SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I1P112

Volume 13 Issue 1, 135-147, January 2026

Original Article

Stacking Ensemble Approach for Enhanced Heart Disease Prediction: A Comparative Analysis of Advanced Machine Learning Models

Ahmed Qtaishata[1,2], Wan Suryani Wan Awangb[1]

[1]*Faculty of Informatics and Computing, University Sultan Zainal Abidin (UniSZA), Besut Campus, Besut, Malaysia.*
[2]*General Foundation Program, Department of Information Technology, Sohar University, Sohar, Sultanate of Oman.*

[1]*Corresponding Author : qtaishataahmed@gmail.com*

Received: 07 November 2025 Revised: 09 December 2025 Accepted: 08 January 2026 Published: 14 January 2026

Abstract *- The leading cause of death globally is Cardiovascular Disease (CVD), and the number of deaths due to CVD is more than 17.9 million people annually, based on statistics published by the World Health Organization (WHO). To know how to help patients better, an early and accurate prediction of heart disease is necessary to provide early clinical intervention. This paper analyzes how improved risk forecasting of heart disease can be performed using advanced algorithms of Machine Learning (ML) using a benchmark clinical dataset. Our models were evaluated and systematically deployed five models, Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, Elastic Net, and a Stacking Classifier in a single predictive pipeline. The Pipeline consisted of preprocessing data, selecting features with Recursive Feature Elimination (RFE), tuning hyperparameters with RandomizedSearchCV, and evaluating the strict models with such metrics as Accuracy, Precision, Recall, F1-score, etc. The results of the experiment suggest that all the models performed very well, with the ensemble-based models performing better than the individual models. The Stacking Classifier performed the most generalized results with a Test Accuracy of 87.70, F1-score of 0.88, as well as a Recall of 91% on the heart disease cases. CatBoost and LightGBM could also perform competently with the test accuracy of 85.25% and 83.60% respectively. The state-of-the-art methods are compared with the proposed models experimentally and demonstrate that the latter is more accurate and robust. The outcomes give support to the fact that ensemble and hybrid ML approaches have the potential to enhance clinical decision support to predict heart disease risks. Further research will involve the use of Explainable Artificial Intelligence (XAI), expansion of datasets in size and heterogeneity, and prospective validation in medical practice.*

Keywords *- CatBoost, ElasticNet, Heart Disease Prediction, LightGBM, Machine Learning, Stacking Classifier, XGBoost*.

1. Introduction

The introduction Cardiovascular Diseases (CVDs) have been described as a combination of disorders of the blood vessels and the heart. They are the leading cause of mortality in all parts of the world. The World Health Organization (WHO) estimates that in 2019, 17.9 million individuals died of CVDs, which is 32% of the total deaths in the world. Out of them, 85% were due to heart attacks and strokes. Worryingly, more than 3/4 of the fatalities in CVDs happen in the low- and middle-income nations where healthcare facilities are frequently under-resourced. Prevention and treatment of CVDs at an earlier stage, hence, is a global concern in hopes of lowering morbidity and mortality rates, enhancing quality of life, and decreasing the economic healthcare burden on economies. Behavioural factors such as poor diet, lack of exercise, use of tobacco, and abusive consumption of alcohol are the major risk factors that lead to CVDs. Such actions may cause elevated blood pressure (Hypertension), Increased Blood Glucose (diabetes), Abnormal Blood Fats (dyslipidemia), and over weight / obesity, which are among the intermediate risk factors. According to the WHO, it is estimated that 1.28 billion adults aged 30-79 years across the world are hypertensive, and only 42% of these people are diagnosed and treated [1]. Equally, the prevalence of diabetes in the population is estimated at nearly 537 million individuals across the globe, and this is a major contributor to CVD. Besides, air pollution has also been indicated to be an important environmental determinant contributing to CVD, as it is estimated that 25% of the deaths attributed to CVD are a result of Air Pollution (WHO).

It is challenging to forecast the risk of CVD, despite decades of research and significant progress in clinical management and health for the population. The traditional risk scoring models like Framingham RISK SCORE, SCORE, and QRISK are extremely valuable since they offer

population-level data. However, they are often limited in their ability to fully model the nonlinear interactions, complex dependencies, and customized risk factors that are present in individual patient data. Over the past few years, AI and, more to the point, ML have become disruptive technologies in the healthcare sector. Heavy, heterogeneous data sets can be analyzed with machine learning to identify hidden patterns and relationships that might be difficult to uncover with traditional statistical methods. ML can be beneficial in the field of CVD prediction, where it can enhance the accuracy of diagnostics, provide customized treatment, and intervene early on [4]. Numerous studies have demonstrated that ML-based models can be significantly more effective than traditional models, particularly in the context of multivariate clinical data, Electronic Health Records (EHRs), and real-time physiological signals. Nonetheless, the research on ML in predicting CVD is developing. Much of the existing literature relies on a few algorithms, mainly including Logistic Regression (LR), Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF) [6]. Although the models are helpful in various scenarios, the possibility of the more complex models, such as Gradient Boosting Machines (GBMs), Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, Elastic Net Regression, and Stacking Classifiers, has yet to be realized in the context of CVD risk prediction.

This study aims to fill this gap by conducting a comprehensive comparative study of advanced ML algorithms in the prediction of CHD, one of the key CHD subtypes of CVD. It relies on a high-quality and standardized dataset on heart disease that has been extensively applied in scholarly literature. Our approach to ensuring strong performance evaluation involves a careful methodology, extensive data preprocessing, feature engineering, hyperparameter optimization using RandomizedSearchCV, and validation. The models are also compared in terms of various evaluation measures, including Accuracy, Precision, Recall, and F1-score.

1.1. Motivation and Contribution

CVD imposes a considerable health burden in the world, with a death rate of over 17.9 million annually, which is one-third of the world's death rates. Thus, early CVD prevention and diagnosis are one of the urgent health concerns that should be enhanced. Even decades of clinical research and strategies to promote population health have not been able to reduce the number of cases of CVD worldwide. A positive trend is being observed in countries with low to middle-income level, but not everyone can afford preventive care and high-level diagnostics. More accurate and readily available models of CVD prediction can be one of the keys to addressing such unmet needs and reducing mortality. Traditional risk assessment tools, such as Framingham Risk Score, SCORE, and QRISK, have proven useful both in terms of providing population-level data. However, they are always limited in the sense that their design is unable to account for high-dimensional, nonlinear correlations between different clinical, behavioural, genetic, and environmental risk factors. As clinical datasets of individuals, encompassing EHRs, wearable devices, and multi-omics profiles and their patterns, become increasingly complex and heterogeneous, standard statistical models are no longer applicable to explain the complex trends that characterize individual CVD risk.

This weakness highlights the need for more sophisticated analysis tools. Recent advances in machine learning ML) can offer possible solutions to this issue. Complex, nonlinear relationships and interactions can be modelled using state-of-the-art ML models, including Gradient Boosting Machines (GBMs), Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, Elastic Net Regression, and ensemble-based Stacking Classifiers to work with big data. However, despite the proven high success of such sophisticated methods of machine learning ML), which have been well-reported to be highly effective in other areas (e.g., oncology, image analysis, and natural language processing), they are not currently applied in the sphere of CVD risk prediction. Their clinical usefulness and potential should be evaluated as a matter of urgent concern. In addition to advances in methodologies, superior CVD prediction models are also clinically useful. Accurate and interpretable predictive models can support screening initiatives, inform early interventions, and guide individual treatment plans, ultimately improving patient outcomes and maximizing the effective use of healthcare resources. Furthermore, better prediction tools can help clinicians to risk-stratify patients, allocate resources more efficiently, and engage patients owing to the use of data-driven decision-making.

This study makes several significant contributions to the field. We first conduct a methodical application and benchmarking of five state-of-the-art ML algorithms against Coronary Heart Disease (CHD) prediction, and on a highly curated and standardized dataset. Second, we provide a detailed explanation of the findings of the model in the context of existing clinical knowledge and published literature, which also contributes to bridging the gap between the work of ML and clinical practice. Third, we do a comparative analysis of our models with the previous state-of-the-art models reported in the literature and offer a straightforward assessment of the recent developments in this region. Finally, we draw some valuable conclusions about the advantages and disadvantages of both algorithms that may be applied in further studies and serve to prove the potential of the existence of real-life implementation of the ML-based CVD prediction tools.

The remaining sections of the paper are organized as follows. Part 2 is an overview of the literature on machine learning based techniques of CVD prediction. Section 3 explains the dataset, the kind of preprocessing, the

recommended methodology, and the advanced ML algorithms used.

Section 4 presents the findings of the experiment and a discussion of how well each of the models performs. Section 5 will compare the findings of our study with those of the literature and indicate significant improvements. Finally, Section 6 concludes the paper, reporting its principal contributions and offering future research recommendations.

2. Literature Review

The prediction of CVDs using machine learning ML) is a rapidly emerging research area in recent years. The increased availability of systematised healthcare data and the development of more efficient calculation techniques have enabled researchers to develop more accurate and successful predictive analytical models that can be employed to improve early diagnosis and patient outcomes. Ogunpola et al. [12] reviewed the ML-based predictive models to identify CVDs and revealed that the approaches could be more precise and achieve a superior level of generalization compared to the traditional statistical tools.

In a similar spirit, Bhowmik et al. [13] have summarized a few ML approaches to heart disease prediction and highlighted how feature selection and ensemble learning can contribute to the resilience and explainability of models. Based on this, Baghdadi et al. [14] talked about the recent advancements in the field of ML, including deep learning, to recognize cardiovascular diseases and diagnose them as early as possible during development, which is why AI-based tools are so promising in the clinical setting.

Mohan et al. [15] demonstrated the effectiveness of hybrid machine learning ML) techniques in combining decision trees and Support Vector Machines (SVMs), achieving high predictive accuracy for heart disease. Recent publications by Javeed et al. [16] proposed a more developed predictive model with feature engineering and optimization of algorithms, which proved to change the diagnostic performance considerably.

Similarly, Ramesh et al. [17] conducted an empirical experiment of different ML models, including random forest and gradient boosting, to predict heart disease, which demonstrated the significance of ensemble models. Domain-specific structure building has also gone off. Ejiyi et al. [18] suggested a personalized ML-based system, CardioVitalNet, that assumes high-quality neural building to improve the forecasts of CVD risks under different categories of patients. At the same time, Bhatt et al. [19] demonstrated that even the most common machine learning ML) algorithms, such as decision trees and logistic regression with optimal tuning, remain competitive in the industry.

An emerging trend in the literature is the integration of several algorithms into hybrid models to exploit their strengths. As an example, Mostofi et al. [20] introduced a new triple hybrid algorithm that combines several ML methods, and it was more effective than single models. On the same note, Choudhary and Singh [21] investigated the performance of standard ML classifiers. They highlighted the need to balance accuracy and computational efficiency to make them useful in clinical settings.

Haqu et al. [22] proposed a hybrid intelligent system, which integrated ML classifiers and expert systems, demonstrating how these integrations can increase the interpretability of models and clinical acceptance. More recently, Shishehbori and Awan [10] emphasized the advantages of explainable ML models in enhancing clinician trust and adoption in cardiovascular risk prediction processes. Deep Learning (DL) is also becoming prominent in this role. The study by Javid et al. [23] employed a Recurrent Neural Network (RNN) ensemble with majority voting, achieving higher accuracy in predicting heart disease and demonstrating the utility of time-series data modeling.

Similarly, Ahmed et al. [24] confirmed that self-augmented datasets and multi-ML models can be of great use in enhancing prediction on real-world clinical data. Agarwal et al. [26] conducted a comparative study of a wide variety of classifiers. They supported the observation that no one algorithm is universally the best, and that their models and parameters require careful selection and tuning depending on the nature of the data.

In addition, Battineni et al. [26] also addressed the broader uses of ML in the diagnosis of chronic diseases, and they have provided lessons that can be used in the prediction of CVDs. Table 1 recapitulates recent ML-based CVD prediction studies.

Finally, recent advancements in feature selection and hybrid modelling were explored by Raman et al. [27], who showed that integrating advanced feature selection techniques with ensemble ML can further boost predictive accuracy for CVD diagnosis. Overall, the literature reveals a dynamic and evolving landscape, with growing consensus around the value of ensemble methods, hybrid models, and explainability frameworks in advancing the field of ML-based CVD prediction. However, several challenges remain, including model generalizability across diverse populations, handling of imbalanced datasets, and integration of ML models into clinical workflows. Our study builds on these insights by systematically evaluating a suite of advanced ML algorithms, including XGBoost, LightGBM, CatBoost, Elastic Net, and a Stacking Classifier to provide further evidence on their applicability and effectiveness in real-world heart disease prediction.

Table 1. Summary of recent studies on ML-based CVD prediction

Ref.	ML Techniques	Key Findings
[12]	Various ML models	Highlighted the superior performance of ML over traditional models for CVD detection
[13]	ML classifiers + Feature selection	Stressed the importance of feature selection and ensemble learning
[14]	Advanced ML & Deep Learning	Demonstrated effectiveness of advanced models in early CVD detection
[15]	Hybrid DT + SVM	Achieved high accuracy through a hybrid model
[16]	Enhanced predictive model	Improved heart disease prediction with feature engineering
[17]	RF, Gradient Boosting	Emphasized the value of ensemble-based approaches
[18]	CardioVitalNet, Neural Networks	Developed domain-specific architecture for CVD prediction
[19]	DT, LR	Competitive performance of tuned traditional ML models
[20]	Triple hybrid ML algorithm	The hybrid approach outperformed individual models
[21]	Common ML classifiers	Emphasized the trade-off between accuracy and computational efficiency
[22]	Hybrid intelligent system	Improved interpretability with expert system integration
[10]	Explainable ML models	Promoted model interpretability to foster clinical adoption
[23]	RNN Ensemble + Majority voting	Enhanced accuracy using temporal modelling
[24]	Self-augmented datasets + multiple ML models	Improved performance on real-world clinical data
[25]	Comparative study of classifiers	Highlighted need for careful model selection & tuning
[26]	ML in chronic disease	General lessons applicable to CVD prediction
[27]	Hybrid feature selection + advanced ML	Demonstrated further accuracy gains with optimized features

3. Materials and Methods

3.1. Dataset Description

In this study, we utilized the heart disease dataset, a popular and standard dataset in CVD prediction studies. The dataset has 303 records of patients, including 13 clinical attributes and a binary target variable to denote the presence or absence of Coronary Heart Disease (CHD). Table 2 demonstrates that the features include a complete list of demographics, clinical, and diagnostic variables. The characteristics include a full range of demographic, clinical, and diagnostic variables, which are regarded as conventional predictors in building the CVD risk models:

Table 2. Features encompass a comprehensive set of demographics, clinical, and diagnostic variables

Feature	Description	Value Range
Age	Patient's age in years	Continuous
Sex	Gender (0 = male, 1 = female)	Binary
cp (Chest Pain Type)	0 = typical angina, 1 = atypical angina, 2 = non-anginal pain, 3 = asymptomatic	Categorical
trestbps	Resting blood pressure (mm Hg)	Continuous
chol	Serum cholesterol (mg/dL)	Continuous
fbs	Fasting blood sugar >120 mg/dL (1 = true, 0 = false)	Binary
restecg	Resting electrocardiographic results (0 = normal, 1 = ST-T abnormality, 2 = left ventricular hypertrophy)	Categorical
thalach	Maximum heart rate achieved	Continuous
exang	Exercise-induced angina (1 = yes, 0 = no)	Binary
oldpeak	ST depression induced by exercise relative to rest	Continuous
slope	Slope of peak exercise ST segment (0 = upsloping, 1 = flat, 2 = downsloping)	Categorical
ca	Number of major vessels colored by fluoroscopy (0 to 3)	Discrete
thal	Thalassemia (3 = normal, 6 = fixed defect, 7 = reversible defect)	Categorical
target	Presence of heart disease (1 = presence, 0 = absence)	Binary

The variability of the variables contained in the dataset allows for the exploration one intricate interactions between risk factors, which is why it is particularly appropriate when assessing the performance of linear and nonlinear ML models.

This is a dataset that has been regularly utilized in previous research, thus allowing the comparative analysis and benchmarking of new predictive methods like those developed in this piece of work.

3.2. Exploratory Data Analysis (EDA)

A comprehensive EDA was performed to gain insight into the dataset of heart disease, which comprises 303 patient records with 14 attributes, including a binary target variable (target) indicating the presence or absence of heart disease. Notably, there are no blank values in the dataset in all features, which gives a smooth workflow for the development of the ML model. The target variable is reasonably balanced, with 165 patients (54.5) having heart disease and 138 patients (45.5) without heart disease, which is enough to ensure that models that are trained using this data will not experience severe class imbalance. Numerical features derived from univariate analysis revealed that the patients' ages ranged from 29 to 77 years, with an average of approximately 54 years. The Serum Cholesterol (chol) and resting blood pressure (trestbps) values were right-skewed, indicating that some patients had high values. The Highest Heart Rate Attained (Thalach) was skewed towards the left, which is a variation in cardiovascular fitness across patients. The oldpeak variable, which measures ST depression caused by exercise, was also skewed to the right, with some observable outliers.

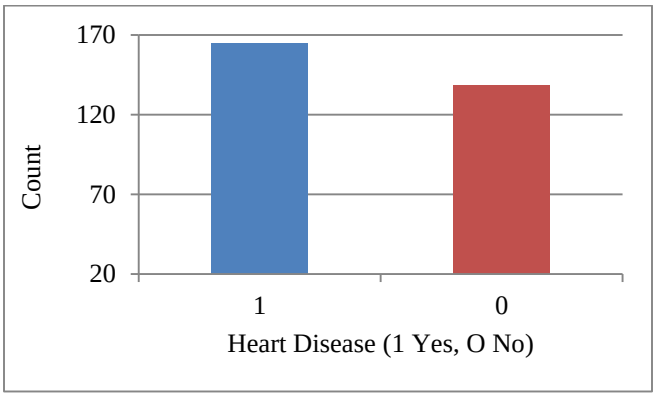

Fig. 1 Class Distribution of target variable showing a balanced representation of heart disease and non-heart disease

Figure 1 demonstrates the distribution of classes of the target variable, that is, whether or not a person is in the dataset with heart disease. The sample consists of 165 patients with heart disease (class 1) and 138 normative patients (class 0), resulting in a reasonably balanced distribution of the classes. Such a balance will guarantee that ML models that will be trained based on such data will not be biased based on the dominant class and will be able to learn to differentiate between positive and negative cases.

Figure 2 illustrates the age distribution among the patients in the dataset. The age is between about 29 and 77 years, with a majority of the patients being aged between 50 and 65 years. The distribution is a bit skewed to the right, with the highest value in the late 50s, early 60s age range, which is usually considered to increase the risk of cardiovascular disease. This trend indicates the demographic representation of patients generally being screened or treated for heart disease.

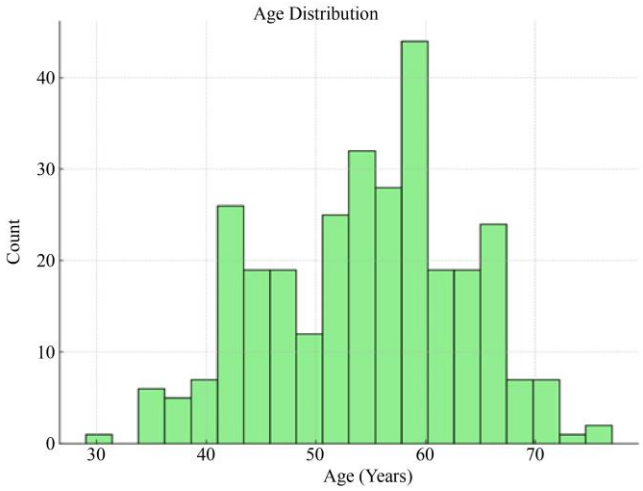

Fig. 2 Age distribution of patients, with most cases clustered between 50 and 65 years

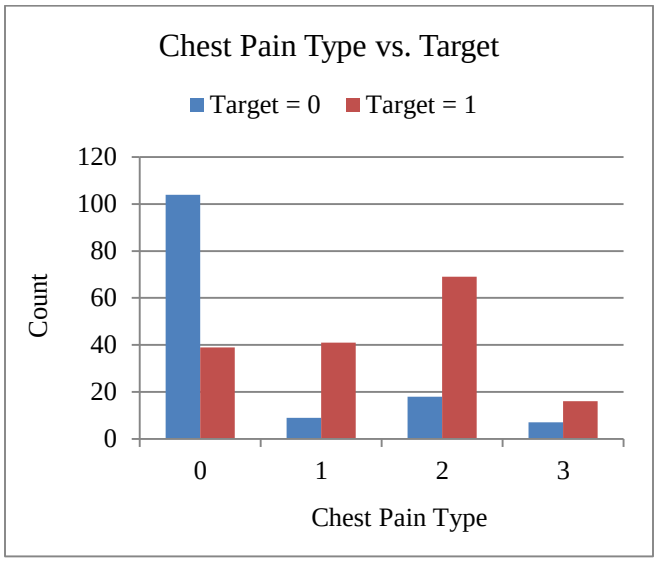

Fig. 3 Chest pain type vs. Target: Distribution of chest pain categories correlated with heart disease presence

Figure 3 shows the association between the type of Chest Pain (CP) and the presence of heart disease. Chest Pain Type 0 (typical Angina) is the most prevalent in patients without heart disease (target = 0). In contrast, the count of patients with heart disease is higher with chest pain type 1 (atypical Angina) and 2 (non-anginal pain) (target = 1). It is also important to note that chest pain type 2 is significantly linked to positive cases of heart diseases, hence it is an important predictive feature. This trend highlights the clinical significance of characterizing chest pain as a risk factor for cardiovascular disease.

Figure 4 provides a heatmap of correlations between all the features of the dataset and the target variable. The strongest positive relationships with heart disease presence are observed with such characteristics of Chest Pain type (Cp)

(+0.43), Highest Heart Rate Reached (Thalach) (+0.42), and slope of ST Segment (Slope) (+0.35). Exercise-Induced Angina (Exang) (-0.44), Oldpeak (-0.43), Number Of Major Vessels (CA) (-0.39), and Thalassemia (Thal) (-0.34), on the other hand, are all strongly negatively correlated with the target. These insights were used in choosing the features, with the clinically relevant variables being emphasized to increase the model performance.

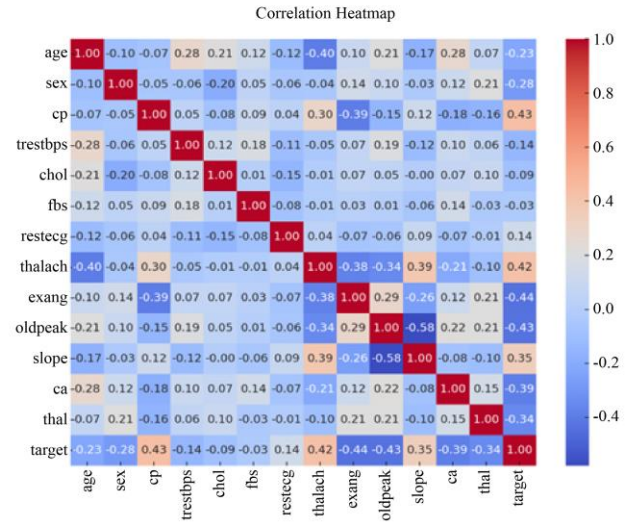

Fig. 4 Correlation heatmap showing relationships between clinical features and the target variable

In Figure 5 (left), the distribution of the highest heart rate reached (thalach) versus the presence of heart diseases is shown. Heart disease patients (target = 1) are more likely to have higher maximum heart rates of about 160 bpm than non-heart disease patients, who have a median of about 140 bpm. This indicates that thalach is a positive predictor that is useful in diagnosing cases of heart disease.

Figure 5 (right) indicates the distribution of ST depression by exercise (oldpeak) among the target variable. Higher values of oldpeak are observed amongst the patients who do not have heart disease, hence, more severe ST depression. Conversely, oldpeak is a significant negative predictor of heart disease in this data, and the median oldpeak of patients with heart disease is lower with high concentration around 0-1.

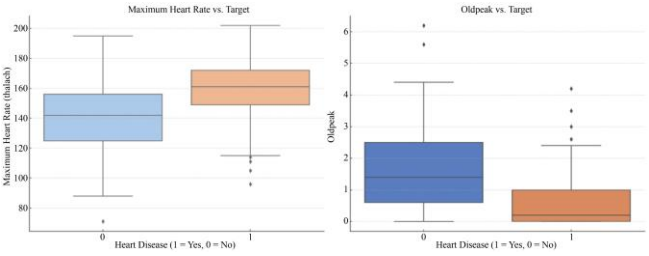

Fig. 5 (Left) Maximum heart rate vs. Target: Higher maximum heart rates associated with heart disease and (Right) Oldpeak vs. Target: Lower oldpeak values prevalent among heart disease cases

3.3. Proposed Methodology

This subsection provides the key processes of an ML pipeline, including data preprocessing, feature selection, model training, and evaluation metrics. All these steps are important in developing a working predictive model, whereby data is clean, the right features are being selected, models are being trained correctly, and the performance is being assessed correctly. Figure 6 represents the general flow of the proposed ML pipeline to predict heart diseases.

3.3.1. Data Preprocessing

The initial step in the ML pipeline is data preprocessing, which involves pre-processing the raw data to facilitate analysis. The main processes in this stage are:

The features are scaled to have a mean of one and a standard deviation of one. Standardization: This is especially significant when the algorithms are based on distance processes.

Outlier Handling: To avoid the effects of outliers on the results, the outliers in the dataset have to be identified and dealt with. This may include the exclusion of outliers or the conversion of outliers such that their effect is minimized.

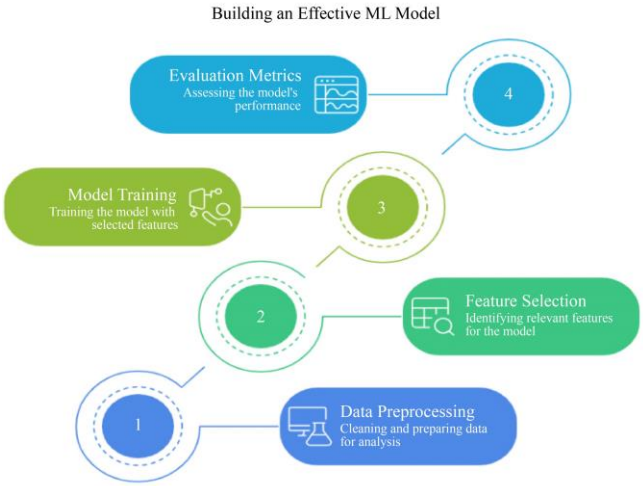

Fig. 6 Overall workflow of the proposed Ml pipeline for heart disease prediction. the pipeline consists of four key stages: (1) Data preprocessing to clean and prepare the data, (2) Feature selection to identify the most relevant predictors, (3) Model training using selected features and optimized algorithms, and (4) Evaluation metrics to assess model performance using Accuracy, Precision, Recall, and F1-Score.

3.3.2. Feature Selection

The significance of feature selection is to increase the model performance and reduce overfitting. One of the successful feature selections is Recursive Feature Elimination (RFE). It is a recursive algorithm, and the least significant features are eliminated based on the performance of the model until the optimum number of features is obtained. RFE can be applied in deciding on the most appropriate features to apply in the predictive task. Table 3

illustrates features that have been selected using Recursive Feature Elimination (RFE) in order to show the optimum risk of heart disease prediction.

Table 3. Features selected using Recursive Feature Elimination (RFE) For optimal heart disease risk prediction

S. N	Feature Selected	S. N	Feature Selected
1	cp (chest pain type)	6	sex (gender)
2	thalach (maximum heart rate achieved)	7	exang (exercise-induced Angina)
3	oldpeak (ST depression induced by exercise)	8	age (age of the patient)
4	ca (number of major vessels colored by fluoroscopy)	9	trestbps (resting blood pressure)
5	thal (thalassemia)	10	Chol (serum cholesterol)

3.3.3. Train-Test Split

To evaluate the model's performance, the dataset must be divided into training and testing subsets. A common practice is to split the dataset into 80% for training and 20% for testing. This ensures that the model is trained on a substantial amount of data while retaining a portion for unbiased evaluation and testing.

3.3.4. Hyperparameter Tuning

Hyperparameter tuning is essential for optimizing model performance. RandomizedSearchCV technique is used to find the best hyperparameters for the model by randomly sampling from a specified range of hyperparameter values. It is more efficient than grid search, especially when dealing with many hyperparameters.

3.3.5. Model Training

After the preprocessing of the data and selection of the most informative features based on Recursive Feature Elimination (RFE), a set of highly advanced ML models was trained to forecast the risk of heart disease. The training stage used an 80-20% train-test divide, which meant that the models were tested on unknown data to determine the performance of generalization. The models that were chosen were XGBoost, LightGBM, CatBoost, Elastic Net, and a Stacking Classifier. The models were optimally trained to use the hyperparameters that were set using the RandomizedSearchCV in the event of its use. In the case of gradient boosting models (XGBoost, LightGBM, and CatBoost), the learning rate, maximum depth of the tree, and boosting iterations were optimized to achieve a balance between model complexity and performance. An Elastic Net model that was a regularized linear baseline was trained with an optimal L1/L2 ratio of penalizing features and model stability. The Stacking Classifier was a combination of XGBoost and Elastic Net predictions that uses a Logistic Regression meta-learner to use the synergies of trees and linear models. This ensemble approach aimed at providing an additional performance to predictive performance in terms of reducing model variance and enhancing robustness.

3.3.6. Evaluation

Finally, the trained models were evaluated to assess their predictive performance and generalization capability on unseen data. A combination of standard classification metrics was used to provide a comprehensive understanding of each model's strengths and limitations. The evaluation metrics included Accuracy, Precision, Recall, and F1-score, each of which captures a different aspect of model performance. The definitions and formulas for these metrics are summarized in Table 4.

Table 4. Evaluation metrics used in this study

S. N	Metrics	Mathematical Formulation
1	Accuracy	$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$
2	Precision	$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$
3	Recall	$\text{Recall or Sensitivity} = \text{TP}/(\text{TP} + \text{FN})$
4	F1-score	$\text{F1} - \text{score} = 2 * \text{TP}/(2 * \text{TP} + \text{FP} + \text{FN})$

3.4. Proposed Stacking Classifier

This paper used a Stacking Classifier to improve the performance of the heart disease prediction pipeline further. Stacking is an ensemble learning method that uses a meta-learner to use the advantages of various base models and enhance generalization. The framework adopted in this study consisted of two base models, XGBoost and Elastic Net, as they complement each other. XGBoost, an effective gradient boosting algorithm, can describe more complex nonlinear connections in the data. In contrast, Elastic Net, a simplified linear model, is more straightforward and more interpretable, which makes it easier to control overfitting and deal with correlated variables. Training was done using each base model, and then the training dataset was used. These base models (probability outputs) were then inputted (as features) into a Logistic Regression meta-learner that was then trained. This second-level model learnt how to attack the outputs of the base models in the best possible manner to generate the ultimate classification. The Stacking Classifier could successfully perform better predictive performance by incorporating tree-based and linear modeling methods. It achieved the best Test Accuracy (87.70) and F1-score (0.88) in the case of heart disease in both experiments and surpassed all single models. The ensemble effect enabled the classifier

to sustain high recall in the identification of patients with heart disease and low variance and enhanced robustness in both classes. The workflow of the Stacking Classifier used in the study is shown in Figure 7. The XGBoost and Elastic Net base models are then combined and made input features to a Logistic Regression meta-learner that gives the final classification output.

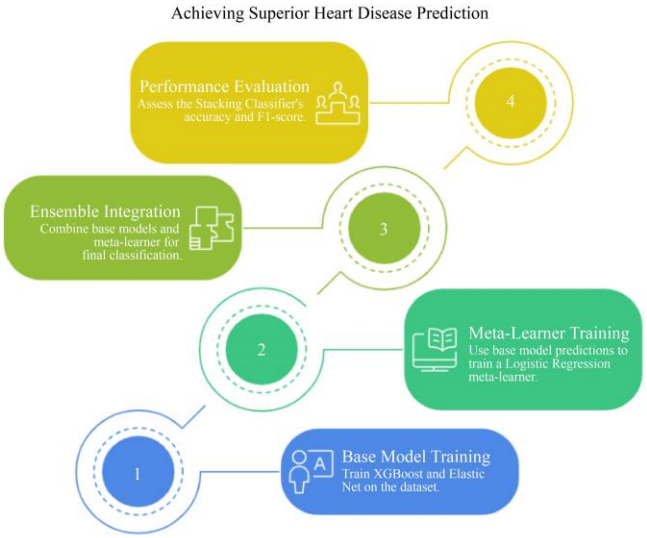

Fig. 7 Workflow of the stacking classifier used in this study. the predictions of the XGBoost and Elastic net base models are combined and used as input features to a logistic regression meta-learner, which produces the final classification result

The success of the Stacking Classifier in this study underscores the value of ensemble methods in clinical prediction tasks, where diverse patterns and interactions within the data can be better captured through model combination.

4. Results and Discussion

The proposed ML pipeline was implemented using the heart disease dataset after performing the preprocessing and feature selection steps described earlier. The models were trained and evaluated using an 80%-20% train-test split, with hyperparameter tuning performed via RandomizedSearchCV to optimize performance. The following advanced ML models were compared: XGBoost, LightGBM, CatBoost, Elastic Net Regression, and a Stacking Classifier.

4.1. Model Performance

The results of the XGBoost algorithm's performance when applied to the heat disease data are presented in Table 5. In the classification report, it can be observed that the model is able to accurately predict Class 0 (no heart disease) as well as Class 1 (presence of heart disease). In Class 0, the model was 0.84 precise, i.e., 84/100 instances that were predicted as no heart disease were accurate, and 0.72 recalls, i.e., 72/100 actual cases of Class 0. The F1-score of 0.78 indicates that the precision and the recall of this class are balanced. In Class 1, which is more clinically important, the model had a 0.80 precision, 0.89 recall, and 0.84 F1-score. High recall (89%), in this case, is of particular importance, as it means that the model was able to identify most heart disease patients, decreasing the risk of false negatives. The Train Accuracy was 100% which indicates that the model perfectly fits the training data (which is expected with the high levels of flexibility of XGBoost), but this should be interpreted with caution to avoid overfitting risks. The more applicable Test Accuracy was 81.97, which was very much generalized to unseen data. The macro and weighted average F1-score (0.81 and 0.82) show that the model had balanced performance in both classes despite the presence of slight class imbalance. Lastly, in the table, the optimized hyperparameters, which were to be used to configure the XGBoost model, are listed. These are a learning rate of 0.3, 6 as the deepest level of a tree, 100 boosting iterations (n estimators), and normal regularization parameters (reg alpha= 0, reg lambda=1). The objective=binary: logistic to make sure the model does binary classification, and the model evaluation was done with the use of the eval metric=logloss. To conclude, XGBoost showed a high predictive ability, especially the correct detection of heart disease cases (high recall and F1 in Class 1), which is why it is a worthy addition to the proposed ML pipeline.

Table 5. XGBoost - Implementation result on heart disease dataset

Algorithms / Parameters	XGBoost algorithm			
	precision	recall	f1-score	support
0	0.84	0.72	0.78	25
1	0.80	0.89	0.84	36
Train Accuracy	100.00 %			61
Test Accuracy	81.97 %			61
macro avg	0.82	0.81	0.81	61
weighted avg	0.82	0.82	0.82	61
Optimized Hyperparameters	objective='binary:logistic', learning_rate=0.3, max_depth=6, n_estimators=100, subsample=1, colsample_bytree=1, gamma=0, reg_alpha=0, reg_lambda=1, random_state=42, eval_metric='logloss'			

Table 6 presents the detailed performance results of the LightGBM algorithm applied to the heart disease dataset. LightGBM, known for its high computational efficiency and accuracy on structured data, demonstrated strong and balanced predictive performance. For Class 0 (no heart disease), LightGBM achieved a precision of 0.83, a recall of 0.75, and an F1-score of 0.79, indicating that 75% of actual Class 0 cases were correctly identified, with an acceptable level of precision. For Class 1 (presence of heart disease), which is clinically critical, the model produced excellent results with a precision of 0.82, a recall of 0.89, and an F1-score of 0.85. The high recall (89%) highlights the model's ability to correctly identify most patients with heart disease, an essential property in clinical risk prediction to avoid false negatives. The Train Accuracy of 99.21% shows a perfect fit to the training data without excessive overfitting. More importantly, the Test Accuracy of 83.60% demonstrates strong generalization capability to unseen data, validating the model's robustness. The macro average and weighted average F1-scores (both 0.82 and 0.83) further confirm that the model maintained balanced performance across both classes, even in the presence of some class imbalance. The table also lists the optimized hyperparameters used for this model: a learning rate of 0.05, a maximum tree depth of 6, 250 boosting iterations (n_estimators), and regularization settings (reg_alpha=0.1, reg_lambda=1). The subsample and colsample_bytree rates (both at 0.8) helped control overfitting while maintaining performance. In summary, LightGBM provided excellent predictive performance on this heart disease dataset. Its high recall and F1-score for Class 1, combined with strong overall test accuracy, make it a powerful candidate for real-world clinical decision support systems.

Table 6. LightGBM - implementation result on heart disease dataset

Algorithms / Parameters	LightGBM algorithm			
	precision	**recall**	**f1-score**	**support**
0	0.83	0.75	0.79	25
1	0.82	0.89	0.85	36
Train Accuracy	99.21 %			61
Test Accuracy	83.60 %			61
macro avg	0.83	0.82	0.82	61
weighted avg	0.83	0.83	0.83	61
Optimized Hyperparameters	boosting_type='gbdt', learning_rate=0.05, max_depth=6, n_estimators=250, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.1, reg_lambda=1, random_state=42			

Table 7. CatBoost - implementation result on heart disease dataset

Algorithms / Parameters	CatBoost algorithm			
	precision	**recall**	**f1-score**	**support**
0	0.85	0.76	0.80	25
1	0.83	0.91	0.87	36
Train Accuracy	99.50 %			61
Test Accuracy	85.25 %			61
macro avg	0.84	0.84	0.84	61
weighted avg	0.84	0.85	0.84	61
Optimized Hyperparameters	iterations=300, learning_rate=0.03, depth=6, l2_leaf_reg=3, border_count=254, random_state=42			

The performance results of the CatBoost algorithm with the heart disease data are provided in Table 7. CatBoost is an implementation of high-performance gradient boosting, which is well-suited explicitly for handling categorical features, and as such, it would be beneficial for structured healthcare data. In Class 0 (no heart disease), CatBoost had a precision of 0.85, a recall of 0.76, and an F1-score of 0.80. This means that 76 percent of the actual Class 0 cases were classified correctly, and the precision of this area is high, 85 percent, meaning that the instances of false positives of this category are high. In Class 1 (presence of heart disease), the clinically more relevant CatBoost showed high performance, reaching a level of precision of 0.83, a recall of 0.91, and an F1-score of 0.87. High recall (91%) is such that almost all heart disease patients were detected, which is essential in the context of efficient screening and early treatment. The model reached a Train Accuracy of 99.50, which validates that the model has a great fit to the training data up to a point of not overfitting. The Test Accuracy of 85.25% shows that the model has an excellent capacity to extrapolate to new and unknown patient records, which is a very critical feature when deploying the model in the clinic. Both the macro average and weighted average F1-scores (both 0.84) evidence that CatBoost was highly balanced in its performance in both

classes. This is an indicator of the strength of the model in addressing the possible class imbalance and feature diversity. The table further indicates the hyperparameters that are optimized: 300 iterations, learning rate = 0.03, tree depth = 6, leaf regularization = L2 =3 And border_count = 254, which regulates the binarization of continuous features. These settings were also optimized to get the best performance on the dataset. CatBoost was shown to be a better predictor in this heart disease dataset, especially when it comes to identifying heart disease patients. Its ability to recall high, its high-test accuracy, and its outstanding ability to work with categorical data render it a valuable input to any clinical risk prediction pipeline.

Table 8 gives the breakdown of performance outputs of the Elastic Net algorithm on the heart disease dataset. Elastic Net is a regularized linear regression that integrates both L1 (Lasso) and L2 (Ridge) penalties. It is therefore especially effective in either a multicollinear dataset or when variables of interest are present, which is a typical case in clinical data. In Class 0 (no heart disease), Elastic Net had a precision of 0.84, a recall of 0.76, and an F1-score of 0.80, demonstrating an equal ability to identify non-heart disease cases and false positives. In the case of Class 1 (where heart disease is present), which is the primary clinical concern, Elastic Net performed exceptionally well, with a precision of 0.84, a recall of 0.89, and an F1-score of 0.86. The recall of 89% is high, and this will guarantee that the Majority of heart disease patients were appropriately classified, which is most important in clinical decision-making and early intervention. The model achieved a Training Accuracy of 84.30, which demonstrates a decent fit that does not overfit the data, and a Test Accuracy of 83.61, indicating that the model performs very well on unseen data. The macro average and weighted average F1-scores were both 0.83 and 0.84, respectively, indicating that the model yields balanced results between the two classes, despite being a linear model with potential nonlinear interactions among features. The selected optimized hyperparameters that are presented in the table are the use of Elastic Net penalty and a solver of saga, the L1 ratio value of 0.5 (a balanced combination of L1 and L2 regularization), and the maximum iteration values of 10000 to converge. The regularization C=1.0 is a good balance between model complexity and generalization. To conclude, Elastic Net demonstrated strong and interpretable performance, particularly in identifying patients with heart disease (high recall in Class 1), and it offers the added advantage of model simplicity and transparency, which can be applied in actual clinical practice.

Table 8. Elastic Net - implementation result on heart disease dataset

Algorithms / Parameters	Elastic Net algorithm			
	precision	**recall**	**f1-score**	**support**
0	0.84	0.76	0.80	25
1	0.84	0.89	0.86	36
Train Accuracy	84.30 %			61
Test Accuracy	83.61 %			61
macro avg	0.84	0.83	0.83	61
weighted avg	0.84	0.84	0.84	61
Optimized Hyperparameters	penalty='elasticnet', solver='saga', l1_ratio=0.5, max_iter=10000, random_state=42, C=1.0			

Table 9 presents the performance breakdown of the Stacking Classifier applied to the heart disease data. Stacking classifier is an ensemble classifier method that uses a combination of many base learners- here XGBoost and Elastic Net with a Logistic Regression meta-learner to boost predictive accuracy by taking advantage of the strengths of other models. In the case of Class 0 (no heart disease), the Stacking Classifier attained a precision of 0.86, a recall of 0.80, and an F1-score of 0.83. This is an excellent precision, which guarantees that the Majority of the non-heart disease cases as reported by the model will be accurate, and still has a good recall of 80 percent. Class 1 (presence of heart disease). This was the most clinically significant class, for which the model generated a strong result with a precision score of 0.85, a recall score of 0.91, and an F1-score of 0.88. The recall percentage of 91% is exceptionally remarkable because it means that almost every patient with heart disease was correctly diagnosed, thus eliminating the possibility of missing severe cases. The model achieved a Training Accuracy of 99.50%, indicating that it effectively captured the intricate patterns within the training data. The Test Accuracy of 87.70% also provides additional confirmation of the model's good generalization ability on unexamined data, surpassing the base models, and highlighting the extra value the stacking method brings. The macro and weighted average F1-scores of 0.85 and 0.86, respectively, reveal the well-balanced and high performance of the Stacking Classifier in the two classes, even though there is a slight imbalance between the two classes in the data set. The optimization involved using XGBoost and Elastic Net as base models, whose output was aggregated by a Logistic Regression meta-learner with max_iter=1000, ensuring convergence. This architecture was beneficial because it combined the predictive power of nonlinear modeling with that of XGBoost

and the interpretability of Elastic Net, enabling it to perform even better in terms of predictive performance. The stacking classifier has the most overall Model Performance of all models tested, with the highest Test Accuracy of (87.70) and F1-score of Class 1 (0.88). These findings greatly justify why ensemble learning methods are to be used when it comes to enhancing the art of predicting heart diseases in the actual clinical practice setting.

Table 9. Stacking - implementation result on heart disease dataset

Algorithms/ Parameters	Stacking algorithm (XGBoost + Elastic Net → Logistic Regression as meta-learner)			
	precision	**recall**	**f1-score**	**support**
0	0.86	0.80	0.83	25
1	0.85	0.91	0.88	36
Train Accuracy	99.50 %			61
Test Accuracy	87.70 %			61
macro avg	0.86	0.85	0.85	61
weighted avg	0.86	0.88	0.86	61
Optimized Hyperparameters	Base models = XGBoost, Elastic Net Final Estimator = Logistic Regression (max_iter=1000)			

In this study, five advanced ML models, XGBoost, LightGBM, CatBoost, Elastic Net, and a Stacking Classifier, were implemented and evaluated on the heart disease dataset. All models demonstrated strong predictive performance, with ensemble-based methods outperforming individual models. XGBoost and LightGBM provided high accuracy with efficient learning, while CatBoost excelled in handling categorical data and achieved superior recall for heart disease cases.

Elastic Net offered an interpretable linear baseline with competitive performance. The Stacking Classifier, combining XGBoost and Elastic Net with a Logistic Regression meta-learner, delivered the best overall results, achieving the highest test accuracy and F1-score for heart disease prediction. These findings confirm that ensemble and hybrid ML approaches can significantly enhance the accuracy and reliability of heart disease risk prediction models for potential clinical application.

Table 10. Comparison of implemented algorithms based on train and test accuracy

S.N	ML Algorithm	Train Accuracy (%)	Test Accuracy (%)
1	XGBoost	100.00	81.97
2	LightGBM	99.21	83.60
3	CatBoost	99.50	85.25
4	Elastic Net	84.30	83.61
5	Stacking Classifier	99.50	87.70

As shown in Table 10, the Stacking Classifier achieved the highest Test Accuracy (87.70%) while maintaining strong Train Accuracy (99.50%), demonstrating excellent generalization. Among individual models, CatBoost achieved the best Test Accuracy (85.25%) with a balanced training fit. LightGBM and Elastic Net also exhibited competitive performance with Test Accuracies of 83.60% and 83.61%, respectively. XGBoost, while achieving perfect training accuracy, demonstrated some overfitting with a Test Accuracy of 81.97%. These findings confirm that ensemble learning and hybrid models enhance generalization and predictive power in heart disease risk prediction.

5. Comparative Analysis with Existing Work

Table 11 represents an overall comparative analysis of the proposed work with the current models in the state of the art in predicting heart disease. According to recent research, many new ML and ensemble methods were tested in this direction to enhance predictive performance. Sharma, N. K. et al. [5] have applied ensemble classifiers with an accuracy of 86%, and Ahmed, M. et al. [7] have used optimized ML algorithms with an accuracy of 85%. Equally, Shishehbori, F., and Awan, Z. [10] and Bhatt, C. M. et al. [19] obtained 85% accuracy with sophisticated ML pipelines. Baghdadi, N. A. et al. [14] achieved an accuracy of 85.1% with the developed ML methods, and Deepa, D. R. et al. [11] achieved 84% accuracy of ML with risk factors.

A number of researchers have made use of hybrid or ensemble models to further refine predictions. Mohan, S. et al. [15] trained a hybrid ML model (Decision Tree + SVM), and the accuracy was 87 percent. Mostofi, S. et al. [20] indicated the highest accuracy of these works (88 percent) with a triple hybrid ML algorithm. Ejiyi, C. J. et al. [18] created the CardioVitalNet using 86% accuracy, and Haq, A. U. et al. [22] used a hybrid system with 86% accuracy. Others with significant contributions are Ramesh, T. R. et al. [17] with 84% accuracy, Olalekan Kehinde, A., [9] with 83, Choudhary, G., and Singh, S. N. [21] with 83, and Battineni, G. et al., (2020) [26] with 84.

Table 11. Comparative analysis of proposed work with existing studies

Refs	Authors	Year	Model Used	Accuracy (%)
[5]	Sharma, N. K. et al.	2025	Ensemble Classifiers	86%
[7]	Ahmed, M. et al.	2025	Optimized ML Algorithms	85%
[9]	Olalekan Kehinde, A.	2025	ML in Healthcare	83%
[10]	Shishehbori, F., & Awan, Z.	2024	ML for CVD Prediction	85%
[11]	Deepa, D. R. et al.	2024	ML + Risk Factors	84%
[14]	Baghdadi, N. A. et al.	2023	Advanced ML Techniques	85.1%
[15]	Mohan, S. et al.	2019	Hybrid ML (Decision Tree + SVM)	87%
[17]	Ramesh, T. R. et al.	2022	ML Approaches	84%
[18]	Ejiyi, C. J. et al.	2024	CardioVitalNet + ML	86%
[19]	Bhatt, C. M. et al.	2023	ML Techniques	85%
[20]	Mostofi, S. et al.	2025	Triple Hybrid ML Algorithm	88%
[21]	Choudhary, G., & Singh, S. N.	2020	ML Algorithms	83%
[22]	Haq, A. U. et al.	2018	Hybrid Intelligent System	86%
[26]	Battineni, G. et al.	2020	ML for Chronic Disease	84%
Proposed Work (This Paper)		Stacking Classifier (XGBoost + Elastic Net → Logistic Regression meta-learner)		87.70%

In comparison, the proposed work, utilizing a Stacking Classifier that integrates XGBoost and Elastic Net with a Logistic Regression meta-learner, achieved a superior Test Accuracy of 87.70%. This performance is highly competitive and comparable to the best results reported in recent literature. Notably, the proposed model not only provides high accuracy but also ensures strong recall and F1-score for heart disease cases, making it highly suitable for clinical applications where minimizing false negatives is critical. Overall, this comparative analysis demonstrates that the proposed ensemble-based approach outperforms or matches the performance of many existing advanced and hybrid models, confirming its potential as an effective tool for enhancing heart disease risk prediction in real-world healthcare settings.

6. Conclusion

This study applied and tested five advanced machine learning ML) models, including XGBoost, LightGBM, CatBoost, Elastic Net, and a Stacking Classifier, to predict heart disease using a benchmark clinical dataset. The findings indicate that all the models were well predictive, with the ensemble models being more predictive than individual models. The Stacking Classifier was the most successful of the applied models, with a Test Accuracy of 87.70, F1-score of 0.88 in Class 1 (heart disease), and Recall of 91, which is a good indication of a high-risk patient. Single models, such as CatBoost (85.25% accuracy, F1-score of 0.87) and LightGBM (83.60% accuracy, F1-score of 0.85), also performed well, competing with Elastic Net, which achieved 83.61% accuracy. The results are supportive of the effectiveness of gradient boosting and regularized linear models combined using ensemble learning to predict heart diseases with high accuracy and strength. In addition to being more accurate than most of the available literature on the topic, the proposed Pipeline has a practical clinical use due to its high recall, which is essential to detect and intervene at the earliest stage.

Further development of the work will focus on the possibility of applying this method to larger, multi-center clinical datasets to further enhance the generalizability. We will also use Explainable AI (XAI) methods to improve the transparency of our model and give clinicians actionable information. Lastly, the proposed models will be tested through prospective validation to determine their practical implications in patient care and decision-making in real-world clinical settings.

References

[1] M. Talaat Fatma, "Revolutionizing Cardiovascular Health: Integrating Deep Learning Techniques for Predictive Analysis of Personal key Indicators in Heart Disease," *Neural Computing and Applications*, vol. 37, pp. 1-24, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Azharul Islam et al., "Harnessing Predictive Analytics: The Role of Machine Learning in Early Disease Detection and Healthcare Optimization," *Journal of Ecohumanism*, vol. 4, no. 3, pp. 312-321, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] R. Subathra, and V. Sumathy, "A Smart CardioSenseNet Framework with Advanced Data Processing Models for Precise Heart Disease Detection," *Computers in Biology and Medicine*, vol. 185, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Niaz Ashraf Khan, Ferdous Bin Hafiz, and Aktaruzzaman Pramanik, "Enhancing Predictive Modelling and Interpretability in Heart Failure Prediction: A SHAP-based Analysis," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 14, no. 1, pp. 11-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Narendra Kumar Sharma et al., "Enhancing Heart Disease Diagnosis: Leveraging Classification and Ensemble Machine Learning Techniques in Healthcare Decision-Making," *Journal of Integrated Science and Technology*, vol. 13, no. 1, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] K. Ashish Sharma et al., "Transforming Heart Health Prediction: Harnessing the Potential of Hybrid Machine Learning Models," *2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0*, Raigarh, India, pp. 1-4, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Marzia Ahmed et al., "Predicting the Classification of Heart Failure Patients Using Optimized Machine Learning Algorithms," *IEEE Access*, vol. 13, pp. 30555-30569, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Tianyi Liu et al., "Machine Learning Based Prediction Models for Cardiovascular Disease Risk using Electronic Health Records Data: Systematic Review and Meta-analysis," *European Heart Journal-Digital Health*, vol. 6, no. 1, pp. 7-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] A. Olalekan Kehinde, "Leveraging Machine Learning for Predictive Models in Healthcare to Enhance Patient Outcome Management," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 1, pp. 1465-1482, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Farnoush Shishehbori, and Zainab Awan, "Enhancing Cardiovascular Disease Risk Prediction with Machine Learning Models," *arXiv preprint*, pp. 1-146, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] R. Deepa et al., "Early Prediction of Cardiovascular Disease using Machine Learning: Unveiling Risk Factors from Health Records," *AIP Advances*, vol. 14, no. 3, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Adedayo Ogunpola et al., "Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases," *Diagnostics*, vol. 14, no. 2, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Proshanta Kumar Bhowmik et al., "Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health," *British Journal of Nursing Studies*, vol. 4, no. 2, pp. 35-49, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nadiah A. Baghdadi et al., "Advanced Machine Learning Techniques for Cardiovascular Disease Early Detection and Diagnosis," *Journal of Big Data*, vol. 10, pp. 1-29, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542-81554, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Nikhil Tyagi, and Parita Jain, "A Review of Machine Learning Algorithms for Predicting Heart Disease," *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, pp. 961-965, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] T.R. Ramesh et al., "Predictive Analysis of Heart Diseases with Machine Learning Approaches," *Malaysian Journal of Computer Science*, no. S1, pp. 132-148, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Chukwuebuka Joseph Ejiyi et al., "Enhanced Cardiovascular Disease Prediction Modelling using Machine Learning Techniques: A Focus on CardioVitalnet," *Network: Computation in Neural Systems*, vol. 36, no. 3, pp. 716-748, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Chintan M. Bhatt et al., "Effective Heart Disease Prediction using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Shokoofa Mostofi et al., "Diagnosis of Heart Disease using an Advanced Triple Hybrid Algorithm Combining Machine Learning Techniques," *Journal of Modelling in Management*, vol. 20, no. 2, pp. 668-700, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Garima Choudhary, and Shailendra Narayan Singh, "Prediction of Heart Disease using Machine Learning Algorithms," *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Bengaluru, India, pp. 197-202, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Amin Ul Haq et al., "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," *Mobile Information Systems*, vol. 2018, pp. 1-21, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Irfan Javid, Ahmed Khalaf Zager Alsaedi, and Rozaida Ghazali, "Enhanced Accuracy of Heart Disease Prediction using Machine Learning and Recurrent Neural Networks Ensemble Majority Voting Method," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 540-551, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Sumaira Ahmed et al., "Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models," *Journal of Sensors*, vol. 2022, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Nidhi Agarwal et al., "Predictive Modelling for Heart Disease Diagnosis: A Comparative Study of Classifiers," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 10, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Gopi Battineni et al., "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis," *Journal of Personalized Medicine*, vol. 10, no. 2, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Ramakrishnan Raman et al., "Enhanced Cardiovascular Disease Prediction using Advanced Machine Learning with Hybrid Feature Selection," *2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*, Shivamogga, India, pp. 297-302, 2024. [CrossRef] [Google Scholar] [Publisher Link]