WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Early Heart Disease Prediction Using Machine Learning

S. Anusuya[1], Mr. S.Barath[2]

[1]Master of Computer Application, krishnasamy college of engineering &Technology,Cuddalore
[2]M.C.A., M.Phil., Assitant Professor, Master of Computer Application, krishnasamy college of engineering &Technology,Cuddalore

ABSTRACT

Heart attack disease is one of the leading causes of the death worldwide. In today"s common modern life, deaths due to the heart disease had become one of major issues, that roughly one person lost his or her life per minute due to heart illness. Predicting the occurrence of disease at early stages is a major challenge nowadays. Machine learning when implemented in health care is capable of early and accurate detection of disease. In this work, the arising situations of Heart Disease illness are calculated. Datasets used have attributes of medical parameters. The datasets are been processed in python using ML Algorithm i.e., Decision Tree Classifier. This technique uses the past old patient records for getting prediction of new one at early stages preventing the loss of lives. In this work, reliable heart disease prediction system is implemented using strong Machine Learning algorithm which is the Random Forest algorithm Which read patient record data set in the form of CSV file. After accessing dataset the operation is performed and effective heart attack level is produced. Advantages of proposed system are High performance and accuracy rate and it is very flexible and high rates of success are achieved.

Keyword: Heart attack, Machine learning, Detection of disease.

1. INTRODUCTION

Heart disease effects the functioning of the heart. World Health Organization had made a survey and made a conclusion that 10 million people are affected with heart disease and lost their lives. The problem that the Healthcare industry faces in today"s life is early prediction of disease after a person is affected. Records or data of medical history is very large and the data in real world might be incomplete and inconsistent. In past predicting the disease effectively and treatment to patients might not be possible for every patient at early stages under these circumstances

Many scientists tried to build a model which is capable of predicting the heart disease in the early stage, but they are not able to build a perfect model. Every proposed system has disadvantages in its own .So to overcome this we are implementing Random forest algorithm in order to achieve accurate results in less time. Machine learning is given a major priority in modern life in many applications and in healthcare sector. Prediction is one of area where machine learning plays a vital role, our topic is to predict heart disease by processing patient"s dataset and a data of patients i.e., user of whom we need to predict the chances of occurrence of a heart disease.

2. LITERATURE SURVEY

Kaan Uyar [1]The World Health Organization (WHO) estimated one third of all global deaths reason as cardiovascular diseases in 2015. Some computational techniques were proposed for investigation of heart diseases. This study proposes a genetic algorithm (GA) based trained recurrent fuzzy neural networks (RFNN) to diagnosis of heart diseases. The University of California Irvine (UCI) Cleveland heart disease dataset is used in this study. Out of total 297 instances of patient data, 252 are used for training and 45 of them are chosen to be the testing. The results showed that 97.78% accuracy was obtained from testing set. In addition to the accuracy, root mean square error, the probability of the misclassification error, specificity, sensitivity, precision and F-score are calculated. The results were found to be satisfying based on comparison.

[2]Ashish Chhabbi,The healthcare industry collects large amounts of Healthcare data, but unfortunately not all the data are mined which is required for discovering hidden patterns and effective decision making. We propose efficient genetic algorithm with the back propagation technique approach for heart disease prediction. This paper has analyzed prediction systems for Heart disease using more number of input attributes. The System uses medical terms such as Gender, blood pressure, cholesterol like13 attributes to predict the likelihood of patient getting a Heart disease.

[3]Theresa Princy ,Nowadays, health disease are increasing day by day due to life style, hereditary. Especially, heart disease has become more common these days, i.e. life of people is at risk. Each individual has different values for Blood pressure, cholesterol and pulse rate. The patient risk level is classified using datamining classification techniques such as Naïve Bayes, KNN, Decision Tree Algorithm, Neural Network. etc., Accuracy of the risk level is high when using more number of attributes.

PROPOSED SYSTEM

To overcome this we are implementing Decision Tree Classifier in order to achieve accurate results in less time. Machine learning is given a major priority in modern life in many applications and in healthcare sector. Prediction is one of area where machine learning plays a vital role, our topic is to predict heart disease by processing patient's dataset and a data of patients i.e., user of whom we need to predict the chances of occurrence of a heart disease. Our aim is to build an application of heart disease prediction system using robust Machine Learning algorithm which is Decision Tree Classifier. A CSV file is given as input. After the successful completion of operation the result is predicted and displayed.

ADVANTAGES OF PROPOSED SYSTEM

- The advantages of proposed model are High performance and accuracy rate.
- It is very flexible and high rates of success are achieved.
- The application when implemented using Decision Tree Classifier has more accuracy rate when compared to other algorithms. In this system, we achieved around 96%.

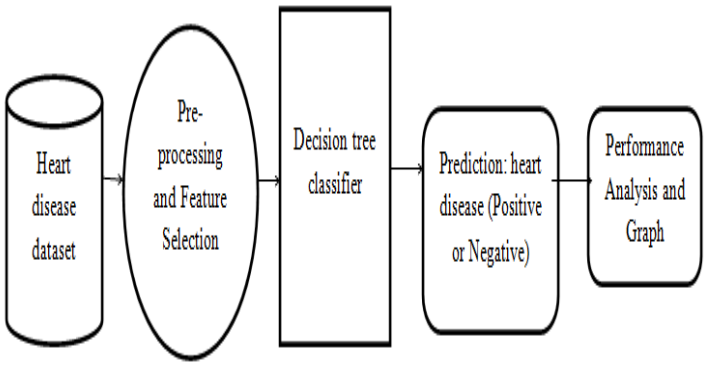


Figure: System Architecture

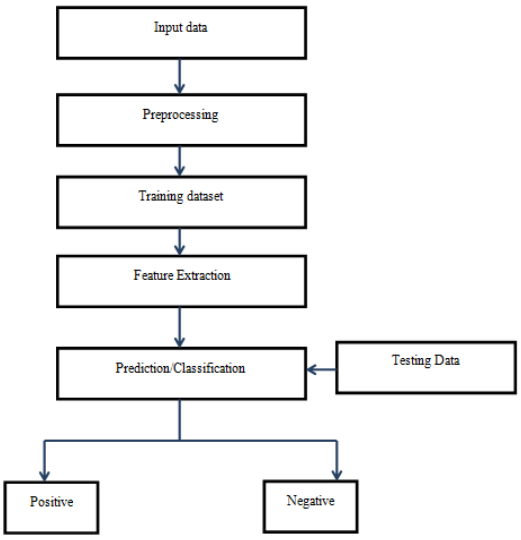


Figure: Data Flow Diagram

DECISION TREE CLASSIFICATION ALGORITHM

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

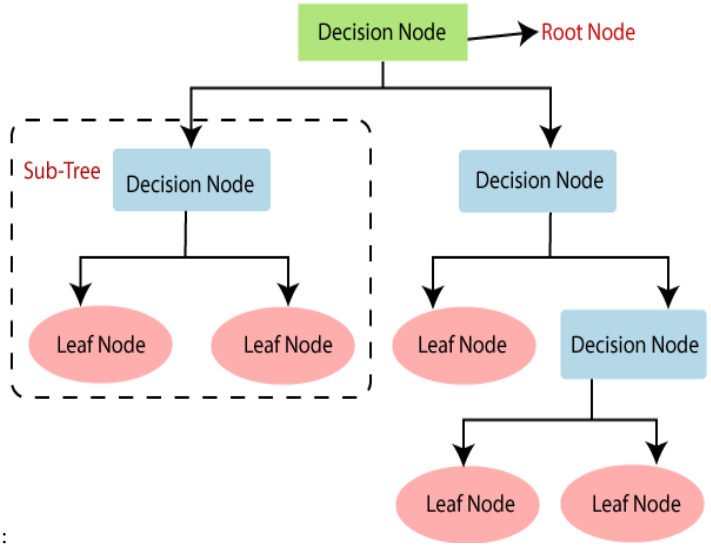
In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset.

It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use the CART algorithm**,** which stands for Classification and Regression Tree algorithm**.**A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.Below diagram explains the general structure of a decision tree



:

ANALYZE AND PREDICTION:

In the actual dataset, we chose only 7 features :

1. **Age**: displays the age of the individual.
2. **Sex**: displays the gender of the individual using the following format :1 = male0 = female
3. **Chest-pain type(cp)**: displays the type of chest-pain experienced by the individual using the following format :1 = typical angina2 = atypical angina3 = non — anginal pain4 = asymptotic
4. **Resting Blood Pressure(trestbps)**: displays the resting blood pressure value of an individual in mmHg (unit)
5. **Serum Cholestrol(chol)**: displays the serum cholesterol in mg/dl (unit)
6. **Fasting Blood Sugar(fbs)**: compares the fasting blood sugar value of an individual with 120mg/dl.If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)
7. **Resting ECG (restecg)**: displays resting electrocardiographic results0 = normal 1 = having ST-T wave abnormality2 = left ventricular hyperthrophy
8. **Max heart rate achieved** : displays the max heart rate achieved by an individual.
9. **Exercise induced angina** :1 = yes0 = no
10. **ST depression induced by exercise relative to rest**: displays the value which is an integer or float.

11. Peak exercise ST segment : 1 = upsloping 2 = flat 3 = downsloping

Accuracy on test set:

We got an accuracy of 96.7% on test set.

Saving the Trained Model:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or . pkl file using a library like pickle .Make sure you have pickle installed in your environment.Next, let's import the module and dump the model into.pkl file

CONCLUSION

Cardiovascular and chronic respiratory diseases are some of the issues faced by health care. Due to the time-independent lifestyle, both diseases affect mortality across the globe. Heart attack occurs without any apparent symptoms. The effective identification of the machine-learning model concerning data is a challenging task. In this paper, we have discussed the details of predicted models for the forecasting of vital sign values that are ultimately helpful for the realization of a machine-learning-based system for the prediction of chronic diseases. This study also aims to facilitate caregivers and medical experts to provide in-time medical assistance to the patients to reduce the fatality rate due to cardiovascular and chronic respiratory complications in indoor patients, particularly after surgical procedures. It is necessary to assess the appropriateness of the prediction model according to the nature of the data.

FUTURE ENHANCEMENTS

DL-based prediction system developed in this study performs well in early diagnosis of CVDs and can be accessed via Internet. This study offers promising results suggesting potential use of DL-based heart disease prediction system as a screening tool to diagnose heart diseases in primary healthcare centres in India, which would otherwise get undetected.

REFERENCES

[1] Ashish Chhabbi, Lakhan Ahuja, Sahil Ahir and Y. K. Sharma, "Heart Disease Prediction Using Data Mining Techniques", © IJRAT Special Issue National Conference "NCPC-2016", pp. 104-106, 19 March 2016.

[2] Ammar Asjad Raja, Irfan-ul-Haq , Madiha Guftar Tamim Ahmed Khan "Intelligence syncope Disease Prediction Framework using DM-techniques" FTC 2016 –Future Technologies Conference 2016.

[3] Ayon Dey, Jyoti Singh, N. Singh "Analysis of supervised machine learning algorithms for heart disease prediction" International journal of computer applications, published in 2006.

[3] JD Berry, DM Lloyd-Jones, DB Garside et al., "Framingham risk score and prediction of coronary heart disease death in young men", Am Heart J., vol. 154, no. 1, pp. 80-6, 2007.

[4] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," Machine Learning,(1997).

[5] M.A. Jabbar, B.L.Deekshatulu, and Priti Chandra, "Intelligent heart disease prediction system using random forest and evolutionary approach", Journal of Network and Innovative Computing, Vol. 4, pp.174-184, 2016.

[6] Kaan Uyar and Ahmet İlhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks" in B.V ICTASC, Elsevier,pp

[7] Kaur H Beant and Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", IJRITCC, vol. 2, no. 10, pp. 3003-08, 2014.

[8] M.M. Kirmani and S.I. Ansarullah, "Prediction of heart disease using decision tree a data mining technique", IJCSN Int. J. Comput. Sci. Netw., vol. 5, no. 6, pp. 885-892, 2016.

[9] Tahira Mahboob, Rida Irfan and Bazelah Ghaffar et al."Evaluating ensemble prediction of coronary heart disease using receiver operating characteristics"

[10] Theresa Princy and R. J. Thomas, "Human Heart Disease Prediction System using Data Mining Techniques", IEEE ICCPCT, 2016.