

## **HEART ATTACK DETECTION USING MACHINE LEARNING**

**Gudimella Venkata Narasimha Dwaraka Vyas, Likhita Pasupureddy, Padma Priya Bikkini, Ramagowri Kenguva, Mr.B.Venkatesh (Assistant Professor)** Department Of Electronics And Computer Engineering, Vignan's Institute of Information Technology, Visakhapatnam-India

### **Abstract**

Heart-related diseases are one of the main reasons for the death rate in any country and across the world over a few years. In India, the death rate has increased by 53% over the past five years and in the world, there is an estimation of 32% death rate out of which 85% belonged to heart-related problems and heart attacks. So, to measure the accuracy of such kinds of heart-related diseases machine learning algorithms have been applied to various parameters to predict the heart attack disease. The heart is one of the major organs in the body and it pumps blood around our body and it will remove waste products. This research paper compares various machine learning algorithms like the Naive Bayes algorithm, Gradient boosting algorithm, and K-nearest neighbor algorithm respectively. It used the dataset from Kaggle and considered the 304 records of the patient dataset with 13 parameters. After the generation of the result, it shows the highest accuracy for the Naive Bayes algorithm, and after that Gradient boosting algorithm and the least accuracy for the KNN algorithm.

**Keywords:** k-Nearest Neighbour, Naïve bayes, Gradient Boosting

### **Introduction**

The chronic disease of heart disease is one of the greatest threats to human health. Heart conditions are characterized by inadequate blood flow to the gut, which makes it difficult to perform normal bodily functions. Cardiovascular disease may be a widespread disease. The causes of this disease are high bp, smoking, overweight, diabetes, cholesterol, and hypertension that has been accumulated because of the unlimited intake of foods, no physical exercise, and fatty foods. This disease may happen to various heart issues like coronary-artery disease, stroke, coronary failure, cardiovascular, and more. Cholesterol, Chest pain, resting electrocardiographic results, resting blood pressure, exercise-induced angina, fasting blood sugar, the slope of the peak exercise ST segment, maximum heart achieved, ST depression induced by exercise relative to rest, number of major vessels colored by fluoroscopy is the major grounds for causing heart problems but we are having data of an individual person like weight, height, diastolic vital sign, systolic vital sign, alcohol, cholesterol, glucose, smoke. There are several amendable risk factors for cardiovascular disease, and many cases can be prevented.

### **Review of Literature**

Ravish D et.al [1] implemented the artificial neural networks for the factors like heart rate ,cholesterol,mean arterial pressure and Ecg of the patient and again it was converted to digital format to avoid the uncertainty in classification. Sasidharan et.al[2] developed a wearable cardio monitoring system by attaching the spo2 sensor and temperature sensor and if there are any fluctuations then the warning is sent before the attack occurs. M. Kavitha et.al[5] Proposed a model by using the machine algorithms named decision tree ,random forest and hybrid model however after using the three models got an accuracy of 88% through out the hybrid model and the author was used the cleveland dataset. A.Pandiaraj et.al[7] the author implemented a genetic algorithm and support vector machine to reduce heart attacks. P.Anuradha et.al[8] the author was suggested a boosting classifiers like catboost and XGboost in a SA heart and cleveland datasets and in this catboost algorithm has got the better results than the other classifiers. Pradeep Raj Savarapu et.al[11] the author proposed a advanced methodology in prediction by using the artificial neural networks and Genetic algorithm and got the best accuracy level when compared to other algorithms.

### **Research Methodology**

Data Visualization:

- Data visualization tools provide access to understand and see patterns in data by making use of

visual elements like graphs, maps, and charts and it very flexible for large and complex datasets.

- The purpose of this paper is to analyse massive amounts of information and make data-driven decisions using data visualization tools and technologies.
- In this paper to analyze the relation between the paramaters heat map plays a crucial role.

Data Cleaning:

- Data cleaning is nothing but detecting and removing or fixing the unwanted data from the dataset. Dataset needs to be consistent so data cleaning helps the remove the duplicate or unwanted data.
- Every dataset needs to maintain validity, accuracy, completeness, consistency, and uniformity. if the data is incorrect then the algorithms are unable to get a reliable output so in order to maintain consistent data, data cleaning is crucial

Data Collection:

- In this research we used a dataset from UCI called Hungarian dataset, which contains the data records after removing incomplete data.
- The dataset collected consists of 304 records of patients' data carries 13 features.

- a. age - in years
- b. sex - (1 = male; 0 = female)
- c. cp - chest pain type
- d. trestbps - resting blood pressure (in mm Hg on admission to the hospital)
- e. chol - serum cholestoral in mg/dl
- f. fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- g. restecg - resting electrocardiographic results
- h. thalach - maximum heart rate achieved
- i. exang - exercise induced angina (1 = yes; 0 = no)
- j. oldpeak - ST depression induced by exercise relative to rest
- k. slope - the slope of the peak exercise ST segment
- l. ca - number of major vessels (0-3) colored by flourosopy
- m. thal - 3 = normal; 6 = fixed defect; 7 = reversable defect
- n. target - 1 or 0

A. Naive bayes algorithm:

Naive bayes algorithm is method to forecast the Probability of independent attributes. It allows us to calculate conditional probabilities. It comes extremely handy because it enables us to use some knowledge that we already have to calculate the probability of a related event. It is not sensitive for uncertain data and it requires less training data.

Naïve Bayes formula :  $P(A|B) = P(B|A) \cdot P(A)/P(B)$

Where, P(A) is nothing but Probability of occurring A

P(B) is nothing but Probability of occurring B

P(B|A) is nothing but Probability of occurring B in given evidence A has already occurred

P(A|B) is nothing but Probability of occurring A in given evidence B has already occurred

B. Gradient Boosting algorithm:

In particular, with large and sophisticated datasets, gradient boosting is a method that stands out for its accuracy and speed of prediction. By using this algorithm, the bias error of the model can be minimized and it also helpful for finding the best estimator by using GridSearchCv.

C. KNN algorithm:

A supervised machine learning algorithm, KNN can be used for both classification and regression tasks and it is a slow learning algorithm, the training data is supplied but no training is performed. K-NN algorithms assume similarity between new cases and existing cases, and place the new cases in the category that most closely matches existing categories.

Accuracy is one of the most important performance evaluation metrics in machine learning .

Accuracy prediction formula:  $1 - (\text{Number of misclassified samples} / \text{Total number of samples})$

D. Machine Learning Algorithms used

The three machine learning algorithms used in this research work are Naïve bayes, Random Forest

Algorithm, K-Nearest Neighbour algorithm. The three techniques are used in predicting the Heart Attack with the help of parameters present in the dataset which is further cleaned from the null values.

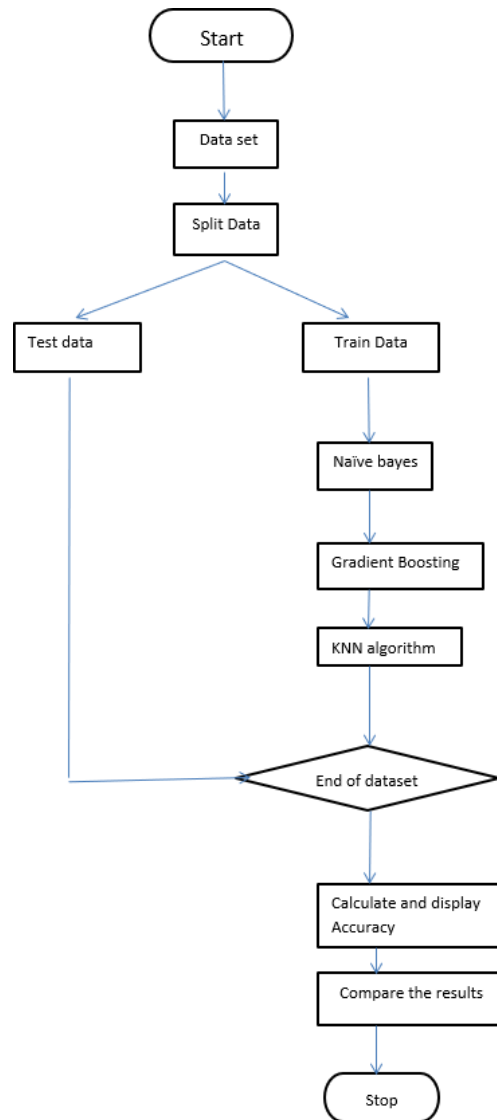


Fig.1 Flow diagram of proposed model

From Fig.1 the flow chart of the proposed model is observed. It is a step wise procedure of the proposed approach in this paper

### Results and Discussion

Based on the output of the system, a prediction can be made about whether the person is suffering from heart disease by using Naïve bayes classifier, the accuracy predicted result is approximately 94% which was highest and after that using Gradient boosting classifier, the accuracy predicted result is approximately 88% . KNN algorithm, the accuracy predicted result is approximately 71% .

By comparing the best case and worst case from the results obtained we are considering the Best-Case result.

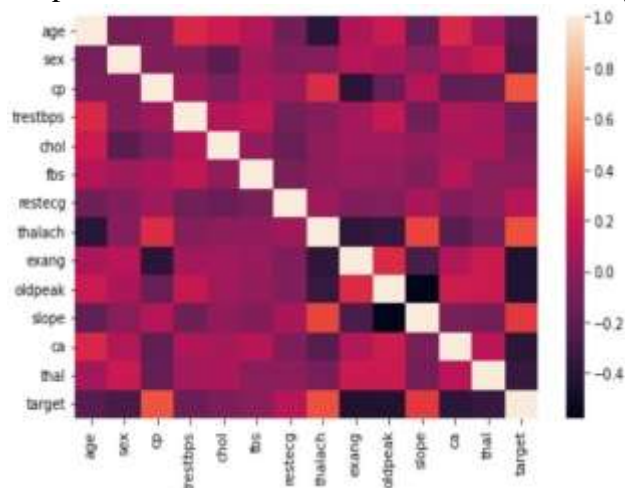
TABLE1. COMPARISON OF ACCURACY SCORE AND ROC-AUC

S.NO.	Algorithms	Accuracy Score	ROC AUC
1	Naive Bayes	94.28	93.37
2	Gradient Boosting	88.57	89.20
3	KNN Algorithm	71.42	64.39

Table 1 is the comparative analysis of the accuracy score and the ROC-AUC i.e probability curve of the three algorithms or methods considered in this research.

Correlation matrix:

Correlation matrices are simply the tables that show correlations among the parameters or factors. A linear relationship between variables is the best case for this type of analysis. If the color



is lighter then better the relationship between the factors.

Fig.2 Correlation Matrix

Scatter matrix:

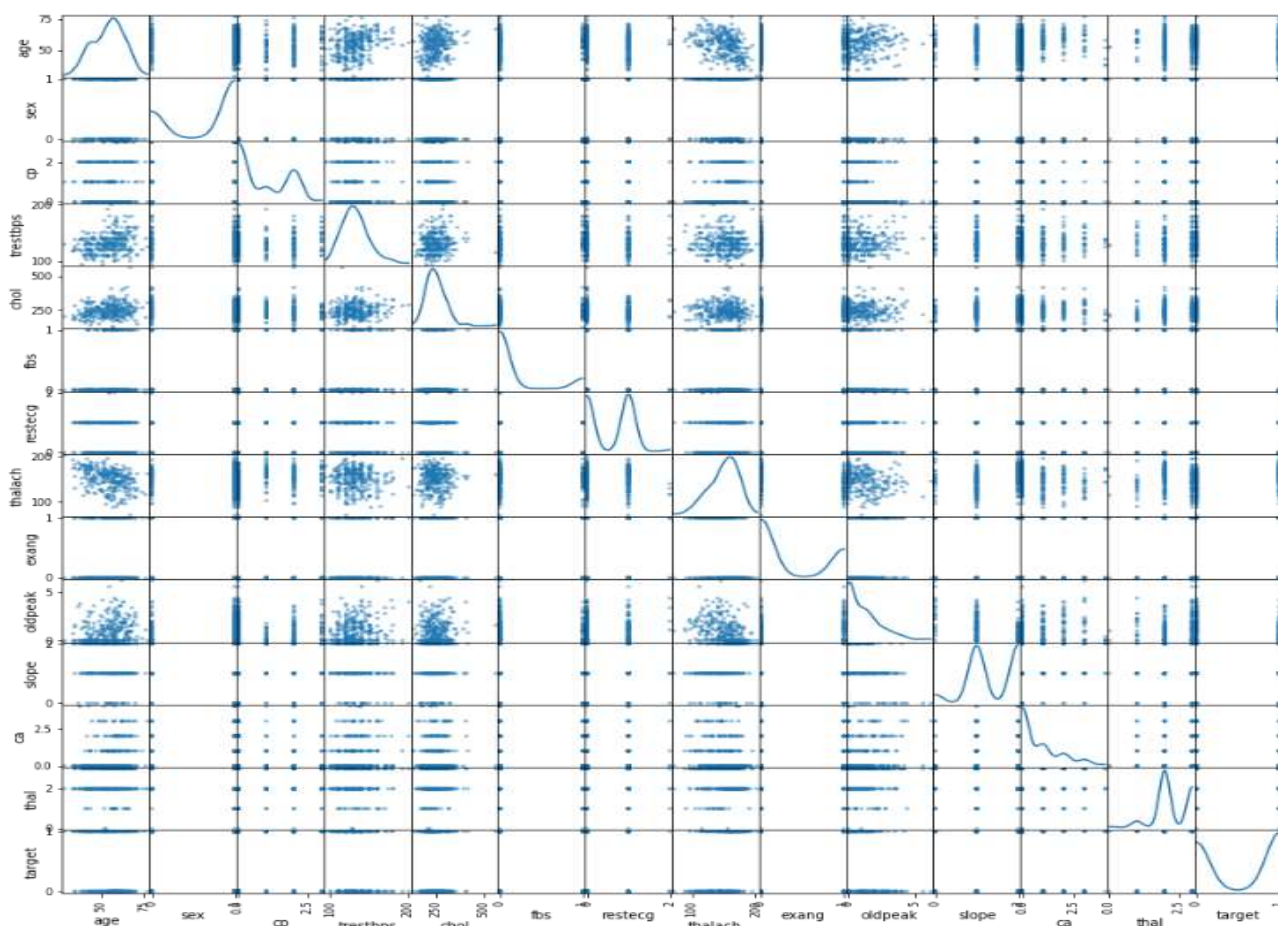


Fig.3 Scatter Matrix

Scatter plots displays how much one variable is affected by another or the connection between them by using dots in two dimensions. Scatter plots are same as line graphs in the concept as they use vertical axes and horizontal axes for plotting the data points.

Axes Subplot:

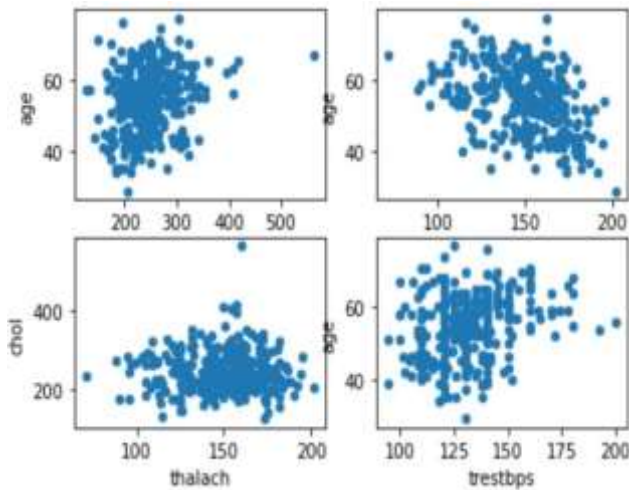


Fig.4 Axes SubplotHistogram Analysis:

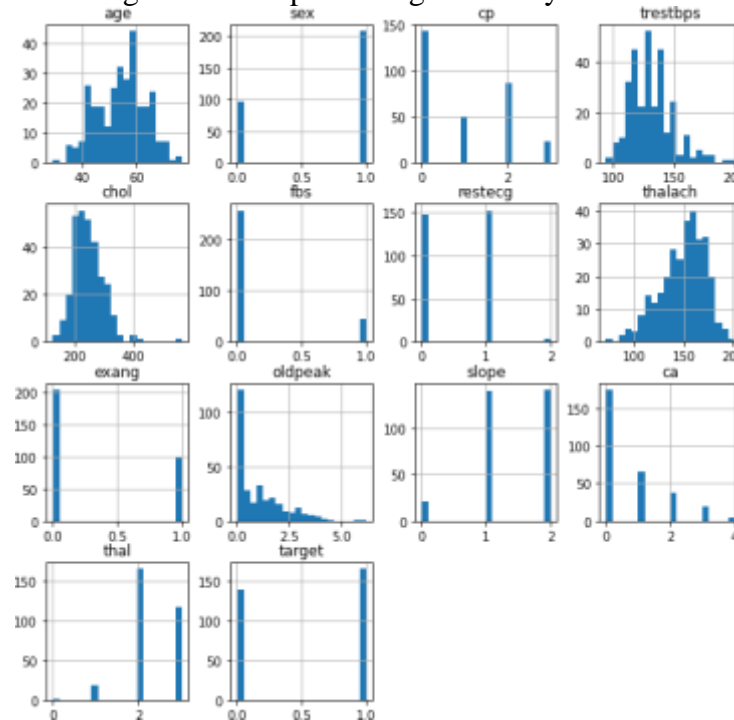


Fig.5 Histogram Analysis

A histogram is a visual representation of continuous or discrete data. The histogram can show a great deal of data as well as the frequency of its values and it can be used to determine the distribution of data and the median. Outliers and gaps in the data can also be highlighted.

### Conclusion

In this Research paper three types of Machine Learning algorithms namely Naïve bayes algorithm, Gradient Boosting algorithm and KNN algorithm are involved for heart attack Prediction. We have considered a dataset from Kaggle and it has 304 records of patient data with 13 parameters like age, cholesterol, sex, etc. Naive Bayes got an accuracy score equal to 94.2% which was far highest for any machine learning algorithm and a roc score of 93.3% after that by using Gradient Boosting Classifier we got the accuracy rate of 88% and roc score of 89% and the least observed accuracy rate was KNN algorithm it has got the accuracy rate of 71% and roc score of 64%. The results show that the Naive Bayes algorithm has got the highest accuracy and roc score.

### References

[1] Ravish, D. K., K. J. Shanthi, Nayana R. Shenoy, and S. Nisargh. "Heart function monitoring, prediction and prevention of heart attacks: Using artificial neural networks." In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 1-6. IEEE, 2014.

- [2] Sasidharan, Priyanka, T. Rajalakshmi, and U. Snehalatha. "Wearable cardiorespiratory monitoring device for heart attack prediction." In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0054-0057. IEEE, 2019.
- [3] Manikandan, Sushmita. "Heart attack prediction system." In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 817-820. IEEE, 2017.
- [4] Ehtiati, Nastaran, Gholamreza Attarodi, Nader Jafarnia Dabanloo, Javid Farhadi Sedehi, and Ali Motie Nasrabadi. "Prediction of ventricular tachycardia using nonlinear features of heart rate variability signal such as poincare plot, approximate and sample entropy, recurrence plot." In *2017 Computing in Cardiology (CinC)*, pp. 1-4. IEEE, 2017.
- [5] Kavitha, M., G. Gnaneswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj. "Heart disease prediction using hybrid machine learning model." In *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pp. 1329-1333. IEEE, 2021.
- [6] Raja, M. Snehith, M. Anurag, Ch Prachetan Reddy, and Nageswara Rao Sirisala. "Machine Learning Based Heart Disease Prediction System." In *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-5. IEEE, 2021.
- [7] Pandiaraj, A., S. Lakshmana Prakash, and P. Rajesh Kanna. "Effective Heart Disease Prediction Using Hybridmachine Learning." In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pp. 731-738. IEEE, 2021.
- [8] Sowmiya, C., and P. Sumitra. "Analytical study of heart disease diagnosis using classification techniques." In *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, pp. 1-5. IEEE, 2017.
- [9] Anuradha, P., and Vasantha Kalyani David. "Feature Selection and Prediction of Heart diseases using Gradient Boosting Algorithms." In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 711-717. IEEE, 2021.
- [10] Junaid, Mohammed Jawwad Ali, and Rajeev Kumar. "Data science and its application in heart disease prediction." In *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, pp. 396-400. IEEE, 2020.
- [11] Savarapu, Pradeep Raj, M. Shankar, Shilpa Itnal, Ram Kumar Madupu, CMAK Zeelan Basha, and E. Sreedevi. "Advanced prediction of Heart diseases using Artificial Neural Network and Genetic Algorithm." In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 875-878. IEEE, 2021.
- [12] Shaji, Shaicy P. "Predictionand Diagnosis of Heart Disease Patients using Data Mining Technique." In *2019 international conference on communication and signal processing (ICCSP)*, pp. 0848-0852. IEEE,2019.
- [13] Ed-Daoudy, Abderrahmane, and Khalil Maalmi. "Performance evaluation of machine learning based big data processing framework for prediction of heart disease." In *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*, pp. 1-5. IEEE, 2019