Middle-East Journal of Scientific Research 24 (1): 192-200, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.24.01.22944

Boosted Apriori: an Effective Data Mining Association Rules for Heart Disease Prediction System

[1]R. Thanigaivel and [2]K. Ramesh Kumar

[1]Department of Computer Science and Engineering, Hindustan University, Chennai, Tamilnadu, India
[2]Department of Information Technology, Hindustan University, Chennai, Tamilnadu, India

Abstract: In health concern business, data mining plays a significant task for predicting diseases. Numeral number of tests must be requisite from the patient for detecting a disease. However using data mining technique can reduce the number of test that are required. Cardiovascular disease is the principal source of deaths widespread and the prediction of Heart Disease is significant at an untimely phase. In order to reduce numbers of deaths from heart diseases there have to be a quick and efficient detection technique. The principle of this study is, hence to extract hidden patterns by applying data mining techniques, which are noteworthy to heart diseases, from a data collected together by an Hospital.

Key words: Data Mining · Heart Disease · Association Rule · Classification Techniques · Simulation

INTRODUCTION

The extraction of significant patterns from the heart disease data warehouse was presented. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse preprocessed to make the mining process more efficient. The first stage of Association Rule used preprocessing in order to handle missing values. Later applied equal interval binning with approximate values based on medical expert advice on Pima Indian heart attack data. The significant items were calculated for all frequent patterns with the aid of the proposed approach. The frequent patterns with confidence greater than a predefined threshold were chosen and it was used in the design and development of the heart attack prediction system. Each data mining technique serves a different purpose depending on the modeling objective. The two most common modeling objectives are classification and prediction. Classification models predict under the categorical labels (discrete, unordered) while prediction models predict continuous-valued functions.

Symptoms of Heart Attack Can Include:

- Discomfort, pressure, heaviness, or pain in the chest, arm, or below the breastbone.
- Discomfort radiating to the back, jaw, throat, or arm.
- Fullness, indigestion, or choking feeling (may feel like heartburn).
- Sweating, nausea, vomiting, or dizziness.
- Extreme weakness, anxiety, or shortness of breath.
- Rapid or irregular heartbeats

Association Rules: Let D = {T1, T2, . . . , Tn} be a set of n transactions and let I bea set of items, I = {i1, i2 . . . im}. Each transaction is a set ofitems, i.e. Ti . An association rule is an implication of the formX ε Y, where X, Y , and X ∩ Y = ε; X is called the antecedentand Y is called the consequent of the rule. In general, a setof items, such as X or Y , is called an itemset. In this work, a transaction is a patient record transformed into a binaryformat where only positive binary values are included as items.This is done for efficiency purposes because transactions represent sparse binary vectors. Let P(X) be the probability ofappearance of itemset X in D and let P(Y |X) be the conditionalprobability of appearance of itemset Y given itemset X appears.For an itemset X ε I, support(X) is defined as the fraction oftransactions Ti ε D such that X ε Ti. That is, P(X) = support(X).The support of a rule Xε Y is defined as support (X * Y) = P(X* Y). An association rule X ε Y has a measure of reliabilitycalled confidence (X * Y) defined as P(Y |X) = P(X*Y)/P (X)= support(X*Y)/ support(X). The standard problem of miningassociation

Corresponding Author: R. Thanigaivel, Department of Computer Science and Engineering, Hindustan University, Chennai, Tamilnadu, India.

rules [1] is to find all rules whose metrics areequal to or greater than some specified minimum supportand minimum confidence thresholds. A k-itemset with supportabove the minimum threshold is called frequent. We use athird significance metric for association rules called lift [25]: lift (X * Y) = P(Y |X)/P (Y) = confidence(X * Y)/support (Y). Lift quantifies the predictive power of Xε Y ; we are interestedin rules such that lift (X * Y) > 1.

Decision Trees: In decision trees [6] the input data set has one attributecalled class C that takes a value from K discrete values 1, . .. , K and a set of numeric and categorical attributes A1, . . ., Ap. The goal is to predict C given A1, . . . , Ap. Decision treealgorithms automatically split numeric attributes Ai into tworanges and they split categorical attributes Aj into two subsetsat each node. The basic goal is to maximize class predictionaccuracy P(C = c) at a terminal node (also called node purity)where most points are in class c and c ◎ {1, . . . , K}. Splittingis generally based on the information gain ratio (an entropybasedmeasure. The splitting process is recursively repeateduntil no improvement in prediction accuracy is achieved witha new split. The final step involves pruning nodes to make thetree smaller and to avoid model overfit. The output is a set ofrules that go from the root to each terminal node consisting of aconjunction of inequalities for numeric variables (Ai <= x, Ai> x)and set containment for categorical variables (Aj ε {x, y, z}) anda predicted value c for class C. In general decision trees havereasonable accuracy and are easy to interpret if the tree has afew nodes. Detailed discussion on decision trees can be foundin [18, 19, 31]. to one item. Our first constraint is the negation of anattribute, which makes search more exhaustive. If an attributehas negation then additional items are created, correspondingto each negated categorical value or each negated interval. Missing values are assigned to additional items, but they arenot used. In short, each transaction is a set of items and eachitem corresponds to the presence or absence of one categoricalvalue or one numeric interval.

The Risk Factor for Heart Disease

Family History of Heart Disease: Most people identify that the heat disease can run in families. That if anybody has a family history of heart disease, he/she may be at greater risk for heart attack, stroke and other heart diseases.

Smoking: Smoking is major cause of heart attack, stroke and other peripheral arterial disease. Nearly 45% of all people who die from smoking tobacco and Chewing Guthka do so due of heart and blood vessel diseases. A smoker's risk of heart attack reduces rapidly after only one year of not smoking.

Cholesterol: Abnormal levels of lipids (fats) in the blood are risk factor of heart diseases. Cholesterol is a soft, waxy substance found among the lipids in the bloodstream and in all the body's cells. High level of triglyceride (most common type of fat in body) combined with high levels of LDL (low density lipoprotein) cholesterol speed up atherosclerosis increasing the risk of heart diseases.

High Blood Pressure: High blood pressure also known as HBP or hypertension is a widely misunderstood medical condition. High blood pressure increase the risk of the walls of our blood vessels walls becoming overstretched and injured. Also increase the risk of having heart attack or stroke and of developing heart failure, kidney failure and peripheral vascular disease and heart related disease.

Obesity: The term obesity is used to describe the health condition of anyone significantly above his or her ideal healthy weight. Being obese puts anybody at a higher risk for health problem such as heart disease, stroke, high blood pressure, diabetes and more.

Lack of Physical Exercise: Lack of exercise is a risk factor for developing coronary artery disease (CAD). Lack of physical exercise increases the risk of CAD, because it also increases the risk for diabetes and high blood pressure.

Diabetes: Diabetes if not controlled can lead to significant heart damage including heart attack and death.

Eating Habits: Healthy diet, intake of low salt in diet, saturated fat in body, Tran's fat, cholesterol and refined sugars will lower our chances of getting heart disease.

Stress: Poorly controlled stress and danger can lead to heart attacks and strokes.

Literature Review: In 2013, Senthil Kumar *et al*., proposed a method that uses components of fuzzy logic like Fuzzification, Advanced Fuzzy Resolution Mechanism and defuzzification. The fuzzification is a process to transfer crisp values into fuzzy values. In the analysis of heart disease a fuzzy resolution mechanism uses predicted value with five layers, each layer has its own nodes. The results are tested with Cleveland heart disease dataset. Fuzzy Resolution Mechanism was developed using MATLAB. Defuzzification process converts the fuzzy set into crisp values. Then the error signal is back propagated in the network. The actual response of the network moves nearer to the desired response by adjusting the synaptic weights in a statistical sense in the network. The generalized delta rule which minimizes the error is used for the weight adjustment in the network. Thus a medical decision support system can be developed particularly in the diagnosing of heart disease.

In 2013, Syed Umar Amin *et al*. [1], developed genetic neural network hybrid system. This system uses the global optimization advantage of genetic algorithm for initialization of neural network weights. A backpropagation algorithm is used to train the networks with optimize initialization of synaptic weights by Genetic Algorithm.

Jesmin Nahar, Tasadduq Imam, Kevin S Tickle and Yi-Ping Phoebe Chen [2] presented a rule extraction model on heart disease using Apriori and Predictive Apriori techniques. They had considered some attributes like age, sex, chest pain, old peak. Based on these attributes they had developed rules for male and female which tells the healthy patterns observed in them. They had demonstrated rule mining to determine interesting knowledge which would analyze the factors causing heart disease.

In 2013, Nabeel Al-Milli, [3], developed heart disease prediction system that uses the back propagation algorithm technique to develop multilayer neural networks in a supervised manner. The error-correction learning rule is the basis for the back propagation algorithm. The algorithm uses a forward pass and a backward pass through the different layers of the network. The forward pass use to fix the synaptic weights of the networks. In the backward pass, the synaptic weights are all adjusted in accordance with an error-correction rule. Error signal is calculated as the difference between the desired output and the actual response of the network.

I.S. Jenzi, P. Priyanka, Dr. P. Alli [4] proposed a new system based on Data Mining for predicting Heart Disease. They had collected patterns from medical data for finding heart disease. They did user friendly application for predicting the disease. They found that decision tree was easy to interpret and had a good accuracy. They found two difficulties like large dataset and user interface should be efficient enough to support data in different format instead of ARTF which can be taken care of in next research.

Abdullah A. Aljumah, Mohammed Gulam Ahamad, Mo-hammad Khubeb Siddique [5] focuses on predictive analysis of diabetic treatment using regression based data mining technique. Authors had collected data from Saudi Arabia consisting of variables like drug, weight, smoke intake and age. They designed a GUI based application which was easy to operate and helpful for providing training to doctors and medical assistants.

Dr. K. Rameshkumar [6] developed a model using ARM (Association Rule Mining) to extract valuable information from database. Author has proposed a new algorithm which would take care of missing values for detecting HIV AIDS. With the help of this proposed algorithm author could able to extract information about CD4 cell counts, RNA levels and treatment given for various patient. The model is lacking of handling data with a very good accuracy.

N. Deepika *et al*. [7] proposed Association Rule for classification of Heart-attack patients . The extraction of significant patterns from the heart disease data warehouse was presented. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse preprocessed to make the mining process more efficient. The first stage of Association Rule used preprocessing in order to handle missing values. Later applied equal interval binning with approximate values based on medical expert advice on Pima Indian heart attack data. The significant items were calculated for all frequent patterns with the aid of the proposed approach. The frequent patterns with confidence greater than a predefined threshold were chosen and it was used in the design and development of the heart attack prediction system. The, Pima Indian Heart attack dataset used was obtained from the UCI machine learning repository. Characteristics of the patients like number of times of chest pain and age in years were recorded. The actions comprised in the preprocessing of a data set are the removal of duplicate records, normalizing the values used to represent information in the database, accounting for missing data points and removing unneeded data fields. Moreover it might be essential to combine the data so as to reduce the number of data sets besides minimizing the memory and processing resources required by the data mining algorithm. In the real world, data is not always complete and in the case of the medical data, it is always true. To remove the number of inconsistencies which are associated with data we use Data preprocessing.

K. Srinivas *et al*. [8] presented Application of Data Mining Technique in Healthcare and Prediction of Heart Attacks. The potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network to the massive Volume of healthcare data. Tanagra data mining tool was used for exploratory data analysis, machine learning and statistical learning algorithms. The training data set consists of 3000 instances with 14 different attributes. The instances in the dataset are representing the results of different types of testing to predict the accuracy of heart disease. The performance of the classifiers is evaluated and their results are analyzed. The results of comparison are based on 10 tenfold cross-validations. According to the attributes the dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing. The comparison made among these classification algorithms out of which the naive Bayes algorithm considered as the best performance algorithm. The performance of various algorithms is listed below.

Diagnosis of heart disease was used Naïve Bayes, K-NN, Decision List in this Naïve Bayes has taken a time to run the data for accurate result when compared to other algorithms.

Sudha *et al*. [9] to propose the classification algorithm like Naïve Bayes, Decision tree and Neural Network for predicting the stroke diseases. The classification algorithm like decision trees, Bayesian classifier and back propagation neural network were adopted in this study. The records with irrelevant data were removed from data warehouse before mining process occurs. Data mining classification technology consists of classification model and evaluation model. The classification model makes use of training data set in order to build classification predictive model. The testing data set was used for testing the classification efficiency. Then the classification algorithm like decision tree, naive Bayes and neural network was used for stroke disease prediction. The performance evaluation was carried out based on three algorithms and compared with various models used and accuracy was measured. While comparing these classification algorithms, the observation shows the neural network performance was more than the other two algorithms

Table 1: Performance Study of Data mining Algorithms

The Algorithms Used	Accuracy	Time Taken
Naïve Bayes	54.33%	613ms
Decision Tree	53.10	816ms
KNN	46.78%	1005ms

Latha Parthiban and R. Subramanian [10] presented Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm . Adaptable based fuzzy inputs are adapted with a modular neural network to rapidly and accurately approximate complex functions. The CANFIS model combined the neural network adaptive capabilities and the fuzzy logic quantitative approach then integrated with genetic algorithm to diagnosis the presence of the disease. Coactive neuro-fuzzy inference system model has good training performance and classification accuracies. Dataset of heart disease was obtained from UCI Machine Learning Repository. Coactive Neuro-fuzzy modeling was proposed as a dependable and robust method developed to identify a nonlinear relationship and mapping between the different attributes.

Dangare *et al*. [11] proposed Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. Prediction System for heart disease used system contains huge amount of data, used to extract hidden information for making intelligent medical diagnosis. The main objective of this research was to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease using historical heart database. To develop the system, medical terms such as sex, blood pressure and cholesterol like 13 input attributes are used. To get more appropriate results, two more attributes i.e. obesity and smoking, as attributes were considered as important attributes for heart disease. A Multi-layer Perceptron Neural Networks (MLPNN) that maps a set of input data onto a set of appropriate. It consists of 3 layers input layer, hidden layer & output layer. There is connection between each layer & weights are assigned to each connection. The primary function of neurons of input layer is to divide input into neurons in hidden layer. The dataset consists of total 573 records in heart disease database. The total records are divided into two data sets one is used for training consists of 303 records & another for testing consists of 270 records. Initially dataset contained some fields, in which some value in the records was missing. These were identified and replaced with most appropriate values using Replace Missing Values filter. The Replace Missing Values filter scans all records & replaces missing values with mean mode method known as Data Preprocessing.

After pre-processing the data, data mining classification techniques such as Neural Networks, is used for classification. Many problems in business, science, industry and medicine can be treated as classification problems. Owing to the wide range of applicability of ANN and their ability to learn complex and nonlinear relationships including noisy or less precise information, neural networks are well suited to solve problems in biomedical engineering. So here use for the neural network technique is classification of medical dataset 14 attributes by considering the single and multilayer neural network models.

Olatubosun Olabode *et al*. [12] to classify the Cerebrovascular disease by using artificial neural network with back propagation error method. The Multi-layer perceptions artificial neural networks with back-propagation error method were feed-forward nets with one or more layers of nodes between the input and output nodes. These additional layers contain hidden units or nodes that were not directly connected to both the input and output nodes. The neural network was trained using back propagation algorithm with sigmoid function on one hidden layer with the 16 input attributes. Predictive models were used in variety of domains for the diagnosis. Dataset for this work were collected 100 records (60 males and 40 females) from federal medical fields. The input values obtained from the records of the forms the input variables in the input layer with 16 nodes. The neural network weights were initialized randomly. This work range of the weights was between [-0.5 and 0.5] and the learning rate was set between 0.1 and 0.9. The training, validation, generalization accuracy was measured.

Data Set: Through our proposed system, we have reduced six attributes to four which will be employed for the prediction of heart conditions. Data of multiple patients is entered into the proposed system and the diagnosed results generated by the system corresponding to those patients have been saved in the database. The resultant data thus obtained is used by the model for calculating the efficiency of the proposed system. Original List of Attributes were as follows:

- Type - Chest Pain Type
- Rbp - Resting blood pressure
- Eia - Exercise induced angina
- Oldpk - Old peak
- Vsl - No. of vessels colored
- Thal -Maximum heart rate achieved

Reduced Input Attributes are as follows:

- Blood Pressure
- LDL – Low Density Lipoprotein (commonly known as Bad Cholesterol)
- HDL – High Density Lipoprotein (commonly known as Good Cholesterol)
- Triglycerides

By reducing the number of attributes as compared to earlier systems we plan on reducing the overall time required by a Doctor/ Physician to diagnose a heart condition of his/ her patient; this in turn reduces the overall cost and time required by a patient for his/her treatment [13-48].

Table 1: Medicaldataset

Attribute	Description
AGE	Ageofpatient
LM	LeftMainnarrowing
LAD	LeftAnteriorDesc.arterynarrowing
LCX	LeftCircumflexarterynarrowing
RCA	RightCoronaryarterynarrowing
AL	Antero-Lateral
AS	Antero-Septal
SA	Septo-Anterior
SI	Septo-Inferior
IS	Infero-Septal
IL	Infero-Lateral
LI	Latero-Inferior
LA	Latero-Anterior
AP	Apical
SEX	Gender
HTA	Hyper-tensionY/N
DIAB	DiabetesY/N
HYPLD	HyperloipidemiaY/N
FHCAD	Familyhist.ofdisease
SMOKE	PatientsmokesY/N
CLAUDI	ClaudicationY/N
PANGIO	PreviousanginaY/N
PSTROKE	PriorstrokeY/N
PCARSUR	PriorcarotsurgY/N
CHOL	Cholesterollevel

Association Rules for Healthy Heart:

Confidence = 1:
IF 0 <= AGE < 40:0 - 1:0 <= AL < 0:2 PCARSUR = n
THEN 0 <= LAD < 50, s=0.01 c=1.00 l=2.1
IF 0 <= AGE < 40:0 - 1:0 <= AS < 0:2 PCARSUR = n
THEN 0 <= LAD < 50, s=0.01 c=1.00 l=2.1
IF 40:0 <= AGE < 60:0 SEX = F 0 <= CHOL < 200

THEN 0 <= LCX < 50, s=0.02 c=1.00 l=1.6
IF SEX = F HTA = n 0 <= CHOL < 200
THEN 0 <= RCA < 50, s=0.02 c=1.00 l=1.8
Two items in the consequent:
IF 0 <= AGE < 40:0 - 1:0 <= AL < 0:2
THEN 0 <= LM < 30 0 <= LAD < 50, s=0.02 c=0.89 l=1.9
IF SEX = F 0 <= CHOL < 200
THEN 0 <= LAD < 50 0 <= RCA < 50, s=0.02 c=0.73 l=2.1
IF SEX = F 0 <= CHOL < 200
THEN 0 <= LCX < 50 0 <= RCA < 50, s=0.02 c=0.73 l=1.8
Confidence >= 0:9:
IF 40:0 <= AGE < 60:0 - 1:0 <= LI < 0:2 0 <= CHOL < 200
THEN 0 <= LCX < 50, s=0.03 c=0.90 l=1.5
IF 40:0 <= AGE < 60:0 - 1:0 <= IL < 0:2 0 <= CHOL < 200
THEN 0 <= LCX < 50, s=0.03 c=0.92 l=1.5
IF 40:0 <= AGE < 60:0 - 1:0 <= IL < 0:2 SMOKE = n
THEN 0 <= LCX < 50, s=0.01 c=0.90 l=1.5
IF 40:0 <= AGE < 60:0 SEX = F DIAB = n
THEN 0 <= LCX < 50), s=0.08 c=0.92 l=1.5
IF HTA = n SMOKE = n 0 <= CHOL < 200
THEN 0 <= LCX < 50, s=0.02 c=0.92 l=1.5

Association Rules for Diseased Arteries:

Confidence = 1:
IF 0:2 <= SA < 1:0 HY PLPD = y PANGIO = y
THEN 70 <= LAD < 100, s=0.01 c=1.00 l=3.2
IF 60 <= AGE < 100 0:2 <= SA < 1:0 FHCAD = y
THEN not(0 <= LAD < 50, s=0.02 c=1.00 l=1.9
IF 0:2 <= IS < 1:0 CLAUDI = y PSTROKE = y
THEN not(0 <= RCA < 50), s=0.02 c=1.00 l=2.3
IF 60 <= AGE < 100:0 0:2 <= IS < 1:0 250 <= CHOL < 500
THEN 70 <= RCA < 100, s=0.02 c=1.00 l=3.2
IF 0:2 <= IS < 1:0 SEX = F 250 <= CHOL < 500
THEN 70 <= RCA < 100, s=0.01 c=1.00 l=3.2
IF 0:2 <= IS < 1:0 HTA = y 250 <= CHOL < 500
THEN 70 <= RCA < 100, s=0.011 c=1.00 l= 3.2

Two items in the consequent:
IF 0:2 <= AL < 1:1 PCARSUR = y
THEN 70 <= LAD < 100 not(0 <= RCA < 50), s=0.01 c=0.70 l=3.9
IF 0:2 <= AS < 1:1 PCARSUR = y
THEN 70 <= LAD < 100 not(0 <= RCA < 50), s=0.01 c=0.78 l=4.4
IF 0:2 <= AP < 1:1 PCARSUR = y
THEN 70 <= LAD < 100 not(0 <= RCA < 50), s=0.01 c=0.80 l=4.5
IF 0:2 <= AP < 1:1 PCARSUR = y
THEN not(0 <= LAD < 50) not(0 <= RCA < 50), s=0.01 c=0.80 l=2.8
confidence>= 0:9:
IF 0:2 <= SA < 1:1 PANGIO = y)
THEN 70 <= LAD < 100, s=0.023 c=0.938 l= 3.0
IF 0:2 <= SA < 1:0 SEX = M PANGIO = y
THEN 70 <= LAD < 100, s=0.02 c=0.92 l=2.9
IF 60 <= AGE < 100:0 0:2 <= IL < 1:1 250 <= CHOL < 500
THEN 70 <= RCA < 100, s=0.02 c=0.92 l=2.9
IF 0:2 <= IS < 1:0 SMOKE = y 250 <= CHOL < 500
THEN 70 <= RCA < 100, s=0.02 c=0.91 l=2.9

Proposed System: Our project has been mainly developed with an aim to efficiently diagnose the presence of heart disease in an individual. For this purpose we are going to use ASP.NET as our front end where in we could create a user interface to accept user details and back end would be Excel. The front end would basically work as.

Register: Firstly, if the patient is not registered or is arriving for the first time to the –doctor he should register himself so that his information can be stored in the database which would be useful in the future for diagnosis. So Initially the patient needs to register himself for the system. But, if the patient is an old user then he might go for the next step as below.

Login: In this step, the patient would login through his user id and access his own profile where in ASP.NET would be useful for giving access to the patient's profile.

User Input: After accessing the profile the doctor would enter the details of the patient as mentioned by him. The doctor would mainly undertake tests considering the four attributes in mind such as Blood Pressure – where in the values observed by the doctor would be entered in the field corresponding to Blood Pressure. Similarly all other values corresponding to the associated attributes such as LDL - Low Density Lipoprotein (commonly known as Bad Cholesterol), HDL - High Density Lipoprotein (commonly known as Good Cholesterol) and Triglycerides observed by the doctor would be entered by him respectively. Thus this would complete all the information required from the patient.

Final Report: After getting the information from the patient, Data mining would be utilized where in the current details of the patient would be compared by his previous details and apriori algorithm would be used to identify if the patient has some symptoms of Heart Diseases or not. Thus, inorder to access the patients history Excel would also be used as the Back end for our System.

Lab Tests	Before taking our product	Test was done 9 months later
Fasting blood sugar	7.0 (3.6 – 6.1 mmol/L)	6.4 mmol/L
Cholesterol	7.87 (0 .2 – 5.2 mmol/L)	6.12 mmol/L
Triglyceride	4.88 (0.20 – 1.90 mmol/L)	did not change
Liver Enzymes		
GGT	184 (13-47 U/L)	131 U/L **reduced**
ALT	50 (0 – 40 U/L)	33 U/L **normal**
Prescription Medication		
Metformin	**4 tablets a day**	**none**
Diamicron	**1 tablet a day**	**none**

CONCLUSION

Heart disease is one of the leading causes of death worldwide and the early prediction of heart disease is important. The computer aided heart disease prediction system helps the physician as a tool for heart disease diagnosis. From the analysis it is concluded that, data mining plays a major role in heart disease classification. Neural Network with offline training is a good for disease prediction in early stage and the good performance of the system can be obtained by preprocessed and normalized dataset. The classification accuracy can be improved by reduction in features.

REFERENCES

1. Amin Syed Umar, KavitaAgarwal and Dr. Rizwan Beg, 2013. Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factor, in Proc. IEEE Conference on Information and Communication Technologies(ICT).
2. Nahar Jesmin and Tasadduq Imam, 2013. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach, Elsevier.
3. Al-Milli N., 2013. Backpropagation Neural Network for Prediction of Heart Disease, Journal of Theoretical and Applied Information Technology, 56(1).
4. Jenzi, I.S., P. Priyanka and P. Alli, 2013. A reliable Classifier Model Using Data Mining Approach for Heart Disease Prediction, International Journal of Advanced Research in Computer Science and Software Engineering, 3(3): 20-24.
5. Aljumah Abdullah A., Mohammed Gulam Ahamad and Mohammad Khubeb Siddique, 2013. Application of data mining: Diabetes helth care in young and old patients, Journal of King Saud University- Computer and Information Sciences, 25: 127-136.
6. Rameshkumar, K., 2012. Association Rules Mining from HIV/AIDS patient's case history database with missing values, International Journal on Data Mining and Intelligent Information Technology Applications, 2(1): 18-24.

7. Deepika, N. and K. Chandra shekar, 2011. Association rule for classification of Heart Attack Patients, International Journal of Advanced Engineering Science and Technologies, 11(2): 253-257.
8. Srinivas, K., B. Kavitha Rani and A. Govrdhan, 2011. Application of Data Mining Techniques in Healthcare and Prediction of Heart Attacks, International Journal on Computer Science and Engineering, 2(2): 250-255.
9. Sudha, A., P. Gayathiri and N. Jaisankar, 2012. Effective Analysis and Predictive Model of Stroke Disease using Classification Methods, International Journal of Computer Applications, 43(14): 0975-8887.
10. Parthiban, Latha and R. Subramanian, 2008. Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm, International Journal of Biological, Biomedical and Medical Sciences, 3: 3.
11. Dangare, Chaitrali S. and Sulabha S. Apte, 2012. Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques, International Journal of Computer Applications, 47(10): 0975-888.
12. Olabode, Olatubosun and Bola Titilayo Olabode, 2012. Cerebrovascular Accident Attack Classification Using Multilayer Feed Forward Artificial Neural Network with Back Propagation Error, Journal of Computer Science, 8(1):18-25.
13. Dent Tom, 2010. Predicting the risk of coronary heart disease, PHG Foundation Publisher.
14. World Health Organization, 2010. Global Status Report on No Communicable Diseases.
15. World Health Organization, 2007. Pocket Guidelines for Assessment and Management of Heart Risk.
16. Shanthi, D., G. Sahoo and N. Saravanan, 2008. Input Feature Selection using Hybrid Neuro-Genetic Approach in the Diagnosis of Stroke Disease, International Journal of Computer Science and Network Security, 8(12): 99-106.
17. Jabbar, M.A., Priti Chandra and B.L. Deekshatulu, 2011. Cluster based association rule mining for heart attack prediction, Journal of Theoretical and Applied Information Technology, 32(2): 197-201.
18. Soni Jyoti, Ujma Ansari and Dipesh Sharma, 2011. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, International Journal of Computer Applications, 17(8): 43-48.
19. Srinivas, K., G. Raghavendra Rao and A. Govardhan, 2011. Survey on prediction of heart morbidity using data mining techniques, International Journal of Data Mining & Knowledge Management Process, 1(3): 14-34.
20. Usha Rani K., 2011. Analysis of heart diseases dataset using neural network approach, International Journal of Data Mining and Knowledge Management Process, 1(5): 1-8.
21. Ephzibah, E.P. and V. Sundarapandian, 2012. A neuro fuzzy expert system for heart disease diagnosis, An International Journal Computer Science & Engineering, 2(1): 17-23.

. Shouman Mai, Tim Turner and Rob Stocker, 2012. Integrating Decision Tree and K-Means Clustering with Different Initial Centroid Selection Methods in the Diagnosis of Heart Disease Patients, Proceedings of the International Conference on Data Mining.

22. John, Peter T. and K. Somasundaram, 2012. Study and development of novel feature selection framwork for Heart disease prediction, International Journal of Scientific and Research Publications.
23. Fal Dessai Indira S., 2013. Intelligent Heart Disease Prediction System Using Probabilistic Neural Network, International Journal on Advanced Computer Theory and Engineering.
24. Akhiljabbar M., Priti Chandra and B.L. Deekshatulu, 2012. Heart Disease Prediction System using Associative Classification and Genetic Algorithm, International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies.
25. Dangare, Chaitrali S. and Sulabha S. Apte, 2012. Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques, International Journal of Computer Applications.
26. Rahman, Rashedur M. and Farhana Afroz, 2013. Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis, Journal of Software Engineering and Applications.
27. Venkatadri, M. and Lokanatha C. Reddy, 2011. a review on data mining from past to the future. International Journal of Computer Applications.
28. Sudhakar, K., 2014. Study of Heart Disease Prediction using Data Mining, 4(1): 1157-1160.
29. Chitra, R. and V. Seenivasagam, 2013. Review of Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques, Journal on Soft Computing (ICTACT), 3(4): 605-609.
30. Sundar, N.A., P.P. Latha and M.R. Chandra, 2012. Performance Analysis of Classification Data Mining Techniques over Heart Disease Data Base, International Journal of Engineering Science & Advanced Technology, 2(3): 470-478.

31. Pattekari, S.A. and A. Parveen, 2012. Prediction System for Heart Disease Using Naive Bayes, International journal of Advanced Computer and Mathematical Sciences, 3(3): 290-294.
32. Ordonez, C., 2006. Association rule discovery with the train and test approach for heart disease prediction., IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society, 10(2): 334-43.
33. Xing, Y., J. Wang, Z. Zhao and A. Gao, 2007. Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease, in 2007 International Conference on Convergence Information Technology (ICCIT 2007), pp: 868-872.
34. Srinivas, K., K. Raghavendra Kao and A. Govardham, 2010 Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques, in The 5th International Conference on Computer Science & Education, pp: 1344-1349.
35. Hsieh Nan-Chen, Lun-Ping Hung and Chun-Che Shih, 2012. Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques, J Med Syst, 2012.
36. Adebayo Peter Idowu, 2013. Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study, International Journal of Applied Information Systems.
37. Monali, Dey and Siddharth Swarup Rautaray, 2014. Study and Analysis of Data mining Algorithms for Healthcare Decision Support System, International Journal of Computer Science and Information Technologies.
38. Bhuvaneswari, R. and K. Kalaiselvi, 2012. Naïve Bayesian Classification Approach in Healthcare Applications, International Journal of computer Science and Telecommunication, 3(1): 106-112.
39. Shouman, M., T. Turner and R. Stocker, 2012. Applying K-Nearest Neighbour in Diagnosing Heart Disease Patients, International Journal of Information and Education Technology, 2(3): 220-224.
40. Kumar, A.S., 2013. Diagnosis of Heart Disease Using Advanced Fuzzy Resolution Mechanism, International Journal of Science and Applied Information Technology, 2(2): 22-30.
41. Ghazavi, Sean N. and Thunshun W. Liao, 2008. Medical data mining by fuzzy modeling with selected features Artificial Intelligence in Medicine, 43: 195-206.
42. Anbarasi, M., E. Anupriya and N.CH.S.N. Iyenga, 2010. Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology, 2(10): 5370-5376.
43. Shanthi, D., G. Sahoo and N. Saravanan, 2008. Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke, International Journal of Biometric and Bioinformatics, 3(1): 250-255.
44. Anooj, P.K., 2011. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules, Journal of King Saud University Computer and Information Sciences, 11: 309-314.
45. Obi, J.C. and A.A. Imainvan, 2011. Decision Support System for the Intelligent Identification of Alzheimer using Neuro Fuzzy logic, International Journal on Soft Computing , 2(2): 25-38.
46. Shanthi, D., G. Sahoo and N. Saravanan, 2009. Evolving Connection Weights of Artificial Neural Network Using Genetic Algorithm With Application to the Prediction Stroke Diseases, International Journal of Soft Computing, 2: 95-101.
47. Kukar Matjaz, 1999. Analysing and improving the diagnosis of ischaemic heart disease with machine learning, Elsevier.
48. Humar, Kahramanli and Novruz Allahverdi, 2008. Design of a hybrid system for the diabetes and heart diseases, Elsevier.