A hybrid system for the prediction of heart disease using machine learning

Prof. B. B. Kotame, Aniket Deogaonkar, Amruta Ghalme, Vaishnavi Gunjal, Mrunali Jagtap, Mrunali Garud

Assistant professor, BE Computer Student
Computer Engineering, Sanjivani College of Engineering, Kopargaon, India

ABSTRACT: Heart disease is one of the most significant causes of humility in the world today. Prediction of cardiovascular disease is a challenge in the area of clinical data analysis. Machine learning (ML) shows to effective in helping in making decisions and predictions from the large quantity of data produced by the health care industry. This proposed system used ML techniques to developments in different areas of the Internet of Things (IoT). This proposed system used a novel method that aims at finding important features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of characteristics and several known classification techniques. This proposed system produce an improved performance level with an accuracy level of 88.7 through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

Keyword:- Machine learning, Decision tree, K-nearest neighbor, Naive bayes.

I. INTRODUCTION

Heart disease is one of the most significant causes of humility in the world today. Prediction of cardiovascular disease is a most challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in helping in making decisions and predictions from the large quantity of data produced by the health care industry. Different techniques in data mining and neural networks have been working to find out the severity of heart disease among humans. The seriousness of the disease is classified based on various methods like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), and Naive Bayes (NB). It is hard to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors.

II. LITERATURE SURVEY

1.   A.sheik Abdullah, and R.R.Rajalaxmi[1] to classify the healthy people and people with heart disease, non-invasive based methods such as machine learning are reliable and efficient. This proposed system to developed a machine-learning based diagnosis system for heart disease prediction by using heart disease dataset. The proposed machine-learning-based decision support system will assist the doctors to diagnosis heart patients efficiently.

2. Tanu Rani and Mr.Narender kumar[2] to provides the efficient machine learning and deep learning algorithms to analyze and predict the disease as soon as possible. The dataset is collected from Kaggle. Four algorithms namely Logistic Regression, Random Forests Classifier Algorithm, Neural network, KNN are used. Of these four algorithms Logistic Regression holds the best accuracy rate in predicting the heart disease. 1

3. Madhavi Veeranki and Jayanag Bayana [3] to investigates the state of the art of various clinical decision support systems for heart disease prediction, proposed by various researchers using data mining and machine learning techniques. Classification algorithms such as the Naive Bayes (NB), Decision Tree (DT), and Artificial Neural Network (ANN) have been widely employed to predict heart diseases, where various accuracies were obtained.

4. Resul Das,Ibrahim Turkoglu and Abdulkadir Sengur[4] a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. This proposed system produce an enhanced performance level with an accuracy through the prediction model for heart disease with the proposed method hybrid random forest with a linear model.

5. Abhishek Taneja[5] to develop a cost effective treatment using data mining technologies for facilitating data base decision support system. This proposed system using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question.

III. METHODOLOGY

Data Pre-processing:

Heart disease data is pre-processed after collection of various record . The dataset contains a total of 303 patient records, where 6 record are with some missing values. Those 6 record have been removed from the dataset and the remaining 297 patient record are used a per-processing.

Feature Selection and Reduction:

Control Flow Diagram:

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain important clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease.

Classification Modeling:

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree(DT) features .Then ,the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.
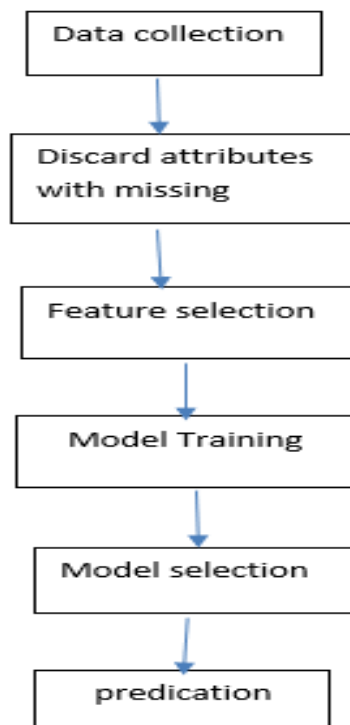

Fig . Flow of System

Accuracies of each Algorithm:

Algorithm	Training accuracy	Cross Validation accuracy	Testing accuracy
Decision Tree Classifier	70%	30%	30%

IV. CONCLUSION

Identifying the transmission of raw cardiac details safety data can continue to preserve individual life in the long run, including early diagnosis of heart disease anomalies. In this propose system, machine learning techniques were used to process raw data and provide a modern and novel discernment of heart disease. Predicting cardiac failure is difficult in the medical profession and is very important. However, once the illness is diagnosed in the early stages, the death rate may be significantly reduced and prevention steps are implemented as quickly as possible.

REFERENCES

[1]. A. Sheik Abdullah, and R.R.Rajalaxmi,"A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier", 2012.

[2]. Tanu Rani, and Mr. Narender Kumar,"A Survey: Hybrid Intelligent Modeling Technique for Data Classification", May 2017.

[3]. Madhavi Veeranki, and Jayanag Bayana "Effective Cardiovascular Disease Prediction using Hybrid Machine Learning Techniques", April 2020.

[4]. Resul Das, Ibrahim Turkoglu and Abdulkadir Sengur "Effective diagnosis of heart disease through neural networks ensembles", May 2009.

[5]. Abhishek Taneja"Heart Disease Prediction System Using Data Mining Techniques", November 2013.

[6]. N. Suganthi, R. Abinavi, S. Deva Dharshini, V. Haritha "Effective Heart Disease Prediction using Distinct Machine Learning Techniques", Mar 2020.

[7]. Xiao Liu, Xiaoli Wang, Qiang Su, Mo Zhang, Yanhong Zhu, Qiugen Wang,and Qian Wang "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method", January 2017.

[8]. Kaan Uyar, Ahmet ˙Ilhan "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks", August 2017.

[9]. G. Jignesh Chowdary , Suganya. G, Premalatha. M "EFFECTIVE PREDICTION OF CARDIOVASCULAR DISEASE USING CLUSTER OF MACHINE LEARNING ALGORITHMS", June 2020.