Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/rb215y02

Ensemble-Based Machine Learning for Early Detection and Risk Prediction of Cardiovascular Diseases

Muhammad Imran[1*], Sadaqat Ali Ramay[2], Tahir Abbas[3]

[1*,2,3]Department of Computer Science, TIMES Institute Multan, Pakistan.

***Corresponding Author:** Muhammad Imran
* Email: m.imran@bzu.edu.pk

Received: April 10, 2024 Accepted: May 02, 2024 Published: September 02, 2024

Abstract:
Cardiac diseases are among the leading causes of death globally. Early and accurate detection can save lives and improve health outcomes. In this research, we used machine learning techniques to predict cardiac diseases. Four widely known datasets were combined to create a diverse dataset with 1,190 instances and 12 attributes. Nine machine learning models were applied, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine, Naive Bayes, XGBoost, AdaBoost, and Gradient Boosting. Ensemble methods like Stacking, Bagging, Voting, Random Subspace, LightGBM, and CatBoost were also implemented to improve performance. The highest accuracy achieved was almost 90% using a ensemble framework. This study analyzes cardiac data based on essential features to provide optimum prediction results. The ensemble framework integrates multiple machine learning classification methods to achieve an optimal solution. The evaluation of the model was performed using precision, recall, and F1-score. In this research, SHAP (SHapley Additive exPlanations) was used to analyze the contribution of each feature towards the prediction of cardiac diseases. SHAP makes machine learning models more understandable by showing how individual features impact predictions. It helped identify important attributes like age, cholesterol levels, and exercise-induced angina that significantly influence the model's outcomes. SHAP summary plots were used to visualize feature importance and their interaction with predictions, improving the model's transparency. This understanding highlights the key factors contributing to heart disease risk. Using SHAP ensures the results are easy to interpret, making the predictions more reliable for medical decision-making. The proposed framework, which combines machine learning and ensemble techniques, outperforms individual models. It demonstrates improved prediction accuracy and effectiveness compared to existing approaches. This research introduces an innovative and practical solution for predicting heart disease in real-world clinical settings. It contributes to reducing the healthcare and societal burden caused by cardiovascular diseases. The study emphasizes the potential of advanced machine learning techniques in improving healthcare outcomes.

Keywords: Heart disease prediction; Machine learning; Ensemble classifier; SHAP; Decision Tree; Naive Bayes; SVM; KNN; logistic regression; RF; Gradient Boosting; XGB

I-Introduction:

Cardiovascular disease is one of the major reasons of mortality in the world. The world is facing several health issues due to a shortage of resources and medical facilities for many people. There are, many cases of heart disease that cannot be treated in time and it increases the mortality rates. Heart diseases (HDs) are among the top causes of death worldwide and a burden on global health systems. It accounts for approximately one-third of all mortalities worldwide, with more than seventeen million deaths annually as reported by the WHO (World Health Organization, 2020). It is necessary to detect and predict accurate risk assessments to improve the health of patients. The quick and right diagnosis reduces the load of heart disease. Medical experts use multiple techniques to identify cardiac issues, relying on medical reports available in the form of images and text. However, the accuracy of their predictions is inconsistent and may lead to incorrect decisions. Therefore, an automated system is needed to avoid such anomalies. Heart diseases, also known as cardiovascular diseases, come in different types. They need to be predicted at an early stage. These diseases are becoming more common worldwide and are causing a high number of deaths. Many persons have lost their lives due to heart disease. There are multiple hazards of this disease, which should be avoided. If someone is already affected, they need to follow preventive measures. Patients with heart disease should take steps as advised by their doctor to decrease the chances of complications.

Cardiac arrest (CA) is the leading reason of mortality worldwide. It is accountable for around 30% of all deaths globally. Recent research suggests that by 2030, deaths from heart disease could rise to around 22 million. According to a report by the American Heart Association (AHA), cardiac arrest affects about 121.5 million persons in the U.S. In Korea, heart disease ranked third as the main reason of death, accounting for 45% of mortalities in 2018. Cardiovascular disease occurs when blood flow is blocked by plaque in the arteries, leading to heart attacks or strokes (Fitriyani et al. 2020).

AI (Artificial Intelligence) is making significant strides in the field of medical sciences. ML (Machine learning) and deep learning, subfields of AI, are renowned for their high performance. Various ML and DL techniques provide optimal solutions for disease diagnosis, especially for cardiac disease.

1.1. Major Types of Cardiovascular Disorders

Cardiac arrest, also called CVD (cardiovascular disease), is a group of conditions that affect the veins and heart. There are different types of cardiac diseases. One type is myocardial infarction, commonly known as a heart attack or angina. Another is coronary heart disease (CHD), which happens when a substance called plaque builds up in the vessels carrying blood towards the heart. These vessels provide oxygen-rich fluids to the cardiac muscles.

This waxy substance grows in the arteries, the condition is called atherosclerosis. This plaque builds up slowly over many years. If not noticed early, it can cause the heart to rupture or harden. Hardened plaque makes the coronary arteries narrow, reducing blood flow to the heart. If plaque continues to harden, it can lead to blood clots forming on its surface. Large clots may completely block blood flow in the bloodlines. Blood flow needs to return quickly. If it doesn't, some portion of the heart begins to die. Rapid handling is needed during a heart disease event. If not treated in time, it can be dangerous for health or even death can occur.

1.2. Primary Risk Factors for Cardiovascular Conditions

Risk factors are the reasons that increase the chances of blocked arteries. These issues are categorized into two classes: modifiable and non-modifiable. The second one, Non-modifiable risk issues include gender, age, and heredity. These cannot be changed and are the main causes of cardiac arrest. Modifiable risk factors can be changed with effort. Some modifiable risk factors are related to habits, stress, diet, and other biochemical factors.

Heart diseases can be coronary, atherosclerosis, rheumatic, congenital, myocarditis, angina, and arrhythmia. Risk factors are habits or reasons that make people more likely to develop heart issues. They can also worsen existing heart disease. Common reasons include cholesterol, smoking, obesity, family background, high blood pressure, cholesterol, diabetes, lack of exercise, and stress.

Cardiovascular disease is a broad term that covers many heart conditions. The word “cardio” refers to the heart in medical terms, which is why these diseases are called cardiovascular diseases. Factors like high blood pressure, cholesterol, diabetes, and obesity often appear together in what is called a metabolic cluster. These clusters, when combined with cardiovascular disease, lead to a higher death rate than what BMI alone would suggest. Education and depression symptoms are also important factors.

1.3. The Need for Heart Disease Prediction

Cardiac failure causes around 17.5 million deaths all over the world, usually in middle and low-income countries. About eighty percent of these deaths are due to heart attacks and strokes. According to the WHO, India is seeing a growing number of cardiovascular disease cases each year. Over 200,000 open-heart surgeries are performed annually, and the number of patients is increasing by 20%-30% every year. This has become a major concern.

Reducing the death rate from cardiovascular diseases is crucial. To address this, researchers are working on identifying heart failure at an early stage. This allows doctors to provide timely treatment and save lives. Advances in technology have made it possible to detect heart disease more effectively. One such technology is data analytics. It supports healthcare systems to utilise data efficiently and find patterns to improve treatments and reduce costs. Studies show that better care and reduced costs can be achieved through data mining, cutting medical spending by up to 30%.

Data mining has already been successful in other fields, like marketing and e-commerce. It is now being used in healthcare to discover new techniques. However, the healthcare sector still has a lot of information but lacks the knowledge to use it fully. There is a large amount of data available, but there aren't enough tools to make sense of it and find hidden patterns.

1.4. Various Prediction Methods for Heart Disease

Several techniques are used for the identification of cardiac arrest. Immediate medical aid is needed when a heart attack occurs to prevent further damage. Doctors use technology to monitor heart patients constantly and give continuous advice to help them recover. The use of computer technology in medicine is growing rapidly. Many methods are used to diagnose and detect cardiopathy. Medical care units produce the data in large amounts, which can be utilised to make decisions. Data analysis techniques help in predicting diseases and can determine if someone is at risk of a heart attack. These techniques provide faster diagnosis, improve accuracy, and reduce the chances of heart attacks.

Data mining analyzes previous data to forecast the future. It combines machine-learning, database technology, and artificial intelligence. It is especially useful in identifying diseases at an initial stage. The process involves extracting important data, training it with a dataset, and then testing it with classification methods. It is famous by the term KDD Knowledge Discovery in Data.

Several algorithms are used to diagnose cardiopathy. It includes Decision Trees, Support Vector Machines (SVM), Neural Networks, and K-Nearest Neighbors (KNN). Machine learning methods follow a general procedure of analysis. This process consists of pre-processing, data normalization, feature extraction, and evaluating predictions based on performance. There are many prediction methods. Mostly used and adopted are data normalization, pre-processing, feature selection, and classification. Performance analysis is done usually after the prediction process to ensure accuracy.

1.5. Feature Selection Method

In data-driven methods, the selection of features is also famous as attribute selection or filtration. It is used to select key fields from raw data to make the model better. This process removes noisy, redundant, and irrelevant data. This improves the accuracy of classification. The goal of this step is to find the smallest set of features that represent the actual data. The results of feature selections depend on the right selection of attributes. The right choice of attributes plays a key role in heart disease diagnosis. Feature selection methods include supervised and unsupervised techniques. It has a variety of algorithms like particle swarm optimization, ant-colony optimization, and grey-wolf optimization.

1.6. Unsupervised Feature Selection Techniques

UFS (Unsupervised Feature Selection) is based on a filter approach. It is split into univariate and multivariate methods. This technique ranks features and selects the best based on their importance. It helps identify useful features and remove irrelevant ones. However, UFS may struggle with removing redundant features. Multivariate methods help by considering the relationship between features. This leads to better accuracy in prediction. UFS is less accurate than other techniques, but its accuracy can be improved by optimizing the method (Solorio-Fernández et al. 2020).

1.7. Convolution Neural Network (CNN)

A Convolution Neural Network (CNN) is designed to handle multi-dimensional data like time series and images. CNN detects patterns and is widely used for image analysis. It is useful for large datasets and can quickly identify heart disease. CNN has multiple layers, each performing a specific function. These layers include convolution, max pooling, and ReLU (Shankar *et al.* 2020). CNN automatically generates features and combines them with classifiers. Its advantage is that it transforms input data into output data efficiently. However, CNN does not encode the orientation or position of objects, and its deep networks take longer to train.

1.8. Classification Techniques

In data analytics, classifiers are accessed to diagnose diseases. Classification identifies and differentiates between data and classes. It is a mostly used technique in machine learning. There are three types of learning which are utilised unsupervised, supervised, and reinforcement learning. Classification algorithms assign labels to data classes. Common techniques include Naive Bayes, Logistic Regression, Decision Tree, Random Forest, Linear Perceptron, and Gradient-Boosted Tree.

Multi-layer Perceptron (MLP) is a feed-forward artificial neural network with three layers. MLP is parallel in nature and has processing elements called nodes. These nodes are connected through signal channels called connections (Nahiduzzaman et al. 2019). MLP is built on layers those are input layer, one or more hidden layers, and an output layer. Each layer has nodes represented by small circles. Information flows from one node to another through lines. The input layer receives signals, which are passed to the hidden layer via weighted links. The hidden layer processes the data and sends the output to the output layer. MLP uses backpropagation for training the extracted data.

II-Literature Review

Cardiac arrest is a type of illness caused by heart problems, which may be predicted based on multiple factors. These include symptoms such as chest pain, shortness of breath, weakness, swollen feet, and exhaustion. The causes of cardiac disease can stem from inactive lifestyles, high cholesterol, smoking, and high blood pressure. Cardiac disease risk can be reduced by addressing these causes. Scientists are working to develop improved models for predicting heart disease. The main purpose of these developed techniques is to build machine algorithms that can retrieve and utilise existing data for diagnosing cardiac disease. The building of these methods depends heavily on the dataset used. There are many datasets available on the internet for research purposes. The researchers employ these to predict disease by examining various ML and DL techniques. Scholars choose different features from these datasets to diagnose the disease.

A study was made to use a dataset of 303 heart patients. The dataset included test reports, previous history, lifestyle, and angiographic information. The researchers used data-driven methods to improve the diagnosis accuracy of heart disease (Sapra et al., 2023). Several scientists have been using multiple techniques for the identification of heart failure. They combine data-driven and DL for the prediction of cardiac disease. They used ensemble classifiers to get extended results (Rath et al., 2022). In a study, the authors used an ensemble classifier with multiple machine learners. They used RF, LR, and SVM as base learners. SMOTE was used for data balancing. They got an accuracy of 93%. They used a dataset of 303 patients from UCI and 4238 records of cardiac patients from Kaggle (Chowdary et al., 2022). Cardiac disease can be managed through exercises and physiotherapy. A study was conducted leading to the invention of an automated rehabilitation training model focused on Human Activity Recognition (HAR) techniques (Yoon et al., 2023).

Many researchers have used Recurrent neural networks (RNNs) to diagnose cardio-pathy. The researchers modified the RNN network to build a Long Short-Term Memory (LSTM) model (Sau et al., 2023). It was built to minimize memory problems encountered during training. It gave better

results. Cardiac disease is not only predictable by applications but it can also be predicted by an electronic stethoscope (Omarov et al., 2022). In a paper, the authors used the Generative Adversarial Networks (GAN) model to balance their heart dataset for predicting cardiac disease (Singhal et al., 2023). They claimed improved accuracy after dataset balancing. In another study, conventional machine learning classifiers (LR, DT, RF, K-NN, SVM, and XGB) were applied, and the logistic regression classifier obtained an accuracy of 85.84% (Ahsan et al., 2022). These conventional classifiers, such as decision trees, logistic regression, support vector machines, and random forest, provide extended results. They performed feature engineering to get the optimum results (Usha et al., 2022).
This was demonstrated in a study where a support vector machine classifier was used to diagnose cardiac failure. They got an accuracy of 87% after applying principal component analysis for feature engineering (Shah et al., 2020). A study utilized a dataset containing electrocardiogram information alongside other attributes for predicting cardiac disease (Dutta et al., 2020). Another study used a dataset featuring electrocardiogram arrhythmia characteristics to diagnose cardiovascular disease. They obtained promising results (Bertsimas et al., 2021). Researchers adopted the N2 Genetic-nu-SVM method for predicting cardiovascular disease (Sarkar et al., 2023). They attained an accuracy of 93% on an Iranian dataset with over a hundred patients. In a different study, authors achieved 89% accuracy using six machine learning classifiers (Herath et al., 2021). They used a dataset of more than 300 patients. They achieved the best results with the SVM classifier. The performance of algorithms was measured through metrics such as precision, F1-score, and recall.

In, the study achieved 90% accuracy using multiple machine-learning techniques (Ware et al., 2020). They identified the random forest classifier as the good finder on a UCI dataset consisting of 303 records. The researchers in another model combined the random forest and linear regression classifiers. They achieved 88.7% accuracy for heart disease prediction (Mohan et al., 2019). Further researchers obtained the highest accuracy by applying DT, KNN, and K-Means classifiers. They found the decision tree gave the best results in a comparative analysis (Golande et al., 2019). A different research proposed an SMO (sequential minimal optimization) classifier, achieving an accuracy of 86.4% (Reddy et al., 2021). In, machine learning techniques were applied to identify cardiac diseases (Alotaibi et al., 2019). The decision tree classifier showed optimal results. In another study, researchers used GNB, LR, and DT to diagnose heart disease. They lessen the strength of features from 13 to four for better results. They applied single-value decomposition functions. They identified that GNB and LR classifiers approached 82.75% accuracy (Ananey et al., 2020).

A convolutional neural network model was utilised to build the multimodal disease risk prediction (MDRP) approach. They applied multi-modal to identify cardiac disease (Chen et al., 2017). In another study, neural network techniques were combined with a decision tree for heart failure identification. They combined the Cleveland and Statlog datasets and achieved 99.9% on the Receiver Operating Characteristic curve (Hassani et al., 2020). The research was performed on a dataset of South Indian hospital dataset with 1,670 records. They concluded that the Random Forest classifier approached the maximum accuracy of 93.8% among other classifiers. They tested 5 machine learning algorithms, including AdaBoost, LR, NB, KNN, and RF (Maini et al., 2021). In a research, authors applied LR and SGD techniques on 303 patients. They accessed the records from the UPI repository. They identified that logistic regression achieved an accuracy of 91.67%, while SGD reached 80.0% (Miranda et al., 2021). Other methods such as Machine learning and Image Fusion algorithms have also been used in predicting cardiac diseases (Diwakar et al., 2021). A work using the Classification and Regression Tree (CART) model was applied for the identification of cardiac disorder. It approached an accuracy of 87% (Ozcan et al., 2023). The scientists use multiple datasets for the identification of heart failure. A comparison of the most used datasets is shown in Table 1.

Table 1: *Analysis of Previously Used Datasets*

DATASET	TECHNIQUE USED	REFERENCE
303 HEART PATIENTS	Machine learning methods	Sapra et al. (2023)
UCI (303), KAGGLE (4238 RECORDS)	Ensemble classifier (RF, LR, SVM) + SMOTE	Rath et al. (2022)
UCI, KAGGLE	Ensemble classifier + SMOTE	Chowdary et al. (2022)
N/A	Human Activity Recognition (HAR) techniques	Yoon et al. (2023)
N/A	Modified RNN, LSTM	Sau et al. (2023)
N/A	Electronic stethoscope	Omarov et al. (2022)
N/A	Generative Adversarial Networks (GAN) + data balancing	Singhal et al. (2023)
N/A	LR, DT, RF, KNN, SVM, XGB classifiers	Ahsan et al. (2022)
N/A	SVM + PCA for feature engineering	Usha et al. (2022)
ELECTROCARDIOGRAM DATASET	SVM classifier	Shah et al. (2020)
ELECTROCARDIOGRAM DATASET	ML classifiers	Dutta et al. (2020)
IRANIAN DATASET (OVER 100 PATIENTS)	N2 Genetic-nu-SVM	Sarkar et al. (2023)
DATASET OF MORE THAN 300 PATIENTS	SVM classifier	Herath et al. (2021)
UCI (303 RECORDS)	Multiple ML techniques, RF as best performer	Ware et al. (2020)
UCI (303 RECORDS)	RF and Linear Regression combination	Mohan et al. (2019)
N/A	DT, KNN, K-Means	Golande et al. (2019)
N/A	SMO classifier	Reddy et al. (2021)
N/A	Decision Tree	Alotaibi et al. (2019)
MULTIMODAL DATASET	CNN-based Multimodal Disease Risk Prediction (MDRP)	Chen et al. (2017)
CLEVELAND AND STATLOG DATASETS	Neural Network + Decision Tree	Hassani et al. (2020)
SOUTH INDIAN HOSPITAL DATASET (1670 RECORDS)	RF classifier	Maini et al. (2021)
UPI REPOSITORY (303 RECORDS)	LR and SGD techniques	Miranda et al. (2021)
CLEVELAND, HUNGARY, SWITZERLAND, LONG BEACH V	Isolation Forest for outlier detection, Lasso for feature selection	Bharti et al. (2021)

CLEVELAND DATASET	Majority Voting Ensemble (SGD, KNN, RF, LR) + min-max normalization	Atallah et al. (2019)
STULONG DATASET	Swarm Whale Optimization (SOW) + LeNet model	Nashif et al. (2018)
UCI DATASET	Weight learning based on density information	Govinda et al. (2023)
UCI MEDICAL REPOSITORY	Fuzzy Rough Set + Multiple Imputation	Xie et al. (2021)
CLEVELAND DATASET	Soft voting ensemble classifier	Rath et al. (2022)
CLEVELAND, UCI, KAGGLE	XGBoost + Optuna	Chandrasekhar et al. (2023)
CLEVELAND DATASET	SVM, XGB, NB, LR classifiers	Srinivas et al. (2022)
CLEVELAND DATASET	MLP-PSO	Karthick et al. (2022)
VA LONG BEACH DATASET	Grey Wolf Horse Herd Optimization-based Shepard CNN	Bataineh et al. (2022)
VA LONG BEACH DATASET	Deep Neural Network (DNN)	Pendala et al. (2023)
CLEVELAND DATASET	Soft Voting Ensemble (Gradient Boost, XGBoost, AdaBoost, etc.)	Alfaidi et al. (2022)
N/A	GWO-KELM method	Pan et al. (2022)
CLEVELAND DATASET	Multi-model approach combining XGB, SVM, RF, and LR	Deepika et al. (2022)
FRAMINGHAM HEART STUDY DATASET	Deep learning techniques	Kapila et al. (2023)
ECHOCARDIOGRAPHIC DATA	CNN-based model	Al Reshan et al. (2023)
KAGGLE DATASET	Ensemble Learning (RF, XGB, DNN)	Rustam et al. (2022)
DATASET OF 1000 PATIENTS	Hybrid Approach (Genetic Algorithms + Deep Learning)	Doppala et al. (2022)

III- Methodology

A – Dataset and Preprocessing

For this study, several datasets were combined to improve the strength and generalizability of the cardiopathy prediction model. These datasets were obtained from the UCI repository (Mohan et al., 2019). The combined data is summarized in Table 2. The datasets contain records of patients with diverse demographic and clinical features, creating a comprehensive resource for training and evaluating the model.

The datasets were integrated by reading each CSV file individually and concatenating them into a single, unified dataframe. A Python script was used for this purpose. The script defined the file paths, read the datasets into separate dataframes, and then merged them into a pooled dataset. This process ensured consistency and completeness in the combined data.

Table 2: *Datasets for Concatenation*

SR#	*DESCRIPTION*
1	Cleveland Dataset
2	Hungarian Dataset
3	Long Beach VA
4	Statlog (Heart) Dataset
5	Switzerland

The main goal of combining datasets was to enhance diversity and minimize biases. By integrating data from multiple sources, the dataset became more representative of a wide range of patient conditions. This improved the model's ability to predict cardiopathy. Additionally, combining datasets increased the sample size, leading to better results.

The benchmark datasets contain 76 attributes, which represent various types of patient information. However, most studies, including this one, focused on 12 key features that are critical for predicting heart disease. The dataset includes 1,190 records of cardiac patients. It has a **Target** column, where 0 indicates the absence of heart disease and 1 indicates its presence.

The 12 selected features are described below:

1. **Age**: The age of the patient in years.
2. **Sex**: Gender of the patient (1 for male, 0 for female).
3. **Chest Pain Type (CP)**: Indicates the type of chest pain, categorized as:

0: Typical angina.
1: Atypical angina.
2: Non-anginal pain.
3: Asymptomatic.

4. **Resting Blood Pressure (Trestbps)**: The patient's blood pressure in mm Hg while resting.
5. **Cholesterol (Chol)**: The serum cholesterol level in mg/dL.
6. **Fasting Blood Sugar (FBS)**: Whether the fasting blood sugar is greater than 120 mg/dL (1: True, 0: False).
7. **Resting ECG Results (Restecg)**: Results of the electrocardiogram test:

0: Normal.
1: ST-T wave abnormality.
2: Left ventricular hypertrophy.

8. **Maximum Heart Rate Achieved (Thalach)**: The highest heart rate achieved during exercise.
9. **Exercise-Induced Angina (Exang)**: Whether the patient experienced angina during exercise (1: Yes, 0: No).
10. **ST Depression (Oldpeak)**: ST segment depression induced by exercise, relative to rest.
11. **ST Slope (Slope)**: The slope of the ST segment during peak exercise, categorized as:

0: Upsloping.
1: Flat.
2: Downsloping.

12. **Target**: The patient is cardiac or not (0–no, 1-yes)

These features are critical for predicting heart disease as they capture essential demographic, clinical, and diagnostic information. The sample instances of dataset are reflected by the Figure 1.

```
dataset information
Head:
   age  sex  chest pain type  resting bp s  cholesterol  fasting blood sugar  \
0   40    1                2           140          289                    0
1   49    0                3           160          180                    0
2   37    1                2           130          283                    0
3   48    0                4           138          214                    0
4   54    1                3           150          195                    0

   resting ecg  max heart rate  exercise angina  oldpeak  ST slope  target
0            0             172                0      0.0         1       0
1            0             156                0      1.0         2       1
2            1              98                0      0.0         1       0
3            0             108                1      1.5         2       1
4            0             122                0      0.0         1       0
Shape:
(1190, 12)
Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1190 entries, 0 to 1189
Data columns (total 12 columns):
```

Figure 1: *Selected Rows and Columns of Dataset*

After uploading the CSV files of the dataset to Google Colab, we normalized it by removing missing and duplicate values. Preprocessing of the dataset is essential for better performance of machine-learning techniques, so we separated the categorical and numerical values. Categorical data is converted to numerical values. Standard scalar is used to scale the values between 0 and 1. Outliers are handled through zscore library of scipy. The dataset is imbalanced as shown in Figure 2. It is balanced by using RandOverSampler, oversampling technique.

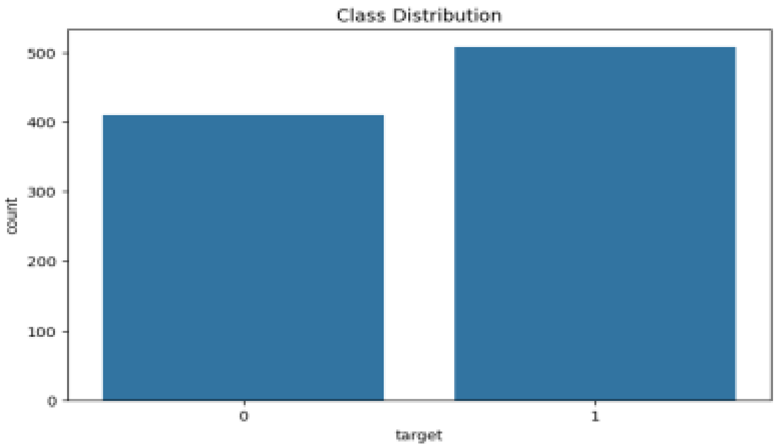

Figure 2: *Class Distribution*

Various machine learning techniques are used to predict cardiovascular disease. The performance of these techniques is evaluated using different metrics, including precision, recall, accuracy, and F1 score. Accuracy represents the overall performance of the model. It shows how well the model predicts both positive and negative cases. Recall (also called sensitivity) measures the model's ability to identify all instances of the target class. It answers whether the model can find all positive cases. Precision evaluates the model's ability to make correct positive predictions. It shows how accurate the positive predictions are. F1 Score is the harmonic mean of precision and recall. It provides a balanced evaluation of the model when both false positives and false negatives are important. Sensitivity is the ratio of correct positive predictions to the total actual positives. It indicates how well the model identifies true positives. These metrics provide a clear understanding of the model's strengths and weaknesses. They help determine how effectively the model identifies and predicts cases. The mathematical formulas for these metrics are as follows:

$$Classification\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$
$$Precision = \frac{TP}{TP+FP} \quad (2)$$
$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$
$$F1 = \frac{2*Precision*Sensitivity}{Precision+Sensitivity} = \frac{2TP}{2TP+FP+FN} \quad (4)$$

Here, FN, TN, FP, and TP stand for false negatives, true negatives, false positives, and true positives, respectively. After converting categorical data, feature scaling is applied to numerical data using the standard scaler. This ensures that the features are balanced and fall within a specific range. Next, the dataset is divided into two parts: 80% for training and 20% for testing. Nine classifiers were then applied to the dataset, and their performance was optimized by adjusting certain hyperparameters. The resulting accuracies are presented in Table 3.

Table 3: *Classification Report of Logistic Regression*

CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
LOGISTIC REGRESSION	0.8529	0.8529	0.8529	0.8529
DECISION TREE	0.8775	0.8819	0.8775	0.8780
RANDOM FOREST	0.8824	0.8828	0.8824	0.8825
ADABOOST	0.8431	0.8469	0.8431	0.8437
GRADIENT BOOSTING	0.8922	0.8934	0.8922	0.8924
SUPPORT VECTOR MACHINE	0.4804	0.5291	0.4804	0.4507
K-NEAREST NEIGHBORS	0.6225	0.6365	0.6225	0.6241
NAIVE BAYES	0.8529	0.8527	0.8529	0.8527
XGBOOST	0.8824	0.8822	0.8824	0.8822

The ensemble model combines the predictions of multiple machine learning classifiers to improve accuracy and reliability. In this study, nine classifiers were used: Logistic Regression, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Support Vector Machine, K-Nearest Neighbors, Naive Bayes, and XGBoost. Each classifier learns patterns from the training data and makes predictions. These predictions are combined in the Stacking Ensemble technique. This approach uses a meta-model, which analyzes the outputs of individual classifiers and makes the final decision. The ensemble approach leverages the strengths of all classifiers, reducing individual weaknesses.
The goal of using the ensemble method is to achieve better accuracy and a balanced performance across key metrics like precision, recall, and F1-score. These metrics reflect how well the model identifies and classifies heart disease cases.

Results of Individual Classifiers
The performance of each classifier is summarized below:
- **Logistic Regression**: Achieved an accuracy of 85.29%. It had balanced precision, recall, and F1-score, all at 85.29%. This indicates consistent performance.
- **Decision Tree**: Delivered slightly better accuracy at 87.75%. Precision and recall were also high, making it a good individual performer.
- **Random Forest**: Performed well with an accuracy of 88.24%. It achieved high scores across all metrics, showing robustness in predictions.
- **AdaBoost**: Had an accuracy of 84.31%. It showed decent precision and recall but was outperformed by some other classifiers.
- **Gradient Boosting**: The best individual performer with an accuracy of 89.22%. It also excelled in precision and recall, making it reliable.
- **Support Vector Machine**: Struggled with an accuracy of only 48.04%. The low F1-score reflects its difficulty in handling the dataset effectively.

- **K-Nearest Neighbors**: Had moderate performance with an accuracy of 62.25%. Precision and recall were also lower compared to top performers.
- **Naive Bayes**: Performed similarly to Logistic Regression with an accuracy of 85.29%. It provided consistent results across all metrics.
- **XGBoost**: Achieved strong performance with an accuracy of 88.24%. Precision and recall were also high, similar to Random Forest.

Gradient Boosting proved to be the most accurate and balanced model. Random Forest and XGBoost also performed well, showing their reliability with complex data. Logistic Regression, Decision Tree, and Naive Bayes gave consistent results. The ensemble model combined the strengths of these classifiers. This improved the overall performance. The Stacking Ensemble technique made robust and accurate predictions for heart disease. This shows the importance of ensemble learning in healthcare. Figure 4 includes a heatmap showing the performance of nine classifiers. It compares their accuracy, precision, recall, and F1-score. Each cell in the heatmap displays the value of a metric for a classifier. Darker shades indicate better performance.

The heatmap makes it easy to spot differences between classifiers. For example, Gradient Boosting has darker shades in all metrics, showing its superior performance. In contrast, Support Vector Machine has lighter shades, showing weaker results. The heatmap highlights trends and individual contributions. Models like Random Forest, XGBoost, and Gradient Boosting show strong and consistent performance. Logistic Regression, Naive Bayes, and Decision Tree perform slightly lower but are still reliable. Underperforming models, like K-Nearest Neighbors and Support Vector Machine, are easy to identify. This helps understand which classifiers contribute most to the ensemble's success. Overall, the heatmap is a valuable tool for comparing and analyzing classifier performance.

The heatmap shows overall trends and individual contributions of each classifier. Random Forest, XGBoost, and Gradient Boosting perform strongly with high scores across all metrics. Logistic Regression, Naive Bayes, and Decision Tree also perform well, but slightly lower than the top models.

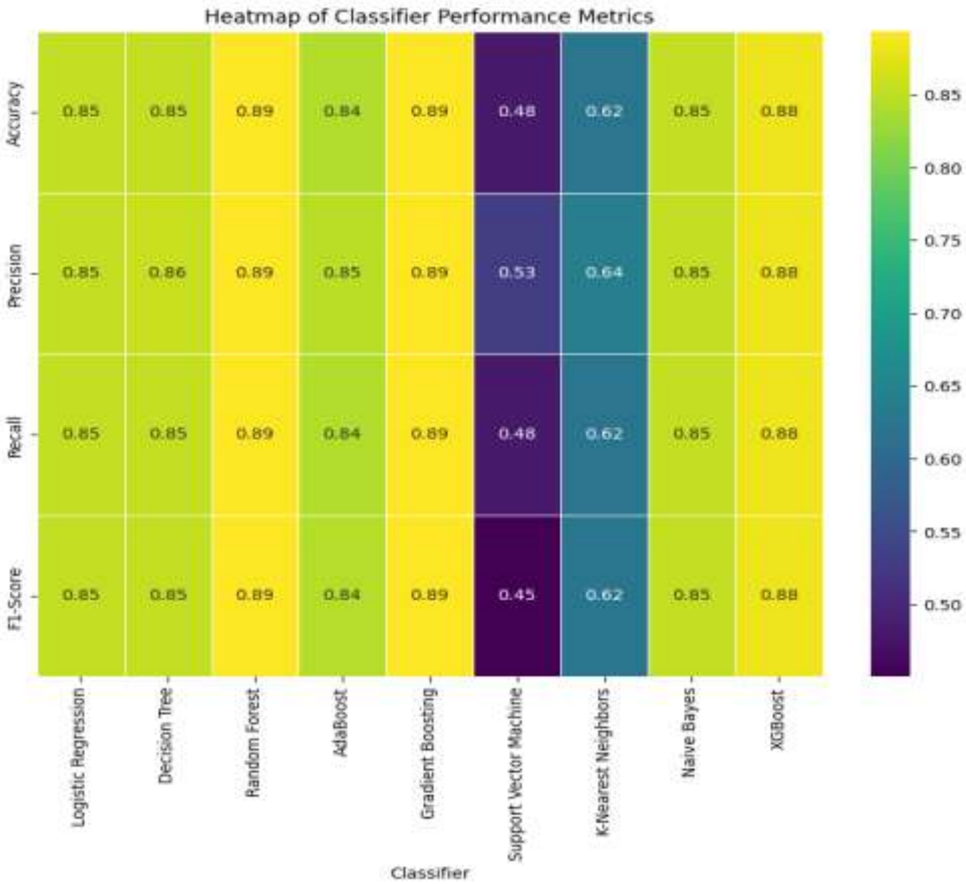

Figure 3: Heatmap of Classifier Performance Metrics

Ensemble classifiers combine the strengths of multiple models to improve performance. They take predictions from several models and merge them. This reduces errors and balances individual weaknesses. Ensemble methods often deliver better accuracy, stability, and robustness. They work well with complex datasets where no single model is ideal. By using diverse models, ensemble classifiers give reliable and general predictions.

Stacking is a powerful ensemble method. It combines outputs from multiple models using a meta-model. While individual models vary in performance, stacking optimizes their predictions. This creates a stronger final model. The main advantage of stacking is its ability to learn complex relationships between models. It leverages their strengths and minimizes weaknesses. Stacking often performs better than simple methods like averaging, making it ideal for tasks like heart disease prediction.

In this study, we used various machine learning models: Decision Tree, Naive Bayes, SVM, KNN, Logistic Regression, Random Forest, AdaBoost, Gradient Boosting, and XGBoost. Each model was fine-tuned for the best results. By analyzing their individual performances, we identified patterns to enhance accuracy and precision in heart disease prediction.

As shown in the results table, this analysis allowed us to propose a robust framework that delivers enhanced and accurate outcomes. Additionally, the Stacking ensemble classifier was applied to further improve the prediction accuracy of the model. This combination of techniques demonstrates the effectiveness of using ensemble learning for better predictions as shown in Figure 4.

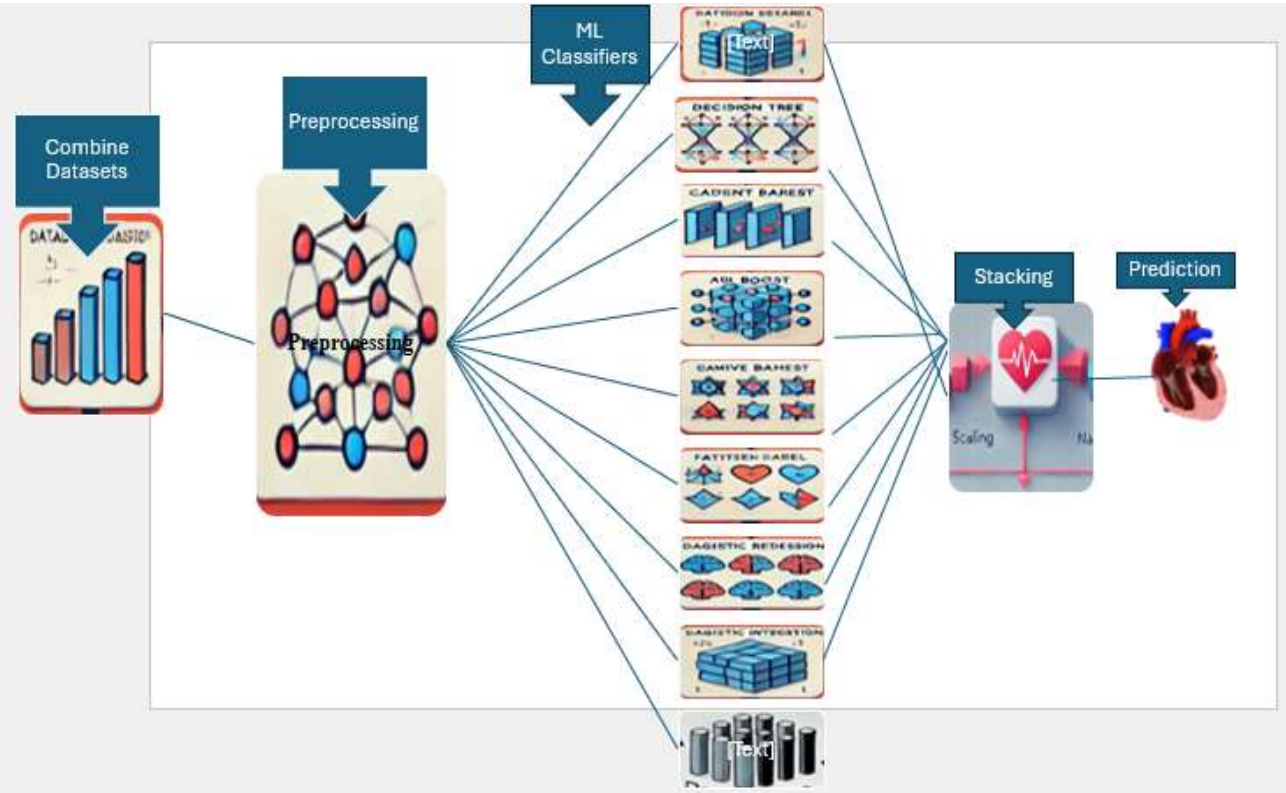

Figure 4: *The Proposed Ensemble Model for Prediction of Heart Disease*

IV. Results

This study aims to predict cardiovascular disease using machine learning techniques. In this research, we have proposed an ensemble framework that has shown the best results among other machine-learning techniques, as shown in Table 4. We have used 9 machine-learning classifiers on the open-access datasets, and our proposed method has demonstrated its superiority over others. We have used an ensemble classifier in conjunction with other ML classifiers to improve the accuracy. Our framework has achieved an accuracy of 89.70%, surpassing all other ML models. We have used the scikit-learn library to import the classifiers to predict heart disease.

Table 4: *Enhanced Accuracy Results of ML Techniques*

SR#	MACHINE-LEARNING CLASSIFIER	ACCURACY
1	Logistic Regression	0.8480392156862745
2	Decision Tree	0.8431372549019608
3	Random Forest	0.8725490196078431
4	XGBoost	0.8823529411764706
5	Gradient Boosting	0.8921568627450981
6	AdaBoost	0.8431372549019608
7	Support Vector Machine	0.5196078431372549
8	K-Nearest Neighbors	0.6225490196078431
9	Naive Bayes	0.8529411764705882
10	Stacking Classifier	0.8970588235294118

Logistic Regression, Naive Bayes, and Decision Tree demonstrated consistent results, with accuracy values ranging from 85% to 88%. These models performed well in balancing precision and recall, indicating their ability to make reliable predictions. Support Vector Machine and K-Nearest Neighbors struggled with this dataset, showing lower accuracy and F1-scores. This highlights their limitations in handling the data's specific features. Ensemble techniques improved the results significantly. AdaBoost showed some improvement, while Gradient Boosting performed the best among individual models. The stacking ensemble combined the strengths of all classifiers, creating a more robust prediction framework. Ensemble methods like Gradient Boosting and Stacking are effective for complex problems like heart disease prediction. These techniques reduce biases and improve model reliability. Metrics like precision and recall confirm the model's ability to identify heart disease cases accurately. The findings emphasize the value of combining multiple classifiers and ensemble methods for better results. The stacking classifier's performance is shown in the confusion matrix (Figure 5), demonstrating its efficient working. It shows that the model correctly identified 109 instances belonging to Class 1. These true positives confirm the model successfully detected the condition it was designed to predict. Similarly, the model also correctly classified 76 instances as Class 0 which are true negatives. The confusion matrix discloses that the model made some errors. It misunderstood 12 instances as Class 1 while they belonged to Class 0. This wrong prediction of false positives indicates cases where the model signaled the presence of the condition. Moreover, the model missed 7 cases where the condition was present but predicted as Class 0. This incorrect prediction is known as false negatives where the model failed to detect the true condition. The confusion matrix shows that the ensemble stacking classifier performed best. The ensemble classifier is effective at correctly identifying both classes. The matrix shows the maximum number of TP and TN.

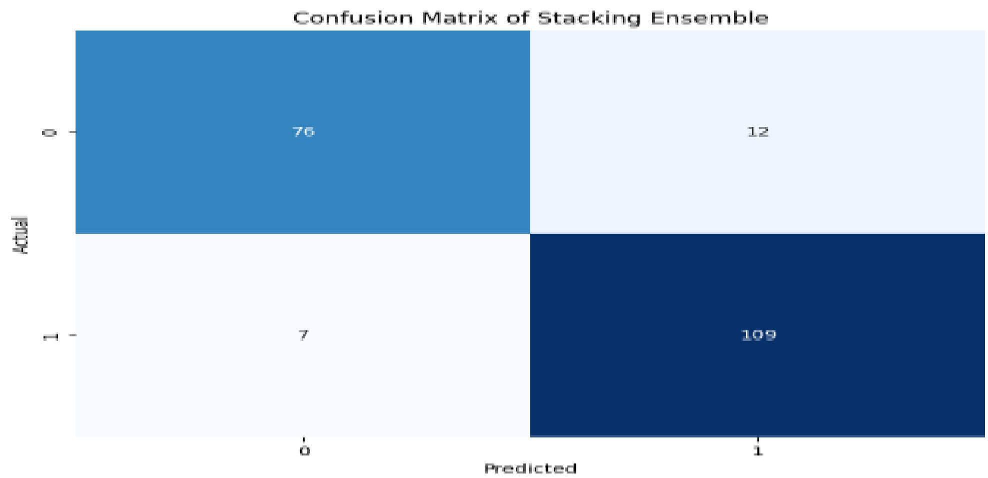

Figure 5: *Confusion Matrix of Stacking Classifier*

The stacking classifier combines the results of nine ML classifiers. These models include LR, DT, RF, XGBoost, Gradient Boosting, AB, SVM, KNN, and NB. Each of these base models learns from the same dataset, and their predictions are combined by a meta-model. In this case, Random Forest is used as the meta-model. The goal is to enhance accuracy by blending the strengths of multiple models. The stacking classifier obtained an accuracy of 89.72% on the TEST data. This is a notable improvement compared to individual models, as stacking allows for better performance by reducing the limitations of any single model. The RF meta-model helped to boost the overall prediction accuracy. It takes advantage of the diverse predictions from the base models. This method is often more effective than relying on a single classifier.

In this study, we applied and evaluated four ensemble classifiers: Stacking, Bagging, Voting, and Random Subspace. These techniques combine the predictions of multiple models to improve accuracy and robustness. Each ensemble method offers a unique way of utilizing the strengths of individual classifiers to achieve better results. The accuracy of each used ensemble model is shown below in Table 5 to reflect the noticeable difference between the used models.

Table 5: *Enhanced Accuracy of ML Ensemble Techniques*

SR#	ENSEMBLE CLASSIFIER	ACCURACY
1	Stacking Ensemble	0.892156863
2	Bagging Ensemble	0.887254902
3	Voting Ensemble	0.887254902
4	Random Subspace	0.87254902

The Stacking Ensemble achieved the highest accuracy of 89.22%. It combines predictions from base classifiers using a meta-model that makes the final decision. This approach uses the strengths of all base models, making Stacking the most effective method in this study. Bagging Ensemble and Voting Ensemble both achieved an accuracy of 88.72%. Bagging trains multiple versions of the same model on different subsets of data and averages their predictions. This reduces variance and improves stability. Voting Ensemble combines outputs from different models and predicts based on majority voting. Its diversity of models enhances performance.

The Random Subspace method achieved an accuracy of 87.25%. It trains classifiers on random subsets of features, capturing different perspectives of the data. Although its accuracy is slightly lower, it still delivers strong and reliable results. These findings show the power of ensemble classifiers in heart disease prediction. Stacking emerged as the best technique, achieving the highest accuracy by leveraging base classifiers' strengths. Bagging and Voting also performed well, showing their reliability. Ensemble methods reduce errors, improve generalization, and are essential for complex datasets.

Figure 5 shows a SHAP (SHapley Additive exPlanations) summary plot. It highlights the impact of different features on heart disease predictions. SHAP values represent how each feature influences the model's output. Positive SHAP values indicate a higher likelihood of heart disease, while negative values indicate a lower likelihood. The features are ranked by importance. The most influential features, like ST slope_2, chest pain type_4, and ST slope_1, are at the top. The color gradient (blue to red) represents feature values. Blue shows low values, and red shows high values. This plot provides insights into how features contribute to predictions. For example:

- High values of *ST slope_2* and *chest pain type_4* have a significant positive impact on predicting heart disease.
- Low values of *cholesterol* and *max heart rate* are associated with a lower likelihood of heart disease.

Each dot represents one instance in the dataset. The horizontal spread of points shows how much a feature's contribution varies across instances. Features with wider spreads have more variability in their impact. The plot highlights the most important features for the model. For example, ST slope and chest pain type are highly influential. This shows their importance in diagnosing heart disease. The visualization helps explain the model's predictions. It shows which features have the most influence and how they drive the model's decisions.

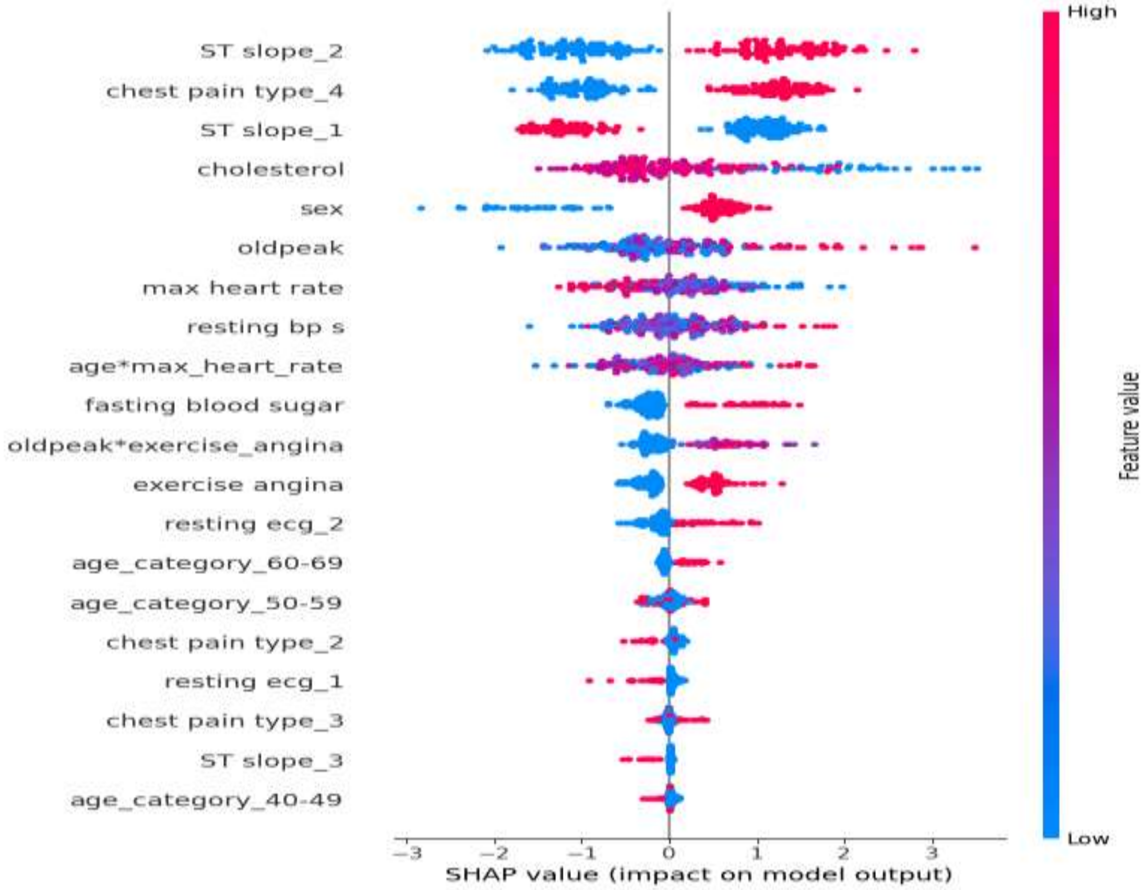

Figure 5: *SHAP Values (Impact on Model Output)*

V. Conclusion

This study shows that Stacking Ensemble classifiers are effective for predicting heart disease. By combining different machine learning models, it uses their strengths to make accurate predictions. The Stacking Ensemble outperformed other methods, achieving 89.22% accuracy. This proves the power of ensemble learning in handling complex healthcare data. The Stacking Ensemble uses a mix of base models and a meta-model for the final prediction. This helps balance out the weaknesses of one model with the strengths of others. The evaluation metrics, like precision, recall, and F1-score, show that the model makes balanced and precise predictions. This makes Stacking a great choice when accuracy and reliability are important. When compared to other ensemble methods like Bagging, Voting, and Random Subspace, Stacking performed the best. Though these methods also worked well, Stacking was better at adapting to the dataset's complexities. Its higher accuracy and consistent performance prove its ability to handle data variations.

In conclusion, the Stacking Ensemble classifier is a strong and effective tool for heart disease prediction. It combines multiple models into one framework, making it ideal for healthcare. This study highlights how ensemble techniques can improve accuracy, reduce errors, and support important decisions in healthcare.

References:

1. Abdellatif, A. H. Abdellatef, J. Kanesan, C.-O. Chow, J.H. Chuah, H.M. Gheni, An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods, IEEE Access 10 (2022) 79974–79985.
2. Abed-Alguni BH & Barhoush, M 2018, ‘Distributed grey wolf optimizer for numerical optimization problems’, Jordanian J. Comput. Inf. Technol. (JJCIT), vol. 4, no. 3, pp. 130-149.
3. Ahsan, M. M. and Z. Siddique, “Machine learning-based heart disease diagnosis: A systematic literature review,” Artificial Intelligence in Medicine, vol. 128, pp. 102289, 2022.
4. Al Reshan, M.S.; Amin, S.; Zeb, M.A.; Sulaiman, A.; Alshahrani, H.; Shaikh, A Networks.

IEEE Access 2023, 11, 121574–121591. [CrossRef]
5. Alfaidi, A. R. Aljuhani, B. Alshehri, H. Alwadei, S. Sabbeh, Machine learning: assisted cardiovascular diseases diagnosis, Int. J. Adv. Comput. Sci. Appl. 13 (2022).
6. Ali, L. A. Rahman, A. Khan, M. Zhou, A. Javeed, J.A. Khan, An automated diagnostic system for heart disease prediction based on statistical model and optimally configured deep neural network, IEEE Access 7 (2019) 34938–34945.
7. Alliance c.h, IEEE standards association and continua health alliance join forces to develop end-to-end, plug-and-play connectivity for personal connected health, Continua Health Alliance, 2013. [Online]. Available: http://standards.ieee.org/news/2013/ieeesa continua.html
8. Alotaibi, F. S. "Implementation of Machine Learning Model to Predict Heart Failure Disease," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 10, no. 6, Art. no. 6, 29 2019, doi: 10.14569/IJACSA.2019.0100637.
9. Alsaeedi AH, Aljanabi, AH, Manna, ME & Albukhnefis, AL 2020, A proactive metaheuristic model for optimizing weights of artificial neural network, Indonesian Journal of Electrical Engineering and Computer Science, vol. 20, pp. 976-984.
10. Ambesange, S., Vijayalaxmi, A., Sridevi, S., Yashoda, B. S. (2020). Multiple heart diseases prediction using logistic regression with ensemble and hyper parameter tuning techniques. In 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), pp. 827-832. https://doi.org/10.1109/WorldS450073.2020.9210404
11. Ambrish G., Ganesh, B., Ganesh, A., Srinivas, C., Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. Global Transitions Proceedings, 3(1): 127-130. https://doi.org/10.1016/j.gltp.2022.04.008
12. Ananey, D., Obiri and E. Sarku, "Predicting the Presence of Heart Diseases using Comparative Data Mining and Machine Learning Algorithms," IJCA, vol. 176, no. 11, pp. 17–21, Apr. 2020, doi: 10.5120/ijca2020920034.
13. Archana K, S Elangovan, Survey of classification techniques in data mining, International Journal of Computer Science and Mobile Applications, vol. 2, pp. 65-71, 2014.
14. Atallah,R. and A. Al-Mousa, Heart disease detection using machine learning majority voting ensemble method, in: Proceedings of the Second International Conference on New Trends in Computing Sciences (ICTCS), 2019, 1–6.
15. Bahrami B., and Mirsaeid Hosseini Shirvani, February 2015, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST), ISSN: 3159- 0040, Vol. 2, Issue 2, pp. 164-168.
16. Bashir, S., Khan, Z.S., Khan, F.H., Anjum, A., Bashir, K. (2019). Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 619-623. https://doi.org/10.1109/IBCAST.2019.8667106
17. Bataineh,A. Al S. Manacek, MLP-PSO hybrid algorithm for heart disease prediction, J. Pers. Med. 12 (2022) 1208
18. Bertsimas, D. L. Mingardi and B. Stellato, "Machine learning for real-time heart disease prediction," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3627–3637, 2021.
19. Bhargava N, G. Sharma, R. Bhargava, and M. Mathuria, Decision tree analysis on j48 algorithm for data mining, Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3,no. 6, 2013.
20. Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. Computational Intelligence and Neuroscience, 2021: 8387680. https://doi.org/10.1155/2021/8387680
21. Boukhatem, C.; Youssef, H.Y.; Nassif, A.B. Heart disease prediction using machine learning. In Proceedings of the 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 21–24 February 2022

22. Cenitta, D. R.V. Arjunan, K. Prema, Ischemic heart disease multiple imputation technique using machine learning algorithm, Eng. Sci. 19 (2022) 262–272.
23. Chandrasekhar,N. S. Peddakrishna, Enhancing heart disease prediction accuracy through machine learning techniques and optimization, Processes 11 (2023) 1210.
24. Charles,V.B. D. Surendran, A. SureshKumar, Heart disease data based privacy preservation using enhanced elgamal and resnet classifier, Biomed. Signal Process. Control 71 (2022), 103185.
25. Chaurasia V, Early prediction of heart diseases using data mining techniques, 2017.
26. Chen Ly, N, C, Qiu, T & Sangaiah, AK 2018, Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in SAR images, Transactions on Industrial Informatics, IEEE, vol. 14, pp. 5530-5538.
27. Chen, M., Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease Prediction by Machine Learning Over Big Data From Healthcare Communities,” IEEE Access, vol. 5, pp. 8869–8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
28. Chhabbi A., Lakhan Ahuja, Sahil Ahir, and Y. K. Sharma,19 March 2016,“Heart Disease Prediction Using Data Mining Techniques”, International Journal of Research in Advent Technology,E-ISSN:23219637,Special Issue National Conference “NCPC-2016”, pp. 104-106.
29. Chowdary, K & Bhargav, P & Nikhil, N & Varun, K & Jayanthi, “Early heart disease prediction using ensemble learning techniques”. Journal of Physics: Conference Series. 2325. 012051. 10.1088/1742-6596/2325/1/012051., 2022.
30. Cinetha K, and Dr. P. Uma Maheswari, Mar.-Apr. 2014,“Decision Support System for Precluding Coronary Heart Disease using Fuzzy Logic.”, International Journal of Computer Science Trends and Technology (IJCST), Vol. 2, Issue 2, pp. 102-107.
31. Das H Naik, B & Behera, H 2020, ‘Medical disease analysis using neuro-fuzzy with feature extraction model for classification’, Informatics in Medicine Unlocked, vol. 8, P. 100288.
32. Deepika, D. N. Balaji, Effective heart disease prediction using novel mlp-ebmda approach, Biomed. Signal Process. Control 72 (2022), 103318.
33. Denison D.G , B. K. Mallick, and A. F. Smith, A bayesian cart algorithm, Biometrika, vol. 85, no. 2, pp. 363-377, 1998.
34. Dilli, M. Babu and M. Sambath, “Heart disease prognosis and quick access to medical data record using data lake with deep learning approaches,” International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 3s, pp. 292–300, 2023.
35. Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. Materials Today: Proceedings, 37: 3213-3218. https://doi.org/10.1016/j.matpr.2020.09.078
36. Dolatabadi Davari, A., Khadem, S. E., & Asl, B. M. (2017). Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Computer Methods and Programs in Biomedicine, 138, 117–126. https://doi.org/10.1016/j.cmpb.2016.10.011
37. Doppala, B.P.; Bhattacharyya, D.; Janarthanan, M.; Baik, N. A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques. J. Healthc. Eng. 2022, 2022, 2585235. [CrossRef] [PubMed]
38. Durairaj M, and V. Ranjani, Data mining applications in healthcare sector: a study, International journal of scientific & technology research, vol. 2, no. 10, pp. 29-35, 2013.
39. Dutta, A. T. Batabyal, M. Basu and S. T. Acton, “An efficient convolutional neural network for coronary heart disease prediction,” Expert Systems with Applications, vol. 159, pp. 113408, 2020.
40. Dutta, A. T. Batabyal, M. Basu, S.T. Acton, An efficient convolutional neural network for coronary heart disease prediction, Expert Syst. Appl. 159 (2020), 113408.
41. Fitriyani , NL, Syafrudin, M, Alfian, G & Rhee, J 2020, HDPM: An effective heart disease prediction model for a clinical decision support system, Access, IEEE, vol. 8, pp. 133034-

133050.

42. Gaidhane PJ& Nigam, MJ 2018, A hybrid grey wolf optimizer and artificial bee colony algorithm for enhancing the performance of complex systems, Journal of Computational Science, vol. 27, pp. 284-302.
43. Golande, A. "Heart Disease Prediction Using Effective Machine Learning Techniques," vol. 8, no. 1, 2019.
44. Govinda,P. moorthi, P. Ranjith Kumar, A likelihood swarm whale optimization based LeNet classifier approach for the prediction and diagnosis of patients with atherosclerosis disease, Comput. Methods Biomech. Biomed. Eng. 26 (2023) 326–337.
45. Gudadhe M., Kapil Wankhade, and Snehlata Dongre, Sept 2010, "Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network", International Conference on Computer and Communication Technology (ICCCT), DOI: 10.1109/ICCCT.2010.5640377, 17-19.
46. Haitao G, Z. Qingbao, and X. Shoujiang, Rapid-exploring random tree algorithm for path planning of robot based on grid method, Journal of Nanjing Normal University (Engineering and Technology Edition), vol. 2, no. 14, 2007.
47. Hao,Y. M. Usama, J. Yang, M.S. Hossain, A. Ghoneim, Recurrent convolutional neural network based multimodal disease risk prediction, Futur. Gener. Comput. Syst. 92 (2019) 76–83.
48. Hassani, M.A., Tao, R., Kamyab, M., Mohammadi, M.H. (2020). An approach of predicting heart disease using a hybrid neural network and decision tree. In Proceedings of the 5th International Conference on Big Data and Computing, pp. 84-89. https://doi.org/10.1145/3404687.3404704
49. Herath, H. M. K. K. M. B., Karunasena, G. M. K. B., Priyankara, H. D. N. S., & Madhusanka, B. G. D. A. (2021, June 22). *High-performance Cardiovascular Medicine: Artificial Intelligence for Coronary Artery Disease*. Research Square. https://doi.org/10.21203/rs.3.rs-642228/v1
50. J., S. K., & S., G. (2019). Prediction of heart disease using machine learning algorithms. 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT). https://doi.org/10.1109/iciict1.2019.8741465
51. Jain D & Singh, V 2018, 'Feature selection and classification systems for chronic disease prediction: A review', Egyptian Informatics Journal, vol. 19, pp. 179-189.
52. Jan M., Awan, A. A., Khalid, M. S., & Nisar, S. (2018). Ensemble approach for developing a smart heart disease prediction system using classification algorithms. Research Reports in Clinical Cardiology, Volume 9, 33–45. https://doi.org/10.2147/rrcc.s172035
53. Javeed A, Rizvi, SS, Zhou, S, Riaz, R, Khan, SU & Kwon, SJ 2020, Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification, Mobile Information Systems, vol. 2020, pp. 1-11.
54. Jayaraman V & Sultana, HP 2019, 'Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification', Journal of Ambient Intelligence and Humanized Computing, pp. 1-10.