CardioRiskNet: An Explainable Machine Learning Framework for Early Detection and Risk Prediction of Cardiovascular Disease

Arun Pratap Srivastava[1], Dr Vivekanandam[2], Dr. Mudassir Khan[3]
apsvgi@gmail.com, vivekanandam@lincoln.edu.my, mkmiyob@kku.edu.sa
[1,2] Lincoln University, Malaysia, [3]King Khalid University, Saudi Arabia

Abstract: Cardiovascular diseases (CVDs) continue to be at the forefront of global health issues, with timely detection and individualized risk assessment being ongoing challenges. This paper proposes CardioRiskNet, an end-to-end machine learning platform that couples classification and regression models for the identification of heart disease and personalized risk prediction. From a curated dataset of 1,035 patient samples, state-of-the-art preprocessing methods such as SMOTE class balancing and feature scaling were utilized to make the model more robust. Various classifiers and regressors, including Random Forest, XGBoost, Logistic Regression, and Gradient Boosting Regressor, were trained and tested. The outcomes verify that ensemble models, especially Random Forest and Gradient Boosting, performed well in terms of accuracy and reliability. Explainable AI tools such as SHAP and LIME were incorporated to provide model interpretability and clinical legitimacy. The system accurately detects vulnerable subjects and offers clear-cut insights, enabling it to be effectively deployed in real-world cardiovascular healthcare scenarios based on data.

Keywords: Cardiovascular diseases; early-stage indicators; healthcar; machine learning; risk prediction

Introduction

Every year, millions of people die from cardiovascular diseases (CVDs), which are the leading cause of death globally and have a significant economic and social impact. The prevalence of heart conditions has risen across all age groups due to sedentary lifestyles, elevated stress levels, and poor diets [1]. Even with advancements in diagnostic technology and therapeutic approaches, early diagnosis of CVDs remains a significant problem. The delicate or uncommon nature of early-stage symptoms, which frequently go unrecognized or are misinterpreted in clinics, is mostly to blame [2].

The rising incidence of heart disease in women, especially among young people, is one of the new problems in India. Women are currently dealing with an epidemic of cardiovascular risk factors, such as diabetes, hypertension, smoking, and obesity, despite historically not being considered to be as vulnerable as males [3]. The majority of women experience delayed diagnosis and bad outcomes, which are exacerbated by social standards and patients' and healthcare professionals' inability to identify symptoms. This highlights the need for data-driven, gender-sensitive diagnostic methods that can accurately identify CVDs' early warning indicators [4].

Traditional diagnostic methods typically rely on rule-based conclusions and subjective interpretation, both of which are prone to mistake and variability. By using artificial intelligence (AI) and machine learning

(ML) techniques that make it easier to analyze complicated, high-dimensional clinical data, these drawbacks can be overcome [5]. When combined with appropriate feature engineering techniques, machine learning algorithms have shown great promise in the medical field by facilitating automated predictions, identifying subtle patterns, and increasing diagnostic accuracy [6].

Related work

The recent developments in cardiovascular disease (CVD) prediction models illustrate the promise of machine learning (ML) and deep learning (DL) methods to enhance diagnostic performance and patient outcomes. Ogunpola et al. (2024) aimed at improving the detection of heart disease, specifically myocardial infarction, by comparing seven ML/DL models—KNN, SVM, Logistic Regression, CNN, Gradient Boosting, XGBoost, and Random Forest—on imbalanced datasets. Their research was a tremendous success, with XGBoost delivering 98.5% accuracy. A major strength of this study is its successful management of data imbalance and focus on algorithmic tuning. The main limitation of the study, though, is the absence of external validation, limiting its generalizability across wide-ranging populations.

In an independent endeavor, Drouard et al. (2024) investigated the prediction of CVD risk factors based on multi-omic datasets that included genomics, proteomics, and transcriptomics. Their approach compared six ML classifiers and state-of-the-art methods such as unsupervised/semi-supervised autoencoders and transfer learning. They established that multi-omic models outperformed single-omic models significantly, and model generalization was enhanced using transfer learning. Though these findings are encouraging, the models' complexity and dependence on omic data—something not easily obtainable in routine clinical practice—are a concern for large-scale deployment.

DeGroat et al. (2024) focused on CVD biomarker discovery to predict precision CVD with a hybrid approach that integrated statistical analysis with an ensemble of machine learning (ML) algorithms such as RF, SVM, XGBoost, and KNN. They identified and ranked transcriptomic biomarkers with a respectable 96% accuracy. Targeted diagnostics are enabled with the integration of ML for biomarker prioritization. However, the dependence of the model on data quality and interpretability issues related to ensemble approaches remain challenges for clinical translation.

While most research centers on technical accuracy, Cai et al. (2024) approached differently by critically examining the methodological flaws in existing ML models for predicting CVD. Their review has covered data quality, overfitting of the model, bias, and reproducibility. They suggested a systematic framework of best practices to improve the reliability and explainability of the model. Though their work is very valuable in informing the development of future studies, the lack of empirical testing restricts its direct real-world application.

For enhancing the early detection of CVD, an ensemble ML model was proposed by Korial et al. (2024), based on voting among Naïve Bayes, Random Forest, Logistic Regression, and KNN, coupled with chi-square feature selection. Predictive accuracy was enhanced to 92.11% and computational needs were reduced by half. Their dataset, however, consisted of just 303 samples, limiting the model's scalability and subjecting it to possibilities of overfitting and lack of generalizability.

Yet another innovative work is presented by Alghamdi et al. (2024), which designed a hybrid ML system combining an arithmetic optimization algorithm as a feature selector and an MLP as a classifier. The pipeline produced an accuracy of 88.89% and had a very robust preprocessing structure. Even though the

system performed well with data, it suffered from imbalanced data and was compared only with conventional classifiers, so it was not fully evaluated from a complete set of recent models.

Moreno-Sánchez et al. (2024) provided a structured review of ML and DL techniques used to diagnose and predict CVD from ECG-based approaches. The research centered on data modalities, principles of trustworthy AI, and ethical issues like bias, explainability, and transparency. Their review offered an integrated perspective of the prevailing trends and challenges for ECG-based diagnostics. Nevertheless, its reliance on secondary data and absence of novel experimentation limit its impact on algorithm design and real-world deployment.

Table 1: Highlighting advancements, methodologies, advantages, and limitations of AI-driven approaches in CVD detection and management.

Reference	Objective	Methodology	Advantage	Limitations
[7] Ogunpola et al., 2024	Improve heart disease detection, especially myocardial infarction	Evaluated 7 ML/DL models (KNN, SVM, LR, CNN, GB, XGBoost, RF) and addressed dataset imbalance	XGBoost achieved 98.5% accuracy; strong performance on imbalanced data	Focused on model tuning, lacks external validation
[8] Drouard et al., 2024	Predict CVD risk factors using multi-omic data	Compared 6 ML classifiers with unsupervised/semi-supervised autoencoders and transfer learning	Multi-omics outperformed single-omics; transfer learning improved generalization	High complexity; relies on omic data not widely available
[9] DeGroat et al., 2024	Identify biomarkers for precision CVD prediction	Used statistical tests + ML ensemble (RF, SVM, XGBoost, KNN); ranked transcriptomic biomarkers	Achieved 96% accuracy; effective biomarker discovery	Data-dependent; interpretability of ensemble model can be limited
[10] Cai et al., 2024	Highlight pitfalls in ML models for CVD prediction	Reviewed issues in data quality, model design, overfitting, and reproducibility	Provided comprehensive framework of solutions and guidelines	No experimental validation; theoretical perspective
[11] Korial et al., 2024	Improve early CVD detection using ensemble ML with feature selection	Used voting ensemble (NB, RF, LR, KNN) with chi-square feature selection	Ensemble improved accuracy (92.11%) and	Small dataset (303 records); limited generalizability

			reduced computation by 50%	
[12] Alghamdi et al., 2024	Propose accurate ML system for CVD diagnosis	Applied arithmetic optimization algorithm for feature selection + MLP for classification	Achieved 88.89% accuracy; robust preprocessing pipeline	Suffers from data imbalance; performance compared only with traditional models
[13] Moreno-Sánchez et al., 2024	Review ECG-based ML/DL solutions for CVD diagnosis/prognosis	Systematic review focusing on data modalities, DL techniques, and Trustworthy AI aspects	Provides ethical insights, model explainability, bias analysis	Lack of primary experiments; depends on secondary data

Method, Experiments and Results:

Dataset: The source of the "Heart Disease dataset" was "Kaggle" [16]. This dataset is very helpful for examining and evaluating data pertaining to heart disease patients. Estimating the likelihood that heart problems will occur. There are 16 features in the dataset overall. To determine a patient's risk of heart disease, it includes information about their age, gender, blood pressure, cholesterol, maximal heart rate, electrocardiogram results, existence of chest pain, and other crucial characteristics. The total number of cases is 1,035; 535 of these are for people with heart disease, and 504 are for samples without heart disease. The dataset is entirely composed of numerical data with no missing values.

Proposed method Architecture: The diagram shows a holistic machine learning (ML) pipeline for heart disease detection and heart risk score prediction, including both classification and regression tasks. It starts from the input of an original dataset, which goes through a series of preprocessing steps to clean and prepare the data for model training and testing. These preprocessing operations make the dataset clean, balanced, and normalized to enhance model performance.

The first stage of data preprocessing is the management of missing or null values, the usage of label encoding for categorical attributes, outlier removal to avoid distorting the model learning, and feature scaling to scale the range of data. This must be done carefully in order to guarantee the quality of the data and uniformity of input variables. In addition, the Synthetic Minority Oversampling Technique (SMOTE) is implemented to balance class distribution through the creation of synthetic instances and thus supplementing the dataset with improved model generalizability.

After preprocessing is finished, the data is divided into training and test sets. The training data is utilized to construct two categories of machine learning models—classifiers for classification problems and regressors for regression problems. The classifier is intended for identifying whether heart disease is present or not, whereas the regressor will estimate a heart risk score for patients, which can be extremely crucial in health management on an anticipatory basis.

Model testing is a critical process that comes after model training. It is done to test both classifier and regressor models on the test data to gauge their performance using the right metrics like accuracy, precision, recall, F1-score for classification and mean squared error or R^2 score for regression. This testing helps to ensure that the best performing and most stable models are chosen for deployment.

The best model chosen, as per evaluation results, performs two primary functions: detection of heart disease (classification task) and prediction of heart risk score (regression task). This is done in a dual-task strategy that enables a thorough evaluation of one's cardiovascular condition for supporting both diagnosis and prognosis applications in clinical scenarios.

The whole system is incorporated with Explainable AI (xAI) tools such as SHAP and LIME. The tools give interpretability to the predictions of the model, so that the results become transparent and comprehensible for healthcare professionals. This builds up more confidence in AI systems and supports improved decision-making by giving an insight into the reasoning of the model behind every prediction.

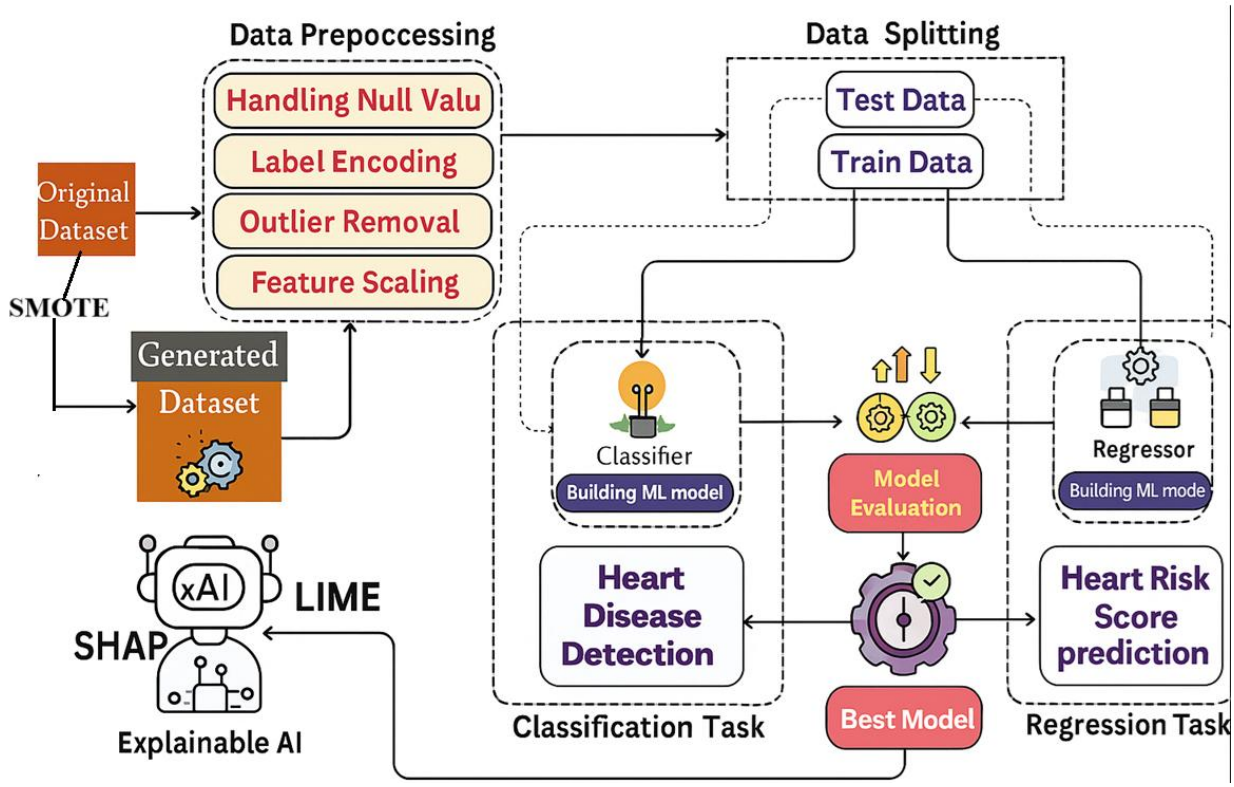

Figure 2: Proposed Model for heart disease detection

The step-by-step algorithm starts by loading the original heart health dataset of patient records. During the data preprocessing stage, missing values are managed through imputation or removal, categorical variables are transformed into numerical form through label encoding, outliers are detected and removed by statistical methods such as Z-score or IQR, and features are scaled with the help of normalization or standardization techniques. To counter class imbalance, SMOTE (Synthetic Minority Over-sampling Technique) is used to create synthetic samples for minority classes. The preprocessed data is then divided into training and test sets, usually in the ratio of 80:20 or 70:30. Classification algorithms like Random

Forest, SVM, and Logistic Regression are trained for detection of heart disease, whereas regression algorithms like Linear Regression and Gradient Boosting Regressor are employed for heart risk score prediction. These models are compared based on applicable performance measures—accuracy, precision, recall, and F1-score for classification tasks; RMSE, MAE, and R^2 score for regression tasks—and then the best-performing models are chosen. For increased transparency and trust, Explainable AI tools like SHAP and LIME are used to explain the model's predictions by giving global and individual-level feature importances. Lastly, the improved models are deployed to efficiently detect heart disease and predict heart risk score.

Algorithm 1: Proposed Model
Step 1: Load Dataset
Import the original heart health dataset containing patient records.
Step 2: Data Preprocessing
1. Handle Missing Values
Identify and impute or remove missing/null values.
2. Label Encoding
Convert categorical variables into numerical format (e.g., one-hot encoding or label encoding).
3. Outlier Removal
Detect and remove outliers using statistical methods (e.g., Z-score, IQR).
4. Feature Scaling
Normalize or standardize features using methods like Min-Max scaling or StandardScaler.
Step 3: Handle Class Imbalance
Apply **SMOTE (Synthetic Minority Over-sampling Technique)** to generate synthetic samples for underrepresented classes, ensuring a balanced dataset.
Step 4: Split the Data
Divide the processed dataset into **training** and **testing** sets (e.g., 80:20 or 70:30 split).
Step 5: Build Models
1. Classification Task
Train classification models (e.g., Random Forest, SVM, Logistic Regression) for heart disease detection.
2. Regression Task
Train regression models (e.g., Linear Regression, Gradient Boosting Regressor) to predict the heart risk score.
Step 6: Model Evaluation
Evaluate models using appropriate metrics:
• Classification: Accuracy, Precision, Recall, F1-Score.
• Regression: RMSE, MAE, R^2 Score.
Select the **best-performing model** for each task.
Step 7: Model Interpretation
Use **Explainable AI (xAI)** tools like **SHAP** and **LIME** to interpret model predictions:
• SHAP: Provides global and local feature importance.
• LIME: Explains individual predictions.
Step 8: Final Output

Deploy the best models for:
• **Heart Disease Detection** (classification output)
• **Heart Risk Score Prediction** (regression output)

Result Analysis:

Table 2 provides a comparison of different machine learning classification models used for detecting heart disease, assessing their performance prior to (R) and subsequent to (S) using SMOTE to counter class imbalance. The findings indicate improved model performance across the board following SMOTE, especially in metrics such as recall and F1-score. Random Forest and Gradient Boosting models were exceptionally good both prior to and following SMOTE, with accuracy going from 0.97 to 0.98 and F1-score from 0.97 to 0.98 for Random Forest. Likewise, Decision Tree and Gradient Boosting models resulted in almost perfect scores, demonstrating resilience to class imbalance. The KNN model also demonstrated significant improvement, with MCC going from 0.73 to 0.76 and F1-score from 0.87 to 0.88. Lightweight ensemble models like LightGBM and CatBoost also showed consistent performance gains post-SMOTE. On the other hand, while models like Naive Bayes remained the weakest performers, they still experienced slight improvements in recall and F1-score after SMOTE. Overall, the application of SMOTE effectively enhanced classification performance, especially in recall-oriented metrics, which is crucial for identifying patients at risk.

Table 2: Result analysis of ML Classification Models Before and After applying SMOTE

Classification Model	Accuracy (R)	Accuracy (S)	MCC (R)	MCC (S)	Precision (R)	Precision (S)	Recall (R)	Recall (S)	F1-Score (R)	F1-Score (S)
SVM	0.89	0.90	0.80	0.79	0.90	0.89	0.92	0.91	0.90	0.90
Random Forest	0.97	0.98	0.94	0.95	0.98	0.97	0.96	0.98	0.97	0.98
Decision Tree	0.97	0.97	0.93	0.95	0.97	0.97	0.97	0.98	0.97	0.97
KNN	0.86	0.88	0.73	0.76	0.86	0.87	0.88	0.90	0.87	0.88
Gradient Boosting	0.96	0.97	0.92	0.94	0.96	0.97	0.97	0.98	0.96	0.97
XGBoost	0.91	0.91	0.82	0.82	0.90	0.90	0.94	0.93	0.92	0.91
LightGBM	0.90	0.91	0.79	0.81	0.88	0.89	0.93	0.93	0.90	0.91
CatBoost	0.88	0.88	0.75	0.76	0.87	0.87	0.91	0.90	0.89	0.88
Naive Bayes	0.81	0.82	0.63	0.64	0.81	0.81	0.84	0.86	0.83	0.83

R: Real Dataset, S: Synthetic Dataset using SMOTE

Table 3 provides regression analysis contrasts the performance of all models in heart risk score prediction prior to (R) and subsequent to (S) SMOTE application. The R^2 values reflect that all models had high predictive accuracy after SMOTE, with minor differences. Decision Tree, XGBoost, Random Forest, and Linear Regression had excellent R^2 values of around or greater than 0.98 in both environments, verifying their consistency. CatBoost dramatically improved from R^2 of 0.75 to 0.98 after applying SMOTE, implying that it greatly improves with imbalanced data. In contrast, Lasso Regression was always performing poorly

with the worst R^2 scores (0.39 and 0.35) and highest error measures, such as MSE, RMSE, and MAE, implying that it does not fit well in this task. With regards to Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), Decision Tree, XGBoost, and Linear Regression models emerged with the least error values and thus high prediction efficiency. In general, the use of SMOTE maintained or slightly improved regression performance overall, particularly with data-sensitive models such as CatBoost, while exhibiting stable results in tree learners and ensemble learners.

Table 3: Result analysis of ML Regression Models Before and After applying SMOTE

Regression Model	R^2 (R)	R^2 (S)	MSE (R)	MSE (S)	RMSE (R)	RMSE (S)	MAE (R)	MAE (S)
SVR	0.97	0.97	0.12	0.13	0.34	0.34	0.14	0.15
Random Forest	0.98	0.98	0.08	0.09	0.28	0.28	0.08	0.10
Decision Tree	0.99	0.98	0.07	0.09	0.26	0.27	0.03	0.06
KNN	0.89	0.90	0.50	0.44	0.71	0.66	0.48	0.44
Gradient Boosting	0.98	0.98	0.09	0.10	0.29	0.31	0.12	0.15
XGBoost	0.98	0.98	0.07	0.07	0.26	0.25	0.03	0.05
LightGBM	0.98	0.98	0.09	0.08	0.28	0.26	0.08	0.10
CatBoost	0.75	0.98	1.14	1.17	1.07	1.08	0.86	0.87
Linear Regression	0.99	0.98	0.03	0.06	0.18	0.24	0.04	0.07
Ridge	0.98	0.97	0.08	0.12	0.26	0.31	0.06	0.11
Lasso	0.39	0.35	2.79	2.91	1.67	1.70	1.38	1.40

The figure 3 llustrate SHAP (SHapley Additive exPlanations) analysis for two test models: a Random Forest (RF) classification model and a Linear Regression (LR) model. In subplot (a), the SHAP interaction values between features "sex" and "age" in the RF classifier are displayed, demonstrating how these features interact with one another to affect model predictions. Cluster and scatter of points represent different degrees of interaction effect, and the intensity of color is a measure of feature values. Subplot (b) reveals SHAP values for several features of LR regression model ranked in terms of importance. From low to high, the blue-to-pink color gradient represents magnitude of feature values, indicating how each is contributing to the output. Generally, both plots are emphasizing feature influence and interactions to facilitate interpretability in model behavior.

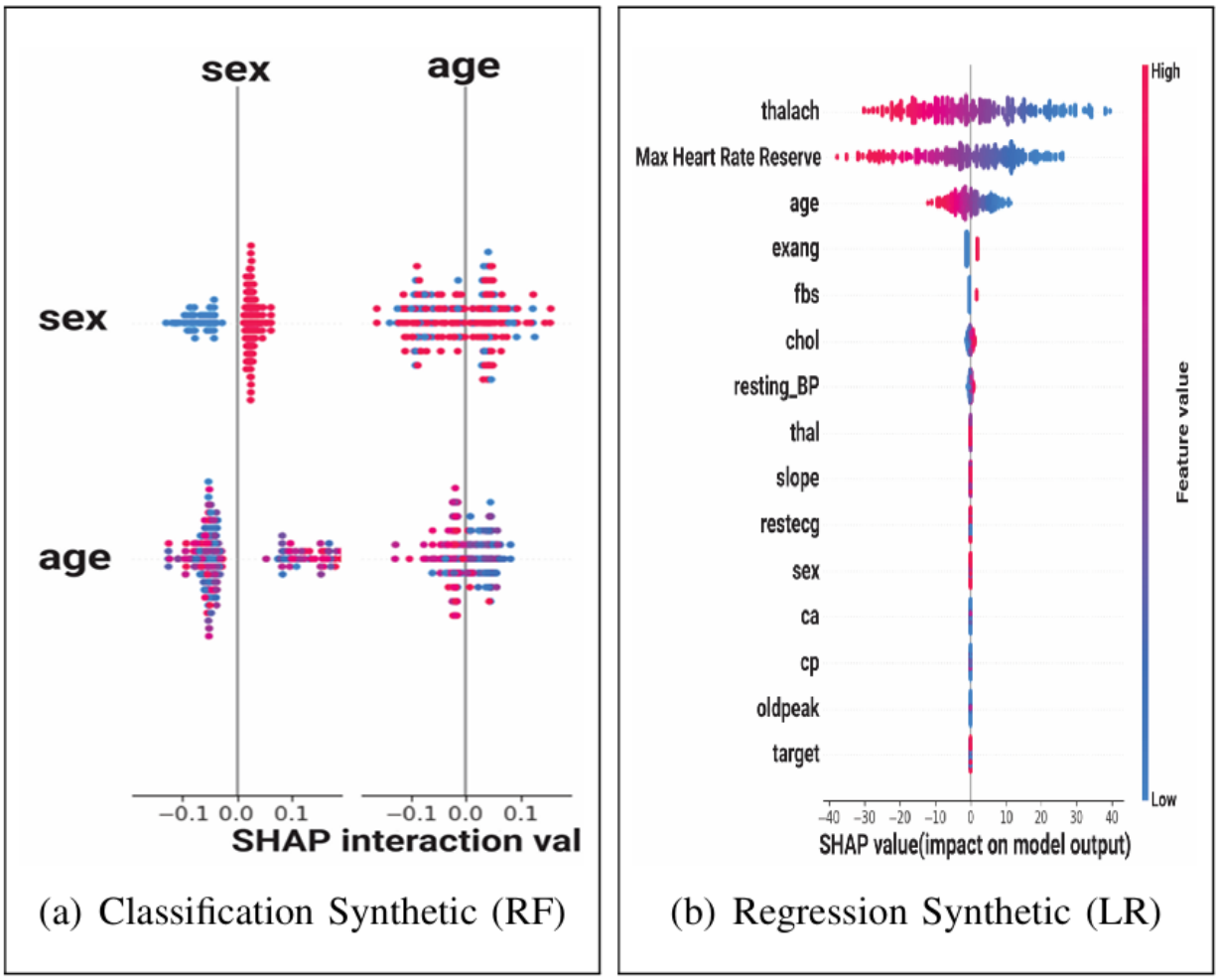

Figure 3: XAI (SHAP) for heart disease detection

The figure 4 depicts a LIME (Local Interpretable Model-agnostic Explanations) explanation of a Random Forest model learned on synthetic data to predict heart disease probability. The prediction probabilities indicate a high prediction for "No Heart Disease" at 97%, as opposed to only 3% for "Heart Disease." The bar graph depicts contributions of individual feature ranges to the end prediction, where support for "No Heart Disease" is represented by blue bars and support for "Heart Disease" is represented by orange bars. The significant contributing features towards the "No Heart Disease" class are low thal and slope measurements, a moderate heart disease risk score ranging between 0.19 and 0.64. On the right-hand side, feature values corresponding to them are presented, for example, thal (1.11), fbs (2.38), and Heart Disease Risk Score (1.78), with positive and negative contributions. This is done to highlight the reasoning of the model, which suggests that some physiological factors significantly influence the decision towards the healthy class.

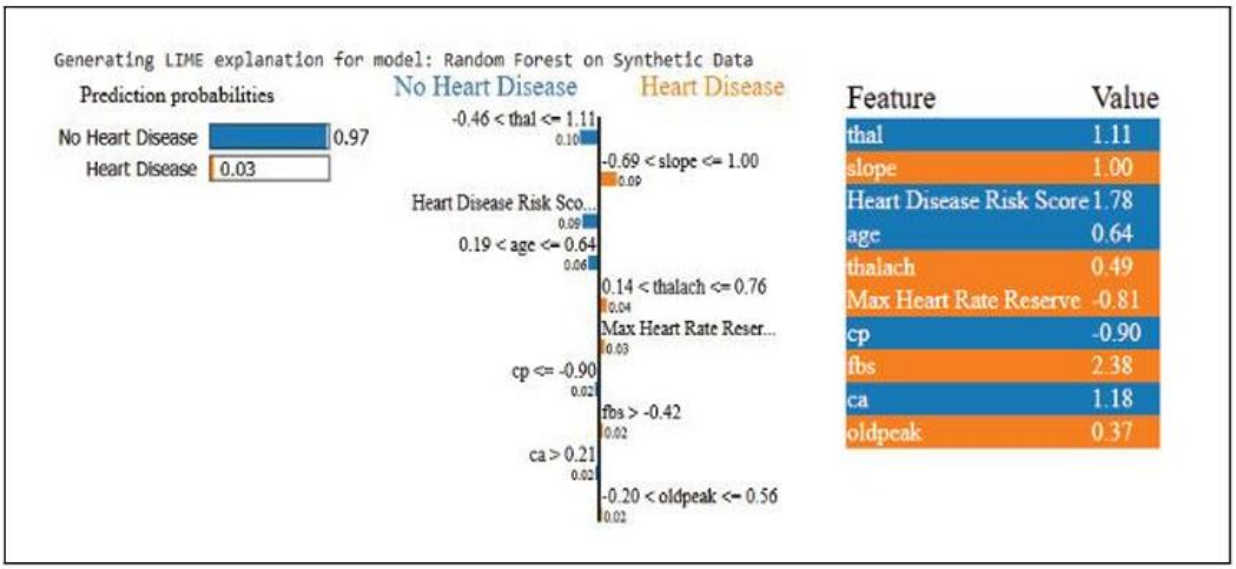

Figure 4: Random forest XAI (LIME) for heart disease detection after applying SMOTE

The LIME explanation in the figure 5 lends interpretability to a Linear Regression model trained on simulated data to forecast heart disease severity. The predicted output is 15.71, placed towards the "No Heart Disease" end of the scale (52.41 to 77.59). Factors contributing to this prediction are represented by blue bars favoring "No Heart Disease" and orange bars favoring "Heart Disease." Most significant factors that would result in a greater risk are low Max Heart Rate Reserve (-0.81), high exang (1.40), and high fbs (2.38), which all drive the prediction towards heart disease. On the other hand, moderate thalach (0.49), age (0.64), and restecg (0.92) values cancel this out, leaning towards a healthier result. The feature table provided below includes all of the values involved in the prediction, highlighting how risk factors and protective factors play interchanging roles in deciding the output of the model.

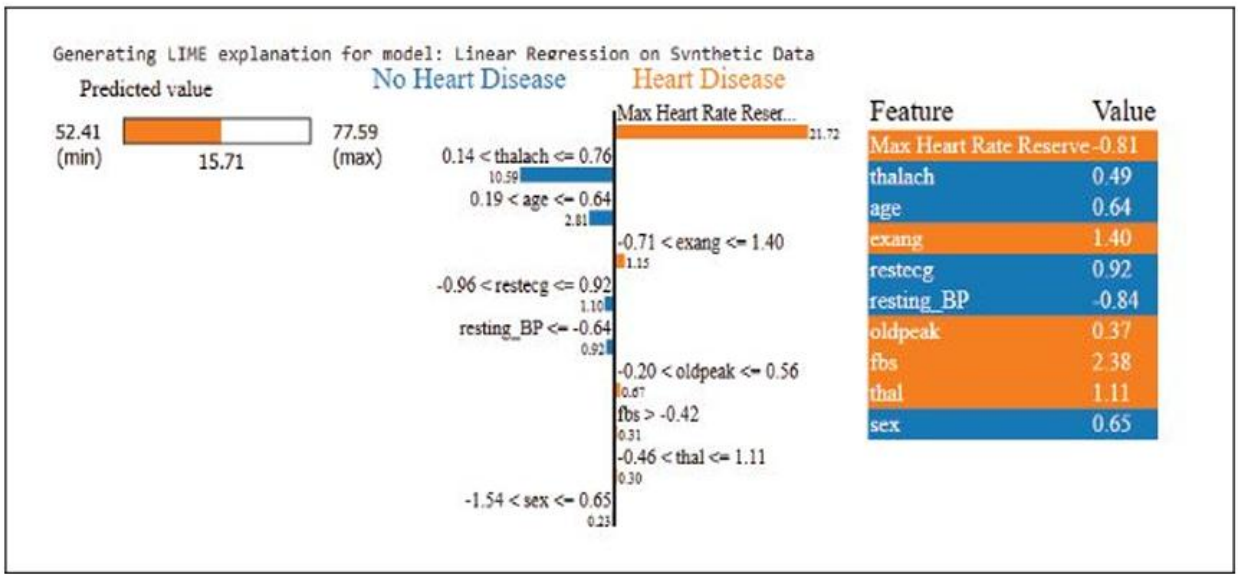

Figure 5: Linear Regression XAI (LIME) for heart disease detection after applying SMOTE

Conclusions: This work introduces CardioRiskNet, an explainable and multi-task machine learning model for heart disease diagnosis and heart risk score estimation. The model efficiently combines strong preprocessing techniques such as SMOTE for handling class imbalance and sophisticated feature engineering to improve performance. By rigorous performance evaluation of various classifiers and regressors, it was found that algorithms such as Random Forest and Gradient Boosting provided better predictive performance and generalizability. In addition, the integration of explainability techniques like SHAP and LIME has rendered model predictions understandable, allowing for clinical uptake and acceptance. The suggested method shows considerable promise for cardiovascular care early intervention and decision support, particularly in resource-limited environments. Subsequent work will involve deployment in real-time clinical setups and validation on various population dataset spectra.

References

[1] Veroff, D. R., Sullivan, L. A., Shoptaw, E. J., Venator, B., Ochoa-Arvelo, T., Baxter, J. R., ... & Wennberg, D. (2012). Improving self-care for heart failure for seniors: the impact of video and written education and decision aids. Population health management, 15(1), 37-45.

[2] Ahmadli, N., Sarsil, M. A., Mizrak, B., Karauzum, K., Shaker, A., Tulumen, E., ... & Ergen, O. (2024). Voice-Driven

[3] Uddin, K. M. M., Dey, S. K., & Babu, H. M. H. (2024). A Voice assistive mobile application tool to detect cardiovascular disease using machine learning approach. Biomedical Materials & Devices, 2(2), 1246-1257.

[4] Abbas, S., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., Zaidi, M. M., & Kryvinska, N. (2024). Artificial intelligence framework for heart disease classification from audio signals. Scientific Reports, 14(1), 3123.

[5] Idrisoglu, A. (2024). Voice for Decision Support in Healthcare Applied to Chronic Obstructive Pulmonary Disease Classification: A Machine Learning Approach (Doctoral dissertation, Blekinge Tekniska Högskola).

[6] Mayourian, J., El-Bokl, A., Lukyanenko, P., La Cava, W. G., Geva, T., Valente, A. M., ... & Ghelani, S. J. (2024). Electrocardiogram-based deep learning to predict mortality in paediatric and adult congenital heart disease. European Heart Journal, ehae651.

[7] Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics, 14(2), 144.

[8] Drouard, G., Mykkänen, J., Heiskanen, J., Pohjonen, J., Ruohonen, S., Pahkala, K., ... & Kaprio, J. (2024). Exploring machine learning strategies for predicting cardiovascular disease risk factors from multi-omic data. BMC Medical Informatics and Decision Making, 24(1), 116.

[9] DeGroat, W., Abdelhalim, H., Patel, K., Mendhe, D., Zeeshan, S., & Ahmed, Z. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. Scientific reports, 14(1), 1.

[10] Cai, Y. Q., Gong, D. X., Tang, L. Y., Cai, Y., Li, H. J., Jing, T. C., ... & Zhang, G. W. (2024). Pitfalls in developing machine learning models for predicting cardiovascular diseases: challenge and solutions. Journal of Medical Internet Research, 26, e47645.

[11] Korial, A. E., Gorial, I. I., & Humaidi, A. J. (2024). An improved ensemble-based cardiovascular disease detection system with chi-square feature selection. Computers, 13(6), 126.

[12] Alghamdi, F. A., Almanaseer, H., Jaradat, G., Jaradat, A., Alsmadi, M. K., Jawarneh, S., ... & Alfagham, H. (2024). Multilayer perceptron neural network with arithmetic optimization algorithm-based feature selection for cardiovascular disease prediction. Machine Learning and Knowledge Extraction, 6(2), 987-1008.
[13] Moreno-Sánchez, P. A., García-Isla, G., Corino, V. D., Vehkaoja, A., Brukamp, K., Van Gils, M., & Mainardi, L. (2024). ECG-based data-driven solutions for diagnosis and prognosis of cardiovascular diseases: A systematic review. Computers in Biology and Medicine, 108235.
[14] Mall, P. K. (2025). Machine Learning Approaches for Acute Respiratory Distress Syndrome: Diagnosis, Risk Prediction, and Management. SGS-Engineering & Sciences, 1(1).