International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 10 No. 4 Mar. 2015, pp. 1312-1322

http://www.ijias.issr-journals.org/

Data Mining in Healthcare for Heart Diseases

Umair Shafique, Fiaz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa

Department of Information Technology,
University of Gujrat,
Gujrat, Pakistan

ABSTRACT: Data Mining is the area of research which means digging of useful information or knowledge from previous data. There are different techniques used for the data mining. Data mining may used in different fields including Healthcare. Heart or Cardiovascular diseases are the very hot issue in Healthcare industry globally. Many patients died due to insufficient amount of knowledge. As Healthcare industry produces a huge amount of data, we may use data mining to find hidden patterns and interesting knowledge that may help in effective and efficient decision making. Data mining in Healthcare is a crucial and complicated task that needs to be executed accurately. It attempts to solve real world health problems in diagnosis and treatment of diseases. This work is also an attempt to find out interesting patterns from data of heart patients. There are three algorithm used with two different scenarios. These implemented algorithms are Decision Tree, Neural Network and Naïve Bayes.

KEYWORDS: Heart Diseases, Decision Tree, Weka, Neural Network, Naïve Bayes.

1 INTRODUCTION

Data Mining or Knowledge Discovery in Databases (KDD) [1] is one of the most motivated areas of research which means finding of interesting patterns and meaningful information of data. Different Data Mining tools and techniques are used to predict behavior and trends in the data which allow the experts to make active and more accurate decision based on the knowledge. World Wide Data mining has successfully been used in different fields of life including marketing, banking, businesses etc; however it may also be used in the field of Healthcare.

Now a day's Heart or Cardiovascular diseases are the very hot issue in Healthcare industry globally. In April 2011, World Health Organization (WHO) published the latest data [2] According to that Coronary Heart Disease Deaths reached 15.36% of total deaths in Pakistan and according to World Health Organization by the year 2030 more than 23 million people will die annually from heart diseases [3].

Healthcare industry produces a huge amount of data about heart diseases but all is waste and nothing as it does not helps in effective and efficient decision making. Doctor's as well health care expert has their own experience on the bases of which they predict about particular heart disease of the patient which may sometime leads to the false results. So there is need to apply data mining by using the patient's data and implemented and analyzing different data mining techniques to find hidden patterns that may support in many ways also each data mining technique have a different purpose depending on the need and use. Although Data mining in Healthcare is an important yet complicated task that needs to be executed accurately. It attempts to solve real world health problems in diagnosis and treatment of diseases.

There will be many advantages of the Data Mining in Healthcare e.g. it may provide benefits of grouping the patients having similar type of diseases or health issues so that they can be provided with effective treatments, check or provide availability of medical solution to the patients at lower cost, safe healthcare treatment, reducing the time for medical treatment, providing detection of causes of diseases and identification of medical treatment methods and efficient use of other resources etc. It also helps the Healthcare organizations and experts in making efficient healthcare policies.

Corresponding Author: Umair Shafique

In our study we will use online available dataset of heart patients with Weka data mining software. We implements three different algorithms named as Decision Tree, Neural Network and Naïve Bayes with and without attributes selection and check the performance of each.

2 OVERVIEW OF DATA MINING

Due to huge amount availability and the need to convert that data into useful knowledge Data mining techniques can be useful. In recent years, Data mining has found significance in almost every field including health care. The abundance of data and the need of powerful analysis tool for that data, described as "data rich but information poor" situation. Knowledge Flow Diagram is shown in Figure 1.

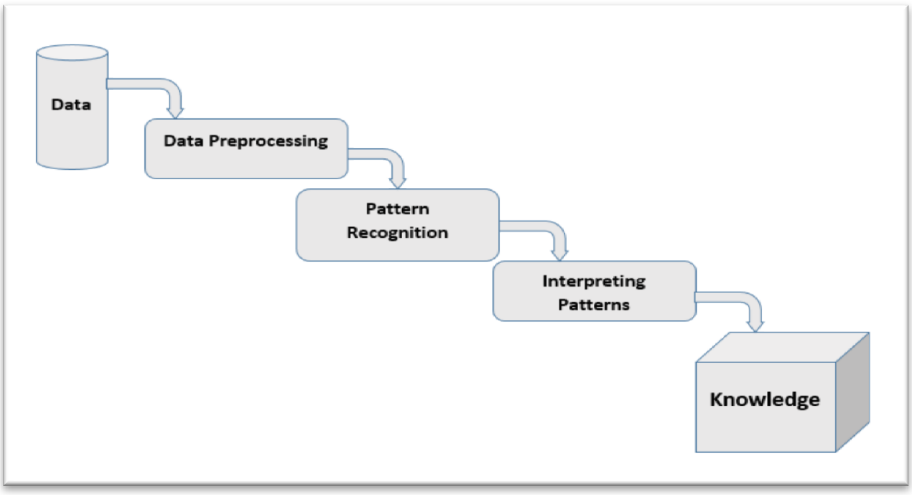

Fig. 1. Knowledge Flow Diagram

The large dataset become data tombs. Data mining has some other names such as knowledge mining from databases, pattern analysis, knowledge extraction, data dredging, and data archaeology. Mining process is more than the data analysis it includes classification, clustering, and association rule discovery etc.

3 DATA MINING APPLICATIONS IN HEALTH SECTOR

The health sector [4] without any argue has more need for data mining today and the applications of data mining are getting popularity in that field. There has been reported of successful data mining applications in health sector and has vast potential. For example data mining applications can evaluate the effectiveness of medical treatments by comparing and contrasting causes, symptoms and can also identify successful standardized treatments for specific diseases. These can also help in health care management by identify and track the states and risk to the patients of the diseases. Quality improvement at low cost in health sector is another important task [5]. The healthcare insurers use data mining to reduce the existence of medical insurance fraud and abuse. The diagnostic and laboratory procedures are costly and painful to patients that increases data mining popularity in the health sector in diagnosis and decision support.

There are some limitation of data mining in health care sector one of them is the integration and collection of data, raw data for data mining exist in different places and systems the solution for this is to collect all the data in a data warehouse. The next one is the quality of data, efficient results needs quality data to be input and missing, inconsistent, or non-standardized data which may be in different formats may cause lack of quality. The domain knowledge and the knowledge of data mining tools and techniques are also very important. Finally, healthcare organizations must make substantial investment on resources, time, effort, and money and mining team should work collaboratively. Data mining requires intensive planning and technological preparation.

4 OVERVIEW OF HEART DISEASES

Human's life fully depends upon the efficient working heart without any break or pause. The term heart or cardio disease refers to such disease that related to the heart and its blood circulatory system. It is a general name for a wide variety of

diseases and disorders that affect the heart and sometimes the blood vessels as well. These are caused by disorder of heart and its pumping system. That may results illness, disability or may be death.

4.1 FACTORS LEADS TO HEART DISEASES AND TREATMENTS TO AVOID THEM

There are number of factors that increase the risk of heart disease. These factors lead to the heart problems and diseases. These factors are Family History, Hyper Tension, Blood Pressure, Cholesterol, Smoking or Tobacco, Poor or Unhealthy Diet, Physical Inactivity etc.

In early stages many Heart Diseases can be avoid by the patients by simply preventing or controlling, prevention measures include regular exercise, healthy and well balanced diet, avoid smoking, maintaining the normal healthy weight etc. Risk factors such as diabetes, cholesterol, hyper-tension etc can also be control or prevented through regular medicine care and by changing life style.

Critical type of heart disease such as heart attack, heart failure or stroke required hospitalization and the treatment for these diseases include supplement amount of oxygen that is deliver to the heart tissues. It also includes monitoring of vital signs and advance life support measures.

5 RELATED WORKS

A lot of work has been done related to heart diseases using data mining techniques. Some of them are given below.

In 2010, Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm [6] was proposed. The Genetic Algorithm is an optimization techniques inspired by natural selection and natural genetics. A Heart disease system was developed to predict accurately the presence of heart disease with reduced number of attributes. In their proposed system they enhanced prediction of heart disease using feature based subset selection in which they used 6 attributes instead of 13 attributes that were originally used. There are three classifiers Decision tree, Classification with clustering and Naive Bayes were used for diagnosis of patients with heart disease and the Weka data mining tool was used for experiments. These experiments show that the Decision tree has highest accuracy as well as construction time as compare to others Naïve Bayes and Classification via clustering.

In 2011, HDPS: Heart Disease Prediction System [7] is introduced in which only one data mining algorithm is used that is artificial neural network (ANN) that is used to classify the heart disease based on 13 different attributes. The data set is used in this system is taken from UCI machine learning repository having 303 instances.

In 2012, data mining neural network approach used for prediction about heart disease [8]. This shows about 100% accuracy. Multilayer Perceptron Neural Network (MLPNN) with Back propagation algorithm (BP) was used in the system and multilayer preceptron is one of the most important models in neural networks. It consists of different levels that are connected with each other by small circles called nodes. The back propagation algorithm is widely used algorithm that calculates the difference between actual and predicted values from output nodes to the previous layer of the nodes. WEKA data mining tool is used for the experiments and the data set for this contains 573 records which are divided into two parts training and testing. Total 15 attributes were used in this to increase the accuracy of the prediction.

In 2013, intelligent and effective heart disease prediction system was introduced using Weighted Associative Classifiers (WAC) [9], in WAC different weights are assigned to the attributes according to their capability about predicting. The system is implemented by using JAVA platform and benchmark data from online available UCI repository. The data set contain 303 records and 14 different attributes. The results of experiments show about 80% accuracy of WAC.

In 2013, different data mining approaches are used to predict heart diseases [10]. Different experiments were conducted using Weka tool. The results of these experiments were compared with each other using 10 cross validations with and without bagging. Bagging means Bootstrap aggregation which is used to increase the accuracy of the classification. Three techniques were used and compare in this research i.e. Naïve Bayes, J48 Decision Tree and Bagging. The data set was taken from Hungarian Institute of Cardiology, that have 76 raw attributes but only 11 attributes were selected for experiments. The experiments also shows that the bagging has the highest accuracy i.e. 85.03 % and J48 decision tree and naïve bayes have 84.35% and 82.31% accuracy respectively.

In March 2014, Frequent Feature Selection method was used for predicting about the heart diseases [11] fuzzy measure and the relevant nonlinear integral are use with this method to give good performance. The experiments were conducted using WEKA data mining tool, 8 attributes are used with 1000 records. Weights are assigned to the attributes to predict about heart attack. There are different stages of mining using feature subset selection method, first is Weighted Support in

which each record contain relative weight of all transaction and the second is MAximal Frequent Itemset Algorithm (MAFIA) which combines old and new algorithmic ideas which mine the frequent item set and at the end significance weight age of each pattern is calculated.

6 MATERIALS AND METHODS

6.1 DATA SET FOR EXPERIMENTS

We took online available heart disease datasets from UCI (University of California, Irvine C.A) Machine Learning Repository [12]. These dataset have the same instance format and attributes. These datasets have 76 raw attributes including predicted attribute but only 14 of them are actually most important. Cleveland Clinic Foundation data set contain 303 patients record and Hungarian Institute of Cardiology data set contains 294 patient records. We integrate both of these datasets. So it means that there are total 597 instances in our dataset. The attributes of the dataset and their description is given in Table 1.

Table 1. Attributes and their Descriptions

Serial #	Attributes	Description	Values	Type
1	Age	Patient's age in years		Numeric
2	Sex	Sex of Patient	Male, Female	Nominal
3	ch_pain	chest pain	Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic	Nominal
4	r_B_Pressure	resting blood pressure (in mm Hg on admission to the hospital)		Numeric
5	Chol	serum cholesterol in mg/dl		Numeric
6	f_B_sugar	fasting blood sugar > 120 mg/dl	1 = true 0 = false	Nominal
7	r_ECG_results	Resting electrocardiographic results	0: normal 1:having ST-T wave abnormality (T wave inversions and/or STelevation or depression of > 0.05 mV) 2: showing probable or definite left ventricularhypertrophyby Estes' criteria	Nominal
8	maxi_heart_rate	maximum heart rate achieved		Numeric
9	Exercise	exercise induced angina	1 : yes 0 : no	Nominal
10	Oldpeak	ST depression induced by exercise relative to rest		Numeric
11	Slope	the slope of the peak exercise ST segment	1: upsloping 2: flat 3: down sloping	Nominal
12	no_of_majorvessels	number of major vessels (0-3) colored by fluoroscopy		Numeric
13	Defectype	defect type	3 = normal 6 = fixed defect 7 = reversible defect	Nominal
14	Class	diagnosis of heart disease	no_heart_disease have_heart_disease	Nominal

6.2 FILLING THE MISSING VALUES

The dataset for our experiments had some missing records, those missing records were find and replaced with appropriate value using ReplaceMissingValues filter from Weka tool. This filter scans all the records and replaces the missing values by mean mode method.

6.3 ATTRIBUTE SELECTION DATA REDUCTION TECHNIQUE

Attribute selection or Feature selection is a data reduction technique that is applied to the dataset. This technique reduces the size of the dataset by removing the irrelevant or redundant attributes. Other than this attribute selection method has other benefits such as the reduce number of attributes helps to make patterns easier and understandable. The Selected Attributes in Weka using BestFirst search method is shown in Figure 1.

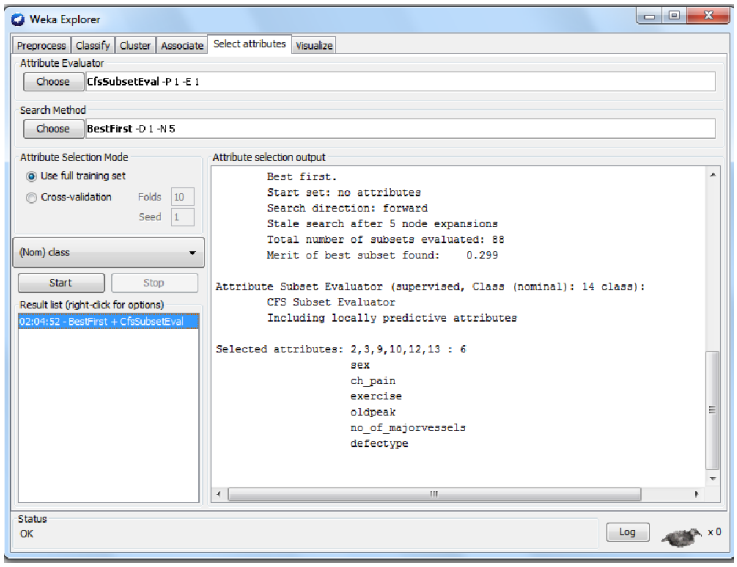

Fig. 2. Attribute Selection in Weka

For our experiments we will select CfsSubsetEval attribute filter. This filter evaluates the worth of subset of attributes in the given dataset by considering the individual predictive ability of each feature and with this filter we will use the Best First searching method. In this method, after each of the iteration the best of the original remaining attributes are added to the dataset. This method selects 7 attributes from the total of 14 attributes. These selected attributes are sex, ch_pain, exercise, oldpeak, no_of_majorvessols and defectype.

6.4 WEKA MACHINE LEARNING SOFTWARE

WEKA (Waikato Environment for Knowledge Analysis) developed at the University of Waikato, New Zealand which is a popular suite of machine learning software. Weka [13] is freely available software under the GNU General Public License that is written in java.

Weka workbench contains a collection of visualization tools and state of the art machine learning algorithms for data analysis and modeling. Different data mining tasks are supported by Weka these tasks are data preprocessing, clustering, classification, regression and visualization. Weka support data set in ARFF (Attribute-Relation File Format) format. ARFF file has two section first section contains Header information and second section contains Data information.

6.5 ALGORITHMS USED FOR EXPERIMENTS

6.5.1 DECISION TREE

Decision Tree is a flowchart like tree structure which contains leafs, nodes and branches [14]. Decision Tree has become popular in knowledge discovery because the construction of decision tree classifier does not require any domain knowledge. Successful decision tree model depends upon the data but in general it has good accuracy. The sample decision tree is shown in Figure 3.

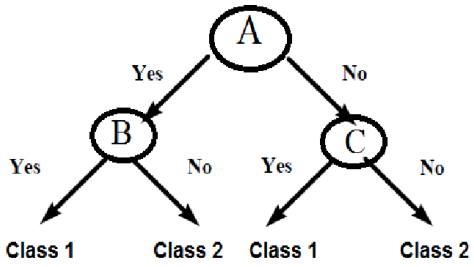

Fig. 3. Sample Decision Tree

J-48 is a type of decision tree algorithm that generate pruned and un-pruned decision tree for classification of data. The un-pruned trees are simply the larger in size where as pruning is a feature of J-48 decision tree that used to correctly classifying the subsets in the dataset. When a decision tree is built many branches shows anomalies in data due to outliers and tree pruning address this problem in data. This method removes the least reliable branches. It follows the facts in which each attribute of data can be used by splitting the data into subsets [15]. It uses the concept of information entropy. To make the decision the attribute with highest information gain is used and information gain is basically the difference in entropy. This algorithm can handle both continuous and continual attributes.

6.5.2 ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks are the human neurons type network structure which consists of number of nodes that are connected through directional links where each node represents a processing unit and the links between them specify the casual relation between them. This classification technique is becoming powerful tool in data mining and may be used for different purposes in descriptive and predictive data mining. The sample artificial neural network is shown in Figure 4.

.

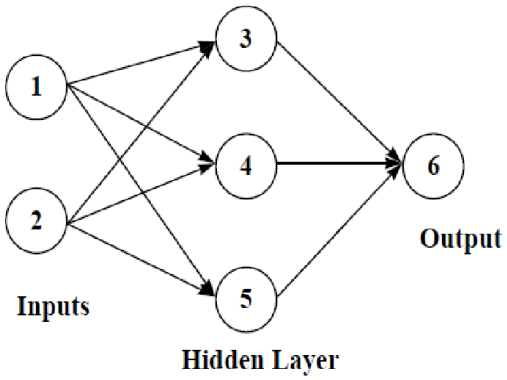

Fig. 4. Sample Artificial Neural Network

Many studies [16] shows that Artificial Neural Networks used in clinical decision making and helps the doctors to analyze and make sense if complex clinical data and medical applications. A Neural Network start with an input layer where each node is corresponds to a predictor variable [17]. The nodes of input layer are connected to the nodes of hidden layers and the nodes of hidden layer may be connects to each other or to an output layer. Multilayer Perceptron is a feed forward neural network with one input and output layer with several possible hidden layers that are totally interconnected.

6.5.3 BAYESIAN CLASSIFIERS

Bayesian classifiers are the classifiers that based on Bayes theorem [18]. They can predict about class probabilities such as which record is belong to a particular class and shows high speed and accuracy when applied to large databases.

Naïve Bayes algorithm is easy to construct and easy to interpret. It generally performs well as compare to other methods. Naïve Bayes classifier is based on conditional probability rule. It uses all the attributes in given dataset and analysis them individually.

7 EXPERIMENTAL SETUP

For our experiments we use Weka machine learning tool and we will select EXPLORER tab. Different experiments are conducted on full dataset. This dataset contains 597 unique attributes. We do experiments under two different scenarios, the first scenario is that in which all the attributes are used for experiments and the second scenario is that in which some most suitable attributes are selected using attribute selection method. The Visualization of the Dataset is shown in Figure 5.

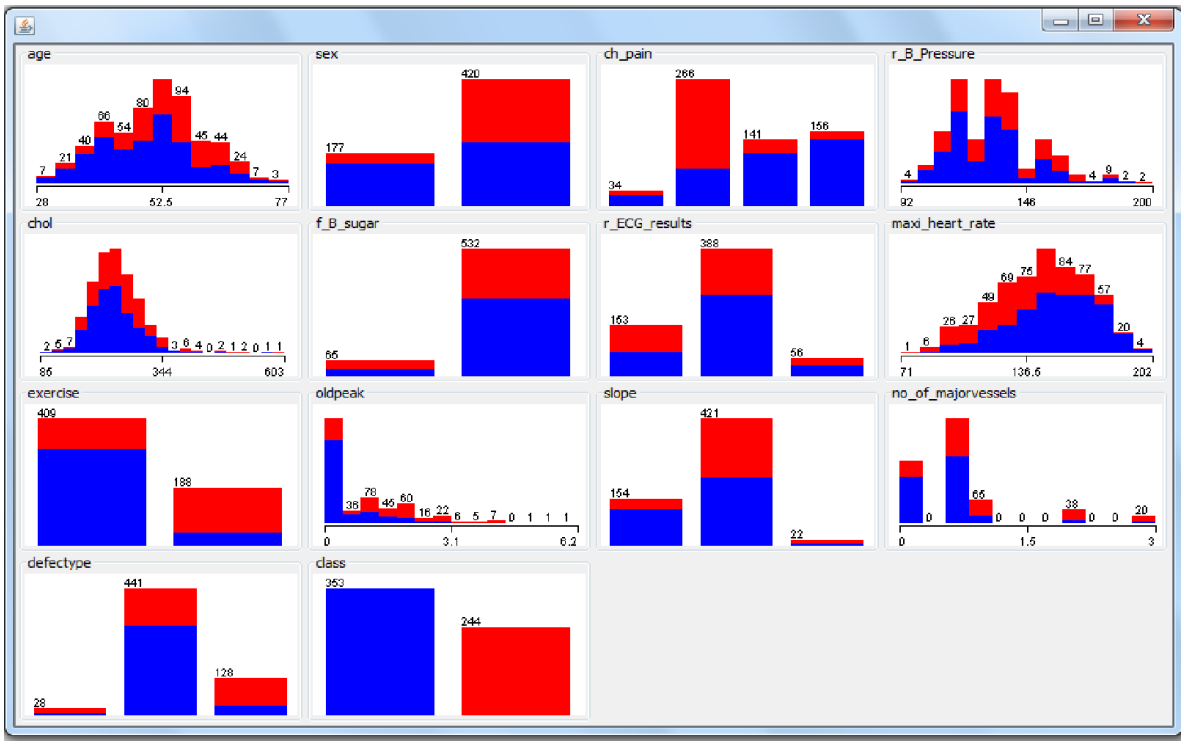

Fig. 5. Visualization of the Dataset

Mainly three algorithms are used in our bench mark data for experiment on both of the scenarios that are Decision Tree, Neural Network and Naïve Bayes. For sampling the training and testing dataset we use 10-fold cross validation. The performance and the accuracy of each experiment is evaluated by standard metrics such as TP rate, TN rate, precision, recall and F-measure which are calculated by Confusion Matrix which is known as predictive classification table. All these measures will be used to compare the performance of these selected and implemented algorithms.

8 DISCUSSION

The focus of our study was on using data mining in healthcare for heart diseases. For this purpose we do some experiments on our data about heart disease patients by applying different data mining algorithms. By implementing different classification algorithms we try to find out which one from the following algorithms is best in predicting about heart diseases.

There are four experiments were performed and these experiment are designed for the different purposes, these purposes are to investigate the difference between pruned and un-pruned decision tree, the affects of attribute selection method on implemented algorithms and to compare the results of J-48 Decision Tree, Neural Networks and Naïve Bayes.

8.1 COMPARISON OF IMPLEMENTED ALGORITHMS

After the experiments the next step is to compare algorithms used in these experiments and select the best one, as we already know every experiment conducted in two different scenarios, in first scenario all the attributes were used and in second scenario only the selected attributes were used. To compare these experiments different performance measures are taken into consideration like accuracy, TN and TP rate, ROC curve and the time taken to built a model. The summery of implemented algorithms is given in Table 2.

Fig. 6. Summary of Implemented Algorithms Performance

Algorithm	Accuracy (%)	TP rate	FP rate	Precision	F-measure	ROC Curve	Time (sec)
J48 decision tree (pruned) with all attributes	78.057	0.781	0.253	0.779	0.778	0.786	0.11
J48 decision tree (pruned) with selected attributes	78.727	0.787	0.246	0.786	0.785	0.778	0.04
J48 decision tree (un-pruned) with all attributes	77.219	0.772	0.257	0.770	0.770	0.788	0.03
J48 decision tree (un-pruned) with selected attributes	79.062	0.791	0.233	0.789	0.789	0.821	0.01
Multilayer perceptron with all attributes	79.89	0.799	0.214	0.799	0.799	0.867	8.24
Multilayer perceptron with selected attributes	80.402	0.804	0.218	0.803	0.803	0.870	3.73
Naïve Bayes with all attributes	82.914	0.829	0.194	0.828	0.828	0.898	0.03
Naïve Bayes with selected attributes	82.077	0.821	0.197	0.820	0.820	0.898	0.01

In general the results of all the implemented algorithms with selected attributes are much better than algorithms with all attributes except Naïve Bayes. The table above presents the great accuracy of implemented algorithms on the given heart disease dataset, the lowest accuracy is 77.219% and the highest accuracy is 82.914%. The Naïve Bayes classifier algorithm with all attributes shows the highest accuracy i.e. 82.914 % and Navie Bayes with selected attributes is nearest to it with 82.077 % accuracy. On the other hand, J-48 decision tree (un-pruned) with all attributes score the lowest accuracy i.e. 77.219 %.

If we look at the time (sec) taken to built a model for selected algorithm we will find that the Neural Networks takes the longest time while Decision Tree and Naïve Bayes takes shortest time in building the model.

Other performances measures like TP rate and FP rate that are also used compare the results also achieve remarkable performance and are shows very close difference. The TP rate and the FP rate were, (0.781, 0.253) for J-48 (pruned) with all attributes, (0.787, 0.246) for J-48 (pruned) with selected attributes, (0.772, 0.257) for J-48 (un-pruned) with all attributes, (0.791, 0.233) for J-48 (un-pruned) with selected attributes, (0.799, 0.214) for multilayer perceptron with all attributes,(0.804, 0.218)for multilayer perceptron with selected attributes, (0.829, 0.194) for naïve bayes with all attributes, and (0.821, 0.197) for naïve bayes with selected attributes respectively.

This shows that the Naïve Bayes with all attributes score the highest TP rate i.e. 0.829 while J-48 (un-pruned) Decision Tree with all attributes score the lowest TP rate. We compare the entire TP rate and FP rate scored by all the algorithms we found that all of these algorithms were better in predicting positive cases as TP rate in them is always greater than FP rate.

Precision score and F-measure of all the algorithms were quite balance, the highest precision scored are 0.828 and 0.820 by Naïve Bayes with all attributes and Naïve bayes with selected attributes respectively while Multilayer Preceptron with selected attributes was closet by 0.803 precision score. If we look at ROC curves we found that the Multilayer Precptron and Naïve Bayes are relatively close and accurate algorithms. ROC curve value for Naïve Bayes classifier was 0.898 which is nearest the “Perfect Classification Point” i.e. 1.

Based on above results and comparisons we found that the Naïve Bayes performs the highest Accuracy, TP rate, Precision, F-measure and ROC curve value. Naïve Bayes also score the fastest execution time as compare to other algorithms. Multilayer Perceptron take the longest time to build a model.

We know heart disease is a fatal disease which causes the death of millions people annually so there is need to keep the number of True Positives high and number of False Positives low. Early diagnosis is the key factor for a successful treatment of a disease, therefore classification algorithms are expected to perform well and emphasis is given to select the algorithms having high TP rate. Accuracy is also worth considering that is accurately identifying heart patients as much as possible.

Based on these results and comparisons we will select Naïve Bayes classifier with all algorithms is selected algorithms to predict the heart diseases for our study. The experiments shown that Naïve Bayes algorithm outclass the Decision Tree and Neural Networks in the domain of predicting of heart diseases.

8.2 Effect of Attribute Selection Method

Our all the experiments are consists of two scenarios, one scenario with all 14 attributes and the other scenario with 8 selected attributes. Attribute selection method helps in increase of classification accuracy and decrease the time and complexity by ignoring irrelevant attributes from given dataset. The results of experiments shows that the accuracy of the implemented algorithms increased and the execution time was decreased by using attribute selection method.

8.3 Selected Algorithm based on Performance

Different objectives were defined at the start of our research. Those were evaluated against our selected algorithm. Results of the experiments show that the selected algorithms had achieved the highest score which suggest that this algorithm better in heart disease diagnosis.

9 Future Recommendation

Our research indicates that data mining can be used and applied in healthcare industry to predict about heart diseases and implemented algorithms are worth considering. Further research should be conducted to increase the classification accuracy by using different classification algorithms such as Bagging, Support Vector Machine or Decision Table etc.

In our experiments we do not change the default parameters of algorithms, in future this study can be enhanced and expand by changing the parameters for experiments. The patterns and relationships found in heart disease dataset can be used to design a complete Knowledge Based system.

In future, more work can be done by using more data set related to heart diseases and by using different data reduction methods to improve the classification. For better accuracy and prediction of heart diseases the datasets that will be used must be quality oriented and free from outliers, inconsistencies and missing values.

10 Conclusion

Our study was focused on the use of data mining techniques in healthcare specifically in Heart Diseases. Heart disease is a fatal disease which may cause life threatening complications such as death. We use online available heart patient's data from UCI repository. There were 597 unique instances in our data set. The classification which is a data mining technique was implemented with following algorithms, Decision Tree, Neural Networks and Naïve Bayes. Some important point were considered to choose suitable tool for mining, on the basis of them Weka machine learning software were used for experiments. To evaluate the performance of the algorithms different performance metrics were considered that are accuracy, precision, F-measure, ROC curve value, TP rate and FP rate.

Four experiments were conducted in two different scenarios, in first scenario all attributes were used and in second scenario selected attributes were used and data set was in ARFF format that is supported by Weka. The experiments show that Naïve Bayes classification algorithms have the highest accuracy among all that is 82.914%. This study shows that the data mining can be used to predict about heart disease efficiently and effectively. The results or the outcomes of our thesis may be used as assistant tool to help in making more consistent diagnosis of heart diseases.

References

[1] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques 2nd Edition" The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, March 2006.

[2] WORLD HEALTH RANKING 2014 [online] available: http://www.worldlifeexpectancy.com/cause-of-death/coronary-heart-disease/by-country/ (May 7, 2014).

[3] WORLD HEALTH ORGANIZATION 2014 [online] available: http://www.who.int/cardiovascular_diseases/en/ (May 7, 2014).

[4] Hian Chye Koh and Gerald Tan, "Data Mining Applications in Healthcare" Journal of Healthcare Information Management — Vol. 19, No. 2, 2005.

[5] Greg Rogers and Ellen Joyner, "MINING YOUR DATA FOR HEALTH CARE QUALITY IMPROVEMENT", SAS Institute, Inc., Cary, NC, January 1, 2009.
[6] M. Anbarasi, E. Anupriya and N.Ch.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm" International Journal of Engineering Science and Technology Vol. 2(10), 2010.
[7] AH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin, "HDPS: Heart Disease Prediction System", IEEE-Computing in Cardiology, September 2011.
[8] Chaitrali S. Dangare and Sulabha S. Apte, "A DATA MINING APPROACH FOR PREDICTION OF HEARTDISEASE USING NEURAL NETWORKS" International Journal Of Computer Engineering & Technology, Volume 3, Issue 3, 2012.
[9] Jyoti Soni, Uzma Ansari and Dipesh Sharma, "Intelligent and Effective Heart Disease Prediction System using WeightedAssociative Classifiers" International Journal on Computer Science and Engineering, Vol. 3 No. 6 June 2011.
[10] Vikas Chaurasia and Saurabh Pal, "Data Mining Approach to Detect Heart Dieses", International Journal of Advanced Computer Science and Information Technology Vol. 2, No. 4, 2013.
[11] S.Saravanakumar and S.Rinesh, "Effective Heart Disease Prediction usingFrequent Feature Selection Method" International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, March 2014.
[12] UCI (University of California, Irvine C.A) Machine Learning Repository, 2014 [online] available: http://repository.seasr.org/Datasets/UCI/arff/ (June 1, 2014).
[13] Remco R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald and David Scuse "WEKA Manual for Version 3-6-8" THE UNIVERSITY OF WAIKATO, August 13, 2012.
[14] Larose, D., "Discovering Knowledge in Data: An Introduction to Data Mining." New Jersey: John Wiley & Sons, Inc, 2005.
[15] Quinlan Ross J.,"C4.5: Programs for Machine Learning". Morgan Kaufmann Publishers, San Mateo, CA.1993.
[16] Qeethara Kadhim Al-Shayea, "Artificial Neural Networks in Medical Diagnosis", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.
[17] Two Crows Corporation,"Introduction to Data Mining and Knowledge Discovery. Third Edition." Two Crows Corporation. 10500 Falls Road, Potomac, USA, 2005.
[18] George John and Pat Langley. "Estimating continuous distributions in Bayesian classifiers" In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, pages 338–345, 1995.

AUTHOR'S BIOGRAPHY

UMAIR SHAFIQUE- Received M.Sc Degree in Information Technology, from University of Gujrat, Gujrat Pakistan in 2014. He is the member of this research work "Data Mining in Healthcare for Heart Diseases".

FIAZ MAJEED- Received MS degree from COMSATS Institute of Information Technology (CIIT) Lahore Pakistan in 2009. He is currently PhD scholar in University of Engineering and Technology (UET) Lahore Pakistan. Further, he is Lecturer in University of Gujrat (UOG), Gujrat, Pakistan and working on couple of research projects. His research interests include data warehousing, data mining, data streams and information retrieval. He has published more than 10 research papers in refereed journals and international conference proceedings in the above areas. He is the supervisor of this project and this paper is part of that research project "Data Mining in Healthcare for Heart Diseases".

HASEEB QAISER- Received M.Sc Degree in Information Technology, from University of Gujrat, Gujrat Pakistan in 2014. He is the member of this research work "Data Mining in Healthcare for Heart Diseases".

IRFAN_UL_MUSTAFA- Received M.Sc Degree in Information Technology, from University of Gujrat, Gujrat Pakistan in 2014. He is the member of this research work "Data Mining in Healthcare for Heart Diseases".