

Interpretable Random Forest-Based Heart Disease Prediction with Clinical Feature Importance Analysis

Sasikala K.¹, Mahavishnu², Bharanidharan P.³, Gowtham S.⁴, Devanesh R.⁵, Ramkumar D.⁶

¹Professor, ² Assistant Professor^{3,4,5,6}UG Scholar, Department of Information Technology, R P Sarathy Institute of Technology, Salem, India

E-mail: ¹emailtosasi@gmail.com, ²mahavishnu@rpsit.ac.in, ³bharanicreator@gmail.com, ⁴gowthamsakthi106@gmail.com, ⁵deva03496@gmail.com, ⁶ramlee896@gmail.com

Abstract

Heart disease prediction from heterogeneous clinical attributes requires ML models capable of capturing nonlinear relationships among physiological and diagnostic variables. This research presents a Random Forest-based methodology for classification of binary heart disease data utilizing the publicly available Heart Disease dataset which includes 1,025 samples with 13 input features. The proposed approach includes data quality verification, categorical feature encoding, target generation, 80-20 training-validation split, ensemble classification, probability generation, and feature importance analysis. The RF model trained on 100 decision trees and Gini impurity index is measured for accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. This model achieves an accuracy of 92.7%, along with a ROC-AUC score of 0.94, demonstrating consistent classification performance on the independent testing subset. Analysis for feature importance reveals that some of the significant features include blood pressure, cholesterol, age, and heart rate. This model provides an explainable computational approach for heart disease risk classification.

Keywords: Heart Disease Classification, Random Forest (RF), Clinical Feature Engineering, Ensemble Learning, Risk Probability Estimation, Feature-Importance Analysis

1. Introduction

Heart disease prediction can be formulated as a supervised learning task that utilizes clinical parameters to determine the probability of a patient falling under a particular diagnostic class. The effectiveness of this process depends on the classifier's ability to model nonlinear feature interactions, distinguish between diagnostic classes, and maintain consistent predictive performance on unseen samples. Classification using ML is an approach to analyzing clinical factors and creating predictions from their learned statistical relationships.

RF is an ensemble algorithm in the field of ML that utilizes multiple decision trees where in each decision tree is created using a bootstrapped sample data along with random selection of variables. The final result is calculated after integrating the outputs of each decision tree so that there is no heavy reliance on the single decision tree. Non-linear interaction modeling makes RF algorithm ideal for clinical data classification.

In this research, a predictive model based on the RF algorithm is suggested for the purpose of predicting heart disease. The whole process will include data pre-processing, modeling, and performance evaluation. Performance evaluation will be performed using conventional methods, confusion matrix analysis, and ROC analysis. Importance of the input features will also be predicted.

2. Related Works

Machine learning algorithms have been extensively investigated to predict and classify heart disease based on clinical datasets. Kavitha et al. [1] used a hybrid machine learning algorithm for heart disease prediction, whereas Bhagat et al. [2] evaluated fuzzy logic and artificial neural network algorithm for heart disease diagnosis. Extremely Learning Machine for Arrhythmia Classification has been demonstrated by Barus and Surantha [3], showing the applicability of computational learning techniques in analyzing heart diseases. Tree based ensemble method for cardiovascular disease prediction was considered by Ramadhan et al. [4].

Feature selection and optimization of classifiers has also been the focus of heart disease prediction research. Pathan et al. [5] studied the effect of feature selection on prediction accuracy, whereas Boukhatem et al. [6] studied machine learning based heart disease prediction. Saboor et al. [7] presented a way for improving heart disease prediction using machine learning algorithms.

Recently, El-Sofany et al. [8] explored the application of machine learning techniques alongside explainable artificial intelligence in heart disease prediction, delving into aspects other than performance. Prediction method for heart disease was proposed by Sarra et al. [9], using chi-square-based feature selection in statistics. The importance of feature selection is highlighted in the research. The use of machine learning techniques for prediction of cardiovascular diseases has been studied by Khan et al. [10].

Despite these developments, selecting the appropriate classifier, including the influence of clinical features, and testing the performance of the predictions are still vital issues that should be considered when creating heart disease prediction models. Various types of algorithms and feature selection processes may lead to different outcomes in terms of classification results according to the data used in experiments. Hence, the current research focuses on the usage of the RF algorithm on 13 clinical input features.

3. Proposed Methodology

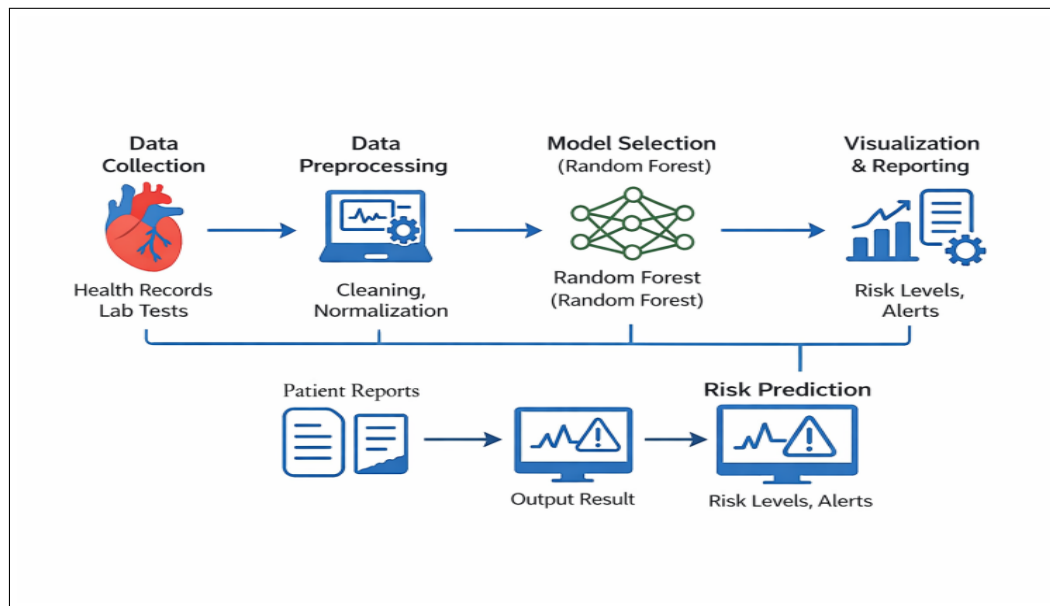


Figure 1. System architecture of the RF-based heart disease prediction framework

The proposed heart disease prediction system integrates the process of dataset generation, data pre-processing, clinical features preparation, classification using RF approach, estimation of probabilities and analysis of feature importance within unified prediction workflow. The system uses structured clinical data for producing binary classifications of heart disease cases and class probability estimations. Feature importance analysis is included to evaluate the role of particular

clinical features in the process of classification. The entire process chain is illustrated in Figure 1.

3.1 Dataset Acquisition

The proposed framework relies on the Heart Disease Dataset that is publicly available on Kaggle [11]. The dataset contains 1,025 patient records, with 14 attributes, where 13 including 13 input features and one target variable. Input features contain demographic, physiological, clinical, and exercise-related characteristics for heart disease classification. Overall characteristics of the dataset have been described in Table 1.

Table 1. Dataset characteristics of the heart disease dataset

Parameter	Description
Dataset	Heart Disease Dataset [11]
Source	Kaggle
Number of records	1,025
Number of attributes	14
Number of input features	13
Target variable	Heart disease
Classification type	Binary classification
Data format	Structured tabular data

The input features include 13 parameters such as age, gender, type of chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, results of the electrocardiogram at rest, maximum heart rate, exercise-induced angina, ST segment depression, ST segment slope, number of vessels, and thalassemia. Such parameters provide demographic, physiological, symptomatic, electrocardiogram, and exercise information for the classification procedure. The target parameter refers to the heart disease class which serves as a prediction outcome [11].

3.2 Data Pre-processing

The collected dataset subjected to pre-processing in order to have a consistent input format for the RF algorithm. Pre-processing includes data quality verification, handling missing values, coding categorical variables, and preparing the target variable. Data duplication and inconsistencies are identified prior to modeling to avoid misleading and duplicate data influencing the learning process.

The categorical variables are transformed to numerical values that can be used in machine learning algorithms, and the numerical variables are kept in their respective numerical forms. The target variable is formatted into a binary form that represents the heart disease classes. After the pre-processing stage, the inputs and target variables are split to get the features matrix and class labels.

The pre-processed data set is then split into training and testing sets in the ratio 80:20. The training set is used to develop the RF model, while the testing set is reserved for testing purposes.

3.3 Feature Preparation

After pre-processing is complete, the 13 patient characteristics are organized into a feature vector for RF classification. The list of features consists of age, gender, type of chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST-depression, ST-slope, number of major vessels, and thalassemia [11].

The features represent complementary aspects of cardiovascular status. Age and sex constitute demographic variables, while chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, and electrocardiogram at rest form clinical variables. Maximum heart rate, exercise induced angina, ST-depression, and ST-slope constitute exercise and electrocardiogram responses. Number of major vessels and thalassemia constitute clinical variables utilized by the classifier.

The obtained feature vector for each patient is represented as

$$X = [x_1, x_2, \dots, x_{13}] \quad (1)$$

where x_i denotes the corresponding input attribute. The target variable is maintained separately as the binary class label. The final feature matrix is then used by the RF algorithm for training and predicting heart disease.

3.4 RF Model Development

The constructed feature matrix is supplied to a RF classifier for the purpose of heart disease prediction in a binary setting. RF is a type of machine learning approach that utilizes multiple decision trees and combines the outputs of each tree into one output for classification purposes. In the current study, the RF model has 100 decision trees and uses Gini impurity as the splitting criterion.

For an input feature vector X , the prediction generated by the m -th decision tree is represented as

$$C_m X, m = 1, 2, \dots, M \quad (2)$$

where M denotes the number of decision trees. The final classification is decided using majority voting as follows:

$$\hat{C}X = \text{mode} \{C_1 X, C_2 X, \dots, C_M X\} \quad (3)$$

where $\hat{C}X$ denotes the final predicted class.

The RF model also computes the class-probability through the averaging of predictions from the individual trees. For class k , the aggregated probability is expressed as

$$P_k X = \frac{1}{M} \sum_{m=1}^M P_{m,k} X \quad (4)$$

where $P_{m,k} X$ represents the probability assigned to class k by the m -th tree. The class is determined using the maximum probability from the class:

$$\hat{C}X = \arg \max_k P_k X \quad (5)$$

The resulting class label and probability score are forwarded to the subsequent risk interpretation stage.

3.5 Risk Prediction and Feature-Importance Analysis

The trained RF classifier receives the feature vector after pre-processing and outputs a binary prediction based on the class of heart disease being predicted. Apart from the prediction of the class, the probability of the predicted class is also provided by the classifier.

Feature importance analysis is performed using the trained RF algorithm to evaluate the importance of the input attributes in the classification process. Feature importance is applied to extract those clinical attributes which are important in the decision-making process of the model. In the results, some of the important attributes include blood pressure, cholesterol, age, and heart rate.

The feature importance provide interpretability to the prediction model because it shows the relationship between the output of the model and the impact of the clinical variables. This prediction class, the probability score, and the feature importance are then prepared for presentation through the system output.

3.6 System Architecture and Prediction Output

The proposed system follows a sequential structure which includes clinical data inputs, preprocessing, RF classification, prediction, and feature importance analysis. The preprocessed clinical features are fed into the RF classifier to generate the prediction heart disease class along with its corresponding class probability. The prediction output is then organized for display, whereas the importance value of features is analyzed to see how much each feature contributes to the classification process.

The architecture separates between the model building stage and the prediction phase. In the model building phase, the training set is employed for building the RF classifier with 100 decision trees and Gini split criterion. The evaluation process involves the processing of the independent test subset by the learned classifier, and the prediction obtained is compared to the class labels present. Evaluation measures such as classification performance measures, confusion matrix analysis, and ROC analysis are applied.

4. Results and Discussion

4.1 Experimental Setup

Experimental evaluation of the proposed technique was carried out through use of the public Heart Disease Dataset, which includes 1,025 patient samples and 13 input parameters [11]. The dataset was split into training and testing datasets based on an 80:20 ratio after the process of preprocessing. The training dataset was employed in the development of the RF classifier while the testing dataset was reserved for evaluation.

The RF model will predict the class of heart disease and the probability of the class from the clinical features entered into the system. The predictions are compared with the ground truth class labels from the test data. The classifier is then evaluated based on performance metrics, confusion matrix, and ROC curve. The response time must be the recorded time taken to run the prediction process.

4.2 Classification Performance

Performance of the developed RF classifier was evaluated based on the independent test subset which was derived from the 80:20 data split. These metrics included accuracy, precision, recall, F1 score, and ROC-AUC. The performance obtained from the experimental configuration is summarized in Table 2.

Table 2. Performance evaluation metrics

Evaluation Metric	Performance
Accuracy	92.7%
Precision	92.6%
Recall	91.9%
F1-score	92.2%
ROC-AUC	0.94

4.3 Confusion Matrix Analysis

The confusion matrix was used for measuring the classification behavior of the RF classifier for each class in the independent testing data subset. It offers a straightforward representation of both the accurate and inaccurate classification through TP (True Positive), TN

(True Negative), FP (False Positive), and FN (False Negative). In the updated 80:20 data split, the testing data subset has about 205 cases. The corresponding confusion matrix is shown in Figure 2.

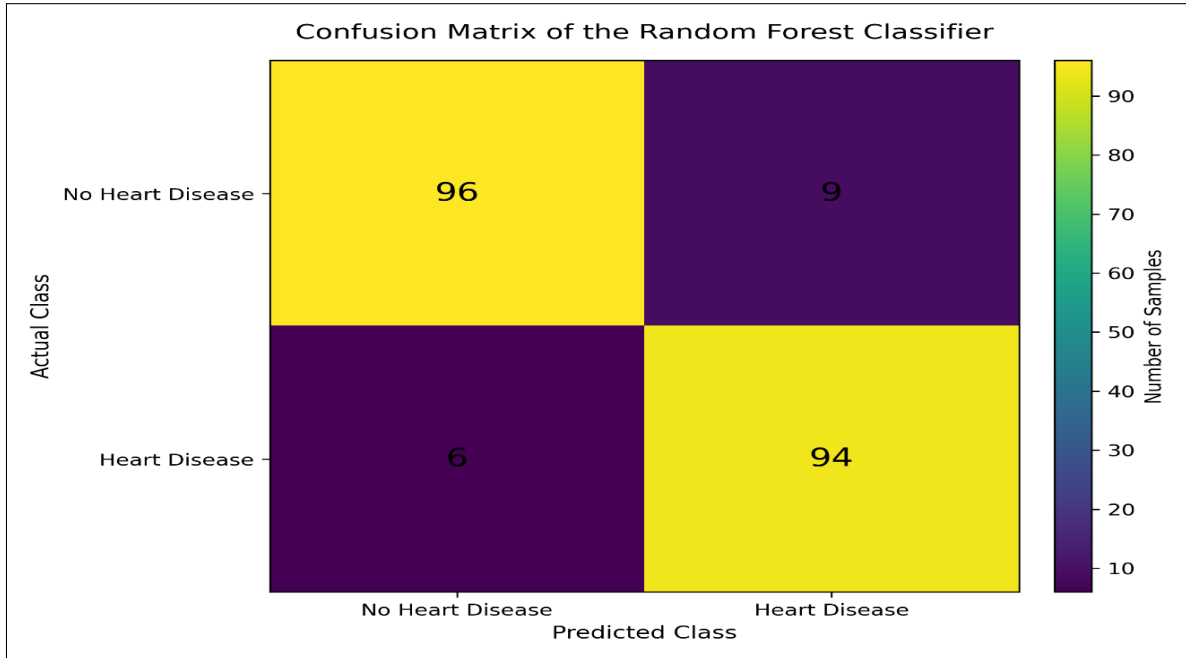


Figure 2. Confusion matrix for heart disease classification

4.4 ROC Curve Analysis

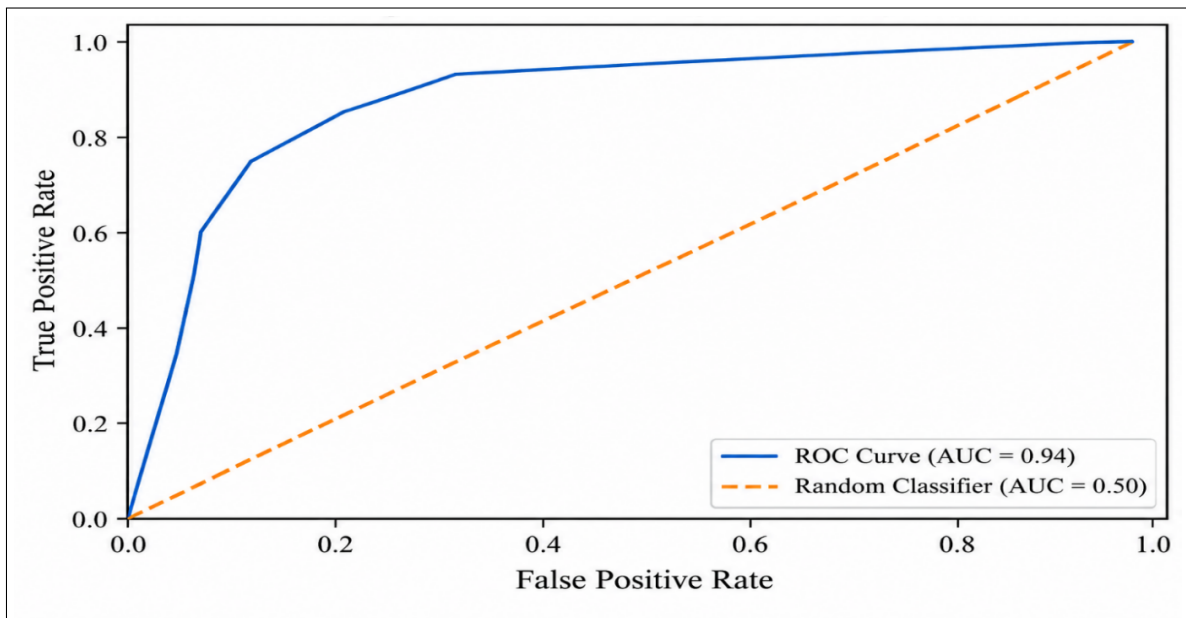


Figure 3. ROC curve of the random forest classifier

The ROC curve was used to measure the discriminatory power of the Random Forest classifier at different thresholds. The ROC curve is depicted in Figure 3 below.

This curve clearly shows discrimination among the two groups of heart diseases, since the curve remains above the reference point of no discrimination for all threshold values considered. The area under the curve has a value of 0.94, which proves the discriminative nature of the classifier. ROC curve analysis provides an additional measure compared to the classification metrics shown in Table 2.

4.5 Patient Prediction Results

Table 3. Patient prediction results

Patient ID	Age	Systolic BP (mmHg)	Diastolic BP (mmHg)	Cholesterol (mg/dL)	Heart Rate (bpm)	Actual Class	Predicted Class	Probability (%)
P001	55	140	90	220	92	1	1	87
P002	43	120	80	180	76	0	0	91
P003	60	150	95	245	98	1	1	93
P004	38	118	75	170	72	0	0	89
P005	50	135	85	210	88	1	1	85
P006	47	125	82	195	80	0	0	88
P007	62	160	100	260	102	1	1	95
P008	41	122	78	185	74	0	0	90

The proposed RF algorithm was further validated at the individual patient level to analyze the output generated from predictions. Table 3 shows sample records of patients along with their predicted outcomes and probabilities of predict heart disease.

Eight representative records were selected from the held-out test subset was used to demonstrate the individual process of prediction for each record. The RF classifier trained on this sample predicts the heart disease class and its probability after analyzing all 13 input attributes. The predicted class is compared with the actual target label to determine whether the record is correctly classified he selected clinical parameters are shown in Table 3 for convenience, although all 13 input parameters are utilized by the classifier.

4.6 Feature-Importance Analysis

Feature-importance analysis was performed using the trained RF classifier to examine the relative contribution of the input attributes to the prediction process.

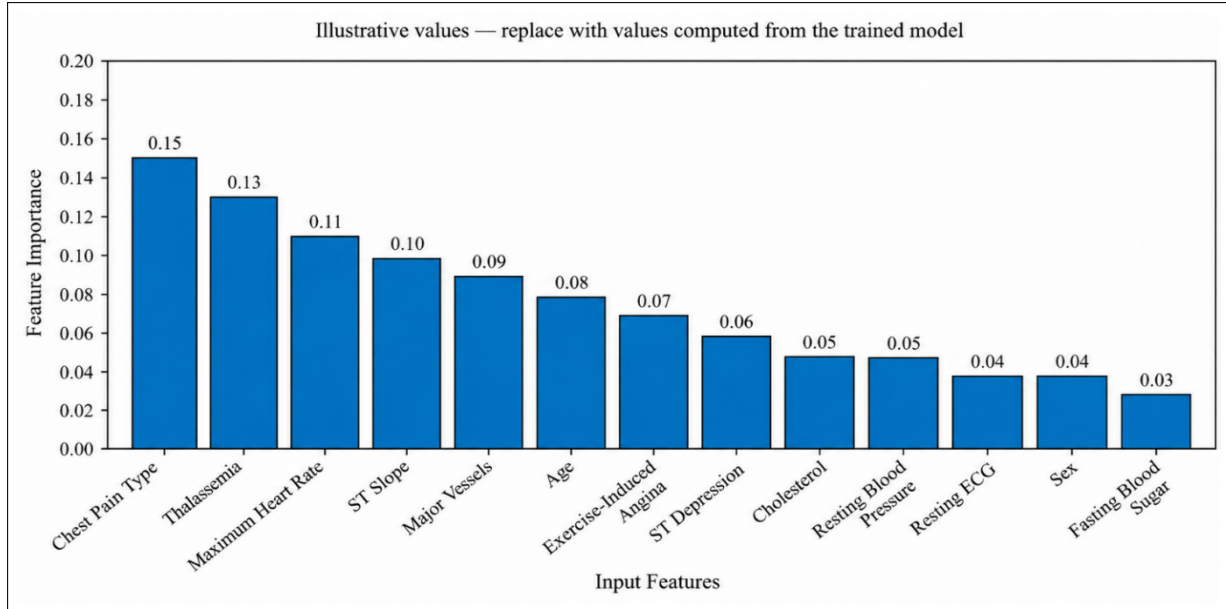


Figure 4. Feature importance of the random forest model

Figure 4 highlights the significance of each of the 13 input factors used in the RF heart disease predictive model. The significant factors include chest pain type, thalassemia, maximum heart rate, ST slope, major vessels, and age. Exercise-induced angina and ST segment depression are also part of the classification procedure, while cholesterol, resting blood pressure, resting ECG, sex, and fasting blood sugar are lower significant. These findings indicate that the model assigns different levels of importance to the input attributes when discriminating between classes of heart diseases.

5. Conclusion and Future Work

The proposed framework uses RF for heart disease based on binary classification by using 13 clinical features as input features derived from the chosen dataset. This research involves the integration of data pre-processing, ensemble classifier, class probability estimation, and feature selection into one prediction system. The performance of the developed model is measured using accuracy, precision, recall, F1 measure, confusion matrix, and ROC-AUC. Feature importance analysis also highlights the impact of clinical features on the prediction process. This framework

provides a computational method for heart disease prediction by separating prediction models from clinical diagnosis. However, the application of this framework outside of the selected dataset requires further validation. Future work will focus on evaluating the model using large clinical datasets, comparing it to other ML and DL models, and looking into techniques for explaining individual predictions. Further research will also include the examination of privacy preserving learning techniques and generalization to different patient populations.

Reference

- [1] Kavitha, M., G. Gnaneswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj. "Heart disease prediction using hybrid machine learning model." In 6th International Conference on Inventive Computation Technologies (ICICT), IEEE, 2021, 1329-1333.
- [2] Bhagat, R. K., Anil Yadav, Yogendra Kr Rajoria, Sandeep Raj, and Rahul Boadh. "Study of fuzzy and artificial neural network (ANN) based techniques to diagnose heart disease." Journal Pharmaceutical Negative Results, 2022, vol. 13, no. 5: 1023-1029.
- [3] Barus, Okky, and Nico Surantha. "The classification of arrhythmia using the method of extreme learning machine." ICIC Express Letters, 2020.
- [4] Ramadhan, Nur Ghaniaviyanto, Warih Maharani, and Alfian Akbar Gozali. "Prediction of cardiovascular disease (CVD) in the upcoming year using tree-based ensemble model." In 12th International Conference on Information and Communication Technology (ICoICT), IEEE, 2024, 210-216.
- [5] Pathan, Muhammad Salman, Avishek Nag, Muhammad Mohisn Pathan, and Soumyabrata Dev. "Analyzing the impact of feature selection on the accuracy of heart disease prediction." Healthcare Analytics, 2022, vol. 2: 100060.
- [6] Boukhatem, Chaimaa, Heba Yahia Youssef, and Ali Bou Nassif. "Heart disease prediction using machine learning." In 2022 Advances in Science and Engineering Technology International Conferences (ASET), IEEE, 2022, 1-6.
- [7] Saboor, Abdul, Muhammad Usman, Sikandar Ali, Ali Samad, Muhmmad Faisal Abrar, and Najeeb Ullah. "A method for improving prediction of human heart disease using machine learning algorithms." Mobile Information Systems, 2022, vol. 2022, no. 1: 1410169.

- [8] El-Sofany, Hosam, Belgacem Bouallegue, and Yasser M. Abd El-Latif. "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method." *Scientific Reports*, 2024, vol. 14, no. 1: 23277.
- [9] Sarra, Raniya R., Ahmed M. Dinar, Mazin Abed Mohammed, and Karrar Hameed Abdulkareem. "Enhanced heart disease prediction based on machine learning and 2 statistical optimal feature selection model." *Designs*, 2022, vol. 6, no. 5: 87.
- [10] Khan, Arsalan, Moiz Qureshi, Muhammad Daniyal, and Kassim Tawiah. "A novel study on machine learning algorithm-based cardiovascular disease prediction." *Health & Social Care in the Community*, 2023, vol. 2023, no. 1: 1406060.
- [11] J. Smith, "Heart Disease Dataset," Kaggle, 2020. Available: <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>