Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 6
June 2023

A Hybrid Model for Predicting Cardiovascular Disease Based on Conventional Machine Learning Classification Algorithms

[1]**Putta Durga**
School of Computer Science and Engineering,
VIT-AP University, Amaravati,
Andhra Pradesh, India
Mail id: pdurga593@gmail.com
[2]**Sudhakar T.**
School of Computer Science and Engineering,
VIT-AP University, Amaravati,
Andhra Pradesh, India
Mail id: tsudhakar105@gmail.com

Abstract

In recent decades, heart disease, also known as cardiovascular disease, has been the leading cause of mortality Heart disease, sometimes called cardiovascular disease, has emerged as the major cause of death during the past few decades. Many different heart conditions fall under its umbrella. While a correct diagnosis of heart disease might lessen the likelihood of serious health issues, an incorrect one can be fatal. It incorporates a variety of cardiovascular disease risk factors with the need for time to acquire accurate, reliable, and reasonable techniques for early identification and speedy illness management. The World Health Organization (WHO) has concluded that cardiovascular disease is the biggest cause of death around the globe. An estimated 17.9 million people worldwide perish each year. Machine learning is widely used in healthcare as a data processing approach. For the purpose of better heart disease prediction, researchers investigate complex medical data using a variety of ML methods. In this paper, we present a number of variables associated with heart disease and a model built using supervised learning techniques such as Naive Bayes (NB), Decision Trees (DT), K-nearest Neighbor (K-NN), Logistic Regression (LR), the Random Forest (RF) algorithm, and Support Vector Machines (SVM). The Cleveland Cardiac Registry Dataset from the University of California, Irvine Heart Disease Patient Repository is used. There are 303 occurrences and 76 characteristics in the set. Only 14 of the 76 attributes listed above are really taken into account while testing algorithms. The purpose of this paper is to determine a patient's risk of developing heart disease. From what we can see, the Decision Tree scores highest in terms of accuracy.

Key Words: *Supervised Machine Learning, Heart Disease, Classification Algorithms, Artificial Neural Networks (ANN)*

I. Introduction

In today's world, heart disease is a major health problem that can be fatal for many people (HD). As a result of the heart's inability to pump sufficient blood to other organs to keep them functioning normally, heart failure is an inevitable consequence of this condition [1]. Because of this, the heart can't pump enough blood to the rest of the body to keep it functioning normally. [2] In the United States, heart disease is more common than in other developed countries. The most common symptoms of heart disease are shortness of breath, weakness, swollen feet, and fatigue [3]. Other symptoms include high jugular venous pressure and peripheral edema, all of which can be caused by functional cardiac or noncardiac issues. The complexity of the diagnostic procedures used in the early stages of heart disease detection and diagnosis is a significant contributor to the expense of living [4]. Because of a lack of diagnostic tools and trained medical professionals, as well as other factors, heart disease is notoriously difficult to diagnose and treat in developing countries [5]. Reduce the likelihood of serious heart problems and increase heart safety by getting an accurate and thorough assessment of a patient's risk for cardiovascular disease [6].

Today's greatest healthcare challenge is the provision of excellent services and quick, accurate diagnoses [7]. The leading cause of death globally in recent years has been identified as heart disease, but it is also the most effectively controlled and monitored disease. The right time of disease detection determines how accurately a disease will be managed overall. The proposed work attempts to identify these heart conditions early to prevent negative outcomes. There are records of sizable collections of medical data compiled by medical specialists that can be examined and used to extract important knowledge. Heart disease can be fatal and shouldn't be ignored. According to Harvard Health Publishing [8], males are more likely than females to develop heart disease. Men were discovered to be rough twice as likely as women to experience a heart attack throughout their lifetime. Even after they took into account the standard risk factors for heart disease, such as high cholesterol, high blood pressure, diabetes, body mass index, and physical activity, there was still a higher risk.

It is feasible to glean essential knowledge from a sizable quantity of data by employing methods that are associated with data mining. The majority of the information in the medical database is in the form of discretized data. The process of making decisions based on discrete data subsequently becomes complicated. Machine learning, or ML for short, is a subset of data mining that makes efficient use of large datasets that have been painstakingly arranged. Machine learning has several applications in the realm of medicine, including the detection, localization, and prognosis of various diseases. The primary objective of this study is to offer cardiologists a diagnostic tool for the early diagnosis of coronary artery disease [9]. As a direct result of this, patients will receive high-quality care, and serious repercussions will be avoided. The analysis of the presented data and uncovering previously concealed discrete patterns are significant roles that machine learning (ML) performs. The early detection and prognosis of cardiac disease can be improved using machine learning approaches, which are applied after data analysis. This study evaluates the efficacy of several machine learning techniques, including K-NN, NB, DT, LR, RF, and SVM, to diagnose heart disease in its earliest stages [10].

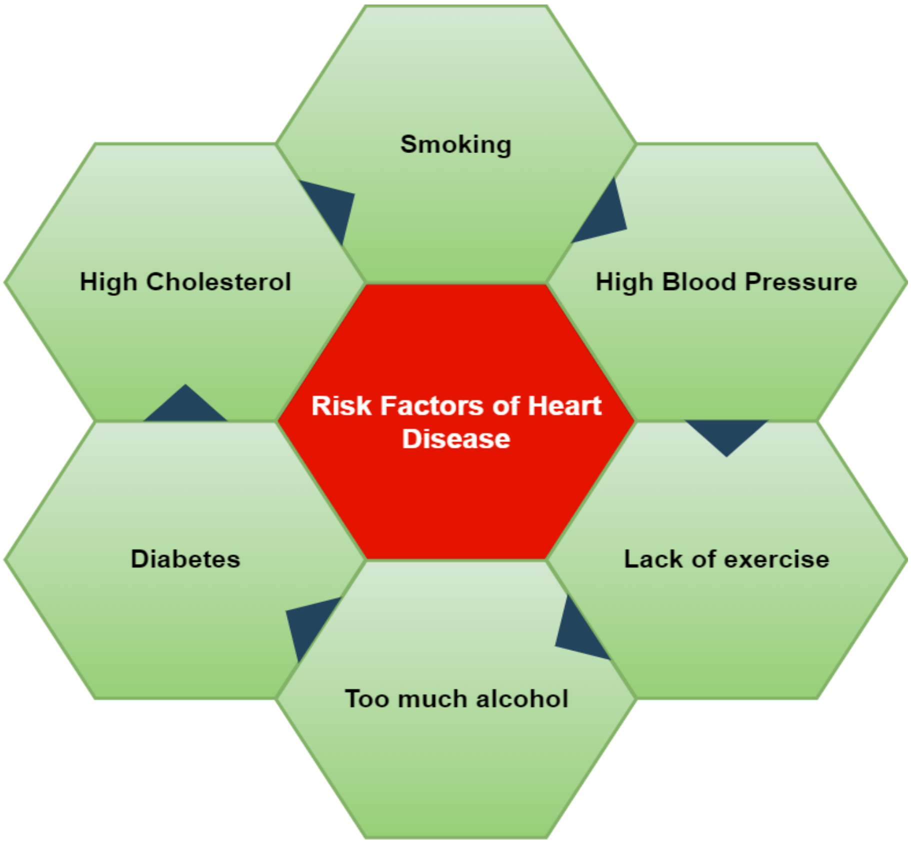


Fig 1: Heart Disease Risk Factors

II. Related Work

Researchers have employed many data mining methods, such as association rules, classification, and clustering, to develop a model for the prediction of cardiovascular disease. Prediction models based on prescriptions were proposed by Shiva Kazempour Dehkordi and Hedieh Sajedi [11] using data mining. To enhance the accuracy of the system, they proposed using the Skating algorithm. Skating, an ensemble technique, is very much like boosting and bagging. They compared the efficacy of DT, NB, K-NN, and staking in four distinct categorization methods. Staking was shown to be the most accurate classifier available. This classifier was 73.17 percent accurate. However, the method's performance is subpar when compared to alternative categorization algorithms and techniques. For instance, in 2018, Jan et al. [12] established an ensemble data mining strategy employing two benchmark datasets from a UCI repository (specifically Cleveland and Hungarian). Five distinct classification methods (radial basis function, neural network, NB, regression analysis, and support vector machine) were employed in an ensemble approach. A very high level of accuracy (98.136%) was provided by RF, while regression techniques were shown to be the least effective algorithm in that investigation.

The authors of the study that was published in 2018 by Le et al. [13] classified the 58 features that were included in the dataset that was taken from the UCI ML Repository using three distinct classification approaches. They demonstrated that a support vector machine (SVM) with a linear kernel performed better by achieving an accuracy of 89.93 percentage points. Tarawneh and Embarak [14] suggested a hybrid method that makes use of 12 features. The performance of this method was compared to that of KNN, J48, GA, DT, artificial neural network (ANN), SVM, and NB. The accuracy achieved using the proposed hybrid technique was 89.2%, which was higher than that achieved using existing algorithms.

Neural networks, decision trees, Nave Bayes, and associative classification have all been shown to be effective in the analysis of data mining techniques for the prediction of heart disease. Comparing associative classification to traditional classifiers, even when handling unstructured data, the results show high accuracy and strong flexibility [15,16].

Prediction of kidney illness was achieved by utilizing SVM and Naive Bayes algorithms [17]. The authors attempted to classify the various phases of renal illness by making use of the ANFIS algorithm that was provided. The goal of the study was to come up with an efficient algorithm for classifying things, and it did so by utilizing a wide variety of metrics for judging its performance, such as accuracy and execution speed. The Naive Bayes method outperformed the SVM algorithm despite the SVM algorithm's greater classification accuracy because the Naive Bayes algorithm provided outcomes more quickly. The results of this study show that the Support Vector Machine (SVM) is superior to the Naive Bayes Approach when it comes to forecasting renal disease.

Olaniyi et al. [18] developed a prediction model and used feedforward multilayer perceptron (MLP) to reach an accuracy of 85% on the UCI ML datasets. They achieved an accuracy of 87.5% using a support vector machine (SVM).

According to Karthick K et al. [19], the Cleveland heart disease (HD) dataset is organized by selecting individual features based on the results of a chi-square statistical test to sort the data. SVM, Gaussian Naive Bayes (GNB), LR, Light gradient-boosting machine (LightGBM), XGBoost, and RF techniques were used to develop models for predicting the risk of cardiovascular disease. These models achieved accuracy levels of 80%, 80.68%, 80%, 77.04%, 73.77%, and 88.5%, respectively. The purpose of developing this data visualization was to illustrate how the characteristics are intertwined with one another. According to the findings of the research, the RF method is able to attain an accuracy of 88.5% throughout the validation process for 303 data instances by making use of 13 features chosen from the Cleveland HD dataset.

Deep learning as well as a variety of other machine learning techniques were utilized by Rohit Bharti et al. [20] to compare the results and conduct an analysis on the UCI ML Heart Disease dataset. The dataset contains 14 main attributes, all of which will be taken into consideration during the research. To achieve and validate a number of promising results, an accuracy and confusion matrix is utilized. The data have been standardized to get better results, and the dataset contains some features that are not important, which have been isolated using Isolation Forest. The possibility of combining the findings of this study with other types of multimedia technology, such as mobile devices, is also considered. An accuracy of 94.2% was achieved with the application of deep learning technology.

The data mining classification technique was used by Mihir et al. [21] to evaluate the parameters. The datasets are handled in Python using a machine-learning technique called Logistic Regression, which exhibits the highest level of accuracy for heart disease.

In [22], O. W. Samuel describes the use of an ML algorithm, an ANN, and Fuzzy analytical hierarchical processing in his system for supporting medical decision-making on the diagnosis of heart disease. Their proposed approach had a classification accuracy of 91.10 percent.

The fundamental objective of the suggested system was to create a heart disease prediction system based on the inputs shown in Table 1, which was done by reciting the aforementioned publications. By comparing DT, RF, LR, and NB, SVM, and K-NN we were able to identify the best algorithm for predicting cardiovascular illness.

III. Methodology

The suggested work predicts cardiac issues by investigating the aforementioned categorization systems and analyzing their performance. The purpose of this research is to determine how well we can predict whether or not a patient has heart disease. The information from the health report is entered by the healthcare provider. The data is entered into a model that estimates the likelihood of developing heart disease. Fig. 1 depicts the full process for building a model to forecast heart disease. These steps need to be taken:

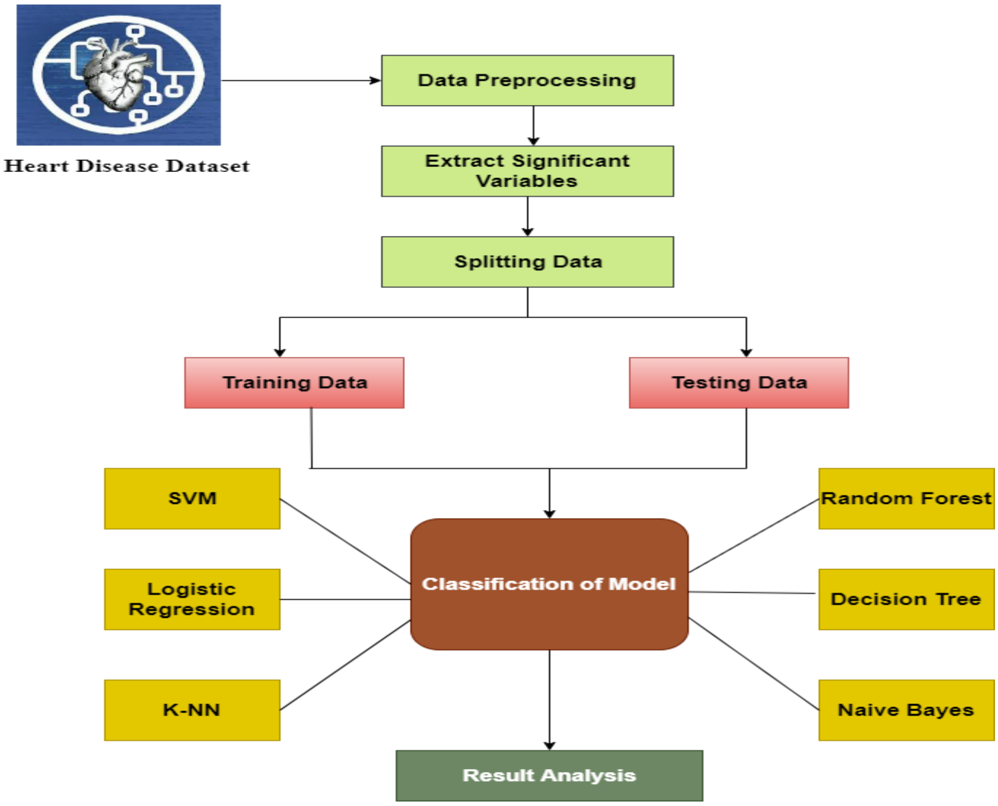


Fig 2: Overview of the Proposed Framework

i. DATA Collection:

The studies made use of a dataset referred known as the Cleveland heart dataset that was located in the ml repository at UCI. The collection is comprised of 14 attributes and 303 instances in total. There are eight different types of categorical attributes and six different types of numerical attributes. Table 1 provides an overview of the dataset that was provided. The ages of these patients, who are included in this dataset, range from 29 to 79. Patients who are male and patients who are female can be differentiated from one another by utilizing the gender values 1 and 0, respectively. There are four distinct types of chest discomfort that can be used to diagnose coronary heart disease. When the blood flow to the heart muscles is limited due to restricted coronary arteries, a condition known as type 1 angina can occur. Chest pain associated with type 1 angina might be a sign of mental or emotional stress. Pain in the chest that is not caused by angina can have a wide variety of origins and is not always related to a real heart condition. The fourth category, known as asymptomatic, might not point to a problem with the heart at all. The reading of blood pressure while the patient is at rest is represented by the character "trestbps." According to what it states, there is chol present in the blood. Fbs is an abbreviation for "fasting blood sugar level," and a value of one is assigned to it if the reading is lower than 120 mg/dl; otherwise, a value of zero is assigned. The abbreviations "restecg" (resting electrocardiographic result), "thalach" (maximum heart rate), "exang" (exercise-induced angina), "old peak" (exercise-induced ST depression), and "slope" (exercise-induced ST segment peak slope) are used in this article. "ca" stands for the number of major vessels stained by fluoroscopy, "thal" stands for the duration of the exercise test in minutes, and "num" Patients diagnosed with cardiovascular disease have a value of 1 for the class attribute, whereas healthy individuals have a value of 0 for this property.

Table 1: Detailed explanation of the dataset's attributes

Name	Type	Description
Age	Continuous	Age in years
Sex	Discrete	1 = male 0 = female
Cp	Discrete	Chest pain type: 1 = typical angina 2 = atypical angina 3 = non-anginal pain 4 =asymptomatic
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Continuous	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar > 120 mg/dl: 1 = true 0 = false
Restecg	Discrete	Resting electrocardiographic results: 0 = normal 1 = having ST-T wave abnormality 2 =showing probable or define left ventricular hypertrophy by Estes'criteria
Thalach	Continuous	Maximum heart rate achieved
Exang	Discrete	Exercise induced angina: 1 = yes 0 = no
Old peak ST	Continuous	Depression induced by exercise relative to rest
Slope	Discrete	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3= down sloping
Ca	Discrete	Number of major vessels colored by fluoroscopy that ranged between 0 and 3.
Thal	Discrete	3 = normal 6= fixed defect 7= reversible defect
Diagnosis	Discrete	Diagnosis classes: 0 = healthy

ii. **Data Pre-processing**

The preparation of data through pre-processing is necessary for the successful training and testing of a machine learning classifier as well as the accurate representation of data. In order for the classifiers to make efficient use of the dataset, pre-processing methods such as the removal of missing values, the application of standard scalar, and the use of MinMax Scalar have been utilized. Because of the 0e standard scalar, which guarantees that each feature will have the same coefficient, the mean value of each attribute will be 0 and the variance will be 1. Similar transformations are applied to the data in MinMax Scalar in order to move every feature to a value between 0 and 1. The missing value of 0e in the feature row that was previously included in the dataset has just been erased. All of these pre-processing techniques were utilized for the data in this research.

iii. **Feature Selection**

Because irrelevant features might affect machine learning classifier performance, feature selection is crucial. Feature selection improves model execution speed and classification accuracy. Our system selected significant features using three well-known FS techniques.

iv. **Classification Algorithms**

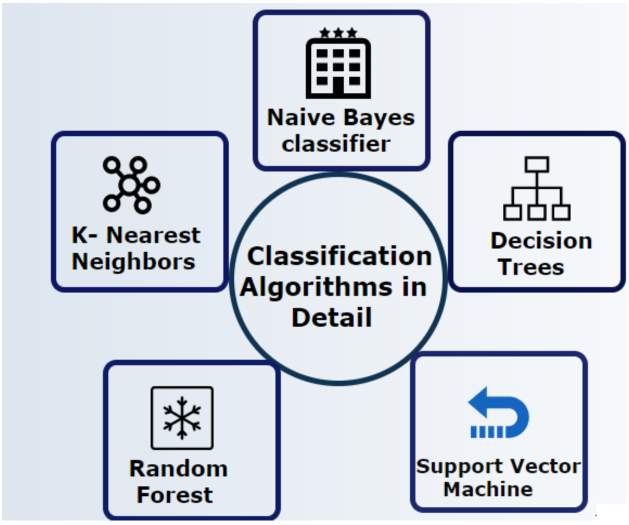


Fig 3: Machine Learning Classification Models

RF, DT, LR, and NB classification methods [12] all take the features mentioned in Table 1 as input. It is common to practice using 80% of the input dataset for training and 20% for testing. A model's training dataset is the information source from which the model is educated. The testing dataset is used to test how well the trained model performs. Calculating and evaluating the efficacy of each algorithm requires the use of a number of different measures, including accuracy, precision, recall, and F-measure scores (all of which are addressed in greater detail below). The study investigated the following sets of algorithms.

1. **Random Forest (RF)**

To solve classification and regression issues, many researchers turn to RF, a popular supervised ML technique. Decision trees are constructed utilizing data from multiple samples, with the majority of samples used for classification and an average used for regression.

2. **Decision Tree (DT)**

Decision tree methods are categorized as supervised learning algorithms. You can use them to fix your classification and regression problems. The most difficult part of using a Decision Tree is determining which attribute should be used for the root node at each level. Attribute selection is the process through which the desired attribute is chosen.

If you're having trouble making sense of the whole process, consider the following algorithm.

Step 1: as instructed by S, the complete dataset is stored in the root node of the tree.

Step 2: you'll want to use the Attribute Selection Measure to zero down on the most important aspect of the dataset (ASM).

Step 3: we'll divide the S into subsets that each contains values that may represent the best characteristics.

Step 4: Build the best attribute node in the decision tree.

Step 5: develop new decision trees by repeatedly using subsets of the dataset produced in step 3. Keep going until you reach the end of the path and the nodes can no longer be sorted into categories; at this point, you have reached a leaf node.

3. **Naïve Bayes**

The Bayes theorem is the basis for the supervised learning algorithm called "naive Bayes," which is used to solve classification problems. It is mostly used to classify text when the training set is big. As a probabilistic classifier, it predicts what will happen based on how likely it is that something will happen. The equation gives the formula for Bayes' theorem (1).

$$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is the posterior likelihood of A given B. Statistical support for hypothesis A given occurrence B

The probability, denoted by P(B|A), is the likelihood that a hypothesis is valid, and thus the probability that it is supported by the data.

4. **Support Vector Machine (SVM)**

SVMs are often utilized to address issues of classification and regression as one of the most popular supervised learning algorithms. However, it is typically used for Machine Learning Classification problems.

In order to efficiently categorize new data points, the SVM algorithm looks for the optimum decision boundary or line that may partition n-dimensional space into classes. This boundary, known as a hyperplane, is the optimal one for deciding what to choose.

5. **Logistic Regression (LR)**

The LR model is a great example of supervised learning. The likelihood of a binary (yes/no) event occurring is calculated. Logistic regression can be used to forecast the value of a categorical dependent variable. Therefore, it is essential that the final value be either discrete or categorical. Instead of giving us hard and fast numbers between 0 and 1, we get probabilistic estimates. Results can be either True or False, 0 or 1, or Yes or No.

6. **K-NN**

The KNN or k-NN algorithm is a supervised learning classifier that uses neighboring data points to predict or label the likely clustering of a single data point. It is applicable to both classification and regression issues, although the idea that comparable points tend to cluster together makes it particularly useful for the former.

IV. Evaluation Metrics

In the following section, the results of employing the RF, DT, NB, LR, SVM, and K-NN models are presented. In order to evaluate the efficacy of the algorithm, we make use of the following performance metrics: accuracy score, precision (P), recall (R), and F-measure. The accuracy metric, which is indicated in equation (2), provides the correct measure of positive analysis. Remember that the quantity o (1) correct positives is specified by the factor indicated in equation (3). The F-measure, which is discussed in equation (4), is used to evaluate how accurate the results are.

Precision (P) = (TP) / (TP +FP)
(2)

Recall (R) = (TP) / (TP+FN)
(3)

F– Measure = (2 * Precision * Recall) / (Precision +Recall) (4)

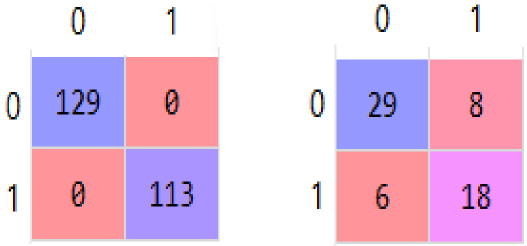
TP Positive diagnostic test results indicate the presence of the disease or ailment being tested for.

FP When a patient who does not actually have the sickness has a positive test result, this is called a false positive.

TN If a person's test results come back negative, it means they do not have the condition.

FN False negatives occur when a patient has the illness but the test comes back negative.

Experiments are run with the dataset after it has been pre-processed, and the aforementioned algorithms are investigated and employed throughout the process. The confusion matrix serves as the basis for the performance measurements that have been discussed thus far. The confusion matrix provides a description of the model's overall performance. Figure 4 demonstrates, for a variety of different methods, the confusion matrix that was produced by the model that was suggested.



a) Training Set b) Test Set

Fig 4: Confusion Matrix for Decision Tree

V. Results & Discussion

Machine learning features are chosen using the correlation matrix. It symbolizes how various attributes are interdependent. It is always preferable to examine the relationships between the features to determine which are negatively correlated and which are positively correlated, as shown in fig. 5.

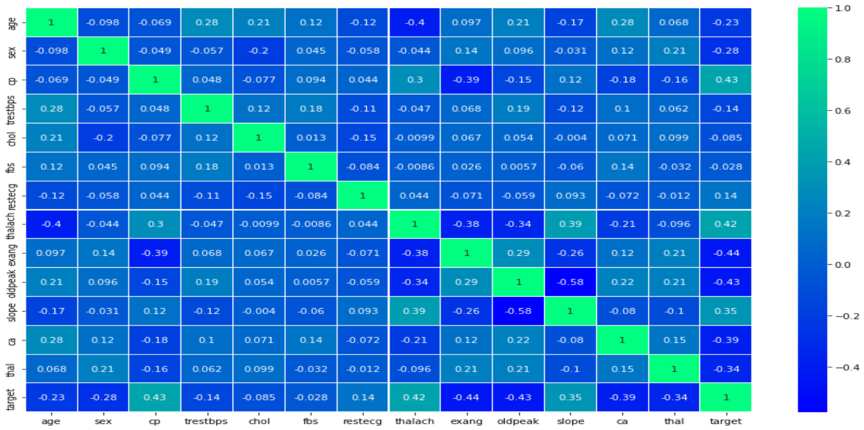


Fig 5: Correlation Matrix

We can infer from the heatmap above that there is a positive correlation between the target and chest pain (cp). It implies that those who experience frequent chest pain are more likely to develop heart disease. The target has a positive correlation with thalach, slope, and resting in addition to chest pain. As a result, there is a negative correlation between the target and exercise-induced angina (exang), which means that while exercising, the heart needs more blood, and the blood flow is slowed down by narrowed arteries. Old peak and thal have a negative correlation with the target, in addition, to ca.

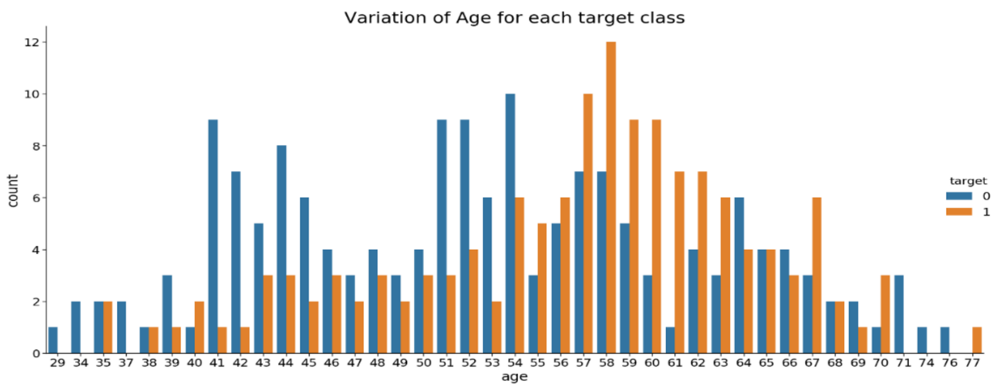


Fig 6: Data Analysis

Figure 6 above displays the ages of those who have the disease and those who do not. Here, target = 1 suggests that the person has heart disease, while target = 0 suggests that the person is not afflicted. We can see that 58 and 57 are the most common ages for those who are suffering. The disease primarily affects people who are 50 years of age or older. The Age feature should be divided into three categories: "Young," "Middle," and "Elder." Fig 7 demonstrates that heart disease affects young people less and older people more.

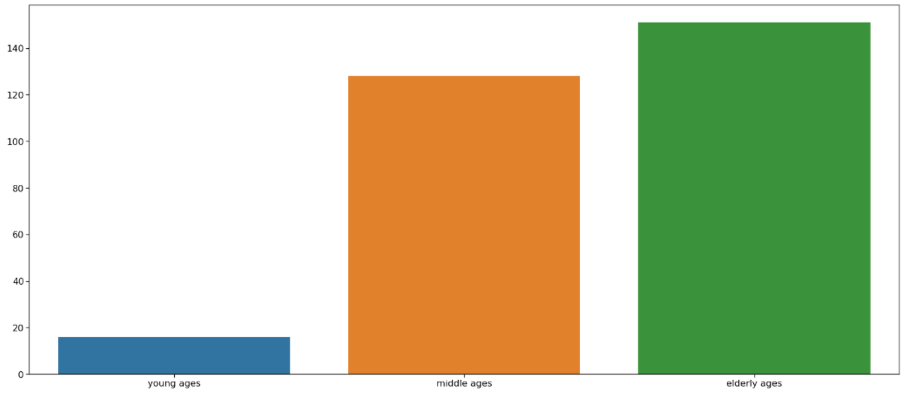


Fig. 7: Heart disease prevalence by age

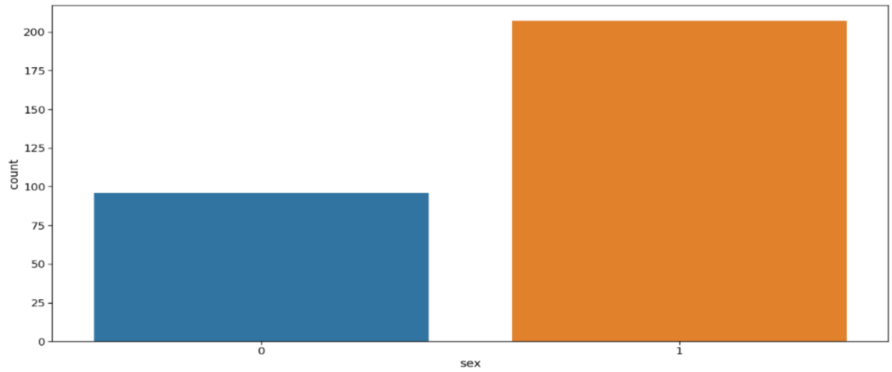


Fig. 8: Heart disease prevalence by gender

Fig. 8 demonstrates that the male-to-female ratio is roughly 2:1.

Table 2: Accuracies of Different Classification Models

S. No	ML Classification Models	Accuracy
1	LR	91
2	SVM	85.5
3	RF	84
4	K-NN	91
5	NB	86.8
6	**DT**	**92.9**

The accuracy of various ML classification models is displayed in Table 2. SVM, K-NN, Logistic Regression, Random Forest, and Naive Baye are just a few of the ML algorithms that we have contrasted. According to the comparison, the Decision Tree is the better option because it provides a better classification rate of 92.9% and better heart disease prediction outcomes.

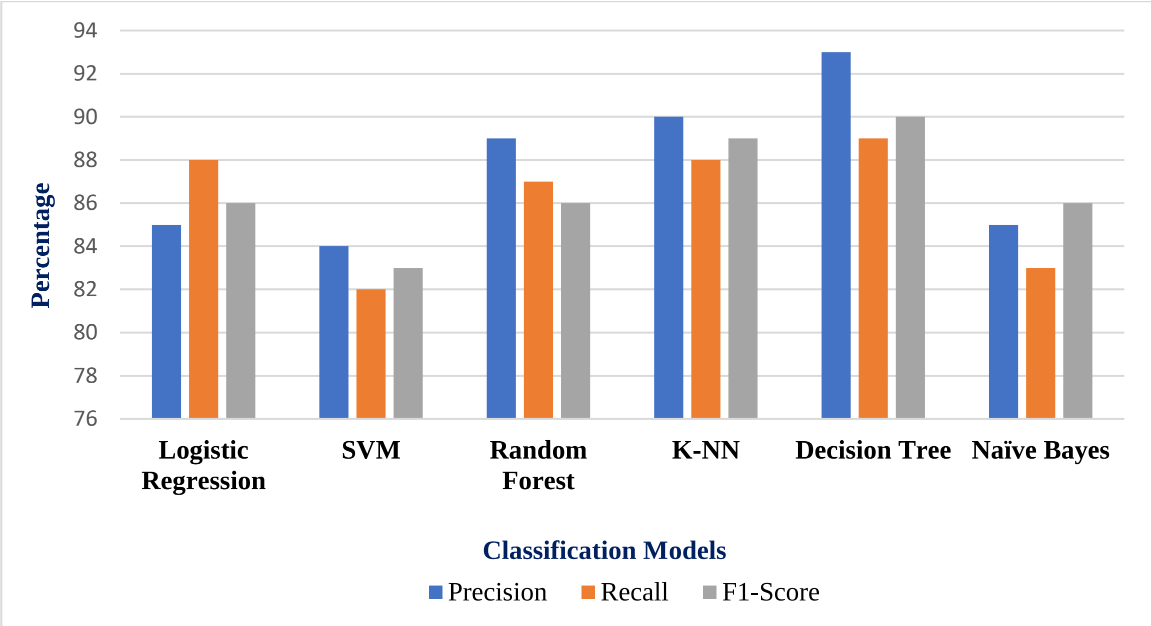


Fig 9: Analysis of ML Models

Fig. 9 displays the Precision, Recall, and F-measure performance metrics for each classification model described in this study. The decision tree produces good outcomes from those.

VI. Conclusion

Heart disease has become the current world's leading health crisis. It might be challenging to manually assess the probability of developing heart disease based on risk factors. However, with the present data, predictions can be made with machine learning methods.

Every year, more and more people learn they have heart disease. Hence, it is important to spread awareness about the need of catching it early and treating it. To that end, patients and the medical community could both benefit greatly from making use of the appropriate technological support. Six different ml algorithms (SVM, DT, RF, NB, LR, and K-NN) have been used to evaluate performance in this study. All of them were put to use in the data collection. The 76-feature dataset covers the known risk factors for cardiovascular disease, and 14 relevant features are selected from it for evaluation. The goal of attribute selection is to boost productivity. To determine which model is more accurate, n features will need to be selected. In order to reduce the number of variables, several features of the dataset are removed because their correlations are so close to one another. Six distinct machine learning methods are tested by building a unified prediction model, and the results are compared for efficacy. The goal is to use many measures of evaluation, such as the confusion matrix, accuracy, precision, recall, and f1-score, to make a reliable disease prediction. When compared to other models, the Decision Tree classifier's accuracy (92%) is the best.

References

[1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," Nature Reviews Cardiology, vol. 8, no. 1, pp. 30–41, 2011.

[2] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou et al., "Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association," Circulation, vol. 123, no. 8, pp. 933–944, 2011.

[3] M. Durairaj and N. Ramasamy, “A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate,” International Journal of Control@Theory and Applications, vol. 9, pp. 256–260, 2016.

[4] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, “Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data,” NeuroImage, vol. 28, no. 4, pp. 980–995, 2005.

[5] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, “Innovative artificial neural networks-based decision support system for heart disease diagnosis,” Journal of Intelligent Learning Systems and Applications, vol. 5, no. 3, pp. 176–183, 2013.

[6] Q. K. Al-Shayea, “Artificial neural networks in medical diagnosis,” International Journal of Computer Science Issues, vol. 8, no. 2, pp. 150–154, 2011.

[7] Avinash Golande, Pavan Kumar T, ”Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[8] Harvard Medical School, “*roughout life, heart attacks are twice as common in men than women,” 2020, https://www. health.harvard.edu/heart-health/throughout-life-heartattacks-are-twice-as-common-in-men-than-women.

[9] Theresa Princy R,J. Thomas,’Human heart Disease Prediction System using Data Mining Techniques’, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[10] Fahd Saleh Alotaibi,” Implementation of Machine Learning Model to Predict Heart Failure Disease”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[11] S.K. Dehkordi, H. Sajedi, Prediction of disease based on prescription using data mining methods, Health Technol. 9 (1) (2018) 37–44.

[12] M. Jan, A.A. Awan, M.S. Khalid, S. Nisar, Ensemble approach for developing a smart heart disease prediction system using classification algorithms, Res. Rep. Clin. Cardiol. 9 (2018) 33–45

[13] H.M. Le, T.D. Tran, L.A.N.G. Van Tran, Automatic heart disease prediction using feature selection and data mining technique, J. Comput. Sci. Cybern. 34 (1) (2018) 33–48.

[14] M. Tarawneh, O. Embarak, February. “Hybrid approach for heart disease prediction using data mining techniques, Acta Sci. Nutr. Health 3 (7) (2019) 147–151, 2019.

[15] Jyoti Soni Ujma Ansari, Sharma Dipesh. Predictive data mining for medical diagnosis: an overview of heart disease prediction‖. Int. J. Comput. Appl. March 2011;17(8). (0975 – 8887).

[16] Sudhakar K. Study of heart disease prediction using data mining. 2014;4(1):1157–60.

[17] H. Barakat, P. Andrew, Bradley, H. Mohammed Nabil Barakat, Intelligible support vector machines for diagnosis of diabetes mellitus, IEEE Trans. Inf. Technol. Bio Med. J. 14 (4) (2009) 1–7.

[18] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, “Heart diseases diagnosis using neural networks arbitration,” International Journal of Intelligent Systems and Applications (IJISA), vol. 7, no. 12, pp. 75–82, 2015.

[19] Karthick K, Aruna SK, Samikannu R, Kuppusamy R, Teekaraman Y, Thelkar AR. Implementation of a Heart Disease Risk Prediction Model Using Machine Learning. Comput Math Methods Med. 2022 May 2;2022:6517716. doi: 10.1155/2022/6517716. PMID: 35547562; PMCID: PMC9085310.

[20] Rohit Bharti, Aditya Khamparia, Mohammad S, Gaurav Dhiman, Sagar Pande, and Parneet Singh “ Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning” Computational Intelligence and Neuroscience Volume 2021,pp.1-11 https://doi.org/10.1155/2021/8387680.

[21] Mihir P, Rahul P, Chaitanya Patil, and Anuradha Kapoor “ Predicting Heart Disease Using Machine Learning Algorithms” International Research Journal of Engineering and Technology(IRJET), vol. 9, no.4, pp. 61-66, 2022.

[22] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, “An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction,” Expert Systems with Applications, vol. 68, pp. 163–172, 2017.