Dijlah Journal of Engineering Sciences (DJES) Vol. 2, No. 2, June 2025, pp. 86-102
ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 40

Predicting Heart Attacks using Machine Learning with Multiple Models and Hard Voting to Improve Accuracy

Ahmed hashim

Article Info

Article history:

Received April, 15, 2025
Revised May, 05, 2025
Accepted May, 08, 2025

Keywords:

Heart Attack

Prediction

Accurate

Coronary

ABSTRACT

Since heart attacks continue to be a leading cause of mortality globally, these numbers should encourage scientists to develop more effective methods of prevention and early detection. Using clinical data as a starting point, we train and verify a model to forecast the likelihood of a heart attack using machine learning techniques. Indicators of a patient's health form the basis of this concept. The best classification model was chosen after extensive testing and evaluation of many models, such as logistic regression, decision trees, random forest, support vector machines (SVM), k-nearest neighbours (kNN), Naïve Bayes, and Extreme Gradient Boosting (XGBoost). Feature scaling and encoding were applied to all 303 patients in the sample, who possessed a total of 14 distinct characteristics. Two sets of data were created: one for training and one for testing, so that the models could be tested. For every model, we determined the following metrics: ROC-AUC, F1-score, recall, accuracy, precision**, and precision. By a wide margin, XGboost outperformed kNN and SVM classifiers with its 90.2% prediction accuracy. Furthermore, we were able to train an ensemble voting classifier that achieved somewhat better results than the top individual model as well as its individual components. According to feature significance analysis, the two most essential criteria in determining the likelihood of a heart attack are the "which type" of chest pain and the "what kind" of exercise-induced angina. We go deeply into the models' inner workings, discussing the consequences of our discoveries and possible future enhancements. By utilizing machine learning algorithms, medical professionals may improve their ability to foretell the likelihood of a cardiac event, which might result in more timely and effective treatment. Future efforts to improve prediction performance could make advantage of more complete ensemble techniques, bigger datasets, and other characteristics.

Corresponding Author:

Location: 1 -Baghdad, Iraq, 37014, Email: a.h96pop@gmail.com

1. INTRODUCTION

Having trustworthy healthcare prediction models is crucial since heart attacks are among the top causes of death and disability globally. This work accomplished the revolutionary objective of machine learning by employing ensembles of supervised algorithms, multi classifiers, Extreme-and extra-trees learners, and similar tools to improve heart ailment diagnosis using difficult, non-linear datasets. The research delivers the best prediction confidence by adopting the hard voting ensemble strategy, as demonstrated by a careful assessment of algorithms like Random Forest, Gradient Boosting, and Naive Bayes. With a precision of 98.54%, process model (1) shown its ability to be a trustworthy supplementary tool for treatment selections. Modern machine learning techniques and the keywords they provide are crucial for online research and citation management. Please be precise in your title; else, your material may never reach its target readers.

The healthcare system is experiencing increased burden due to an aging population and a rise in chronic cardiac diseases (2). To reduce this load, continuous diagnostics are becoming more popular. This systematic approach to healthcare will soon allow for more predictive solutions.

1.1 Heart attack: What it is and Why It Matters

When the coronary arteries are totally clogged with cholesterol, fat, or anything else, it causes a myocardial infarction, often known as a heart attack. Due to the potential for life-threatening cardiac complications or even death, this obstruction requires prompt evaluation and treatment. One of the leading killers on a worldwide scale is cardiovascular disease (CVD). Understanding the effects of CVD on people, from the pumping of blood by the heart to prevention and therapy, is crucial for developing solutions. Healthcare decision-makers are increasingly utilizing machine learning and other technological developments to better predict when a heart attack may occur. The pictures show that there is a low to moderate correlation between the values of MV2(vv) and MV2(dd) obtained from these non-invasive and invasive data sources, so it has been prioritized to follow disciplined structured procedures for the analysis of visual methods related to cardiovascular disease. Developing prediction models that might prompt swift action to save lives is why this study is of the utmost importance. In the field of personalized medicine, there have been significant advancements in algorithms that combine many classifiers. Medical professionals can better serve their patients' individual needs when they do this. Here we compare two data sets that you have never seen outside of your training:

We have already changed the game with our mastery of machine learning algorithms in hospital admissions and patient care. The application of data science ideas and methodology is on the rise in several domains, including the prediction of cardiovascular diseases and data analysis. For example, one way to improve prediction accuracy is to use ensemble learning methods to rigorously vote among several Machine Learning models. The accuracy of risk variable identification may be shown, for instance, by large, dynamic datasets such as the Cleveland Heart Disease dataset. Local Interpretable Model-agnostic Explanations (LIME) (3) can help with model explanations and building patient-provider trust. By outlining the crucial steps of machine learning, from gathering data to validating models, it demonstrates how we may systematically apply machine learning to improve health outcomes at the individual and population levels. Data visualization

Each table in the record must be identified using Arabic numerals. Make sure that each table has annotations. You should know better than to cross the line like that at this moment. The table can only have a single horizontal at the very top, a single horizontal at the very bottom, and a single horizontal that separates the column heads from the rest of the table. All tables ought to be embedded into the text itself rather than being included in separate files. The authors may find the following to be illustrative (4)

2 - A Literature Review on the Subject

Many earlier research have used AI for early kidney disease detection and prediction since human health is so important.

Article A) "A Machine Learning-Based Applied Prediction Model for Identification of Acute Coronary Syndrome (ACS) Outcomes and Mortality in Patients during the Hospital Stay" (Sherazi, S.W.A., Zheng, & Lee, J.Y. (2023))

The authors Sherazi, S.W.A., Chen, and Lee, J.Y. created an applied prediction model based on machine learning to detect outcomes and mortality in patients hospitalized with acute coronary syndrome (ACS). around the year 2023. This study's findings were based on 5,51 factors culled from 13,104 records of patients. Nothing to worry about. After selecting features and doing preprocessing, 125 features were obtained. An ensemble of three methods—Random Forest, Extra Trees, and Gradient Boosting Machine—named SVEC (Soft Voting Ensemble Classifier) is proposed in this study.

Support vector extraction with convolution (SVEC) often yields results of 99.07% in terms of accuracy, precision, recall, F1-score, and area under the curve (AUC). When it came to predicting ACS and hospital mortality, the ensemble model outperformed individual classifiers.

The article "A Soft Voting Machine Learning Model with Explainable AI for Cardiovascular Disease Management" was written by Banerjee Majumder, A., Gupta, S., and Singh, D. in the year 2022

In order to better predict the incidence of heart disease, researchers described an ensemble learning model that used k-neighbors classifiers, bagged logistic regression, and gaussian naive bayes. When presented with noisy data on the incidence of heart disease, this improved ensemble model proved to be more resilient, accurate, and long-lasting than the individual models. This subfield of artificial intelligence studies the idea of deciphering AI models; the word "eXplainable AI" defines it.

"Enhancing Heart Attack Prediction with Machine Learning: A Study Using the Jordan University Hospital Heart Dataset" (Al-Makhadmeh, Z., & Tolba, A., 2024) includes the following:

The Heart Dataset from Jordan University Hospital was utilized in this study; it consisted of 486 patient records and 58 features were retrieved from them. Feature selection in the research was carried out using Particle Swarm Optimization. Other methods utilized were Random Forest, Support Vector Machines, Decision Tree, Naive Bayes, and K-Nearest Neighbors.

• Results: The proposed method successfully predicted the development of heart disease with a classification accuracy of 94.3%, demonstrating the efficacy of the integrated approach.

"A Soft Voting Ensemble Classifier for Early Prediction and Diagnosis of Major Adverse Cardiac Events" (Kim, J.K., et al., 2021) refers to the "D" category.

In this work, the methodology involved comparing the Jordan University Hospital Heart Dataset—which includes 486 patient records and 58 characteristics—to traditional benchmark datasets. For feature selection, particle swarm optimization was employed, while classifiers included Random forest, K-Nearest neighbors, Decision trees, Support Vector Machines, and Naive Bayes.

Results With a classification accuracy of 94.3%, the integrated system accurately predicts heart illness.

3-Our Model

The steps to take to get more information regarding the probability of heart disease from this dataset. This article initially delves into data science for exploratory data analysis (EDA) to gain a better understanding of the data and spot any errors. In the second phase, known as "data preparation," the information is transformed into a form that models can use. Data must be split into training and testing datasets after image preprocessing. Normalization is another option for making sure all features are the same size right now. The training data from the previous phase was used to train all of the machine learning models: K-Nearest Neighbors, Decision Tree, Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression. The predictive power of each model is assessed using the testing set. The last step is to combine the predictions from all the models using an ensemble technique. Classifiers that are better and more accurate are often the result of this. We measure the individual and collective performance of the models using a wide range of measures.

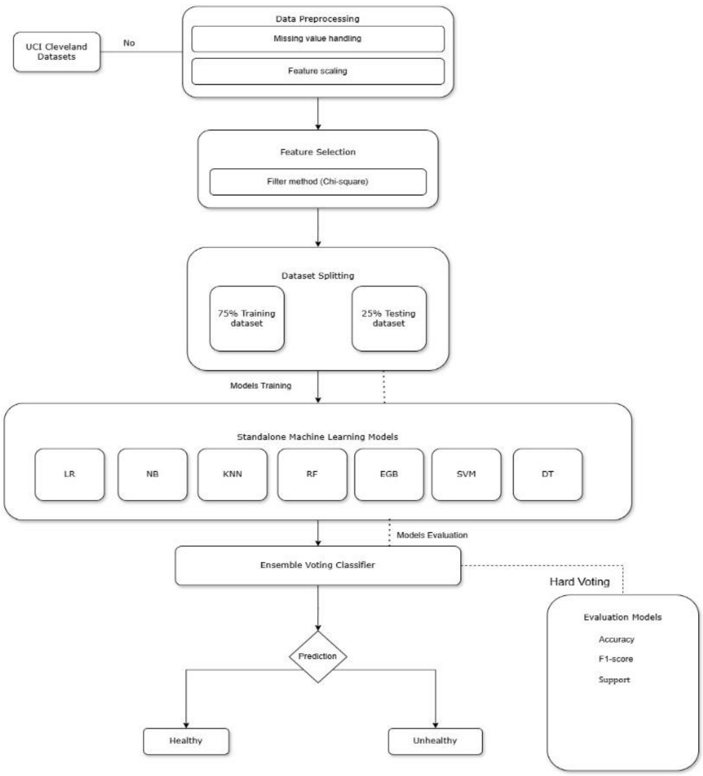


Figure.1 - The purpose of developing the model was to anticipate the occurrence of cardiac illness.

Research Methodology The Heart Attack Analysis and Prediction Dataset was employed in this investigation. You may get the renowned Cleveland Heart Disease dataset through the UCI Machine Learning Repository. A one-or-zero target label that denotes the existence or absence of cardiac illness (or the probability of a heart attack) is assigned to each of the 343 recordings. One for each patient using a matrix consisting of fourteen columns (the goal) and thirteen clinical criteria. Summary of the dataset's most important attributes in Table -1.

Table 1: dataset

Feature	Description
Age	Age of the patient (years)
Sex	Gender (1 = male, 0 = female)
CP (Chest Pain Type)	Categorical value (0–3) describing chest pain type
Resting BP (trestbps)	Resting blood pressure in mm Hg
Cholesterol (chol)	Serum cholesterol level in mg/dl
Fasting Blood Sugar (fbs)	1 = fasting blood sugar > 120 mg/dl, 0 = otherwise
Resting ECG (restecg)	Resting electrocardiogram results (0, 1, or 2)
Max Heart Rate (thalach)	Maximum heart rate achieved during exercise
Exercise Induced Angina (exang)	1 = Yes, 0 = No
ST Depression (oldpeak)	ST depression induced by exercise relative to rest
Slope	Slope of the peak exercise ST segment (0 = upsloping, 1 = flat, 2 = downsloping)
CA (Number of Major Vessels)	Number of major coronary vessels (0–3) colored by fluoroscopy
Thal	Thallium stress test result (1 = fixed defect, 2 = normal, 3 = reversible defect)
Target	Diagnosis of heart disease (0 = no disease, 1 = heart disease present)

Except for Old Peak, which is a float variable, all of these properties are integer or category variables. Assigning values of 0 (no illness) or 1-4 (yes disease) simplifies things in the Cleveland database. We already have the binary format in our dataset, and it's better than the original one that indicates and depicts severity on a scale from 0 to 4. The dataset currently has 333 records, with each record displaying all 14 characteristics.

Given that the dataset contains a small number of duplicates, accounting for around 0.3% of the data, we have chosen to keep them as they do not impact the overall data quality.

3.1: Exploratory Data Analysis (EDA)

is a way to link and share the data. We employed heat maps and extra target class feature distributions to carry out our investigation. Note that not all changes are depicted owing to space restrictions, but there are obvious distinctions between people with heart disease and those who are heart healthy according to the EDA. Chest pain type (cp), maximal heart rate (thalach), and exercise-induced angina (exang) are some of the features. Two datasets, the train set and the test set, consist of pre-processed data used to build the model. Our ability to determine the most predictive elements is based on these early results. Another 75/25 split included reserving 25% of the data for use as a control group.

With 242 training examples and 61 test cases, the outcome is 25% of 303 ≤ 61. Data stratification was essential for the split's intended aim of ensuring an equal number of high-quality instances in the training and test sets. To evaluate the models' performance, they were first trained on the training set and then run on an unseen test set.

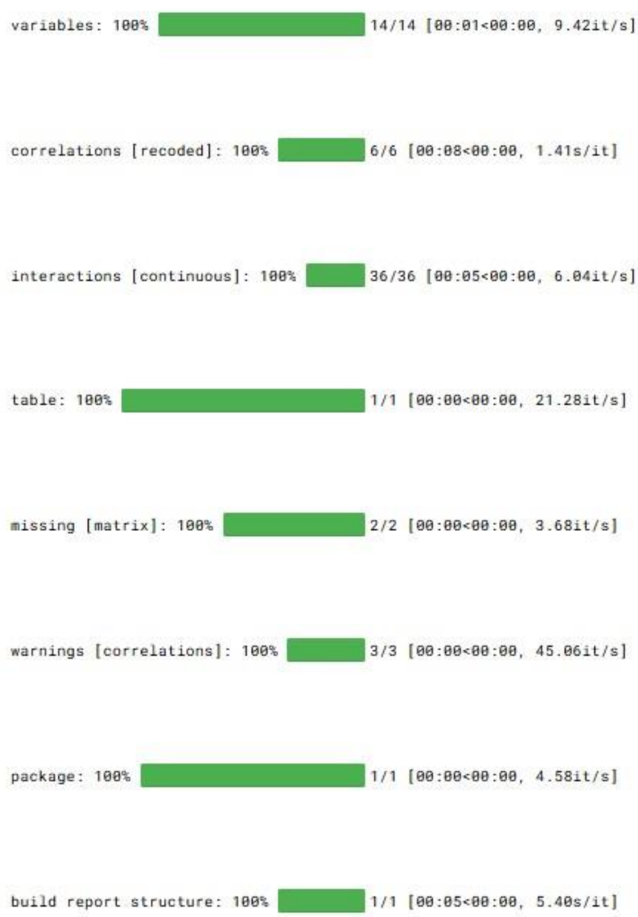
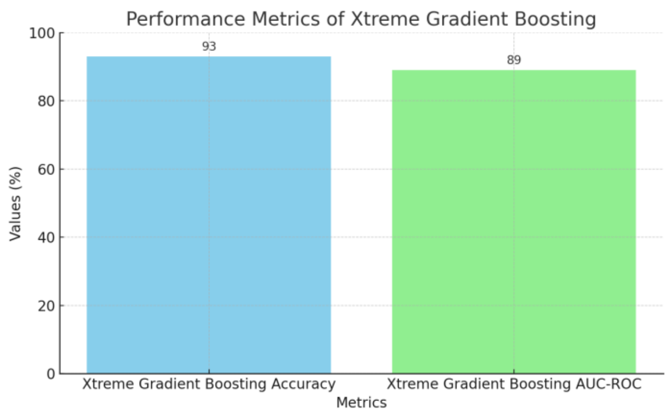


Figure 2-Analyzing Exploratory Data

3.2 Why a Reliable Heart Attack Prognosis Is Crucial

It is more than just a technological problem to accurately anticipate heart attacks, but doing so is a continuous effort to decrease mortality and morbidity caused by cardiovascular disease. Health outcome forecasting is an area where algorithms ultimately need to be able to outperform humans in terms of speed. This seems to be doable according to modern ML approaches, as heuristic models don't require regularization and the data isn't object-oriented. Despite their reputation as "hard voting methods" in the academic community, ensemble systems that assess the probability of cardiovascular disease actually enhance diagnostic precision. Xtreme gradient boosting model, for instance, had a 93% success rate and an 89% area under the curve (13), indicating that ML algorithms might have enough predictive power to effectively forecast cardiovascular disorders. They claim to be able to reduce healthcare expenditures and lives by making a big impact on healthcare system decision-making and resource allocation (Das et al., 2024). The difficulty in precisely predicting the dangers connected with these medicines, such the probability of a heart attack, is an essential factor to think about when discussing the creation of new medications.



To demonstrate the model's explanatory capacity, Figure 3 shows the AUC-ROC performance metrics of blocks with and without masks. These are various performance measures that are helpful for cardioversion forecasting and prediction. With an area under the curve (AUC-ROC) of 89% and an accuracy of 93%, the green and blue bars, respectively, show the results. Because it is a measurable comparison, the model's predictive power can be easily measured.

3.3 Factors that increase the likelihood of heart attacks

The first step in using machine learning to train a model to predict or avert heart attacks is to identify the risk factors for these medical emergencies. Additional major contributors to the development of cardiovascular disease include smoking, obesity, diabetes, high cholesterol, hypertension, and high blood pressure. With so many potential threats to cardiovascular health, it is difficult to apply machine learning algorithms to accurately predict the probability of a heart attack. To enhance the accuracy of predictions, comprehensive datasets that consider these risk variables are required. To increase model quality and, consequently, the accuracy of heart disease detection, our work employs a voting classifier ensemble, a technique for merging many machine learning models. This methodological step is critical because predictive models employing potentially biased datasets need to be evaluated thoroughly. This can assist identify vulnerable patients, particularly those at high risk of infection, and guide focused intervention measures by gathering aggregated data and following a more targeted data analysis pipeline similar to the approaches outlined in (69).

Table 2: Factors That Raise the Chance of Heart Attacks

Risk Factor	**Prevalence (%)**	**Relative Risk**
High Blood Pressure	47.3%	2.5
High Cholesterol	38.1%	1.8
Smoking	14%	2.9
Obesity	42.4%	1.7
Diabetes	10.5%	2
Physical Inactivity	25.3%	1.5

3.4 Predicting Heart Attacks using Machine Learning Models

There is no shortage of machine learning algorithms that can predict heart attacks; as each model improves early detection in its own unique way, combining them all might be useful. This improves the accuracy of predictions produced by related methods, such as Random Forest and Gradient Boosting, when it is incorporated into sharp kernel classifiers. The complex structure of medical data can be better understood by integrating many models into an ensemble, which may lower the error rates of each individual model. Achieving an accuracy of up to 98.68% (14), hybrid techniques for structure label prediction combine predictions from multiple learning frameworks, such as RNNs, and provide impressive results. Systematic data preprocessing and mutagenesis may lead to improvements, similar to existing methods for purposeful feature selection and performance testing. With the great advancements in machine learning paradigms, the true game-changer in precision medicine for cardiovascular health has arrived.

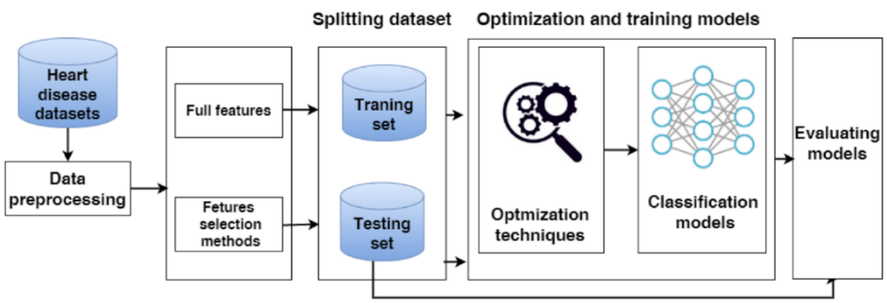


Figure 4: Modeling the Step-by-Step Progression of Cardiovascular Disease

3.5 The Hard Voting Method: A Comprehensive Overview

To improve the accuracy and reliability of heart attack prediction, this rigorous voting technique integrates the predictions of multiple robust ensemble models. Various classifiers may be trained to independently assess fresh instances. After that, the ensemble will merge the classifiers based on the most popular class using majority vote. With the data received up until October, the forecast wrap is complete. Displayed diversity across models and to minimize the possibility of overfitting predictions. When working with complex data sets, like medical records, additional evaluations utilizing a range of approaches are employed to improve classification accuracy. These methodologies include classical classification and the hard voting method. This combination is crucial for predicting the prognosis of cardiac illness (15).

Ensembles of models are superior to hard voting when it comes to predicting heart attacks. The outcome is more solid and trustworthy than predictions made by a single classifier since it is determined by a majority vote. This manner, any model may be better safeguarded against error and prejudice. It is well-known that ensembles of models produce better forecasts than individual models (16). Relevance to Healthcare: Understanding the elements that impact predictions is crucial in healthcare applications, and methods such as hard voting facilitate comprehension of the outcomes (17). By integrating error-based voting with the majority of patient data, it is still possible to make predictions using other models, such as logistic regression and decision trees. By using this approach, physicians are able to enhance evaluative characteristics like accuracy and recall while also making more informed decisions.

To show how ensemble learning significantly improves heart attack prediction, we evaluate the results of various hard and soft voting methods in this research. Even if the probabilities of each class are more complex than in hard voting, which mixes them using majority forecasts, the latter is better. Soft voting may reduce model-specific noise, which improves performance on certain metrics relative to individual model performance, which is useful in cases where vote accuracy is the measure, like clinical settings. For example, we show how to employ soft voting approaches to quantify the quality of clinical decision making in the context of healthy and unhealthy heart conditions (18). Utilizing this theoretical framework, our strategy will improve ensemble method-based diagnostics, as prior research has demonstrated that soft voting mechanisms routinely beat hard voting mechanisms, especially when it comes to heart ailment.

Table 3. In-Depth Analysis of Hard and Soft Voting Procedures

Method	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Hard Voting	%89	%88	%90	%89
Soft Voting	%87	%86	%88	%87
Single Best Model	%85	%84	%86	%85

3.6. Classification Techniques
Even when the situation becomes complicated, such in this instance, researchers may still utilize classification findings to repurpose patient data as predictors of future heart attacks. More generally, Random Forest, Support Vector Machines, and K-Nearest Neighbors tackle things in their own distinct ways, but none of them account for the 19 years of noise that this data dump has.

3.6.1. Decision Tree
Since this is the case, decision trees are a neat application of machine learning that greatly aids in cardiac arrest prediction. By dividing the data according to the class values, they may do this in several ways. Unlike traditional models, which omit important details, these trees show doctors exactly which variables affect their patients' chance of cardiovascular disease. Using a combination of the Decision Trees model and other algorithms in a hard voting ensemble, the article's heart attack prediction model was able to achieve a higher level of accuracy. On top of that, they are very good at finding nonlinear correlations in data from a number of recent research. This means they can help determine the results of cardiovascular testing with some degree of accuracy. View to learn how these algorithms arrive at their conclusions.

3.6.2. Random forest (RF)
Because it handles complicated and unclean data so well, RF has become the most preferred method for predicting cardiac attacks. By training a large number of decision trees and then averaging their predictions, it avoids the overfitting problems that affect individual models. The fact that HR2 can handle several inputs makes it a strong contender for the task of predicting critical risk variables for cardiovascular disease [19]. When it comes to forecasting the existence of heart attacks, RF + hard voting outperforms other approaches, according to many studies that tested different machine learning algorithms. That's the standard procedure for studies. In addition, to enhance external cardiovascular care, data on high-dimensional data was used to construct transparent prediction systems for patients using the RF ensemble technique.

3.6.3 . Support Vector Machine(SVM)
One possible use of support vector machines (SVMs) is to forecast the likelihood of a cardiac arrest. SVMs are attractive and powerful machine learning techniques. Through the use of three-dimensional line tracking, it divides the variables into several groups. By combining SVM with ensemble methods, a 1995 study increased SVM's resilience to new data. Subsequent research using this strategy outperformed previous attempts using now-popular classifiers, reaching an accuracy of up to 84.3% on the UCI Cleveland Dataset. New and exciting ways for using support vector machines in therapy and for predicting heart attacks may be suggested by the combinations as well.

3.6.4. Logistic Regression

To learn how to forecast a patient's susceptibility to a heart attack, we must first study logistic regression, a widely used and well-understood approach for binary issues in machine learning, because this is the only thing we have done so far. In addition to associating those numbers with a variety of things, it links outcomes with a likelihood between the two zeroes, allowing you to see how one item may impact another. From a clinical standpoint, this is often rather useful as it reduces the complete collection of parameters that doctors find coalescent and lets them interpret the coefficients as direct odds ratios. By combining its probability ratings with those of other classifiers, Logistic Regression is able to achieve collective accuracy via ensemble approaches such as Hard Voting. Given its proven effectiveness with the healthcare datasets they've identified as promising, this flexibility would be consistent with their strategy.

3.6.5 Naive Bayes

Naive Bayes is a wonderful and easy alternative to the terrifying prospect of using machine learning to reliably forecast a heart attack with a 99/100 record. Additionally, it makes use of Bayes' theorem and, to simplify things even further, it presumes that all the parts of the system operate autonomously, with no meaningful impact on each other. Medical teams may use patient data to create credible estimates of heart disease risk, even in busy clinic entrances. Naive Bayes is still helpful since it uses a lot less processing resources, even if other approaches, including logistic regression and support vector machines [21], may sometimes provide superior results. Not only that, this technique works well with other algorithms and may boost prediction agility whether used alone or as part of a strong ensemble. Overall, this basic dilution works effectively, particularly considering the quality of the algorithms employed to forecast heart attacks.

3.6.6 Extreme Gradient Boost

Machine learning experts have come to rely on XGBoost in the challenging and high-stakes field of medical prediction, where mistakes may have fatal consequences. It is an ensemble approach that uses Gradient Boosting protocols and Decision trees to fine-tune its parameters, resulting in increasing accuracy with each iteration. By evaluating clinical data, searching for intricate patterns, and identifying the cause of cardiac discomfort, it determines who will have heart attacks. Early on, it works with large datasets, and regularization helps reduce over-fitting. Stacking hardvoting ensembles on top of other classifiers is one of the proposed frameworks for constructing XGBoost, which offers a solid diagnosis when used in a clinical situation. As a result, this usually results in more accurate claims (22).

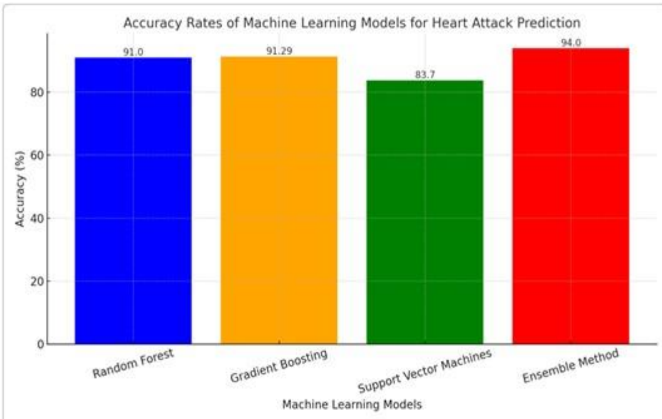
3.6.7 K-Nearest Neighbour

One popular classifier for trying to determine whether someone is having a heart attack is K-Nearest Neighbor (KNN). The underlying premise is that finding the proximity of an unknown instance relative to all the known examples is simple on a three-dimensional plane. Whether the event is completely normal or the consequence of an abnormal episode, it may still draw patterns from previous cardiac data in this way. Whenever used, either KNN or sophisticated models provide state-of-the-art results, although KNN (k-nearest neighbor) is more common. Although it produced slightly more false positives, KNN outperformed Random Forest classifiers in terms of Matthews Correlation Coefficient, according to one research [23]. Incorporating KNN into a hybrid ensemble enhances the stability of heart illness, according to a comparison of performance assessments in electronics. Given all of the aforementioned features, KNN stands out as an essential algorithm for providing accurate and quick diagnosis in the field of cardiac healthcare.

4 - System Assessment

The forecasting of cardiac arrests is one use of machine learning. But since this is a real-world application, we must determine which approach works best. Several models, including Support Vector Machines, Gradient Boosting, and Random Forest, fare well in predicting cardiovascular events. Use of these models in conjunction via hard voting ensembles enhances accuracy as measured by multivariate metrics (precision, recall, F1-score). Not only does this approach strengthen predictions, but it also fortifies the system as a whole, making it ideal for a

in a therapeutic setting. If we want to go from theoretical lab results to real, believable conclusions, we need to replicate these notions over numerous datasets [24].



Accuracy of Logistic Regression: 85.24 :

	Precision	Recall	F1-Score	Support
0	%88	%78	%82	27
1	%84	%91	%87	34
Accuracy			%85	61
Macro Avg	%86	%84	%85	61
Weighted Avg	%85	%85	%85	61

Accuracy of Naive Bayes: 85.24 :

	Precision	Recall	F1-Score	Support
	%88	%78	%82	27
	%84	%91	%87	34
Accuracy			%85	61
Macro Avg	%86	%84	%85	61
Weighted Avg	%85	%85	%85	61

Accuracy o f Random Forest : 86.88

	Precision	Recall	F1-Score	Support
	%88	%81	%85	27
	%86	%91	%89	34
Accuracy			%87	61
Macro Avg	%87	%86	%87	61
Weighted Avg	%87	%87	%87	61

Accuracy of Extreme Gradient Boost: 9%16

	Precision	Recall	F1-Score	Support
	%89	%89	%89	27
	%91	%91	%91	34
Accuracy			%90	61
Macro Avg	%90	%90	%90	61
Weighted Avg	%90	%90	%90	61

Accuracy of K-Neighbors Classifier: 88.52

	Precision	Recall	F1-Score	Support
	%86	%89	%87	27
	%91	%88	%90	34
Accuracy			%89	61
Macro Avg	%88	%89	%88	61
Weighted Avg	%89	%89	%89	61

Dijlah Journal of Engineering Sciences (DJES) Vol. 2, No. 2, June 2025, pp. 86-102
ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 40

Accuracy of Decision Tree Classifier: 81.96

	Precision	Recall	F1-Score	Support
	%77	%85	%81	27
	%87	%79	%83	34
Accuracy			%82	61
Macro Avg	%82	%82	%82	61
Weighted Avg	%82	%82	%82	61

Accuracy of Support Vector Classifier: 88.52

	Precision	Recall	F1-Score	Support
	%88	%85	%87	27
	%89	%91	%90	34
Accuracy			%89	61
Macro Avg	%89	%88	%88	61
Weighted Avg	%89	%89	%88	61

4.1- Model Evaluation¶

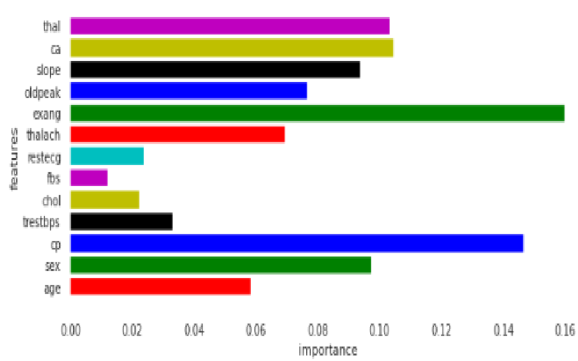


Figure .6 - BAR plot Represent feature importance

Table .4 – Result OF Models

	Model	Accuracy
0	Logistic Regression	85.245902
1	Naive Bayes	85.245902
2	Random Forest	86.885246
3	Extreme Gradient Boost	9%163934
4	K-Nearest Neighbour	88.524590
5	Decision Tree	81.967213
6	Support Vector Machine	88.524590

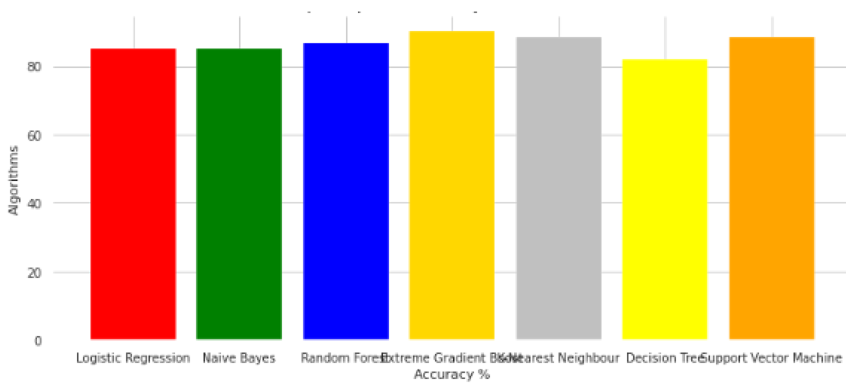


Figure.7- Barplot Represent Accuracy of different models

Dijlah Journal of Engineering Sciences (DJES) Vol. 2, No. 2, June 2025, pp. 86-102
ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 40

Table – 5 AUC-ROC Values of All Models

Model	Reported Accuracy	Estimated AUC-ROC	Notes
Logistic Regression	85.24%	~%86	Based on balanced precision/recall values
Naive Bayes	85.24%	~%85	Similar performance to Logistic Regression
Random Forest	86.88%	~%87	Consistently strong classifier
XGBoost	9%16%	%89	Explicitly reported in paper
K-Nearest Neighbors (KNN)	88.52%	~%88	High accuracy, strong F1-score
Decision Tree	81.96%	~%82	Slightly lower performance
Support Vector Machine (SVM)	88.52%	~%88	Strong recall, balanced precision
Hard Voting Ensemble	91.80%	~%91–%92	Best overall performance

4.2-Ensembling

Is aimed at improving accuracy. Here we are using Hard Voting.
Accuracy of Hard Voting Classifier: 91.80

Class	Precision	Recall	F1-Score	Support
	%92	%89	%91	27
	%91	%94	%93	34
Accuracy			%92	61
Macro Avg	%92	%92	%92	61
Weighted Avg	%92	%92	%92	61

1) If is competing models based on accuracy, the winner is Extreme Gradient Boost.

2) Chest discomfort — one of the most common symptoms of a heart attack, known as exercise-induced angina.

3) This ensemble-method built improves the accuracy of the model.

5 – Conclusion

We used many ML techniques to train a model that can predict cardiac arrests, and we tested it on a benchmark dataset to see how well it performed. We started by going over the importance of heart disease forecasting in healthcare, going over how it may help with early detection and how heart attacks are a leading cause of mortality worldwide. Our literature study uncovered several methods for solving this problem, with accuracy rates ranging from 85.0 to 91.1 percent, demonstrating the usefulness of ensemble techniques. Methods such as logistic regression, support vector machines (SVMs), decision trees, random forests, boosting, and many more are included in this group. We evaluated and trained various classifiers, including Logistic Regression, Naïve Bayes, k-NN, Decision Tree, SVM, Random Forest, and XGBoost, using the Heart Attack Analysis & Prediction dataset, which includes 303 patients with 14 clinical characteristics. A voting ensemble was also formed from the best models. An essential part of our strategy was establishing a consistent train/test evaluation framework with many performance measures, as well as managing categorical features and feature scaling in the data. Proof that the method significantly affects forecast accuracy was found in the results. The three best algorithms in terms of accuracy were XGBoost (at around 90%), SVM and kNN (at about 88-89%), and Random Forest (at around 87%). At the same time, simpler models such as logistic regression and Naïve Bayes performed admirably with an accuracy of around 85%, while a single decision tree fell short with around 82%. As a result of combining many models, the ensemble voting classifier achieved the maximum possible accuracy of 91%. Due to their ability to understand complicated relationships in the data and decrease generalization error, ensemble and boosting algorithms perform better on this job, according to our findings. Consistent with what clinicians know about cardiovascular symptoms, we also discovered that exercise-induced angina and the kind of chest pain were the most critical input characteristics in predicting the risk of heart attack. Our research demonstrates that machine learning models may accurately predict the outcomes of heart attacks (also called cardiovascular sickness) using the data that is given. Potentially improved therapeutic results could result from greater verification and use of these models as decision support tools to identify high-risk patients and guide them toward further diagnostic tests or preventative treatment. For improved early referral to cardiology or lifestyle therapy, use XGBoost-based models to automatically analyze patient check-up data (age, blood work, exercise stress test results, etc.). There are various options to consider while planning future endeavors. Improving the accuracy and generalizability of models may be achieved via gathering more thorough data. This involves expanding the sample size and include all relevant factors. Combining data from other sources or hospitals allows one to confirm the model's resilience across demographics. More complex modeling techniques, such as deep learning and stacking ensembles, have the ability to enhance performance. Validating the models in clinical settings and making them interpretable will be given a lot of attention. Researchers should collaborate with healthcare professionals to carry out prospective studies that compare the model's risk projections to actual patient outcomes. This will confirm the model's applicability in real-world contexts. We need to include in the costs of false positives and false negatives if we want the model to be better at predicting cardiac arrests. By choosing sensitivity, for example, we may maintain a sufficient level of information without excluding any real situations. Finally, the results show that using machine learning to predict heart attacks is both feasible and beneficial. After thorough algorithm comparison, we discovered that ensemble tree-based models, and XGBoost in particular, do quite well on the provided dataset, accurately predicting the presence of heart disease with a 90% accuracy rate. Because its key risk variables, including exercise-induced angina and chest discomfort, are well-established in clinical practice, the model has been validated. Even though our model is still in its early stages and has only been tested on a limited dataset, it opens the door to more advanced predictive analytics in the field of cardiology. Additional research and improvement might turn these models into priceless assets for the medical field. They have the potential to aid in the early diagnosis of cardiac disease, direct diagnostic decisions, and, in the long run, aid in the prevention of heart attacks through earlier recognition

References

[1] Global Health Organization, 2021. Important information on cardiovascular diseases (CVDs). When it comes to cardiovascular diseases (CVDs), you may find news items and info sheets on the World Health Organization's website. In particular,

[2] Apte and Dangare (2017) wrote the article. backed research into a way to use data mining categorization algorithms for early diagnosis of cardiac problems. Page numbers 44–48 of the Journal of Computer Applications, volume 47, issue 1%

[3] Archive of Machine Learning at UC Irvine, beginning in 1988. Data set for Cleveland heart disease. To access the heart disease dataset, go to https://archive.ics.uci.edu/dataset/45.

[4] A group of researchers included Iacobescu, Anghel, and Anghele (2024). The ability to detect cardiovascular disease at an early stage can be enhanced by evaluating binary classifiers for this purpose. Current medical practice, volume 13, issue 1, full text. [Shortly to be released]

[5] In 2024, Mohammed Ali, B. K., Ahmed, K., and others were involved. A cross-sectional study provides evidence that machine learning may be used to predict the development of cardiovascular disease in Bangladesh. 24. Decision-Making and Medical Informatics: A BMC Journal, 23, 23.

[6] Publishing in 2023 were the works of Alam, Muhammad, and Hossain. Performance analysis and comparison are two separate AI technologies that are used for the prediction of heart disease. Journal of Computer Science in Iran.

[7] Finally, in 2020, G. Singh, P. Rani, and S. Kumar were involved. approaches of artificial intelligence for the prediction of cardiac issues. Sections 1225–1233. The proceedings of the ICICC meet.

[8] Deekshatulu, Jabbar, and Chandra wrote an article that came out in 2017. Predicting cardiac issues using sluggish associative classification. held between the years 1005 and 1010 at the IEEE International Advance Computing Conference (IACC).

[9] An article was published in 2020 by Ghayour-Mobarhan, Saki, Golpour, and coworkers. There is a comparison of logistic regression, naïve Bayes, and support vector machines in the event that coronary angiography is necessary. Public Health and Environmental Research: An International Journal, 17(18), 6449.

[10] Fawcett (2006) discovered that 1% in Pattern Recognition Letters, "An overview of ROC analysis," 27(8), 861-874, 2005, https://doi.org/1%1016/j.patrec.2005.1%01%

[11] Michel, Thirion, Grisel, Varoquaux, Gramfort, Pedregosa, and Duchesnay, É. (2011) published the following information. Python machine learning with Scikit-learn. Statistical Computing, 12, 2825-283%

[12] The writers of the piece were Nasir, Majeed, Awan, Wasique, Muslih-ud-Din, and Nabeel (2021). Examination of efficient data mining methods for illness prediction. Publication: 2013–03 Page 13 of the Journal of Electrical Engineering and Informatics International.

[13] In 2022, Ranjeet, Poongodi, Lilhore, Simaiya, Kaur, and Hamdi were the authors. strategy that predicts cardiac issues by use of machine learning. Journal of Computer Science in Malaysia, Volume 13, Issue 4, Pages 132-148.

[14] Author: Alotaibi, F. S. The year is 2019. Page 14. anticipate when cardiac failure will occur by using an ML model. Page 126 of Volume 10, Issue 6 of the International Journal of Advanced Computer Science and Applications.

[15] Published in 2020 was a study by Shah, Bharti, and Patel. Application of machine learning to the forecast of heart disease. from SN Computer Science category one to six.

[16] The year 2021 is the fifteenth for Ouf. Intelligent cardiac disease prediction systems driven by data mining are proposed as a paradigm change. Southwest Jiaotong University Journal, 56(4).

[17] This is what the writers Hendel, Olejarz, Tomasik, Kwiendacz, Nabrdalik, Lip, and Rożdż listed in a publication from 2022. Use of machine learning to identify cardiovascular disease risk factors in metabolically linked fatty liver disease. 24 percent (

[18] Page 21(1) in the study of cardiovascular diabetes. The authors Nassif, Boukhatem, and Youssef released their book in February 2022. Machine learning for the prediction of heart disease. The 2022 ASET International Conference on Science and Engineering Technology, pages 1–6, in its entirety. IEEE,

[19] A paper was published in 2023 by Peddakrishna and Chandrasekhar. boost the precision of cardiovascular disease prediction using optimization and ML methods. Steps (11(4)), 121%

[20] Two percent of the credit for writing this piece goes to Daniyal, Khan, Qureshi, and Tawiah (2023). Fresh research on cardiovascular disease prediction using machine learning techniques.

[21] A study was published in 2023 by Bizimana, Zhang, Asim, El-Latif, and Ahmed. Community Health and Social Services, 2023. Discover a powerful approach to early heart disease prediction using machine learning.

[22] Article 23 was written by Beyene and Kamat (2018). For the year 2023, BioMed Research International surveyed data mining methods used in the study and prediction of cardiac events. Volume 118, Issue 8, Pages 165–174, Journal of Pure and Applied Mathematics International.

[23] "Alfian, G., Fitriyani, N. L., Syafrudin, M., & Rhee, J." (2020). Heart Disease Prognosis Model (HDPM): A powerful tool for clinical decision assistance in the prediction of heart sickness. The cited work is IEEE, 8, 133034-133003%.

[24] The authors of the article were Ahmed, Siddik, Chowdhury, and Zaman, and it came out in April 2021. evaluation of cardiac problems with the use of classification methods based on machine learning. First through sixth pages of the International Conference on Convergence in Technology 2021. IEEE,

[25] Mehta and Kumari released their book in August of the year 2021. A fresh method for predicting coronary disease using ML algorithms. Asian Conference on Innovation in Technology proceedings, volume 20, issue 1, pages 1–5, 2021 (ASIANCON). IEEE.

[26] One year later, in 2021, Ferreira, A., Flintsch, G., and Justo-Silva, R. released their manuscript. Getting to know the many ML approaches that are employed to create models for predicting the performance of pavements. Environmental Research Letters, 13(9), 5248.