Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction

Shantakumar B.Patil
Ph.D. Scholar
Dr. MGR Educational and Research Institute
Chennai

Dr.Y.S.Kumaraswamy
Sr.Prof &HOD,
Dept. of MCA (VTU)
Dayananda Sagar College of Engineering
Bangalore, India

Summary
The diagnosis of diseases is a significant and tedious task in medicine. The detection of heart disease from various factors or symptoms is a multi-layered issue which is not free from false presumptions often accompanied by unpredictable effects. Thus the effort to utilize knowledge and experience of numerous specialists and clinical screening data of patients collected in databases to facilitate the diagnosis process is considered a valuable option. The healthcare industry gathers enormous amounts of heart disease data that regrettably, are not "mined" to determine concealed information for effective decision making by healthcare practitioners. In this paper, we have proposed an efficient approach for the extraction of significant patterns from the heart disease warehouses for heart attack prediction. Initially, the data warehouse is preprocessed to make it appropriate for the mining process. After preprocessing, the heart disease warehouse is clustered using the K-means clustering algorithm, which will extract the data relevant to heart attack from the warehouse. Subsequently the frequent patterns are mined from the extracted data, relevant to heart disease, using the MAFIA algorithm. Then the significant weightage of the frequent patterns are calculated. Further, the patterns significant to heart attack prediction are chosen based on the calculated significant weightage. These significant patterns can be used in the development of heart attack prediction system.

Keywords:
Data Mining, Disease Diagnosis, Heart Disease, Pre-processing, Frequent Patterns, MAFIA (MAximal Frequent Itemset Algorithm), Clustering, K-Means, Significant Patterns.

1. Introduction

The necessity of effective identification of information - contextual data - non obvious and valuable for decision making from a large collection of data has been on a steady increase recently. This is an interactive and iterative process encompassing several subtasks and decisions and is known as Knowledge Discovery from Data. The central process of Knowledge Discovery is the transformation of data into knowledge for decision making, known as Data Mining [2, 20]. Knowledge discovery in databases constitutes numerous distinct clearly illustrated processes. The central process is that of data mining which aids in the recognition of hoarded yet beneficial knowledge from enormous databases. A widely accepted formal definition of data mining is given subsequently "Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data" [4]. Conventionally, the information that is mined is denoted as a as a model of the semantic structure of the dataset. The model might be utilized for prediction and categorization of new data [1].

Diverse fields such as marketing, customer relationship management, engineering, medicine, crime analysis, expert prediction, Web mining, and mobile computing, besides others utilize Data mining [5]. A majority of areas related to medical services such as prediction of effectiveness of surgical procedures, medical tests, medication, and the discovery of relationships among clinical and diagnosis data also make use of Data Mining methodologies [3]. Proffering valuable services at reasonable costs is a chief confront envisaged by the healthcare organizations (hospitals, medical centers). Valuable quality service refers to the precise diagnosis of patients and proffering effective treatment. Poor clinical decisions may result in catastrophes and so are not entertained. It is also necessary that the hospitals reduce the cost of clinical test. This can be attained by the making use of proper computer-based information and/or decision support systems [9].

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 to demanding a more quantitative assessment of information with the of all supporting 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 physicians dealing with an identical case from all over the world [7]. Medical diagnosis is considered to me a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Unfortunately all doctors are not proficient in every sub specialty and besides there is a scarcity of resource persons at certain

Manuscript received February 5, 2009
Manuscript revised February 20, 2009

places. Thus a system of automated medical diagnosis would prove highly beneficial by bringing all of them together [11].

Medical history data constitutes numerous tests necessary to diagnose a particular disease [8]. Clinical databases are a part of the domain where process of data mining has become inevitable owing to the gradual ascent of medical and clinical research data. Healthcare industries can benefit from Data mining by utilizing the same as an intelligent diagnostic tool. It is feasible to obtain knowledge and information regarding a disease from the patient specific stored measurements in case of medical data. As a result, data mining has become a significant domain in healthcare [6]. The effectiveness of medical treatments can be estimated by developing the data mining applications. Data mining is capable of delivering an analysis of which courses of action prove effective [12], achieved by comparing and contrasting causes, symptoms, and courses of treatments. The real-life data mining applications are fascinating as they present data miners with diverse set of problems, time and again. Dealing with heart disease patients databases is one such real-life application. The detection of a disease from numerous factors or symptoms is a many layered problem and might result in false assumptions often accompanied by erratic effects. Hence it seems sensible to attempt employing the knowledge and experience of numerous specialists collected in databases towards supporting the diagnosis process [2], [10].

The researchers in the medical field diagnose and predict the diseases in addition to providing effective care for patients [2, 6, 43, 44, 13] by employing the data mining techniques. The data mining techniques have been employed by numerous works in the literature to diagnose diverse diseases, for instance: Diabetes, Hepatitis, Cancer, Heart diseases and more [39, 40, 41, 42]. The disease related information existing in the form of electronic clinical records, treatment information, gene expressions, images and more were used by all these works. In recent times, the data mining techniques are employed by numerous authors to propose diagnosis approaches for various kinds of heart diseases [14, 9, 22, 23, 24, 25].

The primary intent of our research is to design and develop an efficient approach for extracting patterns, which are significant to heart attack, from the heart disease data warehouses. The approach aims to utilize the data mining techniques: clustering and frequent pattern mining. The heart disease data warehouse consists of mixed attributes containing both the numerical and categorical data. These records are cleaned and filtered with the intention that the irrelevant data from the warehouse would be removed before mining process occurs. Then clustering is performed on the preprocessed data warehouse using K-means clustering algorithm with K value so as to extract data relevant to heart attack. Subsequently the frequent patterns significant to heart disease diagnosis are mined from the extracted data using the MAFIA algorithm. The significant weightage is calculated for each frequent pattern using the approach proposed. Then the patterns with significant weightage greater than a predefined threshold value are chosen. The chosen significant patterns can be used in the development of heart attack prediction system.

The remaining sections of the paper are organized as follows: In Section 2, a brief review of some of the works on heart disease diagnosis is presented. An introduction about the heart disease and its effects is given in Section 3. The extraction of significant patterns from heart disease data warehouse is presented detailed in Section 4. Experimental results are specified in section 5. The conclusions and future work are described in Section 6.

2. Related Works

Numerous works in literature related with heart disease diagnosis using data mining techniques have motivated our work.

A novel technique to develop the multi-parametric feature with linear and nonlinear characteristics of HRV (Heart Rate Variability) was proposed by Heon Gyu Lee et al. [14]. Statistical and classification techniques were utilized to develop the multi-parametric feature of HRV. Besides, they have assessed the linear and the non-linear properties of HRV for three recumbent positions, to be precise the supine, left lateral and right lateral position. Numerous experiments were conducted by them on linear and nonlinear characteristics of HRV indices to assess several classifiers, e.g., Bayesian classifiers [17], CMAR (Classification based on Multiple Association Rules) [16], C4.5 (Decision Tree) [18] and SVM (Support Vector Machine) [15]. SVM surmounted the other classifiers.

A model Intelligent Heart Disease Prediction System (IHDPS) built with the aid of data mining techniques like Decision Trees, Naïve Bayes and Neural Network was proposed by Sellappan Palaniappan et al. [9]. The results illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives. IHDPS was capable of answering queries that the conventional decision support systems were not able to. It facilitated the establishment of vital knowledge, e.g. patterns, relationships amid medical factors connected with heart disease. IHDPS subsists well being web-based, user-friendly, scalable, reliable and expandable.

The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [22]. Experiments were carried out on a sample database of patients' records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and the like. The supervised network has been recommended for diagnosis of heart diseases. Training was carried out with the aid of back propagation algorithm. Whenever unknown data was fed by the doctor, the system identified the unknown data from comparisons with the trained data and generated a list of probable diseases that the patient is vulnerable to.

The problem of identifying constrained association rules for heart disease prediction was studied by Carlos Ordonez [23]. The assessed data set encompassed medical records of people having heart disease with attributes for risk factors, heart perfusion measurements and artery narrowing. Three constraints were introduced to decrease the number of patterns. First one necessitates the attributes to appear on only one side of the rule. The second one segregates attributes into uninteresting groups. The ultimate constraint restricts the number of attributes in a rule. Experiments illustrated that the constraints reduced the number of discovered rules remarkably besides decreasing the running time. Two groups of rules envisaged the presence or absence of heart disease in four specific heart arteries.

Data mining methods may aid the clinicians in the predication of the survival of patients and in the adaptation of the practices consequently. The work of Franck Le Duff et al. [24] might be executed for each medical procedure or medical problem and it would be feasible to build a decision tree rapidly with the data of a service or a physician. Comparison of traditional analysis and data mining analysis illustrated the contribution of the data mining method in the sorting of variables and concluded the significance or the effect of the data and variables on the condition of the study. A chief drawback of the process was knowledge acquisition and the need to collect adequate data to create an appropriate model.

A novel heuristic for efficient computation of sparse kernel in SUPANOVA was proposed by Boleslaw Szymanski et al. [25]. It was applied to a benchmark Boston housing market dataset and to socially significant issue of enhancing the detection of heart diseases in the population with the aid of a novel, non-invasive measurement of the heart activities on basis of magnetic field generated by the human heart. 83.7% predictions on the results were correct thereby outperforming the results obtained through Support Vector Machine and equivalent kernels. The spline kernel yielded equally good results on the benchmark Boston housing market dataset.

In [11] Latha Parthiban et al. projected an approach on basis of coactive neuro-fuzzy inference system (CANFIS) for prediction of heart disease. The CANFIS model diagnosed the presence of disease by merging the neural network adaptive capabilities and the fuzzy logic qualitative approach and further integrating with genetic algorithm. On the basis of the training performances and classification accuracies, the performances of the CANFIS model were evaluated. The CANFIS model is promising in the prediction of the heart disease as illustrated by the results.

In [29] Kiyong Noh et al. put forth a classification method for the extraction of multi-parametric features by assessing HRV from ECG, data preprocessing and heart disease pattern. The efficient FP-growth method was the basis of this method which is an associative. They presented a rule cohesion measure that allows a strong push of pruning patterns in the pattern generating process as the volume of patterns created could possibly be huge. The multiple rules and pruning, biased confidence (or cohesion measure) and dataset consisting of 670 participants, distributed into two groups, namely normal people and patients with coronary artery disease, were employed to carry out the experiment for the associative classifier.

3. Heart Disease

The term Heart disease encompasses the diverse diseases that affect the heart. Heart disease was the major cause of casualties in the United States, England, Canada and Wales as in 2007. Heart disease kills one person every 34 seconds in the United States [28]. 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 severe illness, disability, and death [19]. Narrowing of the coronary arteries results in the reduction of blood and oxygen supply to the heart and leads to the Coronary heart disease (CHD). Myocardial infarctions, generally known as a heart attacks, and angina pectoris, or chest pain are encompassed in the CHD. A sudden blockage of a coronary artery, generally due to a blood clot results in a heart attack. Chest pains arise when the blood received by the heart muscles is inadequate [17].

High blood pressure, coronary artery disease, valvular heart disease, stroke, or rheumatic fever/rheumatic heart disease are the various forms of cardiovascular disease. The World Health Organization has estimated that 12 million deaths occurs world wide, every year due to the cardiovascular 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 [27].

4. Extraction Of Significant Patterns From Heart Disease Data Warehouse

The extraction of significant patterns from the heart disease data warehouse is presented in this section. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse is preprocessed to make the mining process more efficient. The preprocessed data warehouse is then clustered using the K-means clustering algorithm with K=2. This result in two clusters, one contains the data that are most relevant to heart attack and the other contains the remaining data. The frequent patterns are mined from the data, relevant to heart attack, using the MAFIA algorithm. The significant weightage is calculated for all frequent patterns with the aid of the approach proposed. The frequent patterns with significant weightage greater than a predefined threshold are chosen. These chosen significant patterns can be used in the design and development of heart attack prediction system.

4.1 Data Preprocessing

Cleaning and filtering of the data might be necessarily carried out with respect to the data and data mining algorithm employed so as to avoid the creation of deceptive or inappropriate rules or patterns [33]. The actions comprised in the pre-processing of a data set are the removal of duplicate records, normalizing the values used to represent information in the database, accounting for missing data points and removing unneeded data fields. In order for making the data appropriate for the mining process it needs to be transformed. The raw data is changed into data sets with a few appropriate characteristics. Moreover it might be essential to combine the data so as to reduce the number of data sets besides minimizing the memory and processing resources required by the data mining algorithm [37]. In our approach, the heart disease data warehouse is refined by removing duplicate records and supplying missing values. Furthermore it is also transformed to a form appropriate for clustering.

4.2 Clustering Using K-Means Algorithm

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 [31], is known as Clustering. The clustering problem has been addressed in numerous contexts besides being proven beneficial in many applications. Clustering medical data into small yet meaningful clusters 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 [32]. Numerous methods are available in the literature for clustering. We have employed the renowned K-Means clustering algorithm in our approach.

The k-means algorithm [38] 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:

- K points denoting the data to be clustered are placed into the space. These points denote the primary group centroids.
- The data are assigned to the group that is adjacent to the centroid.
- The positions of all the K centroids are recalculated as soon as all the data are assigned.
- 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 warehouse is clustered using the K-means algorithm with K value as 2. One cluster consists of the data relevant to the heart disease and the other contains the remaining data. Later on, the frequent patterns are mined from the cluster relevant to heart disease, using the MAFIA algorithm.

4.3 Frequent Pattern Mining Using Mafia

Frequent Itemset Mining (FIM) is considered to be one of the elemental data mining problems that intends to discover groups of items or values or patterns that co-occur frequently in a dataset [26], [34]. It is of vital significance in a variety of Data Mining tasks that aim to mine interesting patterns from databases, like association rules, correlations, sequences, episodes, classifiers, clusters and the like. Numerous algorithms like the Apriori [21] and FP-Tree [30] have been proposed to support the discovery of interesting patterns. The proposed approach utilizes an efficient algorithm called MAFIA (MAximal Frequent Itemset Algorithm) which combines diverse old and new algorithmic ideas to form a practical algorithm [35] [36]. The proposed algorithm is employed for the extraction of association rules from the clustered dataset besides performing efficiently when the database consists of very long itemsets specifically. The depth-first traversal of the itemset lattice and effective pruning mechanisms are incorporated in the search strategy of the proposed algorithm.

Pseudo code for MAFIA [35]:

MAFIA(**C**, **MFI**, Boolean **IsHUT**) {
name **HUT** = **C**.head $\cup$ **C**.tail;
if **HUT** is in **MFI**
stop generation of children and return
Count all children, use **PEP** to trim the tail, and recorder by increasing support,
For each item **i** in **C**, trimmed_tail{
IsHUT = whether **i** is the first item in the tail
newNode = **C** $\cup$ **I**
MAFIA (newNode, **MFI, IsHUT**)}
if (**IsHUT** and all extensions are frequent)
Stop search and go back up subtree
If (**C** is a leaf and **C.**head is not in **MFI**)
Add **C**.head to **MFI**
}

The cluster that contains data most relevant to heart attack is fed as input to MAFIA algorithm to mine the frequent patterns present in it. Then the significance weightage of each pattern is calculated using the approach described in the following subsection.

4.4 Significance Weightage Calculation

After mining the frequent patterns using MAFIA algorithm, the significance weightage of each pattern is calculated. It is calculated based on the weightage of each attribute present in the pattern and the frequency of each pattern. The formula used to determine the significant weightage (S_w) is as follows:

$$S_{Wi} = \sum_{i=1}^{n} W_i f_i$$

Where W_i represents the weightage of each attribute and f_i denotes the frequency of each rule. Subsequently the patterns having significant weightage greater than a predefined threshold are chosen to aid the prediction of heart attack

$$SFP = \{x : S_w(x) \geq \Phi\}$$

Where SFP represents significant frequent patterns and Φ represents the significant weightage. This SFP can be used in the design of heart attack prediction system.

5. Experimental Results

The results of our experimental analysis in finding significant patterns for heart attack prediction are presented in this section. We have implemented our proposed approach in Java. The heart disease dataset that we have used for our experiments was obtained from [45]. The detailed description of the dataset and the extracted significant patterns are given in the following subsections.

5.1 Heart Disease Dataset Description

The dataset used in our approach contains the parameters like blood pressure, cholesterol, chest pain, maximum heart rate and more. The detailed description of the parameters and their ranges are given as follows:

1 age: age in years
2 sex: sex (1 = male; 0 = female)
3 cp: chest pain type
  Value 1: typical angina
  Value 2: atypical angina
  Value 3: non-anginal pain
  Value 4: asymptomatic
4 trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5 chol: serum cholestoral in mg/dl
6 fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7 restecg: resting electrocardiographic results
  Value 0: normal
  Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8 thalach: maximum heart rate achieved
9 exang: exercise induced angina (1 = yes; 0 = no)
10 oldpeak = ST depression induced by exercise relative to rest
11 slope: the slope of the peak exercise ST segment

Value 1: upsloping
Value 2: flat
Value 3: downsloping

12 ca: number of major vessels (0-3) colored by flourosopy
13 thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14 num: diagnosis of heart disease (angiographic disease status)
Value 0: < 50% diameter narrowing
Value 1: > 50% diameter narrowing

5.2 Results

With the help of the dataset, the patterns significant to the heart attack prediction are extracted. The heart disease data set is preprocessed successfully by removing duplicate records and supplying missing values. The refined heart disease data set, resultant from preprocessing, is then clustered using K-means algorithm with K value as 2. Then the frequent patterns are mined efficiently from the cluster relevant to heart disease, using the MAFIA algorithm. Subsequently, the significant patterns are extracted with the aid of the significance weightage greater than the pre-defined threshold. The extracted patterns from the data set significant to the heart attack prediction and their significance weigthage are given in Table 1:

Table 1**.** Extracted significant patterns and their significance weightage

Significant Patterns	**Significant Weightage**
cholestoral-blood pressure-blood sugar-diagnosis of hd	169.95
thal-sex-blood pressure-blood sugar-diagnosis of hd	167.4
chest pain type-sex-blood pressure-blood sugar-diagnosis of hd	159.53
slope-blood pressure-blood sugar-diagnosis of hd	153.76
exercise induced angina-chest pain type-blood pressure-diagnosis of hd	153.70
thal-chest pain type-blood pressure-diagnosis of hd	152.50
exercise induced angina-blood pressure-blood sugar-diagnosis of hd	151.57
exercise induced angina-chest pain type-blood sugar-diagnosis of hd	142.10
cholestoral-sex-blood sugar-diagnosis of hd	141.51
electrocardiographic-blood pressure-diagnosis of hd	136.4

The values corresponding to each attribute in the significant patterns are as follows: blood pressure range is greater than 140/90 mm Hg, cholestoral range is greater than 240 mg/dl, maximum heart rate is greater than 100 beats/ minute, abnormal ECG and unstable angina. These values are the main cause for the heart attack disease. The extracted significant patterns can be used to develop an efficient heart attack prediction system using artificial intelligence techniques.

6. Conclusions And Future Work

Data mining in health care management is unlike the other fields owing to the fact that the data present are heterogeneous and that certain ethical, legal, and social constraints apply to private medical information. Health care related data are voluminous in nature and they arrive from diverse sources all of them not entirely appropriate in structure or quality. These days, the exploitation of knowledge and experience of numerous specialists and clinical screening data of patients gathered in a database during the diagnosis procedure, has been widely recognized. In this paper we have presented an efficient approach for extracting significant patterns from the heart disease data warehouses for the efficient prediction of heart attack. The preprocessed heart disease data warehouse was clustered to extract data most relevant to heart attack using K-means clustering algorithm. The frequent items have been mined effectively using MAFIA algorithm. Based on the calculated significant weightage, the frequent patterns having value greater than a predefined threshold were chosen for the valuable prediction of heart attack. In our future work, we have planned to design and develop an efficient heart attack prediction system with the aid of these selected significant patterns using artificial intelligence techniques.

References

[1] Sally Jo Cunningham and Geoffrey Holmes, "Developing innovative applications in agriculture using data mining", In the Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999.

[2] Frank Lemke and Johann-Adolf Mueller, "Medical data analysis using self-organizing data mining technologies," Systems Analysis Modelling Simulation , Vol. 43 , no. 10 , pp: 1399 - 1408, 2003.

[3] Tzung-I Tang, Gang Zheng, Yalou Huang, Guangfu Shu, Pengtao Wang, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis", IEMS Vol. 4, No. 1, pp. 102-108, June 2005.

[4] Frawley and Piatetsky-Shapiro, Knowledge Discovery in Databases: An Overview. The AAAI/MIT Press, Menlo Park, C.A, 1996.

[5] Hsinchun Chen, Sherrilynne S. Fuller, Carol Friedman, and William Hersh, "Knowledge Management, Data Mining, and Text Mining In Medical Informatics", Chapter 1, pgs 3-34

[6] S Stilou, P D Bamidis, N Maglaveras, C Pappas , Mining association rules from clinical databases: an intelligent diagnostic process in healthcare. Stud Health Technol Inform 84: Pt 2. 1399-1403, 2001.

[7] T Syeda-Mahmood, F Wang, D Beymer, A Amir, M Richmond, SN Hashmi, "AALIM: Multimodal Mining for Cardiac Decision Support", Computers in Cardiology,pages: 209-212, Sept. 30 2007-Oct. 3 2007

[8] Anamika Gupta, Naveen Kumar, and Vasudha Bhatnagar, "Analysis of Medical Data using Data Mining and Formal Concept Analysis", Proceedings Of World Academy Of Science, Engineering And Technology ,Volume 6, June 2005,.

[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] Andreeva P., M. Dimitrova and A. Gegov, Information Representation in Cardiological Knowledge Based System, SAER'06, pp: 23-25 Sept, 2006.

[11] Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences 3; 3, 2008

[12] Hian Chye Koh and Gerald Tan ,"Data Mining Applications in Healthcare", Journal of healthcare information management, Vol. 19, Issue 2, Pages 64-72, 2005.

[13] L. Goodwin, M. VanDyne, S. Lin, S. Talbert ,"Data mining issues and opportunities for building nursing knowledge" Journal of Biomedical Informatics, vol:36, pp: 379-388, 2003.

[14] Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.

[15] Cristianini, N., Shawe-Taylor, J.: An introduction to Support Vector Machines. Cambridge University Press, Cambridge , 2000.

[16] Li, W., Han, J., Pei, J.: CMAR: Accurate and Efficient Classification Based on Multiple Association Rules. In: Proc. of 2001 Interna'l Conference on Data Mining .2001.

[17] Chen, J., Greiner, R.: Comparing Bayesian Network Classifiers. In Proc. of UAI-99, pp.101–108 ,1999.

[18]Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo 1993.

[19] "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure among Persons with Diabetes", Special report: 2001 – 2003, New Mexico.

[20]M. Chen, J. Han, and P. S. Yu. Data Mining: An Overview from Database Perspective. IEEE Trans. Knowl. Dat. Eng., vol: 8, no:6, pp: 866-883, 1996.

[21] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases", In Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, August 29-September 1994.

[22] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review , Vol. 8, No. 1 (January - June 2007.

[23] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.

[24]Franck Le Duff , Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics ,107(Pt 2):1256-9, 2004.

[25] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel, Lijuan Zhu, "Using Efficient Supanova Kernel For Heart Disease Diagnosis", proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol. 16, pp:305-310, 2006.

[26]Agrawal, R., Imielinski, T. and Swami, A, 'Mining association rules between sets of items in large databases', Proc. ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'93),Washington, DC, July, pp.45–51, 1993.

[27] "Heart Disease" from http://chinese-school.netfirms.com/heart-disease-causes.html

[28] "Heart disease" from http://en.wikipedia.org/wiki/Heart_disease.

[29] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", springer,Vol:345 , pp: 721-727, 2006.

[30] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1–12, 2000,

[31] Zakaria Nouir, Berna Sayrac, Benoît Fourestié, Walid Tabbara, and Françoise Brouaye, "Generalization Capabilities Enhancement of a Learning System by Fuzzy Space Clustering," Journal of Communications, Vol. 2, No. 6, pp. 30-37, November 2007.

[32] F. H. Saad, B. de la Iglesia, and G. D. Bell, A Comparison of Two Document Clustering Approaches for Clustering Medical Documents, Proceedings of the 2006 International Conference on Data Mining (DMIN-06), 2006.

[33] Gerhard Münz, Sa Li, and Georg Carle. "Traffic anomaly detection using k-means clustering", In Proc. of Leistungs-, Zuverlässigkeits- und Verlässlichkeitsbewertung von Kommunikationsnetzen und Verteilten Systemen, 4.

GI/ITG-Workshop MMBnet 2007, Hamburg, Germany, September 2007.

[34] Eric Li, Li Liu, "Optimization of Frequent Itemset Mining on Multiple-Core Processor", pp: 1275-1285, 2007

[35] Douglas Burdick, Manuel Calimlim, Johannes Gehrke, "MAFIA: A Maximal Frequent Itemset Algorithm for Transactional Databases", Proceedings of the 17th International Conference on Data Engineering, pp.443-452, April 02-06, 2001

[36] Doug Burdick, Manuel Calimlim, Jason Flannick, Johannes Gehrke, Tomi Yiu, "MAFIA: A Performance Study of Mining Maximal Frequent Itemsets", Proceedings of the 17th International Conference on Data Engineering, p.443-452, April 02-06, 2001.

[37] Wynne Hsu, Mong-Li Lee, Bing Liu, Tok Wang Ling: Exploration mining in diabetic patients databases: findings and conclusions. KDD 2000: pp: 430-436, 2000

[38] C. Ordonez, "Programming the K-Means Clustering Algorithm in SQL," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining, pp. 823-828, 2004.

[39] A. Bellaachia and Erhan Guven, Predicting Breast Cancer Survivability using Data Mining Techniques , Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining (SDM 2006), Saturday, April 22, 2006.

[40] Abdelghani Bellaachia and David Portnoy, E-CAST: A Data Mining Algorithm for Gene Expression Data , 2nd Workshop on Data Mining in Bioinformatics at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, July 23rd, 2002.

[41] KS Leung,YT Ng, KH Lee, LY Chan, KW Tsui, Tony Mok, CH Tse, Joseph Sung, "Data Mining on DNA Sequences of Hepatitis B Virus by Nonlinear Integrals" Proceedings Taiwan-Japan Symposium on Fuzzy Systems & Innovational Computing, 3rd meeting, pp.1-10 Aug 2006, Japan.

[42] Florin Gorunescu, "Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection," International Journal of Bioengineering, Biotechnology and Nanotechnology, vol.1, no.2, pp.105 - 108, 2008.

[43] Fu-ren Lin, Shien-chao Chou, Shung-mei Pan, Yao-mei Chen, "Mining time dependency patterns in clinical pathways", International Journal of Medical Informatics, 2001.

[44]Margaret R. Kraft, Kevin C. Desouza, Ida Androwich (2002). "Data Mining in Healthcare Information Systems: Case Study of a Veterans' Administration Spinal Cord Injury Population", Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03).

[45] Heart attack dataset from http://archive.ics.uci.edu/ml/datasets/Heart+Disease

Mr. Shantakumar Patil working presently as Professor in Department of Information Science & Engineering at Nagarjuna College of Engineering & Technology, Devanahalli, Bangalore. He has presented more than 6 papers in National & International conferences. His areas of interest are Datamining, datawarehouse, finite Automata, Microprocessor

Dr.Y.S.Kumaraswamy, working presently as Sr.Prof and Head in Department of MCA (VTU) at Dayananda Sagar College of Engineering, Bangalore. He has published more than 150 research papers and guiding 38 Ph.D students in the area of Computer Science/ Mathematics and also selection committee member for ISRO/UGC/DSI profiles.