HEART DISEASE PREDICTION USING MACHINE LEARNING

Naga vara prasad Mella[1], Rayaprolu Anasuya [2]

[1]Assistant Professor MCA, DEPT, Dantuluri Narayana Raju College , Bhimavaram, Andharapradesh
[2]PG Student of MCA, Dantuluri Narayana Raju College , Bhimavaram, Andharapradesh

Abstract Machine Learning is used across many ranges around the world. The healthcare industry is no exclusion. Machine Learning can play an essential role in predicting presence/absence of locomotors disorders, heart diseases and more. Such information, if predicted well in advance, can provide important intuitions to doctors who can then adapt their diagnosis and dealing per patient basis.We work on predicting possible Heart Diseases in people using Machine Learning algorithms. In this project we perform the comparative analysis of classifiers like decision tree, NaïveBayes, Logistic Regression, SVM and Random Forest and we propose an ensemble classifier which perform hybrid classification by taking strong and weak classifiers since it can have multiple number of samples for training and validating the data so we perform the analysis of existing classifier and proposed classifier like Ada-boost and XG-boost which can give the better accuracy and predictive analysis.

Index Term:-Machine Learning,NavieBayes,SVM,Logistic Regression

I Introduction

According to the World Health Organization, every year 12 million death so occur worldwide due to heart disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pin point the most influential factors of heart disease as well as accurately predict the overall risk. Heart Diseases even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aim to predict future heart disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm.Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources,classifying them under suitable headings &finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

2 Literature survey

With growing development in the field of medical science along side machine learning various experiments and researches has been carried out in these recent years releasing the relevant significant papers.

Purushottam,et,ai,proposedsystemforheartdiseaseprediction System"usinghillclimbinganddecisiontreealgo

rithms.TheyusedClevelanddatasetandpreprocessingo fdataisperformedbeforeusingclassificationalgorithms .TheKnowledgeExtractionisdonebasedonEvolutionar yLearning(KEEL),an open-source data mining tool that fills the missing values in the dataset A decision tree follow stop-down order.For each actual node selected by hill-climbing algorithm a node is selected by a test at each level.The parameters and their values used are confidence.Its minimum confidence value is 0.25.The accuracy of the system is about 86.7%.

Santhana Krishnan. J ,et ,al proposed a paper "Prediction of Heart Disease Using Machine Learning Algorithms" using decision tree and Naive Bayes algorithm for prediction of heart disease.In decision tree algorithm the tree is built using certain conditions which gives True or False decisions. The algorithms like SVM ,KNN are results based on vertical or horizontal split conditions depends on dependent variables. structure having root node ,leaves and branches base on the decision ma de in each of tree Decision tree also help in the understating the importance of the attributes in the dataset. They have al so used Cleveland dataset.Dataset splits in 70%training and 30%testing by using some methods. This algorithm gives 91%accuracy.The second algorithm is NaiveBayes,which is used for classification.It can handle complicated,nonlinear,dependent data so it is found suitable for heartdisease dataset as this dataset is also complicated,dependent and on linear in nature. This algorithm gives an 87%accuracy.

Sonam Nikhar et al proposed paper " Prediction of Heart Disease Using Machine Learning Algorithms" their research gives point to point explanation of Naïve Bayes and decisiontree classifier that are used especially in the prediction of HeartDisease. Some analysis has been led to think about the execution of prescient data mining strategy on the same dataset,and the result decided that DecisionTree has highest accuracy than Bayesian classifier.

Aditi Gavhane et al proposed a paper "Prediction of Heart Disease Using Machine Learning",in which training and testing of dataset is performed by using neural network algorithm multi-layer perception. In this algorithm there will be one inputlayer and one output layer and one or more layers are hidden layers between these twoinput and output layers. Through hidden layers each input node is connected to outputlayer. This connection is assigned with some random weights. The other input is calledbias which is assigned with weight based on requirement the connection between then no descent bee fed forwarded or feedback.

Vanish Golandeetal, proposed" Heart Disease Prediction Using Effective Machine Learning Techniques" in which few data mining techniques are used that support the doctors to differentiate the heart disease. Usually utilized methodologies are k-nearestneighbour, Decisiontree and NaïveBayes.Other unique characterization-based strategies utilized are packing calculation ,Part thickness,consecutive negligible streamlining and neural systems, straight Kernel self arranging guide and SVM (BolsterVector Machine).

LakshmanaRao etal,proposed" Machine LearningTechniques for HeartDisease Prediction" in which the contributing elements for heart disease are more. So, it is difficult to distinguish heart disease.To find the seriousness of the heart disease among people different neural systems and data mining techniques are used.

3. Implementation Study

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heartdisease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various

sources, classifying them under suitable headings &finally analysing to extract the desired data we can conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says “Prevention is better than cure”, early prediction & its control can be helpful to prevent& decrease the death rates due to heart disease.

3.1proposed methodology

The proposed system acts as a decision support system and will prove to be an aid for the physicians with the diagnosis. The working of the system starts with the collection of data ,selecting of important attributes, data preprocessing , training and testing of data, finally the model is trained using different classifier the algorithm with highest accuracy is used for prediction.

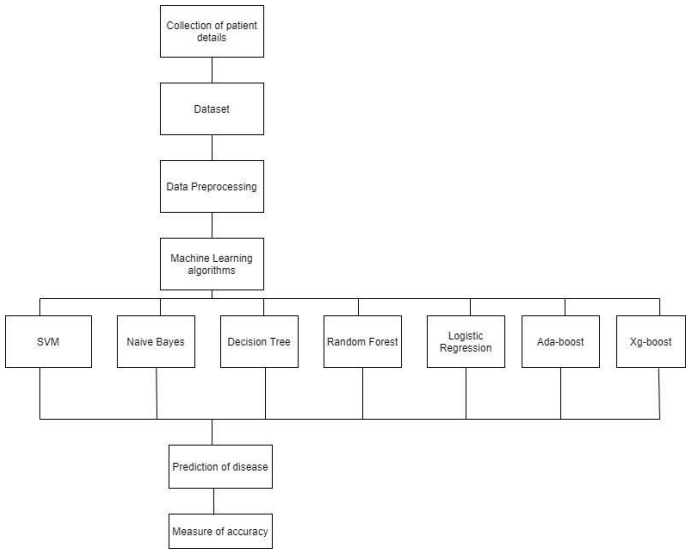


Fig 1: proposed model

3.2 Methodology

3.2 1Collection of dataset:

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system.

3.2.2Pre-processingofData:

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, an missing values of thdataset.Datapre processinghastheactivitieslikeimportingdatasets,split tingdatasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model

3.2.3 Selectionof attributes:

Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, exang, etc are selected for the prediction. The Correlation matrix is used for attribute selection for this model

3.2.4Balancing of Data:

Imbalanced datasets can be balanced in two ways. They are Under Sampling and Over Sampling

(a) Under Sampling:

In Under Sampling, dataset balance is done by the reduction of the size of the ample class. This process is considered when the amount of data is adequate.

(b) Over Sampling:

In Over Sampling, dataset balance is done by increasing the size of the scarce samples. This

process is considered when the amount of data is inadequate.

Figure: Data Balancing

3.2.5Prediction of Disease:

Various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction. for the heart disease prediction. various evaluation metrics like accuracy ,confusion matrix ,precision, recall, and f1-score are considered

4 Results and Evolution Metrics

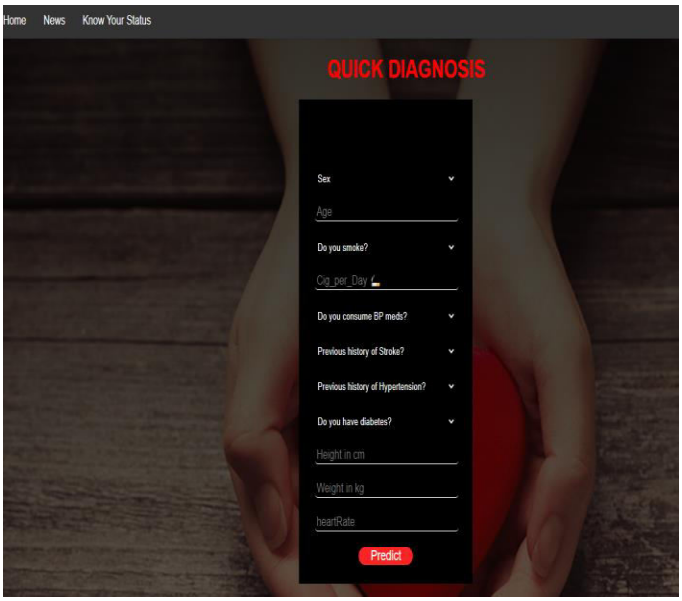
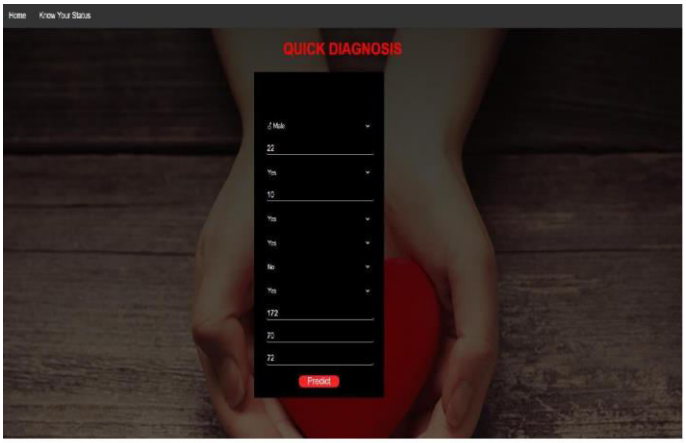


Fig 2:_ Main page



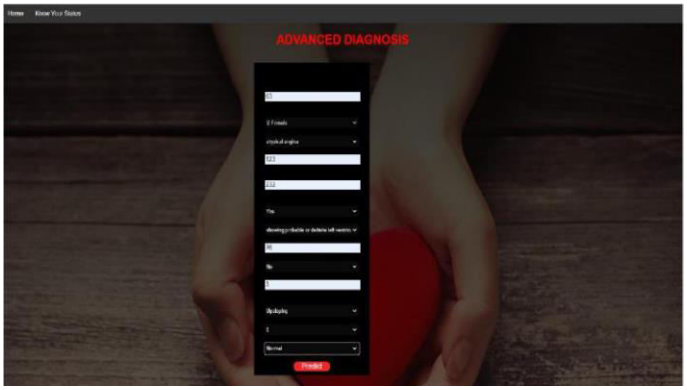


Fig 3 two types of analysis pages

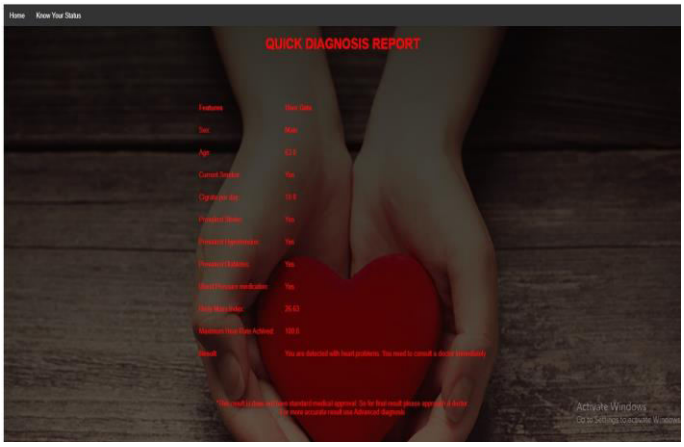


Fig 4:- predicated diseases

5 CONCLUSIONS

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle

changes in high-risk patients and in turn reduce the complications ,which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this paper, the seven different machine learning algorithms used to measure the performance are SVM, Decision Tree, Random Forest, Naïve Bayes, LogisticRegression, Adaptive Boosting, and Extreme Gradient Boosting applied on the dataset.

6 **References**

1. Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis an overview of heart disease prediction. International Journal of Computer Applications,17(8), 43-8

2. Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications,47(10), 44-8.

3. Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology inBiomedicine,10(2), 334-43.

4. Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies,6(1), 637-9.

5. Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp. 259-64). IEEE. ICCRDA 2020 IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012072 IOP Publishingdoi:10.1088/1757-899X/1022/1/012072 9

6. Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heartdisease prediction model: the Korean HeartStudy.BMJopen,4(5),e005025.

7. GannaA,Magnusson P K,Pedersen N L, deFaireU, ReillyM,Ärnlöv J&Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction .Arteriosclerosis ,thrombosis, andvascularbiology,33(9),2267-72.