

Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction

Jyoti Soni
Student, M.Tech (CSE).
Raipur Institute of Technology
Raipur, Chhattisgarh, India

Ujma Ansari
Professor
Raipur Institute of Technology
Raipur, Chhattisgarh, India

Dipesh Sharma
Reader
Raipur Institute of Technology
Raipur, Chhattisgarh, India

Sunita Soni
Sr. Associate Professor
Bhilai Institute of Technology,
Durg-491 001, Chhattisgarh, India

ABSTRACT

The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still 'information rich' but 'knowledge poor'. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. This research paper intends to provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. Number of experiment has been conducted to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The second conclusion is that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

Keywords

KNN, Neural Networks, Bayesian classification, Classification based on clustering, Decision Tree.

1. INTRODUCTION

Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis. However, the available raw medical data are widely distributed, heterogeneous in nature, and voluminous. These data need to be collected in an organized form. This collected data can be then integrated to form a hospital information system. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data.

The World Health Organization has estimated that 12 million deaths occurs worldwide, every year due to the Heart diseases. Half the deaths in the United States and other developed countries occur due to cardio vascular diseases. It is also the chief reason of

deaths in numerous developing countries. On the whole, it is regarded as the primary reason behind deaths in adults. The term Heart disease encompasses the diverse diseases that affect the heart. Heart disease was the major cause of casualties in the different countries including India. Heart disease kills one person every 34 seconds in the United States. Coronary heart disease, Cardiomyopathy and Cardiovascular disease are some categories of heart diseases. The term "cardiovascular disease" includes a wide range of conditions that affect the heart and the blood vessels and the manner in which blood is pumped and circulated through the body. Cardiovascular disease (CVD) results in several illness, disability, and death. The diagnosis of diseases is a vital and intricate job in medicine.

Medical diagnosis is regarded as an important yet complicated task that needs to be executed accurately and efficiently. The automation of this system would be extremely advantageous. Regrettably all doctors do not possess expertise in every sub specialty and moreover there is a shortage of resource persons at certain places. Therefore, an automatic medical diagnosis system would probably be exceedingly beneficial by bringing all of them together. Appropriate computer-based information and/or decision support systems can aid in achieving clinical tests at a reduced cost. Efficient and accurate implementation of automated system needs a comparative study of various techniques available. This paper aims to analyze the different predictive/ descriptive data mining techniques proposed in recent years for the diagnosis of heart disease.

2. METHODOLOGY

Due to resource constraints and the nature of the paper itself, the main methodology used for this paper was through the survey of journals and publications in the fields of medicine, computer science and engineering. The research focused on more recent publications.

3. RESEARCH FINDINGS

3.1 Data Mining in the Heart Disease Prediction.

Three different supervised machine learning algorithms i.e. Naive Bayes, K-NN, Decision List algorithm have been used for

analyzing the dataset in [1]. Tanagra tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared. **Tanagra** is a data mining suite build around graphical user interface algorithms. The main purpose of Tanagra project is to give researchers and students an easy-to-use data mining software, and allowing to analyze either real or synthetic data. Tanagra is powerful system that contains clustering, supervised learning, meta supervised learning, feature selection, data visualization supervised learning assessment, statistics, feature selection and construction algorithms. **Decision Tree** is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high dimensional data. The results obtained from Decision Trees are easier to read and interpret. The drill through feature to access detailed patients' profiles is only available in Decision Trees. **Naïve Bayes** is a statistical classifier which assumes no dependency between attributes. It attempts to maximize the posterior probability in determining the class. The advantage of using naive bayes is that one can work with the naive Bayes model without using any Bayesian methods. Naive Bayes classifiers have works well in many complex real-world situations. The *k*-nearest neighbor's algorithm (**k-NN**) is a method for classifying objects based on closest training data in the feature space. *k*-NN is a type of instance-based learning. The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. But the accuracy of the *k*-NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance. The experiment is performed using training data set consists of 3000 instances with 14 different attributes. The dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing. Based on the experimental results shown in Table 1, it is clear that the classification accuracy of Naive Bayes algorithm is better compared to other algorithms.

Table 1. Performance Study of Algorithm

| Algorithm Used | Accuracy | Time Taken |
|----------------|----------|------------|
| Naive Bayes | 52.33% | 609ms |
| Decision List | 52% | 719ms |
| KNN | 45.67% | 1000ms |

3.1.1 Data Mining and Artificial Neural network

Intelligent Heart Disease Prediction System (IHDP) using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Network. is implemented in [9] using .NET platform . IHDP is Web-based, user-friendly, scalable, reliable and expandable system. It can also answer complex "what if" queries which traditional decision support systems cannot. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease. As a **Data source** a total

of 909 records with 15 medical attributes (factors) were obtained from the Cleveland Heart Disease database. Figure 1 lists the attributes. The records were split equally into two datasets: training dataset (455 records) and testing dataset (454 records) Table 2 summarizes the results of all three models. Naïve Bayes appears to be most effective as it has the highest percentage of correct predictions (86.53%) for patients with heart disease, followed by Neural Network (with a difference of less than 1%) and Decision Trees. Decision Trees, however, appears to be most effective for predicting patients with no heart disease (89%) compared to the other two models.

Table 2. Results Obtained from IHDP

| Techniques | Accuracy |
|---------------|------------|
| Naive Bayes | 86.53 % |
| Decision Tree | 89% |
| ANN | 85.53 % |

1. Diagnosis (value 0: < 50% diameter narrowing (no heart disease); value 1: > 50% diameter narrowing (has heart disease))
- Key attribute**
1. PatientID – Patient's identification number
- Input attributes**
1. Sex (value 1: Male; value 0 : Female)
 2. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic)
 3. Fasting Blood Sugar (value 1: > 120 mg/dl; value 0: < 120 mg/dl)
 4. Restecg – resting electrographic results (value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy)
 5. Exang – exercise induced angina (value 1: yes; value 0: no)
 6. Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3: downsloping)
 7. CA – number of major vessels colored by floursopy (value 0 – 3)
 8. Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)
 9. Trest Blood Pressure (mm Hg on admission to the hospital)
 10. Serum Cholesterol (mg/dl)
 11. Thalach – maximum heart rate achieved
 12. Oldpeak – ST depression induced by exercise relative to rest
 13. Age in Year

Figure 1: List of Attributes

A proficient methodology for the extraction of significant patterns from the heart disease warehouses for heart attack prediction has been presented in [7]. Initially, the data warehouse is pre-processed in order to make it suitable for the mining process. Once the preprocessing gets over, the heart disease warehouse is clustered with the aid of the K-means clustering algorithm, which will extract the data appropriate to heart attack from the warehouse. Consequently the frequent patterns applicable to heart disease are mined with the aid of the MAFIA algorithm from the data extracted. In addition, the patterns vital to heart attack prediction are selected on basis of the computed significant

weightage. The neural network is trained with the selected significant patterns for the effective prediction of heart attack. Multi-layer Perceptron Neural Network with Back-propagation is being used as the training algorithm. In feed-forward neural networks the neurons of the first layer forward their output to the neurons of the second layer, in a unidirectional fashion, which explains that the neurons are not received from the reverse direction. A kind of feed-forward neural network mechanism is the Multi-layer Perceptron Neural Networks (MLPNN) or

Multilayer feed-forward neural network (MFNN). Figure 2 shows the structure of MLPNN. The significant parameters to heart attack prediction have been used with their weightage and the priority levels are advised by the medical experts. The sample combinations of heart attack parameters for normal and risk level along with their values and weightages are shown in Table. 3 In that, lesser value (0.1) of weightage comprises the normal level of prediction and higher values other than 0.1 comprise the higher risk levels.

Tables3: Heart attack parameters with corresponding values and their weightages

| Parameters | Weightage | |
|--------------------------------|--------------------|-----|
| Male and Female | Age < 30 | 0.1 |
| | >30 to <50 | 0.3 |
| | Age>50 and Age <70 | 0.7 |
| | Age>70 | 0.8 |
| Smoking | Never | 0.1 |
| | Past | 0.3 |
| | Current | 0.6 |
| Overweight | Yes | 0.8 |
| | No | 0.1 |
| Alcohol Intake | Never | 0.1 |
| | Past | 0.3 |
| | Current | 0.6 |
| High salt diet | Yes | 0.9 |
| | No | 0.1 |
| High saturated fat diet | Yes | 0.9 |
| | No | 0.1 |
| Exercise | Never | 0.6 |
| | Regular | 0.1 |
| | High If age < 30 | 0.1 |
| | High If age > 50 | 0.6 |
| Sedentary Lifestyle/inactivity | Yes | 0.7 |
| | No | 0.1 |
| Hereditary | Yes | 0.7 |
| | No | 0.1 |
| Bad cholesterol | Very High >200 | 0.9 |
| | High 160 to 200 | 0.8 |
| | Normal <160 | 0.1 |
| Blood Pressure | Normal (130/89) | 0.1 |
| | Low (< 119/79) | 0.8 |
| | High (>200/160) | 0.9 |
| Blood sugar | High (>120&<400) | 0.5 |
| | Normal (>90&<120) | 0.1 |
| | Low (<90) | 0.4 |
| Heart Rate | Low (< 60bpm) | 0.9 |
| | Normal (60 to 100) | 0.1 |
| | High (>100bpm) | 0.9 |

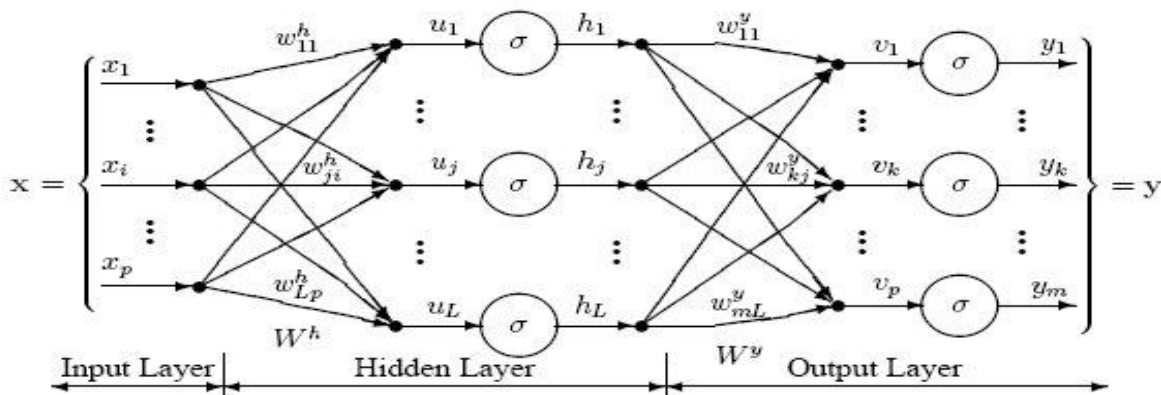


Figure 2: Structure of MLPNN

3.1.2 Data Mining and Genetic 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. **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. Table 4 shows the accuracy of the algorithm obtained from experiment.

Table 4 Results obtained from Experiment

| Data Mining Techniques | Accuracy |
|-------------------------------|----------|
| Naive Bayes | 96.5 % |
| Decision Tree | 99.2 % |
| Classification via clustering | 88.3 % |

3.1.3 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. In [11], four constraints were proposed to reduce the number of rules: item filtering, attribute grouping, maximum itemset 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.

Table 5 shows the 25 attributes selected for the experiments that were the most important. The author has performed the Experiments on a real data set to study the impact of constraints and the elimination of unreliable rules with validation on the test set. Figure 3 shows some of the rules with their support, confidence and lift value.

Table 5. List of attributed used in association rule discovery for Heart Disease prediction.

| Attribute name | Medical meaning | Neg | Constraints | | |
|----------------|-------------------|-----|-------------|-------|----|
| | | | itemFilter | group | ac |
| AGE | Patient age | 0 | 0 | 0 | 1 |
| LM | Left Main | 1 | 1 | 0 | 2 |
| LAD | Left Anter Desc | 1 | 1 | 0 | 2 |
| LCX | Left CircumfleX | 1 | 1 | 0 | 2 |
| RCA | Right Coronary | 1 | 1 | 0 | 2 |
| AL | Antero-Lateral | 0 | 1 | 1 | 1 |
| AS | Antero-Septal | 0 | 1 | 1 | 1 |
| SA | Septo-Anterior | 0 | 1 | 1 | 1 |
| SI | Septo-Inferior | 0 | 1 | 1 | 1 |
| IS | Infero-Septal | 0 | 1 | 1 | 1 |
| IL | Infero-Lateral | 0 | 1 | 1 | 1 |
| LI | Latero-Inferior | 0 | 1 | 1 | 1 |
| LA | Latero-Anterior | 0 | 1 | 1 | 1 |
| AP | Apical | 0 | 1 | 1 | 1 |
| SEX | Gender | 0 | 0 | 0 | 1 |
| HTA | Hypertension Y/N | 0 | 1 | 0 | 1 |
| DIAB | Diabetes Y/N | 0 | 1 | 0 | 1 |
| HYPLPD | Hyperloip Y/N | 0 | 1 | 0 | 1 |
| FHCAD | Faml hist dis Y/N | 0 | 1 | 0 | 1 |
| SMOKE | Smokes Y/N | 0 | 1 | 0 | 1 |
| CLAUDI | Claudication Y/N | 0 | 1 | 0 | 1 |
| PANGIO | Prev angina Y/N | 0 | 1 | 0 | 1 |
| PSTROKE | Prior stroke Y/N | 0 | 1 | 0 | 1 |
| PCARSUR | Prior surgery Y/N | 0 | 1 | 0 | 1 |
| CHOL | Cholesterol | 0 | 0 | 0 | 1 |

3.1.4 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 early 1980. 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.

$c = 1$ and $l \geq 2$:
 IS[0.2,1.0) CLAUDI=y PSTROKE=y \Rightarrow not RCA[0,50) $s=0.02, c=1, l=2.3$
 $c = 1$ and $l < 2$:
 SA[0.2,1.1) CHOL[200,250) \Rightarrow not LAD[0,50) $s=.03, c=1, l=1.98$
 SA[0.2,1.1) SMOK=y CHOL[200,250) \Rightarrow not LAD[0,50) $s=.02, c=1, l=1.9$
 SI[0.2,1.1) SEX=M CHOL[200,250) \Rightarrow not LAD[0,50) $s=.02, c=1, l=1.9$
 $c > 0.9$ and $l > 2$:
 SA[0.2,1.1) CHOL[200,250) \Rightarrow LAD[70,100) $s=0.03, c=0.95, l=3$
 SI[0.2,1.1) SEX=M CHOL[200,250) \Rightarrow LAD[70,100) $s=0.01, c=.91, l=2.9$
 SA[0.2,1.1) PANGIO=y \Rightarrow LAD[70,100) $s=0.02, c=0.91, l=2.9$
 $c \geq 0.9$ and $l < 2$:
 DIAB=y SMOKE=y 0.CHOL[0,200) \Rightarrow not LAD[0,50) $s=.02, c=.93, l=1.8$
 SA[0.2,1.1) CLAUDI=y PSTROK=y \Rightarrow not LAD[0,50) $s=.02, c=.95, l=1.8$
 IL[0.2,1.1) SEX=M CHOL[200,250) \Rightarrow not LAD[0,50) $s=.03, c=.93, l=1.8$
 AP[0.2,1.1) HYPLPD=y CHOL[0,200) \Rightarrow not LAD[0,50) $s=.02, c=.96$
 AGE[60,100) AP[0.2,1.1) SEX=F \Rightarrow not LAD[0,50) $s=.05, c=.94, l=1.8$
 $c < 0.9$ and $l \geq 2$:
 AGE[60,100) SA[0.2,1.1) PANGIO=y \Rightarrow LAD[70,100) $s=.02, c=.88, l=2.8$
 AGE[40,60) SEX=M CHOL[250,500) \Rightarrow not LCX[0,50) $s=.02, c=.85, l=2$
 SA[0.2,1.1) CLAUDI=y \Rightarrow LAD[70,100) $s=0.03, c=0.76, l=2.46$
 AGE[60,100) IL[0.2,1.1) CHOL[250,500) \Rightarrow RCA [70,100) $s=.02, c=.89$
 Two items (simple):
 AP[0.2,1.1) \Rightarrow not(LAD[0,50)) $s=0.24, c=0.78, l=1.54$
 SA[0.2,1.1) \Rightarrow not(LAD[0,50)) $s=0.17, c=0.80, l=1.58$

Figure 3: Some Association rules predicting existence of heart disease using Train and Test approach.

In [5] the authors have developed rule selection method for filtering large number of extracted rules from Coronary Artery Disease (CAD) data set. Two stages rule selection is proposed. Selection based on support of individual rules is applied on the first stage. Rough Set based selection using attribute reduction concept is performed for the second stage. Selection based on unseen data set is carried out to increase the generalization of elected rules. Experiment on CAD data set shows that the proposed method is able to select small number of rules while maintaining the quality of rule based classifier. The proposed method has better quality compared to previous rule selection methods. RST based rule importance measure is modified to select the rules by converting the rules to decision tables. Support filtering as one of rule quality based filtering is applied at the first stage of the selection. It is time consuming and high computation cost to apply only RST based rule selection because the number of attributes is very high. Filtering method based on rule support is applied to reduce the number of rules before applying rule importance measure to select the most importance rules. The modification is proposed by applying this method to decision

system and converting rules to form decision tables based on testing data instead of training data for rule importance measurement. The experimental results show that the proposed method is able to select small number (27) of rules from large number (3881) of rule without significantly reducing the quality of classification. The proposed method is comparable or even better than almost other methods.

4. ISSUES AND CHALLENGES

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 doctor's 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. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

5. CONCLUSION AND FUTURE WORK.

In this paper the problem of constraining and summarizing different algorithms of data mining used in the field of medical prediction are discussed. The focus is on using different algorithms and combinations of several target attributes for intelligent and effective heart attack prediction using data mining. For predicting heart attack, significantly 15 attributes are listed and with basic data mining technique other approaches e.g. ANN, Time Series, Clustering and Association Rules, soft computing approaches etc. can also be incorporated. The outcome of predictive data mining technique on the same dataset reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The second conclusion is that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

The proposed work can be further enhanced and expanded for the automation of Heart disease prediction. Real data from Health care organizations and agencies needs to be collected and all the available techniques will be compared for the optimum accuracy.

6. REFERENCES

- [1] Asha Rajkumar, G.Sophia Reena, *Diagnosis Of Heart Disease Using Datamining Algorithm*, Global Journal of Computer Science and Technology 38 Vol. 10 Issue 10 Ver. 1.0 September 2010.
- [2] Sunita Soni, O.P.Vyas, *Using Associative Classifiers for Predictive Analysis in Health Care Data Mining*, International Journal of Computer Application (IJCA, 0975 – 8887) Volume 4– No.5, July 2010, pages 33-34.
- [3] K.Srinivas, B.Kavihta Rani , A.Govrdhan , *Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks*, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255.
- [4] M. ANBARASI, E. ANUPRIYA, 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, 5370-5376

- [5] N.A. Setiawan, P.A. Venkatachalam, and Ahmad Fadzil M.H. , *Rule Selection for Coronary Artery Disease Diagnosis Based on Rough Set*, International Journal of Recent Trends in Engineering, Vol 2, No. 5, November 2009.
- [6] Sunita Soni , Jyothi Pillai, O.P.Vyas, *An Associative Classifier Using Weighted Association Rule* , IEEE proceedings of the World Congress on Nature and Biologically Inspired Computing (NaBIC'09), December 09-11, 2009, 1492-1496.
- [7] Shantakumar B.Patil, Y.S.Kumaraswamy, *Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network*, European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656
- [8] Ruben D. Canlas Jr., DATA MINING IN HEALTHCARE: CURRENT APPLICATIONS AND ISSUES, August 2009.
- [9] Sellappan Palaniappan Rafiah Awang, *Intelligent Heart Disease Prediction System Using Data Mining Techniques*, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008
- [10] Fadi Thabtah, *A review of associative classification mining*, The Knowledge Engineering Review, Volume 22 , Issue 1 (March 2007),Pages 37-65, 2007.
- [11] Carloz Ordonez, *Association Rule Discovery with Train and Test approach for heart disease prediction*, IEEE Transactions on Information Technology in Biomedicine, Volume 10, No. 2, April 2006.pp 334-343.
- [12] Harleen Kaur , Siri Krishan Wasan and Vasudha Bhatnagar, *THE IMPACT OF DATA MINING TECHNIQUES ON MEDICAL DIAGNOSTICS*, Data Science Journal, Volume 5, 19 October 2006 pp119-126.
- [13] Yin, X., Han, J. *CPAR: Classification based on predictive association rule*. In Proceedings of the SIAM International Conference on Data Mining. San Francisco, CA: SIAM Press, 2003, pp. 369-376.
- [14] Feng Tao, Fionn Murtagh, Mohsen Farid. *Weighted Association Rule Mining using Weighted Support and Significance Framework*, Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining 2003, Pages:661-666 Year of Publication: 2003
- [15] W.Li, J. Han, J.Pei , *CMAR- Classification based on Multiple Association Rules*, ICDM'01, , San Jose, CA, Nov. 2001. pp. 369-376
- [16] W. Wang, J. Yang and P. Yu. *Efficient mining of weighted association rules (WAR)*, Proc. of the ACM SIGKDD Conf. on Knowledge Discovery and Data Mining, 270-274, 2000.
- [17] Liu,B, Hsu. W. Ma, *Integrating Classification and association rule mining* . Proceeding of the KDD, 1998(CBA) pp 80-86.
- [18] Magnus Stensmo, Terrence J. Sejnowski *Automated Medical Diagnosis based on Decision Theory and Learning from Cases*, World Congress on Neural Networks 1996 International Neural Network society pp. 1227-1 231.
- [19] R. Agrawal and R. Srikant. *Fast algorithms for mining association rules*. In VLDB'94, , Santiago, Chile, Sept. 1994. pp. 487-49