SURVEY ON HEART DISEASE PREDICTION APPROACHES

RINI JOY
Toc H Institute of science and Technology, Information Technology
Kochi, Kerala, India

Abstract- The world is cornered by issues that affect the daily life of mankind. Amongst these issues the health of a person promotes itself to be of paramount concern. Many diseases have existed and continue to evolve in the world today. The only means to diagnose a disease before it becomes fatal to a human being is by analyzing its respective symptoms. Even a newly formed disease leaves behind a trail of symptoms by which its future occurrence can be depicted. This paper aims at study of various approaches used for heart disease prediction. These approaches are based on classification, clustering, association rule, neural networks, and algorithm and decision tree. In this paper will explain some existing approaches and scope for future study.

Keywords- Heart disease prediction, classification, association rules

I. INTRODUCTION

A heart attack happens when the flow of oxygen-rich blood to a section of heart muscle suddenly becomes blocked and the heart can't get oxygen. If blood flow isn’t restored quickly, the section of heart muscle begins to die. Acting fast at the first sign of heart attack symptoms can save your life and limit damage to your heart. Treatment works best when it's given right after symptoms occur. Many people aren't sure what's wrong when they are having symptoms of a heart attack. When a heart attack happens, any delays in treatment can be deadly. Knowing the warning symptoms of a heart attack and how to take action can save your life or someone else. To save lives require a mechanism that can diagnose in advance the chances of a heart attack that would occur for an individual and thereby provide the necessary treatment to prevent its occurrence in the future.

II. APPROACHES

In this paper[3] Raghu and et al (2011) developed a Decision Support in Heart Disease Prediction System (DSHDPS) using data mining modeling technique, namely, probability. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It is implemented as web based questionnaire application. It can serve a training tool to train nurses and medical students to diagnose patients with heart disease. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse is preprocessed to make the mining process more efficient. The preprocessed data warehouse is then clustered using the K-means clustering algorithm with K=2. This result in two clusters, one contains the data that are most relevant to heart attack and the other contains the remaining data. The frequent patterns are mined from the data, relevant to heart attack, using the MAFIA algorithm. The significant weightage is calculated for all frequent patterns with probability density model. The frequent patterns with significant weightage greater than a predefined threshold are chosen. These chosen significant patterns can be used in the design and development of heart attack prediction system. They have used some more parameters significant to heart attack with their weightage and the priority levels are advised by the medical experts. The neural network is trained with the selected significant patterns. The designed prediction system employed MLPNN with Back-propagation as training algorithm. With the help of the designed prediction system can predict the different risk levels of heart attack. The outcome of predictive data mining technique on the dataset reveals that Decision Tree outperforms and some time Bayesian classification is having similar
accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The author says that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

In this paper[7] present a prediction system for heart disease using Learning vector Quantization neural network algorithm. The neural network in this system accepts 13 clinical features as input and predicts that there is a presence or absence of heart disease in the patient, along with different performance measures. Dataset is collected from Cleveland database. In this paper author uses learning vector quantization (LVQ) neural network algorithm is used. The main purpose of using LVQ is that it creates prototypes that are easy to interpret for experts in the respective application domain. LVQ network is a nearest neighbor pattern classifier based on competitive learning. It is generally applicable in problems of nonlinear separation and used for data classification in large extent. In this paper the system consists of two steps, in the first step 13 clinical attributes are accepted as input and then the training of the network is done with training data by LVQ algorithm. It gives the presence or absence of heart disease. The system also computes different parameters like accuracy, sensitivity, specificity, training time, testing time, error and ROC curve. To improve the performance of the prediction system, it is trained with different number of neurons and also the numbers of training epochs are varied.

In this paper [6] author presented a lazy data mining approach for heart disease classification. Applied information centric attribute measure PCA to generate class association rules. This class association rules will be used to predict the occurrence of heart disease. Dataset is collected from UCI repository. Associative classification is a rule based new approach which integrates association rule mining and classification. It selects a small set of high quality rules and uses these rules for prediction. Lazy associative classification method induces class association rules specific to test instance. The lazy learning approach projects the training data D, only on those features in the test data. Generally Apriori based rule generation algorithm generates 2k -1 rules for the data set with k items .so it leads to high computation cost. To reduce the no. of rules generated and to improve accuracy author used information centric attribute PCA in lazy associative classification. PCA is dimensionality reduction technique used to find a new set of attributes that better captures the variability of the data.

In this paper [5] proposes the neural network approach is used for analyzing the heart disease dataset. Applying feed forward neural network model and back propagation learning algorithm with variable learning rate and momentum, the heart disease database are trained by the neural network. The input layer contains 13 neurons to represent 13 attributes. It consists of 4 class labels namely normal person, first stroke, second stroke and end of life. The output layer consists of two neurons to represent these four classes. Some of the neural networks are constructed with and without hidden layer that is single and multilayer networks are trained. The dataset was collected from Cleveland database[11]. This dataset classifies the person into normal and abnormal person based on heart diseases. The dataset consists of 414 instances, 13 attributes and a class attribute. Both test and training data are used for performance analysis. In a trained network, the test data is given as the input. With the adjusted weights, the output of the network is calculated. From the experimental results, the author concluded that efficiency of the classification process is increased by applying parallel approach which is adopted in the training phase. In future this work will be enhanced by applying genetic algorithm using neural networks.

In this paper [10] Tanagra tool is used to compare the performance accuracy of data mining algorithms for diagnosis of heart disease dataset. The pre-processed data warehouse is then classified using Tanagra tool. The feature selection in the tool describes the attribute status of the data present in the heart disease. Using unsupervised machine learning algorithm such as k-NN, K-Mean and Entropy based clustering and the results are compared. Tanagra is a collection of machine learning algorithms for data mining tasks. The algorithms can be applied directly to a dataset. Tanagra contains tools for data classification, statistics, clustering, supervised learning, meta-supervised
This paper [3] proposes a Decision Support in Heart Disease Prediction System (DSHDP) using Naive Bayes. Using medical profiles such as age, sex, blood pressure and blood sugar, it can predict the likelihood of patients getting a heart disease. It is implemented as a web-based questionnaire application. It can serve as a training tool to train nurses and medical students to diagnose patients with heart disease. Medical dataset was obtained from the Cleveland Heart Disease database. With the help of the dataset, the patterns significant to the heart attack prediction are extracted. The records were split equally into two datasets: training dataset and testing dataset. To avoid bias, the records for each set were selected randomly. The attribute “Diagnosis” is identified as the predictable attribute with value “1” for patients with heart disease and value “0” for patients with no heart disease. “PatientId” is used as the key; the rest are input attributes. Naive Bayes classifier learns from the “evidence” by calculating the correlation between the target (i.e., dependent) and other (i.e., independent) variables. The Naive Bayes Classifier technique is particularly suited when the dimensionality of the inputs is high.

In this paper [2], predicting heart disease from horoscope of a person using data mining techniques. Horoscope has 12 regions each occupying 30°. There are 12 planets that can occupy any house in a horoscope. Each house deals with different significations and the same set of combinations obtained in a particular house might influence the different significations comprehended by the house, in different ways. Based on house the author predicts the heart disease. The analysis is implemented in Weka. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The algorithms used in this work are Decision table, Multilayer perception, J48 and LWL.

Association rule mining technique proposed by Jabbar and et al (2011) is used for predicting heart attack. In this paper[1], the author proposed a novel method CBARBSN for association rule mining based on sequence numbers and clustering the transactional database for predicting heart attack. The 2 essential steps of this method are, first the scientific statistics is converted into binary and the proposed method is applied to the binary transactional data.

The data is collected from Cleveland database[11]. The medical data contains 14 attributes. In this approach the database is divided into equal size partition and each partition will be called as cluster. Each cluster is examined individually by loading the first cluster into memory and calculating frequent item sets. Then the second cluster is considered similarly and calculating frequent item sets. This approach reduces main memory requirement since it considers only a small cluster at a time and it is scalable and efficient.

This paper [9] presents a computer aided disease diagnosis scheme by a genetic algorithm based fuzzy logic approach. The system uses Cleveland heart disease database. There were 13 attributes involved in the prediction of heart disease. The genetic algorithm is used to determine the attributes that contribute more towards the diagnosis. By genetic search, 13 attributes are reduced to 6 attributes. This paper uses mamdani inference method. Performance of the models evaluated by standard metrics of accuracy, precision, recall and F-measure which were calculated using the predictive classification table, known as Confusion Matrix.

In this paper [8] presents the performance comparison of 10 classification data mining techniques for predicting heart disease survivability. The 10 classification data mining techniques are Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), C4.5 algorithm, k-NN, BLR, Multinomial Logistic Regression (MLR), PLS-DA, The k-means algorithm, Entropy based Mean Clustering (EMC) algorithm, Apriori algorithm. Analyzing the performance by varying performance measures such as computing time, true positive, false negative, true negative, false positive, accuracy, specificity, sensitivity, rate-cross validation error rate, positive precision, negative precision, bootstrap validating error rate. Dataset collected from Cleveland dataset. The records were used for training and visualization. It is also well suited for developing new machine learning schemes. This paper concentrates on functional algorithms like K-NN, K-Mean and Entropy based Clustering.
split equally into two datasets: training dataset and testing dataset. The attribute “Diagnosis” was identified as the predictable attribute with value “1” for patients with heart disease and value “0” for patients with no heart disease. From experiment results PLS-DA classifier is suggested for Heart disease based classification to get better results with accuracy and performance.

III. CONCLUSION

Heart disease is the leading cause of death for both men and women. Know the warning signs and symptoms of a heart attack so that you can act fast if you or someone you know might be having a heart attack. The chances of survival are greater when emergency treatment begins quickly. This paper mainly focuses on the study of various approaches of heart attack disease prediction. Various heart disease prediction research papers are analyzed and studied. Prediction accuracy of existing system can be improved, so In future, new algorithms and techniques are to be developed which overcome the drawbacks of the existing system.

REFERENCES

[1] MA Jabbar, Dr. Priti Chandra, B.L. Deekshatulu “Cluster Based Association Rule Mining For Heart Attack Prediction” JTAIT Vol. 32 No.2 October 2011.
[6] M.Akhil Jabbar, Dr. B.L Deekshatulu, Dr Priti Chandra “Heart Disease Prediction using Lazy Associative Classification “ IEEE publication 2013