# *PERFORMANCE EVALUATION OF MACHINE LEARNING ALGORITHMS TO PREDICT CARDIOVASCULAR DISEASE*

S.Hemalatha -Assistant Professor(Sr.G)
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode
hemalatha@kongu.ac.in

D.Niruba
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode
nirubad.21mca@kongu.edu

S.Nandhakumar
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode.
snandhakumar.21mca@kongu.edu

R.Venkatesh
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode
rvenkatesh.21mca@kongu.edu

***Abstract*** **- Heart disorders are a major factor in many deaths across the world. Many people in their middle or later years suffer from heart conditions, which typically result in significant health problems include heart attacks and strokes. In order to stop any significant health problems before they arise, it is vital to detect and anticipate cardiac illnesses. Various machine learning techniques, including Neural Networks (NN), Decision Trees (DT) and Naive Bayes (NB), Random Forest (RF), Logistic Regression (LR), Support Vector Machines (SVM), K-Nearest Neighbor (KNN) and XG Boost, were used in this paper as a preliminary study and examination to predict heart diseases at various evaluation stages. Results reveal that the Random Forest approach performed better than the other techniques and had a 98.25% prediction accuracy rate.**

*Keywords - Machine Learning, Heart Disease, Decision Trees, Naive Bayes, Neural Networks.*

## I. INTRODUCTION

### Cardiovascular Disease

The term "cardiovascular diseases" (CVDs) refers to a variety of heart and blood vessel conditions. Heart attacks and strokes are often sudden, severe occurrences that are mostly brought on by a blockage that stops the flow of blood to the heart or brain. Fatty deposits that have accumulated on the inner walls of the blood arteries that supply the heart or brain are the most frequent cause of this. Blood clots or haemorrhage from a brain blood artery can both result in strokes. Numerous machine learning (ML) methods are being used more often to forecast cardiovascular disease. Our goal is to evaluate and describe the overall prediction power of ML algorithms in cardiovascular disorders. CVD is classified into four types:

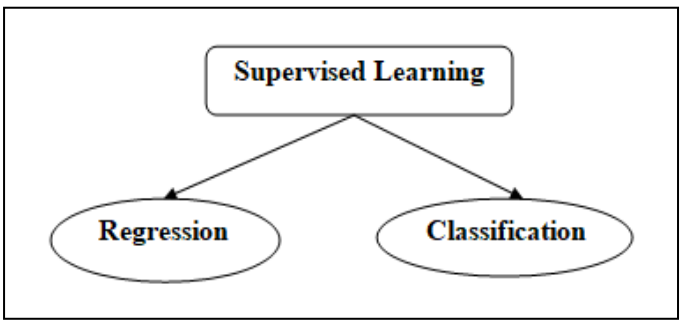
- peripheral arterial disease
- aortic disease
- coronary heart disease
- stroke

### Machine Learning

The most effective testing technique is machine learning. Computers are trained to mimic human talents through a large field of research called machine learning (ML). The field of machine learning is related to artificial intelligence. Testing and training are the foundational ideas of ML. A system gains knowledge from data and experience, and it then applies that information to apply tests to various sorts of demands in accordance with an algorithm. The three different categories of machine learning techniques are supervised, unsupervised, and reinforcement.

### Supervised Learning

Machines are educated using appropriately "labelled" training data, and then utilizing that data to predict the outcome, is known as supervised learning. The term "labelled data" refers to input that has already been assigned the appropriate output. In supervised learning, the computers are taught to accurately anticipate the output by the supervisor, who is the training data that is given to them. Problems with supervised learning may be further separated into two categories:




**Regression:** If there is a correlation between the input and output variables, regression procedures are applied. It is used to forecast continuous variables like weather, market

trends, and other things. Several well-liked regression methods that fall under supervised learning are listed below:

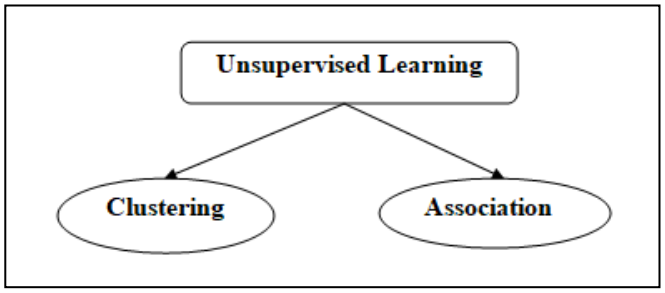
- Regression Trees
- Linear Regression
- Non-Linear Regression
- Polynomial Regression
- Bayesian Linear Regression

**Classification:** When the output variable is categorical, such as Yes-No, Male-Female, True-False, and so on, classification techniques are utilized.

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines

**Unsupervised Learning**

The unsupervised learning is a machine learning approach in which models are not supervised using training dataset. The underlying patterns and insights from the provided data are instead discovered by the models themselves. It may be related to learning, which occurs in the human brain while learning new things. The unsupervised learning method may be divided into two different categories of issues.




**Clustering:** Clustering is a way of organizing the items into clusters so that the objects that have the most similarities stay in one group and share less or no similarities with the objects in another group. Data items are classified based on the existence or lack of commonality found through cluster analysis.

**Association:** Finding the connections between variables in a huge database is done using an unsupervised learning technique called an association rule. It identifies the group of objects that appear collectively in the collection. Marketing strategy is more successful when association rule is used.

Popular unsupervised learning algorithms are listed below:

- K-means clustering
- Hierarchal clustering
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

**Reinforcement**

Reinforcement Learning is a feedback-based Machine Learning approach in which an agent learns how to behave in a given environment by executing actions and seeing the outcomes of those actions. For each positive activity, the agent receives positive feedback; for each poor action, the agent receives negative feedback or a penalty. In contrast to supervised learning, the agent learns autonomously via feedbacks in Reinforcement Learning.

## II. LITERATURE SURVEY

In [1] Aishwarya P, Jeny J.R.V, Reddy N.S, and Samreen has concluded that the main goal of the project is to increase accuracy and diagnose heart disease by implementing machine learning classifier techniques.

In [2] Alfian G, Fitriyani N.L, Rhee J and Syafrudin M. This study proposes an effective heart disease prediction model (HDPM) for a CDSS that includes Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate outliers, a hybrid Synthetic Minority Oversampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to balance the training data distribution, and XGBoost to predict heart disease.

In [3] Amar M, Louridi N and Ouahidi B E. In this research, the quality of cardiovascular disease prediction is improved by adopting a better preprocessing step. A comparison of multiple Machine Learning algorithms employing accuracy, precision, f1-score, and recall performance measures was performed to demonstrate the outcomes.

In [4] Bertsimas D, Mingardi L and Stellato B. In this study, it offer a unique approach for predicting the kind of real-time ECG by extracting information associated to ECG (less than 30 milliseconds). It makes use of the XGBoost algorithm, one of the most effective machine learning techniques, to train models that achieve out-of-sample F1 Scores between 0.93 and 0.99. This is the first analysis, that compares high performance across hospitals, nations, and recording standards.

In [5] Boukhatem C, Youssef H.Y and Nassif A.B. In order to

construct the prediction models, the study exhibited four classification techniques: Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Nave Bayes (NB). Before creating the models, processes for data preprocessing and feature selection were taken. On the basis of accuracy, precision, recall, and F1-score, the models were assessed. The SVM model has the highest accuracy, 91.67%.

In [6] Chakarverti M, Rajan R and Yadav. This proposed study describes clustering and classification of the input information for heart disease is executed with the help of k-means clustering algorithms is applied for clustering data.

In [7] Chaubey A.K, Gogoi A, Mahesh T.R, Rohit S and Sarveshvar M R. In this paper, ML classification techniques built on supervised learning can make diagnosing cardiovascular illness easier. Numerous machine learning (ML) methods, such as Naive Bayes Classifier, Logistic Regression Classifier, and Random Forest, are being used to distinguish between those who have cardiac illness and those who do not. The data is also normalised for better results because the dataset contains certain unnecessary elements that are deleted throughout the data cleaning process.

In [8] G. S and S. K. J. has concluded that by using the classification method and comparing the parameters, this paper system evaluates those parameters using data mining classification techniques.

In [9] Hameetha Begum S and Nisha Rani S.N. This study evaluated and compared various common supervised machine learning approaches for predicting cardiac disorders using medical information from the UCI Machine learning repository. This research investigates the performance of several models, such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Logistic Regression models.

In [10] Kanchana S and S Usha. Throughout the proposed study, machine learning (ML) techniques such as Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), Decision Tree (DT), KNN, Support Vector Machine, and XGBoost will be employed to diagnose heart disease. Implementing these algorithms based on the electronic medical record increases the likelihood of accurate cardiac disease identification and diagnosis. The system improves performance by finding the most accurate model.

In [11] KN L, Kumari N.K.R, N. R, S. N and V. K. The primary goal of this work is to create a basic machine learning model that will help in the accurate identification of heart disease. Machine learning techniques such as Logistic Regression, K-Nearest Neighbor (K-NN), Decision Tree, Naive Bayes, Random Forest, and Support Vector Machine are used in this work to produce better results.

In [12] Kumar Thakkar H and Shukla H. In this research, ML classifiers are built and a comparative analysis is performed for the prediction of robust heart disease. Five ML classifiers are constructed and fully tested on the Cleveland Heart Disease Data set, including Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). A comparison of five performance criteria demonstrates the use of ML classifiers for heart disease prediction.

In [13] Mahalakshmi K and Sujatha P. In this study, decision tree, naive bayes, random forest, support vector machine, K-nearest neighbour, and logistic regression algorithms are used to predict the existence of heart disease. Using metrics like Accuracy, Precision, AUC, and F1-score, the algorithms' performance was examined. According to the experimental findings, Random Forest outperforms other supervised machine learning algorithms for predicting heart disease, with an accuracy of 83.52%. Random forest classifiers' F1-Score, AUC, and accuracy score are 84.21%, 88.24%, and 88.89%, respectively.

In [14] Mohan S, Srivastava G and Thirumalai C This paper proposed a novel method that goal at finding significant features by applying machine learning technique resulting in upgrading the accuracy in the prediction of cardiovascular disease.

In [15] Ouyang S. This study helps in cardiac disease prediction by describing the currently used machine learning diagnostic and prediction methods, highlighting the similarities and peculiarities of these approaches, and discussing the obstacles and future advances. The results show that machine learning approaches may be used to treat a wide range of heart disorder.

## III. PROPOSED METHODOLOGY

The procedure is divided into five key phases. The first and most important stage is to acquire the dataset. Following dataset collection, data preparation occurs, in which the raw dataset is turned to a clean dataset. Following that, feature selection entails selecting acceptable qualities for the prediction system. This is used to improve the system's efficiency. The goal of feature selection was to identify the critical characteristics needed to diagnose the existence of cardiovascular disease. The classification method is applied to the dataset. At the end, the performance of each classification algorithm is evaluated.

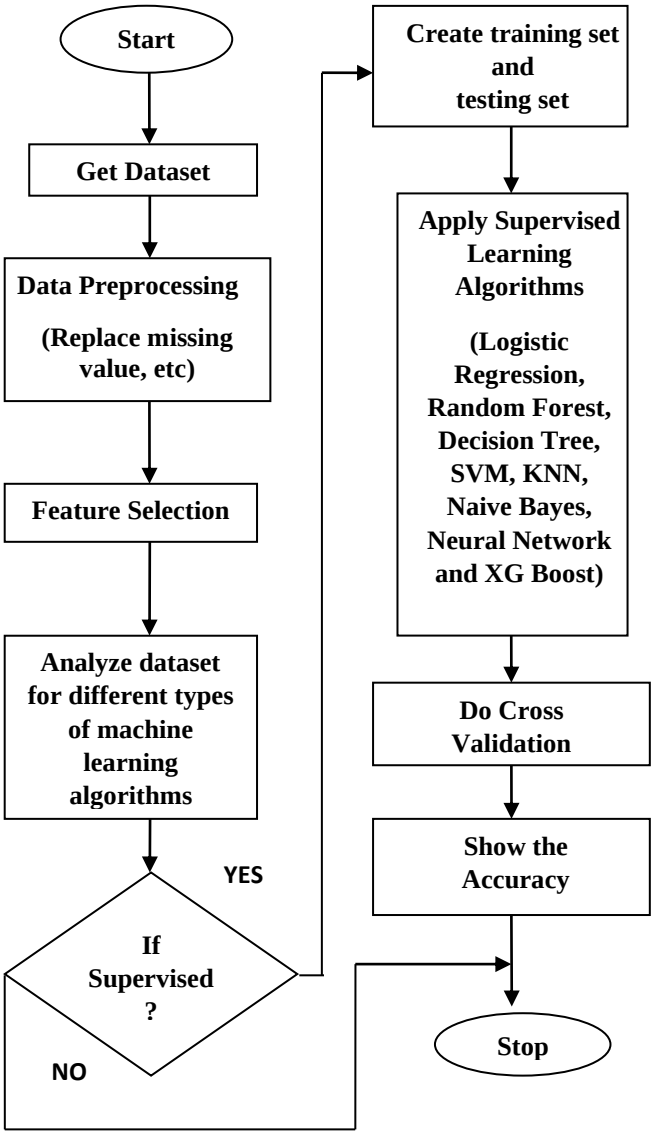



**Fig1 Proposed work**

### A. DATA COLLECTION

With consideration for their history of cardiac issues and in accordance with other medical concerns, a structured dataset of individuals was collected. There are several different disorders that can cause cardiovascular disease is impacted. The majority of deaths among middle-aged persons, according to the World Health Organization (WHO), are caused by cardiovascular illnesses. We use a data set with the medical histories of 690 distinct patients across various age ranges.

The medical features of the patient, such as age, resting blood pressure, fasting sugar level, etc., are included in this dataset, providing us with the much-needed information we need to determine whether or not the patient has been diagnosed with a cardiac condition. The dataset has been collected from Kaggle. This dataset on heart disease was obtained from the UCI repository. The pattern that enables the identification of patients at risk for developing cardiovascular disease is retrieved from this dataset. There are 14 columns and 690 rows in this dataset, where each column represents a single entry.The various attributes are listed in Table 1.

| S. No | Attribute | Description | Type |
|---|---|---|---|
| 1. | Age | Age of the patient | Numerical data |
| 2. | Sex | Patient's Gender | Nominaldata |
| 3. | Cp | Type of chest pain | Nominal data |
| 4. | Trestbps | Resting blood pressure. | Numerial data |
| 5. | Chol | Serum cholesterol | Numerial data |
| 6. | Fbs | fasting blood sugar | Nominal data |
| 7. | Restecg | Resting electrocardio graphic results | Nominal data |
| 8. | Thalach | maximum heart rate achieved | Numerial data |
| 9. | Exang | Exercise induced angina | Nominal data |
| 10. | Old peak | ST depression | Numerial data |
| 11. | Slope | the slope of the peak exercise ST segment | Nominal data |
| 12. | Ca | Number Of major vessels | Nominal data |
| 13. | Thal | thalassemia | Nominal data |
| 14. | Target | 0 or 1 | Nominal data |

**Table 1 : Various Attributes**

### B. DATA PREPROCESSING

The fundamental purpose of data preparation is to transform the unclean dataset into a clean dataset. When addressing missing numbers, noisy data, and outliers, it is helpful. Preprocessing data can assist to cut down on duplicate values. It changes the raw data into a format that is legible and understood. In our dataset, Six entries with missing values were present. We decided to eliminate those six records, making 684 in total. A binary class was generated from the target attribute as well, 1 as the presence of a disease and 0 as absence of disease. Data cleaning, data integration, data reduction, and data transformation are the four primary methods of data preparation.

- Data cleaning - Process of removing incorrect values, duplicate data in the dataset.
- Data integration - Process of combining multiple data to create a unique set of information.
- Data reduction – Process of reducing the number of data records by eliminating the invalid data.
- Data transformation – Process of converting the raw data into structured format without affecting the original content.

### C. FEATURE SELECTION

The selection of appropriate attributes for the prediction system is referred to as attribute or feature selection. This is used to increase the efficiency of the system. Feature selection aimed to recognize the important attributes in order to diagnose the presence of cardiovascular disease. Firstly, the dataset contains 14 attributes; 12 of them were clinical features. While 'age' and 'sex' attributes were patient demographic information. After applying various technique to detect the important attributes. Based on our findings, nine attributes were identified: "sex", "cp", "fbs", "restecg", "exang", "oldpeak", "slope", "ca" and "thal".

### D. CLASSIFICATION ALGORITHMS

The training phase gathers the features (independent variables) from the dataset, and the testing phase (which contains dependent variables) is used to establish how the suitable model performs for prediction. The dataset was split into two phases: training (80%) and testing (20%)..

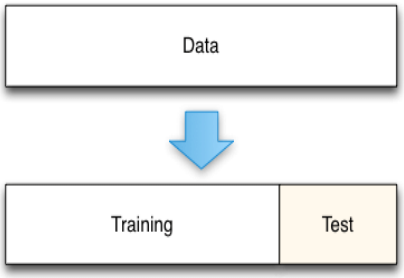



Fig.2 Train test Split

The dataset is classified using a different machine learning technique. They are Logistic Regression, Random Forest, Decision Tree, SVM, KNN, Navie Bayes, Neural Network, and XGBoost.

#### LOGISTIC REGRESSION

Logistic regression is used for classification and regression methods, however it is most commonly employed for classification algorithms. It is used to evaluate the relationship between one or more independent variables and one or more dependent variables.

It is used to solve binary classification issues. In logistic regression, instead of fitting a straight line or hyperplane, the logistic function is used to compress the output of a linear equation between 0 and 1. Because logistic regression has 9 independent variables, it is suitable for classification.

#### RANDOM FOREST

Random forest is a supervised learning technique that may be used for classification and regression issues, but it is most commonly employed for classification. It is based on the ensemble learning strategy, which is a method of combining numerous classifiers to solve a complicated issue in order to increase the efficiency of a performance model.

During training, random forest creates many decision trees, and the output is the modal class or the prediction of individual trees. It helps in classification and prevents overfitting. When the dataset has a large number of features, the accuracy improves. Random forest is appropriate for huge datasets. Similarly, in CVD prediction, the random forest integrates all retrieved information to produce the best accuracy.

#### DECISION TREE

Decision trees are a non-parametric supervised learning approach that may be used for classification and regression issues, but they are most commonly employed for classification. A decision tree classifier builds a tree to classify data by constructing a set of rules. The method begins predicting the class at the root node, compares the root values to the dataset attribute, and then proceeds to the next node depending on the comparison.

#### SUPPORT VECTOR MACHINE

The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space so that we may simply place new data points in the proper category in the future. A hyperplane is the best decision boundary.

SVM selects the extreme points/vectors that help in the creation of the hyperplane. These extreme examples are referred to as support vectors, and the method is known as the Support Vector Machine.

**K NEAREST NEIGHBOR**

The K-NN method assumes similarity between the new case/data and existing cases and places the new case in the category that is most similar to the existing categories. The K-NN method maintains all existing data and uses similarity to classify new data points. This implies that when new data is generated, it may be quickly classified into a well-suited category using the K- NN algorithm.

The K-NN algorithm may be used for both regression and classification, however it is more commonly used for classification tasks. K-NN is a non-parametric method, which means it makes no assumptions about the underlying data.

It is also known as a lazy learner algorithm because it does not immediately learn from the training set; instead, it stores the dataset and then takes an action on it during classification.

**NAÏVE BAYES**

The Nave Bayes method is a supervised learning technique that uses the Bayes theorem to solve classification issues.It is mostly used in text classification with a large training dataset.

The Nave Bayes Classifier is a simple and effective Classification method that helps in the development of rapid machine learning models capable of making quick predictions.The Bayes theorem and the conditional probability rule underpin the Nave Bayes classifier. All properties in a given dataset may be easily interpreted and used, as well as individually analysed.

**NEURAL NETWORK**

The phrase "Artificial Neural Network" comes from biological neural networks, which create the structure of the human brain. Artificial neural networks, like the human brain, include neurons that are interconnected to one another in various levels of the networks. These neurons are referred to as nodes.

In Artificial Neural Networks, dendrites from Biological Neural Networks represent inputs, cell nucleus represents nodes, synapse represents weights, and axon represents output.

**XGBOOST**

Gradient boosted decision trees are implemented in Python via the XGBoost package, which is designed for speed and execution, which is the most essential component of machine learning.Scholars introduced the Python XgBoost (Extreme Gradient Boosting) package at the University of Washington. It is a Python package built in C++ that helps ML model methods by training for Gradient Boosting.

It is the top machine learning package for regression, classification, and ranking tasks, and it supports parallel tree boosting.To understand XGBoost, you must first understand the machine learning ideas and methods on which it is based: supervised machine learning, decision trees, ensemble learning, and gradient boosting.

## E. PERFORMANCE MEASURES

Estimating the performance of a machine learning model is the most important step while building a most effective model. Performance measure is used to evaluate the quality of a model. Accuracy can be calculated as the number of correct predictions to the total number of predictions.

This section displays the outcomes of using Random Forest, Decision Tree, Naive Bayes, and Logistic Regression, SVM, KNN, Neural Network and XGBoost. The metrics Accuracy score, Precision (P), Recall (R), and F-measure are used to analyze the algorithm's performance. The accurate measure of positive analysis is provided by the precision metric (stated in equation (2)). Recall [stated in equation (3)] specifies the amount of actually correct positives. Equation (4) refers to the F-measure. examines accuracy.

The pre-processed dataset is used to carry out the tests in the experiment, and the above mentioned techniques are explored and used. The confusion matrix is used to generate the above mentioned performance metrics. The model's performance is described by the Confusion Matrix. Table 2 shows the confusion matrix produced by the suggested model for various methods. Table 3 displays the accuracy score achieved for the Random Forest, Decision Tree, Logistic Regression, Naive Bayes, SVM, KNN, Neural Network and XGBoost classification techniques.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Precision} = (TP) / (TP + FP) \quad (2)$$

$$\text{Recall} = (TP) / (TP + FN) \quad (3)$$

$$\text{F–Measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

where,

True Positive (TP) : the patient has the disease and the test is positive.
False Positive (FP) : the patient does not have the disease but the test is positive.
True Negative (TN) : the patient does not have the disease and the test is negative.
False Negative (FN) : the patient has the disease but the test is negative.

Confusion matrix – a table which is used to define the performance of a classification algorithm.

| Algorithms | TP | FP | FN | TN |
|---|---|---|---|---|
| **DT** | 95 | 3 | 4 | 69 |
| **LR** | 87 | 5 | 12 | 67 |
| **RF** | 97 | 8 | 2 | 64 |
| **NB** | 82 | 7 | 17 | 65 |
| **KNN** | 96 | 3 | 3 | 69 |
| **NN** | 99 | 3 | 0 | 69 |
| **XGBoost** | 95 | 8 | 4 | 64 |
| **SVM** | 92 | 4 | 7 | 68 |

Table 2 VALUES OBTAINED FOR CONFUSION MATRIX BY USING DIFFERENT ALGORITHM

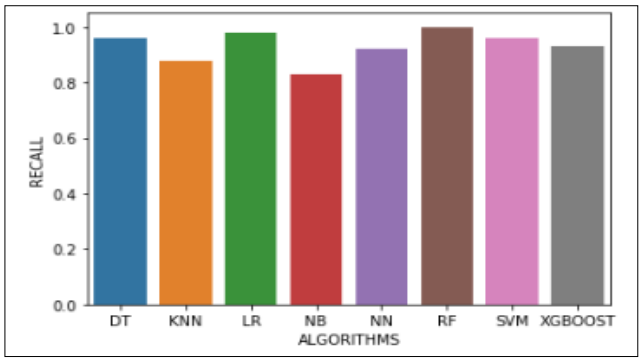



Fig 4 VALUES OBTAINED FOR RECALL BY USING DIFFERENT ALGORITHM

| Algorithms | Precision | Recall | F1-measure | Accuracy |
|---|---|---|---|---|
| **DT** | 0.97 | 0.96 | 0.96 | 95.91 |
| **LR** | 0.92 | 0.98 | 0.95 | 94.15 |
| **RF** | 0.97 | 1.00 | 0.99 | **98.25** |
| **NB** | 0.92 | 0.83 | 0.87 | 85.96 |
| **KNN** | 0.95 | 0.88 | 0.91 | 90.06 |
| **NN** | 0.96 | 0.92 | 0.94 | 96.49 |
| **XGBoost** | 0.96 | 0.93 | 0.94 | 93.57 |
| **SVM** | 0.92 | 0.96 | 0.94 | 92.98 |

Table 3 PERFORMANCE ANALYSIS OF MACHINE LEARNING ALGORITHM

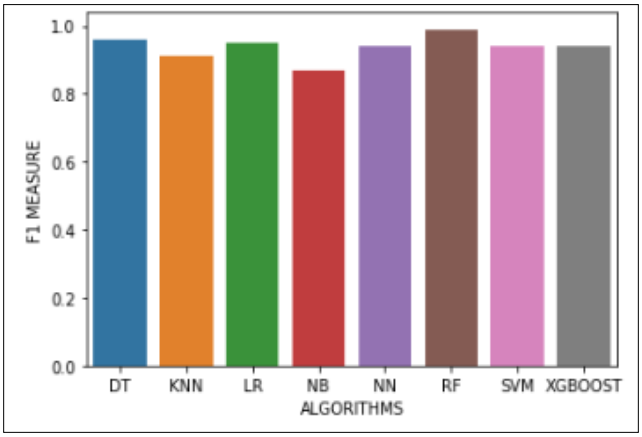



Fig 5 VALUES OBTAINED FOR F1-MEASURE BY USING DIFFERENT ALGORITHM

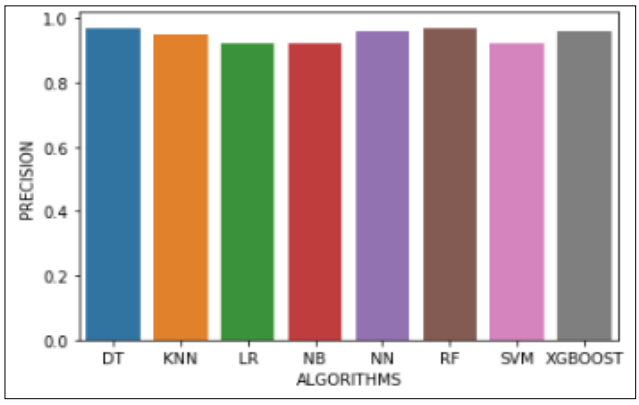



Fig 3 VALUES OBTAINED FOR PRECISION BY USING DIFFERENT ALGORITHM

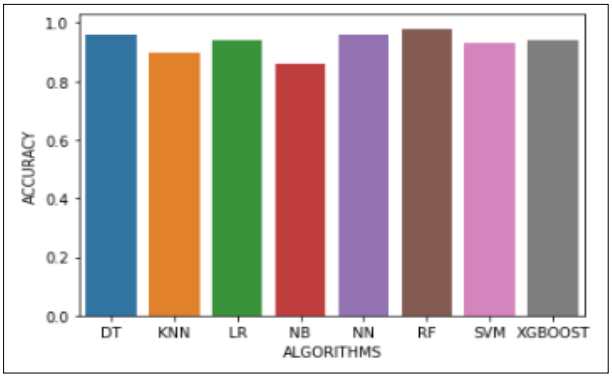



Fig 6 VALUES OBTAINED FOR ACCURACY BY USING DIFFERENT ALGORITHM

## CONCLUSION AND FUTURE WORK

With the rising number of deaths from cardiovascular disease, it has become necessary to create a system that can forecast cardiovascular disease effectively and precisely. The study's objective was to identify the most effective ML algorithm for detecting cardio vascular disorders. Using the UCI machine learning repository dataset, this study analyses the accuracy score of Decision Tree, Logistic Regression, Random Forest, SVM, KNN, Naive Bayes, Neural Network, and XGBoost algorithms for predicting cardiovascular disease. According to the results of the study, the Random Forest algorithm is the most effective algorithm for predicting cardiac disease, with an accuracy score of 98.25%.

According to the results of the study, the Random Forest algorithm is the most effective algorithm for predicting cardiac disease, with an accuracy score of 98.25%. In the future, the work may be improved by creating a web application based on the Random Forest algorithm and employing a larger dataset than that used in this research, which would help to deliver better results and assist health professionals in effectively and efficiently to forecasting cardiovascular problems.

## REFERENCES


**[1]** Aishwarya P, Jeny J.R.V, Reddy N.S, and Samreen (2021), "A Classification Approach for Heart Disease Diagnosis using Machine Learning", IEEE 6th International Conference on Signal Processing, Computing and Control (ISPCC).

**[2]** Alfian G, Fitriyani N.L, Rhee J and Syafrudin M (2020), "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System" in IEEE Access, vol. 8, pp. 133034-133050.

**[3]** Amar M, Louridi N and Ouahidi B E (2019), "Identification of Cardiovascular Diseases Using Machine Learning", IEEE 7th Mediterranean Congress of Telecommunications (CMT).

**[4]** Bertsimas D, Mingardi L and Stellato B (2021), "Machine Learning for Real-Time Heart Disease Prediction", IEEE Journal of Biomedical and Health Informatics.

**[5]** Boukhatem C, Youssef H.Y and Nassif A.B (2021), "Heart Disease Prediction Using Machine Learning", IEEE Advances in Science and Engineering Technology International Conferences (ASET).

**[6]** Chakarverti M, Rajan R and Yadav S (2019), "Classification Technique for Heart Disease Prediction in Data Mining", IEEE 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)

**[7]** Chaubey A.K, Gogoi A, Mahesh T.R, Rohit S and Sarveshvar M R (2021), "Performance of different Machine Learning Techniques for the Prediction of Heart Diseases", IEEE International Conference on Forensics, Analytics, Big Data, Security (FABS).

**[8]** G. S and S. K. J. (2019), "Prediction of Heart Disease Using Machine Learning Algorithms", IEEE 1st International Conference on Innovations in Information and Communication Technology (ICIICT)

**[9]** Hameetha Begum S and Nisha Rani S.N (2021), "Model Evaluation of Various Supervised Machine Learning Algorithm for Heart Disease Prediction", IEEE International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM).

**[10]** Kanchana S and S Usha (2022), "Effective Analysis of Heart Disease Prediction using Machine Learning Techniques", IEEE International Conference on Electronics and Renewable Systems (ICEARS).

**[11]** KN L, Kumari N.K.R, N. R, S. N and V. K (2021), "Heart Disease Detection using Machine Learning Technique", IEEE Second International Conference on Electronics and Sustainable Communication Systems (ICESC).

**[12]** Kumar Thakkar H and Shukla H (2020), "A Comparative Analysis of Machine Learning Classifiers for Robust Heart Disease Prediction", IEEE 17th India Council International Conference (INDICON).

**[13]** Mahalakshmi K and Sujatha P (2020), "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease", IEEE International Conference for Innovation in Technology (INOCON).

**[14]** Mohan S, Srivastava G and Thirumalai C (2020), "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" in IEEE Access, vol.

**[15]** Ouyang S (2022), "Research of Heart Disease Prediction Based on Machine Learning", IEEE 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE).