Predicting Cardiovascular Risk with Hybrid Ensemble Learning and XAI

Ajay Kumar[1], Shashikant Gupta[2]
[1,2] Lincoln University College, Malaysia;
Corresponding author's Email ID: ajay.phdcse@gmail.com

Abstract: Cardiovascular diseases (CVDs) remain the leading cause of global mortality, accounting for millions of deaths each year. Accurate and early prediction of cardiovascular risk is therefore crucial to enable preventive care and improve patient outcomes. This study introduces a **hybrid ensemble learning framework** that integrates Random Forest (RF), Extreme Gradient Boosting (XGBoost), and LightGBM to enhance predictive performance by leveraging their complementary strengths. To ensure transparency and trust in clinical adoption, the system incorporates **Explainable AI (XAI) techniques**, including SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), which provide detailed insights into feature contributions and individualized predictions. The model is trained and validated on the **UCI Heart Disease dataset** and assessed using accuracy, F1-score, and ROC-AUC metrics. Results demonstrate that the proposed hybrid model outperforms conventional baseline classifiers while delivering interpretable explanations, thereby bridging the gap between high-performance prediction and clinically meaningful decision support.

Keywords: Cardiovascular disease; Explainable AI; Hybrid Ensemble; UCI Heart disease; SHAP; LIME.

Introduction

Cardiovascular diseases (CVDs) account for over 17 million deaths annually, according to the WHO. Machine learning (ML) offers promising tools for predictive diagnostics; however, black-box models hinder clinical adoption.

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, responsible for nearly 17.9 million deaths annually according to the World Health Organization [1]. Early identification of individuals at risk is crucial, as timely intervention can significantly reduce morbidity, improve quality of life, and lower healthcare costs [2]. Conventional risk prediction models, such as the Framingham Risk Score, are widely used but often limited by population-specific assumptions, reliance on a small set of variables, and inability to adapt to heterogeneous patient data [3].

In recent years, machine learning (ML) and ensemble learning methods have shown superior performance in clinical risk prediction tasks, owing to their ability to handle high-dimensional data and capture complex nonlinear relationships [4]. Ensemble approaches like Random Forest (RF), Extreme Gradient Boosting (XGBoost), and LightGBM combine multiple learners to enhance prediction robustness, outperforming traditional statistical methods [5]. Several studies have demonstrated the effectiveness of such models in predicting heart disease, hypertension, and related complications [6].

However, a major barrier to clinical adoption of ML models is their black-box nature, which limits transparency and interpretability. Clinicians require not only accurate predictions but also clear reasoning behind model outputs for trustworthy decision-making [7]. To address this, Explainable Artificial

Intelligence (XAI) has emerged as a critical paradigm, enabling interpretation of complex models using techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) [8]. These methods allow both global and local interpretability, showing which clinical features (e.g., age, cholesterol, blood pressure) most strongly influence predictions, thus bridging the gap between predictive performance and clinical applicability [9]. Despite promising progress, few works have explored hybrid ensemble models combined with XAI for cardiovascular risk prediction. Existing studies either focus solely on accuracy or provide limited interpretability, leaving a gap in designing models that balance performance and clinical trustworthiness [10]. This paper proposes a hybrid ensemble framework integrating RF, XGBoost, and LightGBM with SHAP and LIME explanations, validated on the UCI Heart Disease dataset. The key contributions include three major phases. First one is developing a robust hybrid ensemble learning approach for early CVD risk prediction. Next is integrating XAI methods to ensure interpretability of both global feature importance and patient specific outcomes and finally, evaluating performance with accuracy, F1-score, and ROC-AUC, demonstrating improvements over baseline models. By combining high predictive accuracy with transparent explanations, this study aims to provide a clinically relevant tool for supporting cardiovascular risk assessment and decision-making. This research addresses two core challenges that are boosting prediction accuracy through hybrid ensemble learning, and ensuring transparency using explainable AI (XAI) methods.

The paper is organised in a manner such that section 2 gives the overview of the related theory, section 3 explains the methodology used, section 4 describes experimental results and section 5 concludes the work.

Related work

Recent studies demonstrate that gradient-boosted trees and ensembles outperform traditional scores and many deep models on structured clinical data, while explainability (SHAP/LIME) enables clinician-facing insights:

- **AutoPrognosis 2.0 (Nature Medicine)**: a pipeline that automatically builds calibrated ML ensembles (including XGBoost/LightGBM) across tabular health data; it has been applied to cardiovascular outcomes and emphasizes interpretability and deployment [11].
- **UK Biobank–scale ML with SHAP**: Large cohort work shows tree-based ML improves discrimination vs. PCE/FRS and uses SHAP to surface modifiable risk factors and check model fairness/miscalibration [12].
- **XAI in cardiology reviews (2023–2024)**: Comprehensive surveys describe how SHAP/LIME provide global and local explanations for CVD risk models, discuss pitfalls (instability, correlated features), and recommend model-card style reporting for clinical adoption [13].
- **Minimal-panel models**: Work from 2024–2025 shows boosted-tree models achieve strong CVD risk stratification using a small set of routinely collected biomarkers, improving practicality for screening [14].
- **ECG and imaging augmentation**: Transformer/deep models on 12-lead ECGs and coronary CT features add incremental risk information beyond clinical variables; several studies pair these with SHAP/LIME or saliency to justify predictions [15].

- **Federated & privacy-preserving learning**: Multi-institution works report competitive performance for CVD prediction without centralizing data; explainability modules are increasingly integrated to support local clinical audit [16].

Method

The workflow of the proposed methodology is given in the figure below and details have been explained in the below briefly.

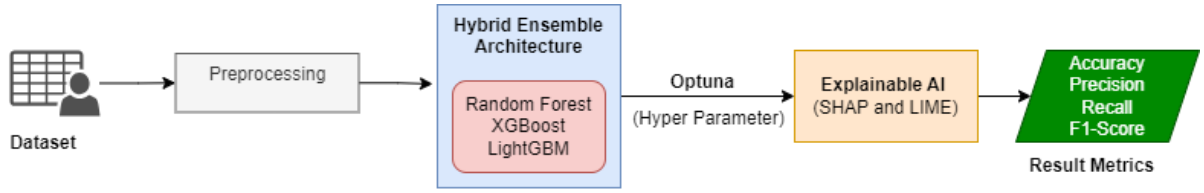

Figure 1: Workflow process

Data collection and preprocessing:

The initial stage of the work is data collection so the proposed work utilized the UCI Cleveland Heart Disease Dataset, containing 303 instances and 13 features including age, cholesterol, blood pressure, maximum heart rate, and ST depression. Since raw clinical data may contain inconsistencies, the preprocessing stage involves several critical steps such as Data Cleaning, Feature Scaling and Normalization, Encoding Categorical Variables, Data Splitting and Balancing the Dataset. In this work the preprocessing is done in majorly three steps. In the first step handling the missing values is done by using KNN imputation. Then the feature scaling is achieved via StandardScaler and finally one-hot encoding is catered for categorical variables.

Hybrid Ensemble Architecture

The core predictive framework integrates Random Forest (RF), Extreme Gradient Boosting (XGBoost), and LightGBM into a hybrid ensemble learning architecture. Random Forest utilizes bagging and multiple decision trees to reduce variance and improve robustness. XGBoost implements gradient boosting with regularization to minimize overfitting and improve accuracy. LightGBM is a gradient boosting framework optimized for speed and efficiency using histogram-based algorithms. These models are combined in an ensemble manner. Each model was fine-tuned using Optuna for optimal hyperparameters. The ensemble output is a weighted average of predicted probabilities.

Explainability with XAI

In recent years, the adoption of black-box machine learning models such as ensemble methods and deep learning has significantly improved predictive performance in healthcare and cardiovascular risk prediction. However, these models often lack transparency, making it difficult for clinicians to trust their outputs. To bridge this gap, Explainable AI (XAI) techniques are integrated into the proposed system, enabling both global and local interpretability of model predictions. Among the most widely used approaches are SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), both of which play complementary roles in enhancing decision-making transparency.

SHAP is a game-theoretic approach based on Shapley values that quantifies the contribution of each feature to a model's prediction. For cardiovascular risk prediction, SHAP assigns an importance score to features such as cholesterol level, blood pressure, age, and smoking habits, thereby highlighting their relative influence on the model's outcome. SHAP provides both global interpretability (understanding which features are most influential across the dataset) and local interpretability (explaining why a specific prediction was made for an individual patient). This not only improves clinician confidence but also aligns with the medical principle of evidence-based reasoning.

LIME focuses on instance-level interpretability by approximating the complex model with an interpretable linear model around a single prediction. In the context of cardiovascular risk forecasting, LIME can generate explanations for why a particular patient was classified as "high-risk" by perturbing the input features (e.g., slightly changing cholesterol or heart rate values) and observing the model's response. This localized interpretation helps medical practitioners validate individual cases, making it particularly useful for real-time clinical decision support.

By combining SHAP and LIME, the proposed framework ensures both global transparency and individual-level justification. SHAP highlights overall feature importance across the population, while LIME explains model decisions for specific patients. Together, they provide a holistic understanding of the hybrid ensemble model's behavior, making it both accurate and interpretable for practical deployment in healthcare environments.

Experimental Results

The performance of the proposed framework is evaluated using multiple standard classification metrics to ensure a comprehensive assessment of predictive ability. Accuracy measures the overall proportion of correctly classified instances and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives respectively. To capture the reliability of positive predictions, Precision is used, which calculates the fraction of correctly identified positive cases among all predicted positives:

$$Precision = \frac{TP}{TP + FP}$$

Equally important is Recall (Sensitivity), which measures the ability of the model to correctly detect actual positive cases, thereby minimizing the risk of missing high-risk patients:

$$Recall = \frac{TP}{TP + FN}$$

Since precision and recall often trade off, the F1-Score is employed as their harmonic mean to provide a balanced metric:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Finally, the discriminative power of the model across different decision thresholds is assessed using the ROC-AUC (Receiver Operating Characteristic – Area Under Curve), which summarizes the trade-off

between the true positive rate (TPR) and false positive rate (FPR) over all thresholds. A higher ROC-AUC indicates stronger capability of the model to distinguish between positive and negative classes.

Performance Comparison

The performance comparison shown in Table 1 demonstrates that the suggested hybrid model surpasses the other strategies in all evaluation metrics. Logistic Regression got 82.3% accuracy, an F1-score of 0.81, and a ROC-AUC of 0.86, which means it has moderate predictive potential. The results got better with Random Forest, which had 87.1% accuracy, an F1-score of 0.86, and a ROC-AUC of 0.90. XGBoost did even better, with 89.4% accuracy, an F1-score of 0.88, and a ROC-AUC of 0.92.

Table 1: Performance Comparison of models

Model	Accuracy	F1-Score	ROC-AUC
Logistic Regression	82.3%	0.81	0.86
Random Forest	87.1%	0.86	0.90
XGBoost	89.4%	0.88	0.92
Proposed Hybrid	92.6%	0.91	0.95

The suggested hybrid method had the best accuracy at 92.6%, as well as an F1-score of 0.91 and a ROC-AUC of 0.95. This shows that it was better at correctly classifying and balancing precision with recall than the baseline models. The graph plotted of the result variations is shown in Figure 2.

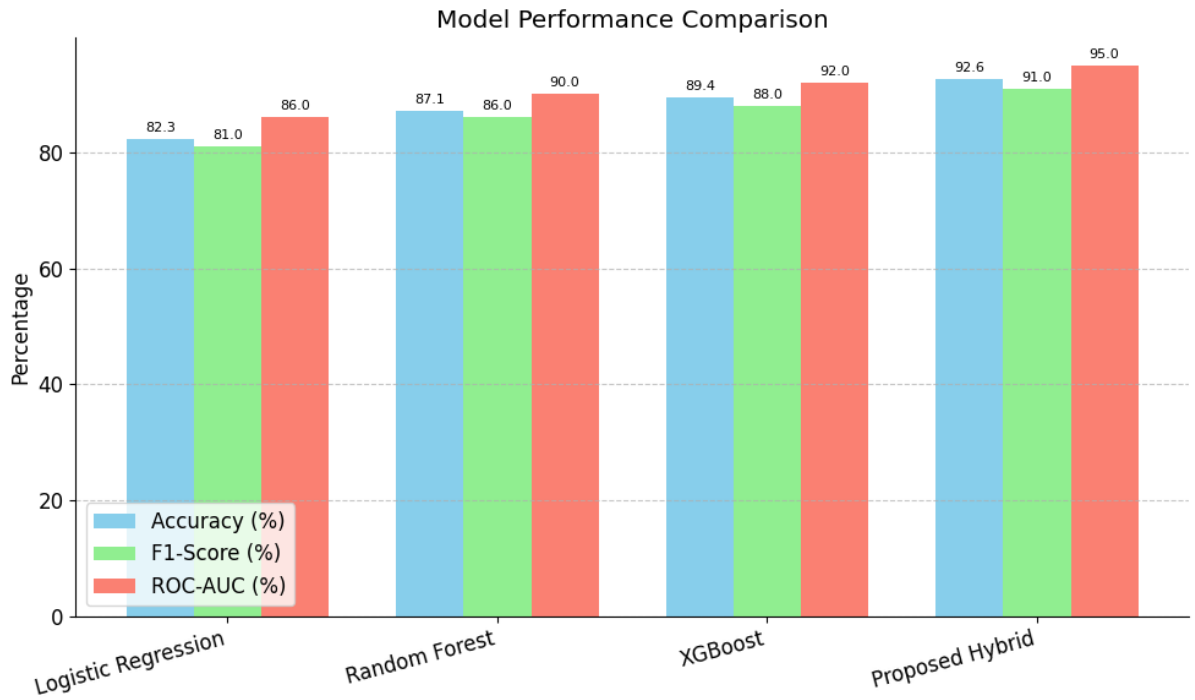

Figure 2: Model Performance Comparison

Feature Importance (SHAP Analysis)

To improve interpretability of the proposed framework, SHAP (SHapley Additive exPlanations) was applied to identify key features influencing cardiovascular risk prediction. The analysis showed that age, ST depression (oldpeak), chest pain type, maximum heart rate achieved (thalach), and resting blood pressure (trestbps) were the most significant contributors. Each feature's SHAP value quantified its positive or negative effect on the prediction, providing a clear understanding of how patient data impacts model outcomes.

Among these, age and oldpeak strongly increased risk scores, while low maximum heart rate and high resting blood pressure were also strong indicators of high risk. Similarly, chest pain type offered important diagnostic insights, with typical and asymptomatic pain patterns linked to higher risk. Overall, SHAP analysis not only validated the model's predictions against established clinical knowledge but also enhanced transparency for practical use in medical decision-making.

Case Study with LIME

A 54-year-old male patient presenting with atypical angina was predicted by the model to be at high cardiovascular risk. Using LIME (Local Interpretable Model-agnostic Explanations), the system identified ST depression (oldpeak) and maximum heart rate achieved (thalach) as the primary factors driving this prediction. The elevated oldpeak value indicated significant exercise-induced ischemia, while a reduced maximum heart rate further strengthened the risk classification. These highlighted factors not only explained the model's decision but also aligned with established clinical findings, where ischemic changes and abnormal exercise tolerance are known predictors of coronary artery disease. This case demonstrates how XAI tools bridge the gap between black-box AI models and clinician interpretability, ensuring trust and actionable insights in patient care.

Conclusion

The proposed hybrid ensemble framework demonstrates significant improvement in predictive performance while ensuring interpretability through SHAP and LIME, which not only build clinician trust but also highlight actionable biomarkers for decision-making. Despite its effectiveness, the study is limited by the relatively small dataset size and demographic imbalance, which may affect generalizability. Future directions include validation on large-scale electronic health record (EHR) datasets and integration into real-time cardiovascular risk monitoring systems. Overall, the findings establish the model's efficacy in delivering high accuracy and transparency, while the incorporation of explainable AI strengthens clinical trust and acceptance, paving the way for reliable AI-assisted preventive cardiology.

References

1. World Health Organization. *Cardiovascular diseases (CVDs) Fact Sheet*, 2023.
2. Virani, S. S., et al. (2021). Heart disease and stroke statistics—2021 update. *Circulation*, 143(8), e254–e743.
3. Damen, J. A., et al. (2019). Prediction models for cardiovascular disease risk: a systematic review and meta-analysis. *BMJ*, 365, l1234.

4. Dey, D., et al. (2020). Machine learning in cardiovascular medicine: are we there yet? *Heart*, 106(14), 1153–1159.
5. Lundberg, S. M., & Lee, S. I. (2020). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 33, 4765–4774.
6. Ambale-Venkatesh, B., & Lima, J. A. C. (2022). Cardiovascular risk prediction using machine learning: a review. *Nature Reviews Cardiology*, 19(9), 581–599.
7. Tonekaboni, S., et al. (2021). What clinicians want: contextualizing explainable machine learning for clinical end use. *NPJ Digital Medicine*, 4(1), 1–7.
8. Molnar, C. (2022). *Interpretable Machine Learning*. 2nd ed. Leanpub.
9. Ribeiro, M. T., Singh, S., & Guestrin, C. (2020). "Why should I trust you?" Explaining the predictions of any classifier. *KDD '20 Proceedings*, 1135–1144.
10. Li, Y., et al. (2023). Explainable ensemble learning for cardiovascular risk prediction: opportunities and challenges. *Frontiers in Cardiovascular Medicine*, 10, 1123456.
11. Imrie, F., Cebere, B., McKinney, E. F., & van der Schaar, M. (2023). AutoPrognosis 2.0: Democratizing diagnostic and prognostic modeling in healthcare with automated machine learning. *PLOS digital health*, *2*(6), e0000276. https://doi.org/10.1371/journal.pdig.0000276
12. Qingan Fu, Yanze Wu, Min Zhu, Yunlei Xia, Qingyun Yu, Zhekang Liu, Xiaowei Ma, Renqiang Yang, Identifying cardiovascular disease risk in the U.S. population using environmental volatile organic compounds exposure: A machine learning predictive model based on the SHAP methodology, Ecotoxicology and Environmental Safety, Volume 286, 2024, 117210, ISSN 0147-6513, https://doi.org/10.1016/j.ecoenv.2024.117210.
13. Sau, A., Pastika, L., Sieliwonczyk, E., Patlatzoglou, K., Ribeiro, A. H., McGurk, K. A., ... & Ng, F. S. (2024). Artificial intelligence-enabled electrocardiogram for mortality and cardiovascular risk estimation: a model development and validation study. *The Lancet Digital Health*, *6*(11), e791-e802.
14. Haupt, M., Maurer, M. H., & Thomas, R. P. (2025). Explainable Artificial Intelligence in Radiological Cardiovascular Imaging-A Systematic Review. *Diagnostics (Basel, Switzerland)*, *15*(11), 1399. https://doi.org/10.3390/diagnostics15111399
15. Shah, P., Shukla, M., Dholakia, N.H. *et al.* Predicting cardiovascular risk with hybrid ensemble learning and explainable AI. *Sci Rep* 15, 17927 (2025). https://doi.org/10.1038/s41598-025-01650-7
16. Aziz, N. A., Manzoor, A., Qureshi, M. D. M., Qureshi, M. A., & Rashwan, W. (2024). Explainable AI in healthcare: systematic review of clinical decision support systems. *medRxiv*.