

Prediction of Heart Disease Using Machine Learning

Poojary Akshay Sudhakar¹, Pratheekraj M Gatti², Mamatha K³

^{1,2}Dept of CSE

³Assistant Professor, Department of CSE

^{1,2,3}Srinivas Institute of Technology

Abstract- Heart is one of the important organs in human body, so its disease is one of the most important causes in today's world. So, the Prediction of Heart Disease using 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 heart disease in most cases depends on a complex combination of clinical and pathological data. 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.

Keywords- Prediction; heart disease; machine learning

I. INTRODUCTION

There is no dearth of records regarding medical symptoms of patients suffering heart strokes. For instance: 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 relevant data pertaining all elements related to our field of study, train 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. Thus, it is of utmost importance that the factors required in the input are most accurate and easily available.

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. 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 [1]. 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 organization 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 [2]. 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 [3]. In the research done by S.Prabhavathi et al, proposed Decision tree based Neural Fuzzy System (DNFS) technique to analyses 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 [4]. 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 [5]. In the search done by K.Gomathi et al, 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. [6]. In this fast moving world people want to live a very luxurious life so they work like a machine in order to earn lot of money and live a comfortable life therefore in this race they forget to take care of themselves, because of this their food habits change their entire lifestyle change, in this type of lifestyle they are more tensed they have blood pressure, sugar at a very young age and they don't give enough rest for themselves and eat what they get and they even don't bother about the quality of the food if sick they go for their own medication as a result of all these small negligence it leads to a major threat that is the heart disease.[7]

III. PROBLEM STATEMENT

The task here is to predict the heart disease of the patient. There is a lot of data available within the healthcare systems on the internet. There is a lack of effective analysis tools to discover hidden relationships and patterns in data. Therefore, a system should be developed where an automated system in medical diagnosis would enhance medical efficiency and reduce costs. This system intends to predict the occurrence of a disease based on data gathered from Cleveland foundation medical research particularly in Heart Disease.

IV. 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 difficult to do without doctor's involvement.

V. METHODOLOGY

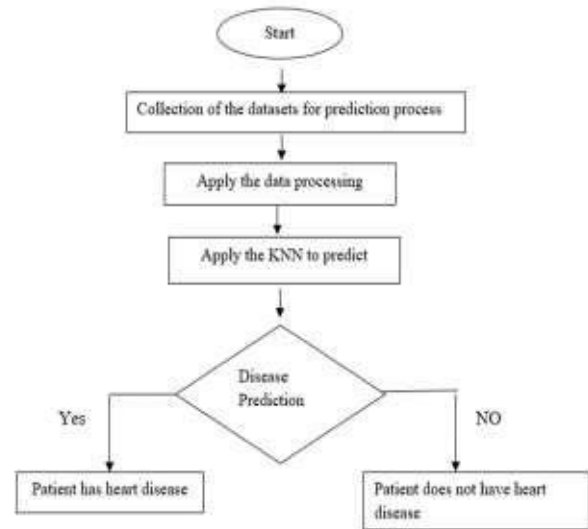


Figure 1: Flowchart for prediction of heart disease using machine learning

A system flowchart symbolically shows the data flows throughout a system and event controlling decisions are made. Initially the collection of datasets is input, and the pre-processing of the input takes place followed by KNN algorithm. Later the prediction of heart disease is classified according to their extracted feature. Finally, the patient with heart disease is detected.

The classifier used in this system for the prediction of Heart Disease using machine learning are KNN ((K- Nearest Neighbor). SVM (Support Vector Machine), Decision Tree Classifier, LGBM Classifier (LightGBM). The KNN Classifier is set to be the best classifier in this system, since it gives better accuracy results based on the algorithm been processed. The KNN algorithm is based on the standard deviation where the values is been set different on each attributes. It usually sets as arrange for values where it has minimum or maximum. So, this algorithm leads to be the highest accuracy level than other three algorithm. Therefore, the user can able to predict the heart disease on such terminologies.

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

VI. RESULT

The output of the system will give a prediction result if the person has a heart disease, in terms of Yes or No. 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

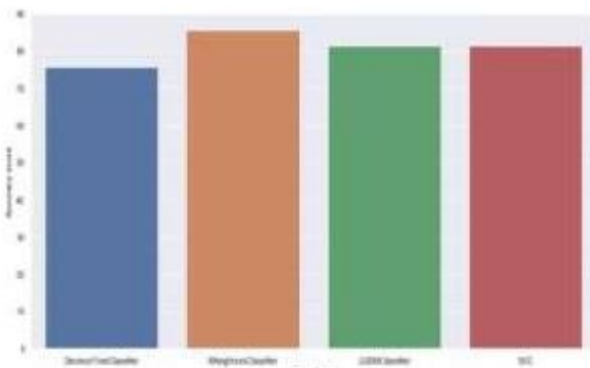


Figure 2: Training loss and accuracy on dataset.

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

VII. CONCLUSION

The proposed method for detecting predicting heart disease can be predicted with excellent accuracy and resilience. 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.

More and more training datasets in various settings and circumstances will be collected in the future. Thus, produces an enhanced performance level with accuracy level above 88% through the prediction model for heart disease. Finally, the Prediction of Heart Disease can be detected.

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/HEARTORG/Conditions/Heart-Health/Screenings_UCM_428687_Article.jsp#. WnsO AeeYPIV
- [6] Scikit Learn, ‘MLPC Lassifier’, [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