ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 2, February 2014

A Literature Review in Health Informatics Using Data Mining Techniques

Dr. D. P. Shukla[1]; Shamsher Bahadur Patel[2]; Ashish Kumar Sen[3]

Prof and Head[1]; Research Scholar[2]; Research Scholar[3]

[123]Dept. of Math/Computer, Govt. PG Science College, Rewa, (M.P.)-India;

ABSTRACT

In this paper we present an overview of the applications of data mining in administrative, clinical, research, and educational aspects of Health Informatics. The current or potential applications of various data mining techniques in Health Informatics are illustrated through some case studies from published literature. . Data mining techniques such as clustering, classification, regression, association rule mining, CART (Classification and Regression Tree) are widely used in healthcare domain. Data mining algorithms, when appropriately used, are capable of improving the quality of prediction, diagnosis and disease classification. The main focus of this paper is to analyze data mining techniques required for medical data mining especially to discover locally frequent diseases such as heart ailments, lung cancer, breast cancer and so on.

Keywords

Data mining, frequent patterns, data mining techniques, medical data mining

1. INTRODUCTION

Health Informatics is a rapidly growing field that is concerned with applying Computer Science and Information Technology to medical and health data. With the aging population on the rise in developed countries and the increasing cost of healthcare, governments and large health organizations are becoming very interested in the potential of Health Informatics to save time, money, and human lives. As a relatively new field, Health Informatics does not yet have a universally accepted definition. The American Medical Informatics Association defined health Informatics as "all aspects of understanding and promoting the effective organization, analysis, management, and use of information in health care"[52]. Similarly, the Canada's Health Informatics Association definition of Health Informatics is "Intersection of clinical, IM/IT and management practices to achieve better health"[53]. These are both broad definitions that cover a wide range of technologies, from developing electronic patient record data warehouses to installing wireless networks in hospitals. A more specific definition is provided by the National Library of Medicine, which defines Health Informatics as "the field of information science concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine"[53]. Note that here, Health Informatics is limited to "analysis and dissemination of medical data", and would not cover pure IT practices such as installing a network in a hospital. Zaiane provides an even more specific definition, which divides Health Informatics into four subfields:

"Health Informatics is the computerization of health information to support and optimize (1) administration of health services; (2) clinical care; (3) medical research; and (4) training. It is the application of computing and communication technologies to optimize health information processing by collection, storage, effective retrieval (in due time and place), analysis and decision support for administrators, clinicians, researchers, and educators of medicine."

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" by Fayyad [52]. 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. 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 methodology of research used in this study in chapter two, we classify them with different criterions in chapter three, then we identify the most used algorithms for disease diagnosis and prognosis, 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], Naive Bayesian [26], combination of K-means, Self Organizing

ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 2, February 2014

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. AN OVERVIEW OF HEALTH INFORMATICS AND APPLICATIONS OF DATA MINING

As mentioned in the introduction, Health Informatics can be divided into four main subfields:

1. Clinical care
2. Administration of health services
3. Medical research
4. Training.

The following subsections present an overview of each subfield of health Informatics, and how data mining is, or can be, applied to extend and improve each subfield.

2.1 Clinical Care

Physicians and nurse practitioners make diagnostic decisions and treatment recommendations based on history, medical imaging, lab results and other text or multimedia records of patients. Health informatics allows doctors to have faster access to more relevant information, and thus make more optimal decisions. For instance, a centralized patient record database will allow a physician in a local clinic to have access to all the relevant medical records of the patient, anywhere in the country. Furthermore, applying data mining techniques on the centralized database will give doctors analytical and predictive tools that go beyond what is apparent from the surface of the data. For instance, a new practitioner can query for all the decisions that previous practitioners have made on a similar case. Similarly, a predictive model can advise doctors whether a certain case would be better treated as an outpatient or an inpatient.

2.1.1 Clinical Decision Support Systems

The applications of Health Informatics in clinical care decision-making are known as (Computer based) Clinical Decision Support System (CDSS)[1] Shortliffe defines a decision support system as "any computer program that is designed to help health professionals to make clinical decisions" [44 as cited in 34]. Applications of Clinical Decision Support Systems can be categorized into:

1. **Information retrieval**: CDDS can offer search capabilities for medical queries. For instance the "antibiotic assistant" of HELP system (introduced in section 2.1.1.1) allow doctors to query the hospital experience with previous infections through the last five years [56].
2. **Alerting systems:** A useful application of CDSS is to monitor inputs and check them for predetermined triggers [57]. These alert systems can be simple, like predefined drug-drug or drug allergy conflicts, or complex, such as alerts based on analysis of various lab results and comparison with expected result protocols.
3. **Reminders**: unlike alerts that are triggered by a specific change in input data, reminders are triggered by passage of time and are used for periodic tasks such as immunization or diabetes tests [57].
4. **Suggestion Systems**: Unlike alerts, which indicate predetermined conditions in input data, suggestion systems are interactive processes that suggest action oriented messages based on their medical knowledge base.
5. **Prediction Models:** CDSS prediction models can be categorized into diagnosis (defined as "aiding in the determination of the existence or nature of a disease" [55] and prognosis (defined as "the forecast of the probable outcome of an illness'' [55]) [57]. An example of a diagnosis predictor is a model that detects *nosocomial* hospital infections based on information from Microbiology laboratory, nurse charting, and other sources. APACHE, introduced in section 2.1.1.2, is an example of a prognosis predictor which predicts ICU mortality based on a number of physiological variables.

2.1.1.1 Case Study: HELP system

Health Evaluation through Logic Processing (HELP) system is an example of a Clinical Decision Support System that includes alerting systems, suggestion systems, and prediction models [56]. An example of an alerting system used in HELP is a model that monitors patient laboratory results, and has simple rule-based triggered to detect anomalies. A suggestion system included in HELP is a set of computerized protocols for managing care of Adult Respiratory Distress Syndrome (ARDS) patients. Both alerting and suggestion systems in HELP are rule-based models, developed by physicians, nurses, and specialists in medical informatics. HELP includes two types of prediction models. One of these models is rule-based models, such as the one used in the Adverse Drug Events (ADE) detection system. The ADE detection system predicts the possibility of a drug reaction based on patient history and a set of predefined protocols. Aside from rule based models, some prediction models in HELP use logistic regression, *e.g.* the model that predicts nosocomial hospital infections based on a number of risk factors. HELP system has been developed and tested for more than 25 years and it is currently in use in many of the 20 hospitals operated by Intermountain Healthcare (IHC) [31 as cited in 9]

2.2 Administration of Health Services

Administrators of health care organizations make hundreds of critical decisions on daily basis. As in any administrative position, the quality of these decisions directly depends on the quality of the information that the decisions are based on. For example, the administrators in a hospital need to decide on the amount of supplies and number of staff and free beds required for an upcoming month. To make this decision, the administrators require an accurate prediction of the number of patients to expect during the coming month, and an approximation of how long each patient will remain in the hospital. As another example, the federal and provincial health administrators need to decide whether a disease outbreak is in progress, and if so, what preventive measures will be most effective against it. To make these decisions, the administration requires a system that can accurately predict a disease outbreak, and also model the cost and benefit of different preventive measures. The following case study illustrates the applications of data mining techniques on epidemic detection. More examples of administrative decision support will be discussed in Section 4, where electronic patient records and various data warehousing techniques are introduced.

2.2.1 Case Study: detecting disease outbreak

In "Decision Theoretic Analysis of Improving Epidemic Detection", Izadi and Buckeridge introduce a method to improve existing threshold-based epidemic detection methods by using POMDPs (Partially Observable Markov Decision Processes) [24]. The main idea is that the potential costs and effects of intervention can be quantified and be used to optimize the alarm function. Furthermore, the intermediate investigation steps, such as asking for more systematic studies, or more investigation done by human expert, can also be quantified in terms of cost and effect. Based on these cost and effects, the system can learn to recommend the optimal action. While the paper concludes that POMDPs can improve the accuracy of the current outbreak detection methods, the current level of false alarms (3 false alarms in every 100 days) seems to be unacceptable for practical use. Similarly, Cooper et al. investigates the use of Bayesian Networks for outbreak detection, focusing on modeling non-contagious outbreak diseases, such as airborne anthrax [13]. The Bayesian network is divided into 3 groups: global (G) interface (I) and people (P). Furthermore, in order to make the algorithm scalable, people with the same attributes are grouped in the same class. The network is evaluated based on data generated by a simulator. Given weather conditions from Historical meteorological conditions for a region, parameters for location and amount of airborne anthrax, a Gaussian plume model derives the concentration of anthrax spores that are estimated to exist in each zip code. The authors compare a no spatial model with a spatial model and conclude that with spatial data they can get better results based on false positive rate.

2.3 Medical Research

Most current successful applications of data mining in Health Informatics are in the subfield of medical research. The reason is that most of the current health related data are stored in small datasets scattered through various clinics, hospitals, and research centers. However, most applications of data mining in clinical and administrative decision support systems require homogeneous and centralized data warehouses (see section 3). On the other hand, data mining methods can still be successfully applied on small and scattered datasets, and help researchers extract insightful patterns, cause and effect relationships, and predictive scoring systems from currently available data.

2.3.1 Case Study: drug exposure side effects from mining pregnancy data

Chen et al. investigate the possible effects of multiple drug exposures at different stages of pregnancy on preterm birth, using Smart Rule, a data mining technique for generating associative rules [11]. In this work, two subsets of Danish National Birth Cohort (DNBC) dataset are used. The first subset contains 4454 records including 1000 women who were depressed and/or exposed to various active drugs. This set is used for finding the side effects of anti-depression drugs. The second subset contains 6231 records, including 414 preterm cases. This set is used for finding side effects of multiple types of drugs. The authors develop a tree hierarchical model for organizing the generated rules, in order to ease the recognition of interesting rules by human experts. Using this system, the authors claim that they are able to find novel and interesting rules.

2.3.2 Case Study: Association rules and decision trees for disease prediction

Ordonez applies different classifiers, associative classifier and decision trees, for predicting the percentage of vessel narrowing (LDA, RCA, LCX and LM) compare to a healthy artery [35]. The dataset contains 655 patient records with 25 medical attributes. Three main issues about mining associative rules in medical datasets are mentioned in this work. A significant fraction of association rules are irrelevant and most relevant rules with high quality metrics appear only at low support. On the other hand, the number of discovered rules becomes extremely large at low support. Hence, association rules are used with constraints. Each item corresponds to the presence or absence of one categorical value or one numeric interval. First constraint is that there is a limit on the maximum item-set size. Second, the items are grouped and in each association, there is at most one from each group. The third constraint is that each item can only appear in antecedent or consequent. The result from associative classifier is compared with two decision tree algorithms: CN4.5 and CART. The authors demonstrate that associative rules can do better than decision trees for predicting diseased arteries.

2.4 Education and Training

The fourth subfield of health informatics is related to educating new healthcare professionals and retraining and keeping the current staff up-to-date with recent advances in technology. The education and training subfield of Health Informatics can be viewed as an instance of the rapidly growing field of e-learning. An increasing interest in applying data mining techniques to e-learning has emerged in recent years, and some of the early applications show promising results [38]. Data mining techniques can benefit all three groups of people who are in contact with a learning system: students, educators, and administrators [38]. Data mining techniques can monitor the success of students at various learning tasks, and recommend relevant resources, materials, and learning paths to achieve a more successful learning experience. For educators, data mining techniques can provide objective feedback of the structure and the content of a course, discover the learning patterns of the students, and cluster learners

ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 2, February 2014

into smaller groups that have similar educational habits and needs. Administrators benefit from data mining techniques by learning about the behavior of their users, so they can optimize the servers, distribute network traffic, and learn about the overall effectiveness of the offered educational programs.

2.4.1 Case Study: Homer, an online learning community

Homer is a centralized e-learning system and an Internet community, developed for the medical students of the University of Alberta [5]. Homer provides online access to a variety of learning materials, including medical dictionaries, demonstration videos, and faculty presentations. One important feature of Homer is the lifetime membership, which grants medical students continued access to learning materials after graduation [18].

3. REVIEW WORK DONE IN THIS FIELD

Data mining is a process of analyzing voluminous data in various perspectives in order to bring about trends or patterns that lead to business intelligence [32]. Data mining plays an important role in IT as it discovers knowledge from historical data of various domains. For instance data mining can be used to mine medical data as Healthcare domain produces huge amount of data about patients, diseases, diagnosis, medicine and so on. By applying data mining techniques in Healthcare domain, the administrators can improve the QoS (Quality of Service) by discovering latent potentially useful trends required by medical diagnosis [33]. Data mining is useful in medical applications such as medications, medical tests, prediction of surgical procedures, and discovery of relationships between pathological data and clinical data [34]. Apriori and FPGrowth are the most widely used frequent pattern mining algorithms [35]. These two algorithms and algorithms based on them are studied in [2], [3], [4], [5], [6], [7], and [8]. These two algorithms are also used in medical data mining. Goodwin et al. [36] applied data mining techniques for birth outcomes. 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.

Hassan and Verma [27] proposed a hybrid approach for classification of medical data which combines K-means, Self Organizing Map (SOM) and Naïve Bayes with NN based classifier. Tsumoto [45] studied multi-stage medical diagnosis using experts' diagnostic rules and diagnostic taxonomy. They focused on automatic grouping of medical knowledge extracted from clinical database. Berlingerio et al. [28] studied Time Annotated Sequences (TAS) algorithm for mining medical data with temporal dimensions. The extracted patterns exhibited the attribute relationships in time domain which helps in accurate diagnosis. Xing et al. [16] developed data mining techniques for predicting the probability of survival of CHD patients. To achieve this they combined three prediction models such as SVM (Support Vector Machine), Artificial Neural Networks (ANN), and Decision trees using C4.5 or ID3, CART and C5.

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. Hai Wang, and Shouhong Wang [49] studied on the role of medical experts in medical data mining. Medical experts can give expert advice that can be

ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 2, February 2014

used as input in medical data mining.

Abdullah et al. [1] applied apriori algorithm for medical data mining. They extracted frequent item sets by analyzing associations between treatments and diagnosis. Saraee et al. [18] applied data mining techniques to medical data pertaining to military with respect to mortality rate in children due to accidents. They used CART algorithm to generate a decision tree. Balakrishnan and Narayanaswamy [31] presented feature selection using SVM for classifying diabetes databases. Drugs and health effects are mined by Froelich and Wakulicz-Deja [29] using adaptive FCM (Fuzzy Cognitive Maps). Their work has led to improved decision support and planning in Healthcare domain.

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 weighting 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 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.

4. SUMMARY OF TECHNIQUES FOR MEDICAL DATA MINING

Data mining techniques have shown significant improvement in medical industry in terms of prediction and decision making with respect to various diseases like cancer, cardio vascular abnormalities, diabetes, and others. Table 1 summary the medical data mining, its areas of application and the utility of the techniques.

Table1. Summary of medical data mining techniques

References	Techniques	Utility	Disease
[1], [2], [3], [4], [5], [6], [7], and [8]	Appriori and FPGrowth	Association rule mining for finding frequent item sets (diseases) in medical databases.	
[9]	Genetic Algorithm	Classification of medical data.	Diabetic Diseases
[9], [10]	Neural Networks	Extracting patterns, detecting trends	
[11], [12]	Association Rule Mining	Finding frequent patterns	
[13]	Bayesian Ying Yang (BYY)	Classification	Liver diseases
[14], [15], [16], [17], [18], [19], [10]	Decision Tree Algorithms such as ID3, C4.5, C5, and CART.	Decision Support	
[21]	Outlier Prediction Technique	For improving classification accuracy	
[22]	Fuzzy cluster analysis	Analyzing medical images	
[17], [23], [24]	Classification Algorithm	Disease classification	Cardio Vascular Diseases
[14], [25]	Bayesian Network algorithm	Modeling and analysis of medical data	Coronary Heart Disease
[26]	Naive Bayesian	Improving classification accuracy.	Coronary Heart Disease
[27]	Combined use of K-means, SOM and Naïve Bayes	Accurate Classification of medical data.	
[28], [29]	Time Series Technique	Medical diagnosis	
[16]	combination of SVM, ANN and ID3	Medical data classification	

ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 2, February 2014

[30]	Clustering and classification	Clustering and classification of biomedical databases	
[16], [31]	SVM	Disease Classification	Diabetes
[29]	Fuzzy Cognitive Maps	Drugs and Health effects classification	
[24]	k-NN	Classification of diseases	Diabetes, Cancer

5. CONCLUSIONS

In this review we identified and evaluated the most commonly used DM algorithms resulting as well-performing on medical databases, based on recent studies. 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. 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 requent 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 Datamining 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*.

[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 definingsymbolic rules, Fisher, Hand, Hoffman, Adams (Eds.) Lecture Notes in Computer Science: Advances inIntelligent 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] Wong Kok Seng, Rosli Bin Besar, Fazly Salleh Abas trosli, (n.d). Collaborative Support for Medical Data Mining in Telemedicine. *IEEE*. p1-6.

[44] M. Ghannad-Rezaie, H. Soltanain-Zadeh, M.-R. Siadat, K.V. Elisevich. (2006). Medical Data Mining using Particle Swarm Optimization for Temporal Lobe Epilepsy. *IEEE*. p1-8.

[45] Shusaku Tsumoto , (n.d). Problems with Mining Medical Data. *IEEE*. p1-2.

[46] Hidenao Abe AND Hideto Yokoi (n.d). Developing an Integrated Time-Series Data Mining Environment for Medical Data Mining. *IEEE*. p1-6.

[47] 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, Chinaliujiquan@gmail.com . *IEEE*. p1-4.

[48] Shusaku Tsumoto (n.d). Problems with Mining Medical Data. *IEEE*. p1-2.

[49] Hai Wang, Shouhong Wang 1. (n.d). Medical Knowledge Acquisition through Data Mining. *IEEE*. 0 (0), p1-4.

[50] Gaurav N. Pradhan AND B. Prabhakaran (n.d). ASSOCIATION RULE MINING IN MULTIPLE, MULTIDIMENSIONAL TIME SERIES MEDICAL DATA. *IEEE*. p1-4.

[51] Oliver Hogl, Michael Müller (2001). On Supporting Medical Quality with Intelligent Data Mining. *IEEE*. p1-10.

[52] American Medical Informatics Association, http://www.amia.org/informatics/.

[53] Canada's Health Informatics Association, http://www.coachorg.com/.

[54] National Library of Medicine, http://www.nlm.nih.gov/tsd/acquisitions/cdm/subjects58.html.

[55] Canadian Institute of Health Research, http://www.mshri.on.ca/colorectalcancer/definitions.html, 05/25/2008

[56] Berner E., "Clinical Decicion Support Systems". Springer Science+Business Media, 2007 .

[57] Greens R., "Clinical Decision Support". Elsevier Inc., 2007.