learning pipeline for heart disease prediction, incorporating SMOTE-based class balancing, stratified cross-validation, hyperparameter tuning, and SHAP-based explainability.

3.3 Class Imbalance Handling with SMOTE

Clinical data often exhibits class imbalance, which can reduce sensitivity to disease-positive cases. This pipeline incorporates SMOTE applied **only to the training data** to mitigate imbalance while avoiding leakage into evaluation. SMOTE synthesizes minority-class examples through interpolation between minority neighbors, improving classifier sensitivity under imbalance.

Implementation note for publication rigor: In a journal submission, it is best practice to apply SMOTE *inside* the cross-validation loop (e.g., using an imbalanced-learn pipeline) to prevent optimistic bias during CV. (Your hold-out test split remains untouched.)

3.4 Feature Scaling Strategy

Feature scaling is performed using standardization (zero mean, unit variance) for models sensitive to feature magnitudes (Logistic Regression, SVM, KNN, MLP). Tree-based ensembles (Decision Tree, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, XGBoost) are trained on unscaled features because splitting criteria are invariant to monotonic transformations in typical implementations.

3.5 Model Development and Evaluation Protocol

A diverse model set is used to compare linear, probabilistic, distance-based, neural, and ensemble learners:

- Logistic Regression, Naïve Bayes, KNN, SVM, MLP
- Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Extra Trees, XGBoost

Performance metrics include accuracy, precision, recall (sensitivity), F1-score, and ROC-AUC. ROC-AUC is emphasized because it evaluates ranking quality across thresholds important for clinical screening contexts where decision thresholds may vary.

3.6 Cross-Validation and Hyperparameter Optimization

To estimate stability and generalization, 5-fold cross-validation is performed using ROC-AUC scoring. Hyperparameter tuning is conducted for Extra Trees using GridSearchCV, optimizing parameters such as number of estimators, maximum depth, and split constraints.

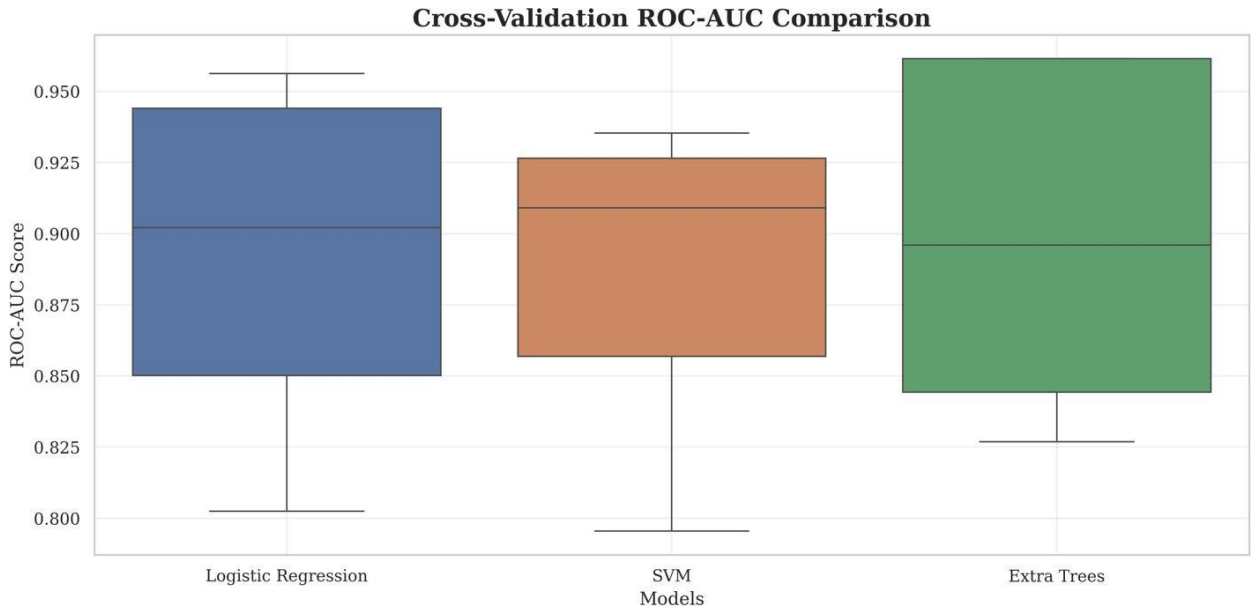

Figure 1. Cross-validation ROC-AUC distribution for selected models using 5-fold stratified validation, illustrating model stability and performance variability.

The cross-validation results demonstrate consistent performance across folds for all models, with the Extra Trees classifier showing the highest median ROC-AUC and relatively stable variance. This indicates strong generalization capability and robustness compared to other models.

RESULTS AND DISCUSSION

4.1 Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	ROC-AUC (%)
Extra Trees	**78.69**	**75.00**	**90.91**	**82.19**	**90.80**
Naive Bayes	**81.87**	**78.95**	**90.91**	84.51	88.53
Random Forest	80.33	75.61	93.94	83.78	88.26
SVM	**83.61**	79.49	93.94	86.11	88.20
Gradient Boosting	81.97	80.56	87.88	84.06	86.90
KNN	80.33	76.92	90.91	83.33	86.74
Logistic Regression	78.69	76.32	87.88	81.69	86.47
AdaBoost	**78.69**	**76.32**	**87.88**	**81.69**	**86.04**
XGBoost	**78.69**	**75.00**	**90.91**	**82.19**	**84.31**
MLP	**73.77**	**72.97**	**81.82**	**77.14**	**82.47**
Decision Tree	**72.13**	**71.05**	**81.82**	**76.06**	**71.27**

Table 2. Hold-out test performance of evaluated models.

The performance of all evaluated models on the hold-out test set is summarized in Table 2. Among the models, the Extra Trees classifier achieved the highest ROC-AUC score of 90.80%, indicating superior discriminative ability. It also demonstrated strong recall performance (90.91%), which is critical in medical diagnosis to minimize false negatives. Ensemble-based methods, including Random Forest and Gradient Boosting, consistently outperformed individual models, highlighting their effectiveness in capturing complex feature interactions. In contrast, simpler models such as Decision Tree and MLP exhibited comparatively lower performance, suggesting limitations in handling the underlying data distribution.

4.2 ROC & Confusion Matrix

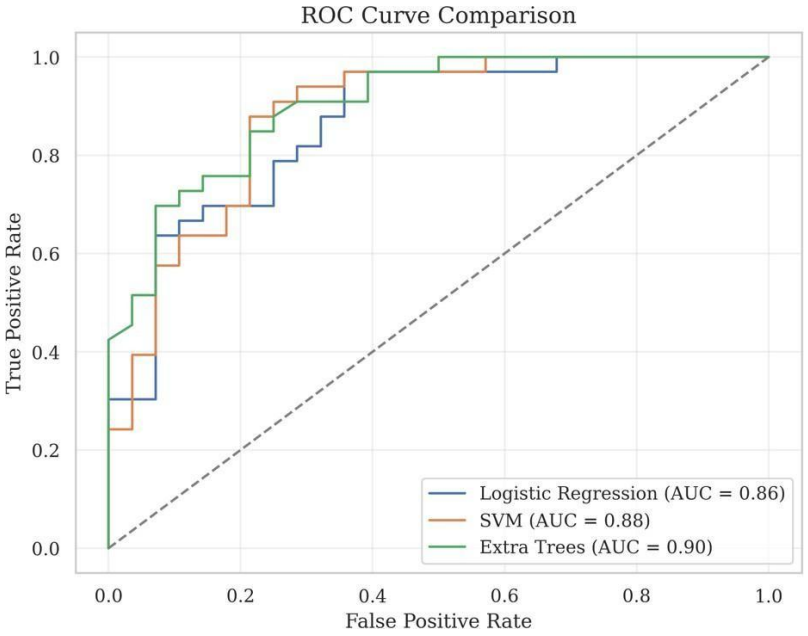

Figure 2 illustrates the ROC curve comparison for selected models. The Extra Trees model demonstrates the highest area under the curve, confirming its superior classification performance. The ROC analysis further validates the robustness of ensemble methods in distinguishing between positive and negative cases across varying thresholds.

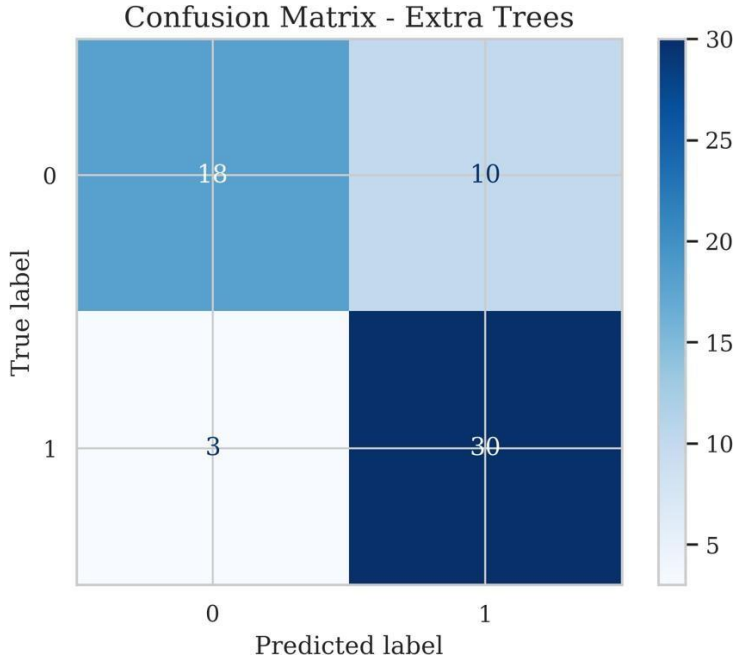

Figure 3. Confusion matrix of the best-performing model (Extra Trees), illustrating the distribution of true positives, true negatives, false positives, and false negatives.

The confusion matrix of the Extra Trees model is shown in Figure 3. The model correctly classified a high number of positive cases, achieving strong recall, which is critical in clinical screening to reduce missed diagnoses. Although a small number of false positives were observed, the model maintains a favorable balance between sensitivity and precision. This performance demonstrates the model's suitability for early-stage heart disease detection, where minimizing false negatives is of primary importance.

4.3 Explainability (SHAP)

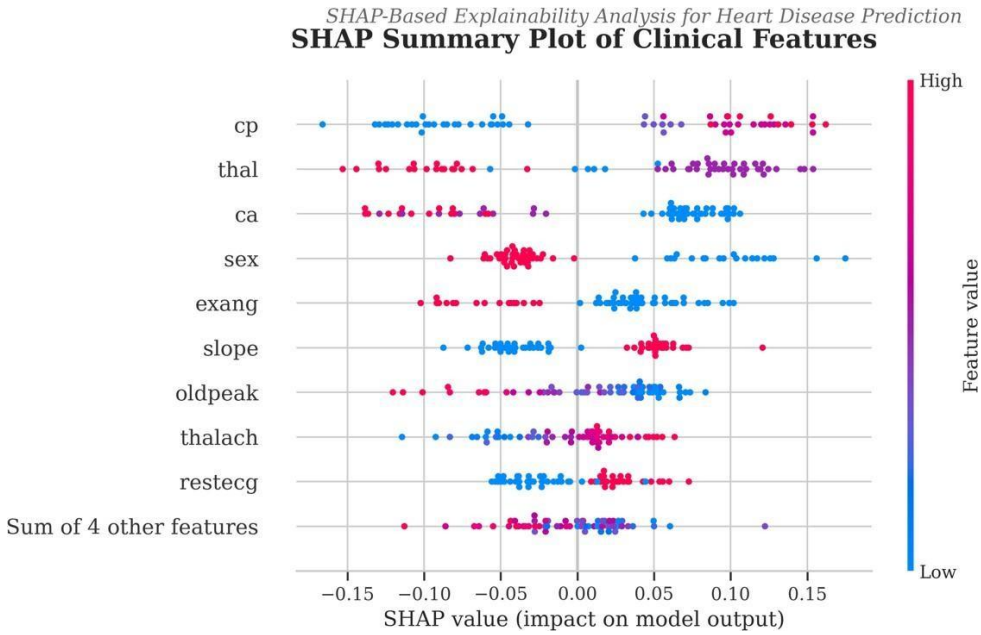

Figure 4. SHAP summary plot illustrating the global impact of clinical features on heart disease prediction. Each point represents an individual instance, where color encodes feature value and position reflects its contribution to the model output.

The SHAP summary plot in Figure 4 provides a global interpretation of feature contributions across all samples. Features such as chest pain type (cp), thalassemia (thal), and number of major vessels (ca) show the highest influence on model predictions, indicating their strong clinical relevance. Positive SHAP values push the prediction toward the presence of heart disease, while negative values indicate lower risk. The distribution of SHAP values further reveals that higher values of certain features are consistently associated with increased disease likelihood. This aligns with established clinical knowledge, reinforcing the reliability of the model. The consistency of these patterns demonstrates that the model is not only accurate but also interpretable, which is essential for deployment in healthcare decision-support systems.

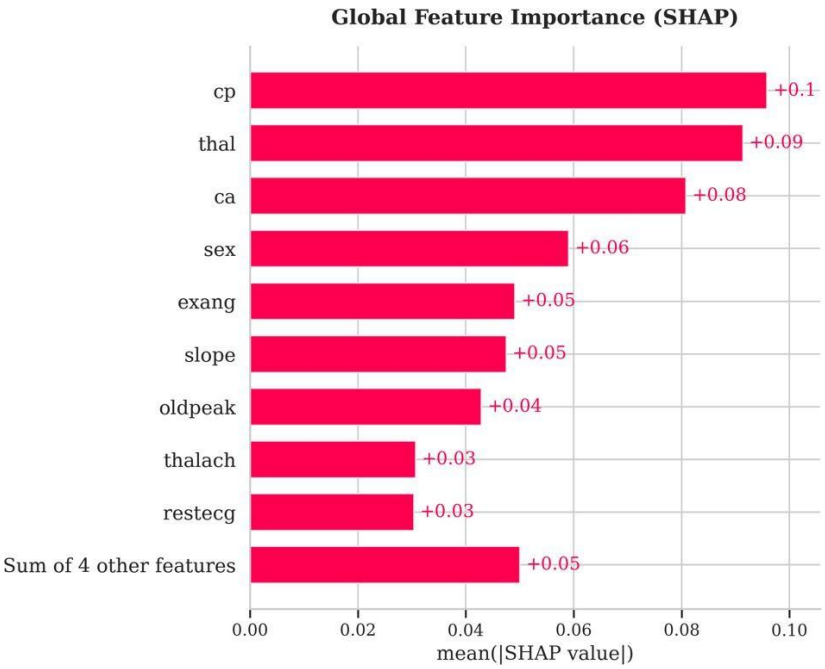

Figure 5. Global feature importance based on mean absolute SHAP values, highlighting the most influential predictors in the model.

To enhance model transparency and clinical interpretability, SHAP (Shapley Additive Explanations) was applied to the best-performing model (Extra Trees). SHAP enables both global and local interpretation by quantifying the contribution of each feature to the model's predictions.

Figure 4 presents the SHAP summary plot, which illustrates how individual feature values influence the prediction output across all samples. Features such as chest pain type (cp), thalassemia (thal), and number of major vessels (ca) demonstrate strong impact, indicating their clinical relevance in heart disease detection.

To further summarize feature importance, Figure 5 shows the mean absolute SHAP values for each feature. The results confirm that cp, thal, and ca are among the most influential predictors, aligning with established clinical risk factors.

4.4 Comparison to Prior Studies

This work differs from large-scale EMR-style studies primarily in scope, but it aligns strongly in methodological intent: transparent, clinically interpretable prediction. Bilal et al. achieve 91.94% accuracy on a much larger dataset (308,737 records) and emphasize XAI as a mechanism for trust and adoption.[9] The present study does not claim direct performance superiority due to major differences in dataset size, feature definitions, and evaluation setup; instead, it demonstrates that even on a benchmark dataset, robust validation (cross-validation + tuning) and SHAP explanations can produce a reproducible, interpretable system suitable as a baseline for further clinical-scale validation.

LIMITATIONS AND FUTURE WORK

The primary limitation is dataset scale and external validity. Benchmark datasets such as the UCI Heart Disease subset are valuable for controlled model comparison, but they cannot substitute for diverse, multi-institutional validation required for deployment.[6] Additionally, while SMOTE is implemented as a training-only balancing mechanism, the strongest methodological practice is to embed resampling into the cross-validation pipeline to avoid biased CV estimates, particularly on small datasets.[3]

Future work should prioritize: (i) evaluating the tuned model on large-scale, real-world clinical datasets similar to those used in population-scale studies;[9] (ii) incorporating multimodal inputs (wearables ECG, imaging) as highlighted in recent cardiovascular AI reviews; (iii) adding statistical significance testing across models (e.g., Wilcoxon signed-rank comparisons) as used in advanced XAI-HD-style frameworks; and (iv) assessing subgroup performance and calibration to support equitable and clinically safe decision support.

CONCLUSION

This paper presented a publication-oriented, explainable machine learning pipeline for early heart disease detection. Using a well-established UCI-style structured dataset, the study compared a broad set of ML models, validated stability through 5-fold cross-validation, optimized an ensemble model via GridSearchCV, and incorporated SHAP explanations to make predictions transparent and clinically interpretable. The results show that ensemble learners provide strong discrimination performance (Extra Trees ROC-AUC 90.80% on the hold-out test set), while cross-validation highlights strong average performance for Random Forest and AdaBoost. Beyond performance, SHAP-based interpretability strengthens trust and usability consistent with current cardiovascular AI literature emphasizing transparency as a requirement for clinical adoption.[9]

REFERENCES

1. World Health Organization, “Cardiovascular diseases (CVDs),” fact sheet, Jul. 31, 2025.
2. A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, *Heart Disease* [Dataset]. UCI Machine Learning Repository, 1989, doi: 10.24432/C52P4X.
3. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
4. S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 4765–4774, 2017.
5. Centers for Disease Control and Prevention, “Heart Disease Facts,” National Center for Chronic Disease Prevention and Health Promotion, Oct. 24, 2024. [Online]. Available: https://www.cdc.gov/heartdisease/facts.htm
6. Talukder, M.A., Talaat, A.S., Kazi, M. et al. XAI-HD: an explainable artificial intelligence framework for heart disease detection. Artif Intell Rev 58, 385 (2025). https://doi.org/10.1007/s10462-025-11385-6
7. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
8. Jakkani, A. K., & Singh, A. P. (2025). Advancements in Artificial Intelligence for Cardiovascular Disease Detection: A Detailed Review of Techniques and Applications. INTERNATIONAL JOURNAL, 9(2).
9. Bilal, A., Alzahrani, A., Almohammadi, K., Saleem, M., Farooq, M. S., & Sarwar, R. (2025). Explainable AI-driven intelligent system for precision forecasting in cardiovascular disease. Frontiers in medicine, 12, 1596335. https://doi.org/10.3389/fmed.2025.1596335
10. Salah, H., Srinivas, S. Explainable machine learning framework for predicting long-term cardiovascular disease risk among adolescents. Sci Rep 12, 21905 (2022). https://doi.org/10.1038/s41598-022-25933-5

11. Patro, S. P., & Padhy, N. (2023). A Secure Remote Health Monitoring for Heart Disease Prediction Using Machine Learning and Deep Learning Techniques in Explainable Artificial Intelligence Framework. *Engineering Proceedings*, *58*(1), 78. https://doi.org/10.3390/ecsa-10-16237
12. Guleria, P., Naga Srinivasu, P., Ahmed, S., Almusallam, N., & Alarfaj, F. K. (2022). XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques. *Electronics*, *11*(24), 4086. https://doi.org/10.3390/electronics11244086
13. Abdulsalam, G., Meshoul, S., Shaiba, H. (2023). Explainable Heart Disease Prediction Using Ensemble-Quantum Machine Learning Approach. Intelligent Automation & Soft Computing, 36(1), 761–779. https://doi.org/10.32604/iasc.2023.032262
14. Laftah, R. H., & Al-Saedi, K. H. K. (2024). Explainable Ensemble Learning Models for Early Detection of Heart Disease. Journal of Robotics and Control (JRC), 5(5), 1412–1421. https://doi.org/10.18196/jrc.v5i5.22448
15. Adalarasu, K., Raghavan, B., Madhavan, B., Venkatesh, S., & Amirtharajan, R. (2026). An explainable machine learning (XAI) framework to enhance types of cardiovascular disease diagnosis and prognosis. Physical and engineering sciences in medicine, 49(1), 115–129. https://doi.org/10.1007/s13246-025-01653-8