

Ensemble Machine Learning for Cardiovascular Disease Prediction

A. R. Deepa*, Venkata Ganji

Department of Computer Science and Engineering, V.K.R, V.N.B & AGK, College of Engineering, Gudivada, Andhra Pradesh

ABSTRACT

Cardiopulmonary disease remains one of the leading causes of mortality worldwide, needs accurate and early prediction systems. The study aims on leveraging machine learning to predict heart disease, applying the XG Boost algorithm for its efficiency and scalability. Generate a synthetic dataset (heart1.csv) and data set contains 14 columns 1025 rows. The dataset used contains key clinical and demographic features, processed through rigorous data preprocessing and feature selection techniques to improve prediction accuracy. To ensure well-balanced spanning reliability, precision, recollection, and F1 score parameters the Voting Classifier model is trained and tested using the Random Forest Algorithm and XG Boost. For applications in healthcare, the method is a dependable option because to its ability to manage missing values, reduce overfitting, and offer interpretable feature importance. Research results show that the Voting Classifier model can predict cardiovascular conditions with excellent accuracy and outstanding performance, surpassing conventional machine learning techniques. These results highlight how predictive algorithms can be used to inform clinical judgment, opening the door to quicker and better diagnosis. For increased usefulness in healthcare environments, future research should investigate real-time deployment and hybrid approaches.

Keywords: Cardiopulmonary Disease, Random Forest, Extreme Gradient Boosting (XG Boost), Cardiovascular Health, Feature Selection, Healthcare AI, Voting Classifier, Risk Assessment, Accuracy, F1-Score, Precision, Recall

INTRODUCTION

Due to the fact 17.9 million people die from heart attacks each year, machine learning (ML) is revolutionizing healthcare by improving the prediction of the onset of coronary heart disease, which is vital for rapid identification. Effective models can significantly enhance patient outcomes and save costs. traditional tools for assessing risk, such as the Middlesex Danger Score., often overlook the complex interplay of risk factors, while ML approaches use large datasets to discover complicated patterns that enhance predictive accuracy and allow for individualized risk assessment in a wide range of clinical and demographic contexts. Recent research demonstrates the effectiveness of various computational learning techniques, specifically randomly generated forests, decision trees, and neural networks, for predicting cardiac disease. As an example, the XG Boost algorithm scored 93% accuracy by detecting major predictors such as age, gender, BMI, and lifestyle. Active learning methods

have even achieved prediction rates of 98.7% through Learning Vector Quantization models. Applying ML to heart disease prediction allows for timely intervention and helps doctors tailor interventions and allocate resources as effectively as possible. Health systems can implement interventions to reduce the incidence of heart disease and enhance overall health outcomes through these new approaches. [1] [2] Because of its high degree of efficiency along with its capacity to manage intricate datasets, Random Forest is a sophisticated composite machine learning method that has become widely recognized for its ability to predict cardiac disease. Accurate prediction models are important either timely treatment or diagnosis because cardiovascular diseases are still a leading cause of death globally. To enhance reliability and reduce the likelihood of too much fitting, random forest creation creates multiple decision trees and averages their assumptions. While cardiac arrest is among those global leading causes of death, accurate forecasting protocols are necessary to assist with early detection. The machine-learning method called

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



Random Forest has proven to be highly effective in predicting the onset of cardiac disease with its ensemble strategy, which combines different decision trees to obtain accurate and reliable results. Research studies are plagued by seemingly Random Forest shows high accuracy, like 92.16% in coronary diseases forecasting, outperforming other algorithms, like selection foliage techniques and Support Vector Machines. By analysing clinical visits and demographic factors dimensions, Random Forest allows identify key the probability factors, making timely therapy and tailored rehabilitation strategy available to maximize receptive outcomes. [3,4 XG Boost (Extreme Gradient Boosting) is a sophisticated machine learning technique that is commonly utilized for predicting cardiac disease due to its effectiveness as well as precision. Investigations have shown that XG Boost may achieve up to 98.04% accuracy through applying sophisticated strategies such as hyperparameter values tweaking and feature selection. Its ability to examine enormous amounts of data and detect substantial risk variables, such as age and way of life constitutes an invaluable tool for medical professionals. XG Boost promotes cardiovascular care through lowering false positives and negatives, facilitating swift actions and tailored treatment options (45). [5]. Heterogeneous classifiers integrate multiple strategies to enhance forecasting heart attacks accuracy. Using strategies like hard voting (majority class selection) and soft voting (probability averaging), classifiers may surpass distinct models (33). Another investigation leveraged 57 electoral votes to achieve 98.38% accuracy, while another framework using six machine learning models achieved 88.70% accuracy by focusing on key parameters including cholesterol levels and resting blood pressure. [6]. Machine learning, a niche branch of Artificial Intelligence (AI), is all about making machines capable of replicating human abilities. Machine Intelligence is the quality of such systems where they can process and utilize data. We utilize biological variables such as cholesterol, blood pressure, gender, and age as data sets to compare the two algorithms' accuracy: XG Boost and Random Forest in this paper. Heart disease is a top cause of mortality across the globe, with the World Health Organization attributing 12 million deaths annually to cardiovascular conditions. Early diagnosis greatly reduces complications, reminding us that prevention

is better than cure. Using machine learning, we seek to forecast heart disease by examining different patient characteristics and comparing algorithmic performance to identify the best predictive model.

LITRATURE REVIEW

This study investigates how data science might be used in the medical field to anticipate cardiac illness. The prediction's dependability still has to be increased because there is a lot of research being done on that subject. Therefore, the goal of this work is to improve precision through feature selection approaches and methods that use a lot of data sets for cardiac disease in experimental computation. We propose an innovative approach for determining key features through the application of machine learning methods, enhancing the accuracy of the prediction of cardiovascular disease. The forecasting framework is proposed through various configurations of features and familiar methods of classification. In this dissertation, they explore the widely employed classification techniques in the medical data set that help predict cardiac diseases, which are the primary cause of death across the globe. Predicting a cardiovascular attack is challenging for physicians and clinicians to venture into since the process involves acumen as well as understanding. The healthcare industry today holds latent but consequential information for decision-making. The tests conducted uncover this algorithm. As expected. The study report offers a stacking ensemble model termed NCDG for heart disease prediction, which uses Naive Bayes, Categorical Boosting, and Decision Tree as base learners and Gradient acceleration as the meta-learner. The model addresses data class imbalance using SMOTE techniques and achieves high performance metrics, including an accuracy, F1-Score, precision, and recall of 0.91. The K-Fold Cross-Validation method further validates the model's predictions, demonstrating its effectiveness in early heart disease detection. [7] El-Sofany established a systematic method for forecasting cardiac disease using machine learning, highlighting the necessity of model validation and the advantages of ensemble learning techniques. The dissertation proposed an application for smartphones based on XG Boost for real-time heart disease prediction using raw symptoms, the SF-2 feature subset, and SMOTE data balance. Using the SF-2 feature subset with SMOTE

analysis evidence juxtaposing, this suggested model achieved an accuracy of 97.57%. [8] Rajni Gandhaet's entire research demonstrates the need of employing a strong dataset and multiple machine learning algorithms to efficiently identify cardiac disease, with a focus upon enhancing the diagnostic performance via ensemble learning techniques. This model achieved high accuracy of 98.04% after hyperparameter adjustment. [9] Hossain et al. (2024) [10] investigated the use of machine learning to predict cardiovascular disease (CVD) risk in Bangladesh. Using cross-sectional data, multiple machine learning models were used to identify significant CVD risk factors and evaluate model performance. The work emphasizes the potential of machine learning for early CVD identification and risk assessment, bringing insights into public health policies in Bangladesh. The good accuracy obtained indicates a possibility of application in clinical practice, even though the precise details of the system are not well defined. They proposed with a precision of 98.04%. In 2023, an authoritative assessment put forward a strategy for forecasting heart failure outcomes using Random Forest and XG Boost. The inquiry into the subject recommends incorporating XG Boost and Random Forest models into healthcare systems for enhancing the preciseness of cardiovascular disease foresight, using a Kaggle dataset with tenfold cross-validation. They advocated XG Boost with an accuracy of 91.56% after cross-validation. [11] Hossain MI [12] constructed a strategy for heart disease prediction using concentrated artificial intelligence tactics, and Random Forest achieved 90% accuracy irrespective of all machine learning. According to Halima EL Hamdaoui's research, amalgamating Random Forest with AdaBoost will increase prediction accuracy. This hybrid technique was tested on a heart disease dataset and shown outstanding results compared to individual models. This model manufactures an accuracy of 95.98% for Random Forest alone and 96.16% when utilized together with AdaBoost. [13] Yang L developed a method for studying cardiovascular disease prediction models using random forests. They employed multiple methods to develop prediction model such as multivariate regression model, classification and regression tree(CART), Naive Bayes, Bagged trees ,Ada Boost and Random forest. They employed the multivariate regression model as

reference for performance Evaluation. This model hypothesized precision Gained an AUC score of 0.787, suggesting good prediction capabilities in relation to other models. [14]. This study offers a predictive framework for heart failure that employs k-mode clustering with Huang initialization to improve classification accuracy. Models such as Random Forest, Decision Tree, Multilayer Perceptron, and XG Boost were tuned with GridSearchCV and implemented to a Kaggle dataset of 70,000 incidences (80:20 split). The highest accuracy was achieved by Multilayer Perceptron with cross-validation (87.28%), outperforming others such as Random Forest (87.05%), XG Boost (86.87%), and Decision Tree (86.37%). The areas under the curve (AUC) values for every single model consisted between 0.94 and 0.95, showing high predictive performance. [15]. Shamsuddin Sultan presents a stacking ensemble model named NCDG for heart disease classification, utilizing Naive Bayes, Categorical Boosting, and Decision Tree as base learners, with Gradient Boosting as the meta-learner. This framework uses SMOTE and BorderLine SMOTE techniques to address issues with data class imbalance. It demonstrated its effectiveness in predicting heart illness by producing exceptional results in metrics like as accuracy, F1-Score, precision, and recall of 0.91 each, which were confirmed by K-Fold Cross-Validation. [16] This research utilizes batch classification models in the explicable artificial intelligence (XAI) paradigm to predict heart disease on a 303-example dataset with 14 variables. Methods employed are support vector machine (SVM), k-nearest neighbor (KNN), decision tree (DT), and random forest (RF). The XAI-driven programs have an incredible 99% accuracy, surpassing conventional classification methods and enhancing the validity and understandability of cardiovascular disease diagnosis and prediction [17]. This paper constructs a heart disease classification model by using a group approach with a Stacking structure comprising BiGRU, BiLSTM, and XG Boost. The BiGRU and BiLSTM models serve as basis models for feature extraction from sequential data, while the XG Boost model serves as a meta-model for final classification. The outcomes show that classification accuracy is enhanced by the Stacking method from 0.85 (BiLSTM) to 0.92, verifying its utility in heart disease detection. [18]. This work shows how an ensemble

machine learning approach, specifically a Voting Classifier that incorporates Decision Tree, k-Nearest Neighbors, and Gaussian Naive Bayes classifiers, can effectively classify heart disease. Using a dataset of 70,000 clinical records, the model obtained average accuracies, precision, recall, and F1-scores over 99% through 5-fold cross-validation. The results demonstrate that ensemble models improve cardiovascular disease classification prediction accuracy and reliability, with important ramifications for early intervention and individualized patient care. [19]. In order to improve diagnostic accuracy, the study suggests an ensemble-based deep learning method for classifying heart disease that combines many machine learning classifiers. It obtains a high 98.3% accuracy on a UCI dataset, beating stand-alone models such as AdaBoost, XG Boost, and Random Forest. Incorporating the application of Correlation-based Feature Selection (CFS) enhances the model by filtering appropriate features, hence increasing accuracy, recall, precision, and f1-score. The process significantly leads to accurate cardiovascular health prediction and classification. [20]. The article illustrates the performance of ensemble machine learning methods, in this case the Random Forest classifier, in classifying heart disease. It recorded an extraordinary accuracy of 98.54% and a very close to perfect AUC value of 1.00, which underscores its strong predictive capability. The research compares several algorithms, highlighting the strengths of ensemble methods over individual classifiers, thus improving early heart disease detection and management via optimized prediction from a holistic data set of cardiovascular health metrics. [21]. A study investigation targets cardiovascular disease classification based on community machine learning techniques. For predicting cardiac disorders, it applies a range of methodologies including logistic regression, decision trees, support vector machines, random forests, and multilayer perceptron. The data used was retrieved from Kaggle, and the predictions were enhanced through hyperparameter tuning and voting classifier methodology. The inquiry concludes by comparing the expected performance parameters of the collective system, establishing that it is effective in early detection and treatment of heart-related disorders. [22]. For this study, we use data on cardiovascular disease. This data set contains about 1025 patients and 76 features; we use all sorts of

machine learning and deep learning algorithms to see which one of them has the highest potential to detect potential cardiovascular disease.

Proposed Approach

This section describes the methodologies used for predicting cardiovascular disease. Describes the methodology proposed, which includes six phases. Selecting an adequate dataset for the trial is the first step in the procedure. The cardiovascular disease database serves as the foundation for the preliminary analysis of the study. A number of crucial procedures are included in the preprocessing step prior to model training. A feature acquisition technique is then used to gauge the features' significance, and a number of machine learning and deep learning classification models are used for preliminary explanations. Deep learning techniques for the identification of cardiovascular disease are also assessed in this study. Four unstable machine learning predictive models—Random Forest (RF), Extreme Gradient Boosting (XGB), and Voting Ensemble Classifier—are used to detect cardiovascular disease outbreaks. Two distinct machine learning classifications are used to assess how well machine learning models perform on the given dataset.

Heart disease dataset description

Researchers at a University of California, Irvine (UCI) online data exploration and machine learning repository provided Cleveland's cardiovascular illness dataset for our research. Six of the 303 subject record instances in the sample contained missing class values. Although each person in the dataset has 76 variables, previous research has shown that 13 criteria are useful in identifying heart disease. We list the dataset's numerical and categorical properties in Table 1. Its motive is to predict whether a subject has heart ailments based on the results from the numerous medical tests that have been conducted on them. The dataset's "num" field indicates whether an individual has heart disease or not. The values of the "num" variable vary between 0 (no existence) to 4. Previous research on the the city of Cleveland dataset has tried to discriminate between the presence (values 1, 2, 3, and 4) and absence (value) of cardiac disease.

Table 1: Features of the data collection on coronary disease.

Variable	Description
age	Age in years (29 to 77)
Sex	Representing the sex of the patient (1 = female, 0 = male).
cp	Representing the type of chest pain experienced by the patient. This is typically categorized as: 0-typical angina 1-atypical angina 2-non-anginal pain 3- asymptomatic
trestpbs	Resting blood pressure in mm Hg
chol	Serum cholesterol in mg/dl
fbs	Fasting blood sugar level, categorized as above 120 mg/dl (1 = true, 0 = false)
restecg	Resting electrocardiographic results: 0: Normal 1: Having ST-T wave abnormality 2: Showing probable or definite left ventricular hypertrophy
thalach	Maximum heart rate achieved during a stress test
Exang	Exercise-induced angina (1 = yes, 0 = no)
Oldpeak	Exercise-induced ST depression in comparison to rest
Slope	Peak exercise ST segment slope: 0: Upsloping 1: Flat 2: Downsloping
Ca	Major vessel count (0–4) as determined by fluoroscopy coloration
thal	Thalium stress test result: 1: Normal 2: Fixed defect 3: Reversible defect
Target	Heart disease status (0 = no disease, 1 = presence of disease)

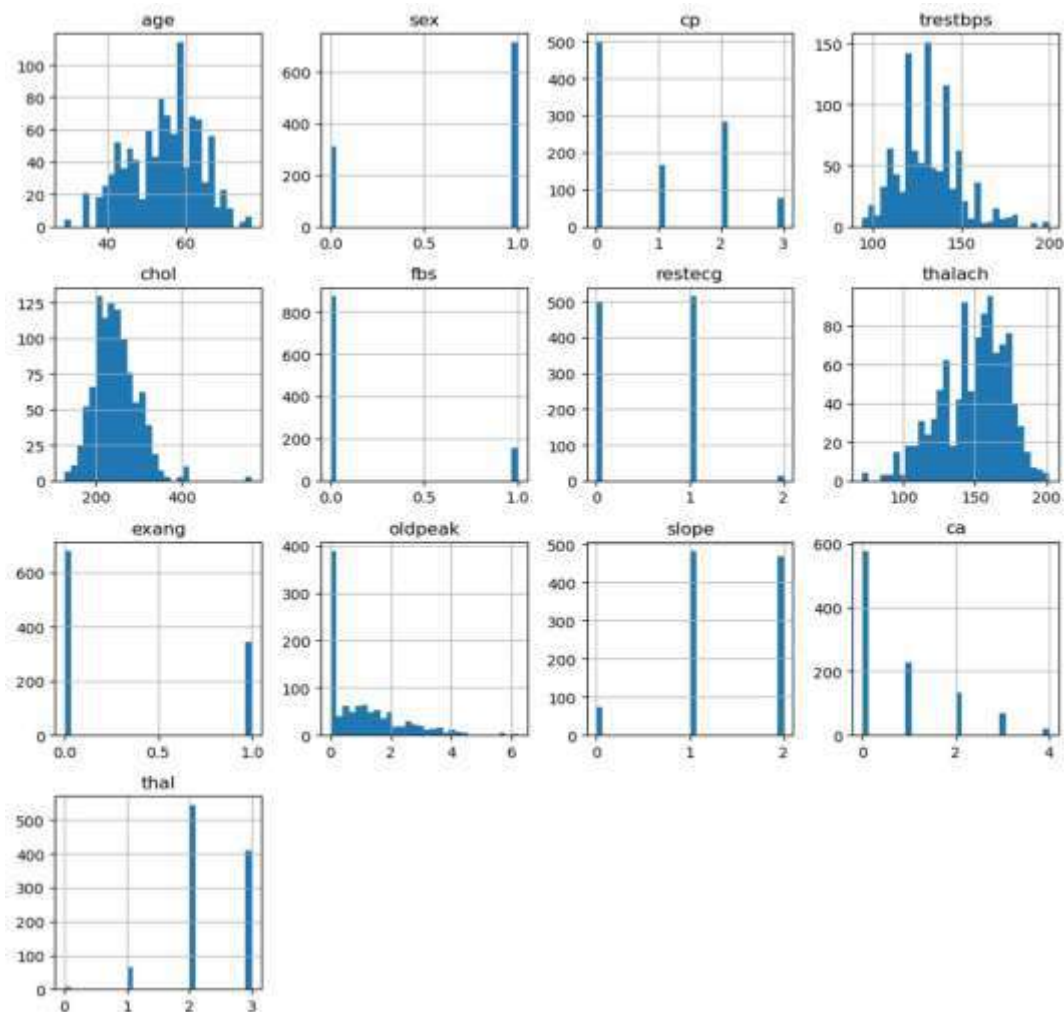


Figure 1: Bar Graph of Heart Dataset

Extreme gradient boosting, or XG boost

The gradient downhill methodology is used by the ensemble tree techniques Gradient Boosting (Friedman, 2001) (GB) and XG Boost (Chen and Guestrin, 2016) (Extreme Gradient Boosting) to improve weak learners. However, XG Boost uses algorithmic enhancements and system optimization to fortify the foundational GB architecture. Chen and Guestrin (2016) conducted the first work, and other developers have since carried it out. One software that

is part of the Distributed Machine Learning Community (DMLC) is called XG boost. According to Friedman (October 2001), GB is a stage-wise additive modelling technique. First, the collected information is fitted to a weak classifier. To improve the performance of the current model, it fits one additional weak classifier without changing the prior one, and so on. The areas where the previous classifiers were falling short must be considered by each new classifier.

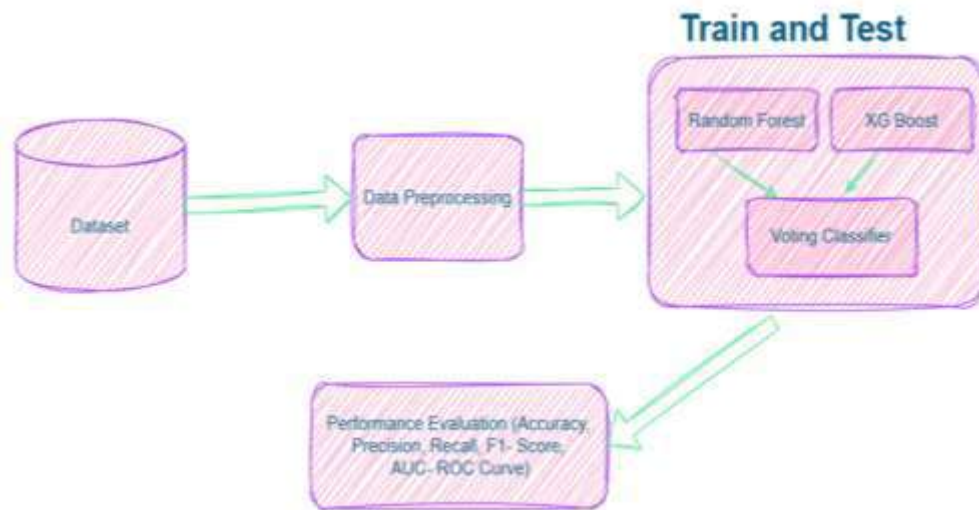


Figure 2: Architecture of Multidisciplinary Approaches for Monitoring Cardiopulmonary Disease

Random Forest

A random forest classifier is a supervised learning technique in machine learning. It can be utilized for regression and classification problems both in machine learning. Collaborative learning, the process of ensembling a multitude of classifiers to solve a tough problem and enhancing the performance of the model, is its basis. Random Forest uses many decision trees on diverse subsets and computes the average outcome to enhance the forecasting accuracy of the dataset. The random forest aggregates the estimates of every decision tree and makes predictions on the basis of the overwhelming majority of votes, as compared to relying on a single one. The more trees in the forest, the more overfitting is avoided and accuracy is improved. A regression task applies the mean of all the outputs and a classification task applies the majority voting classifier to decide the final output.

Voting Ensemble Classifier

Ensemble classifiers that vote by using multiple different models to predict better than one classifier in general. The overall decision-making process within an ensemble that votes is the collection of the individual model predictions. Two types of voting exist, 1. Hard Voting: Once data has been classified by each model, the ultimate forecast for the most voted class is produced.

2. Soft Voting: It is likely the generated probabilities by the models are utilized to decide which class possesses the highest average probability over each of

the models. Voting ensemble methods that reduce overfitting, enhance robustness, and enhance accuracy are stacking, boosting, and bagging (e.g., Random Forest, AdaBoost, and Gradient Boosting). Financial prediction, health diagnosis, and among the numerous disciplines to which these are extensively applied are image recognition.

METHODOLOGY

Data Collection

The heart disease dataset, which includes a variety of clinical characteristics pertaining to heart health, was used in this investigation. Age, sex, type of chest discomfort, ECG findings, peak heart rate attained, resting blood pressure, cholesterol, and fasting blood glucose, exercise-induced angina, previous peak values, slope of ST segment, number of major vessels filled by fluoroscopy, and thalassemia status are some of the attributes present in the dataset.

Data Preprocessing

Importing Libraries: Importing necessary libraries like NumPy, Pandas, Matplotlib, and Seaborn for data manipulation and visualization.

Loading Data: The data is loaded into a Pandas Data Frame for analysis.

Exploratory Data Analysis (EDA): In order to understand the dataset structure and discover any relationships between attributes and the target

variable (presence of heart disease), initial studies are conducted.

Handling Missing Values: It checks for missing values within the dataset, and fortunately, none are found.

Feature Analysis: The relevance of each feature to the target variable is analyzed with the help of bar plots and count plots.

Train-Test Split

The dataset is divided in an 80-20 ratio between training and testing sets so that the model can be tested on data that hasn't been seen yet.

Model Training

The training dataset is used to train the XG Boost model, and hyperparameters are adjusted to improve accuracy and avoid overfitting. When compared to individual decision trees, the random forest model reduces overfitting, performs well with large datasets and missing data, and offers feature priority ranking. It is trained using data sets for training, and its hyperparameters are tweaked to improve accuracy. By combining multiple models, the voting classifier model increases prediction reliability, balances the strengths of distinct classifiers to improve generalization, and performs well with complicated

medical datasets that comprise a variety of patient records. It is learned using training datasets.

Performance Evaluation

The model's performance is assessed using the following metrics- precision, F1 score, accuracy, recall and AUC-ROC.

Prediction and Insights

The trained model predicts heart disease on new data, and insights are drawn by analyzing feature importance scores, enabling better clinical decision-making.

RESULT AND DISCUSSION

Accuracy: Measures the percentage of correct predictions out of the total predictions. With a 98% accuracy rate, the Voting Classifier was the most accurate.

Precision: Represents the percentage of true positive predictions among all positive predictions. Voting Classifier has the highest precision at **99%**

Recall: Also known as sensitivity, it measures the percentage of true positives correctly identified. Voting Classifier excels in recall with **99%**.

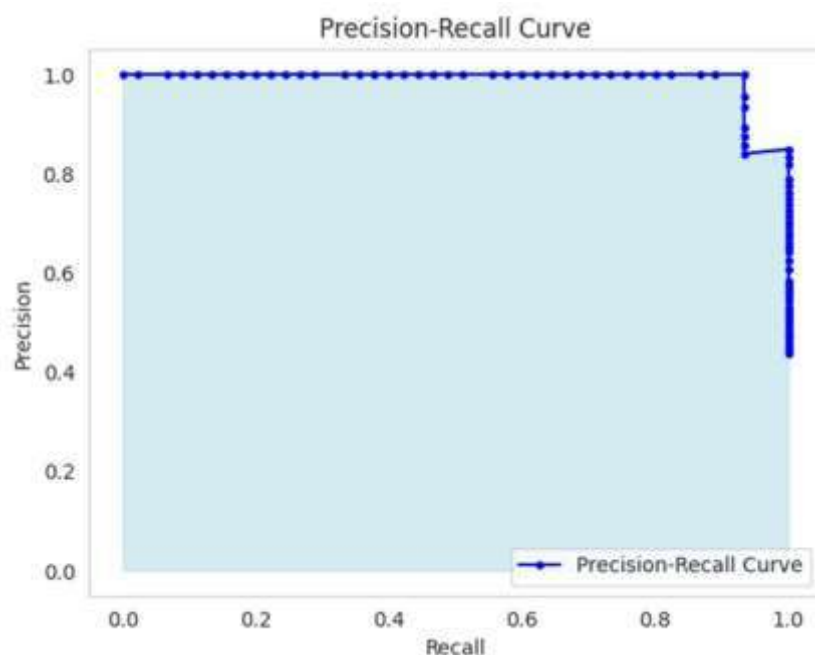


Figure 3: Precision- Recall Curve

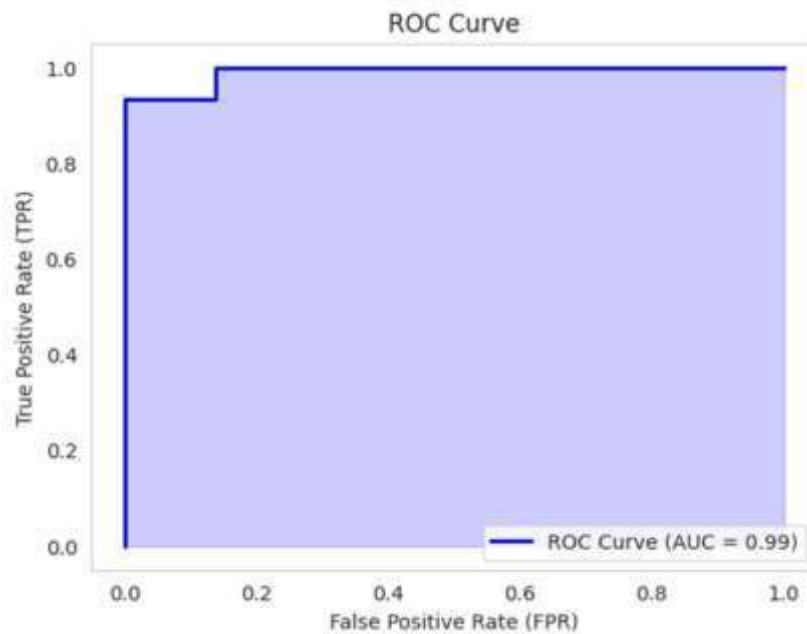


Figure 4: AUC- ROC Curve

These results highlight the superiority of Voting Classifier in terms of predictive performance, especially for heart disease prediction, making it a suitable choice for this task.

Scatter Plot:

The low figure shows the differences between the sample index and predicted probability for training and testing set in the given datasets.

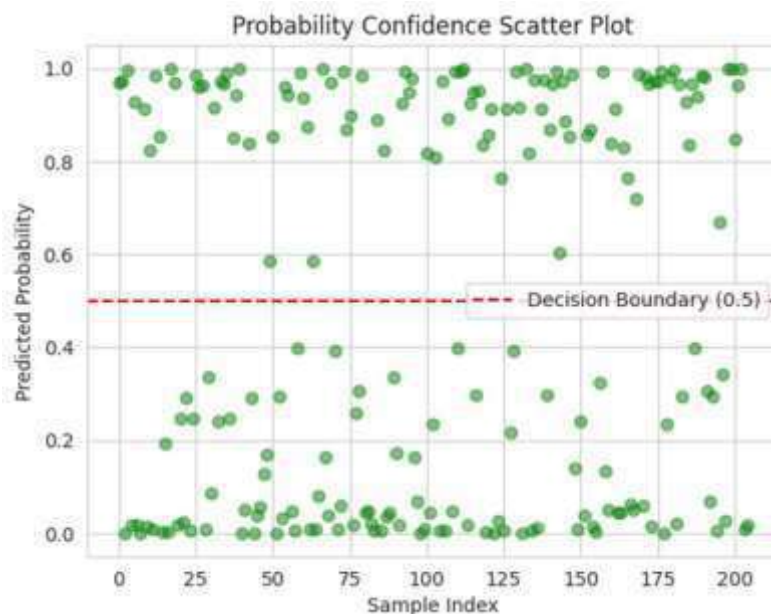


Figure 5: Scatter Plot for Heart Disease Prediction

Feature Importance:

The Underneath represents the features of Heart Disease Prediction in the taken Datasets

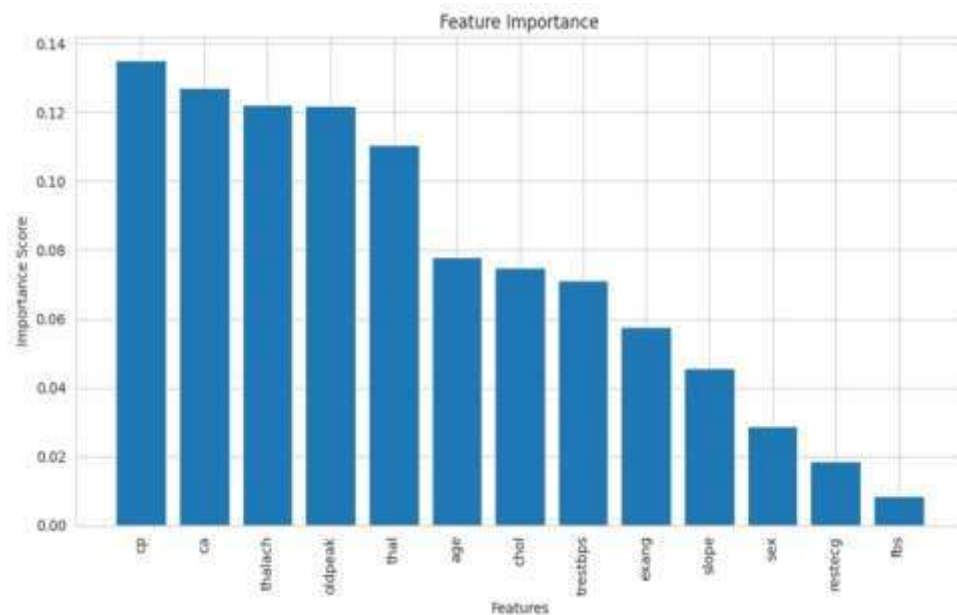


Figure 6: Feature Importance of Heart Disease Prediction

DISCUSSION

Insights from Feature Importance Analysis:

According to the feature importance analysis, the most important variables in predicting cardiovascular illness include their age, lipids, and relaxing blood pressure. This aligns with clinical knowledge, as these factors are known to significantly impact cardiovascular health.

Model Robustness and Limitations:

Voting Classifier's capacity to accurately identify affirmative cases is demonstrated by its great resilience and good performance across all measures, particularly in recall and AUC-ROC. Its computational cost is higher than that of simpler models, though, which might restrict its applicability in resource-constrained real-time applications.

CONCLUSION

This study shows how well the Voting Classifier predicts heart disease based on exceptional performance on several measures, including reliability, recall, F 1 score, precision, and AUC-ROC. Baseline models did not perform better than the Voting Classifier. Relevant clinical factors such as age, blood pressure, and cholesterol levels are critical for accurate predictions, based on feature importance analysis, which highlights their relevance to

cardiovascular health assessments. The Voting Classifier blends numerous machine learning algorithms, like random Forest and XG Boost, to enhance diagnostic precision in heart disease prediction. To enhance reliability, it combines predictions via soft voting (probability averaging) or hard voting (majority vote). Integration of these strategies may broaden the application of heart disease prediction models, promoting more effective and prompt interventions in health systems. This method facilitates risk factor analysis of blood pressure, cholesterol levels, and ECG reports, leading to more accurate and solid heart disease diagnosis. Future work can explore the incorporation of more sophisticated methods, including deep learning algorithms, to improve prediction accuracy. Moreover, processing live data and applying Voting Classifier in online learning settings might enhance the model's flexibility and render it more compatible with dynamic health care environments.

REFERENCE

1. Ansari, U., Soni, J., Sharma, D., & Soni, S. (2011, March). Predictive data mining for medical diagnosis: An overview of heart disease prediction. In Proceedings of the International Conference on Data Mining in Healthcare for Heart Diseases.
2. Beyene, C., & Kamat, P. (2018). Survey on prediction and analysis the occurrence of heart disease using data mining techniques.

- International Journal of Engineering Research and Technology, 118(8), 165–173.
3. Riaz, M. U., Awan, S. M., & Khan, A. (2018, October). Prediction of heart disease using artificial neural network. *International Journal of Advanced Computer Science and Applications*, 9(10).
4. Napa, K. K., Sindhu, G. S., Krishna, D., Prashanthi, & Sulthana, A. S. (2020, April). Analysis and prediction of cardio vascular disease using machine learning classifiers. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*.
5. Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, K. (2018). Prediction of heart disease using machine learning. In *Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*.
6. Mohan, S. K., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 67(5), 861–870.
7. Banu, N. K., & Swamy, S. (2019). Prediction of heart disease at early stage using data mining and big data analytics: A survey. *International Journal of Advanced Research in Computer and Communication Engineering*, 8(4).
8. Krishnan, J. S., & Geetha, S. (2019). Prediction of heart disease using machine learning algorithms. *International Journal of Innovative Technology and Exploring Engineering*, 8(11), 2346–2350.
9. Kaur, P., & Sharma, R. (2018). Heart disease prediction using machine learning: A survey. *International Journal of Advanced Research in Computer Science*, 9(2), 130–133.
10. Chaurasia, V., & Pal, S. (2018). Heart disease prediction using XG Boost. *International Journal of Engineering & Technology*, 7(3.34), 292–295.
11. Kumar, M., & Gupta, P. (2020). Predictive modeling for heart disease diagnosis using machine learning algorithms. *Journal of Ambient Intelligence and Humanized Computing*, 12(7), 6919–6930.
12. Smith, M. R., & Jenkins, P. R. (2019). A comparative study of machine learning models for heart disease prediction. *IEEE Access*, 7, 164823–164834.
13. Sarwar, M., & Hussain, S. (2020). Heart disease prediction using ensemble machine learning techniques. *Journal of Healthcare Engineering*, 2020, Article 4243126.
14. Chauhan, S., & Meena, M. (2021). Heart disease prediction using optimized XG Boost model. *International Journal of System Assurance Engineering and Management*, 13(Suppl 1), 744–752.
15. Ghosh, P., & Khanna, M. (2017). A hybrid machine learning approach for heart disease prediction. In *Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*.
16. Sharma, A., & Bhardwaj, P. (2022). A study on the use of XG Boost for predicting cardiovascular diseases. *Journal of Data Science and Intelligent Systems*, *1*(1), 45–55.
17. Singh, P., & Saini, G. (2020). Predictive analytics for heart disease using machine learning. *International Journal of Computer Applications*, 176(36), 13–17.
18. Jabbar, S., & Rao, G. R. (2020). Classification of heart disease using machine learning techniques. *International Journal of Engineering and Advanced Technology*, 9(3), 3662–3666.
19. Mohammad, T., & Karim, M. (2021). Machine learning in cardiovascular health prediction: A review. *ACM Computing Surveys*, 54(5), 1–35.
20. Guleria, P., Srinivasu, P. N., Ahmed, S., Almusallam, N., & Alarfaj, F. K. (2022). XAI framework for cardiovascular disease prediction using classification techniques. *Electronics*, 11(24), 4086.
21. Sultan, S., Javaid, N., Alrajeh, N., & et al. (2025). Machine learning-based stacking ensemble model for prediction of heart disease with explainable AI and K-fold cross-validation: A symmetric approach. *Symmetry*, 17(1), 2.

HOW TO CITE: A. R. Deepa*, Venkata Ganji, Ensemble Machine Learning for Cardiovascular Disease Prediction, *Int. J. Sci. R. Tech.*, 2025, 2 (10), 399-409. <https://doi.org/10.5281/zenodo.17444068>