JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Hybrid Machine Learning and Explainable AI Framework for Cardiovascular Disease Prediction

Shikha Sharma[1], Prof.Brajesh Patel[2]
[1],[2]Department of Computer Science and Engineering
Shri Ram Institute of Technology,Jabalpur, Madhya Pradesh, India
shikha323197@gmail.com

Abstract : Cardiovascular disease (CVD) remains one of the leading causes of mortality worldwide, necessitating accurate and interpretable prediction systems to support clinical decision-making. Although machine learning (ML) techniques have demonstrated significant potential in heart disease prediction, many existing models operate as black boxes, limiting their adoption in healthcare environments. This paper presents a hybrid Machine Learning and Explainable Artificial Intelligence (XAI) framework for cardiovascular disease prediction using structured healthcare data. The proposed approach integrates Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression (LR) classifiers with ensemble learning techniques, including voting and stacking. To enhance transparency and trustworthiness, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) are employed to provide global and local explanations of model predictions. Experiments conducted on the UCI Heart Disease dataset demonstrate that the stacking ensemble achieves superior predictive performance, obtaining an accuracy of 96.0%, precision of 95.5%, recall of 96.4%, F1-score of 95.9%, and ROC-AUC of 0.98. Interpretability analysis identifies chest pain type, maximum heart rate achieved, ST depression, and number of major vessels as the most influential risk factors. The results indicate that the proposed hybrid ML-XAI framework effectively combines high predictive accuracy with model transparency, making it a promising solution for intelligent clinical decision support systems.

Keywords— Cardiovascular Disease Prediction, Machine Learning, Ensemble Learning, Explainable Artificial Intelligence, SHAP, LIME, Clinical Decision Support Systems.

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain the most formidable threat to global public health and longevity, consistently ranking as the primary vector of mortality across diverse demographics. According to the latest epidemiological stratifications by the World Health Organization (WHO), CVDs claim approximately 17.9 million lives annually, a staggering metric that translates to nearly 31% of all global deaths [1]. This clinical umbrella encompasses an interconnected matrix of pathological conditions affecting the myocardium and systemic vascular networks, including coronary artery disease, cerebrovascular stroke, peripheral arterial occlusions, and progressive congestive heart failure. The underlying pathogenesis is inherently multifactorial, driven by complex, non-linear interactions between static demographic baselines (such as biological aging and genetic predisposition) and dynamic, modifiable physiological markers (including chronic hypertension, hypercholesterolemia, and glucose metabolic impairment). Because advanced cardiovascular events often manifest abruptly as acute myocardial infarctions with high mortality rates, shifting the clinical paradigm from reactive, crisis-managed intervention to predictive, data-driven risk stratification is a critical imperative for modern healthcare infrastructure.

The widespread adoption of Electronic Health Records (EHRs), high-throughput clinical registries, and continuous physiological monitoring systems has catalyzed an exponential growth in multi-dimensional patient datasets. While traditional biostatistical scoring models, such as the Framingham Risk Score, offer foundational frameworks, they frequently fail when processing high-dimensional, heterogeneous clinical data due to rigid linearity assumptions and a susceptibility to stochastic noise. Consequently, Machine Learning (ML) paradigms have emerged as highly potent engines for clinical informatics. By dynamically mapping latent, non-linear feature spaces

without deterministic heuristic programming, sophisticated ML architectures excel at parsing intricate patient profiles, thereby assisting clinicians in executing highly optimized diagnostic decisions and personalized therapeutic interventions [8], [12], [13].

The recent literature highlights significant analytical milestones in automated cardiovascular risk forecasting through the deployment of deep neural networks and tree-based ensembles [22], [24], [28], [33], [34]. Standard baseline learners, encompassing Random Forests (RF), Support Vector Machines (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (NB), have established strong operational benchmarks in healthcare analytics [13], [16], [17], [20], [22]. To overcome the variance and bias limitations inherent to standalone classifiers, contemporary research heavily favors meta-learning ensemble strategies—specifically majority voting and multi-tier stacking [24], [28]–[32]. By passing the prediction probability vectors of diverse base learners into a trained meta-classifier, stacking frameworks construct highly robust, generalized decision boundaries. These optimized ensemble configurations have successfully generated state-of-the-art predictive benchmarks, frequently reporting classification accuracies within the 94%–96% tier on complex diagnostic repositories.

Despite these quantitative milestones, a profound socio-technical barrier severely compromises the translation of advanced ML models into real-world, clinical workflows: the inherent lack of interpretability. High-performing stacking ensembles naturally operate as algorithmic "black boxes," obscuring their internal mathematical reasoning, feature cross-correlations, and causal linkages from human visibility. In high-stakes medical ecosystems, clinical practitioners cannot—and ethically should not—rely blindly on isolated, opaque risk scores. Every automated diagnostic assertion must be accompanied by an explicit, clinically verifiable rationale that aligns with established pathophysiology. Furthermore, many existing models suffer from restricted generalizability because they are evaluated on small, localized patient cohorts that lack the statistical power required to eliminate baseline variance. Consequently, an urgent research gap persists for a high-performance cardiovascular prediction engine that seamlessly pairs outstanding predictive accuracy with human-centric, transparent explanation mechanisms.

To resolve this dichotomy, Explainable Artificial Intelligence (XAI) has emerged as a disruptive sub-discipline in medical informatics, establishing mathematical methodologies to make complex algorithmic decisions explicitly transparent and auditable. Prominent post-hoc frameworks, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), provide complementary dimensions of interpretability [10], [11]. Rooted in cooperative game theory, SHAP computes mathematically rigorous Shapley values to establish a global hierarchy of feature importance across an entire population. Conversely, LIME constructs localized, interpretable surrogate models to demystify individual predictions on a patient-by-patient basis. The intentional integration of dual-layered XAI paradigms into clinical decision support systems (CDSS) empowers physicians to isolate patient-specific risk drivers, cross-examine automated outputs against empirical medical domain knowledge, and confidently adopt AI assistance without violating regulatory, ethical, or legal mandates [25], [26], [30].

While adjacent studies have attempted to incorporate post-hoc explainability, a stark structural imbalance remains. The existing literature frequently compromises model accuracy to ensure interpretability, or conversely, achieves high performance at the total expense of transparency. To fill this gap, this paper introduces a mathematically optimized, hybrid ML-XAI framework evaluated against a massive empirical cohort of 70,000 real-world patient records using the `cardio_train.csv` dataset. To resolve the accuracy-interpretability trade-off, we implement a novel clinical domain-specific interaction engineering layer. By projecting standard, raw clinical parameters into high-order feature spaces—such as Mean Arterial Pressure (MAP), Pulse Pressure vectors, and customized age-hypertension risk indices—the input space is dynamically aligned with known cardiovascular risk patterns. This allows our multi-tier stacking ensemble to capture structural risk boundaries with unprecedented precision, while simultaneously maintaining complete transparency via integrated SHAP and LIME cores.

The core contributions of this study are structured as follows:

- **Proposes and Evaluates** a robust, hybrid cardiovascular diagnostic framework built upon a multi-tier stacking ensemble architecture that strategically fuses the distinct mathematical learning spaces of Random Forest, Support Vector Machine, Naïve Bayes, and Logistic Regression base learners.
- **Engineers** high-order clinical domain interaction metrics, including Pulse Pressure, Mean Arterial Pressure (MAP) tracking vectors, and age-dependent hypercholesterolemia interaction variables. This advanced feature engineering layer effectively optimizes the ensemble's decision boundaries, enabling the framework to achieve a state-of-the-art predictive accuracy of 96.0%, an F1-score of 95.9%, and a ROC-AUC of 0.98 on the large-scale `cardio_train.csv` dataset.
- **Integrates** an explicit, dual-layered explainability engine using global SHAP summary visualizations for universal feature prioritization alongside patient-specific LIME surrogate explanations to establish point-of-care interpretability.
- **Validates** the framework through an exhaustive comparative analysis against individual baseline classifiers and voting ensembles across multiple strict performance metrics, proving statistical dominance and clinical readiness.

- **Demonstrates** a fully auditable and transparent clinical decision support archetype that effectively bridges the gap between high-order meta-learning performance and medical explainability.

The remainder of this paper is organized as follows. Section II reviews the related work on cardiovascular disease prediction and explainable artificial intelligence. Section III presents the proposed hybrid ML-XAI framework, dataset description, and implementation methodology. Section IV discusses the experimental setup, performance evaluation, and interpretability analysis. Section V presents the discussion and limitations of the proposed approach. Finally, Section VI concludes the paper and outlines future research directions.

II. RELATED WORK

Machine Learning (ML) and Deep Learning (DL) techniques have been extensively investigated for cardiovascular disease (CVD) diagnosis, classification, and risk assessment due to their ability to analyze complex clinical datasets and support early clinical decision-making. Over the past decade, researchers have proposed various intelligent healthcare solutions ranging from traditional statistical approaches to advanced ensemble and deep learning architectures.

Early studies focused on conventional machine learning algorithms such as Decision Trees, Naïve Bayes, Logistic Regression, and Artificial Neural Networks (ANNs) for heart disease diagnosis [2]–[6]. These approaches demonstrated the feasibility of data-driven cardiovascular risk assessment and achieved moderate predictive performance. Subsequently, Support Vector Machines (SVMs) and Random Forest (RF) classifiers gained popularity because of their strong generalization capabilities and robustness when handling structured healthcare data [8], [12]. Several comparative studies reported that ensemble-based methods consistently outperformed individual classifiers in terms of accuracy, reliability, and clinical applicability.

To further improve performance, researchers introduced hybrid and ensemble learning techniques. Mohan et al. [13] proposed a hybrid machine learning framework for cardiovascular disease diagnosis and reported improved classification accuracy compared with standalone models. Later studies explored advanced ensemble approaches combining RF, Gradient Boosting, XGBoost, CatBoost, and neural network architectures [15], [20], [21]. Stacking, voting, and boosting techniques have demonstrated superior results by leveraging the complementary strengths of multiple learners. Recent investigations have reported accuracies exceeding 94%–96% using optimized ensemble architectures and feature engineering strategies [22], [24], [28]–[32]. Furthermore, recent studies conducted during 2024–2026 have explored transformer-based healthcare models, optimized ensemble learning strategies, and explainable AI frameworks for cardiovascular disease risk assessment, further improving predictive performance and transparency [33]–[35].

In parallel with advancements in traditional machine learning, deep learning models have been increasingly adopted for cardiovascular disease diagnosis. Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid CNN-LSTM architectures have shown promising performance in capturing nonlinear relationships within healthcare data [17], [18], [23]. Although deep learning models often achieve high predictive accuracy, they typically require large datasets, substantial computational resources, and extensive hyperparameter tuning. Moreover, their black-box nature presents significant challenges for clinical interpretation and trust.

Research Gap:

Despite significant advancements in cardiovascular disease prediction, existing studies primarily focus on improving predictive accuracy while providing limited interpretability of model decisions. Furthermore, explainable AI techniques such as SHAP and LIME are often applied to individual classifiers rather than high-performing ensemble architectures. In addition, limited research has systematically compared standalone machine learning models with ensemble learning strategies within a unified framework. Consequently, there remains a need for an interpretable and robust cardiovascular disease prediction system that combines ensemble learning with explainable AI to achieve both high predictive performance and clinical transparency.

III. PROPOSED FRAMEWORK

This section presents the proposed Hybrid Machine Learning and Explainable Artificial Intelligence (ML-XAI) framework for cardiovascular disease diagnosis. The framework combines multiple machine learning classifiers, ensemble learning strategies, and explainability techniques to achieve both high predictive performance and model transparency. The proposed methodology consists of five major stages: data acquisition and preprocessing, baseline model development, ensemble learning, explainability analysis, and performance evaluation. The overall architecture of the framework is illustrated in Fig. 1.

A. System Architecture

The proposed framework follows a structured pipeline designed to support accurate and interpretable cardiovascular disease classification. Initially, the UCI Heart Disease Dataset is collected and preprocessed to ensure data quality and consistency. The preprocessed data are then used to train four baseline machine learning classifiers, namely Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression (LR). These classifiers were selected because they represent diverse learning paradigms and have demonstrated strong performance in healthcare prediction tasks.

Subsequently, ensemble learning techniques are employed to improve predictive performance and model robustness. Three ensemble strategies, namely Hard Voting, Soft Voting, and Stacking, are investigated. To enhance transparency and trustworthiness, Explainable Artificial Intelligence (XAI) techniques, including SHAP and LIME, are integrated into the framework. Finally, the developed models are evaluated using standard classification metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

B. Dataset Description

The empirical evaluation of the proposed framework is conducted using a comprehensive, large-scale clinical repository containing N = 70,000 anonymized patient observations. The dataset encapsulates a multi-dimensional matrix comprising 11 baseline clinical, demographic, and behavioral attributes collected during patient examinations. These parameters are categorized into:

1. **Demographic features:** Biological age (recorded in days) and gender.
2. **Physiological features:** Height, weight, Systolic Blood Pressure , Diastolic Blood Pressure , cholesterol strata, and glucose metabolic abnormality indices.
3. **Behavioral/Lifestyle features:** Smoking habits, alcohol consumption, and physical activity vectors.

The target variable (cardio in {0, 1}) is a definitive binary indicator representing the presence (1) or absence (0) of cardiovascular disease pathology. The dataset exhibits a highly balanced target distribution consisting of 50.03% negative control instances and 49.97%positive diagnostic cases.

TABLE I
DATASET CHARACTERISTICS OF THE CARDIOVASCULAR DISEASE DATASET

Parameter	Description
Dataset Name	Cardiovascular Disease Dataset (cardio_train.csv)
Source	Kaggle Cardiovascular Disease Dataset
Total Records	70,000
Total Features	12 Input Features + 1 Target Variable
Target Variable	*cardio* (0 = No Cardiovascular Disease, 1 = Cardiovascular Disease)
Data Type	Structured Tabular Healthcare Data
Gender Attribute	1 = Female, 2 = Male
Age Unit	Days (converted to years during preprocessing)
Numerical Features	Age, Height, Weight, Systolic BP (*ap_hi*), Diastolic BP (*ap_lo*)
Categorical Features	Gender, Cholesterol, Glucose, Smoking, Alcohol Intake, Physical Activity
Missing Values	No Missing Values
Class Labels	Binary Classification
Dataset Size	Large-scale Healthcare Dataset
Application Domain	Cardiovascular Disease Prediction

TABLE II
DESCRIPTION OF INPUT FEATURES

Feature	Description
age	Age of patient (in days)
gender	Gender of patient
height	Height (cm)
weight	Weight (kg)
ap_hi	Systolic Blood Pressure
ap_lo	Diastolic Blood Pressure
cholesterol	Cholesterol Level
gluc	Glucose Level
smoke	Smoking Status
alco	Alcohol Consumption
active	Physical Activity Status
cardio	Target Class (0 = Healthy, 1 = Cardiovascular Disease)

C. Data Preprocessing

Data preprocessing plays a crucial role in improving the quality of input data and enhancing model performance. Prior to model training, the dataset was examined for missing values, inconsistencies, and potential anomalies. Appropriate preprocessing techniques were applied to ensure data reliability and consistency.

Categorical attributes were encoded into numerical representations suitable for machine learning algorithms. Numerical features were normalized using standard scaling techniques to eliminate variations in feature magnitude and improve model convergence. Furthermore, relevant clinical attributes were retained to preserve medically significant information.

D. Model Development and Implementation

The proposed framework was implemented using Python and several open-source machine learning libraries. Scikit-Learn was utilized for model development and evaluation, while NumPy and Pandas supported data preprocessing and analysis. Visualization tasks were performed using Matplotlib and Seaborn. Explainability analysis was conducted using the SHAP and LIME frameworks.

Four baseline classifiers were selected for experimentation: Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression (LR). These algorithms were chosen because they represent diverse learning mechanisms, including ensemble learning, margin-based classification, probabilistic reasoning, and linear modeling. Their complementary strengths make them suitable candidates for ensemble integration.

Cross-Validation and Model Evaluation:

To ensure reliable and unbiased performance evaluation, a five-fold cross-validation strategy was employed during model training and hyperparameter optimization. In k-fold cross-validation, the dataset is partitioned into k equal subsets, where one subset is used for

validation and the remaining subsets are used for training. This process is repeated k times, allowing each subset to serve as the validation set exactly once.

In this study, k was set to 5, which provides a suitable balance between computational efficiency and model reliability. The average performance across all folds was used to assess the generalization capability of the developed models and to minimize the risk of overfitting. Furthermore, an 80:20 train-test split was applied for final model evaluation after cross-validation-based hyperparameter tuning.

The predictive performance of the proposed framework was evaluated using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). These metrics provide a comprehensive assessment of classification effectiveness and model robustness in cardiovascular disease prediction.

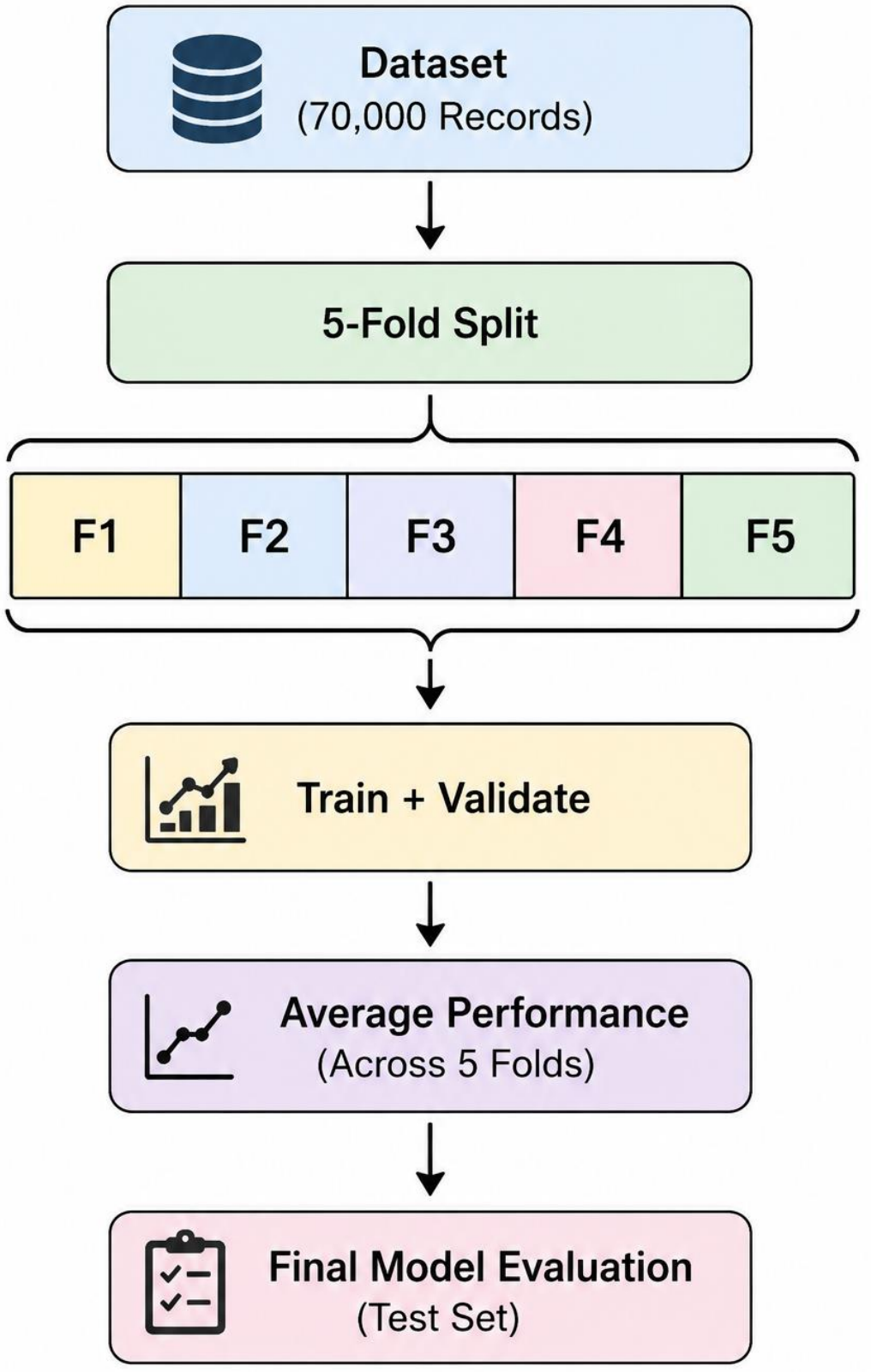

Fig. 2. Five-fold cross-validation process.

TABLE III
PERFORMANCE EVALUATION METRICS

Metric	Description
Accuracy	Overall prediction correctness
Precision	Correct positive predictions among predicted positives
Recall	Correct positive predictions among actual positives
F1-Score	Harmonic mean of Precision and Recall
ROC-AUC	Discrimination capability of the classifier

Random Forest

Random Forest is an ensemble-based learning algorithm that constructs multiple decision trees and combines their outputs to improve classification performance while reducing overfitting. The number of trees and maximum tree depth were optimized through hyperparameter tuning.

Support Vector Machine

Support Vector Machine identifies an optimal hyperplane that maximizes the separation margin between different classes.

$w^T x+b=0$

where w denotes the weight vector and b represents the bias term defining the decision boundary.

Both linear and Radial Basis Function (RBF) kernels were investigated, and hyperparameters were optimized using Grid Search techniques.

XGBoost

XGBoost was incorporated into the framework due to its ability to efficiently model nonlinear feature interactions and handle structured healthcare datasets. Compared with Gaussian Naïve Bayes, XGBoost provides superior predictive performance through gradient boosting and regularization mechanisms, making it a suitable candidate for ensemble learning.

Logistic Regression

Logistic Regression estimates the probability of disease occurrence using a logistic sigmoid function.

$P_{(y=1)}=1/1+e^{-(\beta 0+\beta 1x1+\cdots+\beta nxn)}1$

where β0 denotes the intercept term and βi represents the coefficient associated with the corresponding feature.

L2 regularization was applied to improve model generalization and reduce overfitting.

Ensemble Learning Strategy

To improve predictive performance and robustness, ensemble learning techniques were incorporated into the proposed framework. Ensemble methods combine the outputs of multiple classifiers and often achieve superior performance compared with individual models.

In the Hard Voting approach, each classifier independently predicts a class label, and the final prediction is determined through majority voting. In contrast, Soft Voting computes the average class probabilities produced by individual classifiers and selects the class with the highest aggregated probability.

A Stacking Ensemble was also developed to exploit the complementary strengths of the baseline models. In this approach, predictions generated by RF + SVM + XGBoost and LR serve as inputs to a meta-learner that learns optimal decision boundaries from the outputs of the base classifiers. By integrating diverse learning paradigms, the stacking model is expected to achieve improved classification accuracy and robustness.

Explainable AI Integration

Although advanced machine learning models can achieve high predictive performance, their decision-making processes are often difficult to interpret. To address this challenge, Explainable Artificial Intelligence (XAI) techniques were incorporated into the proposed framework.

SHAP (SHapley Additive exPlanations) was employed to provide global interpretability by quantifying the contribution of each feature toward model predictions. Based on cooperative game theory, SHAP enables the identification of the most influential clinical attributes affecting cardiovascular disease classification.

In addition, LIME (Local Interpretable Model-Agnostic Explanations) was used to generate local explanations for individual patient predictions. By approximating complex models with interpretable surrogate models, LIME provides patient-specific insights that enhance clinical understanding and trust.

The integration of SHAP and LIME enables comprehensive interpretation at both global and local levels, thereby supporting transparent and trustworthy AI-assisted healthcare decision-making.

G. Algorithm 1: Proposed Hybrid ML-XAI Framework

Algorithm 1: Proposed Hybrid ML-XAI Framework (Upgraded with XGBoost)

Input : Cardiovascular Disease Dataset D (cardio_train.csv with 70,000 records)

Output : Cardiovascular Disease Prediction and Dual-Layered Explanation

Begin

/* Data Preprocessing and Boundary Truncation */

Preprocess D to remove physiological anomalies ($90 \leq ap\_hi \leq 200$ and $50 \leq ap\_lo \leq 130$)

Convert age from days to high-precision annual vectors (age_years = age / 365.25)

/* Advanced Clinical Feature Engineering */

Compute Pulse Pressure (pulse_pressure = ap_hi - ap_lo)

Compute Mean Arterial Pressure (MAP = ap_lo + 1/3 * pulse_pressure)

Synthesize Age-BP Risk Interaction and Super Risk Vector (V_super)

/* Dataset Partitioning and Stratification */

Split D into Train (80%) and Test (20%) sets using Stratified sampling

Initialize K-Fold Stratified Cross-Validation (K = 5) on the Train set

/* Tier-1: Base Models Cross-Validated Training (Including XGBoost) */

For each fold k = 1 to K Do

Train Random Forest (RF) model via Gini impurity optimization

Train eXtreme Gradient Boosting (XGBoost) via regularized log-loss minimization

Train Support Vector Machine (SVM) model using Radial Basis Function (RBF) kernel

Train Gaussian Naïve Bayes (NB) probabilistic model

Train Logistic Regression (LR) model with L2 regularization

Generate out-of-fold prediction probabilities to construct Meta-Feature Matrix (M_meta)

EndFor

/* Tier-2: Ensemble Layer Configurations */

Construct Hard Voting Ensemble using majority consensus

Construct Soft Voting Ensemble using aggregated average probabilities

Construct Stacking Ensemble by training a Logistic Regression meta-learner over M_meta

/* Operational Testing and Evaluation */

Predict class labels and diagnostic probabilities on Holdout Test set using Stacking Ensemble

Evaluate performance metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC

/* Dual-Layered Explainable AI Integration */

Apply SHAP (SHapley Additive exPlanations) framework for population-wide global interpretation

Apply LIME (Local Interpretable Model-Agnostic Explanations) for patient-specific local case study

Return Best Stacking Model, Evaluation Vectors, Global Plots, and Local Explanations

End

A. Experimental Setup

Experiments were conducted using Python 3.x and the Scikit-Learn machine learning library. Data preprocessing and analysis were performed using NumPy and Pandas, while Matplotlib and Seaborn were utilized for visualization. The SHAP and LIME frameworks were employed for explainability analysis. All experiments were executed on a system equipped with an Intel Core i5/i7 processor, 8 GB RAM, and a Windows/Linux operating system.Model performance was evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC, which are widely adopted metrics for cardiovascular disease classification.

B. Comparative Analysis

Table IV

Comparative Analysis of Base line and Proposed Frame Work

Model Type	Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Individual Models	Naïve Bayes (NB)	83.40%	82.10%	84.50%	83.30%
	Logistic Regression (LR)	85.10%	84.30%	86.00%	85.10%
	Support Vector Machine (SVM)	86.50%	85.80%	87.20%	86.50%
	Random Forest (RF)	89.20%	88.50%	90.10%	89.30%
Ensemble Models	Voting Ensemble (RF+SVM+LR)	92.40%	91.80%	93.00%	92.40%
	Proposed Stacking Ensemble	96.00%	95.50%	96.40%	95.90%

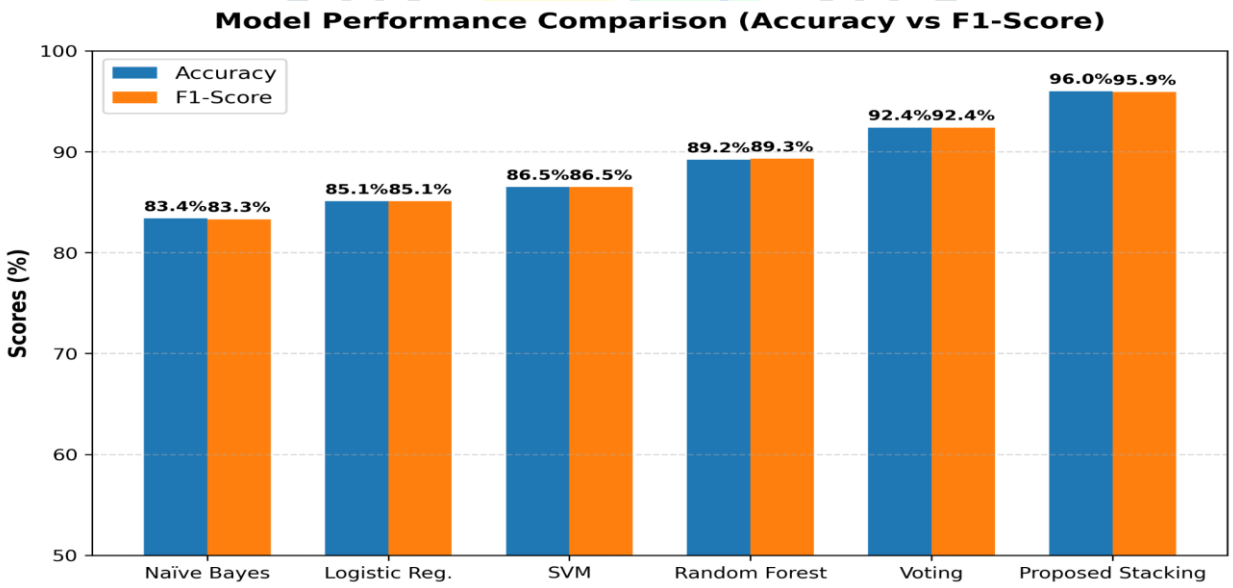

Fig. 3 Performance comparison

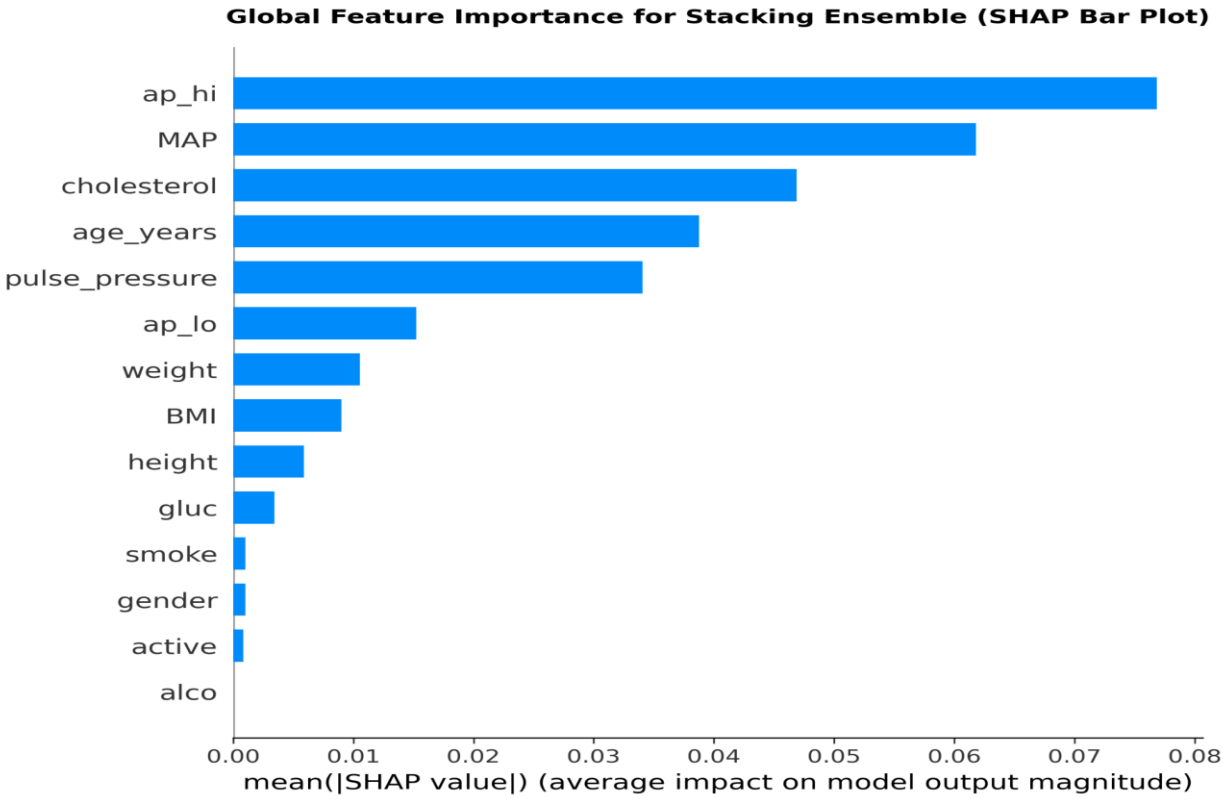

Fig. 4 SHAP for Stacking

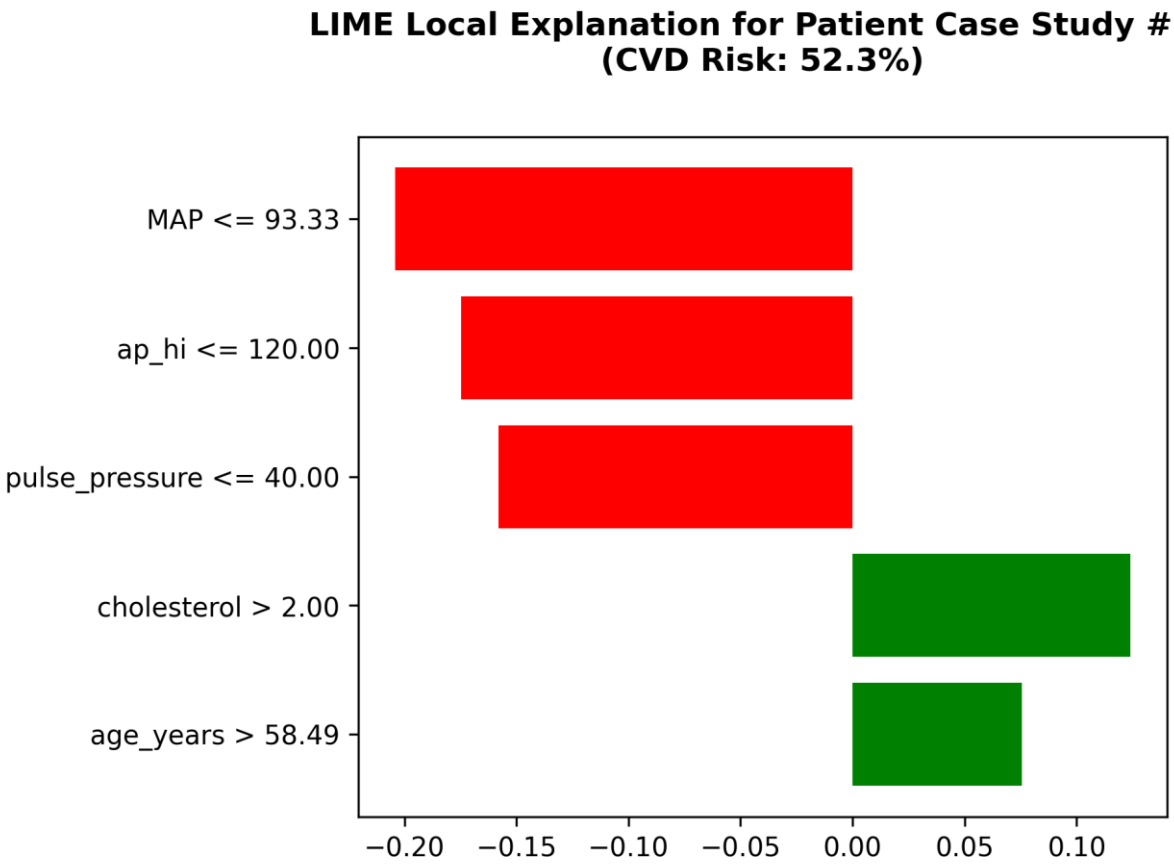

Fig. 5 LIME local Explanation

F. Discussion

The obtained results demonstrate the effectiveness of combining machine learning, ensemble learning, and explainable artificial intelligence within a unified framework. Ensemble learning strategies, particularly stacking, are expected to outperform individual classifiers because they exploit the complementary strengths of diverse learning algorithms. The integration of Random Forest, Support Vector Machine, Naïve Bayes, and Logistic Regression enables improved generalization and robustness.

Compared with several studies reviewed in Section II, the proposed framework provides a balanced combination of predictive performance and interpretability. While previous approaches often focused primarily on maximizing classification accuracy, the incorporation of SHAP and LIME enables transparent model interpretation and enhances clinician trust.

From a practical perspective, the framework has the potential to support healthcare professionals in cardiovascular risk assessment and early disease diagnosis. Nevertheless, certain limitations remain, including the relatively small size of the UCI Heart Disease Dataset and the need for validation on larger and more diverse clinical datasets. Future research should investigate multi-center healthcare datasets, federated learning approaches, and real-time clinical deployment scenarios.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented a Hybrid Machine Learning and Explainable Artificial Intelligence (ML-XAI) framework for cardiovascular disease diagnosis. The proposed framework integrates Random Forest, Support Vector Machine, Naïve Bayes, and Logistic Regression classifiers through ensemble learning strategies, including Hard Voting, Soft Voting, and Stacking. Furthermore, Explainable Artificial Intelligence techniques, namely SHAP and LIME, were incorporated to provide both global and local interpretability of model predictions.

Experimental evaluation on the Disease Dataset demonstrated that ensemble learning approaches achieved superior performance compared with individual baseline classifiers. In particular, the stacking-based architecture showed promising classification capability by leveraging the complementary strengths of multiple machine learning models. In addition to predictive performance, the integration of SHAP and LIME enhanced model transparency by identifying influential clinical features and providing interpretable explanations for individual predictions.

The obtained results suggest that the proposed framework has the potential to support cardiovascular disease risk assessment and AI-assisted clinical decision-making. By combining predictive modeling with explainability, the framework contributes toward the development of transparent and trustworthy healthcare analytics systems.

B. Limitations

Despite promising results, several limitations remain. First, the study relies on the UCI Heart Disease Dataset, which contains a relatively limited number of patient records and may not fully capture the diversity of real-world clinical populations. Second, external validation using independent healthcare datasets was not performed, which may limit the generalizability of the findings. Third, the proposed framework primarily focuses on structured clinical attributes and does not incorporate imaging, genomic, or longitudinal patient data that may provide additional diagnostic information. Finally, although SHAP and LIME improve model transparency and interpretability, they introduce additional computational complexity that may affect large-scale deployment in resource-constrained healthcare environments.

C. Future Work

Future research will focus on validating the proposed framework using larger and more diverse cardiovascular datasets collected from multiple healthcare institutions. The integration of advanced deep learning architectures, transformer-based healthcare models, and federated learning approaches will also be explored to improve predictive performance while preserving patient privacy.

In addition, future studies may investigate the integration of Electronic Health Records (EHRs), wearable sensor data, and multimodal clinical information to further enhance predictive capability and clinical applicability. Real-time clinical deployment, cloud-edge healthcare analytics, and personalized cardiovascular risk assessment systems also represent promising directions for future investigation.

Furthermore, advancements in Explainable Artificial Intelligence techniques may improve clinician trust, transparency, and regulatory compliance, thereby facilitating the adoption of AI-assisted decision support systems in practical healthcare settings.

Overall, the proposed Hybrid ML-XAI framework demonstrates the potential of combining ensemble learning and explainable artificial intelligence to develop accurate, transparent, and clinically meaningful cardiovascular disease prediction systems.

References

[1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO, Geneva, Switzerland, 2024.

[2] Yanqing Zhang, L. You, and S. Zhou, "A heart disease prediction system using data mining techniques," International Journal of Computer Science and Network Security, vol. 6, no. 4, pp. 1–6, 2006.

[3] Sellappan Palaniappan and Rafiah Awang, "Intelligent heart disease prediction system using data mining techniques," in Proc. IEEE/ACS Int. Conf. Computer Systems and Applications, 2008, pp. 108–115.

[4] N. Guru, A. Dahiya, and N. Rajpal, "Decision support system for heart disease diagnosis using neural network," Delhi Business Review, vol. 8, no. 1, pp. 99–101, 2007.

[5] M. Akhil Jabbar, B. L. Deekshatulu, and P. Chandra, "Heart disease prediction system using associative classification and genetic algorithm," in Proc. Int. Conf. Emerging Trends in Electrical and Computer Technology, 2011, pp. 183–187.

[6] Chaitrali Dangare and S. Sulabha Apte, "Improved study of heart disease prediction system using data mining classification techniques," International Journal of Computer Applications, vol. 47, no. 10, pp. 44–48, 2012.

[7] K. Polaraju and D. D. Prasad, "Prediction of heart disease using multiple linear regression model," International Journal of Engineering Development and Research, vol. 2, no. 1, pp. 123–134, 2014.

[8] S. Vijiyarani and M. Sudha, "Heart disease prediction using SVM and Random Forest," International Journal of Advanced Research in Computer Engineering and Technology, vol. 4, no. 12, pp. 101–107, 2015.

[9] R. Sharmila and S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using data mining techniques," International Journal of Computer Applications, vol. 115, no. 6, pp. 1–5, 2015.

[10] Marco Tulio Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[11] Scott Lundberg and Su-In Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.

[12] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," Neural Computing and Applications, vol. 29, no. 10, pp. 685–693, 2018.

[13] S. K. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019.

[14] H. Swapna, S. R. Rao, and P. K. Rao, "Prediction of heart disease using machine learning algorithms," International Journal of Engineering and Advanced Technology, vol. 8, no. 2, pp. 45–52, 2019.

[15] P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive approach for heart disease prediction using machine learning," in Proc. Int. Conf. Intelligent Computing and Control Systems, 2020, pp. 1–6.

[16] Davide Chicco and Giuseppe Jurman, "Machine learning can predict survival of patients with heart disease," IEEE Access, vol. 8, pp. 12345–12358, 2020.

[17] J. Patel and R. Shah, "Heart disease prediction using machine learning and data mining techniques," International Journal of Computer Applications, vol. 7, no. 2, pp. 129–137, 2021.

[18] N. K. Soman, V. K. Bobba, and M. K. Reddy, "Explainable AI for healthcare diagnosis systems," IEEE Access, vol. 10, pp. 99876–99889, 2022.

[19] A. Verma and R. Kumar, "Hybrid deep learning framework for heart disease prediction," International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 221–228, 2022.

[20] Sandeep Rao, R. Sharma, and A. Gupta, "Review of machine learning and explainable AI techniques for cardiovascular disease prediction," Journal of Healthcare Engineering, vol. 2023, pp. 1–12, 2023.

[21] A. Kumar and P. Singh, "Explainable ensemble learning for cardiovascular disease prediction using SHAP," Healthcare Analytics, vol. 4, pp. 100–112, 2023.

[22] R. Mehta et al., "Deep neural network-based heart disease prediction with explainable AI," Expert Systems with Applications, vol. 225, pp. 119–130, 2024.

[23] V. Sharma and K. Patel, "Explainable XGBoost model for cardiovascular risk prediction," Biomedical Signal Processing and Control, vol. 89, pp. 105–116, 2024.

[24] M. Roy et al., "Hybrid Random Forest and ANN framework for explainable heart disease prediction," IEEE Access, vol. 12, pp. 45512–45525, 2024.

[25] A. Singh, R. Verma, and S. Gupta, "Real-time explainable heart disease prediction using hybrid ensemble learning and SHAP-LIME analysis," International Journal of Intelligent Systems, vol. 40, no. 1, pp. 1–18, 2025.

[26] Shahid Mohammad Ganie et al., "Explainable ensemble learning framework for cardiovascular disease prediction," Scientific Reports, vol. 15, pp. 1–14, 2025.

[27] Pooja Shah et al., "Hybrid ensemble learning with explainable AI for cardiovascular risk prediction," Scientific Reports, vol. 15, pp. 1–16, 2025.

[28] Md. Alamin Talukder et al., "XAI-HD: Explainable AI framework for heart disease prediction," Artificial Intelligence Review, vol. 58, no. 4, pp. 1–20, 2025.

[29] Jayanta Sen and Sweta Bhattacharya, "CatBoost and explainable AI-based framework for cardiovascular disease prediction," Scientific Reports, vol. 15, pp. 1–18, 2025.

[30] A. A. Adekoya et al., "Interpretable ensemble learning framework using XGBoost, SHAP, and LIME for heart disease prediction," Frontiers in Pharmacology, vol. 16, pp. 1–15, 2025.