RESEARCH ARTICLE   OPEN ACCESS

Heart Disease Prediction Using Supervised Machine Learning Algorithms

Yogesh Gedam, Shivraju Bomble, Uma Kurwade, Bhavana Parchake, Hemant Uike

Information Technology and Engineering Government College of Engineering - Amravati

ABSTRACT

Now a days heart diseases are the major cause of death in the world. More than three million Indians die from heart related disease every year. So it is very necessary to predict the heart disease at a very early stage. The accurate and early prediction will result in the early treatment of the patient and may save many lives. Machine Learning is the best technology in today's world to detect the heart related disease using various attributes and algorithms. We have used supervised machine learning algorithms such as Random Forest algorithm, Naive Bayes algorithm, Decision Tree algorithm, Support Vector Classifier algorithm and KNN algorithm for the accurate prediction of heart disease. We have used UCI repository dataset for training and testing the data. For python programming we have used jupyter notebook as it contain many libraries and header files that make the work easy, precise and accurate.

Keywords:- Supervised Learning, Random Forest, Support Vector Classifier, Decision tree, K-Nearest(KNN), Naive Bayes, Jupyter Notebook, UCI repository.

I.   INTRODUCTION

Heart disease is the major cause of death in Indian from a very long time. Around The annual number of deaths from cardio vascular disease in India is projected to rise from 2.26 million (1990) to 4.77 million (2020). As the rate is increasing rapidly, it is very important that the disease should be predicted at an early stage to get a proper treatment which will surely decrease the death rate. The machine techniques can do this prediction efficiently. Analyzing data from different perspectives and summarizing it into useful information can be done using machine learning. It is an iterative and interactive process of discovering valid, useful, comprehensive and understandable patterns and models in limited data sources. Machine Learning brings a set of tools and techniques that can be applied to processed medical data to discover hidden patterns that provide health-care professionals an additional source of knowledge for making decisions. Using different supervised machine learning algorithms such as Naive Bayes, Random forest, Decision tree algorithms, etc we can find the maximum possibility that whether the person is suffering from heart disease or not.

II.   OBJECTIVES

A. To provide effective method -

By using different machine learning algorithms, heart disease prediction can be done in a very effective way with very less errors and high accuracy.

B. Overcome less number of heart disease experts -

As the number of heart disease experts is less in our country, the model will overcome this problem as it doesn't require experts for the detection of heart disease.

C. Faster Diagnosis -

As the results are provided in very less time, it will help

for the faster diagnosis of heart disease and the patient can go on for further treatments and minimize the impact of disease.

III.   OVERVIEW OF HEART DISEASE PREDICTION

To deal with heart disease problem there is an essential need of prediction system for awareness about heart diseases and for that, we are going to build a machine learning model to check whether a person is having heart disease or not.

A. *Attributes[13]*

Age	Patient's age
Sex	female – 0 and Male -1
Cp	Chest pain type
Trestbps	Blood pressure
Chol	Cholesterol values from 126 to 564
Fbs	Fasting blood sugar > 120 mg/dl, true-1 false-0
Restecg	Resting ECG result (0 to 1)
Thalach	Max heart beat rate (71 to 202)
Exang	Exercise (1-yes 0-no)
Oldpeak	ST depression introduced by exercise relative to rest (0 to .2)
Slope	The slop of the peak exercise ST segment (0 to 1)
Ca	Number of major vessels (0-3)
Thal	3 - normal

Table. 1

B. *Machine Learning Techniques*

B.1 Random Forest algorithm

Building multiple decision trees and merging them together to get a more accurate and stable prediction is done using random forest algorithm. [4]
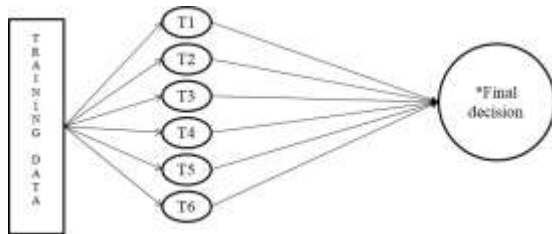

Fig. 1: Random Forest process

B.2 Naive Bayes algorithm

This algorithm predicts on the basis of the probability of an object. [5]
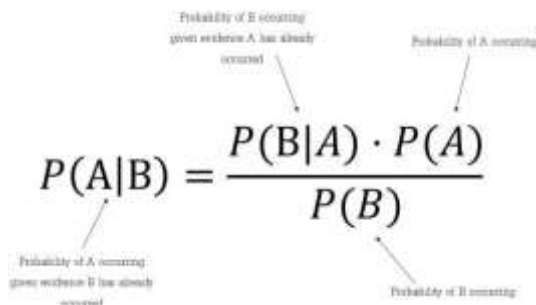

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Fig. 2: Naive Bayes process

B.3 Decision Tree algorithm

It is a supervised machine learning algorithm where the data is continuously split according to a certain parameter. In this, there are two entities namely decision nodes and leaves. [4]
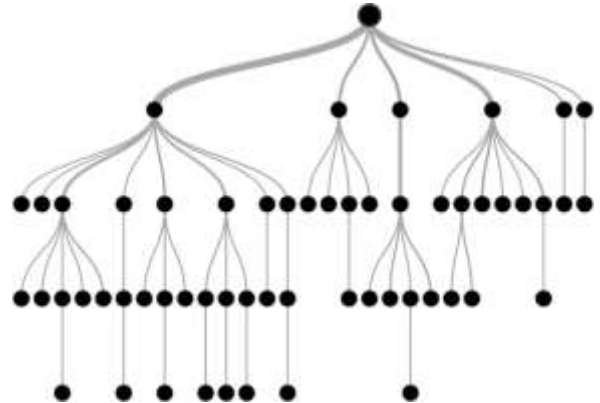

Fig. 3: Decision Tree process

B.4 Support Vector Classifier algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems.[4]
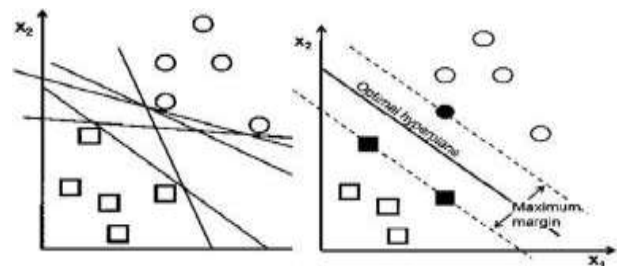

Fig. 4: Support Vector process

B.5 K- Nearest Neighbors algorithm

K-nearest neighbors (KNN) algorithm uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set.[4]

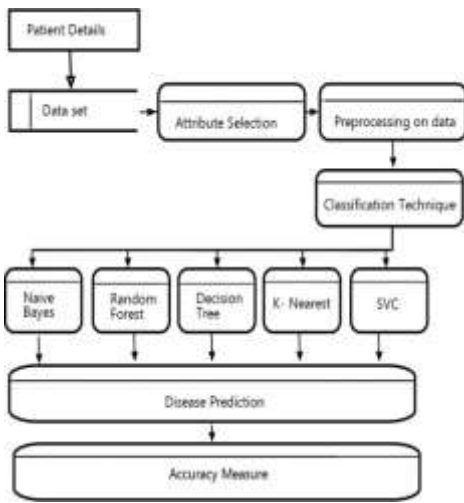

Fig. 5: K-nearest neighbors process

C. Data Flow Diagram

Fig. Data flow diagram

LITERATURE SURVEY

.

This paper is building the system using UCI database that gives diagnosis. 13 attributes with 304 records are used from the heart disease database. According to recommendations 13 attributes have considered for building the system. Using this model to predict whether the patient have heart disease or not based on this values of 13 attributes. The scope of machine learning algorithms is increasing in predicting various diseases. Its capacity to think like a human being is making this concept important and versatile. The heart disease dataset is used for applying random forest to get an accuracy of 85.81%. Using random forest algorithm with well-set attributes gives us the more accurate result. Rajesh Nichenametla et al, [3] this is the most common disease nowadays. We have used different attributes which can relate to heart diseases efficiently to find the best method to predict the heart disease. In this paper we have used Naive Bayes Algorithm which analyzes the dataset based on the risk factors. Also we used decision tree algorithm and combination of various other algorithms to predict the heart disease based on the 13 attributes. The results shows that when the dataset is small, naive Bayes algorithm gives the accurate results and when the dataset is large decision trees gives the accurate results. In this way heart disease is predicted by exploring the various effective algorithms such as Naive Bayes, random forest, Decision tree, k-nearest, support vector classifier etc. The objective of this paper is to predict if the persons suffers from heart disease or not based on results given by various algorithms that we have used. The user enters the input values from the patient's health report. The data set used for model which predicts the probability of having heart disease. In this paper used UCI data set. This database consists of total 13 attributes. The main aim of this paper is to effective and properly prediction of heart diseases at the earliest stage and reduces the cost of medical expenses. This can be achieved by various types of techniques and analyzes their performance to give the accurate result. Models based on supervised learning algorithms such as Support Vector Classifier (SVC), K-Nearest Neighbor (KNN), Naive Bayes, Decision Trees (DT), and Random Forest (RF) are used.

Fig: Comparative Chart.

Table. 2

	Reference		Author	Propose	Algo. used	Acc. %
1	International Journal of Engineering Research and Technology (IJERT) 2019-Conference Proceedings	2017	Yashendra K. Singh	Efficient heart disease prediction system using Random Forest.	Random forest	85.8%
2	International Journal of Engineering and Technology 7(2,8) (2018)	2018	V.V Ramalingam	Heart disease prediction system using ML techniques survey	Support vector machine k-nearest Random forest Decision tree	95.56%
3	International Journal of Engineering and Technology 7(2)	2018	Rajesh Nichenametla	Prediction of heart disease using ML algorithms	Naive Bayes KNN	83% 80%.
4	International Journal of Advanced Science and Technology, Vol 29, No.3s, pp.78-87	2020	A. Rajdhan	Heart disease prediction system using ML.	Decision tree Naive Bayes Random forest	short time results
5	International Journal of Innovation Research in Science, Engineering and Technology, Vol.7	2018	Prof. Shubhangi Patil.	Disease Prediction Using ML Over Big Data.	Naive Bayes	94.80%.
6	International Conference for Emerging Technology(INCE)	2020	Sneha Grampurohit.	Disease Prediction using ML Algorithm.	Decision Tree Naive Bayes Random Forest	95%
7	International Conference on Electrical and Electronics (ICE3-2020	2020	Archana Singh	Heart Disease Prediction Using ML Algorithm	Decision Tree Naive Bayes	87%

machine learning algorithms. V. V Ramalingam el at,[6]. This paper presents a survey of various algorithms and

D. Results

ACCURACY COMPARISON

Algorithm	Accuracy
Random Forest	85.71%
Naive Bayes	81.32%
Support Vector machine	63.74%
Decision tree	71.43%
KNN	61.54%

Table. 3

CONCLUSION

It is concluded that some algorithm such as naive Bayes and the random forest is more accurate than support vector and k nearest neighbor. So that more than one combination of algorithms will give more accurate results so that ensemble learning is important to predict more accurate results.

Based on the above results, it can be concluded that using a machine learning algorithm we can identify cardiovascular diseases. Using appropriate data and machine learning algorithm it is very accurate to confirm the disease but there's always scope to improve the machine learning model so that it improves results when over fitting data is provided. In this paper we have studied various supervised machine learning algorithms such as random forest, naive bayes, decision tree, support vector machine and k-nearest neighbor. Among all these algorithms, random forest is the best algorithm to use as it gives highest accuracy, easy to implement and gives fast result.

REFERENCES

[1] Archna sigh, Rakesh Kumar ,Heart Disease Prediction Using Machine Learning Algorithms, Institute of Electrical and Electronics Engineers, 2020.

[2] Apurb Rajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli. Heart Disease Prediction using Machine Learning, International Journal Of Engineering Research & Technology (IJRT), 2020.

[3] Abhijeet jagtap, Priya malewadkar, Omkar Baswat, harshali Rambade, Heart Disease Prediction using Machine learning, International Journal Of Research in Engineering Science and Management, 2019.

[4] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja Heart disease prediction using machine learning techniques: A survey, International Journal of Engineering & Technology, 2018.

[5] N Rajesh,Maneesha T,Shaik Hafeez,Hari Krishna, Prediction of Heart Disease Using Machine Learning Algorithms in Science, International Academic Journals, 2018.

[6] Mangesh Limbitote, Dnyaneshwari Mahajan, Kedar Damkondwar, Pushkar Patil, A Survey on Prediction Techniques of Heart Disease using Machine Learning, International Journal Of Engineering Research & Technology, 2020.

[7] Rajesh N,T Maneesha, Shaikh Hfeez,hari krishna, Prediction Of Heart Disease Using Machine Learning Algorithms, International journal of engineering & Technology, 2018.

[8] P.Jayapratha, S.G.Mirtthula, D. K. Heart disease prediction using ML techniques, International Journal of Advanced Science and Technology, 2020.

[9] Aditi Gavhane, Gouthami Kokkula, Isha Panday and Kailash Devadkar, Prediction of Heart Disease using Machine Learning, Proceedings of the 2 nd International conference on Electronics Communication and Aerospace Technology(ICECA), 2018.

[10] Himanshu Sharma and M A Rizvi, Prediction of Heart Disease using Machine Learning Algorithms: A Survey, International Journal on Recent and Innovation Trends in Computing and Communication, 2017.

[11] Senthil Kumar Mohan, Chandrasegar Thirumalai and Gautam Srivastva, Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, IEEE Access, 2019.

[12] Sonam Nikhar, A.M. Karandikar, Prediction of Heart Disease Using Machine Learning Algorithms, International Journal of Advanced Engineering, Management and Science (IJAEMS), 2016.

[13] https://archive.ics.uci.edu/ml/datasets/Heart+Disease.