

Prediction of Heart Disease Using Machine Learning

Prasanna M¹, Shrijith Shetty P², Mamatha K³

^{1,2} UG Scholar, Department of Computer Science and Engineering, Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangalore, Karnataka, India

³ Assistant Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangalore, Karnataka, India

ABSTRACT

Article Info

Volume 7, Issue 4

Page Number: 213-217

Publication Issue :

July-August-2021

Article History

Accepted : 10 July 2021

Published : 19 July 2021

Heart disease is one of the most significant causes of mortality in the world today. Prediction of Heart disease is a critical challenge in the area of clinical data analysis. Machine learning has been shown to be effective in assisting in making decision and predictions from large quantity of data produced by the health industry. Several types of heart diseases are expanding day by day because of way of life, genetic problem, blood pressure, cholesterol level, pulse rate etc. So the diagnose of disease plays important role for the prevention of heart related problems. Researchers received different methods to analyze it. This system aims at Predicting heart disease with various machine learning techniques and increasing the accuracy of the system. Classification approach consists of two algorithms such as KNN classification algorithm and Decision tree algorithm. The result of classification shows 86% accuracy by using n number of neighbors in this approach.

Keywords : Prediction, Heart Disease, Machine Learning

I. INTRODUCTION

Now-a-days death there is increased death rate in heart related diseases. Heart Attack is the main reason for the cause death of the people. There are so many reasons and factors which involve in the occurring of the heart diseases. The survey analysis describe the death of the male is more than the female due to the heart diseases because of the smoking and drinking habit of the male. The human life is depending on the working functionality of the heart because heart supply the blood to the all the organs of the body. As per the Indian Heart

Association, 50% of heart strokes occur under 50 years of age and 25% of all heart strokes occur under 40 years of age in Indians. Urban population is thrice as vulnerable to heart attacks as rural population.

The purpose of collecting relevant data pertaining all elements related to our field of study, by training the data as per the proposed algorithm of machine learning and predict how strong is there a possibility for a patient to contract a heart disease. For the purpose of patients entering data, we suggest making use of the easily available sensors in watches and cell phones to measure the simple factors. Our main aim

behind developing the system is to make it user-friendly so that regular monitoring of the patient is made possible.

The user inputs its specific medical details to get the prediction of heart disease for that user. The algorithm will calculate the probability of presence of heart disease. The implementation of work is done on Cleveland heart disease dataset from the University of California Irvine (UCI) by machine learning techniques. The result will be displayed on the webpage itself. Thus, minimizing the cost and time required to predict the disease. Format of data plays crucial part in this application.

II. LITERATURE SURVEY

In the research done by ShivaPrakash and JPrerana Support vector machine (SVM) and artificial neural network (ANN). A multilayer perceptron neural network (MLPNN) with three layers was employed to develop a decision support system for the diagnosis of heart disease. This multilayer perceptron neural network was trained by back-propagation algorithm which is computationally an efficient method. Results showed that a MLPNN with back-propagation technique can be successfully used for diagnosing heart disease.

In the research done by Chala Beyene et al, recommended Prediction and Analysis the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in short time. The proposed methodology is also critical in healthcare organisation with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex are some of the attributes are included to identify if the person has heart disease or not. Analyses of dataset are computed using WEKA software

In the research done P.Sai Chandrasekhar Reddy et al, proposed Heart disease prediction using ANN algorithm in data mining. Due to increasing expenses of heart disease diagnosis disease, there was a need to develop new system which can predict heart disease. Prediction model is used to predict the condition of the patient after evaluation on the basis of various parameters like heart beat rate, blood pressure, cholesterol etc

In the research done by S.Prabhavathi et al, [23] proposed Decision tree based Neural Fuzzy System (DNFS) technique to analyse and predict of various heart disease. This paper reviews the research on heart disease diagnosis. DNFS stand for Decision tree based Neural Fuzzy System. This research is to create an intelligent and cost effective system, and also to improve the performance of the existing system. Specifically in this paper, data mining techniques are used to enhance heart disease prediction. The result of this research shows that the SVM and neural networks results highly positive manner to predict heart disease. Still the data mining techniques are not encouraging for heart disease prediction.

In the research done by Sairabi H.Mujawar et al, used k-means and naïve bayes to predict heart disease. This paper is to build the system using historical heart database that gives diagnosis. 13 attributes have considered for building the system. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used. 13 attributes with total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

In the search done by K.Gomathi et al, [16] suggested multi disease prediction using data mining techniques. Nowadays, data mining plays vital role in predicting multiple disease. By using data mining techniques the number of tests can be reduced. This paper mainly concentrates on predicting the heart disease, diabetes and breast cancer etc.,

III. OBJECTIVE

The objective of Prediction of heart disease is to develop a platform which will be simple and easy to use, as here one must provide the patient's medical details and based on the features extracted the algorithm will then detect the heart disease and spot its type. The algorithm is well trained model to make errors in predicting the heart disease and its type hence, in short accuracy is improved and thereby it also saves time and makes easier for doctors as well as patients to predict whether they are prone to any type of heart disease or not, which is otherwise we difficult to do without doctor's involvement.

IV. IMPLEMENTATION

System Implementation is the stage where the theoretical design is converted into a working system, the new system, replacing an existing system. The system is implemented using Anaconda Navigator and data set.

K-Nearest neighbor (KNN) is a simple and nonparametric classifier. KNN is preferred when all the features are continuous. KNN is also called as case- based reasoning and has been used in many applications like pattern recognition, statistical estimation. Classification is obtained by identifying the nearest neighbor to determine the class of an unknown sample. KNN is preferred over other classification algorithms due to its high convergence speed and simplicity.

Support vector is a supervised learning algorithm which can be used for binary classification or regression. SVM is a coordinate of individual observations. It is based on decision planes which defines decision boundaries. It also separated the set of objects having different class. An SVM classifier is a linear classifier where the separating hyper plane is chosen to

minimize the expected classification error of the unseen test patterns in finding significant patterns for heart attack prediction.

Decision Tree Classifier use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub- nodes. LightGBM is a gradient boosting framework that uses tree based learning algorithms. It is designed to be distributed and efficient with the following advantages: Faster training speed and higher efficiency. Lower memory usage. Better accuracy.

DETECTION OF Heart Disease

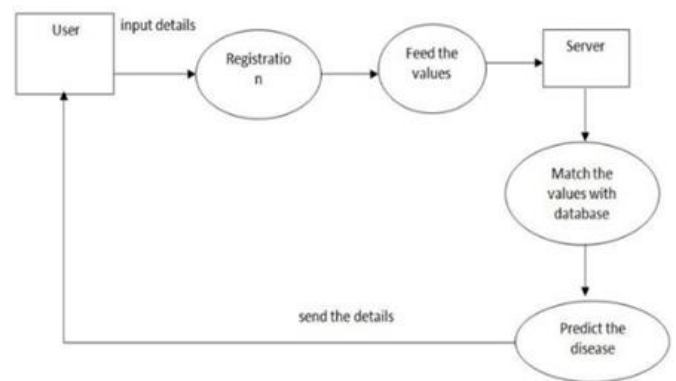


Figure 1: Detection of Heart Disease

ALGORITHM

K-Nearest Neighbor

Step 1: Start

Step 2: Input the heart disease data set

Step 3: Apply pre-processing techniques-Fill in missing values

Step 4: Select the features based on values obtained after applying PSO as FSS

Step 5: Discard redundant features

Step 6: Apply (KNN+IQR) on Predominant features

Step 7: Measure the performance of the KNN+PSO model

Step 8: Classification of data set into patients with heart disease and normal and predict

Step 9: Stop

V. ANALYSIS

The output of the system will give a prediction result if the person has a heart disease, in terms of Yes or No. The system gives an idea about the heart status leading to CAD beforehand. If the person is prone to have heart disease then the result obtained will be Yes and vice versa.

In case of a positive output, he needs to consult a cardiologist for further diagnosis. The statistics of the results obtained during the testing of the dataset is shown in the following table.

Table 1: Result Analysis Table

Decision tree	KNeighbors	LGBM	SVC
78	88	84	84

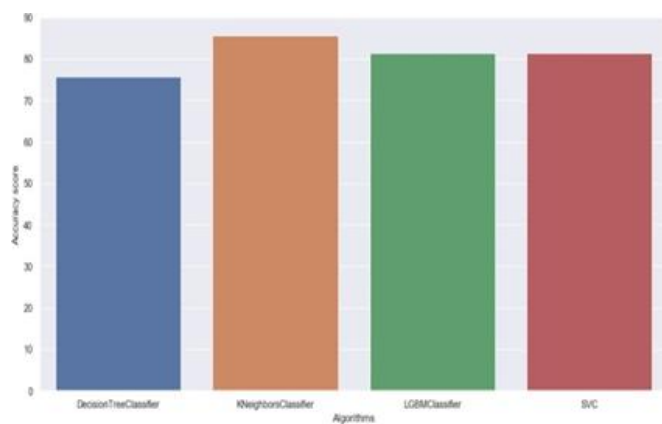


Figure 2 : Training loss and accuracy on dataset

The Figure 2 describes accuracy data on each classifier in order to know which is best classifier on its data.

VI. CONCLUSION

There are so many machine learning techniques detection and prediction of the heart diseases. KNN provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence we use K-Nearest Neighbors in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known and the rate of people dying due to heart diseases will reduce eventually.

In the future the performance of the KNN can be improved by using the combination of the other methods and pruning of the given data. We can use the other machine learning techniques to get the better accuracy.

VII. REFERENCES

- [1]. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review Online: 25 March 2017 DOI: 10.1007/s10462-01
- [2]. Prerana T H M, Shivaprakash N C et al "Prediction of Heart Disease Using Machine Learning Algorithms- Naïve Bayes, Introduction to PAC Algorithm, Comparison of Algorithms and HDPS", Vol 3, PP: 90-99 ©IJSE, 2015
- [3]. Salam Ismaeel, Ali Miri et al "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis", IEEE Canada International Humanitarian Technology Conference, DOI:10.1109/IHTC.2015. 72380 43, 03 September 2015

- [4]. F BrainBoudi,'Risk Factors for Coronary Disease',2016.Online]Available:<https://emedicine.medscape.com/article/164163-overview>.
- [5]. National Health Council,'Heart Health Screenings',2017.Online]Available:http://www.heart.org/g/HEARTORG/Conditions/Heart-Health-Screenings_UCM_428687_Article.jsp#.WnsOAeYPIV
- [6]. ScikitLearn,'MLPClassifier',Online] Available:http://scikitlearn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html
- [7]. Prediction System for heart disease using Naïve Bayes *Shadab Adam Pattekari and Asma Parveen Department of Computer Science and Engineering Khaja Banda Nawaz College of Engineering

Cite this article as :

Prasanna M, Shrijith Shetty P, Mamatha K, "Prediction of Heart Disease Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7, Issue 4, pp.213-217, July-August-2021. Available at
doi : <https://doi.org/10.32628/CSEIT217355>
Journal URL : <https://ijsrcseit.com/CSEIT217355>