

# Heart Attack Prediction System Using Logistic Regression

<sup>1</sup>Student, Ayushi Chouksey, G H Rasoni University, Saikheda(M.P.)  
<sup>2</sup>Student Anurag Burange, G H Rasoni University, Saikheda(M.P.)

<sup>3</sup>Student, Ashwini Choudhary, G H Rasoni University, Saikheda(M.P.)  
Corresponding Author: Prof. Sonia Bajaj, Head of Department, Computer Science Engineering  
G H Rasoni University, Saikheda(M.P.)

Submitted: 15-04-2022

Revised: 27-04-2022

Accepted: 30-04-2022

**ABSTRACT:** Heart disease is a unusual condition of the heart and the heart circulation. Heart disease is also known as cardiovascular disease which is our country's main executioner. Nowadays it is the leading cause of death in whole world. The chances of certain heart diseases may be increased by smoking, high blood pressure, high cholesterol, unhealthy diet, lack of exercise etc. There are a type of disease that affects the heart. The most common heart disease is coronary artery disease which can lead to chest pain, heart attacks or stroke. Due to the increase in diseases, the health care industry is producing more data. But this data is not being used properly. It is very challenging to identify the symptoms of cardiovascular disease. It requires careful understanding and analysis on patient's medical records and identifies parameters that cause heart disease. This diagnosis or analysis is a difficult task i.e., it should be performed precisely and efficiently. There is a need to predict the disease to avoid risks associated with it and aware the patient well in advance. In this project we have prepared a heart disease prediction system to predict whether the person is diagnosed with heart disease or not using the medical history of the individual. The implementation of work is done on Cleveland heart disease dataset from the University of California Irvine. We used different algorithms of machine learning such as Logistic Regression, Decision tree, K Nearest Neighbours algorithm, Support vector machine. Thus, this report presents the comparative based study by analysing the performance of different machine learning algorithms.

**KEYWORDS:** Heart Disease Prediction, Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbour.

## I. INTRODUCTION

Machine learning is a method of data exploration that automates analytical model building. It is the study of computer algorithms that can improve automatically through experience and by use of data. It is a diverse field and its scope and implementation is increasing day by day. Machine learning enables a machine to automatically learn from data, improve performance from experiences and predict things without being explicitly programmed. Here, prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome. We can take knowledge from it and use it in our project Heart disease prediction system as it will be beneficial for lots of people. Cardiovascular disease are the leading cause of death globally. As estimated in current scenario approximately 17.9 million people died from CVDs, representing 32% of all

global deaths, 85% were due to heart attack and stroke. Heart is most important organ in the body. It keeps your blood flowing and your organs going. It is a tireless muscle that pumps more than two thousand gallons of blood everyday. Any irregularity to heart can cause distress in other parts of the body. Any kind of disturbance to normal functioning of the heart may cause serious health issues which can be classified as heart disease. In

today's world heart disease is one of the key reasons for mortality in India. There are the following heart disease risk factors such as unhealthy

lifestyle, smoking, alcohol, high blood pressure, cholesterol

which can cause heart disease. Therefore, it is necessary to remain healthy and should be aware of the heart disease at initial stage. There are the large set of medical created by medical experts that are

available for analyzing and extracting valuable knowledge from it. Cardiovascular disease identification techniques are to complicated. For this

reason there is a need to implement a support system to predict heart disease through a machine learning model. Our project can help to predict the people who are likely to diagnose with heart disease by

using their medical history. This project focuses on mainly machine learning techniques which is subfield of data mining which handles large scale dataset efficiently. We are using the following algorithms - Logistic regression, KNN, SVM and decision tree. The highest accuracy achieved is from Logistic regression 89% which is better than the previous system. The objective of our project is to check whether the patient is diagnosed with heart disease or not based in their medical attributes such as gender, age, chest pain, fasting sugar level. The

dataset taken is from the University of California Irvine (UCI) machine learning repository to test on different data mining techniques. Using this dataset we predict whether the patient can have heart disease or not. Here in this dataset there are 14 medical attributes of a patient and one class which is named as Target it will predict whether the patient is likely to have a heart disease. These medical

attributes are trained with the following algorithms such as Logistic Regression, KNN, SVM and Decision tree. We found out that the logistic regression algorithm is most efficient for our model as it gives the highest accuracy among all the algorithm. We finally made a detector tool which can be able to detect heart disease at early stage which will help to take effective treatment to patients on time and avoid severe consequences.

## II. LITERATURE REVIEW

There are number of works done related to disease prediction systems using different machine learning algorithms.

[1] Baban Uttamrao Rindhe, Nikita Ahire, Rupali Patil, Shweta Gagre, Manisha Darade (2021) developed a predicting model for heart disease prediction. Its main aim is to predict the heart disease of a patient using machine learning

algorithms. It deals with various machine learning algorithms Random Forest, Support Vector Machine, Artificial Neural Network. The highest accuracy is achieved by Support Vector Classifier. Anupama Yadav, Levish Gediya, Adnanuddin Kazi (2021) proposed a predictive system. The major objective is to obtain improved accuracy for detecting heart disease using

algorithms in which target output calculates whether the person has heart disease or not. Here it converts categorical variables like sex, cp, fbs to dummy variables. Equating machine learning algorithms such as SVM, Decision Tree, Random Forest and KNN. For every algorithm they trained their model. The predicted highest accuracy is 86.2% using Gaussian SVM Kernel.

[2] Dr. Poonam Ghuli and students recommended the proposed work which predicts the chances of heart disease and classifies patients risk level by implementing data mining techniques such as Naïve Bayes, Decision Tree, Logistic Regression and Random Forest. It compares the other ML algorithms and found that the Random Forest gives the highest accuracy of 90.16%. Here they also used the performance analysis metrics using accuracy score, precision (P), recall (R) and F-measure.

[3] Harshit Jindal, Sarthak Agarwal, Preeti Nagrath gives us the model in finding the probability of the classifier to correctly and accurately identify the heart disease. It uses the following techniques - Logistic Regression, KNN and Random Forest classifier. The prediction is based on the medical attributes such as gender, age, chest pain, fasting sugar level etc. The highest accuracy is 88.52% by KNN algorithm. Use of more training data ensures the higher chances of model to predict accuracy.

[4] Boshra Brahmi et al, developed different data mining techniques to evaluate the prediction and diagnosis of heart disease. The main objective is to evaluate different classification techniques such as J48, Decision Tree, KNN, SVM and Naive Bayes. After this evaluating some performance in measure of accuracy, precision, sensitivity, specificity are evaluated and compared J48 and Decision Tree gives the best techniques for heart disease predictor.

[5] Xian-Yan Gao, Abdelmegeid Amin Ali (2021) ensembles learning methods used to enhance the performance of predicting heart disease. Here it uses two features of extraction methods: Linear discriminant analysis (LDA) and principal component analysis (PCA). The data is split 75% for training and 25% for testing and also ninefold cross-validation is utilized in the training set. Different algorithm applied are KNN, Decision Tree, Random Forest and Naive Bayes. The model

evaluation is performed focusing on accuracy score, recall, precision, f-score, ROC and AUC. Decision tree has achieved the best performance.

[6] Dhai Eddine Salhi, Abdelkamel Tari and M-Tahar Kechadi (2021) used data analytics to detect and predict diseases patient. They selected the most relevant features by the correlation matrix and then applied ML algorithms Neural Networks, SVM and KNN. The highest accuracy is obtained by Neural Networks is 93%. We also notice that the results are close with simple progression of Logistic Regression. The authors also tested the diagnosis of heart patients by applying two techniques: genetic algorithms and KNN algorithm but found satisfaction with KNN algorithm.

### III. METHODOLOGY

The methodology is a process which

includes steps that transform given data into recognized data patterns for user's knowledge. As per the data and information we have gathered, we found that the following tasks must be carried out to get much accurate predictions.

#### 1. Data Collection:

Here, we have taken Cleveland Heart Disease dataset from University of California Irvine (UCI) machine learning repository to test on different data mining techniques. It contains 303\*14 rows and columns. Having 14 attributes, including the predicted attribute i.e., Target. The 13 attributes or feature were the inputs and the class i.e., Target is the output which will predict result in binary form. If its integer value is 0, then it means the patient is healthy he is not suffering from heart disease. The output is totally dependent on the features which user or patient is giving. The following are the attributes given which are used in this research. They are:

Sr no.	Attribute Description	Representative Icon
1.	Age : Age of the patient at the current time of examination.	Age
2.	Sex : Categorized as either 0 or 1 such as 0=female and 1=male.	Sex
3.	Chest Pain : Represents the severity of chest pain patient is suffering.	CP
4.	Resting Blood Pressure : Resting blood pressure value of patient in mmHg(unit).	Trestbps
5.	Cholestrol : Cholestrol of patient in mg/dl(unit).	Chol
6.	Fasting Blood Sugar : Categorized as either 0 or 1 such as 1=if fbs>120 mg/dl (true) else 0 (false).	Fbs
7.	Resting ECG : It shows the result of ecg.	ECG
8.	Max Heart Rate : Maximum heart rate achieved by patient.	Thalach
9.	Exercise induced angina : Categorized as either 0 or 1.	Exang
10.	Oldpeak : Displays the value of ST depression of any patient induced by exercise w.r.t rest.	Oldpeak

11.	Slope : Describes patient condition during peak exercise.	Slope
12.	No. of major vessels : Result of fluoroscopy.	CA
13.	Thalassemia : It is classified into three ranges from 1 to 3, where 1=normal, 2=fixed defect, 3=reversible effect.	Thal
14.	Target : This is the prediction column for diagnosis of heart attacks. It is classified into two types 0 and 1, where 0 means no possibility of heart attack and 1 means possibilities of heart attack.	Target

Table 1.Attributes table

## 2. Preprocessing of the dataset:

It is the most important step in the implementation of the model. The higher the quality of the data higher accuracy it will predict. Therefore, it is extremely important to check or pre-process the data before feeding it into our model. In this we check for the NA values. We also differentiate the categorical values and the continuous values. We also check for the missing values in the dataset and if found it should be removed.

## 3. Features selection:

Feature scaling is the most important for machine learning algorithms that calculate distances between data. If we do not scale, the features with a higher value range it will start dominating when we will perform the calculations on distances. chances of heart disease. The following are the algorithm given below:

**A. Logistic Regression** – Logistic Regression is a machine learning algorithm which is used for the classification problems, which is used for the classification problems. It is predictive analysis algorithm and based on the concept of probability.

We can also call logistic regression a linear regression model but the logistic regression uses more complex cost function, this cost function can be defined as the sigmoid function. The hypothesis of logistic regression tends it to limit the cost function between 0 and 1.

Sigmoid function,  $f(x) = \frac{1}{1 + e^{-x}}$

## 4. Checking the distribution:

In this section we check the distribution of the target i.e., 0 and 1. We plotted a graph in which it is showing that the number of person having the healthy heart is more than that of defective heart.

## 5. Data splitting:

In this step, the heartdisease dataset is splitted into two parts: 70% is for training set and 30 % is for the testing set. The training set is used for training the model and the testing set is utilized to evaluate the model.

**6. Training Model:** In this model we have applied various machine learning algorithms so that we can find the most suitable and accurate one to predict the The maximum accuracy achieved is 89.0%.

**B. Decision Tree** – Decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems. It uses tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. The accuracy achieved Is 70.3%.

**C. K-Nearest Neighbors** – It is based on supervised learning technique. It assumes the similarity between new case/data and available cases and put

the new case into the category that is most similar to the available category. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. The accuracy achieved is 62.64%.

**D. Support Vector Machine** –Support vector machine is one of the most popular supervised learning algorithms. The goal is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. The accuracy achieved is 85.7%.

**7. Evaluating models :**

We used the accuracy score for the evaluation of accuracy. It is one metric for evaluating classification models. For binary classification accuracy can also be calculated in terms of positives and negatives.

**8. Predictive Model :**

After applying the following steps on dataset we train our model and also tested it. Now, we developed a predictive model which will show output for the new patient's input.

```

import pandas as pd
new_data=pd.DataFrame({'age':12,
                       'sex':'M',
                       'trestbps':100,
                       'chol':100,
                       'fbs':100,
                       'restecg':1,
                       'thalach':100,
                       'exang':1,
                       'oldpeak':1,
                       'slope':1,
                       'ca':1,
                       'max':1,
                       'inter':[1]}
                       )
new_data
    
```

Fig1. Predictive Model

**9. Plotting Graph :**

After applying all the machine learning algorithms i.e., Logistic Regression, Decision Tree, Support vector Machine and K-Nearest Neighbour we plot the overall graph of the algorithm showing the accuracies level of the algorithms.

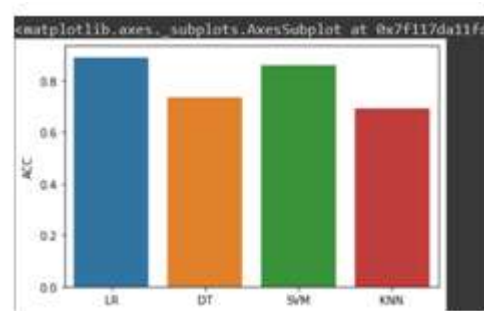


Fig2. Comparison Graph

**IV. RESULT & ANALYSIS**

As the health related issues is growing day by-day. One such issue is heart disease. Curing the disease on time has become the most important. There is a need to predict the diseases on time which will be cost efficient and helpful for the doctors. We come up with the conclusion that as nowadays heart disease problem is the major cause of death. So we will make a predictive system which will predict in advance whether the person having the heart disease or not in binary form. We have taken ideas from base paper in which they have used the heart disease prediction dataset from University of California Irvine (UCI) machine learning repository to test on different data mining techniques. They split the data into 75% for training and 25% for testing. They used four algorithms – SVM, Decision tree, Naive Bayes and Logistic Regression. The highest accuracy achieved is 84%. We also took the same dataset and applied various machine learning algorithms. We used the following algorithms – Logistic Regression, Decision Tree, Support Vector Machine and K-Nearest Neighbors. Our main focus was on Logistic Regression as we are working on discrete values. We used the sigmoid function and also got the highest accuracy of 89% from logistic regression. We also build the predictive system in which when we are giving new inputs it will predict whether the patient is suffering from heart disease or not.

Sr No.	Algorithms	Accuracy	Proposed accuracy
1.	Support Vector Machine	81.57%	85.7%
2.	Logistic Regression	82.89%	89.0%
3.	Decision Tree	80.43%	70%

Table 2. Result Table



The following image shows the predictive model we obtained for “Heart Attack Prediction System”.

```
prediction=lg.predict(new_data)
if prediction[0]==0:
    print("The Person does not have a Heart Disease")
else:
    print("The Person have Heart Disease")
The Person does not have a Heart Disease
```

Fig.3 Predictive Model

## V.CONCLUSION

We concluded that our project “Heart Attack Prediction System Using Logistic Regression” can be useful in everyone’s day-to-day life. As it is most important for the health care system. With the increase in number of deaths due to heart disease it is very important to build a system which can easily predict the disease just by using some inputs such as age, blood pressure, cholesterol, ecg etc. We applied several machine learning algorithms such as – Support Vector Machine, Decision Tree, K -Nearest Neighbor and Logistic Regression. We also found out the best algorithm which is giving more accuracy of 89.0% in comparison to other algorithms. The Logistic Regression algorithm is the most efficient algorithm we have concluded from the result. This will help professional in predicting the heart diseases effectively and efficiently. The working model can also help in reducing treatment cost by providing initial diagnostics in time. If the person doesn’t want to visit to the doctor, he/she can also check for his or her health inputs, so just by entering the symptoms one can be able to know whether he/she will suffer from disease or not. We can also update this project in future by adding more attribute to the data set. And we can also make it more interactive by designing its GUI. After building its GUI we can put it into hospital’s website which will help user to access it from anywhere and get to know their result.

## BIBLIOGRAPHY

- [1]. Xiao-Yan Gao,<sup>1</sup> Abdelmegeid Amin Ali,<sup>2</sup> Hassan Shaban Hassan,<sup>2</sup> and Eman M. Anwar “Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method” <https://doi.org/10.1155/2021/6663455>.
- [2]. Dr. Poonam Ghuli et al, “Heart Disease Prediction using Machine Learning” International Journal of Engineering Research & Technology (IJERT), ISSN:

- 2278-0181, Vol. 9 Issue 04, April-2020.
- [3]. Rishabh Magar, Rohan Memane, Suraj Raut, 1 Prof. V. S. Rupnar, 1 Computer Department, 1 MMCOE, Pune, India “HEART DISEASE PREDICTION USING MACHINE” 2020 JETIR June 2020, Volume 7, Issue 6 ‘www.jetir.org (ISSN-2349-5162)’.
- [4]. American Heart Association, Heart Failure, American Heart Association, Chicago, IL, USA, 2020, <https://www.heart.org/en/health-topics/heart-failure>.
- [5]. World Health Organization, Cardiovascular Diseases, WHO, Geneva, Switzerland, 2020, [https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1).
- [6]. M. Sanz, A. Marco del Castillo, S. Jepsen et al., “Periodontitis and cardiovascular diseases: consensus report,” Journal of Clinical Periodontology, vol. 47, no. 3, pp. 268–288, 2020. View at: [Publisher Site | Google Scholar](#).