Explainable Machine Learning Framework for Cardiovascular Disease Diagnosis and Prognosis

Md. Emon Akter Sourov[a], Md. Sabbir Hossen[a], Pabon Shaha*[b], Muhammad Minoar Hossain[b], and Md Sadiq Iqbal[a]
[a]Bangladesh University, Dhaka-1207, Bangladesh.
[b]Mawlana Bhashani Science and Tecnology University, Tangail-1902, Bangladesh.
{emonsourov, sabbirhossen5622, pabonshahacse15, minoarhossain16005}@gmail.com, sadiq.iqbal@bu.edu.bd

***Abstract*—Heart disease continues to pose a critical worldwide health issue, more specifically in areas with insufficient access to healthcare infrastructure and diagnostic systems. Conventional diagnostic approaches often fall short in accurately detecting and managing heart disease risks, resulting in unfavorable outcomes. Machine learning presents a powerful means to boost the precision and reliability of cardiovascular disease prognosis and diagnosis. In this research, we introduced a unified approach incorporating classification techniques for detecting heart disease and regression techniques for forecasting associated risks. The analysis utilized the dataset, named Heart Disease, containing 1,035 instances. To mitigate the problem of data disproportion, the SMOTE was implemented, producing 100,000 additional synthetic samples. Evaluation metrics such as F1-score, recall, precision, accuracy, MAE, RMSE, MSE, and R^2 were adopted to evaluate the performance of the models. Among the classification algorithms, Random Forest delivered the most notable results, attaining an accuracy of 0.972 on actual data and 0.976 on artificially generated data. For prediction modeling, for both synthetic and real samples, linear regression produced the best R2 values of 0.992 and 0.984, respectively, along with the least amount of measurement errors. Furthermore, Explainable AI methods were utilized to improve the comprehensibility of the model outcomes. This paper emphasizes the transformative capabilities of machine learning for diagnosing cardiovascular disease and estimating risk levels, thereby supporting timely interventions and enhancing clinical settings.**

***Index Terms*—Disease Detection, Heart Diseases, Cardiovascular Diseases, Explainable AI, and Machine Learning**

I. INTRODUCTION

The global incidence of heart disease continues to have a startlingly high prevalence. The phrase "Heart Disease Risk" refers to the likelihood of individuals facing complications or negative health outcomes from cardiovascular issues, including coronary artery disease, heart failure, or arrhythmias. These hazards not only endanger personal health and lives but also place immense strain on healthcare infrastructures and national economies [1]. Heart illnesses are progressively prevalent in low-resource environments, where diagnostic capabilities and specialist treatments are often unavailable [2]. As noted by the World Health Organization (WHO) in 2023 [3], Heart disease is a significant public health hazard in Bangladesh, resulting in 273,000 deaths annually. Of these, cardiac attack is the leading cause, responsible for 34% of total national mortality. Risk factors for cardiovascular illness appear in multiple forms, including high cholesterol, diabetes, obesity, hypertension, and lifestyles like unhealthy diets and smoking. These are frequently intensified by deeper contributors like hereditary traits, insufficient preventive care, and limited awareness or delayed intervention [4]. Since these dangers frequently develop covertly, early identification and consistent monitoring are essential [5]. Machine learning has become a potent tool in recent years for identifying a variety of illnesses across multiple domains, such as healthcare and agriculture [6], [7]. It has shown potential in forecasting cardiovascular risk and identifying heart disease in early phases. This provides a novel method for improving cardiac healthcare outcomes [8]. This study presents a machine learning framework combining classification and regression models to evaluate heart disease risk, offering an integrated solution for immediate diagnosis and prognosis. A set of ten classification and eleven regression models, including sophisticated models such as Lasso, Ridge, TabNet, LightGBM, and CatBoost, was utilized. The Synthetic Minority Oversampling Technique (SMOTE) was used to rectify the imbalance in the data. The ML algorithms underwent a thorough evaluation using performance indicators like F1-score, recall, precision, accuracy, MCC, R^2, MSE, RMSE, and MAE for comprehensive assessment. Furthermore, Explainable AI methods were introduced to make the models' forecasts more interpretable and trustworthy for clinical use. The following headings are used to arrange the remaining portions of our study. Section II explores previous literature and summarizes ongoing methods and their drawbacks. Section III details the dataset and methodology used for cardiovascular disease detection and risk analysis utilizing machine learning. Section IV analyzes and interprets the outcomes from the applied methods. The results are condensed in Section V, which also suggests possible directions for future research.

II. LITERATURE REVIEW

To predict the risk of cardiovascular disease, many researchers have proposed various models and strategies through machine learning, reflecting the importance of this area in enhancing clinical outcomes. Despite considerable attention, it remains a growing and actively evolving field. To develop a comprehensive understanding of existing progress and persistent obstacles in diagnosing and predicting heart disease, several prior studies were reviewed and summarized to illustrate the current research landscape. *Rabbi et al.* [9] constructed an ensemble-based approach combining GNB, DT, LR, KNN, SVM, and RF with advanced techniques like stacking, bagging, voting, and boosting. Evaluated on the Cleveland, Indicators of Heart Disease, and Framingham datasets, the bagging ensemble achieved a top accuracy

of 97% on Framingham and Indicators, while the voting ensemble reached 92% on Cleveland. The proposed BEMLA consistently outperformed individual classifiers, offering a robust solution for heart disease prediction. *Ganie et al.* [10] enhanced cardiovascular disease forecasting using voting and stacking ensembles derived from 15 base algorithms trained on a couple of datasets. Six optimal models were combined into meta-ensembles, with stacking achieving the best performance. Statistical tests, including Friedman and Holm's post-hoc, validated the models' superiority. SHAP-based XAI was employed for interpretability, showing how feature contributions impact predictions. *Rohan et al.* [11] conducted an extensive evaluation involving 11 feature selection techniques and 21 classifiers for heart disease prediction. Models included CNN, LSTM, GRU, BiLSTM, RF, SVM, XGBoost, and more. XGBoost attained the most extraordinary performance, with 97% accuracy, 98% sensitivity, and an F1-score of 0.98, outperforming all others across multiple metrics. *Nissa et al.* [12] emphasized fast classification using boosting models like AdaBoost, LightGBM, Gradient Boosting, RF, and CatBoost. AdaBoost achieved the top performance in their study with 95% accuracy, though tuning and evaluation issues were acknowledged that might impact performance generalizability. *Singh et al.* [13] used machine learning methods to forecast the occurrence of CHF utilizing a smaller collection of features, to reduce diagnostic expenses and improve the diagnosis of congestive heart failure. To improve data quality and impute missing data, their method combines KNN with the C4.5 technique for feature optimization and anomaly elimination. The study contrasts DNN with six machine learning algorithms: RF, SVM, NB, DT, LR, and KNN, assessing metrics like F1-score, specificity, and accuracy. The DNN outperformed other methods, recording 95.30% accuracy. *Husnain et al.* [14] showcased the potential of artificial intelligence to forecast heart diseases using methods like neural networks SVM, and RF. The neural network surpassed conventional diagnostic tools, achieving 92% accuracy for high-risk patient identification. *Mienye et al.* [15] introduced a method that integrates SHAP-based interpretability, Bayesian hyperparameter optimization, and robust ensemble techniques. Ensemble models, including AdaBoost, RF, and XGBoost, were evaluated. Their optimized XGBoost achieved notable results on the Cleveland dataset, with 0.971 specificity and 0.989 sensitivity. Nonetheless, the study's strong reliance on Bayesian optimization may not ensure peak results across all data. *Abuhaija et al.* [16] explored seven classifiers, multilayer perceptron, artificial neural network, LR, SVM, DT, KNN, and Naïve Bayes. A correlation-based filter was used to determine key features. Their evaluation based on precision, accuracy, specificity, and sensitivity showed that the J48 decision tree attained the greatest accuracy of 95.76%. *Bhatt et al.* [17] developed a predictive method to lower heart diseases mortality. They used Huang initialization with k-modes clustering and tested models such as RF, DT, XGBoost, and Multilayer Perceptron on a Kaggle dataset of 70,000 entries. The cross-validated MLP model achieved the highest accuracy of 87.28%, outperforming others. *Chandrasekhar et al.* [18] utilized six ML models for heart disease forecasting using the IEEE Dataport and Cleveland datasets. AdaBoost attained 90% accuracy on the IEEEDataport, whereas LR attained 90.16% accuracy on the Cleveland. A soft voting ensemble increases the accuracy to 95% and 93.44%, respectively.

III. MATERIALS & METHODS

This study focuses on detecting heart disease and predicting related risks through both classification and regression methods using the heart disease dataset. The methods used in this study are depicted in Figure 1. The overall methodology covers model training, data preprocessing, and the entire structure of the suggested approach.

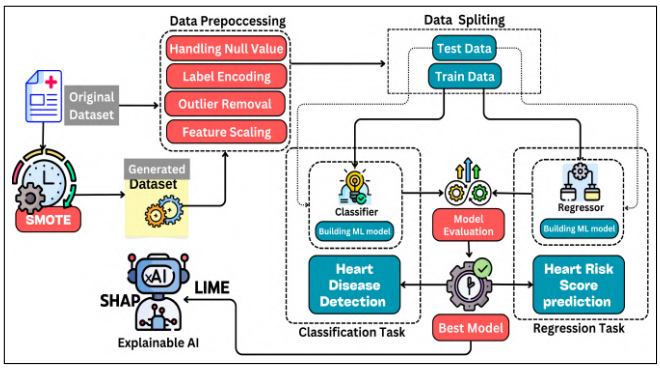


Fig. 1. Suggested Workflow Diagram for Machine Learning-Based Cardiovascular Disease Diagnosis and Prognosis

A. Dataset

The "Heart Disease dataset" utilized in this research was sourced from "Kaggle" [19]. It comprises patient-related information that is essential for evaluating and forecasting the probability of cardiovascular disease. The dataset includes 16 attributes, offering insights such as peak heart rate, cholesterol, blood pressure, sex, age, ECG readings, chest pain occurrence, and other vital features instrumental in assessing cardiac risk. It consists of 1,035 entries in total, 504 representing healthy cases, and 531 indicating cardiovascular illness. The dataset is complete without any null values, and every feature is provided as numerical values.

B. Data Preprocessing

We conducted several preprocessing steps to prepare the dataset effectively. To ensure interoperability with machine learning models, label encoding was used to transform categorical data into numerical form. Although the dataset did not contain any missing values, we defined an imputation strategy for potential gaps: median values for numerical fields and mode values for categorical fields. The Interquartile Range (IQR) approach was used to identify and eliminate outliers in continuous characteristics. Additionally, feature scaling was applied to normalize the dataset, ensuring that all attributes contributed proportionately during model training and no single feature dominated due to its scale.

C. Synthetic Data Generation

To mitigate the imbalance in class distribution, we applied the Synthetic Minority Over-sampling Technique (SMOTE). This method synthesizes new examples by interpolating between samples from the minority class and their nearest neighbors. Using SMOTE, we created an additional 100,000 synthetic instances to balance the dataset better. Subsequently,

we divided both the original and artificially balanced datasets into testing and training sets using a 20:80 ratio.

D. Model Training

This section explains the process used to train both regression and classification models. To generate accurate and consistent forecasts, we initially performed the required preprocessing procedures. SMOTE was applied to address the skew in class distribution, ensuring that the model would not be disproportionately influenced by the dominant class. The training was organized into two main stages: one focused on classification and the other on regression. Each stage was further broken into two scenarios: before and after the application of SMOTE. For classification tasks, we trained multiple classifiers to effectively separate the classes. We utilized both traditional and ensemble methods to boost prediction accuracy and overall model reliability. In the regression phase, several algorithms were used to estimate continuous output values and capture intricate patterns in the data. The classification phase involved ten ML classification models, including Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), Gradient Boosting (GB), K-Nearest Neighbors (KNN), TabNet, CatBoost, Decision Tree (DT), Random Forest (RF), Gaussian Naive Bayes (Gaussian NB), and Support Vector Machine (SVM). For prediction heart disease, we utilized eleven regression models: CatBoost Regressor, LightGBM Regressor, Extreme Gradient Boosting Regressor (XGBR), K-Nearest Neighbors Regressor (KNNR), Decision Tree Regressor (DTR), Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Lasso, Ridge, and Support Vector Regressor (SVR).

E. Performance Parameters

To assess the suggested approach, a comprehensive set of evaluation metrics was employed. Classifiers were measured using balanced accuracy (Acc), accuracy, precision, recall, and F1-score. For regression models, the evaluation relied on metrics such as Matthews Correlation Coefficient (MCC), R-squared (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE), capturing both prediction quality and reliability. The definitions and mathematical expressions for each metric are detailed below.

Accuracy is the percentage of accurately identified outcomes relative to the total predictions, serving as a general metric for classification performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision reflects the classification model's effectiveness in detecting each class by computing the fraction of true positives among predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall, expressed as the ratio of true positives to the sum of false negatives and true positives, assesses the model's capacity to detect positive results.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The F1 Score is a statistic that calculates the harmonic mean of recall and precision to assess the classifier's overall performance.

$$\text{F1 Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{4}$$

The Matthews Correlation Coefficient, or MCC, measures the accuracy of binary classifications by taking into account every component of the confusion matrix to produce a fair measurement even when datasets are unbalanced.

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

The mean squared error (MSE), which represents the total prediction error, is calculated as the average of the squared discrepancies between actual and projected values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

The coefficient of determination, or R2 score, quantifies the percentage of the target variable's variance that the model can account for; values nearer 1 denote better performance.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{7}$$

The Root Mean Squared Error (RMSE), which is the square root of MSE, evaluates the distribution of residuals to provide an interpretable indicator of prediction accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

Mean Absolute Error (MAE) provides information about the average error of the model by calculating the mean absolute difference between actual and forecasted data.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

F. Explainable AI

We used Explainable AI (XAI) methods, such as SHAP and LIME, to make the machine learning classification and regression models more straightforward to understand. LIME produced localized interpretations by perturbing input attributes and observing their influence on predictions. In contrast, SHAP delivered global insights by assigning contribution scores to each feature according to how it affects the model's outputs [20], [21]. These tools helped clarify how various attributes shaped the model's behavior, ensuring interpretability and offering clarity for domain experts in the decision-making process.

IV. RESULT & DISCUSSION

This section presents the outcomes obtained from the applied machine learning methods for cardiovascular risk estimation and disease identification. Both regression and classification algorithms were assessed using real-world and synthetically generated datasets to ensure comprehensive and reliable conclusions. The detailed analysis for each model is presented below, supplemented with Explainable AI visualizations, residual plots, ROC curves, confusion matrices, and tables.

A. Result Analysis

Table I summarizes the performance metrics for different classifiers in identifying cardiovascular disease, where Random Forest demonstrated the highest effectiveness. It attained top scores in F1-score (0.977), recall (0.981), precision (0.974), MCC (0.952), and accuracy (0.976) on the SMOTE data, showcasing its strength in balanced classification tasks. Gradient Boosting and Decision Tree followed closely with 0.970 and 0.974, accuracy scores, respectively. On the other hand, TabNet performed the worst, achieving an accuracy of

TABLE I
CLASSIFICATION MODELS PERFORMANCE FOR DIAGNOSIS OF CARDIOVASCULAR DISEASES BEFORE AND AFTER UTILIZING SMOTE

Classifiers	F1-Score		Recall		Precision		MCC		Accuracy	
	Original	SMOTE	Original	SMOTE	Original	SMOTE	Original	SMOTE	Original	SMOTE
TabNet	0.461	0.474	0.484	0.521	0.441	0.437	-0.204	-0.182	0.404	0.418
Naive Bayes	0.825	0.831	0.842	0.858	0.812	0.806	0.628	0.642	0.814	0.822
CatBoost	0.886	0.883	0.905	0.904	0.869	0.865	0.754	0.760	0.876	0.880
LightGBM	0.903	0.908	0.932	0.925	0.877	0.894	0.792	0.812	0.896	0.906
XGBoost	0.917	0.914	0.942	0.928	0.896	0.902	0.824	0.824	0.910	0.910
Gradient Boosting	0.960	0.972	0.966	0.976	0.955	0.969	0.916	0.942	0.956	0.970
KNN	0.871	0.884	0.881	0.896	0.863	0.874	0.726	0.760	0.862	0.878
Decision Tree	0.970	0.972	0.968	0.983	0.971	0.965	0.934	0.948	0.968	0.974
Random Forest	**0.973**	**0.977**	**0.964**	**0.981**	**0.983**	**0.974**	**0.944**	**0.952**	**0.972**	**0.976**
SVM	0.904	0.897	0.915	0.911	0.896	0.887	0.796	0.790	0.894	0.896

TABLE II
REGRESSION MODELS PERFORMANCE FOR PREDICTING CARDIOVASCULAR DISEASES BEFORE AND AFTER UTILIZING SMOTE

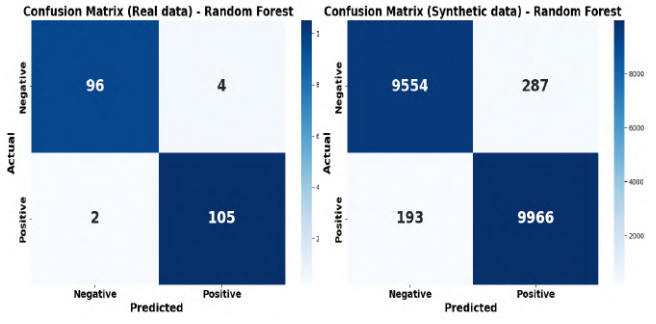
Regressors	MAE		RMSE		MSE		R^2	
	Original	SMOTE	Original	SMOTE	Original	SMOTE	Original	SMOTE
Lasso	1.382	1.400	1.670	1.704	2.793	2.907	0.388	0.354
Ridge	0.062	0.108	0.262	0.314	0.082	0.118	0.982	0.974
Linear Regression	**0.036**	**0.066**	**0.184**	**0.238**	**0.034**	**0.061**	**0.984**	**0.992**
CatBoost	0.858	0.868	1.068	1.078	1.144	1.167	0.750	0.982
LightGBM	0.076	0.098	0.278	0.262	0.087	0.081	0.978	0.982
XGBoost	0.030	0.054	0.264	0.254	0.074	0.073	0.984	0.984
Gradient Boosting	0.120	0.150	0.292	0.308	0.091	0.104	0.978	0.976
KNN	0.484	0.436	0.706	0.656	0.504	0.440	0.888	0.902
Decision Tree	0.030	0.060	0.256	0.274	0.069	0.087	0.986	0.980
Random Forest	0.084	0.102	0.278	0.278	0.083	0.086	0.980	0.980
SVR	0.140	0.152	0.340	0.344	0.118	0.129	0.974	0.972

0.418 on the artificial dataset. The TabNet is susceptible to hyperparameter settings and relatively low effectiveness on tiny structured data, which may be the cause of this underperformance, impairing its capacity to generalize. Overall, the outcomes reinforce the superiority of the Random Forest model and demonstrate how class balancing via SMOTE enhances predictive performance.

Table II outlines the regression model outcomes for heart disease risk estimation. Among all models, Linear Regression yielded the best performance, achieving R2 scores of 0.984 for original data and 0.992 for SMOTE data, as well as the lowest MAE of 0.036 on original and 0.066 on SMOTE data, and MSE values of 0.034 on original and 0.061 on SMOTE data, demonstrating excellent accuracy and minimal prediction error. Models like XGBoost and Random Forest also performed well, with Random Forest upholding consistent R^2 scores of 0.980 on both datasets. On the other hand, CatBoost produced the least favorable results, having the highest error metrics, MAE of 0.858 and ,MSE of 1.144 and an R2 score of 0.750 on actual data. These results highlight the superior capability of Linear Regression in forecasting heart disease risk.

B. Confusion Matrix Representations

Figure 2 shows the Random Forest's confusion matrix both before and after SMOTE was applied. In Figure 2(a), while the model misclassified two positive samples as negative and four negative samples as positive before SMOTE, it accurately predicted 105 positive and 96 negative samples. In contrast, Figure 2(b) illustrates improved performance after SMOTE, accurately identifying 9966 samples as positive and 9554 samples as negative. The model incorrectly classified 287 negatives as positives and 193 positives as negatives. The use of SMOTE enhances prediction balance by reducing false negatives, thus improving sensitivity and overall model effectiveness in recognizing positive cases. This reflects strong classification capability with minimal misclassification.



(a) Real (b) SMOTE

Fig. 2. Confusion Matrices of Random Forest for Heart Disease Diagnosis

C. ROC Curve Representations

Figure 3 shows the Random Forest's ROC curve both before and after SMOTE was applied. The ROC curve displays the False Positive Rate (FPR) versus the True Positive Rate (TPR) across various threshold settings, serving as a tool to evaluate classification performance. The curves for real and synthetic datasets closely align, each achieving an AUC score of 0.99. This elevated AUC suggests how well the model can

differentiate across classes. The slight variation between the curves implies that SMOTE effectively handles the issue of class distribution without degrading the model's predictive quality.

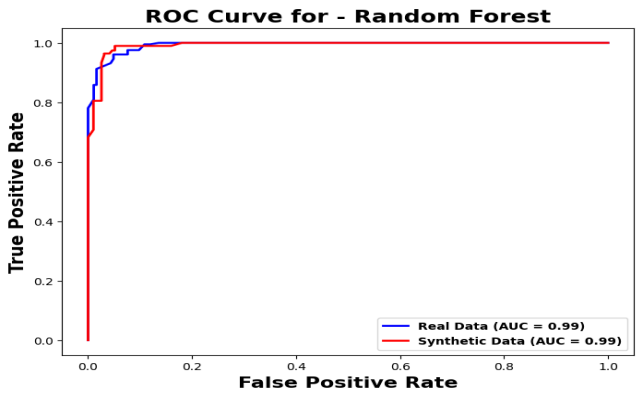
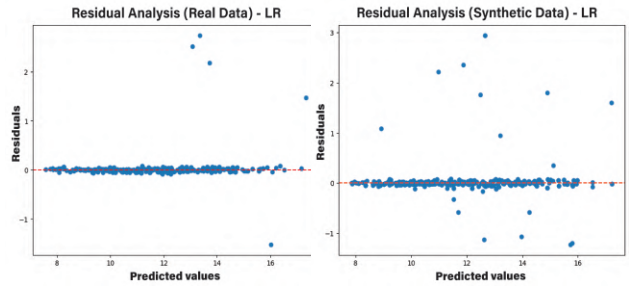


Fig. 3. ROC Curve of Random Forest for Heart Disease Diagnosis on Both Real & Synthetic Data

D. Residual Analysis Representations

Figure 4 presents the Linear Regression model's residual analysis for both Real and SMOTE-generated data. In Figure 4(a), residuals—defined as the differences between predicted and actual values, are plotted against predicted outputs. Similarly, Figure 4(b) displays a scatter plot comparing predicted values with residual differences, where the majority of points lie near zero, signifying minimal prediction bias. The visualizations demonstrate that residuals are mostly concentrated around zero, indicating effective model performance. Nonetheless, a few scattered outliers exist, reflecting instances of higher prediction error.



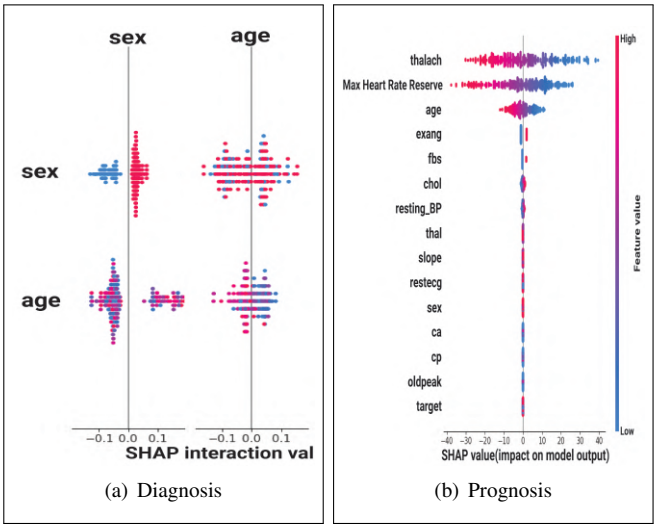
(a) Real (b) SMOTE

Fig. 4. Residual Analysis of Logistic Regression for Heart Disease Prognosis

E. Explainable AI (SHAP and LIME) Representations

Figure 5(a) depicts the SHAP visualization for the RF model after balancing the dataset, showing how significant features like "sex" and "age" influence the model's forecasts in detecting cardiovascular disease. The SHAP interaction values, which show how each feature affects the model's output, are displayed on the horizontal axis. Positive predictions are pushed toward the heart disease positive class, while negative predictions are pushed toward the negative class.

Figure 5(b) illustrates the SHAP plot for the Linear Regressor model applied to the regression task after SMOTE. The figure displays, on the y-axis, the relative relevance of each feature's contribution to the model's predictions. The x-axis displays the SHAP values, indicating both the intensity and direction of feature impact. Features such as age, Max Heart Rate Reserve, and thalach (maximum heart rate) appear as the most significant influential contributors to the regressor's prediction behavior.



(a) Diagnosis (b) Prognosis

Fig. 5. SHAP for Random Forest (Diagnosis) and Logistic Regression (Prognosis) After Balancing the Dataset with SMOTE

Figure 6 shows the Random Forest's LIME summary plot following SMOTE classification, indicating the interpretability of the classifier's forecast for the identification of heart disease. The likelihood of forecasting "Heart Disease" (0.03) and "No Heart Disease" (0.97) for a particular case is displayed. The prediction's primary determinants, together with their corresponding values and contribution weights, are highlighted in the right panel. The model's choice is influenced by the values of several key features, including "thal," "slope," and "Heart Disease Risk Score," which can have a favorable or negative effect.

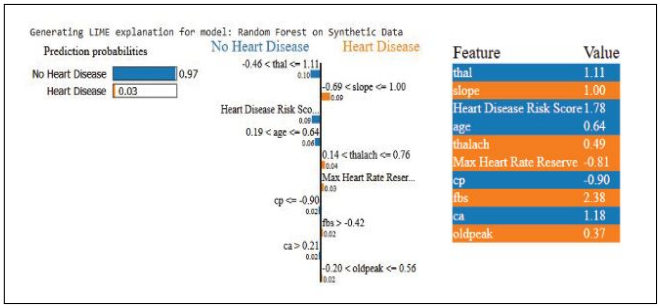


Fig. 6. LIME for Random Forest with SMOTE for Heart Disease Diagnosis

Figure 7 illustrates the LIME of the Linear Regressor model after utilizing SMOTE in the prediction phase. The instance's anticipated value, which is heavily influenced by important characteristics, falls between "No Heart Disease" and "Heart Disease." Features like age, thalach, and Max Heart Rate Reserve are highlighted in the bar along with their positive and negative contributions, demonstrating their importance in influencing the model's result. To put the forecast in context, the feature values are listed in the accompanying table.

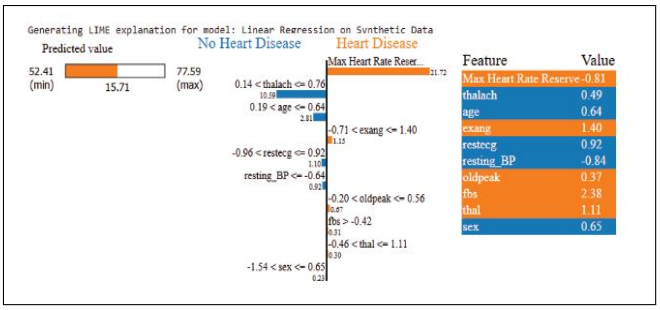


Fig. 7. LIME for LR with SMOTE for Heart Disease Prognosis

F. Discussion & Limitation

Through performance comparison among multiple models and techniques, Random Forest emerged as the most reliable method for cardiovascular disease classification, while Linear Regression proved most effective for risk estimation. The model's strong performance metrics indicate its potential for dependable use in practical applications. We also benchmarked our framework against prior research discussed in the related work section. As reflected in Table III, our model surpasses all previously examined models. Nonetheless, a notable limitation of the proposed framework lies in its reliance on the underlying dataset's quality and variability. Even though SMOTE was utilized to counteract data imbalance, the data may still lack the diversity found in actual clinical conditions, which could restrict the model's capacity to generalize effectively.

TABLE III
COMPARISON WITH STATE-OF-THE-ART

Ref.	Model	Findings
[12]	Adaptive Boosting (AdaBoost)	95% Accuracy
[13]	Deep Neural Network	95.30% Accuracy
[14]	Neural Network	92% AccuracyAccuracy
[16]	Decision Tree	95.76% Accuracy
[17]	Multilayer Perceptron	87.28% Accuracy
Proposed Model	**Random Forest for Diagnosis Logistic Regression for Prognosis**	**97.6% with RF and 99.2% R^2 with LR**

V. CONCLUSION AND FUTURE WORK

This research introduces a dependable and interpretable machine learning approach for identifying and forecasting Heart conditions, achieving an accuracy of 0.976 using Random Forest for detection and a 0.992 R^2 score via Linear Regression for risk prediction. By integrating powerful algorithms with interpretability tools like SHAP and LIME, the approach ensures high accuracy while shedding light on the most influential risk indicators. SMOTE was instrumental in addressing class imbalance, leading to enhanced results on the synthetic dataset and reinforcing the model's practical relevance. This work emphasizes machine learning's role in early detection and prognosis, and the value of explainable AI for trustworthy medical decision-making. Looking ahead, future work may involve enriching the dataset with additional clinical features, conducting long-term studies for progression modeling, and evaluating the model's effectiveness across different demographic and healthcare settings.

REFERENCES

[1] T. Chandola *et al.*, "Work stress and coronary heart disease: what are the mechanisms?," *European heart journal*, vol. 29, no. 5, pp. 640–648, 2008.

[2] J. I. Hoffman, "The global burden of congenital heart disease," *Cardiovascular journal of Africa*, vol. 24, no. 4, pp. 141–145, 2013.

[3] T. Desk, "Experts: One in four adults in bangladesh suffer from hypertension." Dhaka Tribune, 2024. Available at: https://www.dhakatribune.com/bangladesh/health/360195/experts-one-in-four-adults-in-bangladesh-suffer.

[4] F. L. Mondesir *et al.*, "Diabetes, diabetes severity, and coronary heart disease risk equivalence: Reasons for geographic and racial differences in stroke (regards)," *American heart journal*, vol. 181, pp. 43–51, 2016.

[5] A. O. Mocumbi *et al.*, "Challenges on the management of congenital heart disease in developing countries," *International journal of cardiology*, vol. 148, no. 3, pp. 285–288, 2011.

[6] M. S. Hossen *et al.*, "An efficient deep learning framework for brain stroke diagnosis using computed tomography (ct) images," 2025. arXiv, cs.CV, Available at: https://arxiv.org/abs/2507.03558.

[7] M. S. Hossen *et al.*, "A hybrid machine learning approach utilizing cnn feature extraction with traditional classifier to identify strawberry leaf diseases," in *4th International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, 2025.

[8] V. Chang *et al.*, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthcare Analytics*, vol. 2, p. 100016, 2022.

[9] M. S. H. Rabbi *et al.*, "Performance evaluation of optimal ensemble learning approaches with pca and lda-based feature extraction for heart disease prediction," *Biomedical Signal Processing and Control*, vol. 101, p. 107138, 2025.

[10] S. M. Ganie *et al.*, "Ensemble learning with explainable ai for improved heart disease prediction based on multiple datasets," *Scientific reports*, vol. 15, no. 1, p. 13912, 2025.

[11] D. Rohan *et al.*, "An extensive experimental analysis for heart disease prediction using artificial intelligence techniques," *Scientific Reports*, vol. 15, no. 1, p. 6132, 2025.

[12] N. Nissa *et al.*, "A technical comparative heart disease prediction framework using boosting ensemble techniques," *Computation*, vol. 12, no. 1, p. 15, 2024.

[13] A. K. Singh *et al.*, "Predictive analytics of congestive heart failure using deep neural networks: a comparative study," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 2165–2181, 2021.

[14] A. Husnain *et al.*, "Harnessing ai for early detection of cardiovascular diseases: Insights from predictive models using patient data," *International Journal for Multidisciplinary Research*, vol. 6, no. 5, 2024.

[15] I. D. Mienye *et al.*, "Optimized ensemble learning approach with explainable ai for improved heart disease prediction," *Information*, vol. 15, no. 7, p. 394, 2024.

[16] B. Abuhaija *et al.*, "A comprehensive study of machine learning for predicting cardiovascular disease using weka and spss tools," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, p. 1891, 2023.

[17] C. M. Bhatt *et al.*, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023.

[18] N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy through machine learning techniques and optimization," *Processes*, vol. 11, no. 4, p. 1210, 2023.

[19] S. Mahsa, "Heart disease dataset." Kaggle, Available at: https://www.kaggle.com/datasets/snmahsa/heart-disease, 2020.

[20] R. Dwivedi *et al.*, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.

[21] M. S. Hossen *et al.*, "Social media sentiments analysis on the july revolution in bangladesh: A hybrid transformer based machine learning approach," in *Proceedings of the 17th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, IEEE, 2025.