International Journal of Data Engineering (IJDE)
Singaporean Journal of Scientific Research(SJSR)
Vol.6.No.6 2014 Pp. 301-310
available at:www.iaaet.org/sjsr
Paper Received :02-10-2014
Paper Accepted:18-10-2014
Paper Reviewed by: 1. Dr. P. Surya Prakash 2. Dr.R. Kumutha
Editor : Dr. Binod Kumar

A SURVEY OF DATA MINING TECHNIQUES ON MEDICAL DIAGNOSIS AND RESEARCH

V.Gayathri
Asst. Professor
G.T.N Arts & Science College,
Karur Road
Dindigul -624 005, Tamilnadu, India

M.Chanda Mona
Professor & Head
G.T.N Arts & Science College,
Karur Road
Dindigul -624 005, Tamilnadu, India

S.Banu Chitra
Asst. Professor
G.T.N Arts & Science College,
Karur Road
Dindigul -624 005, Tamilnadu, India

ABSTRACT

In a Human Life Span disease is a major cause of illness and death in the modern society. Medical diagnosis is an important but complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are unfortunately not equally skilled in every sub specialty and they are in many places a scarce resource. A system for automated medical diagnosis would enhance medical care and reduce costs. This 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.
Keywords: Medical Diagnosis, Data Mining, Automation, Classification Technique.

1. INTRODUCTION

Data mining is defined as "a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database". Healthcare databases have a huge amount of data but however, there is a lack of effective analysis tools to discover the hidden knowledge. Appropriate computer- based information and/or decision support systems can help physicians in their work. The efficient and accurate implementation of an automated system needs a comparative study of various techniques available. Here we present an overview of the current research being carried out using the DM techniques for the diagnosis and prognosis of various diseases, highlighting critical issues and summarizing the approaches in a set of learned lessons. The rest of this paper is organized as follows: First we show the overview of Data Mining in this study in chapter two, we study the literature study of existing concepts is in chapter three, then we identify the advantages and disadvantages of data mining in medical domain. and finally we show the conclusions of our work.

The data mining techniques that have been applied to medical data include Apriori and FPGrowth [1], [2], [3], [4], [5], [6], [7], and [8], unsupervised neural networks [9][10], linear genetic programming[9],Association rule mining [11], [12], Bayesian Ying Yang [13], decision tree algorithms like ID3, C4.5, C5, and CART [14], [15], [16], [17], [18], [19], [20], outlier prediction technique [21], Fuzzy cluster analysis [22],classification algorithm[17],[23],[24], Bayesian Network algorithm [14], [25], Naïve Bayesian [26], combination of K-means, Self Organizing Map (SOM) and Naïve Bayes [27], Time series technique [28], [29], combination of SVM, ANN and ID3 [16], clustering and classification [30],SVM [16], [31], FCM [29],k-NN [24], and Bayesian Network [14]. This review provides the summary of all these techniques in terms of the problem they solve or their utility in medical data mining or the tools which are implemented over them and so on.

2. OVERVIEW OF DATA MINING

Data Mining Techniques Used in Diagnosis System Classification technique is the most frequently used data mining tasks with a majority of the implementation of Bayesian classifiers, neural networks, and Association Rule. Myriad of quantitative performance measures were proposed with a high proportion of accuracy, sensitivity, specificity, and ROC curves. The latter are usually associated with qualitative evaluation. Classification methods maps the data in to predefined targets. This approach is a supervised learning as targets are predefined. The goal of the classification method is to build a classifier based on some cases with some attributes to describe the objects or one attribute to describe the group of the objects. Then, these classifiers are used to predict the group attributes of new cases from the domain based on the values of other attributes. The following techniques are used in the disease diagnosis.

A. Decision Trees (DT's)

Decision tree is a tree where each non-terminal node represents a test or decision on the considered data item. Selection of a certain branch depends upon the outcome of the test. To classify a particular test data item, we start at the root node and follow the assertions down until we reach a terminal node or leaf node. Decision is made when a terminal node is approached. Decision trees that use recursive data partitioning can also be interpreted as a special form of a rule set. The Decision Tree algorithm, is based on conditional probabilities and unlike naïve Bayes, decision trees generate rules. Rules are the conditional statement that can easily be

understood by humans and easily used within a database to identify a set of records.

In some other applications of data mining, the accuracy of a prediction is the only thing that really matters. It may not be very important to know how the model works. In others, the ability to explain the reason for a decision can be crucial. For eg, a Marketing professional would need complete descriptions of customer segments in order to launch a successful marketing campaign. The Decision Tree algorithms are ideal for this type of application.

B. Neural Networks

Neural networks were recently the most popular artificial intelligence-based data modeling algorithm used in clinical medicine. Neural networks (NN) are those systems modeled based on the working of human brain. The power and speed of modern digital computers is truly astounding. Human cannot ever hope to compute a million operations a second. Still, there are some tasks for which even the most powerful computers cannot compete with the human brain, perhaps not even with the intelligence of an earthworm. This is a simple model and consists of a single `trainable' neuron. The term 'Trainable' means that its threshold and input weights are modifiable. Inputs are presented to the neuron and each input has a desired output (determined by us). If the neuron doesn't give the desired output, then it has made a mistake. To rectify this, its threshold and/or input weights must be changed. How this change is to be calculated is determined by the learning algorithm. Artificial Neural networks may be able to model complex non-linear relationships, comprising an advantage over simpler modeling methods like the Naïve Bayesian classifier or logistic regression.

C. Naive Bayes

The Microsoft Naive Bayes algorithm is a classification algorithm provided by Microsoft SQL Server Analysis Services for use in predictive modeling. The term(name) Naive Bayes derives From the fact that the algorithm uses Bayes theorem, but does not take into account dependencies that may exist, and therefore its assumptions are said to be naive.

This algorithm is less computationally intense than other Microsoft algorithms, and therefore is useful for quickly generating mining models to discover relationships between input columns and predictable columns. You can use this algorithm to do initial explorations of data, and then later you can apply the results to create additional mining models with other algorithms that are more computationally intense and more accurate. The Microsoft Naive Bayes algorithm calculates the probability of every state of each input, given each possible state of the predictable column. You can use the Microsoft Naive Bayes Viewer in Business Intelligence Development Studio to see a visual representation of how the algorithm distributes states.

D. Association Rule

In the process of generating the class association rules, instead of considering all the attributes, apply PCA and rank all attributes. The attribute with the highest ranking is used to generate the class association rules. This approach generates n*m rules for a single test instance with n non class attributes and m classes in the entire data set. If t Test cases are to be predicted the no. Of rules generated will be t*n*m. After identifying the principle 43 component attribute, the subsets are generated. For each generated subset, probability values are calculated. The decision of which class will be assigned to test instance X is based on the analysis of the subsets of attributes values with the highest posterior probabilities. Find the accuracy of the data set.

3. Literature Review

Evans et al. [37] stated that hereditary syndromes can be detected automatically using data mining techniques.

Doron Shalvi and Nicholas DeClaris, [10] discussed medical data mining through unsupervised neural networks besides a method for data visualization. They also emphasized the need for preprocessing prior to medical data mining.

In the year 2000 Krzysztof J. Cior [38], bioengineering professor, identified the need for data mining methods to mine medical multimedia content.

Tsumoto [39] identified problems in medical data mining. The problems include missing values, data storage with respect to temporal data and multi-valued data, different medical coding systems being used in Hospital Information Systems (HIS).

Brameier and Banzhaf [9] explored and analyzed two programming models such as neural networks, and linier genetic programming for medical data mining.

Abidi and Hoe [40] proposed and implemented a symbolic rule extraction workbench for generating emerging rule-sets.

Abidi et al. [41] explored the usage of rule-sets as results of data mining for building rule-based expert systems.

Olukunle and Ehikioya [11] proposed an algorithm for extracting association rules from medical image data. The association rule mining discovers frequently occurring items in the given dataset.

Shim and Xu [13] proposed a classification method based on Bayesian Ying Yang (BYY) which is a three layered model. They applied this model to classify liver disease through automatic discovery of medical trends.

Brunie et al. [42] proposed architecture for mining geno-medical data in heterogeneous and grid-based distributed infrastructures.

Mahmud Khan et al. [15] focused on decision tree data mining algorithm for medical image analysis. Especially they studied on lung cancer diagnosis through classification of x-ray images.

Podgorelec et al. [21] presented an outlier prediction method for improving performance of classification as part of medical data mining.

Wang et al. [22] applied fuzzy cluster analysis for medical images. They used decision tree algorithm to classify mammography into normal and abnormal cases.

Cheng et al. [17] applied classification algorithm to diagnose cardio vascular diseases. For classification effectiveness they focused on two feature extraction techniques namely automatic feature selection and expert judgment.

Seng et al. [43] introduced web based data mining for the application of telemedicine.

Ghannad-Rezaie et al. [44] presented an approach to integrate PSO rule mining methods and classifier on patient dataset. They used Particle Swarm Optimization technique as well. The results revealed that, their approach is capable of performing surgery candidate selection process effectively in epilepsy.

Bethel et al. [12] developed an association rule learner which is based on the criteria collected from past breast cancer patients. The rule learner is used in a tool by name “Clinical Trial Assignment Expert System”.

Xue et al. [25] proposed and applied Bayesian Network algorithm for diagnosis of an ailment known as Coronary Heart Disease (CHD).

Abraham et al. [26] proposed discrimination techniques to improve the accuracy of classification of medical data using Naive Bayesian classifier algorithm.

Jyoti Soni et al[1] proposed three different supervised machine learning algorithms. They are Naïve Bayes, k-NN, and Decision List algorithm. These algorithms have been used for analyzing the heart disease dataset. Data mining tool is used for classifying these data. These classified data is evaluated using 10 fold cross validation and the results are compared.

Mohammad Taha Khan et al[2] proposed a prototype model for the breast cancer as well as heart disease prediction using data mining techniques. The data used is the Public- Use Data available on web, consisting of 297 records for heart disease and 297 for breast cancer. There are two decision tree algorithms C4.5 and the C5.0 have been used on these datasets for prediction and performance of both algorithms is compared.

D. Chen, K. Xing et al [3] investigates the statistical analysis of the SEER data and computes survival percentage based on gender, race, geographic area, cancer stage, etc. used SEER data containing records of lung cancer patients diagnosed from 1988 through 1998. They examined the following attributes: AJCC stage, grade, histological type and gender. For each of the first three attributes, they considered four popular values that are generally used in lung cancer studies. The attribute gender had two values: male and female. This gave them 128 (4_4_4_2)possible combinations of values. They applied ensemble clustering on those combinations to get seven clusters and studied survival patterns of those clusters.

SEER data is used by D. Chen, K. Xing et al [4], for patients diagnosed of cancer of lung or bronchus from the year 1988 through 2001. They studied 8 months survivability of lung cancer. They compared penalized logistic regression and SVM for survival prediction of lung cancer, and found that logistic regression resulted in better prediction performance. They also note that SVM modeling is significantly slow, taking hours to train.

Vararuk et al[5]. Have studied the application of data mining techniques on HIV/AIDS data with the purpose of utilizing the data mining results for the management of HIV/AIDS. For the study a total of 2,50,000 records from HIV/AIDS patients in Thailand are used. IBM's Intelligent Miner is used for clustering and association rule discovery. As the researchers indicated, clustering is used in order to identify characteristics of categories of people affected symptoms that may follow a set of existing ones.

The findings of Vararuk et al. Have showed clustering of patients with common characteristics and errors in the data.

A study conducted by Teklu[6] has attempted to investigate the application of data mining techniques on Anti-Retroviral Treatment (ART) service with the purpose of identifying the determinant factors affecting the termination/continuation of the service. This study applied classification and association rules using, J48 and apriori algorithms respectively, on 18740 ART patients' datasets. The methodology employed to perform the research work is CRISP-DM. Finally the investigator proved the applicability of data mining on ART by identifying those factors causing the continuation or termination of the service.

Abe et al. [46] proposed an integrated time-series data mining environment for mining huge amount of medical data for extracting more valuable rule-sets.

Jiquan et al. [47] proposed a framework known as term-mapping to combine multiple medical data sources for data mining.

Barnathan et al. [30] presented a framework for clustering, classification and similarity search of biomedical images or 2D and 3D in nature.

Shusaku et al. [48] proposed multi-scale matching and clustering technique on medical data. Their results revealed that their technique is capable of grouping hepatitis data based on temporal covariance of choline esterase, albumin and platelet.

Maria-Luiza Antonie et al[7] analysed the Breast cancer is the second leading cause of cancer deaths in women today and it is the most common type of cancer in women. This paper also presents some experiments for tumour detection in digital mammography. They examine the use of different data mining techniques, neural networks and association rule mining, for anomaly detection and classification.Results are shows that the two approaches are performed well, obtaining a classification accuracy reaching over 70% percent for both techniques. Additionally, the experiments we conducted demonstrate the use and effectiveness of association rule mining in image categorization.

C4.5 is a well-known classification technique in decision tree induction which has been used by Abdelghani Bellaachia and Erhan Gauven[8] along with two other techniques i.e. Naïve Bayes and Neural Network. They conduct the investigation of the prediction of survivability rate of breast cancer patients using above data mining techniques and it is used in the new version of the SEER Breast Cancer Data. However, the author found the model generated by C4.5 algorithm for the given data has a much better performance than the other two techniques.

Evans et al. [37] stated that hereditary syndromes can be detected automatically using data mining techniques.

Doron Shalvi and Nicholas DeClaris, [10] discussed medical data mining through unsupervised neural networks besides a method for data visualization. They also emphasized the need for preprocessing prior to medical data mining.

In the year 2000 Krzysztof J. Cior [38], bioengineering professor, identified the need for data mining methods to mine medical multimedia content.

Tsumoto [39] identified problems in medical data mining. The problems include missing values, data storage with respect to temporal data and multi-valued data, different medical coding systems being used in Hospital Information Systems (HIS).

Brameier and Banzhaf [9] explored and analyzed two programming models such as neural networks, and linier genetic programming for medical data mining.

Abidi and Hoe [40] proposed and implemented a symbolic rule extraction workbench for generating emerging rule-sets.

Abidi et al. [41] explored the usage of rule-sets as results of data mining for building rule-based expert systems.

Olukunle and Ehikioya [11] proposed an algorithm for extracting association rules from medical image data. The association rule mining discovers frequently occurring items in the given dataset. Shim and Xu [13] proposed a classification method based on Bayesian Ying Yang (BYY) which is a three layered model. They applied this model to classify liver disease through automatic discovery of medical trends.

Brunie et al. [42] proposed architecture for mining geno-medical data in heterogeneous and grid-based distributed infrastructures. Mahmud Khan et al. [15] focused on decision tree data mining algorithm for medical image analysis. Especially they studied on lung cancer diagnosis through classification of x-ray images.

Podgorelec et al. [21] presented an outlier prediction method for improving performance of classification as part of medical data mining. Wang et al. [22] applied fuzzy cluster analysis for medical images. They used decision tree algorithm to classify mammography into normal and abnormal cases. Cheng et al. [17] applied classification algorithm to diagnose cardio vascular diseases. For classification effectiveness they focused on two feature extraction techniques namely automatic feature selection and expert judgment.

Seng et al. [43] introduced web based data mining for the application of telemedicine. Ghannad-Rezaie et al. [44] presented an approach to integrate PSO rule mining methods and classifier on patient dataset. They used Particle Swarm Optimization technique as well. The results revealed that, their approach is capable of performing surgery candidate selection process effectively in epilepsy. Bethel et al. [12] developed an association rule learner which is based on the criteria collected from past breast cancer patients. The rule learner is used in a tool by name "Clinical Trial Assignment Expert System". Xue et al. [25] proposed and applied Bayesian Network algorithm for diagnosis of an ailment known as Coronary Heart Disease (CHD). Abraham et al. [26] proposed discrimination techniques to improve the accuracy of classification of medical data using Naive Bayesian classifier algorithm.

Pradhan and Prabhakaran [50] proposed an approach through association rule mining to mine high-dimensional, time series medical data for discovering high confidence patterns. Karegowda and Jayaram [23] proposed a model to classify diabetic database using two techniques in cascading fashion for classification accuracy. The techniques are known as Correlation based Feature Selection (CFS) and Genetic Algorithm (GA). CHAO and WONG [19] proposed a decision tree learning methodology which could interpret attributes in medical data classification for higher accuracy when compared with Incremental Tree Induction (ITI) algorithm. TANG and TSENG [24] studies three classifiers for medical data mining. They are weighted fuzzy k-NN, fuzzy k-NN, and crisp k-NN to classify diabetic and cancer datasets. Tu et al. [20] proposed an intelligent medical decision support system which provides diagnosis of heart diseases through a decision tree algorithm C4.5 and bagging algorithm Naïve Bayes. Su et al. 2011 [14] explored three techniques namely Back Propagation Network (BPN), C4.5 (decision tree algorithm), and Bayasian Network (BN) for mining medical databases. Hogl [51] introduced a language known as Knowledge Discover Question Language for preparing questions that are used to discover knowledge from medical data. They explored ways and means for intelligent medical data mining.

Accuracy = ———————————

4. ADVANTAGES AND DISADVANTAGES OF DATA MINING IN MEDICAL DOMAIN

4.1 Advantages

Data mining can be advantageous as it can help

- Healthcare insurers detect fraud and abuse.
- Physicians identify effective treatments and best practices.
- Patients receive better and more affordable healthcare services.
- One of the key advantages of using data mining is their speed in working with large data

sets. Generation of quicker report and faster analysis that can increase operational efficiency and reduce operating cost.

• Data Mining can extract predictive information from large databases which is a very important feature of Data Mining.

4.2 Disadvantages

• Heterogeneity of medical data Volume and complexity Physician's interpretation Poor mathematical categorization
• Ethical, Legal and Social Issues
• Data Ownership Lawsuits
• Privacy and Security of Human Data Administrative Issues
• Privacy issues
• Security issues
• Misuse of Information/inaccurate information

5. CONCLUSION

In this paper we survey various data mining techniques that have been employed for medical data mining. Data mining techniques have higher utility in medical data mining as there is voluminous data in this industry. Due to the rapid growth of medical data, it has become indispensable to use data mining techniques to help decision support and predication systems in the field of Healthcare. The medical mining yields required business intelligence to support well informed diagnosis and decisions. This paper has provided the summary of data mining techniques used for medical data mining besides the diseases they classified. It also throws light into the importance of locally frequent patterns and the mining techniques used for the purpose.

6. REFERENCES

[1] Umair Abdullah (2008). Analysis of Effectiveness of Apriori Algorithm in Medical Billing Data Mining1. IEEE.p1-5.

[2] Cong-Rui Ji and Zhi-Hong Deng. (n.d). Mining Frequent Ordered Patterns without Candidate Generation. IEEE. 0 (0), P1-5.

[3] Hai-Tao He and Shi-Ling Zhang. (2007). A New method for Incremental Updating Frequent patterns mining. IEEE. 0(0), p1-4.

[4] Carson Kai-Sang Leung • Christopher L. Carmichael and Boyu Hao. (2007). Efficient Mining of Frequent Patterns from Uncertain Data. IEEE. 0 (0), p489-494.

[5] Shariq Bashir, Zahid Halim, A. Rauf Baig. (2008). Mining Fault Tolerant Frequent Patterns using Pattern Growth Approach. IEEE. 0 (0), p172-179.

[6] Sunil Joshi and Dr. R. C. Jain. (2010). A Dynamic Approach for Frequent Pattern Mining Using Transposition of Database. IEEE. 0 (0), p498-501.

[7] Thanh-Trung Nguyen. (2010). An Improved Algorithm for Frequent Patterns Mining Problem. IEEE. 0 (0), p503-507.

[8] Xiaoyong Lin and Qunxiong Zhu. (2010). Share-Inherit: A novel approach for mining frequent patterns. IEEE. 0 (0), p2712-2717.

[9] Markus Brameier and Wolfgang Banzhaf. (2001). A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining. IEEE.p1-10.

[10] Doron Shalvi and Nicholas DeClaris., (n.d). An Unsupervised Neural Network Approach to Medical Data Mining Techniques. IEEE. 0 (0), p1-6.

[11] Adepele Olukunle and Sylvanus Ehikioya, (n.d). A Fast Algorithm for Mining Association Rules in Medical Image Data. IEEE. p1-7.

[12] Cindy L. Bethel and Lawrence O. Hall and Dmitry Goldgof (n.d). Mining for Implications in Medical Data. IEEE. p1-4.

[13] Jeong-Yon Shim, Lei Xu (n.d). MEDICAL DATA MINING MODEL FOR ORIENTAL MEDICINE VIA BYY BINARY INDEPENDENT FACTOR ANALYSIS. IEEE. p1-4.

[14] Jenn-Lung Su, Guo-Zhen Wu, I-Pin Chao (2001). THE APPROACH OF DATA MINING METHODS FOR MEDICAL DATABASE. IEEE. p1-3

[15] Safwan Mahmud Khan Md. Rafiqul Islam Morshed U. (n.d). Medical Image Classification Using an Efficient Data Mining Technique. IEEE, p1-6.

[16] Yanwei Xing, Jie Wang and Zhihong Zhao (2007). Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease. IEEE. p1-5.

[17] Tsang-Hsiang Cheng, Chih-Ping Wei, Vincent S. Tseng (n.d). Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches . IEEE. p1-6.

[18] Mohammad Saraee, George Koundourakis, Babis Theodoulidis. (n.d). EASYMINER: DATA MINING IN MEDICAL DATABASES. IEEE. p1-3.

[19] SAM CHAO, FAI WONG, "AN INCREMENTAL DECISION TREE LEARNING

METHODOLOGYREGARDING ATTRIBUTES IN MEDICAL DATA MINING". Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
[20] My Chau Tu AND Dongil Shin (2009). A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms. IEEE. p1-5.
[21] Vili Podgorelec, Marjan HerikoMaribor, (n.d). Improving Mining of Medical Data by Outliers Prediction. IEEE. p1-6.
[22] Shuyan Wang Mingquan Zhou Guohua Geng (n.d). Application of Fuzzy Cluster Analysis for Medical Image Data Mining. IEEE. p1-6.
[23] Asha Gowda Karegowda M.A.Jayaram (2009). Cascading GA & CFS for Feature Subset selection in Medical Data Mining. IEEE. p1-4.
[24] Graduate Institute of Applied Information Sciences (2009). MEDICAL DATA MINING USING BGA AND RGA FOR WEIGHTING OF FEATURES IN FUZZY K-NN CLASSIFICATION. IEEE. p1-6.
[25] Weimin Xue, Yanan Sun, Yuchang Lu (n.d). Research and Application of Data Mining in Traditional Chinese Medical Clinic Diagnosis. IEEE.p1-4.
[26] Ranjit Abraham, Jay B.Simha, Iyengar (n.d). A comparative analysis of discretization methods for Medical Data mining with Naïve Bayesian classifier. IEEE. p1-2.
[27] Syed Zahid Hassan and Brijesh Verma,(n.d). A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases. IEEE. p1-6.
[28] Michele Berlingerio (n.d). Mining Clinical Data with a Temporal Dimension: a Case Study. IEEE. p1-8.
[29] Wojciech Froelich, Alicja Wakulicz-Deja (2009). Mining Temporal Medical Data Using Adaptive Fuzzy Cognitive Maps. IEEE. P1-8.
[30] Michael Barnathan, Jingjing Zhang, Vasileios (n.d). A WEB-ACCESSIBLE FRAMEWORK FOR THE AUTOMATED STORAGE AND TEXTURE ANALYSIS OF BIOMEDICAL IMAGES. IEEE. p1-3.
[31] Sarojini Balakrishnan (n.d). SVM Ranking with Backward Search for Feature Selection in Type II Diabetes Databases. IEEE. p1-6.
[32] Arun K Pujari "Data Mining Techniques", Edition 2001.
[33] M. Ilayaraja Department of Computer Science & Engineering Alagappa University Karaikudi, India ilayarajaalu@gmail.com. (2013). Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm. IEEE. 0 (0), p1-6.
[34] J. C. Prather, D. F. Lobach, L. K. Goodwin, J. W.Hales , M. L. Hage¸ W. Edward Hammond, "MedicalData Mining: Knowledge Discovery in a Clinical DataWarehouse", 1997.
[35] HAI-BING MA, JIN ZHANG, YlNG-JIE FAN, YUN-FA W. (2004). MINING FREQUENT PATTERNS BASED ON IS+-TREE. IEEE. 0 (0), P1208-1213.
[36]. Goodwin L, Prather J, Schlitz K, Iannacchione My Hammond W, Grzymala J, DataMining Issues for Impproved Birth Outcomes, Biomed. Science Instrum, 34, 1997, pp. 291-296.
[37]. Evans S, Lemon S, Deters C, Fusaro R and Lynch H, Automated Detection of hereditary Syndromes Using Data Mining, Computers and Biomedical Research 30, 1997, pp. 337-348.
[38] Krzysztof J. Cior , Medical Data Mining and Knowledge Discovery. (n.d). From the guest Editor. IEEE. p1-2
[39] Shusaku Tsumoto (n.d). Problems with Mining Medical Data. IEEE. p1-2.
[40] Syed Sibte Raza Abidi Kok Meng (n.d). Symbolic Exposition of Medical Data-Sets: A Data Mining Workbench to Inductively Derive Data-Defining Symbolic Rules. p1-6.
[41] S. S. R. Abidi, K. M. Hoe, A. Goh, "Analyzing data clusters: A rough set approach to extract cluster defining symbolic rules, Fisher, Hand, Hoffman, Adams (Eds.) Lecture Notes in Computer Science: Advances in Intelligent Data Analysis, 4th Intl. Symposium, IDA-01. Springer Verlag: Berlin, 2001.
[42] Lionel Brunie, Maryvonne Miquel, Jean-Marc Pierson, and Anne Tchounikine, "Information grids: managing and mining semantic data in a grid infrastructure; open issues and application to geno-medical data. 2003, 14th International workshop on Database and Expert Systems Applications.
[43]Michael Barnathan, Jingjing Zhang, Vasileios (n.d). A WEB-ACCESSIBLE FRAMEWORK FOR THE AUTOMATED STORAGE AND TEXTURE ANALYSIS OF BIOMEDICAL IMAGES. IEEE.
[44] Liu Jiquan Deng Wenliang Xudong Lu (n.d). Liu Jiquan Deng Wenliang Xudong Lu Huilong Duan College of Biomedical Engineering & Instrument Science Zhejiang University Hangzhou 310027, IEEE. p1-4.
[45] Shusaku Tsumoto (n.d). Problems with Mining Medical Data. IEEE. p1-2.
[46] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer science and Engineering, Vol. 3, No. 6, June 2011.
[47] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent, F. Massin, "A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining", International Journal of Applied Engineering Research, Vol. 7 No. 11 (2012), pp. 1-6.
[48] D.Chen, K. Xing, D. Henson, L. Sheng, A. Schwartz, and X.Cheng. "Developing prognostic systems of cancer patients by ensemble clustering". Journal of Biomedicine and Biotechnology, 2009.

[49] F. D. "Machine learning methods in the analysis of lung cancer survival data". DIMACS Technical Report 2005-35 February 2006.
[50] Ruben, D.C.J.,"Data Mining in Healthcare: Current Applications and Issues". 2009.
[51]Maria-LuizaAntonie ,Osmar R.Za¨iane, Alexandru Coman, "Application of Data Mining Techniques for Medical Image Classification",Auguest26, 2001.
[52] Eldon Y. Li," Artificial neural networks and their business applications" Information & Management 27 (1994) 303- 313.
[53] Bellaachia Abdelghani 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,06
[54] A.Shameem Fathima,D.Manimegalai and Nisar Hundewale," A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, November 2011.

Author Biograhy

V. Gayathri working as Assistant Professor in G.T.N. Arts College, Karur Road, Dindigul. Tamilnadu. She has 8 years of experience in Industry and academic fields. She completed her MCA in Anna University and Master of Philosopy(Computer Science) in PRIST University, Thanjavur, Tamilnadu.Her area of Interest includes Data Mining. She has also life member for several association and society.

M.Chandamona working as Assistant Professor and Head of BCA Department in G.T.N.Arts College, Dindigul. Have 3 years of Industrial experience in Odyssey Technologies ,Coimbatore.Have 15+ years of experience in academic field. Worked in Karpagam College for 1 year and 6 months and CMS College for 3 years, Coimbatore. Working for past 11 years in G.T.N Arts college, Dindigul. Did graduation in MVM College,Madurai Kamaraj University.Did Post graduation in Alagappa University, Karaikudi. Did Master of Philosophy in Sri Vasavi College,Bharathiar Universiy,Erode. Area of interest includes Network security, data mining and Image Processing.

S.Banu Chitra working as Assistant Professor in G.T.N. Arts College, Karur Road, Dindigul. Tamilnadu. She has 8 years of experience in Industry and academic fields. She completed her M.Phil(Computer Science) in Madurai Kamaraj University, Madurai. and MCA in Bharathiar University, Coimbatore. Her area of Interest includes Digital Image Processing . She has also life member for several association and society.