Data Mining Techniques to Find Out Heart Diseases: An Overview

Aqueel Ahmed, Shaikh Abdul Hannan

Abstract : Heart disease is a major cause of morbidity and mortality in modern society. Medical diagnosis is extremely important but complicated task that should be performed accurately and efficiently. Although significant progress has been made in the diagnosis and treatment of heart disease, further investigation is still needed. The availability of huge amounts of medical data leads to the need for powerful data analysis tools to extract useful knowledge. There is a huge data available within the healthcare systems. However, there is a task of effective analysis tools to discover hidden relationships and trends in data. Knowledge discovery and data mining have found numerous application in business and scientific domain. Researchers have long been concerned with applying statistical and data mining tools to improve data analysis on large data sets. Disease diagnosis is one of the applications where data mining tools are proving successful results. This research paper proposed to find out the heart diseases through data mining, Support Vector Machine (SVM), Genetic Algorithm, rough set theory, association rules and Neural Networks.

In this study, we briefly examined that out of the above techniques Decision tree and SVM is most effective for the heart disease. So it is observed that, the data mining could help in the identification or the prediction of high or low risk heart diseases.

Keywords : Data Mining, Heart Disease, SVM, rough sets techniques, association rules & clustering.

I. INTRODUCTION

1.1 Overview of Data Mining

Knowledge discovery in databases is well-defined process consisting of several distinct steps. Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. A formal definition of Knowledge discovery in databases is given as follows: "Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data" [17]. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the healthcare administrators to improve the quality of service. The discovered knowledge can also be used by the medical practitioners to reduce the number of adverse drug effect, to suggest less expensive therapeutically equivalent alternatives. Anticipating patient's future behavior on the given history is one of the important applications of data mining techniques that can be used in health care management.

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Health care data is massive. It includes patient centric data, resource management data and transformed data. Health care organizations must have ability to analyze data. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care.

The availability of integrated information via the huge patient repositories, there is a shift in the perception of clinicians, patients and payers from qualitative visualization of clinical data by demanding a more quantitative assessment of information with the supporting of all clinical and imaging data. For instance it might now be possible for the physicians to compare diagnostic information of various patients with identical conditions. Likewise, physicians can also confirm their findings with the conformity of other physicians dealing with an identical case from all over the world [18]. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial.

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et al proposed that integration of clinical decision support with computer based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome [19]. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making.

Data mining is a process of extraction of useful information and patterns from huge data.

Revised Manuscript Received on September, 2012.

Aqueel Ahmed, Department of Computer Science,Vasantrao Naik Mahavidyalaya, Aurangabad (M.S.).

Shaikh Abdul Hannan, Department of Computer Science, Vivekanand College, Aurangabad (M.S.).

It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown.

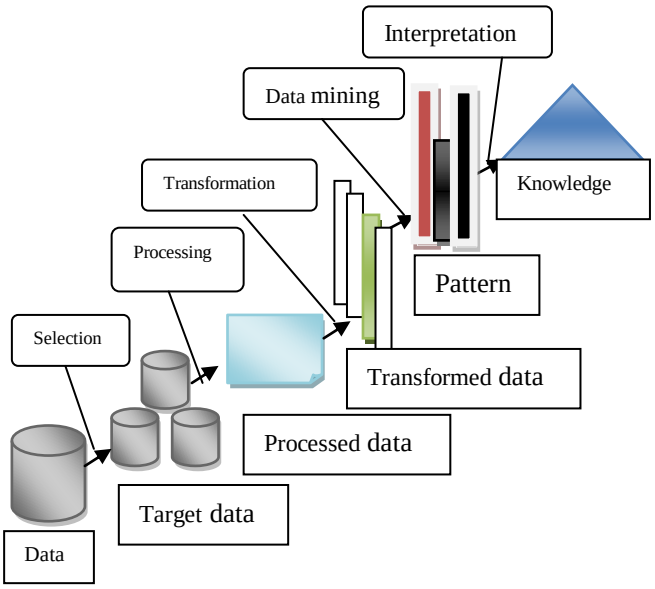

Fig. 1 : KDD Process

Once these patterns are found they can further be used to make certain decisions for development of their businesses.

Three steps involved are

- Exploration
- Pattern identification
- Deployment

Exploration: In the first step of data exploration data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.

Pattern Identification**:** Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction. Deployment: Patterns are deployed for desired outcome.

1.2 Causes and impact of heart diseases

According to WHO report Global atlas on cardiovascular disease prevention and control states that cardiovascular disease (CVDs) are the leading causes of death and disability in the world. Although a large proportion of CVDs is preventable, they continue to rise mainly because preventive measures are inadequate. Over 17.3 million An estimated 17.3 million people died from CVDs in 2008, Over 80% of CVD deaths take place in low- and middle-income countries, 23.6 million By 2030, almost 23.6 million people will die from CVDs.

1.2.1 Protect heart health

- Tobacco use, an unhealthy diet, and physical inactivity increase the risk of heart attacks and strokes.
- Engaging in physical activity for at least 30 minutes every day of the week will help to prevent heart attacks and strokes.
- Eating at least five servings of fruit and vegetables a day, and limiting your salt intake to less than one teaspoon a day, also helps to prevent heart attacks and strokes.

1.2.2 Cardiovascular Diseases (Cvds) Key Facts

- CVDs are the number one cause of death globally: more people die annually from CVDs than from any other cause.
- An estimated 17.3 million people died from CVDs in 2008, representing 30% of all global deaths. Of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to stroke.
- Low- and middle-income countries are disproportionally affected: over 80% of CVD deaths take place in low- and middle-income countries and occur almost equally in men and women.
- By 2030, almost 23.6 million people will die from CVDs, mainly from heart disease and stroke. These are projected to remain the single leading causes of death.

1.2.3 Cardiovascular Diseases

Cardiovascular disease is caused by disorders of the heart and blood vessels, and includes coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. The major causes of cardiovascular disease are tobacco use, physical inactivity, an unhealthy diet and harmful use of alcohol. These are the three causes of heart diseases (1) chest pain (2) stroke and (3) heart attack.[14]

To prevent and identification of these diseases different techniques of data mining is used through this easily find out heart related diseases and this is the aim of this research studies.

Heart disease is the leading cause of death all over the world in the past ten years. Several researchers are using statistical and data mining tools to help health care professionals in the diagnosis of heart disease. [11]

II. DATA MINING ALGORITHMS AND TECHNIQUES

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

2.1. Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules.

If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis.

The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

2.2. Clustering

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification. For example, to form group of customers based on purchasing patterns, to categories genes with similar functionality.

Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative(divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

2.3. Predication

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

2.4. Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

2.5. Neural networks

Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. For example handwritten character reorganization, for training a computer to pronounce English text and many real world business problems and have already been successfully applied in many industries.

Shaikh Abdul Hannan et.al describes aiming to develop a expert system for diagnosing of heart disease using support vector machine and feed-forward back-propagation technique. Now a days neural network are being used successfully in an increasing number of application areas. This work includes the detailed information about patient and reprocessing was done. The Support Vector Machine (SVM) and feed-forward Back-propagation technique have been applied over the data for the expert system. [20].

Shen, Z., Clarke et.al present possible interpretations of the weights of these neurons and show how they can be used as a selection criteria for which questions to use as inputs. The technique is compared against other statistical methods [15]. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs. [1]

III. SURVEY OF LITERATURE (DIFFERENT DATA MINING TECHNIQUES TO FIND OUT HEART DISEASES)

3.1 Decision Tree Classification Algorithm

Heart disease or coronary artery disease (CAD) or coronary heart disease (CHD) or ischemic heart disease (IHD)[4] is a broad term that can refer to any condition that affects the heart. For developing clinical decision support systems, literature presents a number of researches that have made use of artificial intelligence and data mining techniques. Till now, several studies have been reported on heart disease diagnosis.

These studies have applied different approaches to the given problem and achieved high classification accuracies, of 77% or higher, using the dataset taken from the UCI machine learning repository. Experimental results showed a correct classification accuracy of approximately 77% with a logistic-regression-derived discriminate function. The John Gennari LASSIT conceptual clustering system achieved a 78.9% accuracy on the Cleveland database. A Fuzzy Support Vector Clustering to identify heart disease was used in. Resul Das introduced a methodology that uses SAS base software 9.13 for diagnosing heart disease. Zheng Yao applied a new model called R-C4.5 which improved the efficiency of attribution selection and partitioning models. Gang Kou partition applied data separation-based techniques to preserve privacy in the classification of medical data.

Karolis M.A. et.at has developed a data mining system for the assessment of heart event related risk factors targeting in the reduction in the reduction of CHD events using the risk factors i) Before the event a) non modifiable – age, sex etc. The analysis was carried out using C4.5 decision tree algorithm for the mentioned three events using five different splitting criteria. The highest percentage of correct classification archived were 66%, 75% and 75% for the myocardial Infarction (MI) Percutaneous Coronary Intervention (PCI) and Coronary Artery Bypass Graft Surgery (CABG) [16].

CHD have reached epidemic proportions among Indians. India is undergoing a rapid health transition with rising burden of CHD. Further, the long-term case fatality following acute coronary syndrome is considerably higher among Indians as compared to other populations. In addition, a reversal of socio-economic gradients for CHD risk factors has emerged in the Indian population. In this work, we have identified a system for automated medical diagnosis of heart disease risk using decision tree classifier. The success of population based interventions, addressing multiple risk factors for CHDs, through lifestyle linked community programmes was demonstrated initially in North Karelia study2.

In developing countries such as India such measures may indeed work due to several reasons. First, the risk factor levels are high among Indians conferring a higher risk. Interventions are likely to have a higher impact on high risk population [8]. CVD are the leading cause of death and disability in both developed and developing countries. A paradigm shift away from the biomedical model is therefore required in the perspective of the existing health care system while responding to the rapidly increasing burden of CVD morbidity and mortality in India [6]. Uneducated and less educated people in rural India have a higher prevalence of coronary heart disease and of the coronary risk factors smoking and hypertension [5].Analysis of data suggest that the risk for CVD and stroke is at epidemic proportions in a cohort of well-educated physicians who are in the highest quintile of income [7,10]. CVD affects people of all income levels[5,7,10].

3.2. UCI Database Description about Decision Tree Classification

The heart disease database from the University of California Irvine. UCI archive is used. This database contains four data sets from the Cleveland Clinic Foundation, Hungarian Institute of Cardiology, V.A. Medical Center and University Hospital of Switzerland. It provides 920 records in total. Originally, the database had 76 raw attributes. However, all of the published experiments only refer to 13 of these: Age, Sex, P, Trstbps, Chol, Fbs, estecg, Thalach, Exang, OldPeak, Slope, Ca, Thal and Num.[10]

3.3 Clustering D.M. Technique Using K- Means Algorithms

The categorization of objects into various groups or the partitioning of data set into subsets so that the data in each of the subset share a general feature, frequently the proximity with regard to some defined distance measure, is known as Clustering. The clustering problem has been identified in numerous contexts and addressed being proven beneficial in many medical applications. Clustering the medical data into small with meaningful data can aid in the discovery of patterns by supporting the extraction of numerous appropriate features from each of the clusters thereby introducing structure into the data and aiding the application of conventional data mining techniques. Numerous methods are available in the literature for clustering and employed the renowned K-Means clustering algorithm in this approach. The k-means algorithm is one of the widely recognized clustering tools that are applied in a variety of scientific and industrial applications. k-means groups the data in accordance with their characteristic values into k distinct clusters. Data categorized into the same cluster have identical feature values. k, the positive integer denoting the number of clusters, needs to be provided in advance. The steps involved in a k-means algorithm are given subsequently:

Prediction of heart disease using K – Means clustering technique

1. K points denoting the data to be clustered are placed into the space. These points denote the primary group centurions.
2. The data are assigned to the group that is adjacent to the centurion.
3. The positions of all the K centroids are recalculated as soon as all the data are assigned.
4. Steps 2 and 3 are reiterated until the centroids stop moving any further. This results in the segregation of data into groups from which the metric to be minimized can be deliberated.

The preprocessed heart disease data is clustered using the Kmeans algorithm with the K values. Clustering is a type of multivariate statistical analysis also known as cluster analysis, unsupervised classification analysis, or numerical taxonomy. K-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. It is well suited to generating globular clusters. The K-Means method is numerical, unsupervised, non-deterministic and iterative.

3.4 K-Means and derivatives

The k-Means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster — that is, its

coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. Example: The data set has three dimensions and the cluster has two points X = (x1, x2, x3) and Y = (y1, y2, y3). Then Z becomes Z = (z1, z2, z3), where z1= x1+y1 , z2= x2+y2 and z3= x3+y3

3.5 Advantages to using this technique

a. The main advantages of this algorithm are its simplicity and speed which allows it to run on large datasets.
b. With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small).
c. K-Means may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.[3]

3.6 Data Mining Through Genetic Algorithms

We start out with a randomly selected first generation. Every string in this generation is evaluated according to its quality, and a fitness value is assigned. Next, a new generation is produced by applying the reproduction operator. Pairs of strings of the new generation are selected and crossover is performed. With a certain probability, genes are mutated before all solutions are evaluated again. This procedure is repeated until a maximum number of generations are reached. While doing this, the all time best solution is stored and returned at the end of the algorithm.

Genetic algorithm have been used in [4], to reduce the actual data size to get the optimal subset of attributed sufficient for heart disease prediction. Classification is a supervised learning method to extract models describing important data classes or to predict future trends. Three classifiers e.g. Decision Tree, Naïve Bayes and Classification via clustering have been used to diagnose the presence of heart disease in patients. Pairs of strings of the new generation are selected and crossover is performed. With a certain probability, genes are mutated before all solutions are evaluated again. This procedure is repeated until a maximum number of generations are reached.[9]

3.7 Classification via clustering

Clustering is the process of grouping similar elements. This technique may be used as a preprocessing step before feeding the data to the classifying model. The attribute values need to be normalized before clustering to avoid high value attributes dominating the low value attributes. Further, classification is performed based on clustering. Experiments were conducted with Weka 3.6.0 tool. Data set of 909 records with 13 attributes. All attributes are made categorical and inconsistencies are resolved for simplicity. To enhance the prediction of classifiers, genetic search is incorporated. Observations exhibit that the Decision Tree data mining technique outperforms other two data mining techniques after incorporating feature subset selection but with high model construction time. Naïve Bayes performs consistently before and after reduction of attributes with the same model construction time. Classification via clustering performs poor compared to other two methods.

3.8 Association Rule Discovery

Association rules represent a promising technique to improve heart disease prediction. Unfortunately, when association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical., four constraints were proposed to reduce the number of rules: item filtering, attribute grouping, maximum item set size, and antecedent/consequent rule filtering. When association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical. A more important issue is that, in general, association rules are mined on the entire data set without validation on an independent sample. To solve these limitations, the author has introduced an algorithm that uses search constraints to reduce the number of rules, searches for association rules on a training set, and finally validates them on an independent test set. Instead of using only Support and confidence, one more parameter i.e. lift have been used as the metrics to evaluate the medical significance and reliability of association rules. Medical doctors use sensitivity and specificity as two basic statistics to validate results. Sensitivity is defined as the probability of correctly identifying sick patients, whereas specificity is defined as the probability of correctly identifying healthy individuals. Lift was used together with confidence to understand sensitivity and specificity. To find predictive association rules in a medical data set the algorithm has three major steps. First, a medical data set with categorical and numeric attributes is transformed into a transaction data set. Second, four constraints mentioned above are incorporated into the search process to find predictive association rules with medically relevant attribute combinations. Third, a train and test approach is used to validate association rules.

3.9 Rough Set Theory

The result of knowledge discovery process can be decision tree, association rules, decision rules, sequential pattern, etc. The most comprehensive and interpretable knowledge extracted is in the form of rules. Some rule induction algorithm such as rough set theory results in large number of rules. This large number makes interpretability of the knowledge becomes low. Lacking of interpretability will cut down the advantages of rule based systems. The resulting large number of rules is because of noise redundancy in input and/or training data sets. Rule pruning is the method to reduce the number of rules while maintaining the quality of the system. Rough set theory (RST) is a relatively new mathematical and artificial intelligent technique developed by Zdzislaw Pawlak, Warsaw University of Technology, in the early1980. RST is especially useful to discover relationships in data. The discovery of relationship in the data is called knowledge discovery or data mining. The result of knowledge discovery is understandable and meaningful knowledge from data. RST method emerged as mathematical tool to manage uncertainties, ambiguity and vagueness from incomplete, inexact and noisy information.

Through this technique it discovers rules evaluated with support, confidence, and lift. Association rules are applied on a real data set containing medical records of patients with heart disease.[12,13]

IV. CONCLUSION

This paper examines the classification techniques in data mining and shows the performance of classification among them. In these classification accuracy among these data mining techniques has discussed. The result shows the difference in error rates. However there are relatively differences in different techniques. Decision tree and SVM perform classification more accurately than the other methods. Data mining application in heart disease Author name et.al. reported that the major advantage of data mining technique shows the 92.1 % 91.0 % accuracy for the heart disease.

We suggest that the age, sex, chest pain, blood pressure, personnel history, previous history, cholesterol, fasting blood sugar, resting ECG, Maximum heart rate, slope, etc. that may be used as reliable indicators to predict presence of heart disease. We also suggest that data should be explored and must be verified from the team of heart disease specialist doctors.

In future, we will try to increase the accuracy for the heart disease patient by increasing the various parameters suggested from the doctors by using different data mining techniques.

REFERENCES

1. Mrs. Bharati M. Ramageri, "Data Mining Techniques And Applications",Bharati M. Ramageri / Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305
2. Heart disease from "http://www.nhlbi.nih.gov/educational/hearttruth/lower-risk/what-is-heart-disease.htm"
3. Bala Sundar V, Bharathiar, "Development of a Data Clustering Algorithm for Predicting Heart" International Journal of Computer Applications (0975 – 888) Volume 48– No.7, June 2012
4. http://heart-disease.emedtv.com/ coronary-artery-disease/coronary-artery-disease.html
5. R. Gupta, V. P. Gupta, and N. S. Ahluwalia, "Educational status, coronary heart disease, and coronary risk factor prevalence in a rural population of India", BMJ. pp 1332–1336, 19 November 1994.
6. Panniyammakal Jeemon & K.S. Reddy, "Social determinants of cardiovascular disease outcomes in Indians", pp 617-622, November 2010.
7. A Mathavan, MD, A Chockalingam, PhD, S Chockalingam, BSc, B Bilchik, MD, and V Saini, MD, "Madurai Area Physicians Cardiovascular Health Evaluation Survey (MAPCHES) – an alarming status", The Canadian Journal of Cardiology; 25(5): 303–308, May 2009.
8. Vamadevan S. Ajay & Dorairaj Prabhakaran, "Coronary heart disease in Indians: Implications of the INTERHEART study", Indian J Med Res 132, pp 561-566, November 2010.
9. K.S.Kavitha,K.V. Ramakrishnan, Manoj Kumar Singh **"**Modeling and design of evolutionary neural network for heart disease detection" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, 1694-0814, September 2010.
10. Rajeswari K, Vaithiyanathan V, P. Amirtharaj "Application of Decision Tree Classifiers in Diagnosing Heart Disease using Demographic Data" American Journal of Scientific research ISSN 2301-2005 pp. 77-82 EuroJournals Publishing, 2012.
11. Yanwei Xing "Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease" IEEE Transactions on Convergence Information Technology, pp(868 – 872), 21-23 Nov. 2007
12. Ordonez C," Association rule discovery with the train and test approach for heart disease prediction", IEEE Transactions on Information Technology in Biomedicine, P(334 – 343), April 2006
13. AbuKhousa, E "Predictive data mining to support clinical decisions: An overview of heart disease prediction systems", IEEE Transaction on Innovations in Information Technology (IIT), pp(267 - 272) March 2012.
14. Srinivas, K.,"Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", IEEE Transaction on Computer Science and Education (ICCSE), p(1344 - 1349), 2010.
15. Shen, Z., Clarke, M. Jones, R. Alberti, T. "A neural network approach to the detection of coronary artery disease",IEEE Transaction on Computers in Cardiology, 5-8 Sep 1993.
16. Karaolis, M.A. Moutiris, J.A. Hadjipanayi, D. Pattichis, C.S., "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees" IEEE Transactions on Information Technology in Biomedicine, pp 559 -566, May 2010.
17. Frawley and Piatetsky-Shapiro, 1996. Knowledge Discovery in Databases:An Overview. The AAAI/MIT Press, Menlo Park, C.A.
18. Miller, A., B. Blott and T. Hames, 1992. Review of neural network applications in medical imaging and signal processing. Med. Biol. Engg. Comp., 30: 449-464.
19. Chen, J., Greiner, R.: Comparing Bayesian Network Classifiers. In Proc. of UAI-99, pp.101–108,1999.
20. Shaikh Abdul Hannan, V. D. Bhagile R. R. Manza, R. J. Ramteke, "Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feedforward Backpropagation technique", International Journal on Computer Science and Engineering, pp 2150-2159, 2010.