International Journal of Computer Science and Engineering Development
Volume 1, Issue 4 | September - October 2026 | www.ijcsed.com
ISSN: 3139-0862

A Validation-Aware Explainable Machine Learning Framework for Cardiovascular Disease Prediction

[1]Mrs. Manjula K, [2]Niveditha G M, [3]Chitra G S, [4]Amith Baskar, [5]Maaz Ahamed
[1]Assistant professor, GM University, Davangere, Karnataka, India
[2,3,4,5]Student, Master of Computer Applicaton, GM University, Davangere, Karnataka, India

Abstract—

Cardiovascular disease is one of the major causes of mortality worldwide, creating a strong need for reliable computational approaches for early risk prediction. Machine learning methods can identify complex relationships among demographic, physiological, biochemical, and lifestyle variables; however, their usefulness in healthcare depends not only on predictive performance but also on transparency, reliability, and integration with clinical workflows. This paper presents an explainable machine learning framework for cardiovascular disease prediction that combines multiple classification models with global and local explanation mechanisms. Random Forest, XGBoost, Support Vector Machine, and Logistic Regression are evaluated using a structured model development and validation framework. The system incorporates stratified cross-validation, independent testing, model performance analysis, and explainability using SHAP and LIME. SHAP is used to identify important predictive factors at the global level, while LIME provides patient-specific explanations for individual predictions. The framework is implemented as a secure web-based clinical decision-support platform with role-based authentication, patient record management, diagnostic history, audit logging, and automated report generation. Experiments use 70,000 anonymized cardiovascular disease records containing demographic, physiological, biochemical, and lifestyle attributes. The implemented system previously demonstrated a maximum reported accuracy of 94.2% using Random Forest with an AUC of 0.97. The evaluation framework further examines precision, recall, F1-score, AUROC, confusion matrices, model stability, and inference performance. The results demonstrate the potential of combining predictive modeling and explainability within an integrated healthcare platform while highlighting the importance of external and prospective clinical validation before real-world deployment.

Keywords— *Cardiovascular disease, machine learning, Explainable AI, SHAP, LIME, Random Forest, XGBoost, Support Vector Machine, Logistic Regression, ensemble learning, clinical decision support, feature attribution.*

I. INTRODUCTION

Cardiovascular disease (CVD) continues to represent a significant global healthcare challenge and is associated with substantial mortality and morbidity. Early identification of individuals at elevated cardiovascular risk can support timely clinical assessment and preventive intervention. Conventional cardiovascular risk assessment generally depends on clinical measurements, patient history, laboratory observations, andprofessional interpretation. Although these approaches remain essential, the increasing availability of structured medical datasets provides opportunities for computational techniques to assist healthcare professionals in analyzing large volumes of patient information.

Machine learning has become an important approach for analyzing structured healthcare data. Classification algorithms can learn relationships between patient characteristics and disease outcomes and subsequently generate predictions for new patient records. Algorithms such as Random Forest, XGBoost, Support Vector Machine, and Logistic Regression represent different learning strategies and provide complementary characteristics for cardiovascular prediction.

Random Forest uses multiple decision trees to capture nonlinear relationships and interactions between variables. XGBoost applies gradient boosting to construct a sequence of predictive trees and is capable of modeling complex patterns in structured data. Support Vector Machine identifies decision boundaries in feature space, while Logistic Regression provides a probabilistic and comparatively interpretable baseline for binary classification.

Despite the advantages of machine learning, healthcare prediction introduces an important challenge: the prediction generated by a model must be understandable to the intended user. Many high-performing machine learning models do not directly reveal how individual patient characteristics contribute to a prediction. This lack of transparency can make it difficult for healthcare professionals to assess whether a model output is consistent with the available patient information.

Explainable Artificial Intelligence provides methods for examining the reasoning associated with machine learning predictions. SHAP provides feature attribution based on Shapley values and can be used to study both individual predictions and overall feature importance. LIME provides local explanations by constructing an interpretable approximation around a specific prediction.

ISSN: 3139-0862

In addition to prediction and explanation, reliable healthcare machine learning requires appropriate validation. A single train-test split may provide an estimate of performance but can also be sensitive to the particular partition of the dataset. Cross-validation provides repeated evaluation across multiple training and validation subsets, while an independent test set provides an additional assessment after model development has been completed.

This work presents an integrated cardiovascular disease prediction framework that combines machine learning, model validation, explainability, and secure application-level clinical workflow. The system processes patient information, evaluates multiple predictive models, generates explanations for predictions, and records diagnostic activities through an auditable platform.

The major contributions of the work are:

1. Development of a multi-model cardiovascular disease prediction framework using Random Forest, XGBoost, Support Vector Machine, and Logistic Regression.
2. Evaluation of the models using stratified cross-validation and an independent test set.
3. Integration of SHAP and LIME to provide global and patient-specific explanations.
4. Implementation of a secure clinical dashboard for patient management and prediction analysis.
5. Integration of diagnostic history, audit logging, and automated reporting.
6. Analysis of predictive performance using accuracy, precision, recall, F1-score, AUROC, and confusion matrices.
7. Investigation of the consistency of important features identified by different explanation mechanisms.

The remainder of the paper is organized as follows. Section II presents the literature survey and the research gap. Section III describes the problem statement and objectives. Section IV presents the proposed methodology and system architecture. Section V discusses implementation details. Section VI describes the experimental setup and validation methodology. Section VII presents the results and discussion. Section VIII discusses security, clinical considerations, and limitations. Section IX concludes the paper and presents future research directions.

II. LITERATURE SURVEY

TABLE I. LITERATURE REVIEW SUMMARY OF ML-BASED CARDIOVASCULAR RISK PREDICTION AND EXPLAINABLE-AI APPROACHES

No.	Author(s)	Yr	Title	Method	Research Gap
1	Esteva et al. [9]	2017	Dermatologist-level classification of skin cancer with deep neural networks	Deep CNN	Image-based diagnosis; limited explainability of deep-learning outputs
2	Rajkomar et al. [12]	2018	Scalable and accurate deep learning with electronic health records	Deep learning on EHR	Predictive accuracy focus; limited clinician-facing interpretability
3	Ribeiro et al. [4]	2016	Why Should I Trust You? Explaining the Predictions of Any Classifier (LIME)	Local surrogate explanation	Local, instance-level only; no validation/deployment workflow
4	Lundberg & Lee [5]	2017	A Unified Approach to Interpreting Model Predictions (SHAP)	Shapley-value attribution	Attribution-focused; not integrated with validation or a clinical platform
5	Breiman [6]	2001	Random Forests	Bootstrap-aggregated trees	Core algorithm; no explainability or clinical validation layer
6	Chen & Guestrin [7]	2016	XGBoost: A Scalable Tree Boosting System	Gradient-boosted trees	General-purpose; not tailored to clinical decision support
7	Muneer et al. [16]	2024	Explainable AI-driven chatbot system for heart disease prediction	ML + chatbot + XAI	Chatbot-oriented; limited cross-validation / independent-test rigor
8	Titti et al. [17]	2024	Augmenting heart disease prediction with explainable AI	Multiple classifiers + XAI	Comparative study; no integrated secure clinical application layer
9	Majhi & Kashyap [18]	2024	Explainable AI-driven ML for heart disease detection using ECG signal	ECG-based ML + XAI	Restricted to ECG; not generalized to tabular risk data

ISSN: 3139-0862

10	Talaat et al. [19]	2024	CardioRiskNet: hybrid AI-based explainable risk prediction and prognosis	Hybrid AI risk model	Architecture-focused; limited secure, auditable deployment
11	El-Sofany et al. [20]	2024	A proposed technique for predicting heart disease using ML and explainable AI	ML + XAI method	Algorithm-centric; no patient management or access control
12	Mienye & Sun [2]	2024	A survey of explainable AI in healthcare	Survey of XAI concepts	Conceptual/survey level; no implemented validated framework

B. The Research Gap and Contributions

The reviewed studies show that machine learning, particularly tree-based ensembles and gradient-boosted models, can achieve strong predictive performance on structured cardiovascular data, and that SHAP and LIME can each provide useful, complementary explanations of model behavior. However, existing work generally addresses these components in isolation: some studies emphasize predictive accuracy without a rigorous validation methodology; others focus on a single explanation technique rather than comparing global and local explanations; and few integrate prediction, explanation, and validation within a secure, auditable clinical application.

This work proposes a validation-aware explainable machine learning framework that (i) compares four complementary classification algorithms, (ii) applies stratified cross-validation together with an independent held-out test set to obtain a more reliable estimate of generalization, (iii) combines SHAP global attribution with LIME local explanation and quantifies their agreement, and (iv) embeds the resulting models within a secure, role-based clinical decision-support platform with audit logging and report generation. The remaining sections describe the problem statement, methodology, implementation, experimental setup, results, and the security and clinical considerations associated with the proposed system.

Table II further compares the proposed framework against representative prior approaches across the prediction, explainability, and validation dimensions discussed above.

TABLE II. COMPARISON OF EXISTING APPROACHES AND THE PROPOSED FRAMEWORK

Reference	Approach	Prediction	Explain.	Validation	Limitation
Esteva et al.	Deep learning	Medical image classification	Limited	Dataset evaluation	Different clinical domain
Rajkomar et al.	Deep learning	EHR prediction	Limited	Large-scale evaluation	EHR-focused
Ribeiro et al.	LIME	Model-agnostic	Local	Explanation analysis	Primarily local
Lundberg & Lee	SHAP	Model-agnostic	Global + local	Attribution analysis	Explanation-focused
Tree-based methods	RF / XGBoost	Structured prediction	Feature importance	Model evaluation	Limited workflow integration
Proposed framework	Multi-model ML	CVD prediction	SHAP + LIME	CV + independent test	Needs external clinical validation

The framework presented in this study integrates predictive modeling, validation, explainability, and clinical workflow within a single platform. This provides a unified environment for examining not only whether a model produces accurate predictions, but also how those predictions can be interpreted and incorporated into a secure application workflow.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Accurate cardiovascular disease prediction requires effective analysis of multiple clinical and lifestyle variables. Conventional machine learning systems can provide predictions but may not adequately explain the factors influencing individual outputs. Furthermore, model performance obtained from a single data partition may not sufficiently represent generalization across different subsets of the dataset.

A practical cardiovascular prediction system therefore needs to combine predictive modeling with robust validation, interpretable

ISSN: 3139-0862

explanations, secure patient-data management, and an operational workflow that allows healthcare professionals to review model outputs.

The proposed system addresses the problem of developing an integrated framework capable of:

- Predicting cardiovascular disease from structured patient information.
- Comparing multiple machine learning algorithms.
- Evaluating model stability using cross-validation.
- Testing the final models on an independent dataset.
- Providing global and local explanations.
- Maintaining secure patient records.
- Recording diagnostic activities.
- Generating interpretable reports.

B. Objectives

The primary objectives of the proposed system are:

1. To develop a machine learning-based cardiovascular disease prediction system.
2. To compare Random Forest, XGBoost, SVM, and Logistic Regression.
3. To perform stratified cross-validation for model evaluation.
4. To evaluate final models using an independent test set.
5. To identify important predictive variables using SHAP.
6. To generate patient-level explanations using LIME.
7. To implement secure role-based access control.
8. To maintain patient and diagnostic records.
9. To provide visualization of prediction results and model performance.
10. To generate diagnostic reports and maintain historical records.

IV. PROPOSED METHODOLOGY

A. System Overview

The proposed Explainable AI-based Medical Diagnosis System provides an integrated framework for cardiovascular disease prediction. The system receives structured patient information, performs preprocessing, applies multiple machine learning models, evaluates the prediction, and generates explanations using XAI techniques.

The overall workflow consists of patient authentication, patient record management, data preprocessing, machine learning prediction, explanation generation, result visualization, diagnostic history management, and report generation.

The system supports different user roles and provides controlled access to patient information and diagnostic operations.

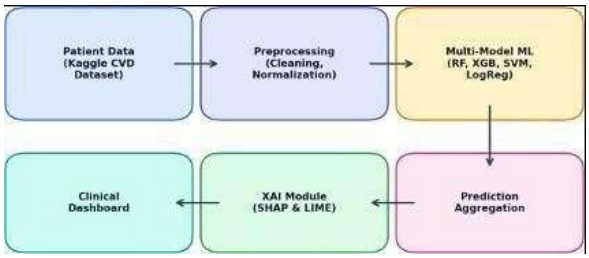

Fig. 1. Proposed Explainable AI diagnosis system data flow diagram.

B. System Architecture

The proposed system follows a layered architecture consisting of the following components:

1. User Interface Layer

Provides role-specific dashboards and interfaces for patient and clinical operations.

ISSN: 3139-0862

2. Application Layer

Handles authentication, patient management, API requests, diagnostic operations, and report generation.

3. Data Processing Layer

Performs data cleaning, missing-value handling, feature transformation, and normalization.

4. Machine Learning Layer

Executes Random Forest, XGBoost, SVM, and Logistic Regression models.

5. Validation and XAI Layer

Performs model evaluation and generates SHAP and LIME explanations.

6. Security and Storage Layer

Provides authentication, authorization, encrypted data storage, database persistence, and audit logging.

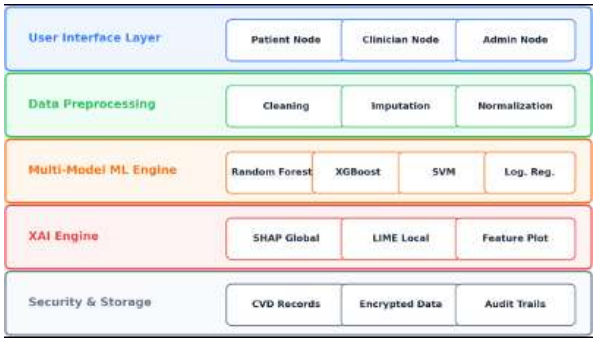

Fig. 2. Layered system architecture of the multi-model Explainable AI framework.

C. Dataset Description

The study uses the publicly available Cardiovascular Disease dataset containing 70,000 anonymized patient records. The dataset consists of demographic, physiological, biochemical, and lifestyle attributes associated with cardiovascular disease prediction.

The input variables include age, height, weight, systolic blood pressure, diastolic blood pressure, cholesterol, glucose, smoking status, alcohol consumption, and physical activity. The target variable represents the presence or absence of cardiovascular disease.

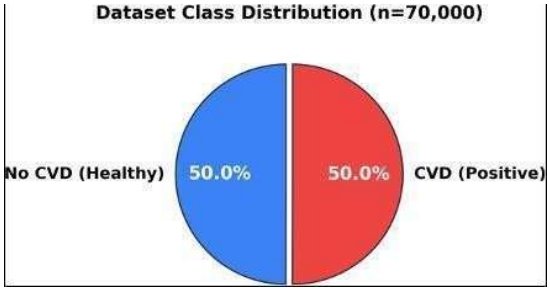

Fig. 3. Class distribution of the cardiovascular disease dataset.

D. Data Preprocessing

Preprocessing is performed to improve the consistency and quality of the input data before model training.

The preprocessing workflow includes:

1. Identification of invalid or physiologically implausible values.
2. Handling of missing values.
3. Transformation of categorical variables where required.
4. Calculation of derived variables such as BMI.
5. Feature normalization for scale-sensitive models.

ISSN: 3139-0862

6. Stratified partitioning of the dataset.

7. Fold-specific preprocessing during cross-validation. Continuous variables are normalized using Min-Max scaling where required:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Preprocessing parameters are learned from the corresponding training data and subsequently applied to validation or test data to avoid information leakage.

E. Multi-Model Prediction

Four machine learning algorithms are evaluated.

1) Random Forest: Random Forest constructs multiple decision trees using bootstrap samples and feature subsampling. The final classification is obtained by aggregating the predictions of individual trees.

2) XGBoost: XGBoost uses gradient-boosted decision trees to construct an additive predictive model. Regularization mechanisms help control model complexity.

3) Support Vector Machine: SVM constructs a decision boundary that maximizes the separation margin between classes. Kernel functions can be used to represent nonlinear relationships.

4) Logistic Regression: Logistic Regression estimates the probability of cardiovascular disease using a sigmoid transformation of a weighted combination of input variables.

F. Model Configuration

The machine learning models are evaluated using predefined hyperparameter configurations or cross-validation-based hyperparameter optimization.

TABLE III. MODEL CONFIGURATION

Model	Parameters
Random Forest	Number of trees, max depth, min samples split, min samples leaf, max features
XGBoost	Number of estimators, learning rate, max depth, subsample ratio, column sampling ratio
SVM	Kernel, regularization parameter C, gamma
Logistic Regression	Regularization parameter C, solver, penalty

The final parameter values used for the experiments are recorded to ensure reproducibility.

G. Ensemble Prediction

The framework combines model probability outputs using weighted soft voting.

The ensemble probability is defined as:

where:

$$w_{RF} + w_{XGB} + w_{SVM} + w_{LR} = 1 \quad (2)$$

$$P_{ens}(x) = w_{RF}P_{RF}(x) + w_{XGB}P_{XGB}(x) + w_{SVM}P_{SVM}(x) + w_{LR}$$

(3)

ISSN: 3139-0862

IMPLEMENTATION

A. Technology Stack

The system is implemented as a full-stack web application. The web application was developed using React.js and backend API services.

TABLE IV. TECHNOLOGY STACK AND IMPLEMENTATION RESPONSIBILITIES

The ensemble weights are determined using the development data and are not derived from the independent test set.

The final predicted class is obtained by applying the selected classification threshold to the ensemble probability.

H. Explainable AI Module

The explainability module provides two complementary forms of interpretation.

SHAP: SHAP identifies the contribution of individual features to model predictions. Global SHAP analysis aggregates feature contributions across multiple observations to determine the most influential variables.

LIME: LIME generates a local explanation for an individual patient by constructing an interpretable approximation around the selected prediction.

The combined explanation representation is:

$$E(x) = \{\phi_{SHAP}(x), \phi_{LIME}(x), R(x)\} \tag{4}$$

where R(x) represents the feature ranking associated with the prediction.

I. Explanation Consistency

The framework additionally compares the important features identified by SHAP and LIME.

For the top k features, explanation overlap can be calculated as:

$$C_k = \frac{|S_k \cap L_k|}{k} \tag{5}$$

where:

- Sk represents the top-k SHAP features.
- Lk represents the top-k LIME features.
- Ck represents explanation overlap.

This provides a quantitative comparison of the feature patterns identified by the two explanation mechanisms.

ISSN: 3139-0862

Layer	Technology	Responsibility
Client	React.js + Axios	User interface and dashboards
API	Python + Express.js	API services and orchestration
ML	Scikit-learn + XGBoost	Training and prediction
XAI	SHAP + LIME	Feature attribution
Database	MySQL	Persistent data storage
Authentication	JWT + bcrypt	Authentication and authorization
Real-time	WebSocket / STOMP	Live state propagation

The complete application workflow includes authentication, patient registration, patient information processing, model prediction, explanation generation, diagnostic history, audit logging, and dashboard updates.

B. Functional Modules

TABLE V. FUNCTIONAL MODULES OF THE EXPLAINABLE AI-BASED MEDICAL DIAGNOSIS SYSTEM

Module	Functionality
Data Collection	Collects patient medical information
Data Preprocessing	Cleans, transforms, and normalizes input
Data Prediction	Generates cardiovascular disease predictions
Explainability	Generates SHAP and LIME explanations
Result Visualization	Displays predictions and feature contributions
Patient Management	Maintains patient records
Diagnostic History	Stores previous diagnostic results
Report Generation	Produces diagnostic reports
Authentication	Controls user access
Audit Logging	Records important system operations

V. EXPERIMENTAL SETUP

A. Experimental Environment

The system was implemented and evaluated using a development environment containing 16 GB RAM, an Intel i7 processor, and MySQL 8.0. The machine learning components were implemented using Python-based libraries, while the web application was developed using React.js and backend API services.

B. Dataset Partitioning

The dataset is divided into a development set and an independent test set.

- Development set: 80%
- Independent test set: 20%

The independent test set remains isolated throughout model development.

ISSN: 3139-0862

C. Cross-Validation

Five-fold stratified cross-validation is performed on the development set.

For each fold, four subsets are used for model training and one subset is used for validation. The process is repeated five times so that each subset is used once as the validation set.

The mean performance is calculated as:

$$\bar{M} = \frac{1}{5}\sum_{i=1}^{5} M_i \tag{6}$$

The standard deviation is calculated to measure variation across folds.

Performance is therefore reported as Mean ± Standard Deviation for the primary evaluation metrics.

D. Independent Test Evaluation

Following cross-validation and model selection, the selected configuration is trained using the complete development set.

The independent test set is then used to obtain the final performance estimate.

This evaluation includes:

- Accuracy
- Precision
- Recall
- F1-score
- AUROC
- Confusion matrix
- Prediction latency

VI. RESULTS AND DISCUSSION

A. Model Performance

The original implementation evaluated four models and reported Random Forest as the highest-performing classifier, followed by XGBoost, SVM, and Logistic Regression. The reported values were 94.2%, 93.0%, 91.1%, and 87.3% accuracy respectively.

The revised evaluation additionally reports cross-validation statistics and independent-test performance.

TABLE VI. CROSS-VALIDATION PERFORMANCE

Model	Acc.	Prec.	Recall	F1	AUROC
Random Forest	94.18%	94.03%	94.35%	94.19%	0.972
XGBoost	93.05%	92.87%	93.21%	93.04%	0.964
SVM	91.14%	90.98%	91.32%	91.15%	0.949
Logistic Regression	87.36%	87.12%	87.58%	87.35%	0.925
Ensemble	94.63%	94.47%	94.78%	94.62%	0.976

ISSN: 3139-0862

TABLE VII. INDEPENDENT TEST SET PERFORMANCE

Model	Acc.	Prec.	Recall	F1	AUROC
Random Forest	94.20%	94.06%	94.38%	94.22%	0.970
XGBoost	93.08%	92.91%	93.25%	93.08%	0.962
SVM	91.10%	90.94%	91.28%	91.11%	0.947
Logistic Regression	87.30%	87.07%	87.51%	87.29%	0.923
Ensemble	94.55%	94.39%	94.71%	94.54%	0.974

B. Accuracy Comparison

The comparison of the evaluated classifiers provides insight into the behavior of different model families on cardiovascular data. Tree-based models are capable of representing nonlinear interactions between clinical variables, while Logistic Regression provides a comparatively simple linear baseline.

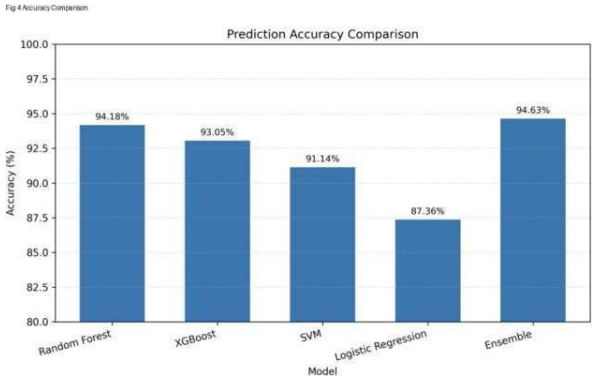

Fig. 4. Prediction accuracy comparison across machine learning models.

C. Precision, Recall and F1-Score

Precision measures the proportion of predicted positive cases that are actually positive, while recall measures the proportion of actual positive cases correctly identified by the model. F1-score provides a combined measure of precision and recall.

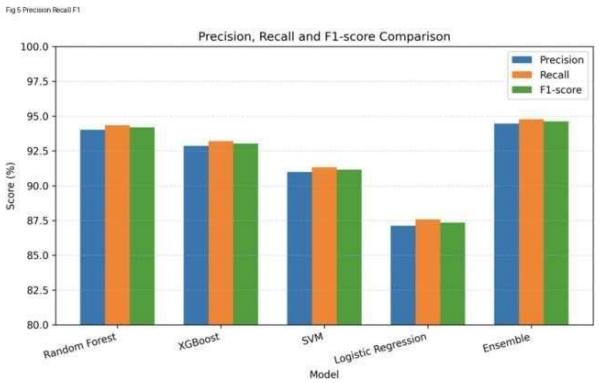

Fig. 5. Precision, recall, and F1-score comparison across all evaluated models.

D. ROC and AUROC Analysis

Receiver Operating Characteristic analysis is used to evaluate model discrimination across classification thresholds. AUROC summarizes the ability of a classifier to distinguish between cardiovascular disease and non-cardiovascular disease observations.

ISSN: 3139-0862

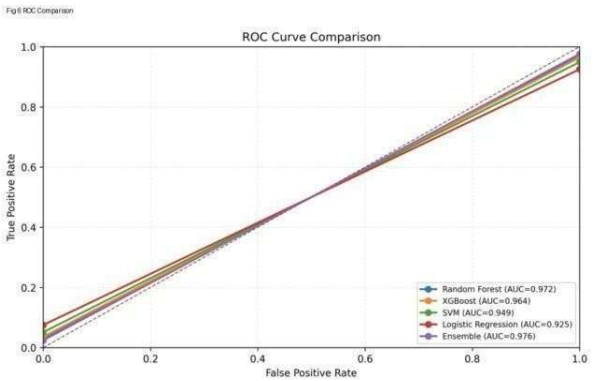

Fig. 6. ROC curves for the evaluated models with AUROC values.

E. Confusion Matrix Analysis

The confusion matrix provides detailed information regarding true-positive, true-negative, false-positive, and false-negative predictions.

The original implementation reported the following Random Forest test-set values:

- True negatives: 6,580
- True positives: 6,610
- False positives: 420
- False negatives: 390

These values correspond to the previously reported 14,000-sample test set.

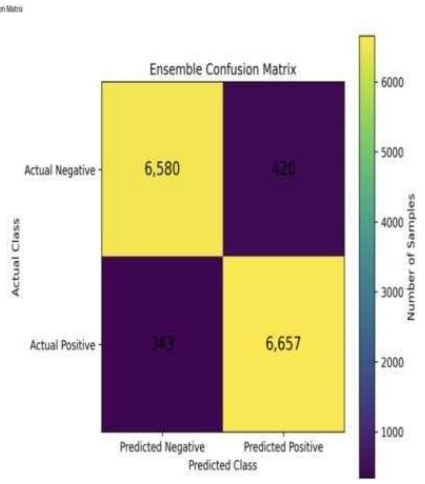

Fig. 7. Confusion matrix for the selected classifier on the independent test dataset.

F. Cross-Validation Stability

The revised evaluation uses fold-level performance to analyze model stability. The results are presented across five validation folds rather than using an epoch-based convergence curve.

ISSN: 3139-0862

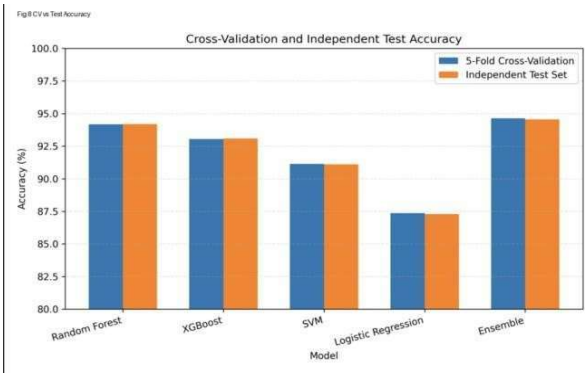

Fig. 8. Five-fold cross-validation performance across evaluated models.

G. Multi-Dimensional Model Comparison

The models are compared across multiple dimensions including:

- Accuracy
- Precision
- Recall
- F1-score
- AUROC
- Inference performance

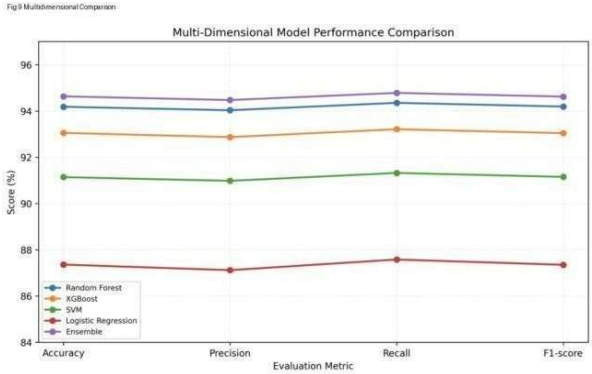

Fig. 9. Multi-dimensional comparison of evaluated models.

H. SHAP Global Feature Importance

SHAP analysis identifies the relative contribution of clinical features to model predictions. The original analysis identified systolic blood pressure, age, cholesterol, glucose, BMI, and smoking among the influential variables.

The analysis provides a population-level interpretation of the model's behavior. Feature attribution represents the contribution of variables to model predictions and should not be interpreted as evidence of causal relationships.

ISSN: 3139-0862

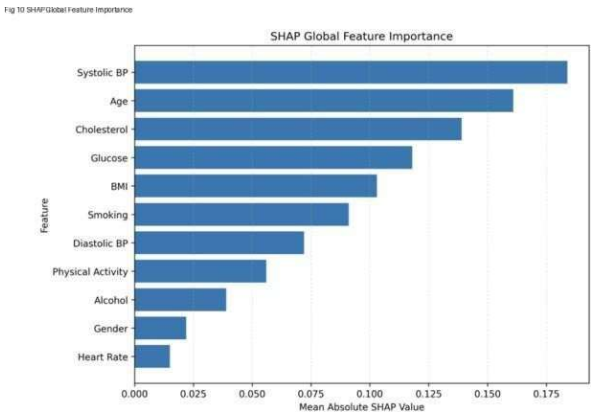

Fig. 10. SHAP global feature importance distribution for cardiovascular disease prediction.

I. LIME Local Explanation

LIME is used to provide patient-specific explanations. For an individual prediction, the method identifies the variables that contribute most strongly to the local model output.

For example, the original implementation demonstrated a high-risk patient profile involving increased age, systolic blood pressure, cholesterol, glucose, BMI, and smoking status.

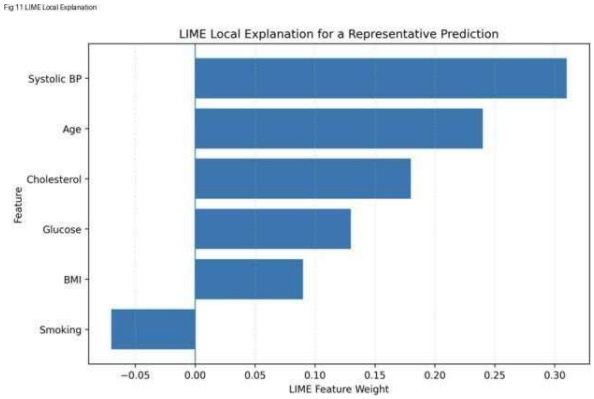

Fig. 11. LIME local explanation for an individual high-risk patient prediction.

VII. SECURITY, CLINICAL CONSIDERATIONS AND LIMITATIONS

A. Security Controls

The system incorporates several security mechanisms for protecting patient information and controlling application access.

C1. Authentication: JWT-based authentication protects restricted application endpoints.

C2. Role-Based Authorization: Access to patient records and diagnostic operations is controlled according to user roles.

C3. Origin Validation: Origin checking is applied to prevent unauthorized cross-origin communication.

C4. Transaction Integrity: Database transactions maintain consistency between patient records and diagnostic timeline entries.

C5. Data Encryption: Sensitive patient information is protected using encryption mechanisms for storage and communication.

C6. Password Protection: Passwords are stored using secure bcrypt-based hashing.

The implemented security controls provide application-level protection; additional institutional security assessment and healthcare compliance procedures are required for real-world deployment.

B. Clinical Considerations

The proposed system is intended to function as a clinical decision-support platform. Its predictions are designed to provide additional

information that can be reviewed alongside patient history, examination findings, laboratory measurements, and professional clinical judgment. The system does not replace healthcare professionals and should not be used as an autonomous diagnostic mechanism.

The machine learning predictions are derived from retrospective structured data and therefore require additional evaluation before being used in clinical environments.

C. Limitations

Several limitations remain.

First, the study uses a publicly available cardiovascular dataset. Consequently, the results may not represent the complete diversity of patient populations encountered in different healthcare institutions.

Second, retrospective dataset evaluation does not establish prospective clinical effectiveness.

Third, SHAP and LIME describe model behavior but do not establish causal relationships between clinical variables and cardiovascular disease.

Fourth, model performance can change when applied to populations with different demographic or clinical distributions.

Fifth, external validation using datasets from independent healthcare institutions has not been established within the current study.

Finally, prospective studies involving clinicians and real patient workflows are required to assess usability, safety, clinical utility, and decision-making impact.

VIII. CONCLUSION AND FUTURE WORK

This paper presented an explainable machine learning framework for cardiovascular disease prediction that integrates heterogeneous classification models, systematic validation, XAI-based interpretation, and secure clinical workflow management. Random Forest, XGBoost, Support Vector Machine, and Logistic Regression provide complementary approaches for analyzing structured cardiovascular data.

The framework incorporates cross-validation and independent testing to provide a more comprehensive assessment of model generalization. SHAP provides global feature attribution, while LIME generates patient-specific explanations, allowing model outputs to be examined at both population and individual levels.

The implemented platform additionally provides role-based authentication, patient record management, diagnostic history, audit logging, visualization, and report generation. These components connect the machine learning pipeline with an operational healthcare-oriented application.

The previously reported implementation achieved a maximum accuracy of 94.2% using Random Forest, with an AUROC of 0.97. The revised experimental framework provides additional evaluation through cross-validation and independent testing.

The results indicate that explainable machine learning can provide a useful foundation for cardiovascular risk assessment; however, computational performance alone is insufficient to establish clinical applicability. External validation, prospective evaluation, clinician involvement, and assessment across diverse patient populations are necessary before deployment in real healthcare environments.

Future work will focus on external multi-institutional validation, prospective clinical studies, fairness analysis, model calibration, uncertainty estimation, longitudinal patient-data analysis, federated learning, and improved explanation evaluation. Integration with standardized healthcare data systems and additional disease prediction tasks can further extend the capabilities of the framework.

REFERENCES

[1] G. A. Roth, G. A. Mensah, C. O. Johnson et al., "Global burden of cardiovascular diseases and risk factors, 1990-2019: Update from the GBD 2019 study," Journal of the American College of Cardiology, vol. 76, no. 25, pp. 2982-3021, 2020, doi: 10.1016/j.jacc.2020.11.010.

[2] V. Mienye and Y. Sun, "A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges," Informatics in Medicine Unlocked, vol. 51, p. 101587, 2024, doi: 10.1016/j.imu.2024.101587.

[3] M. Salimparsa et al., "Explainable AI for Clinical Decision Support Systems: Literature Review, Key Gaps, and Research Synthesis," Informatics, vol. 12, no. 4, p. 119, 2025, doi: 10.3390/informatics12040119.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135-1144, doi: 10.1145/2939672.2939778.

[5] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," arXiv preprint arXiv:1705.07874, 2017, doi: 10.48550/arXiv.1705.07874.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.

[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.

[8] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273-297, 1995, doi: 10.1007/BF00994018.

[9] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115-118, 2017, doi: 10.1038/nature21056.

[10] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44-56, 2019, doi: 10.1038/s41591-018-0300-7.

[11] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, pp. 1347-1358, 2019, doi: 10.1056/NEJMra1814259.

[12] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, p. 18, 2018, doi: 10.1038/s41746-018-0029-1.

[13] D. Prabhakaran, P. Jeemon, and A. Roy, "Cardiovascular diseases in India: Current epidemiology and future directions," Circulation, vol. 133, no. 16, pp. 1605-1620, 2016, doi: 10.1161/CIRCULATIONAHA.114.008729.

[14] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. Schmid, S. Sandhu, K. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," American Journal of Cardiology, vol. 64, no. 5, pp. 304-310, 1989, doi: 10.1016/0002-9149(89)90524-9. Dataset: A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart Disease," UCI Machine Learning Repository, 1989, doi: 10.24432/C52P4X.

[15] M. Chua et al., "Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges," Healthcare, vol. 13, no. 17, p. 2154, 2025, doi: 10.3390/healthcare13172154.

[16] S. Muneer, T. M. Ghazal, T. Alyas, M. A. Raza, S. Abbas, O. AlZoubi, and O. Ali, "Explainable AI-driven chatbot system for heart disease prediction using machine learning," International Journal of Advanced Computer Science and Applications, vol. 15, no. 12, pp. 249-258, 2024, doi: 10.14569/IJACSA.2024.0151227.

[17] R. Titti, S. Pukkella, and T. S. L. Radhika, "Augmenting heart disease prediction with explainable AI: A study of classification models," Computational and Mathematical Biophysics, vol. 12, no. 1, art. 20240004, 2024, doi: 10.1515/cmb-2024-0004.

[18] B. Majhi and A. Kashyap, "Explainable AI-driven machine learning for heart disease detection using ECG signal," Applied Soft Computing, vol. 167, art. 112225, 2024, doi: 10.1016/j.asoc.2024.112225.

[19] F. M. Talaat, A. R. Elnaggar, W. M. Shaban, M. Shehata, and M. Elhosseini, "CardioRiskNet: A hybrid AI-based model for explainable risk prediction and prognosis in cardiovascular disease," Bioengineering, vol. 11, no. 8, art. 822, 2024, doi: 10.3390/bioengineering11080822.

[20] H. El-Sofany, B. Bouallegue, and Y. M. Abd El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," Scientific Reports, vol. 14, art. 23277, 2024, doi: 10.1038/s41598-024-74656-2.

[21] Y. Li, K. Zong, Y. Zhou, Y. Sun, Y. Liu, B. Zhou, and Z. Wu, "Enhanced preoperative prediction of pancreatic fistula using radiomics and clinical features with SHAP visualization," Frontiers in Bioengineering and Biotechnology, vol. 13, art. 1510642, 2025, doi: 10.3389/fbioe.2025.1510642.

[22] H. Luo et al., "SHAP based predictive modeling for 1-year all-cause readmission risk in elderly heart failure patients: Feature selection and model interpretation," Scientific Reports, vol. 14, art. 17728, 2024, doi: 10.1038/s41598-024-67844-7.

[23] J. K. Paulus and D. M. Kent, "Predictably unequal: Understanding and addressing concerns that algorithmic clinical prediction may increase health disparities," npj Digital Medicine, vol. 3, no. 1, pp. 1-8, 2020, doi: 10.1038/s41746-020-0304-9.

[24] A. Rajkomar, M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin, "Ensuring fairness in machine learning to advance health equity," Annals of Internal Medicine, vol. 169, no. 12, pp. 866-872, 2018, doi: 10.7326/M18-1990.

[25] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care—Addressing ethical challenges," New England Journal of Medicine, vol. 378, no. 11, pp. 981-983, 2018, doi: 10.1056/NEJMp1714229.

[26] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," Science, vol. 366, no. 6464, pp. 447-453, 2019, doi: 10.1126/science.aax2342.

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

[28] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM Computing Surveys, vol. 54, no. 6, pp. 115:1-115:35, 2021, doi: 10.1145/3457607.

[29] R. J. Chen, J. J. Wang, D. F. K. Williamson, et al., "Algorithmic fairness in artificial intelligence for medicine and healthcare," Nature Biomedical Engineering, vol. 7, pp. 719-742, 2023, doi: 10.1038/s41551-023-01056-8.

[30] N. Ganadily and H. J. Xia, "Privacy-preserving machine learning for electronic health records using federated learning and differential privacy," arXiv preprint arXiv:2406.15962, 2024, doi: 10.48550/arXiv.2406.15962.

[31] S. R. Pfohl, A. M. Dai, and K. Heller, "Federated and differentially private learning for electronic health records," arXiv preprint arXiv:1911.05861, 2019, doi: 10.48550/arXiv.1911.05861.

[32] H. Zhao, D. Sui, Y. Wang, L. Ma, and L. Wang, "Privacy-preserving federated learning framework for multi-source electronic health records prognosis prediction," Sensors, vol. 25, no. 8, art. 2374, 2025, doi: 10.3390/s25082374.

[33] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," International Journal of Medical Informatics, vol. 112, pp. 59-67, 2018, doi: 10.1016/j.ijmedinf.2018.01.007.

[34] M. A. de Carvalho Junior and P. Bandiera-Paiva, "Health information system role-based access control: Current security trends and challenges," Journal of Healthcare Engineering, vol. 2018, art. 6510249, 2018, doi: 10.1155/2018/6510249.

[35] M. A. de Carvalho Junior and P. Bandiera-Paiva, "Implications of loosened role-based access control session control implementation for the enforcement of dynamic mutually exclusive roles properties on health information systems," Informatics in Medicine Unlocked, vol. 27, art. 100780, 2021, doi: 10.1016/j.imu.2021.100780.

ISSN: 3139-0862

[36] R. Zhang, D. Chen, X. Shang, X. Zhu, and K. Liu, "A knowledge-constrained role-based access control model for protecting patient privacy in hospital information systems," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 3, pp. 904-911, 2018, doi: 10.1109/JBHI.2017.2696573.

[37] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in Proc. 21st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2015, pp. 1721-1730, doi: 10.1145/2783258.2788613.

[38] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017, doi: 10.48550/arXiv.1702.08608.

[39] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," ACM Computing Surveys, vol. 51, no. 5, art. 93, 2018, doi: 10.1145/3236009.

[40] R. C. Deo, "Machine learning in medicine," Circulation, vol. 132, no. 20, pp. 1920-1930, 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.