Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Ensemble Machine Learning Framework for Predicting Cardiovascular Disease: A Clinical, Behavioral, and Socio-demographic Insights from Bangladesh

Mohammad Kamrul Hasan

kamrul.ice@nstu.edu.bd

Noakhali Science and Technology University

Ilias Uddin Rifat

Noakhali Science and Technology University

Sanjida Akter

Noakhali Science and Technology University

Shourav Dey

Noakhali Science and Technology University

Nazmun Nahar

Noakhali Science and Technology University

Md Sabbir Ejaz

Noakhali Science and Technology University

Md Ashikur Rahman Khan

Noakhali Science and Technology University

Md Mahbubul Alam

Noakhali Science and Technology University

Research Article

Keywords: Cardiovascular Disease, ML, DL, XAI, SHAP, LIME

Posted Date: May 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-7484162/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable Ensemble Machine Learning Framework for Predicting Cardiovascular Disease: A Clinical, Behavioral, and Socio-demographic Insights from Bangladesh

Mohammad Kamrul Hasan[1*], Ilias Uddin Rifat[1*], Sanjida Akter[1], Shourav Dey[1], Nazmun Nahar[2], Md Sabbir Ejaz[1], Md Ashikur Rahman Khan[1], Md Mahbubul Alam[1]

[1]Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali-3814, Bangladesh.

[2]Institute of Information Technology, Noakhali Science and Technology University, Noakhali-3814, Bangladesh.

*Corresponding Author:

Mohammad Kamrul Hasan (Primary)

Email: kamrul.ice@nstu.edu.bd

Ilias Uddin Rifat

Email: rifat1115@student.nstu.edu.bd

Abstract

Cardiovascular diseases (CVD) continue to be a leading cause of illness and death globally, with increasing prevalence in Bangladesh. Deep learning and machine learning techniques have emerged as highly effective techniques in the medical field, particularly for the prediction of CVDs. The study employs a dataset comprising 1,222 individuals characterized by 19 clinical (already existing data) and demographic features, utilizing sophisticated preprocessing techniques to improve prediction accuracy. We experimented with several algorithms to find the best option for heart disease prediction. These included Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Deep Neural Networks (DNN), and a Stacking Ensemble Classifier. The proposed approach uses a Stacking Ensemble Classifier made up of several base learners and a meta-model. Hyperparameter tuning is used to improve the model, and 5-fold cross-validation is used to test it. The individual classifiers were more than 95% accurate, with SVM and RF being the best. Our stacking model was always better than single classifiers, with an accuracy of 99.18%. SHAP and LIME were used to guarantee interpretability, which is essential for applications in the healthcare industry. The study demonstrates the effectiveness of ensemble learning models in enhancing the predictability and interpretability of cardiovascular disease.

Keywords: Cardiovascular Disease, ML, DL, XAI, SHAP, LIME.

1. Introduction

The World Health Organization (WHO) states that cardiovascular diseases (CVDs) cause one-third of all deaths that occur early in life. In 2017, heart disease was responsible for 14.31% of all deaths in Bangladesh [3]. Every year, CVDs kill about 17.9 million people. Cardiovascular disease is one of the main causes of death around the world, affecting about 26 million people each year. Several cardiovascular diseases primarily affect populations in developing and middle-income countries, including Bangladesh [4, 5].

Cardiovascular disorders encompass all conditions that affect the heart, veins, and the vascular system. These encompass coronary diseases, including myocardial infarctions, HF, cerebrovascular diseases, including strokes, and some cancers [1]. Cardiovascular diseases encompass a broad range of conditions that affect the heart and blood vessels. This includes cardiac disease, characterized by insufficient oxygenated blood flow to the heart, and cerebrovascular illnesses, which interfere with blood flow to the brain. Furthermore, the irreversible damage to the heart chambers may impair the heart's function, as seen in chronic heart failure [2].

Cardiovascular illness causes several symptoms, including chest tightness, pressure, angina, and pain. Individuals may also experience numbness, dyspnea, weakness, or a cold sensation in their limbs if blood flow to those areas is compromised. Additional symptoms may encompass nausea, fatigue, chills, and pain radiating to the neck, elbows, shoulder, jaw, upper abdomen, or back [6]. Tobacco use, drinking too much alcohol, and lack of physical activity have all been found to be significant behavioral risk factors for cardiovascular disease, including heart disease and stroke [7]. Hypertension, obesity, elevated cholesterol levels, familial predisposition, tobacco and alcohol use, along with factors such as gender, smoking, age, dietary habits, diabetes, stress, physical activity, chest pain history, troponin levels, diastolic blood pressure, and sedentary lifestyle, have been identified as critical predictors of cardiovascular disease across diverse age groups [8]. Assessing these risk factors often involves examining a patient's symptoms and identifying how lifestyle choices, occupational stress, and food habits can influence them. These factors are contributing to the growing global incidence of cardiovascular diseases, particularly myocardial infarctions, commonly referred to as heart attacks [9].

The differences in cardiovascular disease death rates between Asian and Western countries may be attributed to variations in life expectancy, population characteristics, genetic and environmental risk factors, as well as the effects of urbanization and lifestyle changes during different stages of development [11]. Patients with heart disease initially exhibit no symptoms but manifest them in its advanced stages, typically when they are unmanageable or untreatable. Consequently, despite the challenges, the early diagnosis and prediction of CVD in low-risk individuals with hypersensitivity are important for prognostic evaluation [12].

An early diagnosis and correct prediction of CVDs will help healthcare practitioners to reduce the burden on individual patients and healthcare facilities and improve the patient outcomes [13]. Recently, Healthcare sector showed interest in machine learning, deep learning, and ensemble techniques that proved to be very effective and powerful approaches. These modern

techniques have shown considerable promise in enhancing CVD risk assessment and prediction [14–16]. These systems have shown a lot of promises for classifying and predicting cardiovascular disease by using a wide range of clinical and demographic characteristics [17,18]. ML and DL techniques have been effectively utilized across various domains, including medical diagnostics for liver disease detection [19], language processing [20], and even sports-related activities such as cricket player selection [21]. In machine learning, ensemble methods combine several models to make predictions that are more reliable and correct [22]. Ensemble learning integrates predictions from multiple models, leading to a final prediction that is more precise and robust [23]. Explainable AI is a promising new field of research aimed at creating transparent and interpretable machine learning models. XAI seeks to explain how the features of a model co-relate with the output. XAI achieves a good balance between accuracy, effectiveness, and explainability [24].

Prior research indicates that ensemble methods, deep learning, and machine learning provide a promising way to predict CVD. These cutting-edge methods make it possible to analyze the complex relationships between several risk factors, such as blood pressure, diabetes, smoking, alcohol use, obesity, and poor eating habits. Compared to conventional statistical methods, these techniques offer a more comprehensive and accurate evaluation of CVD risk by capturing these realtionship. These technologies analyze large and complex datasets, find the patterns, and predict heart disease in individual patients. Several studies have explored machine learning (ML) and deep learning (DL) approaches for predicting cardiovascular disease (CVD). Islam et al. developed a CVD prediction system using Logistic Regression (LR), Decision Tree (DT), and Naïve Bayes on a dataset of 301 individuals with 12 clinical features, achieving the highest accuracy of 86.25% with LR [25]. Studies using LR, DT, Naïve Bayes, and SVM on clinical datasets have consistently identified SVM and LR as strong performers in early diagnosis [26,38]. Other works have explored ANN, RF, and CNN-based methods, with CNN models providing effective results for risk assessment [27,29]. Hybrid statistical ML approaches, such as QDA integrated with ML pipelines, have also been investigated [28].

A comparative analysis involving multiple classifiers, including KNN, RF, Naïve Bayes, AdaBoost, and LR, has demonstrated that ensemble methods are highly effective across diverse datasets [30, 32, 45]. Population-specific research, particularly in Bangladesh, has demonstrated strong predictive capability using both traditional ML and ensemble techniques [34, 45]. Large-scale studies applying DT, XGBoost, RF, and MLP have further validated the robustness of ensemble and neural network approaches in healthcare analytics [35-37]. DL architectures, such as ANN, CNN, LSTM, and DNN, have shown promising results in complex feature learning for CVD detection [39, 43].

Recent trends emphasize ensemble and hybrid strategies that combine boosting, bagging, and stacking with explainable AI tools, such as SHAP, to enhance both predictive performance and clinical interpretability [32, 40, 41, 42, 44]. Numerous previous studies have relied on limited, publicly accessible datasets, such as UCI or Kaggle, despite advancements in predicting cardiovascular illnesses. This limits the generalizability of their findings, particularly in the Bangladeshi context. Many focus only on clinical features, overlooking lifestyle and environmental factors. While DL models often achieve high accuracy, they are computationally

demanding and lack external validation, raising concerns about robustness. Limited interpretability and class imbalance further reduce clinical applicability, and the validity of existing methods across algorithms raises questions about their consistency. Nevertheless, despite the increasing research in this field, we found an evident lack of studies on the prediction of CVD, which is a burden that is progressively increasing. By utilizing machine learning and deep learning models to forecast cardiovascular diseases from a Bangladeshi perspective, this work aims to bridge the gap. The following summarizes this study's main contributions:

- To determine the prevalence of important cardiovascular risk factors in the people of Bangladesh.
- To develop an effective and intelligent model for CVD prediction by integrating these risk factors.
- To improve forecast accuracy, many base models are combined to create an ensemble model.
- To assess each model's performance in relation to the stacking ensemble model.
- To compare the proposed approach with other predictive models and evaluate its performance on actual Bangladeshi datasets.
- To guarantee the interpretability of the model, employ Explainable AI (XAI) techniques.

In this study, we have investigated the main cardiovascular risk factors in a Bangladeshi population and developed an ensemble predictive model which combines cardiovascular risk factors for risk assessment. SHAP and LIME are applied to make sure interpretability, transparency, and clinical reliability of the model, validated on real-world datasets and cross-benchmark with existing methods. Finally, we created an API that analyzes early heart disease risk factors.

2. Methodology

This study aims to develop a cardiovascular disease prediction model using ML and deep learning (DL) techniques. A well-defined methodological process was followed for this, starting from data collection, through data preprocessing, algorithm selection and construction of the model, and performance evaluation. Figure 1: The most basic workflow diagram. It starts with collecting data and deciding on important attributes, then goes on to preprocessing and encoding and eventually splitting the data set into train and test data sets. Multiple classification algorithms are then created and trained, to maximize detection efficiency. The test set is used to evaluate performance, and the top-performing classifier is selected based on key metrics, including F1 score, accuracy, precision, and recall. Additionally, explainable AI techniques, including SHAP and LIME, are applied to ensure transparency, interpretability, and clinical reliability of prediction. Finally, the selected models are employed to predict cardiovascular disease, aiding in the identification of high-risk patients. As noted earlier, early prediction of CVD is vital for clinical intervention.

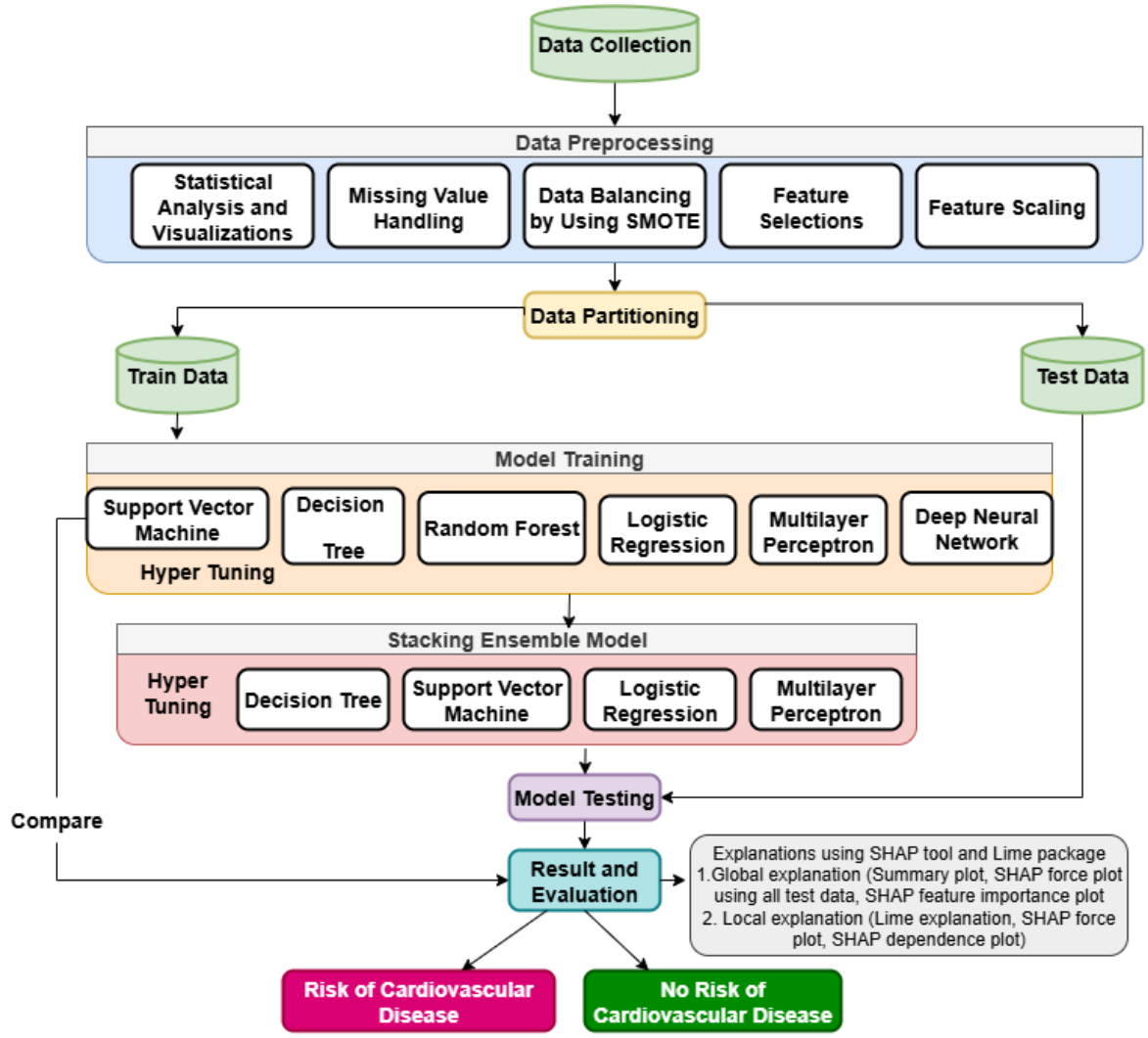


Figure 1: Workflow Diagram.

2.1 Data Collection

This research incorporates both primary and secondary data to meet the study's objectives thoroughly. The survey "Machine Learning Approach for Predicting Cardiovascular Disease in Bangladesh: Evidence from a Cross-Sectional Study in 2023" by Sorif Hossain et al. provides a supplemental dataset for this investigation. The dataset, collected in Bangladesh through a cross-sectional survey, provides information on the risk factors and prevalence of cardiovascular disease. It includes 651 medical records: 391 with cardiovascular disease and 260 without, for comparison. Explicitly collected for this study, focusing on Bangladeshi individuals aged 17 and above, including both patients with and without cardiac conditions. Data were collected from questionnaires at multiple hospitals, including Dhaka Medical College, NICVD, Shaheed Suhrawardy Medical College, Mymensingh Medical College, and Sherpur Sadar Hospital, which represent all regions of Bangladesh. The primary dataset has 415 records of patients with cardiac failure and 156 without, collected from May to November 2024 using random sampling. This dataset is unique and directly related to the research objectives. Table 1 below summarizes the properties of the whole dataset, which includes 30 features necessary for predicting cardiovascular disease.

Table 1: Dataset Description

Questionnaires	Variables	Category (encoding)	Sample (%)
Heart disease	Heart_Disease	No (0) Yes (1)	416(34.04) 806(65.96)
Gender	Gender	Female (0) Male (1)	367(30.04) 855(69.96)
Education Level	Education	No education (0) Primary (1) Secondary (2) Higher secondary (3)	200(16.367) 450(36.82) 346(28.31) 226(18.49)
Division/Province	Division	Dhaka (0) Chattogram (1) Khulna (2) Rajshahi (3) Barisal (4) Sylhet (5) Mymensingh (6) Rangpur (7)	318(27.16) 268(22.88) 79(6.75) 67(5.72) 90(7.69) 68(5.80) 266(22.71) 15(1.28)
Socio-economic status	Eco_Cond	< 20k (0) 20-40k (1) >40k (2)	327(26.76) 663(54.26) 232(18.99)
Residence	Residence	Urban (0) Rural (2)	521(42.64) 701(57.36)
Do you exercise regularly?	Physical_Exercise	No (0) Yes (1)	435(35.60) 787(64.40)
Do you have sound sleep at night?	Sound_Sleep	No (0) Yes (1)	546(44.69) 676(55.31)
Do you take two or more servings of vegetables or fruits per day?	Vegetables_Consume	No (0) Yes (1)	169(13.83) 1053(86.17)
Regularly consume junk food?	Junk_Food	No (0) Yes (1)	423(34.62) 799(65.38)
Do you consume too much salt in your diet?	Salt	No (0) Yes (1)	651(53.27) 571(46.73)
Do you feel bad about yourself?	Feel_Bad	No (0) Yes (1)	320(26.19) 902(78.81)
Don't find pleasure in doing anything?	Feel_zeroInt	No (0) Yes (1)	481(39.36) 741(60.64)
Do you feel hopeless?	Feel_Hopeless	No (0) Yes (1)	274(22.42) 948(77.58)

Do you have a smoking habit?	Smoke	No (0) Yes (1)	582(47.63) 640(52.37)
Do you have a habit of drinking alcohol?	Drink_Alcohol	No (0) Yes (1)	1141(93.37) 81(6.63)
Do you have high blood pressure?	BP	No (0) Yes (1)	553(45.25) 669(54.75)
Do you have high cholesterol in your blood?	Colestrol	No (0) Yes (1)	389(31.83) 833(68.17)
Do you have any family history of heart failure disease?	Heartdisease_family	No (0) Yes (1)	550(45.00) 672(54.00)
Have you experienced anemia?	Animia	No (0) Yes (1)	783(64.08) 439(35.92)
Have any diabetes?	Diabeties	No (0) Yes (1)	477(39.03) 745(60.97)
Do you have a hypertension problem?	Hypertension	No (0) Yes (1)	534(43.69) 688(56.31)
Do you have a sleep apnea problem?	Sleepapnea	No (0) Yes (1)	410(33.55) 812(66.45)
Do you have irregular heart rhythms?	IrregularHeartRhythms	No (0) Yes (1)	565(46.24) 657(53.76)
Do you have coronary artery disease?	CoronaryDisease	No (0) Yes (1)	800(65.47) 422(34.53)
Do you have angina symptoms?	Angina	No (0) Yes (1)	548(44.85) 674(55.15)
Do you have a kidney, lung, or other significant disease?	KidneyLungsDisease	No (0) Yes (1)	1042(83.47) 180(14.73)
Should you take a statin to lower your cholesterol levels?	Take_Colestrol Medication	No (0) Yes (1)	709(58.02) 513(41.98)
Creatinine level	Creatinine	Normal (0) High (1)	413(51.94) 382(48.06)
Sodium level	Sodium	Low (0) Normal (1) High (2)	110(14.36) 398(51.96) 288(37.60)
BMI	BMI	Underweight (0) Normal weight (1) Overweight (2)	69(5.65) 761(62.27) 392(32.07)

2.2 Data Distribution and Visualization

This process involves visualizing the distribution of data and identifying patterns within it. Visualizations help to detect trends, anomalies, and correlations among the features. Histograms display the frequency of numerical data within specified intervals, providing

insights into the data's shape, dispersion, and central tendency. They serve as effective summary measures of the overall distribution. Figure 2 presents a series of histograms for the dataset, where each histogram illustrates the frequency of binary values (0 or 1) for specific attributes, such as junk food consumption, hypertension, and family history of heart disease.

The degree and direction of the relationship between the variables were assessed using correlation analysis; correlation coefficients between -1 and +1 indicate a considerable perfect negative and positive relationship, while values close to 0 indicate little to no association. It is crucial for identifying significant predictors and understanding interdependencies among features. The correlation heatmap for the cardiovascular disease dataset is shown in Figure 3. The color intensity of each hue represents the strength of the link, with purple signifying a negative correlation and yellow a positive correlation. Key findings include a strong positive association between Cholesterol, Feel_Bad, and Feel_Hopeless, as well as between Heart Disease Family History and multiple other health-related variables.

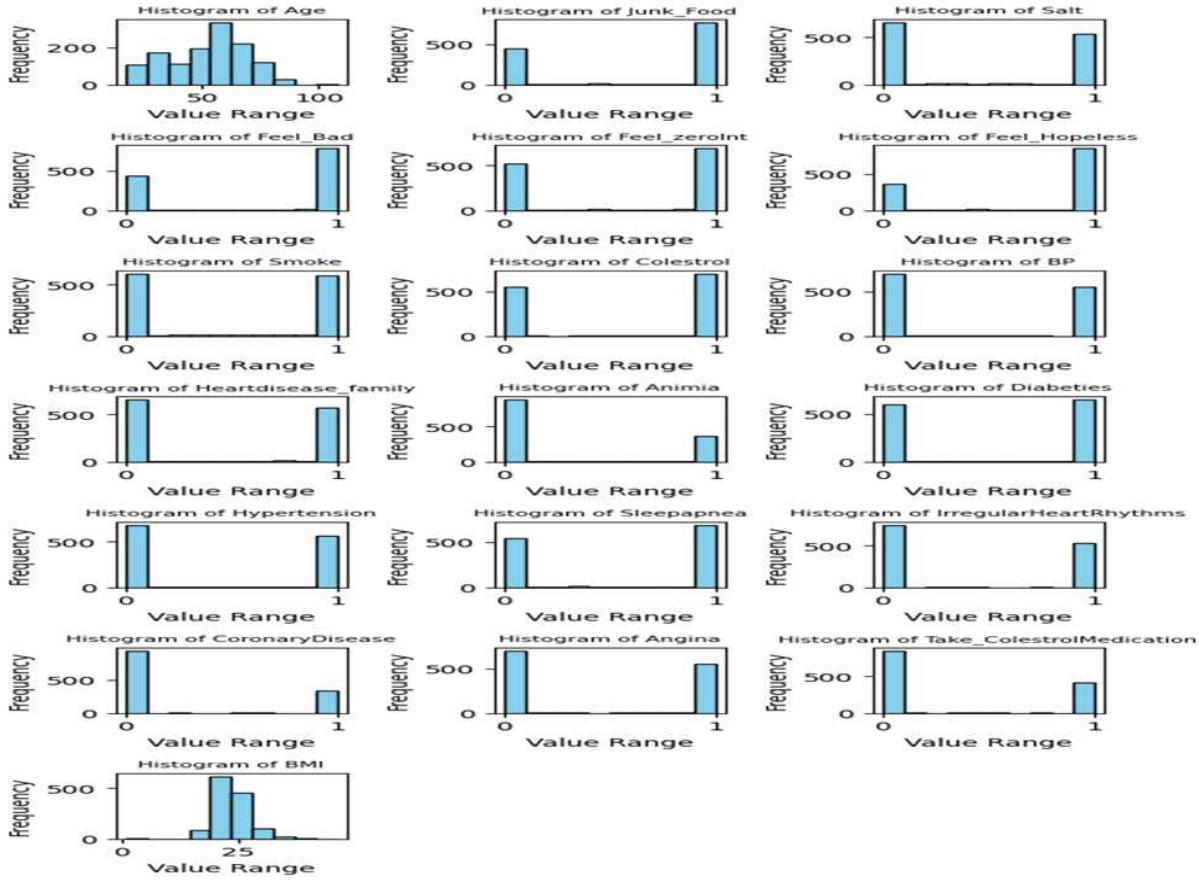


Figure 2: Histogram Analysis of the Dataset

2.3 Data Preprocessing

Data preprocessing is a crucial step in developing a robust prediction model in the healthcare and medical analytics domains. The consistency and caliber of the data can strongly impact the accuracy and efficiency of predicting cardiovascular disease. The following steps made up the preprocessing pipeline in this investigation.

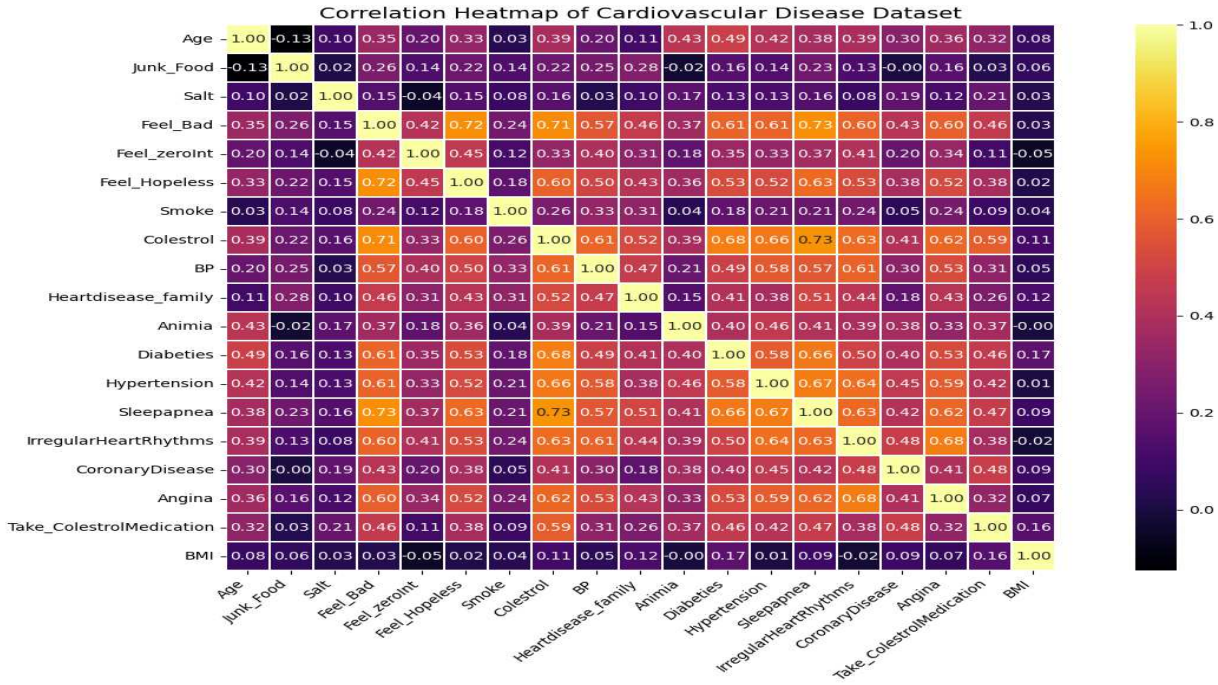


Figure 3: Correlation Heatmap of Dataset

2.3.1 Missing Values Handling

Missing values can severely affect the performance of classification models, as many algorithms either do not support missing entries or yield unreliable results when they are present. Our dataset contained records with missing values for specific variables. To address this, we applied the mean imputation method, replacing each missing value with the mean of its respective feature. This ensured data completeness without introducing bias.

2.3.2 Data Splitting

The dataset was split into two sections at random to assess model generalization: Models are trained using the training set (80%), which identifies patterns in the input variables. Testing Set (20%): Reserved for the final analysis, this set ensures an objective evaluation of the model's functionality. By ensuring that performance measurements accurately reflect the model's capacity to handle unseen data and closely mimic real-world application circumstances, splitting helps to avoid overfitting.

2.3.3 Data Balancing

An imbalanced dataset can bias classification models toward the majority class, reducing predictive accuracy for minority class samples. Initially, our training dataset consisted of 644 samples with cardiovascular disease and 333 without, resulting in a significant class imbalance. To address this, we employed the Synthetic Minority Oversampling Technique (SMOTE), which interpolates existing data to generate synthetic examples for the minority class. This process balanced the dataset and reduced classification bias, enabling the model to learn both classes effectively. Figure 4 illustrates the balance of the target variable.

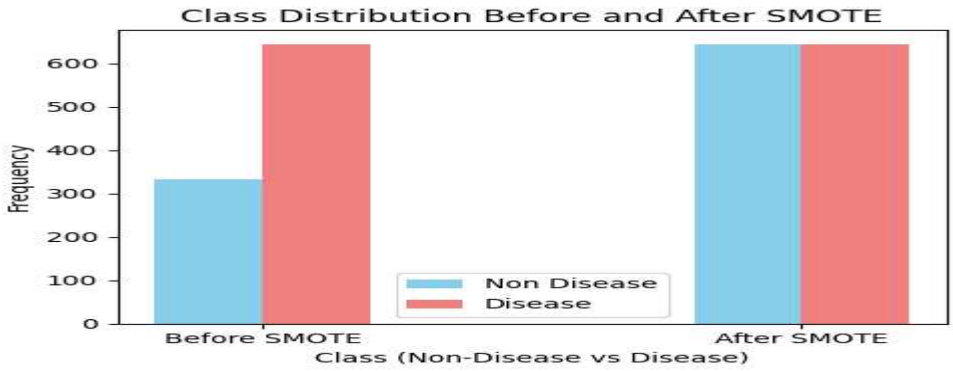


Figure 4: Target Data Balance

2.3.4 Feature Selection

Feature selection is important for making models more accurate, reducing overfitting, and speeding up computations by eliminating of unnecessary or redundant variables. Our primary feature selection technique was Recursive Feature Elimination (RFE). Until the most pertinent predictors are left, RFE iteratively eliminates the least significant characteristics and refines the model. This approach ensured that our final model focused on the features with the highest predictive power for cardiovascular disease, 19 features, thereby improving generalization to unseen data.

2.3.5 Feature Scaling

By preventing features with wider numeric ranges from dominating those with similar ranges, feature scaling guarantees that all variables contribute equally to the model. We used a program called "StandardScaler" to change the features in our dataset that had different scales so that they all had a mean of 0 and a standard deviation of 1. This standardization made the model more stable, made learning faster, and made it less sensitive to differences in feature magnitudes. The equation for standardization is as follows:

$$X_{stand} = \frac{X - mean(X)}{standard\ deviation(X)}$$

2.4 Machine Learning Classification Models

2.4.1 Support Vector Machine

The Support Vector Machine (SVM) is a robust supervised machine learning technique primarily utilized for tasks relating to classification, though it is also applicable to regression problems. It is especially effective for binary classification, leveraging various kernel functions to handle both linearly separable and complex non-linear cases in higher-dimensional spaces. The linear discriminant function in n-dimensional space for pattern recognition is written as: $g(x) = \omega \cdot x + b$, where the equation of the classification hyperplane can be written as $(\omega.x) + b=0$. The discriminant function g(x) was normalized in situations when the data is linearly separable so that every training sample meets the requirement $|g(x)| \geq 1$. This normalization ensures that samples on the classification boundary have $|g(x)| = 1$, and the class interval is

defined as $\frac{2}{||w||}$. Consequently, this interval is equivalent to $||w||$ or $||w||^2$. To achieve correct classification of all samples, the following criteria must be fulfilled:

$$y_i\,[(\omega.x_i)+b]-1\geq 0\ ,\ i=1,2,\ldots..n$$

By solving the above equation and minimizing $\|w\|^2$, we can determine the ideal classification surface. Because they define and support the perfect classification surface, the points that are located on this hyperplane are known as support vectors [4].

2.4.2 Decision Trees

Although they are primarily used in classification settings, decision trees are supervised learning approaches that can be applied to both regression and classification tasks. This classifier uses a hierarchical structure including two categories of nodes to categorize unknown data. Decision nodes, with multiple branches, represent tests or choices based on feature values, while leaf nodes denote the final class labels or outcomes. Tree construction relies on Attribute Selection Measures (ASM), such as information gain and the Gini index, to determine the most suitable attribute for the root node and subsequent splits. The tree is formed by partitioning nodes according to information gain, reflecting the degree to which a feature aids in class differentiation. Attributes exhibiting greater information gain are favored over those with lesser values, and the computation of information is expressed as follows:

Information gain = Entropy(S) - [(Weighted Average) * Entropy (Each feature)].

Entropy=-P(yes) log2 P(yes)-P(no) log2 P(no).

The Gini index measures node purity or impurity, where lower values indicate purer splits. It is calculated as,

Gini Index = 1- $\sum_j^n P_j^2$

Where P_j is the proportion of instances where nodes belong to class j [47], the primary benefits of decision trees are their interpretability; the hierarchical framework facilitates clear visualization and comprehension of the decision-making process.

2.4.3 Random Forest

To handle complex problems and enhance model efficacy, the Random Forest classifier employs ensemble learning, which combines multiple classifiers to improve overall performance. To increase accuracy, it utilizes several decision trees, each trained on a distinct subset of the data. The predictions are then averaged. Rather than depending on a single decision tree, it aggregates the forecasts of several trees and uses majority vote to determine the result. Augmenting the quantity of trees typically enhances accuracy and reduces the likelihood of overfitting. The procedure comprises two primary stages: initially, the creation of a random forest by generating N decision trees, followed by the prediction phase utilizing the trees established in the first stage. Each decision tree utilizes factors such as knowledge gain or the Gini index to pick attributes [48].

Gini = 1 - $\sum_{i=1}^n P_i^2$

The value of P_i represents the probability that an object will fall into a particular category.

2.4.4 Multi-layer Perceptron

Information moves unidirectionally from input nodes to output nodes via one or more hidden layers in this kind of artificial neural network (ANN), much like it does in the human brain. Input, secret, and output are the three primary node kinds that make up this system. It is a perfect, supervised learning classifier for predicting chronic diseases. Multiple interconnected brain units or perceptrons arranged into layers and connected by weighted linkages make up a Multilayer Perceptron (MLP). An MLP diagram is shown in Figure 5. The backpropagation algorithm, which iteratively adjusts the weights to minimize prediction errors, is employed to train the network. Equations [49] demonstrate that the hidden layers typically employ activation functions such as the sigmoid and tanh functions. The equation describes the weight updating procedure for minimizing errors.

Sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$

tanh(x) = 2σ(x) (2x) -1

Mean square error = $\frac{1}{p}\sum_{i=1}^{p}(m_i - \hat{m})2$ Where m_i represents the actual value, $\hat{m}$ Denotes the predicted value, and P signifies the number of samples.

*$w_j = w_j - \alpha(\frac{dE}{dw_j})$

Where E is the error term ($m - \hat{m}$), w_j is the previous weight, * w_j is the updated weight, and α is the learning rate ($0 < \alpha < 1$).

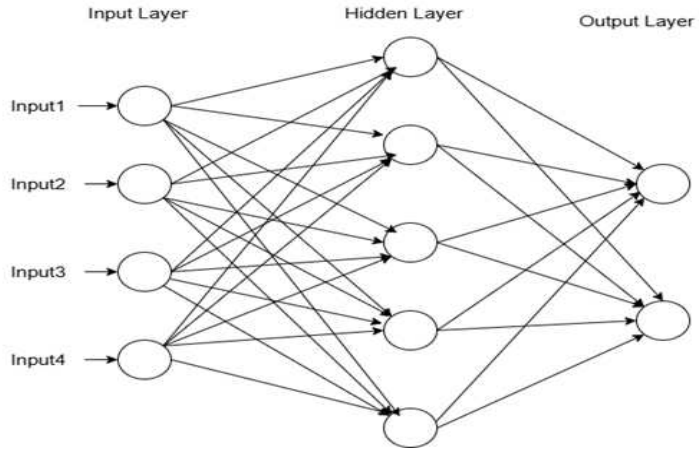


Figure 5: Schematic Diagram of MLP

2.4.5 Deep Neural Network

A "deep neural network" is a type of artificial neural network that has many layers of neurons that are all connected to each other. Because they have many levels, DNNs are good at finding and encoding complex data patterns. A typical DNN structure has an input layer, several hidden layers, and an output layer. Every neuron in a layer gets data from the layer above, uses an activation function, processes the data with weights, and then sends the data to the layer below.

To train DNNs, the backpropagation algorithm changes the weights repeatedly by sending errors back through the network. Optimization techniques like Stochastic Gradient Descent (SGD) are often used to lower the loss function and make the model work better [50]. The following is a mathematical way to show a DNN with L layers. To find the l-th layer's output for an input vector x, do the following:

$$h^{(l)} = f(W^{(l)}h^{(l-1)} + b^{(l)})$$

The activation function, like ReLU or sigmoid, is shown by $f(\cdot)$, the weight matrix is shown by $W^{(l)}$, and the bias vector is shown by $b^{(l)}$. The last output layer makes a prediction, $\hat{y}$, which is a loss function that compares to the real value y. The training strategy seeks to minimize the loss function [51] by progressively adjusting the network parameters.

2.4.6 Stacking Classifier

Stacking is an ensemble method that uses a diverse set of machine learning models to boost the predictive accuracy of your final outputs as well as general performance. Stacking consists of multiple base models to capture different types of data patterns and outputs of base models are fed into meta-model to produce final prediction. This strategy is different from bagging or boosting since those algorithms often use only one type of learner. The base Learners of this MLP, DT, SVM and LR etc. Considering logistic regression is also interpretable and simple, it was simultaneously employed to develop the meta-learner. We stacked heterogeneous base models (Decision Trees, MLP, and Logistic Regression) to learn diverse patterns, and a Logistic Regression meta-model that learned on the outputs of the base models to further improve final predictions. This ensemble helps in improving the generalization, reducing overfitting, and managing complex or imbalanced data. Figure 6 shows how our method's stacking classifiers work. Figure 7 shows the stacking ensemble model that was used in this study again. It combines base classifiers that have been trained separately, like MLP, DT, LR, and SVM. A LR meta-classifier uses their predictions to make the final prediction. This layered method aims to improve accuracy, reduce bias, and make the most of each model's strengths.
Stacking operates in a two-level structure, involving base models that learn complementary aspects of the data, and a meta-model that aggregates the outputs of these base models to enhance generalization. The trained meta-model produces output by extracting and integrating the most valuable insights from the base models, ensuring robust and accurate final predictions. Stacking tends to be useful when there is no a priori best model at the data level and a reasonable estimate of the global model is desired. The resulting ensemble model, known as a stacking model, often outperforms the base models and is therefore a widely used and powerful technique in predictive modeling [52].

Algorithm 1: Proposed Stacking Model

Input:

- *Training set: X_train_scaled and Y_train_res*
- *Testing set: X_test_scaled and Y_test*
- *Base Models: Logistic Regression, Decision Tree, Multilayer Perceptron*
- *Meta Model: Logistic Regression*

Output:

- *Final predicted result Y_pred_stacked*
- *Show, Precision, Recall, F-score, Accuracy, Confusion Matrix, ROC curve*

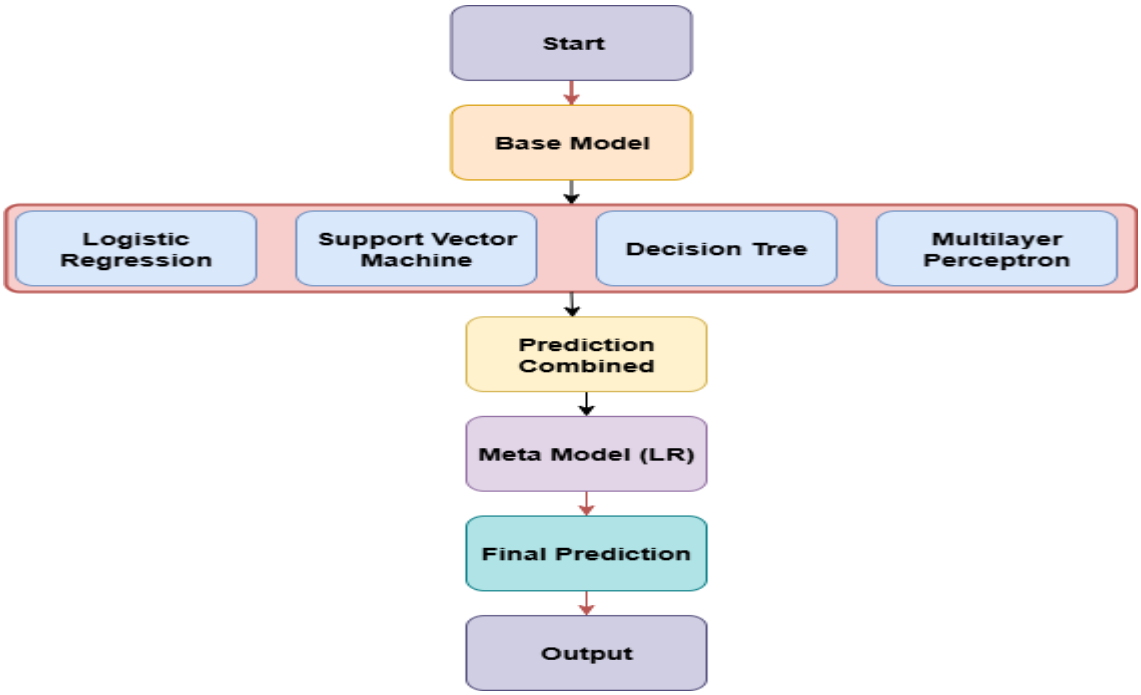


Figure 6: Workflow Diagram of Stacking Ensemble Model

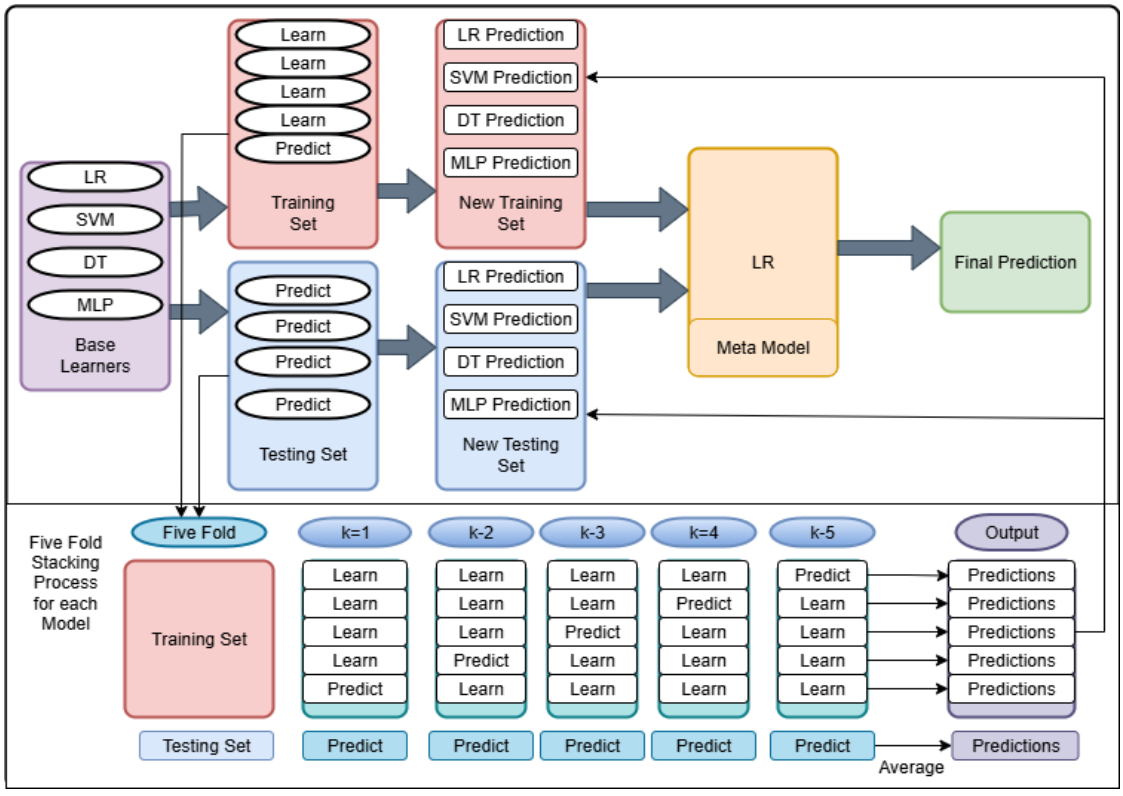


Figure 7: Stacking Model

Steps:

1. Initialize the base models as B {Logistic Regression, Support Vector Machine, Decision Tree, Multilayer Perceptron}

2. Set the meta model as M {Logistic Regression}

3. Build the stacking model using 5-fold cross-validation: Stackmodel StackingClassifier(B, M,cv=5)

4. Train the stacking model using scaled training data: Stackmodel. fit(X_train_scaled, Y_train_res)

5. Predict testing set: Y_pred_stacked=Stackmodel.predict(X_test_scaled)

6. Return the final predicted result by Y_pred_stacked.

2.5 Explainable Artificial Intelligence (XAI)

This study combines two explainable AI methods, SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), to improve interpretability and guarantee clinical dependability [53]. Based on cooperative game theory's Shapley values, SHAP assigns a score to each attribute that indicates its importance to the model's predictions. It offers both local interpretability, which explains how features affect predictions for specific patients, and global interpretability, which demonstrates which features have the most significant influence on predictions across the dataset. SHAP emphasizes in this study how variables like blood pressure, cholesterol, and irregular heart rhythms influence cardiovascular disease outcomes. LIME is a method that works with any model and explains predictions on a case-by-case basis. It works by changing the data around a certain instance and making a simpler local model that tries to guess how the complex classifier will act. This study employs LIME to produce patient-specific explanations, facilitating physicians' comprehension of the rationale underlying each prediction. After the model was created and tested, SHAP was used to look at how each feature contributed, and LIME was used to look at how each case could be understood. These methods made the predictions clearer, understandable, and trustworthy, which allowed doctors to check the model's outputs and use them to make better decisions.

2.6 Model Evaluation

This section includes the metrics employed to assess the effectiveness of the proposed model on predicting early cardiovascular disease. The results, demonstrating very good predictive ability and high generalization to new data, suggest potential for clinical applications. A confusion matrix evaluates how well a classification model has performed by comparing the actual (true) class labels against the class labels that were predicted. The matrix includes TP, FP, TN, and FN, with the rows corresponding to true labels and the columns corresponding to predicted class labels [54]. There are the sequential performance matrices assessed for all the models from confusion matrix below.

i. ***Accuracy*:** It is the number of accurate predictions compared to all predictions, which includes both correct and incorrect ones.

$$Accuracy = \frac{number\ of\ \ correct\ prediction}{Total\ Number\ of\ predictions}$$

ii. ***Precision*:** It is the proportion of false positives (FP) or true positives (TP) among all positive predictions.

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

iii. ***Recall or True Positive Rate*:** It indicates the ratio of actual positive samples in the collection to predicted positive cases, or true positives.

$$Recall = \frac{number\ of\ true\ positives}{Total\ number\ of\ positive\ samples}$$

iv. ***F1-Score*:** A fair assessment of a model's performance in binary classification tasks that considers both false positives and false negatives is provided by the F1-score. It represents the harmonic mean of precision and recall. When weighing the trade-off between accuracy and recall, the F1-score is extremely helpful, especially when the costs of false positives and false negatives are not equal.

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

v. ***Area Under the Receiver Operating Characteristic Curve:*** The model's capacity to differentiate between classes—usually positive and negative—across a range of classification thresholds is assessed using the AUC-ROC curve. A higher AUC value indicates better categorization performance.

3. Results

The results section describes the performance of each machine learning model in relation to the matrices mentioned earlier. Each model was evaluated using the metrics of precision, recall, training accuracy, and testing accuracy, and the ROC-AUC is described in the comparison section.

3.1 Performance Analysis of ML Models

3.1.1 Support Vector Machine

The SVM was trained on 1,288 balanced samples and evaluated on 245 test samples using the train-test split method. As shown in Figure 8, it achieved 98.21% training accuracy and 97.96% testing accuracy. The confusion matrix shown in Figure 9 was used to assess the model's performance. When the model's output matches the actual target values, the yellow and green cells in this matrix indicate correct predictions. In contrast, the purple cells represent areas where the predictions diverged from the actual values. Figure 9 also demonstrates that the SVM model accurately predicted 82 instances where cardiovascular disease (CVD) was absent and made a single incorrect prediction. It also correctly identified 158 cases of CVD, with only four misclassifications. Five misclassifications and 240 accurate predictions resulted from this. Consequently, the model's overall accuracy was determined to be 97.96%, as illustrated in Figure 10.

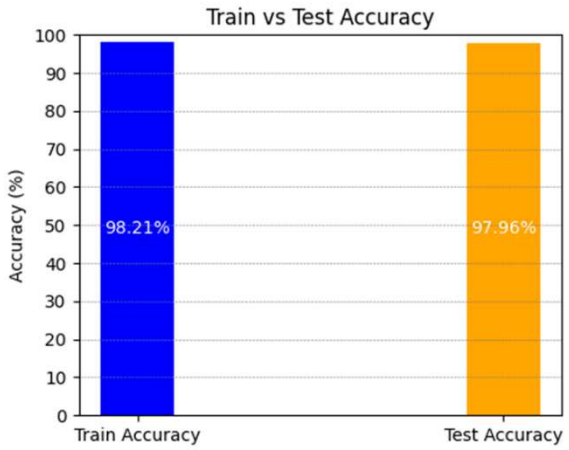


Figure 8: Train and Test Accuracy of SVM

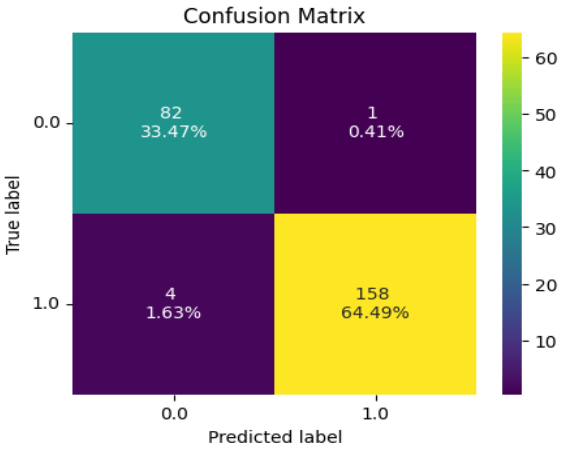


Figure 9: Confusion Matrix of SVM

The model showed a 97.53% recall and a 98.37% precision. The model achieved an effective balance between precision and recall, as evidenced by an F1 score of 98.44%, as shown in Figure 10. This implies that it can successfully detect actual positive cases and produce incredibly accurate forecasts. The SVM's ROC curve result is displayed in Figure 11. According to the evaluation, the SVM model achieves a 0.99 ROC-AUC score, indicating an excellent classification model. This value implies that there is an ideal distinction between the two categories that the model can make. The high AUC is indicative of both the model's strength and reliability regarding the specified task.

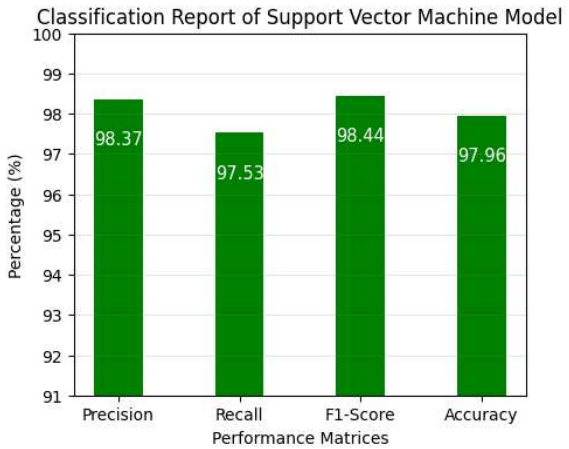


Figure 10: Classification Report of SVM

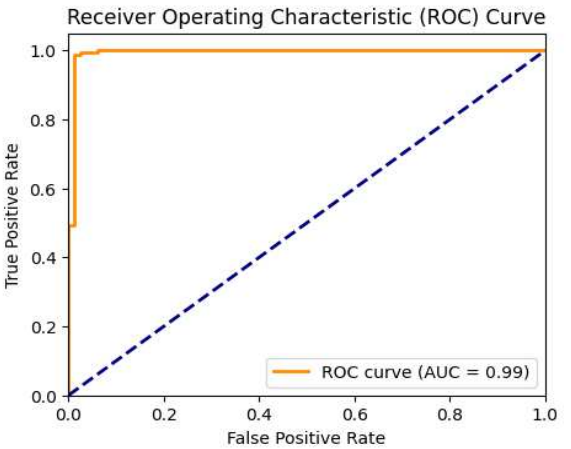


Figure 11: ROC Curve of SVM

3.1.2 Decision Tree

Figure 12. The chart shows the train accuracy is 95.65%, while the test accuracy is 95.10%. The slight difference between the two indicates that the model has good generalization and no significant overfitting issue; both have close performance results. Figure 13 shows that the DT model accurately predicted 78 cases without cardiovascular disease (CVD) and made five incorrect predictions. It correctly identified 155 CVD cases, with only seven misclassifications.

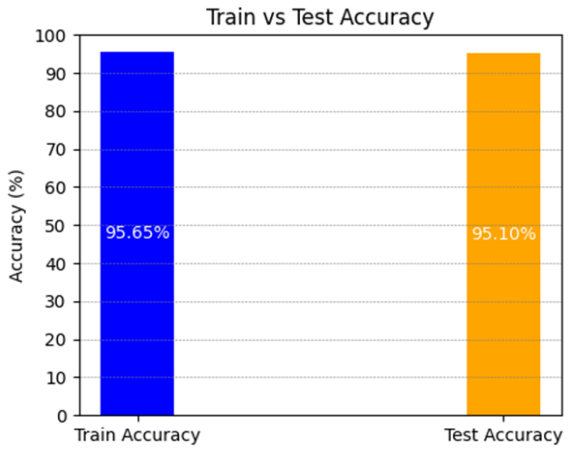


Figure 12: Train and Test accuracy DT

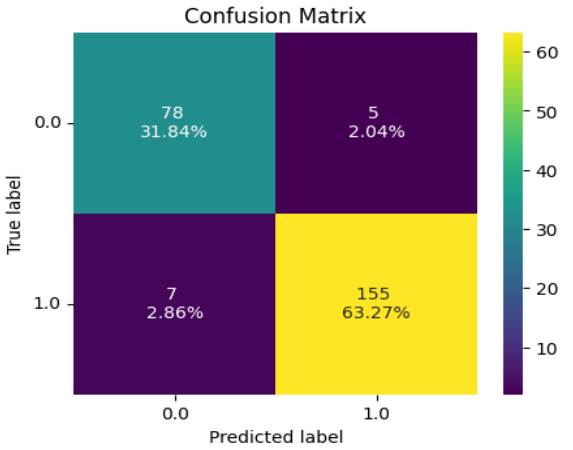


Figure 13: Confusion Matrix of DT

Overall, the Decision Tree classifier made 12 errors but successfully identified 233 of 245 cases. The model's excellent performance is highlighted in the classification report in Figure 14, particularly its high true-positive rate of 95.68%, which shows that it can accurately identify positive situations. This model's accuracy is 95.10 percent. Furthermore, the precision rate was remarkably high at 96.88%. With an overall accuracy of 96.27%, the model demonstrated its dependability and efficacy. The Decision Tree model's ROC curve result is displayed in Figure 15. The ROC-AUC displayed an outstanding classification score of 0.98. This demonstrates that the model can be relied upon because it accurately distinguishes between the two classes. This type of outcome reflects the effectiveness of the Decision Tree in predicting the chosen data set.

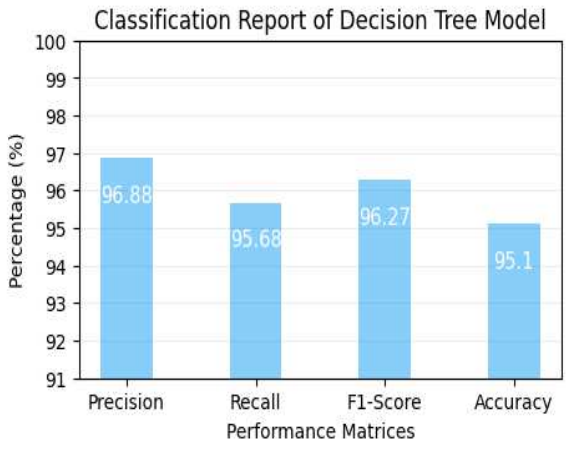


Figure 14: Classification Report of DT

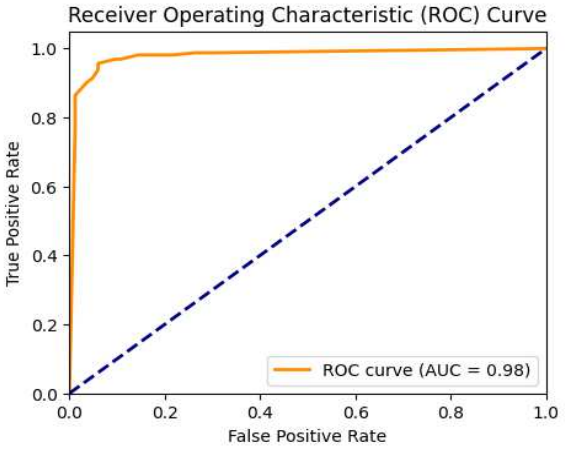


Figure 15: ROC Curve of DT

3.1.3 Random Forest

A Random Forest classifier was trained and its performance evaluated using the dataset. The model's accuracy on the training and test data is contrasted in Figure 16. According to the chart, the accuracy for the test data is 98.37%, while the train accuracy is 99.92%. The model performed well on the unseen test data, as indicated by the slight difference between the two.

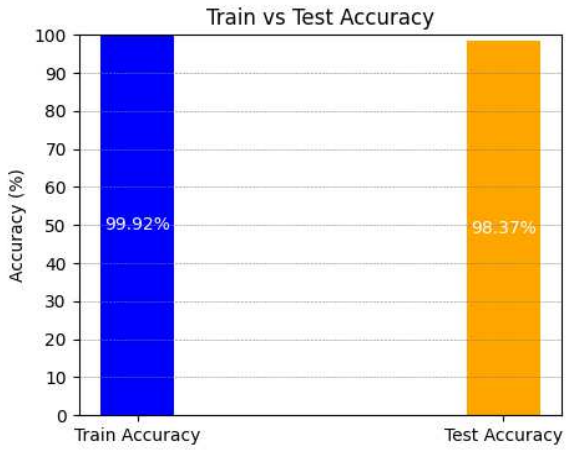


Figure 16: Train and Test Accuracy of Random Forest

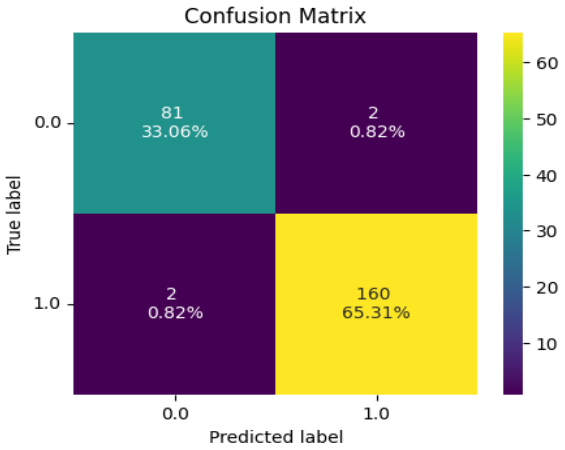


Figure 17: Confusion Matrix of Random Forest

The confusion matrix produced by the model's output is shown in Figure 17. When the anticipated class and the actual class match, the yellow and green cells in this matrix indicate accurate predictions. The purple cell, on the other hand, denotes misclassifications in which the actual class and the anticipated class are different. For class 1, the model correctly predicted 160 out of 162 samples, and for class 0, it correctly predicted 81 out of 83 samples, with two misclassified samples. With four misclassifications, the RF model's overall accuracy of 98.37% was remarkable, with 241 out of 245 cases being accurately predicted. The Random Forest classifier's classification report, shown in Figure 18, shows an outstanding precision rate of 98.77% and a perfectly true positive rate of 98.77%. With a rate of 98.77%, the model demonstrated a balanced mix of accurate and inaccurate predictions and was found to be reliable and helpful in this study. The outcome of the RF's ROC curve is displayed in Figure 19. It scored at 1.00 in ROC-AUC, which is an ideal score, implying that the model perfectly classifies the two classes. This score indicates that the classification is error-free and is indicative of high accuracy in the model, as well as potential over-fitting to the data.

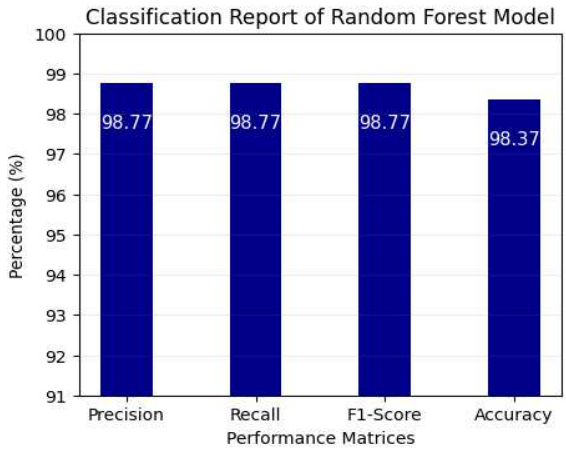


Figure 18: Classification report of RF

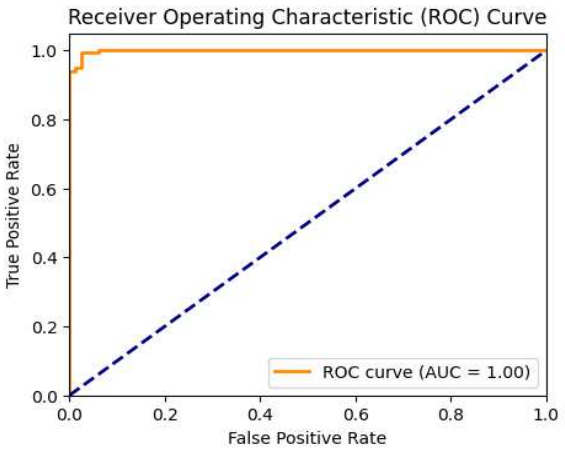


Figure 19: Classification report of RF

3.1.4 Multilayer Perceptron

The Multilayer Perceptron model was tested on both the training dataset and the testing dataset. The plot of the accuracy of performance on training versus test is shown in Figure 20. The model achieved an impressive accuracy of 99.84% on the training dataset, while the testing dataset accuracy was slightly lower at 97.55%. This marginal loss in performance indicates good generalization of the model and no heavy overfitting.

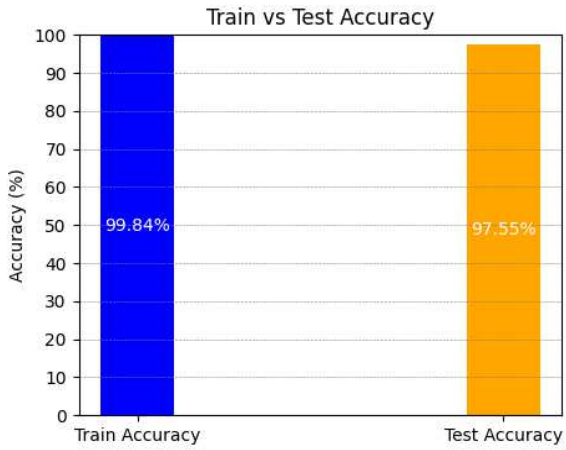


Figure 20: Train and Test Accuracy of MLP

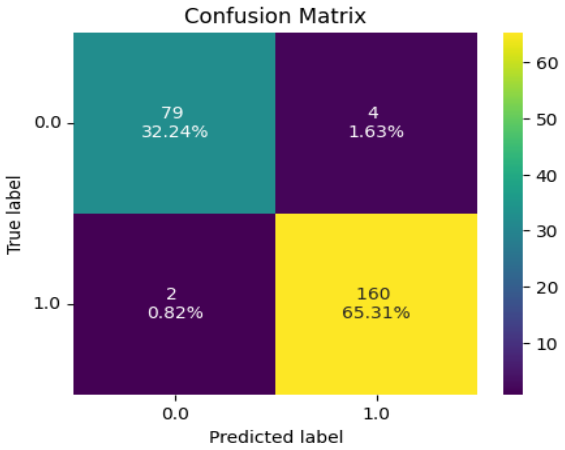


Figure 21: Confusion Matrix of MLP

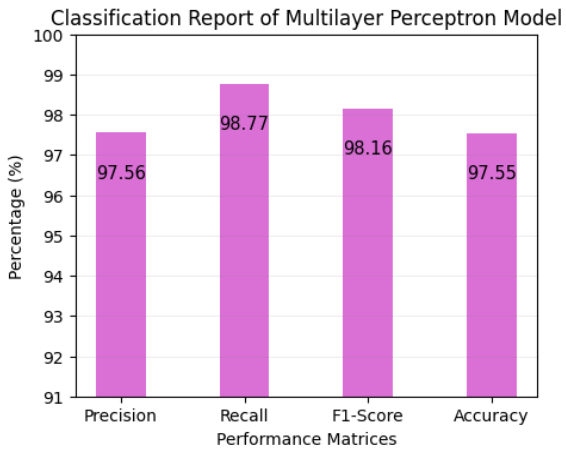


Figure 22: Classification Report of MLP

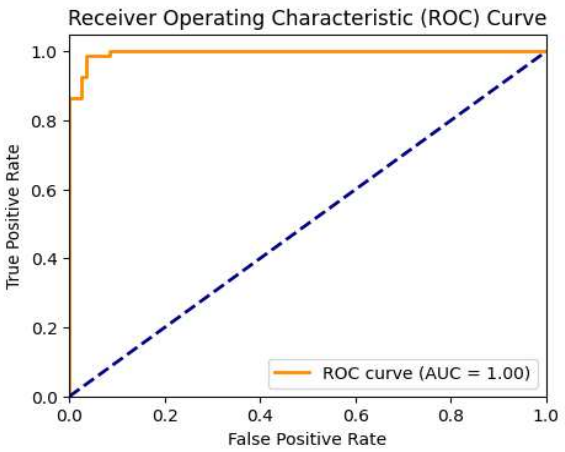


Figure 23: ROC Curve of MLP

Figure 21 presents the confusion matrix for the MLP model applied to the dataset. This matrix provides a straightforward way to visualize our model's classification performance. MLP was able to predict 160 samples of class 1 correctly, achieving a classification accuracy of 65.31% with only two samples misclassified (0.82%). In class 0, it correctly predicted 79 average samples with an accuracy of 32.24% and four wrong predictions. Among 245 instances, the model made only six misclassifications, which reveals its good performance in handling the classification task. The final performance metrics of the model are summarized in the

classification report shown in Figure 22. With an F1-score of 98.16%, 98.77% recall, and 97.56% precision, the MLP achieved a solid balance between detecting true positive cases and preventing false positives. For predicting cardiovascular disease, the model is dependable and stable, with an overall prediction accuracy of 97.55%. The outcome of the MLP ROC curve is displayed in Figure 23. Its faultless classification was revealed by its superb ROC-AUC score of 1.00. It indicates that the model's ability to differentiate between the two classes in the given data is flawless. This kind of performance indicates a very high degree of predictive ability.

3.1.5 Deep Neural Network

Figure 24 shows a comparison of the model accuracy acquired on these two datasets. The model achieved an accuracy of 99.53% on the training data and a slightly lower accuracy of 97.54% on the test data. This slight difference suggests that the model performs effectively when applied to fresh, untested data. Figure 25 shows the confusion matrix of DNN. From the figure, we can see that for class 0, it correctly predicted 78 samples and misclassified five samples. And for class 1, it correctly predicted 161 samples and misclassified 1 sample. The chart in Figure 26 represents the precision, recall, F-score, and accuracy of 97.97%, 98.77%, 97.86%, and 97.54% respectively. Figure 27 again shows the ROC curve of DNN. It showed a ROC-AUC value of 0.99, indicating almost perfect classification ability. The high value illustrates that the model is highly efficient in separating the two classes. The result confirms the strength and high predictive power of DNN in this task.

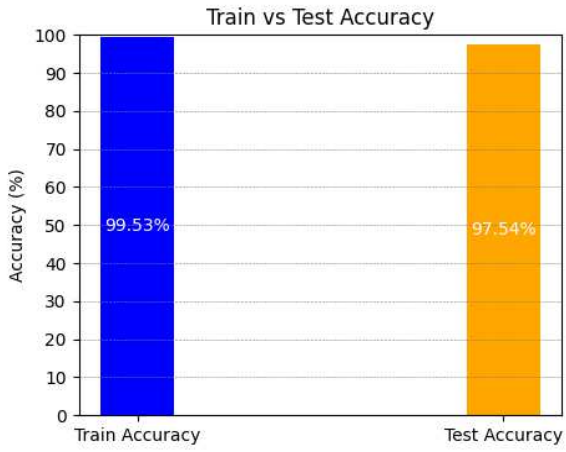


Figure 24: Training and Testing Accuracy of DNN

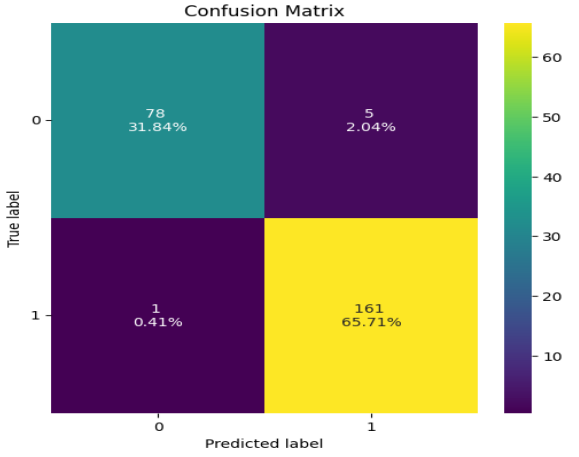


Figure 25: Confusion matrix of DNN

3.1.6 Stacking Ensemble Model

The accuracy of the model on the training and testing datasets is shown in Figure 28. On the train data, the model's accuracy was a high 99.77%; however, on the test data, it was somewhat lower at 99.18%. This suggests that the model can successfully generalize to test data that haven't been observed yet. Figure 29 displays the confusion matrix, highlighting the strong performance of the ensemble model. For Class 1, the model achieved perfect classification, correctly identifying 162 instances with no misclassifications, resulting in an accuracy rate of 66.12%.

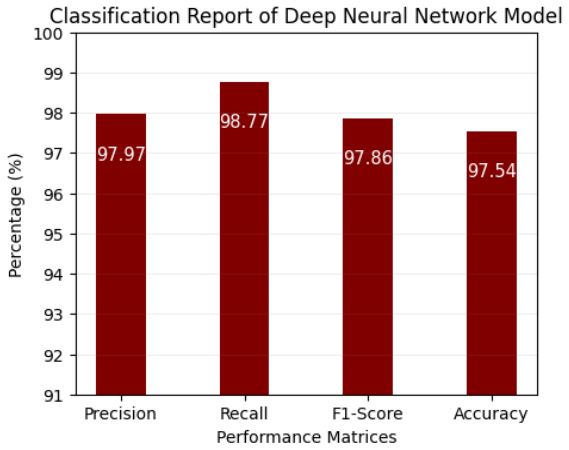


Figure 26: Classification report of DNN

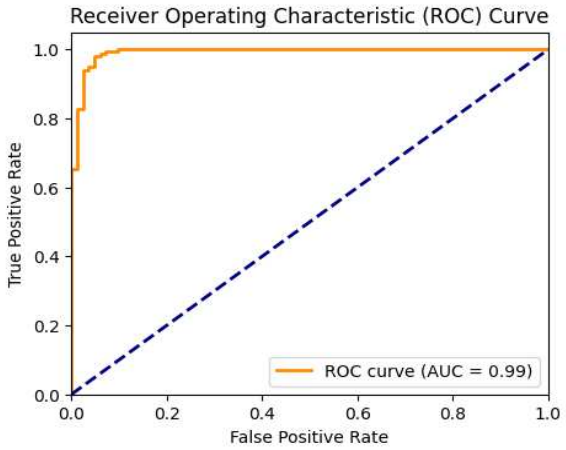


Figure 27: ROC Curve of DNN

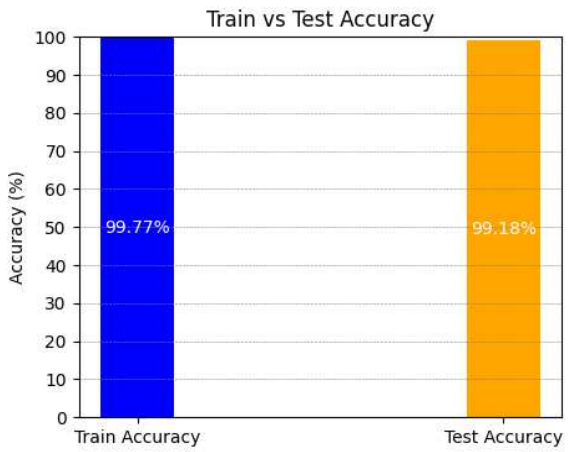


Figure 28: Train and Test Accuracy of Stacking Ensemble Classifier

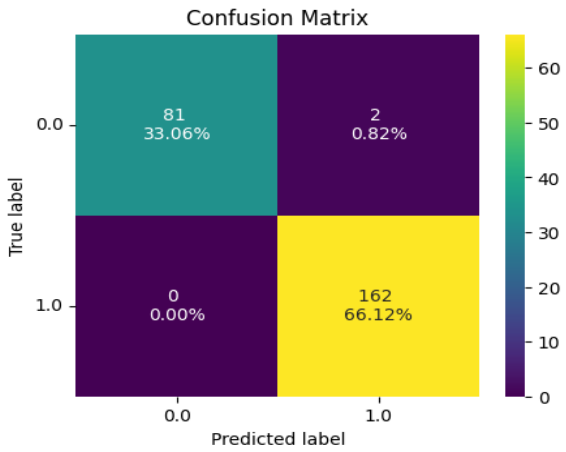


Figure 29: Confusion matrix of Ensemble Model

Similarly, for class 0, it accurately classified 81 samples (33.06%), while misclassifying only two instances, corresponding to a minimal error rate of 0.82%. Out of 245 total samples, the model correctly predicted 243, indicating highly reliable performance. The classification report in Figure 30 provides additional evidence of the model's efficacy. With a precision of 98.78%, a recall rate of 100%, and an F1-score of 99.39%, the stacking model achieved a remarkable overall accuracy of 99.18%. According to the metrics, the model performs exceptionally well in detecting positive situations while preserving high precision and reducing false positives. The stacking classifier's ROC curve value is displayed in Figure 31. By displaying a ROC curve value of 1.00, this model demonstrates perfect classification. It indicates that the model's ability to differentiate between the two classes in the given data is flawless. This kind of performance indicates a very high degree of predictive ability.

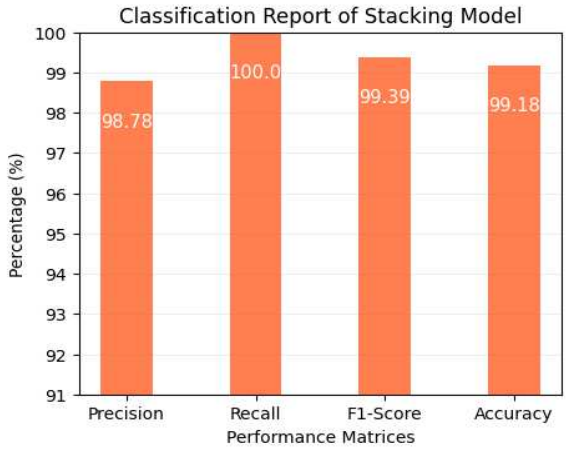


Figure 30: Classification report of Stacking Ensemble Model

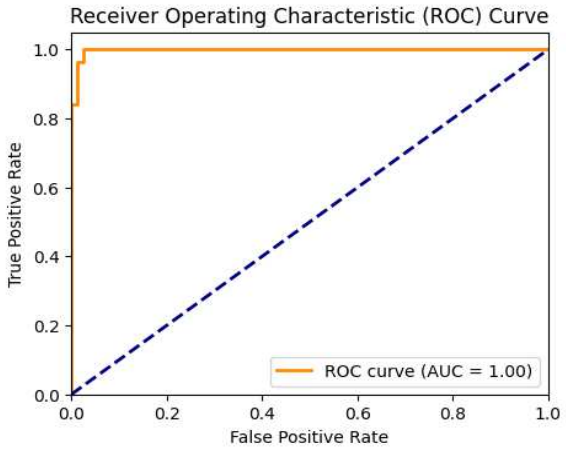


Figure 31: ROC Curve of Stacking Ensemble Model

Table 2: Description of the Hyperparameters of all Models

Model	Best Hyperparameters
SVM	'C': 1, 'gamma': 'scale', 'kernel': 'poly', CV:5
Decision Tree	max_depth: 12, min_samples_split: 2, min_samples_leaf: 13, max_leaf_nodes:30, criterion: 'gini', cv: 10
Random Forest	n_estimators: 200, criterion: 'entropy', max_depth: 20, min_samples_split: 2, min_samples_leaf: 1, cv: 5
MLP	hidden_layer_sizes: 100, activation: 'tanh', solver: 'lbfgs', alpha: 0.1, learning_rate: 'constant', cv: 10
DNN	Layers: 2, epochs: 1-100, loss: 0.4911 to 0.021, units_per_layer: 100, learning_rate: 0.1, dropout: [0.2, 0.3, 0.4, 0.5], cv: 20 trials, 5-fold
Stacking Classifier	Cv: 5, dt_max_depth: 3, final_estimator_c: 3.92, lr_c: 9.72, mlp_alpha: 0.0085, mlp_hidden_layeer_sizes: 50, Svm_C: 5.71

3.2 Comparative Analysis

All the models are tuned with their hyperparameters in search of finding the best combination of parameters to produce the best outcome of each model. We employed the "GridSearchCV" and "RandomSearchCV" tools to initially set the parameters and identify the best parameters for training the models, as shown in Table 2. This study evaluates various machine learning and deep learning models, especially SVM, DT, RF, MLP, DNN, and the Stacking Ensemble Model, using key performance metrics. The comparison chart for the performance matrices is shown in Figure 31. To facilitate comparisons, Table 3 presents the accuracy, precision, recall, F1-score, ROC, MAE, MSE, and RMSE values for each model. With the highest scores on all metrics—accuracy (99.18%), precision (98.78%), recall (100%), F1-score (99.39%), and ROC score (0.997)—the Ensemble Model (Stacking) performed exceptionally well. Out of all the individual models in Table 3, this one has the lowest error value (MAE = 0.0082, MSE = 0.0082, and RMSE = 0.0904). By combining the predictions of several classifiers, the Stacking

model, an ensemble technique, improves performance through higher generalization. With an accuracy of 98.37%, precision of 98.77%, and recall of 98.77%, Random Forest also demonstrated remarkable performance, striking a good balance between identifying real positives and reducing false positives. The ROC curve comparison is depicted in Figure 32. The ROC value of this model is 0.998. SVM, with a precision of 98.37%, performed excellently in predicting the positive class, and its accuracy (97.96%) and F1-score (98.44%) were also impressive. However, the SVM model demonstrated a slightly lower recall (97.53%) compared to Random Forest and Ensemble, suggesting it may miss a few actual positive cases. MLP and DNN showed similar performance, with F1-scores of 98.16% and 97.86%, respectively, and recall values of 98.77%. The ROC values of these two models are 0.995 and 0.996, respectively. While these models performed slightly below Random Forest in terms of recall, they remained competitive in accuracy and precision. The Decision Tree model showed the lowest performance, achieving an accuracy of 95.10%, a precision of 96.88%, and an F1-score of 96.27%. This model gained a 0.978 ROC value. While it had a decent performance, it lagged other better-performing models, particularly in recall and F1-score. Additionally, Figure 30 displays the combined ROC curve of all models, along with the stacking ensemble model. It demonstrates that all models achieved excellent performance, with AUC values exceeding 0.97. Among them, the best classifier with an AUC of 0.98 is Random Forest. Stacking, DNN, MLP, and SVM also demonstrated near-perfect results. The optimization helped stabilize the performance and reduce overfitting across various classifiers. Lastly, the Stacking Classifier was used to further build upon these fine-tuned models, utilizing properly chosen meta-parameters, resulting in improved accuracy and generalization.

The Stacking Classifier offers improved performance compared to other models, as it leverages the strengths of various base learners, including Decision Trees, SVM, LR, and MLP, thereby enhancing its discriminative power. The base models represent varying patterns within the data, and they are combined to obtain a final prediction of the pieces in a meta model that stacks the final predictions. By using 5-fold cross-validation, this method may help avoid overfitting and guarantee that models are generalizable. In regards to accuracy and F1-score, the stacking classifier continuously outperformed individual classifiers, such as SVM, DT, and RF. Better recall results in fewer false negatives, which is crucial for any application that predicts cardiovascular disease.

Table 3: Comparison of Performance and Error Matrix

Model	Accuracy	Precision	Recall	F1-Score	ROC	MAE	MSE	RMSE
SVM	97.96%	99.37%	97.53%	98.44%	0.994	0.0204	0.0204	0.1429
DT	95.10%	96.88%	95.68%	96.27%	0.978	0.0490	0.0490	0.2213
RF	98.37%	98.77%	98.77%	98.77%	0.997	0.0163	0.0163	0.1278
MLP	97.55%	97.56%	98.77%	98.16%	0.995	0.0245	0.0245	0.1565
DNN	97.54%	97.97%	98.77%	97.86%	0.996	0.0286	0.0286	0.1690
Ensemble Model	**99.18%**	**98.78%**	**100.0%**	**99.39%**	**0.997**	**0.0082**	**0.0082**	**0.0904**

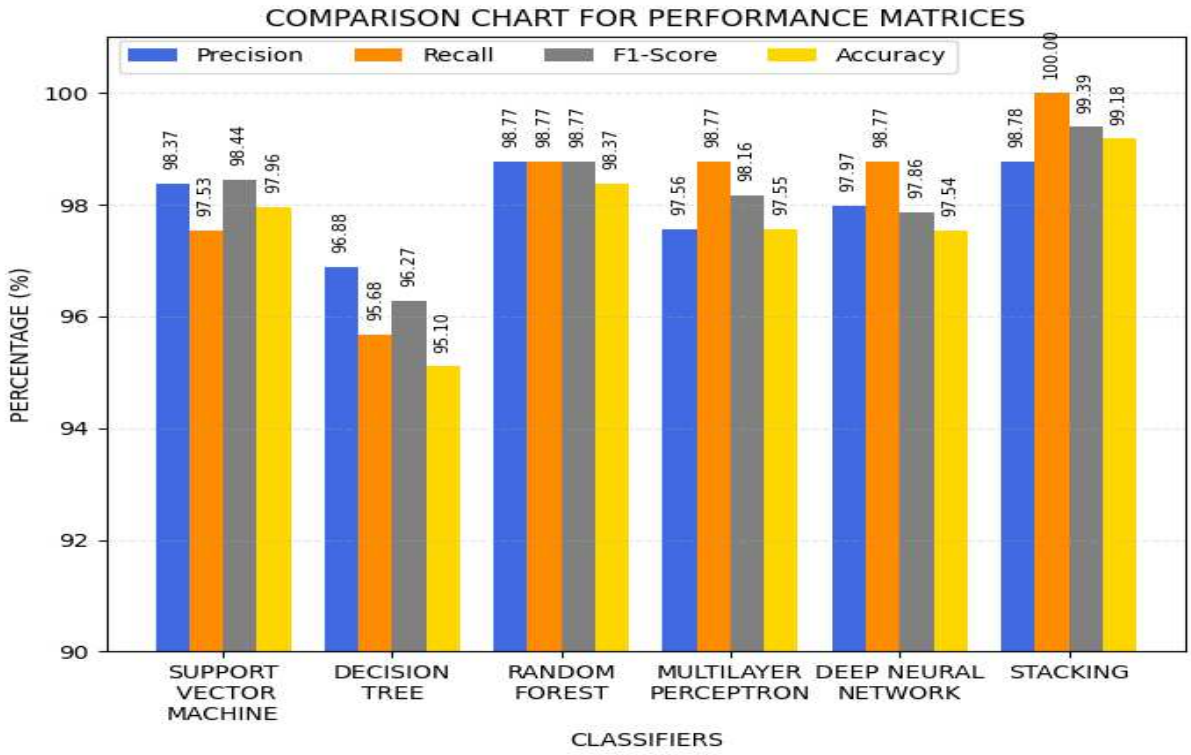


Figure 31: Comparison Chart for Performance Matrix of all Models

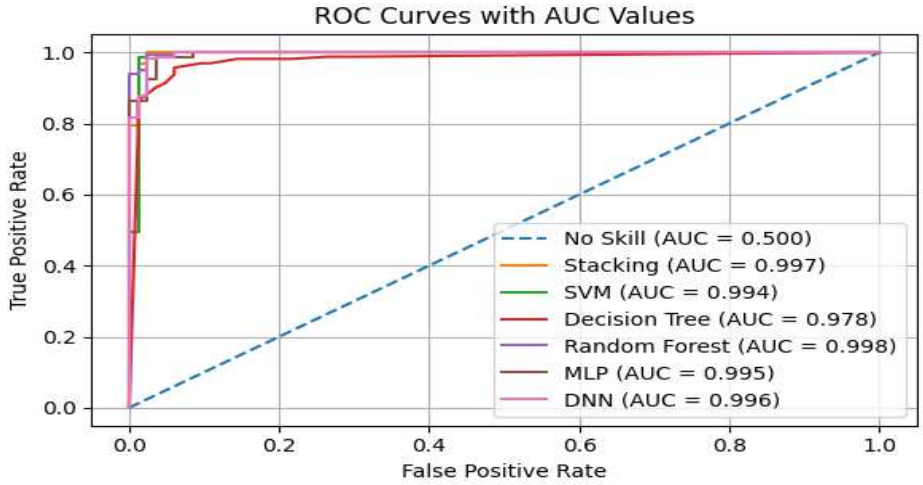


Figure 32: Comparison of ROC Curve

4. Proposed Classifier

The Stacking Classifier was chosen as the final model for this project study because it has good predictive performance and the benefits of the ensemble. This model makes predictions by combining multiple base classifiers, such as DT, LR, MLP, and SVM. This ensemble model is better at generalizing because it can use the different biases and patterns that the base learners found.Through our stacking model, we achieved surprisingly good results, with an accuracy of 99.18%, precision of 98.78%, recall of 100.00%, F1-score of 99.39%, and ROC-AUC of 0.997, which demonstrated the effectiveness of our stacking model in reducing false negatives and maintaining the clarity of class separation. A 100% recall rate is significant in healthcare, where missing positive cases can be costly and lead to adverse outcomes. Trained with 5-fold cross-

validation using scaled and resampled data, the model is both robust and fair as heterogeneity is handled. In comparison with each model in the experiment, Stacking achieved better prediction results, which balanced relatively well in terms of precision, recall, and overall accuracy. The findings reveal the efficacy of ensemble techniques, such as stacking, which utilizes models as base learners to leverage their capabilities and attain the desired results, and remains the best predictive model for cardiovascular disease in our study.

4.1 XAI

4.1.1 Global XAI

Figure 33 shows how different features contribute to predicting cardiovascular disease using a Stacking Classifier. The features with the most significant impact on the model's predictions are Cholesterol, Irregular Heart Rhythms, and Blood Pressure (BP). These features are crucial for assessing the likelihood of a cardiovascular event. Other important features that play a substantial role include Feel_Bad, Sleep_Apnea, Hypertension, and Coronary Disease. On the other hand, features such as Salt, Smoke, and Anemia seem to have a minor impact, indicating that they contribute less to the overall prediction model. Figure 34 presents the SHAP values for multiple features used in predicting cardiovascular disease with the Stacking Classifier. It illustrates the impact of each feature on the model's output, using SHAP values on the x-axis and corresponding features on the y-axis.

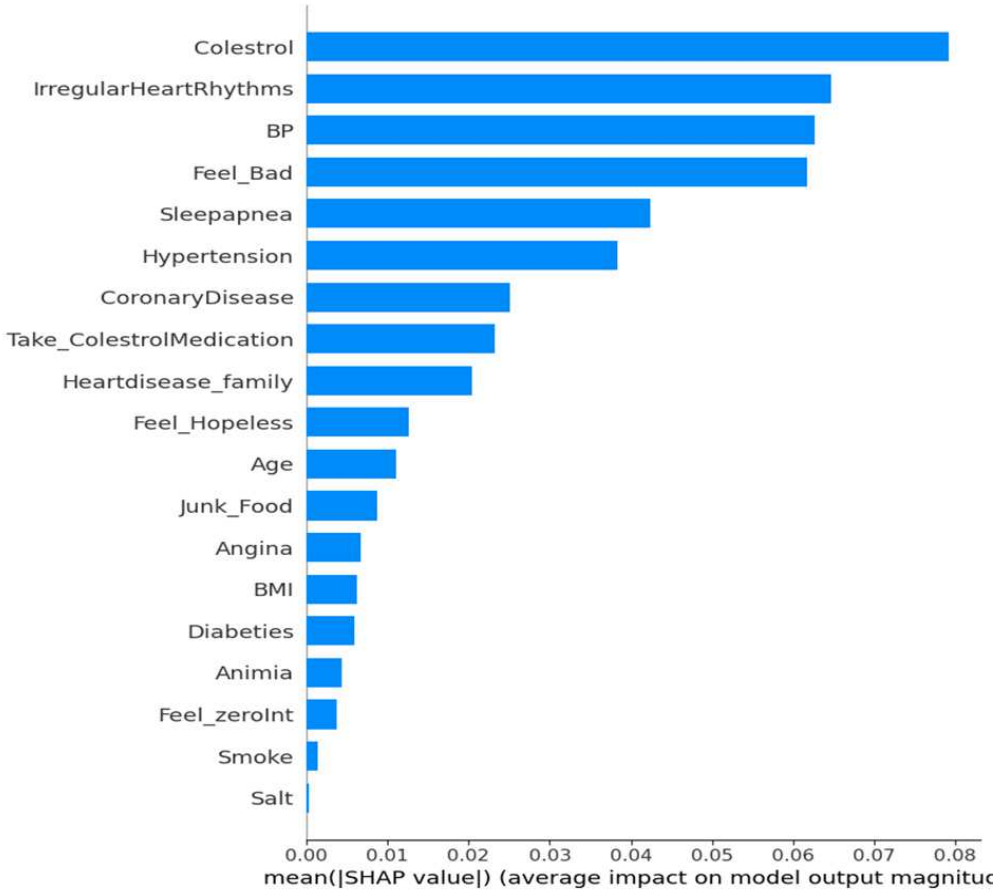


Figure 33: Average Impact of Each Feature on Model Output Magnitude

The color gradient represents the feature values, ranging from blue (low values) to red (high values), providing a clear indication of how each feature's strength affects the model's predictions. From the plot, we can see the relative importance of each feature. Features such as cholesterol, Irregular Heart Rhythms, and Blood Pressure have the most significant impact on the model's predictions, as indicated by their SHAP values extending further along the x-axis. This means they play a crucial role in determining the likelihood of developing cardiovascular disease. On the other hand, features like Salt, Smoke, and Diabetes have a minimal effect on the model, as seen by their SHAP values clustering around zero. The graphic also shows the distribution of SHAP values for each feature.

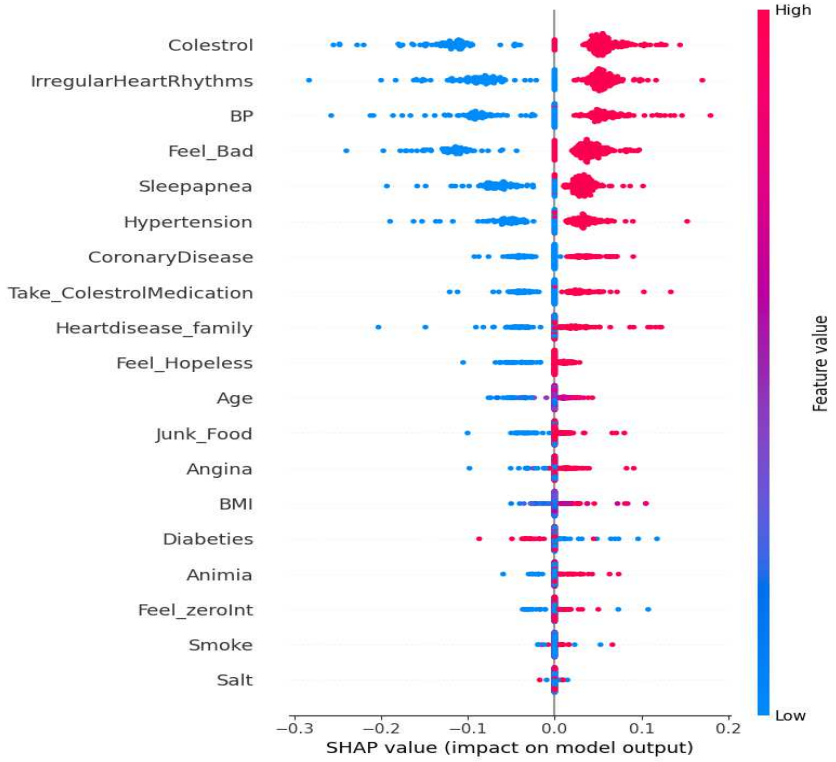


Figure 34: Impact on Model Output of Each Feature

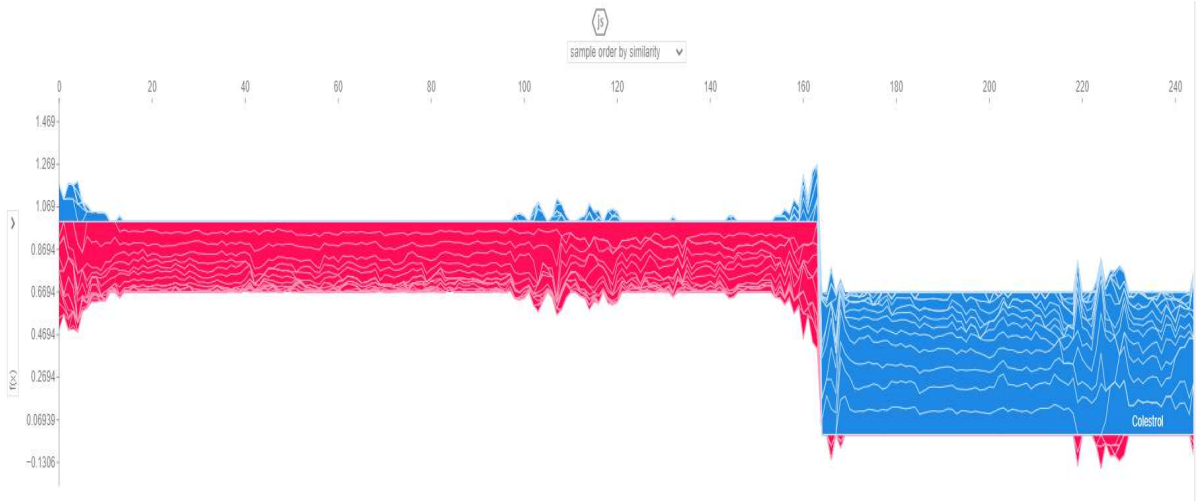


Figure 35: SHAP Force Plot

Figure 35 represents a SHAP force plot for this analysis, illustrating the role of each feature, specifically Cholesterol (Cholesterol), in the model's forecast for an individual occurrence. The figure visualizes the impact of the Cholesterol feature from blue to red, depicting lower to higher Cholesterol values. Positive SHAP values increase the prediction, while negative values decrease it. Here, the Cholesterol attribute has a significant influence on the decision, with the distribution shifting from a high probability of blue (low risk of disease) to a high likelihood of red (high risk of illness) as the Cholesterol score increases. This visualization provides a helpful insight into how cholesterol affects the model's decision-making process on a personalized level, making it easier to understand its role.

4.1.2 Local XAI

Local Explainability aims to clarify the reasoning behind individual model predictions. Figure 36 presents a SHAP force plot that shows an explanation for a single Stacking Classifier model's prediction. In this plot, the base value displayed in the center represents the model's expected prediction before considering the attributes. Each feature's impact is indicated by the blue and red bars, with blue bars representing features that have a pessimistic prediction and red bars showing features that have an optimistic prediction. In this prediction, factors like cholesterol (Colestrol), blood pressure (BP), and irregular heart rhythms are very important. These factors have a big effect on the model's results. Their presence makes the prediction increase up, which means that cardiovascular disease is more likely to happen. Other features, such as Feel_Bad and Sleepapnea, also influence the forecast, but their contribution is comparatively more minor.

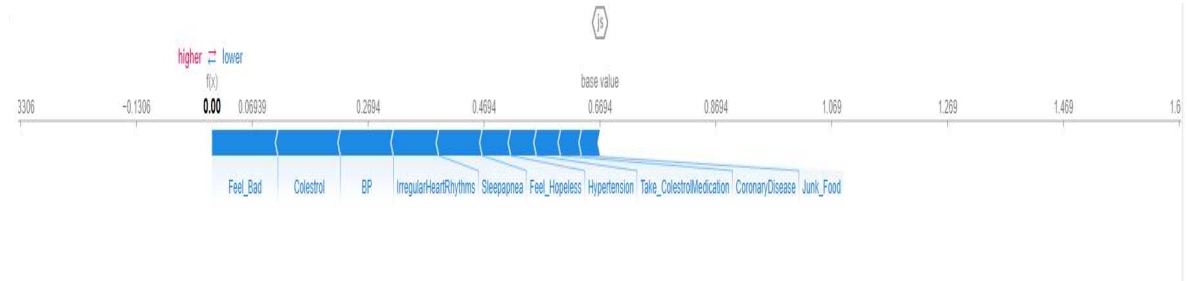


Figure 36: SHAP Force Plot of a Stacking Classifier

Figure 37 presents a SHAP dependence plot demonstrating how Cholesterol (Colestrol) values influence the model's predictions. The x-axis represents the cholesterol levels, while the y-axis displays the SHAP values. The proper hue encoding illustrates the values of another feature, IrregularHeartRhythms, and how its value impacts the relationship between Cholesterol and the prediction. From the plot, we observe that as cholesterol levels increase, the SHAP value also rises, suggesting that higher Cholesterol levels increase the likelihood of cardiovascular disease. The color of the dots indicates that IrregularHeartRhythms also influences the relationship, with higher values of IrregularHeartRhythms being linked to higher SHAP values for Cholesterol. Figure 38 represents an explanation of an individual prediction generated using LIME for a cardiovascular disease prediction instance from the Stacking Classifier model. The bar chart on the left illustrates the prediction probabilities, with the model predicting a 91% probability for class 0 and 9% for class 1. On the right side of the figure, you'll find the most essential features that affected the model's decision, along with their corresponding SHAP

values. These values indicate the extent to which each feature contributed to the prediction of the outcome. For instance, IrregularHeartRhythms has a negative contribution (-0.85), steering the forecast towards class 0, while features such as CoronaryDisease (+1.66) and Colestrol (+0.91) push the prediction towards class 1. The values in the chart denote the actual values of the features for this instance.

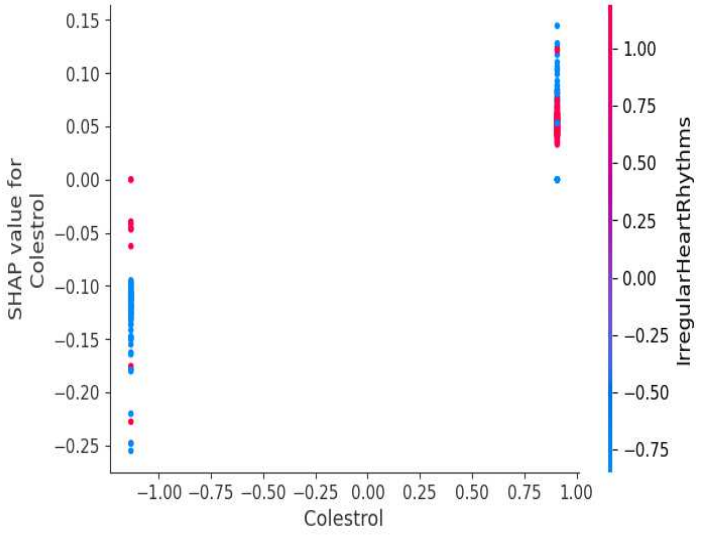


Figure 37: SHAP Dependence Plot Relationship Between Features

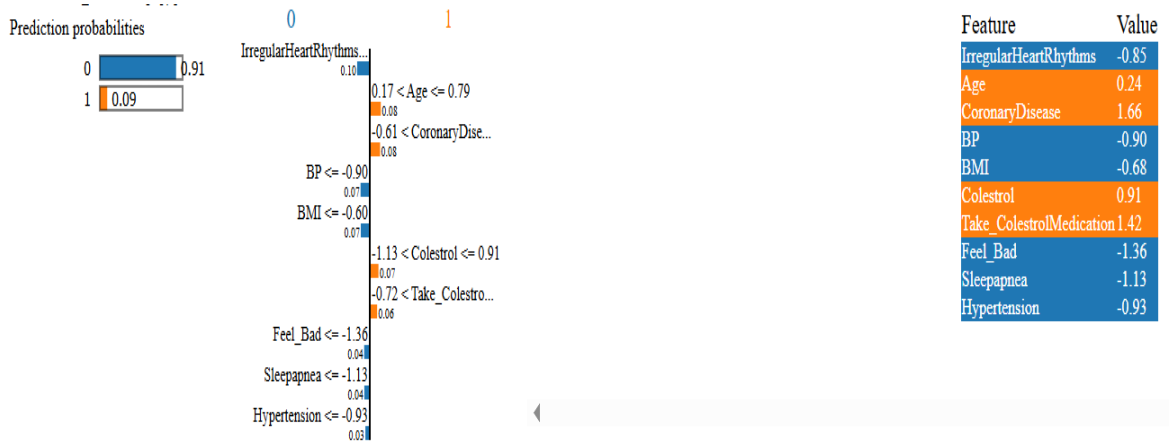


Figure 38: LIME For Local Explainability Stacking Ensemble Model

5. Discussion

Cardiovascular disease (CVD) is one of the critical health issues worldwide and in the BRICS countries, especially in developing ones, such as the case of Bangladesh, where its prevalence is escalating at a very high rate. Current methods of risk assessment can be a time- and resource-intensive task, making it impossible to perform large-scale screening and early prevention in low-resource healthcare environments. However, machine learning-based methods have recently presented attractive alternatives, offering improved speed, scalability, and accuracy in predicting CVD risk in populations.

Consistent with the Bangladeshi context described earlier, this paper contributes a procedure of ML that can perform faster, scalable, and accurate analysis. In our study, individual

classifiers achieved an accuracy of more than 95%, with RF and SVM producing the best results. The proposed Stacking Ensemble classifier, however, achieves an accuracy of 99.18%, which is significantly higher than that of each model, thus demonstrating the strength of combining multiple learners. In addition, the model quality was high with good recall and F1-score, which illustrates its strength and competence in outperforming individual classifiers.

Table 4: Comparison With Previous Studies

Authors	**Year**	**Dataset**	**Classifier**	**Accuracy**
Shivangi Diwan et al.[57]	2022	Cleveland Heart Disease Dataset	Ensemble methods (K-NN, Random Forest, Logistic Regression)	87% (Ensemble Model)
Abdullah Alqahtani et al. [58]	2022	Public cardiovascular disease dataset	Voting Ensemble	88.7%
David Opeoluwa Oyewola et al. [59]	2024	Kaggle Cardiovascular Disease Dataset	Ensemble Learning	98.45%
Narendra Kumar Sharma et al. [60]	2024	IEEE Data Port Heart Disease Dataset	Decision Tree, KNN, Logistic Regression, Random Forest, Ensemble Model	84.23% (Iterative Ensemble Approach)
Abdullah Al Maruf et al.[61]	2023	UCI Cleveland Dataset	ANN, Stacking Ensemble	90% (Ensemble Model)
E. Akkur [62]	2023	IEEE Data Port Dataset	KNN, SVM, DT,RF, HVE, SVE	93.7%(Soft Voting Ensemble Classifier)
A. Tiwari et al. [63]	2022	IEEE Data Port Dataset	RF, XGBoost, Stacking Ensemble Classifier	92.34% (Stacking Ensemble Classifier)
B.Y.Kazangirler et al. [64]	2024	Cleveland Dataset	SVM, DT, Naïve Bayes	91%(Naïve Bayes)
S. Hossain et al. [45]	2024	Collected from the Bangladeshi Hospital	LR, NB, DT, AdaBoost, RF, Bagging Tree	96.15%(Random Forest)
Proposed Model	***2025***	***Collected***	***SVM, DT, RF, MLP, DNN, Stacking Ensemble Model (LR, SVM, DT, MLP)***	***99.18%(Stacking Ensemble Model)***

In comparison to previous works presented in Table 4, our results are more comprehensive and demonstrate a notable improvement in predictive performance. For example, Shivangi Diwan et al. achieved 87% precision on the Cleveland Heart Disease Dataset, suggesting that ensemble

learning can be effective, though with accuracy limitations [57]. Similarly, Abdullah Alqahtani et al. achieved 88.7% identification accuracy using a Voting Ensemble on a publicly available CVD dataset, which represents a modest improvement but leaves considerable room for further enhancement [58]. A slightly better performance was reported by David Opeoluwa Oyewola et al., who achieved an accuracy of 98.45% through training on the Kaggle CVD dataset with the inclusion of ensemble learning [59]. In comparison, Narendra Kumar Sharma et al. achieved 84.23% accuracy using an iterative ensemble on the IEEE Data Port Heart Disease Dataset, with reduced optimization due to the dataset's complexity [60]. Abdullah Al Maruf et al. employed ANN and Stacking Ensemble techniques on the UCI Cleveland Dataset, achieving an accuracy of 90%, which indicates that stacking techniques can indeed enhance the results [61]. Similarly, E. Akkur recorded 93.7% accuracy using a Soft Voting Ensemble Classifier, which again justifies the robustness of this classifier framework [62]. A. In another study by Tiwari et al. on the IEEE data port dataset, 92.34% accuracy was achieved again using Random Forest, XGBoost, and Stacking [63]. B. Y. Kazangirler et al. also evaluated several types of classifiers with 91% accuracy; however, in their case, the best performer was Naive Bayes [64]. Additionally, based on data available from a Bangladeshi hospital, S. Hossain et al. trained LR, NB, DT, AdaBoost, RF, and Bagging Tree with an accuracy of 96.15% using Random Forest, which demonstrates the effectiveness of ensemble approaches in the local context [45].

In combination, these findings suggest that although ensemble learning algorithms have shown promise in various studies, their performance is highly dependent on dataset characteristics, data preprocessing, and the choice of classifier. Previous studies on predicting CVD have mainly used small, publicly available datasets (like UCI Cleveland and Kaggle), which limits their usefulness for diverse populations. Although these datasets are useful for testing algorithms, they don't capture the unique demographic, lifestyle, and clinical characteristics of Bangladeshi patients. Our study instead uses hospital-based data from Bangladesh, making the model more relevant to the local population. By addressing this limitation, our work provides a more reliable foundation for developing clinical decision-support tools tailored to the Bangladeshi healthcare system. Our proposed Stacking Ensemble classifier achieved 99.18% accuracy on a raw dataset collected from hospitals in Bangladesh. The model used DT, SVM, LR, and MLP as base learners and LR as the meta-learner to make better use of various styles of learning and minimize the number of errors in individual predictions. The results show that our carefully designed stacking model, along with thorough preprocessing and hyperparameter adjustment, is a practical, scalable, and efficient way to predict cardiovascular disease in a practical healthcare setting.

Additionally, our model will incorporate explainability techniques like SHAP and LIME to improve the transparency of predictions. Vital cardiovascular disease parameters illustrated the significant influence of specific clinical and demographic factors in forecasting cardiovascular disease during risk evaluation. A web-based framework is also built around the proposed model, making it easy and quick to predict CVD risk. The platform will help healthcare professionals quickly figure out a patient's risk level, which will let them act at the right time, especially in places where resources are limited.

In general, the proposed Stacking Ensemble model does a satisfactory task of predicting, generalizing, and explaining, which all indicate it becoming an appropriate option for finding CVD in the Bangladeshi population. Using SHAP and explicit LIME makes it easier to understand. A web-based framework makes it easy to set up quickly. All of these features make the system beneficial to vital interventions, help doctors make decisions, and give policymakers a place to start when making new rules.

6. API Design

We built our API to use a Stacking Ensemble model to predict cardiovascular disease. It takes user input through a front-end interface, processes the data, and then gives a prediction about whether a person is at risk of getting heart disease. The model learns from several health-related factors, such as age, BMI, cholesterol levels, blood pressure, and lifestyle choices.

Key Features

i. ***Backend:*** The API is built with Flask, a small web framework for Python. It has one POST endpoint (/predict) where individuals may upload their health data. The backend executes a stacked model (stack_model.pkl) and a scaler (scaler.pkl) to handle data that comes in. The stacked model utilizes the predictions of several base models (SVM, DT, MLP, LR) to enhance prediction accuracy.

ii. ***Frontend:*** The web interface, developed in HTML, CSS and Javascript, provides a method for users to input health data. When submitting the form, the frontend takes this data and sends it to the backend created with Flask, where it is processed then the prediction is returned: 'Risk of Cardiovascular Disease' OR 'No Risk of Cardiovascular Disease'. The prediction result is displayed on the frontend which gives the user a sense of action on the output of the model.

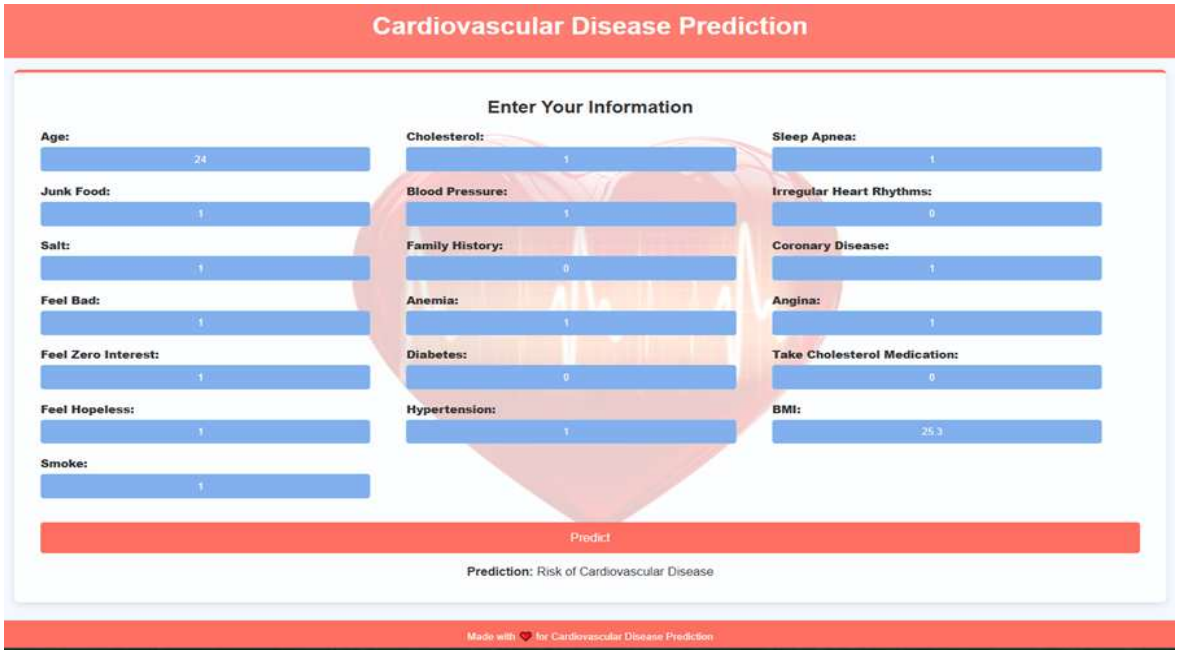


Figure 39: CVD prediction API

iii. Technology Stack:

- ***Backend*:** Python, Flask, scikit-learn, Flask-CORS.
- ***Frontend*:** HTML, CSS, JavaScript.
- ***Model*:** Stacked Ensemble model using classifiers like LR, SVM, DT, and MLP. The model is saved using pickle.

This API provides an automated and accessible way for individuals to assess their risk of cardiovascular disease based on their health inputs. It demonstrates the integration of ML models with web technologies to deliver real-time predictions in an easy-to-use format. Figure 39 illustrates the input interface where user data is provided and shows the predicted result. The API is designed for local hosts.

Strength

This research developed and compared a Stacking Ensemble Classifier for forecasting cardiovascular illnesses using several machine learning methodologies. The model was built by combining multiple base learners, including DT, SVM, LR, and MLP, alongside an LR meta-model. It is worth noting that, while the stacking model outperformed the individual classifiers in accuracy, precision, recall, and F1-score, it achieved an exceptionally high accuracy rate of 99.18%. The model's performance and robustness both became improved when hyperparameter adjustment and 5-fold cross-validation were used. Also, using SHAP and LIME made sure that the model's predictions were transparent and trustworthy, which is very important in healthcare. The proposed classifier has a lot of potential for use in real-world clinical settings because it is very accurate, has fewer false negatives, and is easy to understand. It could be a useful tool for early diagnosis and prevention of cardiovascular disease.

Limitation

The dataset utilized in this work is limited in terms of both size and scope. The model's ability to generalize and overall performance would probably improve if it had a larger and a broader dataset. The server couldn't host the API that was designed.

Advantage

This research paper provides important guideline to the policy maker and healthcare institutions in Bangladesh with regards to early diagnosis and management of cardiovascular disease. The Stacking Ensemble model successfully predicts the CVD risk and finds the most important features like Cholesterol, Irregular Heart Rhythms and Blood Pressure (BP) to identify if a patient will have a cardiovascular disease or not. This web-based framework allows clinicians to deliver timely assessments, providing personalized interventions, while also supporting rapid and scalable assessments. Also, the app can be updated for mobile phones, as it will make the app, more accessible throughout the country. Finally, the key features highlighted will assist policymakers to develop targeted preventive interventions, allocate resources, and maximize the dividends to societal residents. The most important thing is that the early risk assessment for heart disease can be done independently by anyone using the API, which means that one can take early measures for their health conditions.

7. Conclusion

Predicting cardiovascular disease (CVD) is essential in current healthcare, as it promotes early diagnosis and facilitates immediate treatment. This study provides important findings into CVD prediction, particularly within the context of Bangladesh. The study employs both deep learning and machine learning techniques to evaluate CVD risk. This study involved the development and evaluation of a Stacking Ensemble model for predicting cardiovascular diseases. We consider several learners as base including DT, LR, SVM, and MLP, and we combine them using an LR as meta-model. The results indicate that the stacking model scored higher than all classifiers in terms of accuracy, precision, recall and F1 score. The stacking model achieved a very high accurate rate of 99.18%. Adding hyperparameter tuning and 5-fold cross validation to increase the model accuracy in a more robust way helping in overfitting. Finally, we also employed SHAP and LIME for further improvements in model interpretability to ensure that the predictions made by the model are interpretable and reliable, essential in healthcare. This study addresses an emerging area in healthcare for predictive modeling, where ensemble learning techniques exhibit a clear potential to improve the accuracy and transparency of cardiovascular disease prediction.

Future Scope

In the future, using the model with larger and broader datasets will be extremely advantageous to improve the generalization and overall robustness of the model. This approach would ensure the model works effectively across different populations in various countries, considering various demographic, genetic, and lifestyle factors that impact cardiovascular health. Additionally, we aim to develop a mobile app in our future studies, utilizing a diverse dataset alongside semi-supervised learning. The app will also be integrated with primary medications and lifestyle-changing information, provided by nutritionists and healthcare experts.

Acknowledgements

We are very grateful to the patients who give information about their health conditions to conduct our study.

Availability of data and materials

Data will be available upon request to the corresponding author.

Declarations

Funding

The Noakhali Science and Technology University Research Cell funded this research.

Clinical Trial

Clinical trial number: not applicable.

Ethics approval and Consent to Participate

This study did not involve any experiments with humans or animals. Ethical approval was granted by the Ethics Committee of Noakhali Science and Technology University (reference no. NSTU/ET/EC/2024/101), which conducted a meticulous oversight of the study's ethical standards. All patients with CVD and/or their legal guardians provided informed consent before data collection. This study is based on existing clinical data, which doesn't require strict compliance with the Helsinki Declaration. Nonetheless, given the low-risk nature of the data collection and the ethics oversight reported, the study's practices appear to uphold the core principles of the Declaration of Helsinki.

Consent for publication

Participants consented verbally for publication without revealing identifiable details; no signed consent was collected. Patients were asked verbally for permission to use their data for research, and no personal or sensitive information was included.

Competing interests

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

The writers used "ChatGPT," "Grammarly," and "QuillBot" to improve the language and readability of this work. The authors take full responsibility for the content of the published paper and have reviewed and revised it as needed after use.

References

[1] "World Heart Day 2017: At the heart of health! - On Medicine," *On Medicine*, Sep.28,2017. https://blogs.biomedcentral.com/on-medicine/2017/09/28/world-heart-day-2017-at-the-heart-of-health/, (accessed June 11, 2024).

[2] A. A. Almazroi, "Survival prediction among heart patients using machine learning techniques," *Mathematical Biosciences and Engineering*, vol. 19, no. 1, pp. 134–145, 2022. https://doi.org/10.3934/mbe.2022007.

[3] WHO, "Cardiovascular Diseases,"*www.who.int*,2022.https://www.who.int/health-topics/cardiovascular-diseases/(accessed June 12, 2025).

[4] H. Dokainish, K. Teo, J. Zhu, A. Roy, …"Global mortality variations in patients with heart failure: results from the International Congestive Heart Failure (INTER-CHF) prospective cohort study," *The Lancet Global Health*, vol. 5, no. 7, pp. e665 e672, Jul. 2017. https://doi.org/10.1016/s2214-109x(17)30196-1.

[5] A. M. Islam, A. Mohibullah, and T. Paul, "Cardiovascular Disease in Bangladesh: A Review," *Bangladesh Heart Journal*, vol. 31, no. 2, pp. 80–99, Apr. 2017. https://doi.org/10.3329/bhj.v31i2.32379.

[6] C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," *2017 IEEE International Conference on Intelligent*

Techniques in Control, Optimization and Signal Processing (INCOS), Mar. 2017. https://doi.org/10.1109/itcosp.2017.8303115.

[7] World Health Organization, "Cardiovascular Diseases (CVDs)," *World Health Organization*, Jun.11, 2021. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[8] Md. A. A. R. Asif, M. M. Nishat, F. Faisal, Dip, R. R. Dip, M. H. Udoy, Md. F. Shikder, and R. Ahsan, 2021. Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease. *Engineering Letters*, *29*(2). https://www.engineeringletters.com/issues_v29/issue_2/EL_29_2_42.pdf

[9] Md. I. Hossain, M. H. Maruf, Md. A. R. Khan, F. S. Prity, S. Fatema, Md. S. Ejaz, Md. A. S. Khan, "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison," Jun. 2023. https://doi.org/10.1007/s42044-023-00148-7.

[10] N. Anjum, C. U. Siddiqua, M. Haider, Z. Ferdus, Md. A. H. Raju, T. Imam, and Md. R. Rahman, "Improving Cardiovascular Disease Prediction through Comparative Analysis of Machine Learning Models," *Journal of Computer Science and Technology Studies*, vol. 6, no. 2, pp. 62–70, Apr. 2024. https://doi.org/10.32996/jcsts.2024.6.2.7.

[11] Sasayama, S. (2008). Heart disease in Asia. Circulation, 118(25), 2669–2671. https://doi.org/10.1161/CIRCULATIONAHA.108.837054.

[12] N. A. Baghdadi, S. M. F. Abdelaliem, A. Malki, I. Gad, A. Ewis, and El-Sayed Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," *Journal of Big Data*, vol. 10, no. 1, Sep. 2023. https://doi.org/10.1186/s40537-023-00817-1.

[13] J. Li, "Prediction of cardiovascular disease based on machine learning," Applied and Computational Engineering, no. 1, pp. 60–66, Mar. 2024. https://doi.org/10.54254/2755-2721/46/20241090.

[14] Z. Zhang and J. Zhu, "Auxiliary prediction of cardiovascular diseases based on deep learning," Applied and Computational Engineering, no. 1, pp. 202–212, Mar. 2024. https://doi.org/10.54254/2755-2721/45/20241122.

[15] V. V. R. Karna, V. R. Karna, […] "A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms," Archives of Computational Methods in Engineering, no. 3, pp. 1763–1795, Oct. 2024. https://doi.org/10.1007/s11831-024-10194-4.

[16] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, "Cardiovascular Disease Detection using Ensemble Learning," Computational Intelligence and Neuroscience, pp. 1–9, Aug. 2022. https://doi.org/10.1155/2022/5267498.

[17] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure,"

American Heart Journal, pp. 1–17, Nov. 2020. https://doi.org/10.1016/j.ahj.2020.07.009.

[18] Farooqui, N.A., Haleem, M., Ahmad, M., Murugesan, P. (2024). Classification and Prediction of Cardiovascular Disease Using Machine Learning Techniques. In: Malik, H., Mishra, S., Sood, Y.R., García Márquez, F.P., Ustun, T.S. (eds) International Conference on Signal, Machines, Automation, and Algorithm. SIGMAA 2023. Advances in Intelligent Systems and Computing, vol 1460. Springer, Singapore. https://doi.org/10.1007/978-981-97-6349-8_31.

[19] Md. A. R. Khan et al., "An effective approach for early liver disease prediction and sensitivity analysis," Iran Journal of Computer Science, no. 4, pp. 277–295, Mar. 2023. https://doi.org/10.1007/s42044-023-00138-9.

[20] M. K. Hasan, S. A. Islam, Md. S. Ejaz, Md. M. Alam, N. Mahmud, and T. A. Rafin, "Classifying Bengali Newspaper Headlines with Advanced Deep Learning Models: LSTM, Bi-LSTM, and Bi-GRU Approaches," Asian Journal of Research in Computer Science, no. 4, pp. 372–388, Dec. 2023. https://doi.org/10.9734/ajrcos/2023/v16i4398.

[21] Md. Robel, Md. Ashikur Rahman Khan, Ishtiaq Ahammad, Md. Mahbubul Alam and Kamrul Hasan, "Cricket Players Selection for National Team and Franchise League using Machine Learning Algorithms," Cloud Computing and Data Science, pp. 108–139, Dec. 2023. https://doi.org/10.37256/ccds.5120243741.

[22] P. M. Goad and P. J Deore, "Predicting and Analyzing Cardiovascular Disease through Ensemble Learning Approaches," International Research Journal of Multidisciplinary Technovation, pp. 153–163, Sept. 2024. https://doi.org/10.54392/irjmt24510.

[23] S. Kumar and B. Thakur, "Heart Disease Prediction Using a Stacked Ensemble Learning Approach," SN Computer Science, no. 1, Dec. 2024. https://doi.org/10.1007/s42979-024-03499-5.

[24] Kabir, M.H., Hasan, K.F., Hasan, M.K., Ansari, K. (2022). Explainable Artificial Intelligence for Smart City Application: A Secure and Trusted Platform. In: Ahmed, M., Islam, S.R., Anwar, A., Moustafa, N., Pathan, A.S.K. (eds) Explainable Artificial Intelligence for Cyber Security. Studies in Computational Intelligence, vol 1025. Springer, Cham. https://doi.org/10.1007/978-3-030-96630-0_11.

[25] S. Islam, N. Jahan and M. E. Khatun, "Cardiovascular Disease Forecast using Machine Learning Paradigms," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 487-490. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00091.

[26] M. N. R. Chowdhury, E. Ahmed, M. A. D. Siddik and A. U. Zaman, "Heart Disease Prognosis Using Machine Learning Classification Techniques," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-6. https://doi.org/10.1109/I2CT51068.2021.9418181.

[27] R. Katarya and P. Srinivas, “Predicting Heart Disease at Early Stages using Machine Learning: A Survey,” *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, July 2020. https://doi.org/10.1109/icesc48915.2020.9155586.

[28] K. Mehrabani-Zeinabad, A. Feizi, M. Sadeghi, H. Roohafza, M. Talaei, and N. Sarrafzadegan, “Cardiovascular disease incidence prediction by machine learning and statistical techniques: a 16-year cohort study from eastern Mediterranean region,” *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Apr. 2023. https://doi.org/10.1186/s12911-023-02169-5.

[29] M. T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, “Heart disease risk prediction using deep learning techniques with feature augmentation,” Mar. 2023. https://doi.org/10.1007/s11042-023-14817-z.

[30] E. Maini, B. Venkateswarlu, B. Maini, D. Marwaha, "Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India," *Med. J. Armed Forces India*, vol. 77, no. 3, pp. 302-311, 2021. https://doi.org/10.1016/j.mjafi.2020.10.013.

[31] I. M. Pires, G. Marques, N. M. Garcia, and V. Ponciano, “Machine learning for the evaluation of the presence of heart disease,” *Procedia Computer Science*, vol. 177, pp. 432–437, 2020. https://doi.org/10.1016/j.procs.2020.10.058.

[32] A. Kondababu, V. Siddhartha, BHK. B. Kumar, and B. Penumutchi, “A comparative study on machine learning based heart disease prediction,” *Materials Today: Proceedings*, Feb. 2021. https://doi.org/10.1016/j.matpr.2021.01.475.

[33] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, “Analysis and Prediction of Cardiovascular Disease using Machine Learning Classifiers,” *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2020. https://doi.org/10.1109/icaccs48705.2020.9074183.

[34] M. Mandava, Dr. S. R. Vinta, H. Ghosh, and I. S. Rahat, “An All-Inclusive Machine Learning and Deep Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population,” *EAI endorsed transactions on pervasive health and technology*, vol. 9, Oct. 2023. https://doi.org/10.4108/eetpht.9.4052.

[35] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, “Risk prediction of cardiovascular disease using machine learning classifiers,” *Open Medicine*, vol. 17, no. 1, pp. 1100–1113, Jan. 2022. https://doi.org/10.1515/med-2022-0508.

[36] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques,” *Algorithms*, vol. 16, no. 2, p. 88, Feb. 2023. https://doi.org/10.3390/a16020088.

[37] M. N. Uddin and R. K. Halder, “An ensemble method-based multilayer dynamic system to predict cardiovascular disease using a machine learning approach,” *Informatics in*

Medicine Unlocked, vol. 24, p. 100584, 2021. https://doi.org/10.1016/j.imu.2021.100584.

[38] M. D. A. Hossen, T. Tazin, S. Khan, E. Alam, H. A. Sojib, M. M. Khan, A. Alsufyani, "Supervised Machine Learning-Based Cardiovascular Disease Analysis and Prediction," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–10, Dec. 2021. https://doi.org/10.1155/2021/1792201.

[39] CMM. Mansoor, S. K. Chettri, and HMM. Naleer, "Efficient Prediction Model for Cardiovascular Disease Using Deep Learning Techniques," *Migration Letters*, vol. 20, no. S13, pp. 449–459, Dec. 2023. https://doi.org/10.59670/ml.v20is13.6475.

[40] E. Nikookar and E. Naderi, "Hybrid Ensemble Framework for Heart Disease Detection and Prediction," International Journal of Advanced Computer Science and Applications, no. 5, 2018. https://doi.org/10.14569/ijacsa.2018.090533.

[41] S. Brindha and T. Ajisha, "An Efficient Ensemble Classifier for Heart Disease Diagnosis and Early Prediction," International Journal of Recent Technology and Engineering (IJRTE), no. 4, pp. 109–114, Nov. 2020. https://doi.org/10.35940/ijrte.d4824.119420.

[42] I. K. Sifat and Md. K. Kibria, "Optimizing hypertension prediction using ensemble learning approaches," PLOS ONE, no. 12, p. e0315865, Dec. 2024. https://doi.org/10.1371/journal.pone.0315865.

[43] S. Gasmi, A. Djebbar, and H. F. Merouani, "Enhancing Heart Disease Prediction using Deep Learning Model," All Sciences Abstracts, no. 2, p. 5, July 2023. https://doi.org/10.59287/as-abstracts.1195.

[44] Deneshkumar, V., Jithu, G., Jebitha, R. (2023). Ensemble Machine Learning Algorithms for Predicting Cardiovascular Disease. In: Roy, S., Sinwar, D., Dey, N., Perumal, T., Tavares, J.M.R.S. (eds) Innovations in Computational Intelligence and Computer Vision. ICICV 2022. Lecture Notes in Networks and Systems, vol 680. Springer, Singapore. https://doi.org/10.1007/978-981-99-2602-2_32.

[45] Hossain, S., Hasan, M.K., Faruk, M.O. et al. Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. BMC Cardiovasc Disord 24, 214 (2024). https://doi.org/10.1186/s12872-024-03883-2

[46] Zhang, Y. (2012). Support Vector Machine Classification Algorithm and Its Application. In: Liu, C., Wang, L., Yang, A. (eds) Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34041-3_27

[47] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021. https://doi.org/10.38094/jastt20165.

[48] M. Pal, “Random Forest classifier for remote sensing classification,” *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, Jan. 2005. https://doi.org/10.1080/01431160412331269698.

[49] Chieh-Chen Wu, Wen-Chun Yeh, Wen-Ding Hsu, Md. M. Islam, P. A. Nguyen, T. N. Poly, Yao-Chin Wang, Hsuan-Chia Yang, Yu-Chuan Li, "Prediction of fatty liver disease using machine learning algorithms," *Computer. Methods Programs Biomedicine*, vol. 170, pp. 23–29, 2019. https://doi.org/10.1016/j.cmpb.2018.12.032.

[50] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015. https://doi.org/10.1038/nature14539.

[51] J. Heaton, “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning,” *Genetic Programming and Evolvable Machines*, vol. 19, no. 1–2, pp. 305–307, Oct. 2017. https://doi.org/10.1007/s10710-017-9314-z.

[52] “Stacking in Machine Learning - GeeksforGeeks,” GeeksforGeeks, May 20, 2019. https://www.geeksforgeeks.org/machine-learning/stacking-in-machine-learning. (Last accessed June 20, 2024).

[53] A. M. Salih et al., “A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME,” Advanced Intelligent Systems, no. 1, June 2024. https://doi.org/10.1002/aisy.202400304.

[54] “Evaluation Metrics in Machine Learning” https://www.geeksforgeeks.org/metrics-for-machine-learning-model. Last Accessed on: June. 15, 2024.

[55] P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithyavishnupriya, "Heart Disease Prediction Using Deep Neural Network," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 666-670. https://doi.org/10.1109/ICICT48043.2020.9112443.

[56] M. Ashraf, M. A. Rizv, and H. Sharma, “Improved Heart Disease Prediction Using Deep Neural Network,” *Asian Journal of Computer Science and Technology*, vol. 8, no. 2, pp. 49–54, May 2019. https://doi.org/10.51983/ajcst-2019.8.2.2141.

[57] S. Diwan, G. S. Thakur, S. K. Sahu, M. Sahu, and N. K. Swamy, “Predicting Heart Diseases through Feature Selection and Ensemble Classifiers,” Journal of Physics: Conference Series, no. 1, p. 012027, May 2022. https://doi.org/10.1088/1742-6596/2273/1/012027.

[58] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, “Cardiovascular Disease Detection using Ensemble Learning,” Computational Intelligence and Neuroscience, pp. 1–9, Aug. 2022. https://doi.org/10.1155/2022/5267498.

[59] D. O. Oyewola, E. G. Dada, and S. Misra, “Diagnosis of Cardiovascular Diseases by Ensemble Optimization Deep Learning Techniques,” International Journal of Healthcare Information Systems and Informatics, no. 1, pp. 1–21, Nov. 2023. https://doi.org/10.4018/ijhisi.334021.

[60] N. K. Sharma, A. S. Chauhan, S. Fatima, and S. Saxena, “Enhancing heart disease diagnosis: Leveraging classification and ensemble machine learning techniques in healthcare decision-making,” Journal of Integrated Science and Technology, no. 1, Nov. 2024. https://doi.org/10.62110/sciencein.jist.2025.v13.1016.

[61] Al Maruf, A., Golder, A., Al Numan, A., Haque, M.M., Aung, Z. (2024). Prediction of Heart Disease and Heart Failure Using Ensemble Machine Learning Models. In: Udgata, S.K., Sethi, S., Gao, XZ. (eds) Intelligent Systems. ICMIB 2023. Lecture Notes in Networks and Systems, vol 728. Springer, Singapore. https://doi.org/10.1007/978-981-99-3932-9_41

[62] E. AKKUR, “Prediction of Cardiovascular Disease Based on Voting Ensemble Model and SHAP Analysis,” Sakarya University Journal of Computer and Information Sciences, no. 3, pp. 226–238, Dec. 2023. https://doi.org/10.35377/saucis...1367326.

[63] A. Tiwari, A. Chugh, and A. Sharma, “Ensemble framework for cardiovascular disease prediction,” Computers in Biology and Medicine, p. 105624, July 2022. https://doi.org/10.1016/j.compbiomed.2022.105624.

[64] B. Y. Kazangirler and E. Özkaynak, “Conventional Machine Learning and Ensemble Learning Techniques in Cardiovascular Disease Prediction and Analysis,” Journal of Intelligent Systems: Theory and Applications, no. 2, pp. 81–94, Sept. 2024. https://doi.org/10.38016/jista.1439504.

[65] J. Ivan and S. Y. Prasetyo, “Heart Disease Prediction Using Ensemble Model and Hyperparameter Optimization,” International Journal on Recent and Innovation Trends in Computing and Communication, no. 8s, pp. 290–295, Aug. 2023, doi: https://doi.org/10.17762/ijritcc.v11i8s.7208.

[66] A. E. Korial, I. I. Gorial, and A. J. Humaidi, “An Improved Ensemble-Based Cardiovascular Disease Detection System with Chi-Square Feature Selection,” Computers, no. 6, p. 126, May 2024. https://doi.org/10.3390/computers13060126.

[67] W. Huang et al., “Application of ensemble machine learning algorithms on lifestyle factors and wearables for cardiovascular risk prediction,” Scientific Reports, no. 1, Jan. 2022. https://doi.org/10.1038/s41598-021-04649-y.

[68] CMM. Mansoor, S. K. Chettri, and HMM. Naleer, “Efficient Prediction Model for Cardiovascular Disease Using Deep Learning Techniques,” Migration Letters, no. S13, pp. 449–459, Dec. 2023. https://doi.org/10.59670/ml.v20is13.6475.

[69] Shrivastava, R., Vigneshwaran, P., Soni, A.K. (2022). Heart Disease Prediction Analysis Through Ensemble Learning Model. In: Goyal, V., Gupta, M., Mirjalili, S., Trivedi, A. (eds) Proceedings of International Conference on Communication and Artificial Intelligence. Lecture Notes in Networks and Systems, vol 435. Springer, Singapore. https://doi.org/10.1007/978-981-19-0976-4_37

[70] I. Nadheer, “Heart Disease Prediction System using hybrid model of Multi-layer perceptron and XGBoost algorithms,” BIO Web of Conferences, p. 00047, 2024. https://doi.org/10.1051/bioconf/20249700047.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HeartDiseasePredictionSurveyGoogleForms.pdf