



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Heart Disease Prediction

**Dr. Rakesh Roshan<sup>1</sup>, E.Tejaswini<sup>2</sup>, D.Varun<sup>3</sup>, V.Srivatsa<sup>4</sup>, B.Sai Suhas Reddy<sup>5</sup>**

Assistant Professor, Anurag University, Hyderabad, India<sup>1</sup>

Student, Anurag University, Hyderabad, India<sup>2,3,4,5</sup>

**ABSTRACT:** According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analyzing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

Our Heart disease prediction System is intended to assist patients in recognizing their heart state early and receiving treatment at an earlier stage, allowing them to avoid any serious conditions by using Machine Learning Algorithms.

**KEYWORDS:** Cardiovascular disease, Mortality rate, Expertise, Machine learning, Prediction system

## I. INTRODUCTION

Heart disease is a leading cause of morbidity and mortality worldwide, making it a critical public health concern. Timely and accurate diagnosis of heart disease is crucial for effective treatment and prevention. Machine Learning (ML) offers a powerful tool to enhance the prediction and early detection of heart disease by analyzing complex patterns and relationships in patient data.

This introductory guide explores the application of ML in predicting heart disease, providing an overview of the key concepts and steps involved in developing a predictive model. By leveraging ML techniques, healthcare professionals can make more informed decisions, potentially saving lives and improving patient outcomes.

### **Importance of Heart Disease Prediction:**

Understanding why early detection and prediction of heart disease are vital for improving patient care and reducing healthcare costs. The importance of heart disease prediction lies in its capacity to save lives, reduce healthcare costs, and enhance the quality of life for individuals. Early prediction enables timely interventions and preventative measures, mitigating the risk of heart disease and its associated complications. By allocating healthcare resources more efficiently, it optimizes healthcare delivery, making it more accessible and cost-effective. Moreover, heart disease prediction informs public health strategies, allowing for targeted interventions and policy development to combat this leading cause of mortality.

### **What is Machine Learning ?**

Machine Learning (ML) is a subset of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computer systems to learn and make predictions or decisions based on data without explicit programming. In other words, ML allows machines to improve their performance on a specific task through experience

and exposure to data.

### Why Machine Learning ?

Machine learning is employed in heart disease prediction due to its ability to analyze complex patient data, extract valuable insights, and offer accurate risk assessments. This technology is crucial in handling the ever-growing volume of healthcare data, automating risk assessments, and continuously improving prediction accuracy. It aids healthcare providers in making informed decisions, optimizing resource allocation, and contributing to ongoing research in cardiology, ultimately leading to better patient care and outcomes.

## II. RELATED WORK

Heart disease is difficult to recognize due to a variety of risk factors such as high blood pressure, cholesterol, and abnormal pulse rate. Because of the disease's complexity, it must be handled with care. Otherwise, the effects of heart or death may occur. Accurate decisionmaking and optimal treatment are required to address cardiac risk in less Time. The problem at hand is to develop a machine learning model that accurately predicts the risk of heart disease in individuals based on a set of clinical and demographic features. The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications.

Prediction of this heart disease is regarded as one of the most important subject. So our primary objective in this project is to develop a machine learning-based predictive model that can accurately assess the risk of heart disease in individuals. Explore the heart disease datasets using exploratory data analysis (EDA) and to design the system using the Machine Learning model. To predict the future possibility of heart disease by implementing the Logistic Regression algorithm (Which is highly accurate). This model will utilize a diverse set of patient attributes and historical health data to make reliable predictions.

## III. METHODOLOGY

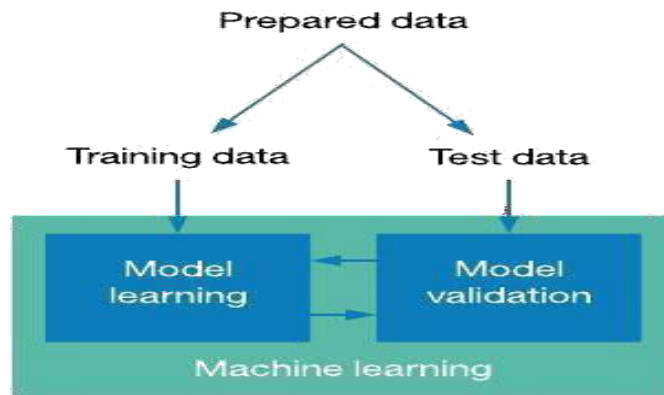
The implementation of the project has been carried out in a step-by-step manner. A detailed description of each module is given below and it is followed by an introduction to the technologies used in implementing the project.

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

- 1.) Collection of Dataset
- 2.) Selection of attributes
- 3.) Data Pre-Processing
- 4.) Balancing of Data
- 5.) Disease Prediction

### Collection of dataset :

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system.



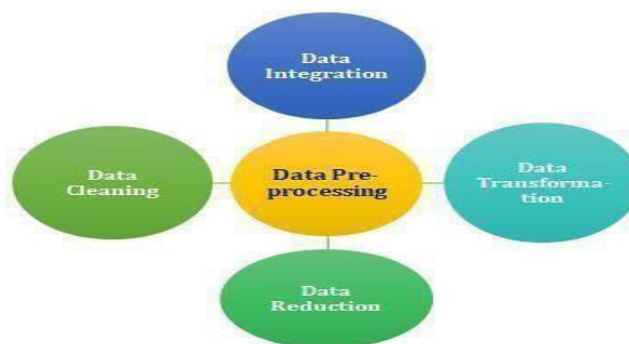
**Selection of attributes:**

Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, exang, etc are selected for the prediction. The Correlation matrix is used for attribute selection for this model.



**Pre-processing of Data :**

Data pre processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data preprocessing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Pre processing of data is required for improving the accuracy of the model.



### Balancing of Data:

Sampling

(a) Under Sampling:

In Under Sampling, dataset balance is done by the reduction of the size of the ample class.

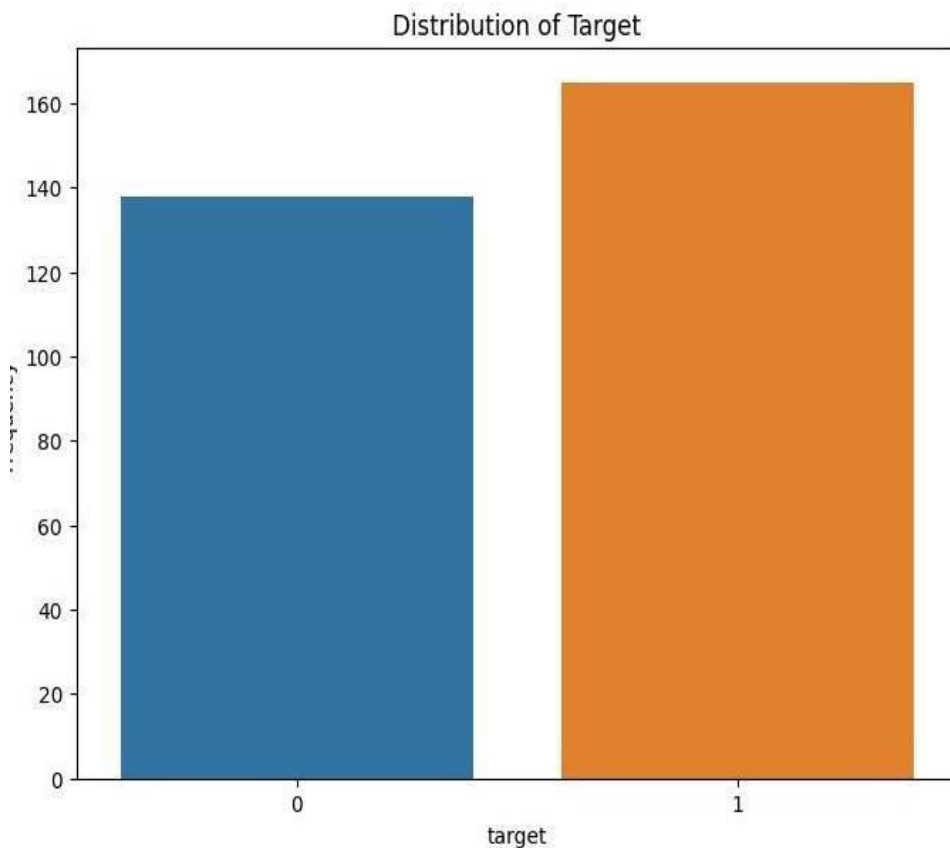
This process is considered when the amount of data is adequate. (b) Over Sampling: In Over Sampling, dataset balance is done by increasing the size of the scarce samples. This process is considered when the amount of data is inadequate.

### Prediction of Disease:

Machine learning algorithm like Logistic Regression, is used for classification. This algorithm that gives the highest accuracy is used for heart disease prediction.

## IV. EXPERIMENTAL RESULTS

After performing the machine learning approach for training and testing we find that accuracy of the Logistic Regression is better compared to other algorithms. Accuracy is calculated with the support of the confusion matrix of the algorithm, here the number count of TP, TN, FP, FN is given and using the equation of accuracy, value has been calculated and it is concluded that extreme gradient boosting is best with 85% accuracy





**Heart Disease Prediction using ML**

Age:  Sex:  Chest Pain types:

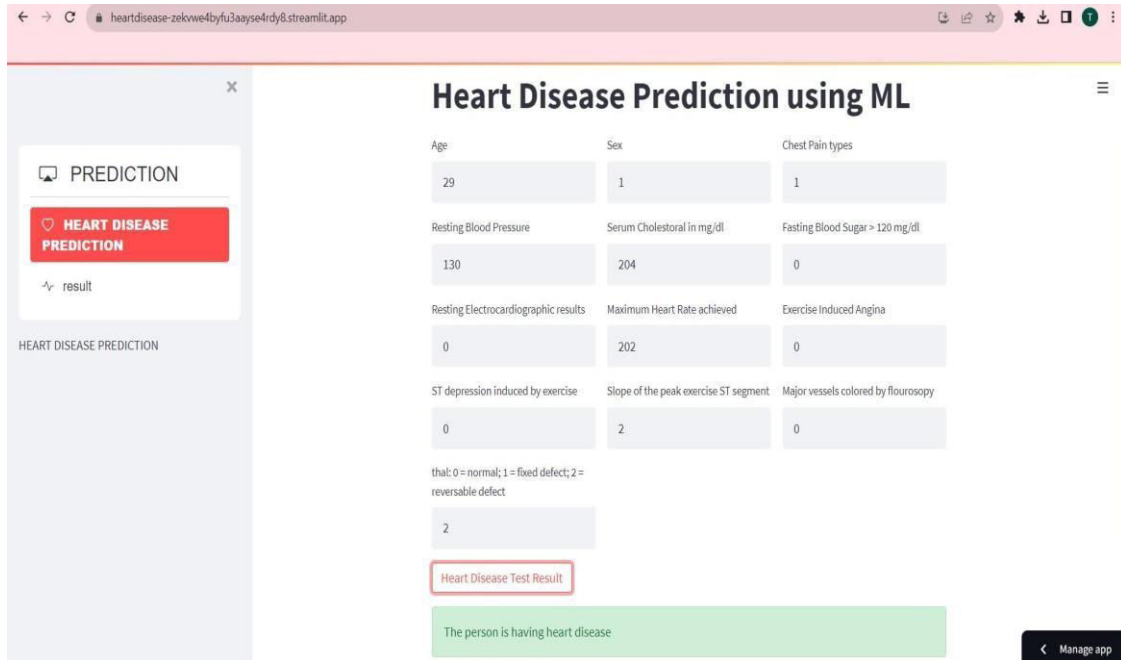
Resting Blood Pressure:  Serum Cholesterol in mg/dl:  Fasting Blood Sugar > 120 mg/dl:

Resting Electrocardiographic results:  Maximum Heart Rate achieved:  Exercise Induced Angina:

ST depression induced by exercise:  Slope of the peak exercise ST segment:  Major vessels colored by fluoroscopy:

thal: 0 = normal; 1 = fixed defect; 2 = reversible defect

Heart Disease Test Result



The screenshot displays a web application interface for heart disease prediction. On the left, there is a sidebar with a 'PREDICTION' button and a 'HEART DISEASE PREDICTION' label. The main area contains a form with the following fields and values:

Age	Sex	Chest Pain types
29	1	1
Resting Blood Pressure	Serum Cholesterol in mg/dl	Fasting Blood Sugar > 120 mg/dl
130	204	0
Resting Electrocardiographic results	Maximum Heart Rate achieved	Exercise Induced Angina
0	202	0
ST depression induced by exercise	Slope of the peak exercise ST segment	Major vessels colored by fluoroscopy
0	2	0

Below the form, there is a legend: "that: 0 = normal; 1 = fixed defect; 2 = reversible defect". A 'Heart Disease Test Result' button is present, and a green box at the bottom states: "The person is having heart disease". A 'Manage app' button is located in the bottom right corner.

## V. CONCLUSION

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients.

After performing the machine learning approach for training and testing we find that accuracy of the Logistic Regression is better compared to other algorithms. Accuracy is calculated with the support of the confusion matrix of each algorithm, here the number count of TP, TN, FP, FN is given and using the equation of accuracy, value has been calculated and it is concluded that logistic Regression is best with 85% accuracy.

Finally comparing the results We get that accuracy of logistic Regression is about 85% . Heart diseases are one of the major concerns of society and the number of people affected by these diseases is increasing day by day and it is important to find a solution so with the help of data analytics and machine learning models ,we can determine these diseases and have a better chance of treating it.

## REFERENCES

1. M. Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), 43-8
2. Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. *International Journal of Computer Applications*, 47(10), 44-8.
3. Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine*, 10(2), 334-43.
4. Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. *International Journal of Computer Science and Information Technologies*, 6(1), 637-9.
5. Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based
6. decision support framework for intelligent heart disease diagnosis. In *International Conference on Information Society (i-Society 2014)* (pp. 259-64). IEEE. ICCRDA

7. 2020 IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012072 IOP Publishing doi:10.1088/1757-899X/1022/1/012072 9
8. Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A
9. coronary heart disease prediction model: the Korean Heart Study. *BMJ open*, 4(5), e005025.
10. Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J &
11. Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. *Arteriosclerosis, thrombosis, and vascular biology*, 33(9), 2267-72.
12. Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease
13. prediction using lazy associative classification. In 2013 International Mutli-
14. Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details