

# A Stacking Based Ensemble Model for Early Prediction of Cardiovascular Disease

Durga Thokala, Rajasekhar Reddy M\*

*Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amaravati Campus, Andhra Pradesh, 522503, India.*

\*Corresponding authors: [t\\_durga@av.students.amrita.edu](mailto:t_durga@av.students.amrita.edu)

<https://doi.org/10.55674/ias.v15i2.267559>

Received:29 April 2026; Revised:26 June 2026; Accepted:30 June 2026; Available online: 27 August 2026

## Abstract

Cardiovascular disease is remains the first cause of death world-wide and hence advanced prediction models need to be developed to facilitate early detection and risk mitigation. In this work we explore and analyze a machine learning approach to cardiovascular disease prediction based on patient's clinical and ECG information (Age, Cholesterol level, Arterial Pressure measurements and ECG signals). The data set recruited for experiment has 303 samples. Several classification and ensemble learning models were developed and evaluated in this study. For the sake of performance comparison. Feature normalization and class balancing via resampling are also part of the data preprocessing and cross validation. The ensemble models also have better performance evaluated with accuracy, precision and recall compared to single classifier especially stacking models and gradient boosting. The best accuracy of stacking model is 97.33% with AUC-ROC as 99.00%, followed by gradient boosting 97.00% with AUC-ROC as 99.16%. The findings indicate that ensemble-based learning approaches can accurately identify cardiovascular risk which would lead to an effective framework for supporting in patient care and timely risk identification through AI analytics.

**Keywords:** Cardiovascular disease; Machine learning; Ensemble learning

© 2026 Center of Excellence on Alternative Energy reserved

## 1. Introduction

Cardiovascular diseases (CVDs) have a substantial burden on public health systems and an estimation of one-third of the deaths in the world is due to CVDs every year by the World Health Organization (WHO). They are a class of heart diseases and blood vessel diseases and give rise to a significant impact on public health worldwide. According to WHO reports, these diseases are accounting for nearly one-third of annual global deaths. Early and accurate diagnosis plays an important role in effective treatment and prognosis. Recent studies have also investigated biomarkers for early mortality risk stratification in patients with cardiovascular disease [1]. Recently, machine learning has demonstrated considerable capability in supporting cardiovascular disease identification and clinical decision making. By integrating multiple ML algorithms and learning knowledge from massive unstructured healthcare data, a complex hybrid scheme may be realized with high precision to predict CVD.

CVDs are a broad group of diseases, such as coronary artery disease and congestive heart failure, as well as disorders of heart rhythm. Timely and accurate diagnosis is critical to prevent adverse outcomes and reduce the mortality rate. However, traditional methods for diagnosis can be for the most part invasive or require expensive and sophisticated instruments, and many of them are highly reliant on expert interpretation, which may influence their generalizability and applicability. Recently, AI-driven predictive analytics approaches in general have emerged are the most-watched technologies with a great potential to improve the diagnostic accuracy by enabling the early detection of diseases. [2] Machine learning (ML) schemes are gaining a considerable attention in healthcare with the promise of being able to uncover complex hidden relationships within clinical datasets that are difficult to identify using traditional with conventional diagnostics protocols. Numerous investigations have explored data driven techniques for cardiovascular risk prediction. Tiwari A et al. (2022) performed experiments on different algorithms, such as Random Forest, XGBoost and Extra Trees Classifier, on cardiovascular datasets and concluded that the

XGBoost with ensemble stacking classifier achieved the highest accuracy as well as recall, thus it can be used as a reliable prediction model. [3] HEV classifier and feature engineering are also related to the outcome in another paper performance of early detection is improved by the application of classification algorithms including Decision Tree, Random Forest, Support Vector Machines (SVM), and logistic regression, as well as AdaBoost, gradient boosting and extra trees [4]. Predictive accuracy is still hindered by effects such as imbalanced datasets, limited feature relevance, and reduced model interpretability, despite the increasing use of ML-based methods. To deal with these challenges, in this paper, we investigate the effectiveness of a range of solutions for the task from the predictive modelling framework, such as logistic regression, decision tree classifiers, random forests, SVMs and gradient boosting algorithms, as well as a couple of ensemble-based ones.

The dataset employed in this study is composed of 303 patients' records, with the pertinent clinical data such as age, blood cholesterol levels, systolic and diastolic blood pressures, and features derived from ECG. These data preprocessing steps were performed: feature normalization, class balancing by Synthetic Minority Oversampling Technique (SMOTE) approach and cross validation to predict more accurately. The major contribution of this work is to find the best ML model for cardiovascular disease prediction through comparative evaluation among several models on the basis of performance parameters like accuracy, precision, recall, F1-score and AUC-ROC. With the idea of promoting the reliability of cost-effective and AI-assisted diagnostic systems for early cardiovascular risk prediction, this paper considers ensemble approaches and improved feature selection techniques. The outcome of this work appears to offer pragmatic assistance to medical doctors and data scientist to develop automated accurate diagnostic tools for early screening and enhanced patient care.

The key offerings of this study are:

- The present work is validated on the heart disease dataset, the heart data were taken from kaggle.
- This paper presents a novel comprehensive stack-based ensemble learning model which integrates a set of machine learning classifiers.
- Our Framework implemented the preprocessing methods such as normalization, SMOTE based class balancing, Bayesian hyperparameter optimization and cross-validation.
- An empirical analysis among individual and ensemble classifiers towards the best cardiovascular risk predictor is compiled.

### *Literature Survey*

generalized to predictions of cardiovascular disease (CVD), as there are already several studies investigating machine learning-based methods, especially with regard to the advantages of ensemble methods and a systematic data preprocessing, as well as the choice of appropriate models. Earlier research indicates that predictive learning algorithms can significantly improve clinical decision support when appropriately optimized and deployed. Pandey et al. (2024) investigated the association between high-density lipoprotein parameters and the risk of heart failure using a multicohort analysis, highlighting the relevance of clinical biomarkers in cardiovascular risk assessment [5]. Valarmathi and Sheela (2021) investigated heart disease prediction using hyperparameter optimization techniques, including Grid Search, Randomized Search, and Genetic Programming, to optimize Random Forest and XGBoost classifiers. Their study demonstrated the effectiveness of hyperparameter optimization for improving cardiovascular disease prediction performance [6]. In this chapter, Gul et al. (2026) provide a systematic review on machine learning and ensemble techniques for predicting their investigation exposed key performance trends, challenges, and the rising prominence of ensemble-based approaches for enhancing predictive reliability and robustness in cardiac diagnosis [7]. Gupta et al. (2023) also found that ensemble-based approaches that combine various base models consisting of decision trees, logistic regression has better predictive performance than single models. Reported findings suggest that ensemble-learning strategies, including bagging, boosting, and stacking, improve predictive accuracy, model stability, and overall reliability, which makes these methods superior to individual classifiers. In addition, ensemble methods based on Extreme Gradient Boosting (XGBoost) and Adaptive Boosting (AdaBoost) are also very powerful methods dealing with noisy and incomplete medical datasets. [8]. Clinical feature selection and data preprocessing play an important role in improving the quality of cardiovascular disease prediction models. Dubey et al. (2024) investigated the use of clinical attributes and SMOTE-based class balancing to improve coronary heart disease prediction. Their findings demonstrate that addressing class imbalance through SMOTE can enhance the predictive performance and generalization of machine learning models [10]. Relevant research on emerging deep learning architecture has significantly broadened the potential of applications in predicting cardiovascular diseases. Works on CNNs for ECG signal processing have demonstrated potential results for the early detection of disease [11]. Also, deep learning and machine learning methodologies can be combined into a hybrid approach that is claimed to be more effective in predicting over complex medical datasets [12]. EML for Cardiovascular classification using XAI-driven ensemble classifiers outperforms most of the traditional classification algorithms. Besides, [13] had found to be more interpretable, making more trustworthy decisions, by leveraging the XAI

framework. These findings are in line with the recent increases in the relevance of machine learning in healthcare predictive systems. Niveditha et al. have shown that interpretive analysis integrated machine learning models enhanced prediction reliability and clinical decision support in healthcare domain [14]. Raghupathy et al. also illustrated that the introduction of explainable AI to ensembles greatly enhanced model transparency and reliability predictions within medical systems [15]. In general, the above literature clearly provides evidence that means it would be highly advantageous to apply a machine learning technique with ensemble strategies to enhance the prediction accuracy of cardiovascular disease models.

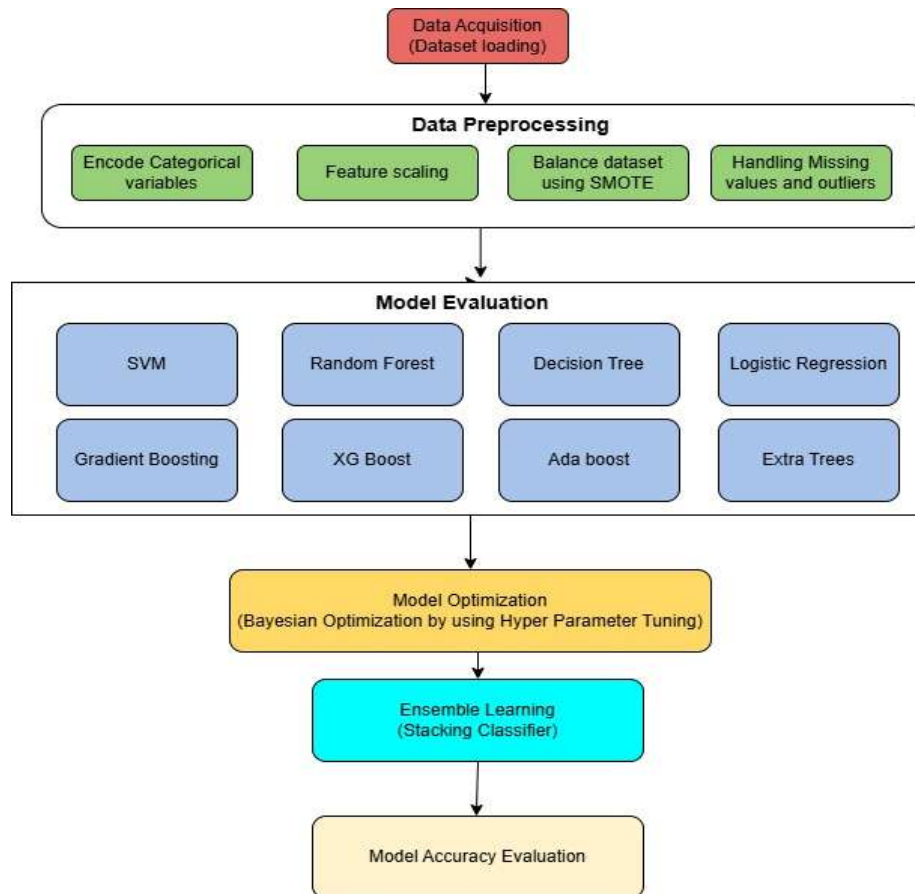
## 2. Materials and Methods

### 2.1 Dataset

A total of 303 patient records are available: the clinical records of 303 patients between 17 and 97 years of age with heart disease were obtained from a publicly available repository. The features include necessary cardiovascular risk indicators including age, gender, chest pain category, blood pressure measurements, and cholesterol levels fasting blood sugar, ECG results, maximum heart rate, exercise-induced angina, exercise-induced ST depression, etc. This database has been extensively utilized in works concerning cardiovascular disease prediction, which guarantees trustability as it holds significant risk factors. To save space, we split the dataset into two subsets: 70% was used for training and for 30% for testing to evaluate the models' performance.

### 2.2 Proposed Methodology

Method: The flowchart of the proposed method is depicted in the methodological diagram in Fig. 1. Prior to the start of the process (i.e., at the data acquisition stage), a cardiac disease-related structured dataset covering all pertinent clinical aspects is obtained. The relevant demographic information and physiological data including diagnostic markers to be utilized during model building are included in the dataset.



**Fig. 1** Workflow of the proposed methodology for cardiovascular disease prediction, comprising data acquisition, preprocessing, model evaluation, optimization, and ensemble-based classification.

The second phase, called Data Pre-processing Phase, makes use of multi-stage procedure of raw data transformation to retain high-quality data for applying machine learning algorithms. Initially, we transform categorical features such as type of chest pain, and thalassemia into numeric values by using one-hot and label encoding, respectively, so that the algorithms can utilize these attributes. Next, min-max normalization is performed for the continuous attributes (age, trestbps, chol, thalach, and oldpeak) to scale the attribute values in the range of [0, 1] as the attributes having larger values may dominate the model training procedure. Subsequently, outlier handling is performed To examine no missing data is clear of outlier for chol and TRESTBPS by using Z-score and interquartile range (IQR) with the exception of wrong data will be discarded, otherwise clinically feasible extreme value will remain. To address unequal class representation, SMOTE was employed to generate a more balanced distribution of samples, which results in the improvement of recall and predictive performance. Finally, because the objective is to be able to predict the generalisation performance of the models, the data is split so that we can both train and test the models we build with alacrity. This configuration enables us to make the input to the machine learning clean, balanced and standardized and then as a result, it fortifies the robustness and interpretability of the proposed cardiovascular disease classification model.

The dataset was checked for missing values, null values, and duplicate values. Categorical variables are converted into numerical values using encoding techniques. Min-max scaling techniques are applied to convert values into a uniform range to prevent dominating nature. Outlier values are present in the data; those are removed using Z-score and interquartile range (IQR) methods. Imbalanced data is addressed using the Synthetic Minority Oversampling Technique (SMOTE) to improve classification performance. Finally, the dataset was divided into a 70% and 30% ratio for training and testing, providing a balanced and standardized dataset for model development and its evaluation.

## 2.2.1 Model Development

Several established supervised learning algorithms be applied for cardiovascular disease prediction based on training and testing data splits. The considered models are SVM, RF, DT, LR, GB, XGBoosting, AdaBoost, and Extra Trees. All models were then entirely optimized and tested for high-accuracy prediction. In order to get better performance and generalize well, ensemble-based methods were used. "Accuracy," "sensitivity," and "F1-score" were the key performance metrics, along with the F1-score, which served as the primary evaluation measure.

### A. Support Vector Machine Classifier (SVM)

The SVM classifier was used in a binary classification scheme to separate the heart disease and non-heart disease subjects. SVM tries to find out a separating hyperplane which segregates two classes maximally in the feature space. Kernel functions are used to generalize the approach, capturing nonlinear relationships between features.

Decision Function (Linear Kernel): Optionally, with linear kernel classification defined as Eq. (1);

$$f() = \text{sign}(w^T x + b) \quad (1)$$

The sign function returns +1 for positive classes and -1 otherwise.

Kernelized Support Vector Machine: For non-linear class separation, kernel functions are employed, and the classification function is expressed as Eq. (2);

$$f(x) = \text{sign}(\sum_i \alpha_i y_i K(x_i, x) + b) \quad (2)$$

where  $\alpha_i$  denotes the large coefficients that represent the contribution weights of the support vectors. The terms  $y_i$  correspond to the binary class labels (1 or 0) associated with support vectors. The function  $K(x_i, x)$  refers to the kernel mapping function, which projects the input data into a higher-dimensional feature space to capture nonlinear relationships (such as linear or RBF kernels), thereby measuring similarity between a support vector used to map inputs into a higher-dimensional space. (linear/RBF as discussed in the notebook) that quantifies the similarity among support vector  $x_i$  and the input sample  $x$ .

### B. Random Forest (RF)

The random forest ensemble learning technique was applied to enhance the accuracy and stability of classification of heart disease. It builds many decision trees from bootstrapped training samples and combines them by aggregating their predictions

through averaging. Using the Random Forest model, the prediction for a single observation is mathematically equal to the average of the scores of all the trees Eq. (3);

$$\hat{y} = \text{average}\{h_t(x): t = 1 \text{ to } T\} \quad (3)$$

$\hat{y}$  indicates the predicted probability of heart disease,  $T$  represents the total number of trees in the ensemble, and  $h_t(x)$  denotes the prediction generated by the  $t$ -th decision tree for a given input sample  $x$ .

### C. Gradient Boosting

Gradient boosting is an ensemble-based stage-wise additive learning algorithm that aims to build a strong learner by linearly combining weak learners, e.g., decision trees, along a sequence of iterations to result into a robust predictive model. Every next tree tries to minimize the errors of previous trees and it leads to better predictive performance of the final model, particularly in separating heart disease.

### D. XGBoost

The more sophisticated version of gradient boosting known as XGBoost was used to enhance the robustness of prediction even more. It embeds explicit regularization terms in the objective function to control over fitting and improve generalization ability in prediction of cardiovascular diseases. Objective of the optimization can be represented as Eq. (4);

$$\text{Obj}(\Theta) = \sum_i L(y_i, \hat{y}_i) + \Omega(\Theta) \quad (4)$$

$\text{Obj}(\Theta)$  denotes the total objective to minimize,  $L(y_i, \hat{y}_i)$  measures the discrepancy between the actual label  $y_i$  and the predicted probability  $\hat{y}_i$ , and  $\Omega(\Theta)$  is a regularization component that discourages excessively complex models like very deep trees.

### E. AdaBoost

The AdaBoost (Adaptive Boosting) algorithm was also used as an ensemble method to improve the prediction of heart disease. It adjusts itself by the re-weighting of training samples in each iteration according to the classification performance in the last iteration, and it integrates multiple weak classifiers into one strong classifier. Each iteration we train a weak learner with an updated distribution (more focus on the hard examples) over the training set. The weight of a weak learner is inversely proportional to its error rate and is given by Eq. (5);

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1-e_t}{e_t} \right) \quad (5)$$

where  $\alpha_t$  is the weight of  $t$ -th learner and  $e_t$  is its error rate. Base learners with smaller errors are weighted more in the ensemble, and the ensemble predictive power is improved.

### F. Extra Trees

As an alternative approach for heart disease prediction, we employed the Extra Trees (Extremely randomized trees) technique, which is an ensemble of decision trees. Entering the world of decision tree ensembles, it creates a number of trees like Random Forest however it adds extra randomness when building the trees which can result in better generalization. Extra Trees Model Predicts Heart Disease: We calculate Extra Trees model predicted probability of heart disease by taking the voting of each tree in model and average them Eq. (5).

$$\hat{y} = \text{average}\{h_t(x) : t = 1 \text{ to } T\} \quad (6)$$

$\hat{y}$  represents the estimated likelihood that an individual is affected by heart disease.

$T$  refers to the overall count of decision trees present in the ensemble,

$h_t(x)$  denotes the output produced by the  $t$ -th tree for the given patient input vector  $x$ .

### G. Decision Tree

For patients with heart disease, a standard decision tree is used as the baseline model. This approach builds a hierarchical, tree-structured set of decision rules derived from patient attributes.

"The decision tree operates as follows:"

- Recursive Partitioning: The dataset is repeatedly divided according to feature values to form subsets that exhibit greater uniformity with respect to the target class.
- Optimal split selection: At every internal node, the method selects an attribute and corresponding split that most effectively distinguishes patients with and without heart disease, using evaluation measures such as information gain (entropy) or Gini index.
- Stopping Criteria: The partitioning procedure proceeds recursively until predefined termination conditions are satisfied, including constraints like maximum tree depth or a minimum sample threshold at terminal nodes.
- Prediction: For an unseen patient record, the algorithm traverses the tree based on the patient's attribute values until a terminal node is reached, which yields the final class prediction.

### H. Logistic Regression

Logistic regression was used to estimate the likelihood of heart disease from a patient feature vector. The model calculates a linear decision score from the input variables, which is then converted to a probability estimate using a logistic function.

Probability Prediction Formula: "The probability of heart disease presence is modeled as:" Eq. (7).

$$P(y = 1 | x) = 1 / (1 + \exp(-(w \cdot x + b))) \quad (7)$$

Where  $P(y = 1|x)$  denotes the estimated probability of that an individual is affected by cardiovascular disease ( $y = 1$ ) given the input attribute  $x$ .

$w$  represents the weight vector corresponding to patient's attributes.

$x$  denotes the patient's feature representation.

$b$  indicates the bias (intercept) component.

$\exp()$  refers to the exponential function.

### I. Ensemble Model

The Stacking Classifier from `sklearnensemble` was applied to construct the ensemble model. This ensemble is built by combining the predictions of the best base single models (all except the logistic regression) with the logistic regression as the meta-learner. Stacking exploits the complementary strengths of heterogeneous models to improve predictive power and generalization. The original features were included in the training of the meta-learner by setting `pass-through` to `True`.

#### 2.2.2. Hyper Parameter Tuning

Tuning of hyperparameters for all models was conducted using Bayesian optimization with Bayes Search CV from the `scikit-optimize` library. Bayesian optimization efficiently traverses the hyperparameter space by evaluating a sequence of potential hyperparameter configurations as predicted by a probabilistic model. In the optimization, a 5-fold stratified cross-validation (Stratified Fold, `n_splits=5`) was used, with the accuracy score as the optimization metric.

Hyperparameter tuning was performed with different values to improve the performance of the planned model using `bayessearch` CV. Parameters such as the number of estimators, maximum tree depth, minimum samples split, and learning rate are considered for optimization. Fivefold stratified cross-validation techniques were used during the training process for improving model performance. Performance metrics like accuracy, precision, recall, and f1-score are considered for model comparison. Accuracy is used as the main function for selecting the best model. For Support Vector Machines, the regularization parameter ( $C$ ), kernel type, and gamma values were tuned.

**Table 1** Shows parameter configuration for prediction models.

| Algorithm / Classifier     | Selected Parameter Configuration                                                                          |
|----------------------------|-----------------------------------------------------------------------------------------------------------|
| Proposed Stacking Ensemble | Meta-classifier = Logistic Regression, passthrough enabled, base models tuned using Bayesian optimization |
| XGBoost                    | $\eta = 0.2419$ , tree depth = 6, boosting rounds = 103                                                   |
| Gradient Boosting          | $\eta = 0.2905$ , tree depth = 3, boosting rounds = 150                                                   |
| Random Forest              | tree depth = 3, minimum leaf samples = 6, minimum split samples = 11, trees = 32                          |
| AdaBoost                   | $\eta = 0.2662$ , boosting rounds = 127                                                                   |
| Extra Trees                | tree depth = 7, trees = 30                                                                                |
| Decision Tree              | tree depth = 6, minimum split samples = 8                                                                 |
| Logistic Regression        | regularization coefficient (C) = 4.1069                                                                   |

The stacking ensemble got put together from the base learners after bayesian optimization finished on them. Logistic regression worked as the meta learner and passthrough was enabled to use both original features and the predictions from the bases. They kept the random seed at forty two to keep reproducibility.

### 2.2.3. Evaluation Metrics

The following assessment criteria were used to evaluate each base model's and the stacking ensemble's overall performance on the test data:

**Accuracy:** Accuracy gives the overall correctness of the classifier by estimating how many samples are assigned to the correct category out of the total number of evaluated instances Eq. (8);

$$\text{Accuracy} = \frac{\text{Correctly Classified Instances}}{\text{All Tested Instances}} \quad (8)$$

**Precision:** Precision indicates the quality of positive predictions by measuring the fraction of instances identified as positive that truly belong to the positive class Eq. (9);

$$\text{Precision} = \frac{\text{Correctly Identified Positive Samples}}{\text{All Samples Predicted as Positive}} \quad (9)$$

**Recall:** Recall evaluates the sensitivity of the model by determining how effectively it detects actual positive cases present in the dataset Eq. (10);

$$\text{Recall} = \frac{\text{Correctly Identified Positive Samples}}{\text{Total Actual Positive Samples}} \quad (10)$$

**F1-Score:** The F1-score offers a single, consolidated metric by integrating both precision and recall their harmonic mean, thereby providing a balanced measure of a classifier's performance Eq. (11);

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (11)$$

**AUC-ROC:** Refers to the area under the receiver operating characteristic curve and serves as an overall indicator of model performance, reflecting how well a model differentiates between positive and negative outcomes.

**Cross-Validation Accuracy:** The well-generalizing model was evaluated using the mean accuracy from the five-fold cross-validation on training data. For each trained model, we choose an optimal probability threshold to transform predicted probabilities into class labels by maximizing the F1-score on the precision-recall curve. Subsequently, this threshold was employed to produce the final predictions for accuracy, precision, recall, and F1-score. Some supervised machine learning

algorithms were used in this work to predict cardiovascular disease. Each of the following sub-sections describes a model and how it is customised to the particular requirements of the proposed framework.

### 3. Results and Discussions

#### 3.1 Environment Definition

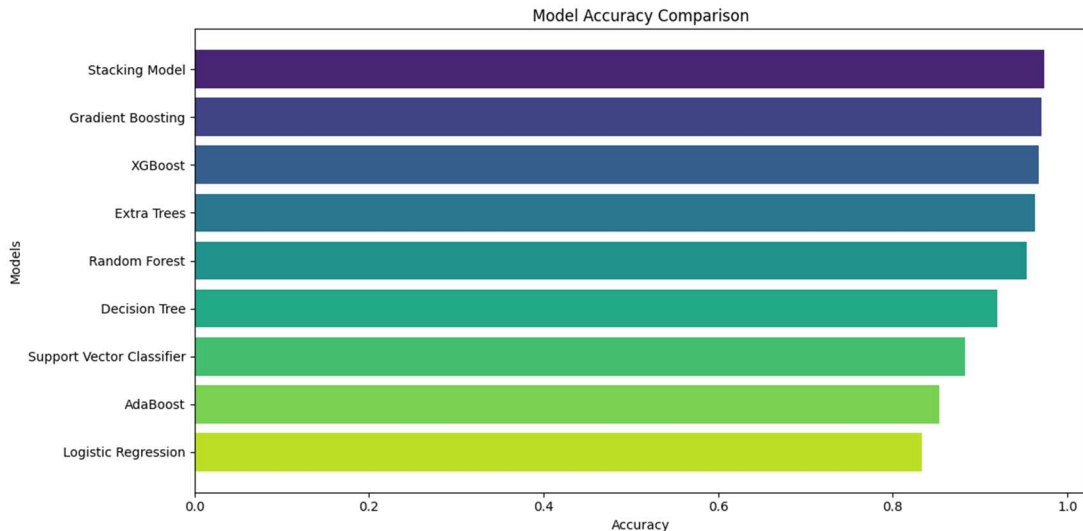
The performance was done on “500 GB SSD”, “1TB HDD”, “8GB RAM” and “Intel Core i5 8th generation” @ 2.4-3.0 GHz. All of our experiments use Python based machine learning libraries such as Scikit-learn, XGBoost, Pandas, NumPy, Matplotlib, and Scikit-learn. The same five-fold stratified cross-validation method is used to test all the models on the same train and test sets.

#### 3.2 Result

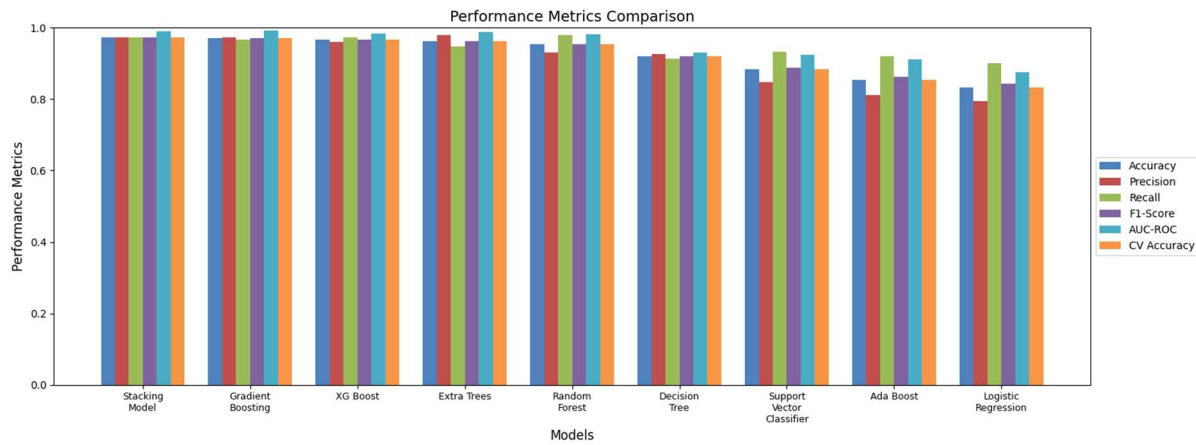
Among all evaluated approaches, stacking and gradient boosting demonstrated the strongest predictive performance for cardiovascular disease assessment. With 97.33% and 99.00% as AUC-ROC value, stacking model has been the best accuracy for the overall performance, and gradient boosting is the second best with 97.00% as accuracy and 99.16% as AUC-ROC. XGBoost and Extra Trees also performed well, unlike the traditional single classifiers. On the other hand, standard single-model methods such as the decision tree-based classifier, the support vector machine-based classifier, and logistic regression had relatively low predictive performances. These findings indicate that ensemble learning methods are better suited for clinical risk prediction and that the combination of multiple models for integration can lead to further robustness and stability. The metric performance comparison in terms of graph is presented in Fig. 2 and Fig.3.

**Table 2** Shows comparison table.

| Classifier                      | Classification Accuracy | Precision Score | Recall Rate | F1 Measure | Area Under ROC Curve |
|---------------------------------|-------------------------|-----------------|-------------|------------|----------------------|
| Stacking-Based Ensemble Model   | 0.9733                  | 0.9733          | 0.9733      | 0.9733     | 0.9900               |
| Gradient Boosting               | 0.9700                  | 0.9732          | 0.9667      | 0.9699     | 0.9916               |
| Extreme Gradient Boosting       | 0.9667                  | 0.9605          | 0.9733      | 0.9669     | 0.9836               |
| Extra Trees                     | 0.9633                  | 0.9793          | 0.9467      | 0.9627     | 0.9887               |
| Random Forest-based Classifier  | 0.9533                  | 0.9304          | 0.9800      | 0.9545     | 0.9809               |
| Decision Tree Classifier        | 0.9200                  | 0.9257          | 0.9133      | 0.9195     | 0.9307               |
| Support Vector-based Classifier | 0.8833                  | 0.8485          | 0.9333      | 0.8889     | 0.9249               |
| Ada Boost                       | 0.8533                  | 0.8118          | 0.9200      | 0.8625     | 0.9124               |
| Logistic Regression             | 0.8333                  | 0.7941          | 0.9000      | 0.8438     | 0.8748               |

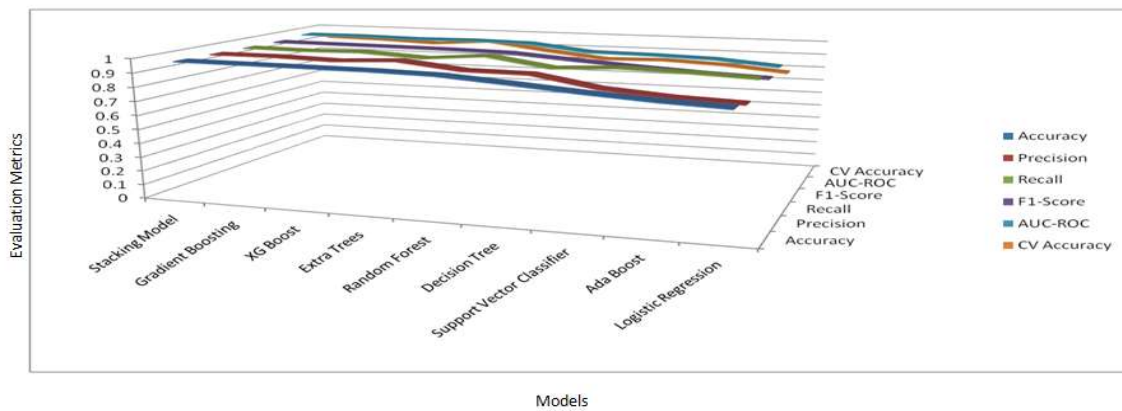


**Fig. 2** Performance of comparison models.



**Fig.3** Performance metrics for models.

This bar chart illustrates a comparative evaluation of multiple machine learning approaches applied for heart disease classification, evaluated using performance indicators including accuracy, precision, recall, F1-measure, AUC-ROC, and cross-validated accuracy.



**Fig. 4** All performance metrics.

Fig. 4 Presents our 3D visualization of the comparative analysis for different machine learning algorithms to heart disease classification evaluated by a set of measures (i.e., accuracy, precision, recall, F1-measure, AUC-ROC, and cross-validated accuracy). The results indicate that ensemble methods including stacking, gradient boosting and XGBoost consistently outperform single models in all evaluation measures showing that they are more stable and effective for heart disease prediction problems.

#### 4. Comparison Study

To evaluate how well the proposed stacking-based ensemble framework works, we compared its performance with existing methods for predicting cardiovascular disease. The comparison with recent methods is shown in Table 3. The comparison shows that the framework we developed offers better predictive performance than that of other methods. The proposed framework achieves overall classification accuracy of 97.33%, outperforming some classical and ensemble based approaches from literatures. The enhanced performance implies that the stacking method can successfully improve the prediction accuracy by combining multiple classifiers. This result suggests that the proposed model is a promising candidate for early, reliable cardiovascular disease prediction.

**Table 3** Shows an evaluation between the suggested strategy with earlier approaches.

| Research Method          | Dataset name                      | Models/Classifier                                                                                           | Accuracy                |
|--------------------------|-----------------------------------|-------------------------------------------------------------------------------------------------------------|-------------------------|
| V. Shorewala et al. [9]  | Cardiovascular disease dataset    | Neural Network (MLP)<br>Bagged Decision Tree (J48)<br>KNN + Random Forest + SVC (Meta: Logistic Regression) | 75.10%                  |
| A. Alqahtani et al. [16] | Cardiovascular disease dataset    | ML ensemble Stacked                                                                                         | 88.70%                  |
| S. Tuly et al. [17]      | Cleveland                         | Stacked ensemble (Meta-MLP &Base-GB, AdaB, XGB, RF)                                                         | 93.00%                  |
| I.D. Mienye et al. [18]  | Cleveland and Framingham datasets | Ensemble machine learning                                                                                   | 91.00%<br>and<br>93.00% |
| S. Mohapatra et al. [19] | UCI dataset                       | XGB, GB, Stacking                                                                                           | 92.60%                  |
| A. Arun et al. [20]      | Framingham dataset                | Ensemble bagging, Boosting                                                                                  | 92.10%                  |
| Proposed model           | Heart disease dataset             | Ensemble stacking model                                                                                     | 97.33%                  |

Table 3 shows how effective the proposed stacking ensemble framework is when compared to recent methods for predicting cardiovascular disease. The framework achieved an accuracy of 97.33%. This result indicates better predictive ability than several traditional and ensemble-based machine learning algorithms.

This improvement in performance benefits from the stacking ensemble technique, which combines multiple classifiers in an intelligent manner by taking advantages of them and avoiding/minimizing their weaknesses. It can be seen that these results certify that the proposed framework can be a powerful, dependable and efficient tool for early cardiovascular disease prediction and therefore it can be applied to support clinical decision application.

#### 4. Conclusion

In this paper, we propose a stacking-based ensemble method, which utilizes the average of all base classifiers in the predictions to improve the estimation of the precision, with the stacking classifier achieving a magnificent 97.33% classification accuracy. It is superior to individual machine learning models. The study is vital because of the immense requirement of early detection as a ground for timely action or treatment. Leveraging sophisticated ensemble learning methods, the investigation not only enhances diagnostic precision but can also have a profound impact on patient quality of life and decrease in mortality associated with heart disease and is deemed breakthrough in the med-tech arena. The biggest pitfalls of this work is that the dataset is limited in size, hence the model may not generalize to much larger more heterogeneous datasets. The future work will enhance the model by incorporating larger and more dissimilar datasets and by investigating new modalities and means for making the diagnosis extra accurate and clinically useful. This work contrasted the individual ML model and its ensemble variant via "stacking." We also want to extend this comparison to all the ensemble techniques in the coming works such as voting, blending and cascading. The data preprocessing pipeline ie feature scaling, SMOTE-based class balancing and Bayesian search hyperparameter optimization were instrumental to the models' performance enhancement. These results are in line with previous studies which reported that preprocessing and data quality matter when applying machine learning in the health domain. For future work, we plan to investigate the application of deep architectures, such as CNNs, to handle more complex data types like ECG signals, which have been shown to be promising in recent works. Furthermore, to improve interpretability of the models through Explainable AI (XAI) methods and to deploy the developed models in real clinical environment for a trial will definitely represent an important future work. More research on hybrid systems and on-line data may be needed to advance the system's performance so that it can generalize better to novel populations.

#### References

- [1] M. Lei, Y. Li, L. Cheng, N. Tang, J. Song, M. Song, Q. Su, M. Liu, S. Fu, F. Lin, Y. Gao, Stress hyperglycemia ratio as a biomarker for early mortality risk stratification in cardiovascular disease: a propensity-matched analysis, *Cardiovasc. Diabetol.* 24(1) (2025) 286.

- [2] A. Sharma, S. Dhanka, A. Kumar, M. Nain, B. Dhanka, V.K. Bhardwaj, S. Maini, A.S. Arora, A systematic review on machine learning intelligent systems for heart disease diagnosis, *Arch. Comput. Methods Eng.* 32(7) (2025) 4303 – 4329.
- [3] A. Tiwari, A. Chugh, A. Sharma, Ensemble framework for cardiovascular disease prediction, *Comput. Biol. Med.* 146 (2022) 105624.
- [4] H. Mohammedqasim, R.A. Mohammedqasem, O. Ata, E.I. Alyasin, Diagnosing coronary artery disease on the basis of hard ensemble voting optimization, *Medicina.* 58(12) (2022) 1745.
- [5] A. Pandey, K.V. Patel, M.W. Segar, M.D. Shapiro, C.M. Ballantyne, S.S. Virani, V. Nambi, E.D. Michos, M.J. Blaha, K. Nasir, M. Cainzos-Achirica, C. Ayers, B.D. Westenbrink, J.L. Flores-Guerrero, S.J.L. Bakker, M.A. Connelly, R.P.F. Dullaart, A. Rohatgi, Association of high-density lipoprotein parameters and risk of heart failure: a multicohort analysis, *JACC Heart Fail.* 12(7) (2024) 1242 – 1253.
- [6] R. Valarmathi, T. Sheela, Heart disease prediction using hyper parameter optimization (HPO) tuning, *Biomed. Signal Process. Control.* 70 (2021) 103033.
- [7] G. Gul, I.A. Korejo, D.N. Hakro, H. Alqahtani, A. Abbasi, M. Babar, O. Al Rahbi, N.I. Ali, Machine learning and ensemble methods for cardiovascular disease prediction: a systematic review of approaches, performance trends, and research challenges, *Computers.* 15(1) (2026) 25.
- [8] S. Mondal, R. Maity, Y. Omo, S. Ghosh, A. Nag, An efficient computational risk prediction model of heart diseases based on dual-stage stacked machine learning approaches, *IEEE Access.* 12 (2024) 7255 – 7270.
- [9] V. Shorewala, Early detection of coronary heart disease using ensemble techniques, *Inform. Med. Unlocked.* 26 (2021) 100655.
- [10] M. Dubey, J. Tembhurne, R. Makhijani, Improving coronary heart disease prediction with real-life dataset: a stacked generalization framework with maximum clinical attributes and SMOTE balancing for imbalanced data, *Multimed. Tools Appl.* 83(37) (2024) 85139 – 85168.
- [11] S.T. Aarthy, J. Mazher Iqbal, A novel deep learning approach for early detection of cardiovascular diseases from ECG signals, *Med. Eng. Phys.* 125 (2024) 104111.
- [12] M. Swathy, K. Saruladha, A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques, *ICT Express.* 8(1) (2022) 109 – 116.
- [13] P. Guleria, P. Naga Srinivasu, S. Ahmed, N. Almusallam, F.K. Alarfaj, XAI framework for cardiovascular disease prediction using classification techniques, *Electronics.* 11(24) (2022) 4086.
- [14] G. Niveditha, B.U. Maheswari, R.P. de Prado, Elevating maternal healthcare: synergy of cardiotocography, machine learning models and interpretive analysis, *Procedia Comput. Sci.* 258 (2025) 2787 – 2797.
- [15] B.K. Raghupathy, M.R. Reddy, P. Theeda, E. Balasubramanian, R.K. Namachivayam, M. Ganesan, Harnessing explainable artificial intelligence (XAI) based SHAPLEY values and ensemble techniques for accurate Alzheimer's disease diagnosis, *Eng. Technol. Appl. Sci. Res.* 15(2) (2025) 20743 – 20747.
- [16] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, T. Javed, Cardiovascular disease detection using ensemble learning, *Comput. Intell. Neurosci.* 2022(1) (2022) 5267498.
- [17] S. Tuly, S. Majumder, A. Chowdhury, A. Habib, S.A. Suha, Predicting heart disease risk with lifestyle factors: uncovering vital predictors via feature selection and machine learning techniques, *Proceedings of the 3rd International Conference on Computing Advancements.* (2024) 466 – 473.
- [18] I.D. Mienye, Y. Sun, Z. Wang, An improved ensemble learning approach for the prediction of heart disease risk, *Inform. Med. Unlocked.* 20 (2020) 100402.
- [19] S. Mohapatra, S. Maneesha, P.K. Patra, S. Mohanty, Heart diseases prediction based on stacking classifiers model, *Procedia Comput. Sci.* 218 (2023) 1621 – 1630.

- [20] A. Arun, R.P. Singh, S. Kanchan, Predicting the coronary heart disease risk over the next decade through ensemble methodologies, 2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT). (2024) 1 – 6.