



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 1, January 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Review on Heart Disease Prediction Using Machine Learning

Chaitrali Sorte, Raghav Sharma, Saurabh Shinde, Eshan Vasaikar, Mahesh Yewle

Engineering Student, Dept. of Information Technology, PCCOE, Pune, Maharashtra, India

ABSTRACT: Heartdisease instances are increasing at an alarming rate, and it's critical and important to be able to predict such diseases in advance. Heart disease prognosis has become one of the most difficult challenges in the medical sector in recent years. In the modern period, about one person dies from heart disease every minute. In the realm of healthcare, data science is critical for analysing massive amounts of data. Due to the difficulty of forecasting cardiac sickness, it is vital to automate the procedure in order to reduce the risks associated with it and to inform the patient well in advance. We created a heart disease prediction system that uses the patient's medical history to forecast whether or not the patient would be diagnosed with heart disease.

The heart disease dataset from the UCI machine learning repository was used in this work. Using data mining techniques such as Decision Tree, Logistic Regression, and Random Forest, the suggested work predicts the likelihood of HeartDiseaseand analyses the risk level of patients. As a result, this study compares the performance of several machine learning methods. The experiment results show that the Logistic Regression method has the highest accuracy of 81.96 percent when compared to other machine learning techniques. This investigation has provided us with valuable information that can aid in the prediction of heart disease patients. It's based on the. pybn file format.

KEYWORDS: Random Forest(RF), Logistic Regression(LR), Decision Tree (DT), Heart Disease Prediction

I. INTRODUCTION

The work described in this paper focuses on several data mining techniques used in the prediction of heart disease. The heart in human body is the very important organ in the body. In a nutshell, it controls blood flow throughout our bodies. Any heart irregularity might induce discomfort in other parts of the body. Heart disease is defined as any interruption in the normal functioning of the heart. In today's society, heart disease is one of the most common causes of death. Heart disease can be caused by an unhealthy lifestyle, such as smoking, drinking too much alcohol, or eating too much fat, which can lead to hypertension. Each year, more than 10 million people die around the world, according to the World Health Organization. as a result of heart disease. The only way to avoid heart-related diseases is to live a healthy lifestyle and catch them early.

1.1 Aim and Objective:

The project's goal is to properly forecast cardiac illness so that doctors can deliver the best treatment available.

The major goal of this research is to give clinicians a tool to help them diagnose cardiac disease at an early stage. As a result, it will be easier to deliver appropriate treatment to patients while avoiding serious effects. The importance of machine learning in detecting hidden discrete patterns and analysing the data is critical. Following data analysis, machine learning approaches aid in the prediction and early diagnosis of cardiac disease.

We're working on a model that employs a machine learning algorithm to predict cardiac disease, and we're evaluating different machine learning models based on their accuracy. Our data set is trained using decision trees, logistic regression, and a random forest model. We anticipate which algorithm is the best based on the results

1.2 Motivation of Project

In today's healthcare, the key difficulty is to provide high-quality services and precise diagnoses. Even though heart disease has been identified as the leading cause of death worldwide in recent years, it is also one of the diseases that may be effectively controlled and managed. The precise timing of disease discovery determines the entire accuracy of disease management.

1.3 Organization of report

For the first chapter introduction to the project is given. Individual seminar topic, introduction to the topic is given. Next is survey paper information is mentioned where three research papers mentioned along with objective, summary, results, limitations and technologies used in each paper is written in brief. Following the survey, one comparison table



is there which includes common factors observed in survey papers. Conclusion and references are there at the end of the report.

II. LITERATURE REVIEW

Using the UCI Machine Learning dataset, a lot of work has been done to predict cardiac disease. Varied data mining approaches have been used to achieve various levels of accuracy, as discussed below.

AvinashGolande and colleagues look into a number of machine learning algorithms for classifying cardiac illness. In this study, the accuracy of the Decision Tree, KNN, and K-Means algorithms for classification were examined[1]. This study found that Decision Tree had the highest accuracy, and that it may be made more efficient by combining several methodologies and perfect parameters.

Fahd Saleh Alotaibi created a machine-learning technique that analyzes five different approaches. [2]. The Rapid Miner tool achieved higher accuracy when compared to Matlab and Weka. This study assessed the accuracy of the classification methods Decision Tree(DT), Logistic Regression(LR), Random Forest(RF), Naive Bayes, and SVM. The decision tree algorithm was found to be the most accurate.

To predict heart illness, Nagaraj M Lutimath and colleagues applied Naive Bayes classification and SVM. Mean Absolute Error (MAE), Sum of Squared Error, and Root Mean Squared Error are the performance measures used in the analysis, and it has been shown that SVM leads Naive Bayes in terms of accuracy. [3]

SP.Rajamhona, In 2018, conduct an assessment of a neural network-based system for predicting cardiac disease.[4]and dataset use is "Cleveland Heart Disease Data set".Algorithm use in these paper is Artificial Neural Network and they have Accuracy 80.46%

In 2016, Salma Banu N.K used data mining and big data analytics to predict cardiac disease at an early stage.[5] He used Naive Bayes, Decision Trees, and SVM machine learning algorithms to obtain an accuracy of 84.1 percent.

III. METHODOLOGY

This work used the SVM, LR, DT, and RF machine learning algorithms to predict heart disease. All of these algorithms were written in the Python programming language and tested on a collection of data. The python programming language was used to generate these models. The dataset is split into two parts: training and testing. The SVM, LR, DT, and RF models are then trained appropriately.

The model is trained using techniques from the SVM and its prediction accuracies are calculated using the testing set.

The proposed model has been evaluated on a data set of 303 patients using the Python programming language. The data collection, titled heart disease data set for study, is publicly available on Kaggle. The data collection is offered as a CSV file, which is best suited for use with the Python programming language.

The accuracy of each technique is calculated, and it is determined that logistic regression provides the best results.

Data Set Feature

| Sr.No. | Attribute Description | Different Values Of Attribute |
|--------|--|---|
| 1 | Age- age of a patient | Values of age are ranging from 29 to 71 |
| 2 | Sex-The sex refers to a person's gender (0 for Female, 1 for Male) | 0,1 |
| 3 | CP- denotes the level of the person's chest discomfort. | 0,1,2,3 |



| | | |
|----|--|-----------------------------------|
| 4 | RestBP-It represents the patient’s Blood Pressure | Values ranging from 94 to 200 |
| 5 | Chol-It displays the patient's cholesterol level. | Values ranging from 126 and 564 |
| 6 | FBS-It stands for the patient's fasting blood sugar. | 0,1 |
| 7 | Resting ECG-The result of an ECG is shown in the resting ECG. | 0,1,2 |
| 8 | Heartbeat- displays the patient's maximum heart rate. | Values are ranging from 71 to 202 |
| 9 | Exang- is a test that can be done to see if you have exercise-induced angina. If yes=1 and no=0, | 0,1 |
| 10 | The level of Multiple Depression in the patient is described by OldPeak. | Values in between 0 and 6.2 |
| 11 | Slope-The patient's state during peak exercise is described by the slope. It's broken down into three segments (Unsloping, Flat, Down sloping) | 1,2,3 |
| 12 | CA- Result of fluoroscopy | 0,1,2,3 |
| 13 | Thaltest is required for patient who is experiencing chest pain, asthma or difficulty in breathing. There are four different types of values that indicate the Thallium test. | 0,1,2,3 |
| 14 | Target-It Last column of the dataset. It is class or label Colum. It represents the no. of classes in dataset. This dataset has a categorize in binary, which means there are only two options (0,1). Class "0" indicates that there is a low risk/healthy of heart , whereas class "1" indicates that there is a significant risk/unhealthy of heart . The value "0" or "1" is determined by the 13 other coloumn/attributes. | 0,1 |

IV. RESULTS ANALYSIS

Following table shows result of different algorithm of ML accuracy to detect person having heart problem or helthey.

Table showing result

| Name Of Algorithm | Heart Disease Accuracy |
|---------------------|------------------------|
| Decision tree | 72.13% |
| Random forest | 72.13% |
| Logistic regression | 81.96% |
| SVM | 78.73% |



V. CONCLUSION

With the rising number of deaths due to heart disease, it is becoming increasingly important to build a system that can effectively and accurately forecast heart disease. The study's purpose was to find the most efficient machine learning method for detecting heart issues. Using the UCI machine learning repository dataset, this study analyses the accuracy score of Logistic Regression (LR), Decision Tree(DT) and Random Forest(RF) for predicting heart disease. According to the findings of this study, the LR is the most effective algorithm for predicting heart disease, with an accuracy score of 81.96 percent. The work could be enhanced in the future by developing a system based on the LR algorithm and using a larger dataset than the one used in this analysis, which would enable health professionals in successfully and efficiently anticipating cardiac disease.

REFERENCES

1. Avinash Golande and Pavan Kumar T, they work on "Heart Disease Prediction Using Effective Machine Learning Techniques," and do the publication in International Journal of Recent Technology and Engineering, Vol. 8, No. 4, pp.944-950.
2. In 2019 ,Fahd Saleh Alotaibi, "Implementation of a Machine Learning Model to Predict Heart Failure Disease," done publication in International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 10, No. 6.
3. In 2019,Nagaraj M Lutimath, Chethan C, and Basavaraj S Pol., "Prediction Of Heart Disease Using Machine Learning," In International Journal Of Recent Technology and Engineering, vol. 8, no. 2S10, pp. 474-477.
4. SP.Rajamhona,2018.Analysis of neural network based heart disease prediction system. Cleveland and statlog heart disease database , UCI repository.
5. In 2016, Salma Banu N.k. published Prediction of cardiac disease at anlow level of disease using big data analytics and data mining. Repository at the University of California, Irvine.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details