IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HEART ATTACK PREDICTION USING MACHINE LEARNING

[1]ALEENA PRAKASH, [2]AMBILY JACOB

[1]Msc Scholar, [2]Assistant Professor

[1,2]Department of Computer Science,

[1,2]St.Joseph's College (Autonomous), Irinjalakuda, Thrissur, India

Abstract: In our busy day to day life going after money and living status we forget about our health and many of us are least bothered about ourselves. Health issues are increasing tremendously these days. Many people start going for complete medical check-ups during their late 40's or 50's of age. But our lifestyle has a great impact on health causing heart attack and other health problems. Early detection of heart attack can prevent death rates. As most of our health care industries are aiming at diagnosing these diseases in early stages, by using Data Mining technique, we can detect disease at an early stage and help to cure it completely at affordable cost. By using a proper database, a trained ML model is created which can diagnose a normal person and generate an output report which shows whether the person is affected by heart attack or not.

Data mining is an advanced technology, which is the process of discovering actionable information from a large set of data, which is used to analyse large volumes of data and extract patterns that can be converted to useful knowledge. From the medical profiles fourteen attributes are extracted such as age, sex, blood pressure and blood sugar etc. can predict the patient getting heart disease. These attributes are fed into a ***Random forest classifier*** for heart attack prediction; it can provide as reliable performance as that achieved in diagnosing heart attack. We are also using ***Support vector machine*** (SVM) algorithm to compare the accuracies between random forest classifier and SVM. This will help the medical industries to offer better diagnosis and treatment of the patient to attain a good quality of services. The main advantages of this paper are: early detection of heart attack and its diagnosis correctly on time and providing treatment with affordable cost.

Keywords- Random Forest classifier, Support Vector Machine, Heart attack prediction.

I. INTRODUCTION

We are in the 20th century, we don't have time even to take care of ourselves. Even if we are aware that heart attack are the main cause of death from the last few decades, we are least bothered about checking the status of our heart. We all know the treatment cost of heart disease is very high and is not affordable by everyone. Thus people find it very much difficult and reluctant to go for treatment in the final stages. Unlike other diseases heart disease does not show any visible symptoms. Therefore, checking the heart status in its early stages is a very important task and here comes the need of this system. Data Mining is used in such areas where we have a large amount of raw data to generate relationships in large medical datasets. With

the help of Data Mining techniques, we can detect disease at an early stage and help to cure it completely at an affordable cost. Create a ML model which can diagnose a normal patient and generate an output which shows the person is affected by heart attack or not. This can provide a reliable performance as such achieved in diagnosing heart attack. Here we make use of an algorithm (Random Forest) to train the extracted features.

II. RELEATED WORK

Lot of work has been carried out to predict heart attack using UCI Machine Learning dataset. Different levels of accuracy have been attained using various data mining techniques which are explained as follows.

- Avinash Golande studies various different ML algorithms that can be used for classification of heart disease. Research was carried out to study Decision Tree, KNN and K-Means algorithms that can be used for classification and their accuracy were compared. This research concludes that accuracy obtained by Decision Tree was highest further it was inferred that it can be made efficient by combination of different techniques and parameter tuning.
- T. Nagamani have proposed a system which deployed data mining techniques along with the Map Reduce algorithm. The accuracy obtained according to this paper for the 45 instances of testing set, was greater than the accuracy obtained using conventional fuzzy artificial neural network. Here, the accuracy of algorithm used was improved due to use of dynamic schema and linear scaling.
- Fahd Saleh Alotaibi has designed a ML model comparing five different algorithms. Rapid Miner tool was used which resulted in higher accuracy compared to Matlab and Weka tool. In this research the accuracy of Decision Tree, Logistic Regression, Random forest, Naive Bayes and SVM classification algorithms were compared. Decision tree algorithm had the highest accuracy.
- Anjan Nikhil Repaka proposed a system that uses NB (Naïve Bayesian) techniques for classification of dataset and AES (Advanced Encryption Standard) algorithm for secure data transfer for prediction of disease.
- Theresa Princy. R executed a survey including different classification algorithm used for predicting heart disease. The classification techniques used were Naive Bayes, KNN (K-Nearest Neighbour), Decision tree, Neural network and accuracy of the classifiers was analysed for different number of attributes.
- Nagaraj M Lutimath has performed the heart disease prediction using Naive bayes classification and SVM (Support Vector Machine). The performance measures used in analysis are Mean Absolute Error, Sum of Squared Error and Root Mean Squared Error, it is established that SVM was emerged as superior algorithm in terms of accuracy over Naive Bayes.

The main idea behind the proposed system after reviewing the above papers was to create a heart attack prediction system based on the inputs as shown in Table 1. We analysed the classification algorithms namely Random Forest and Support Vector Machine based on their Accuracy and identified the best classification algorithm which can be used in the heart attack prediction.

III. PROPOSED SYSTEM

The standardised medical data collected from the patients having like 14 attributes like, Sex, Age, Blood pressure, Blood sugar etc. can predict the nature of the patient getting heart disease. These attributes are fed into a machine learning algorithm (Random Forest). It is really a difficult task to gather all these 14 attributes of the patient. In some cases we cannot get all the details. In such cases, we need to skip that record and gather details of the patient only if there are all these 14 attributes present. Out of the whole data collected, we are taking 4/5 of the data for training the model and the rest 1/5 of the data is used for testing purposes

to confirm its accuracy. It is further gone through the processing step where we can classify the details in its 14 attributes and have an extra 1 group where we have the Label where the status of that patient is given. This helps to get the picture more clear .The training set data is moved through a Machine Learning Algorithm after the processing step is completed to get a 'Trained Model'. This Trained Model is now able to analyse the patient by providing the 14 attributes.

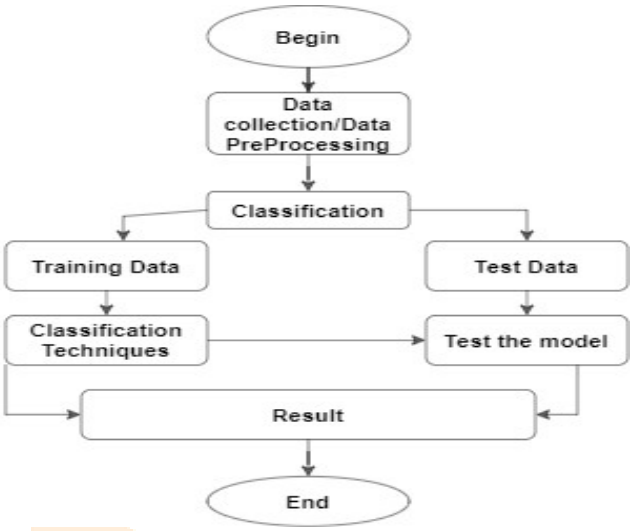

Figure 1. Generic Model Predicting Heart attack

A. Data Collection and Pre-processing

The dataset used was the Heart attack Dataset which is a combination of 4 different database, but only the UCI Cleveland dataset was used. This database consists of 14 features. We are using processed UCI Cleveland dataset available in the UCI repository for our analysis. The complete description of the 14 attributes used in the proposed work is mentioned in Table 1 shown below.

Table 1. Features selected from dataset

Sl. No.	Attribute Description
1.	*Age*- represent the age of a person
2.	*Sex*- describe the gender of person (0Feamle, 1-Male)
3.	*CP*- represents the severity of chest pain patient is suffering.
4.	*TrestBPS*-It represents the patient's BP.
5.	*Chol*-It shows the cholesterol level of the patient.
6.	*FBS*-It represent the fasting blood sugar in the patient.
7.	*Resting ECG*-It shows the result of ECG
8.	*Thalach*- shows the max heart beat of patient

9.	*Exang*- used to identify if there is an exercise induced angina. If yes=1 or else no=0
10.	*OldPeak*- describes patient's depression level.
11.	*Slope*- describes patient condition during peak exercise. It is divided into 3 segments(Unsloping, Flat, Down sloping)
12.	*CA*- Result of fluoroscopy.
13.	*Thal*- test required for patient suffering from pain in chest or difficulty in breathing. There are 4 kinds of values which represent Thallium test.
14.	*Target*-It is the final column of the dataset. It is class or label Colum. It represents the number of classes in dataset. This dataset has binary classification i.e. two classes (0,1).In class "0" represent there is less possibility of heart attack whereas "1" represent high chances of heart attack. The value "0" Or "1" depends on other 13 attribute.

B. Classification

The attributes mentioned in Table 1 are provided as input to the ML algorithms such as Random Forest and Support vector machine. The input dataset is split into 80% of the training dataset and the remaining 20% into the test dataset. Training dataset is the dataset which is used to train a model. Testing dataset is used to check the performance of the trained model. For each of the algorithms the performance is computed and analysed based on accuracy. The algorithms explored in this paper are listed as below.

i. Random Forest

Random Forest algorithm is used for classification as well as regression. It creates a tree for the data and makes prediction based on that. Random Forest algorithm is used on large datasets and produce the output. The generated samples from the decision tree can be saved so that it can be used on other data. There are two stages in random forest, firstly create a random forest then it will make a prediction using a random forest classifier created in the first stage.

ii. Support Vector Machine

Support Vector Machine or SVM is one of the popular Supervised Machine Learning algorithm, which is used for classification as well as Regression problems. The goal of the SVM algorithm is to create the best line or decision boundary that divides the n-dimensional space into classes so that we can easily put the new data in the correct class. This best decision boundary is called a hyper plane. SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors and hence algorithm is termed as Support Vector Machine.

C. Software and Hardware Requirements

Hardware Requirements:

The minimum configuration for hardware is given below:

- Intel® Pentium® or higher processor.
- 65 MB RAM or higher

Software Requirements:

- Anaconda-Spyder.
- Python
- Django

IV. RESULT AND ANALYSIS

The results obtained by applying Random Forest and Support Vector Machine are shown in this section. The metrics used to carry out performance analysis of the algorithm are Accuracy score.

Accuracy = (correct_count/total)*100

- Correct_count: Number of true predictions.
- Total: Total number of data.

Table ii. Analysis of machine learning Algorithm

Algorithm	Accuracy
Random Forest	91.707%
Support Vector Machine	67.804%

```
svm accuracy: 67.8048780487805
Radomforest accuracy: 91.70731707317074
```

Figure 2. Accuracies of both algorithms in SPYDER-IDE

V. CONCLUSION

With the increasing number of deaths due to heart attacks, it has become mandatory to develop a system to predict heart attack effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of heart attack. This study compares the accuracy score of Random Forest and Support Vector Machine algorithms for predicting heart attack using UCI machine learning repository dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 91.707% for prediction of heart attack. In future the work can be enhanced by developing a web application based on the Random Forest algorithm as well as using a larger dataset as

compared to the one used in this analysis which will help to provide better results and help health professionals in predicting the heart attack effectively and efficiently.

VI. REFERENCES

[1] Avinash Golande, Pavan Kumar T, "Heart Disease Prediction Using Effective Machine Learning Techniques", *International Journal of Recent Technology and Engineering*, Vol 8, pp.944-950,2019.

[2] T.Nagamani, S.Logeswari, B.Gomathy," Heart Disease Prediction using Data Mining with Mapreduce Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Fahd Saleh Alotaibi," Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, "Design And Implementation Heart Disease Prediction Using Naives Bayesian", *International Conference on Trends in Electronics and Information(ICOEI 2019).*

[5] Theresa Princy R,J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', *International Conference on Circuit Power and Computing Technologies,Bangalore*,2016.

[6] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,'Prediction Of Heart Disease using Machine Learning', *International journal Of Recent Technology and Engineering,8,(2S10), pp* 474-477, 2019.

[7] UCI, ―Heart Disease Data Set.[Online]. Available (Accessed on May 1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.

[8] Sayali Ambekar, Rashmi Phalnikar,"Disease Risk Prediction by Using Convolutional Neural Network",2018 Fourth International Conference on Computing Communication Control and Automation.

[9] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, ―Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field,‖ in Machine Learning Paradigms, 2019, pp. 71–99.

[10] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018

[11] Fajr Ibrahem Alarsan., and Mamoon Younes 'Analysis and classification of heart diseases using heartbeat features and machine learning algorithms',Journal Of Big Data,2019;6:81.

[12] Internet source [Online].Available (Accessed on May 1 2020): http://acadpubl.eu/ap