ISSN No. 0976-5697

Volume 8, No. 5, May-June 2017

International Journal of Advanced Research in Computer Science

REVIEW PAPER

Available Online at www.ijarcs.info

Significance of Data Mining in Disease Classification and Prediction for Mining Clinical Data : A Review

Dr. Narander Kumar
Department of Computer Science
Babasaheb Bhimrao Ambedkar University
Lucknow, India

Sabita Khatri
Department of Computer Engineering
University Institute of Engineering & Technology
Babasaheb Bhimrao Ambedkar University
Lucknow, India

***Abstract*:** Data Mining is a well established area of research that has become increasingly popular in health domain in recent years. It plays a vital role the health care towards uncovering latest trends specifically in early disease predictions. Data Mining is now becoming helpful for researchers and scientist towards gaining novel and deep insights of any large biomedical datasets. Uncovering new biomedical and healthcare related knowledge in order to support clinical decision making, is another dimension of data mining. Early disease prediction has now become the most demanding area of research in health care sector. As health care domain is bit wider domain and having different disease characteristics, different techniques have their own prediction efficiencies, which can be increased and tuned in order to get into most optimize way. In this research work, authors have comprehensively compared different data mining techniques and their prediction capability on different set of diseases datasets. Authors also have discussed different Data Mining Techniques such as classification, clustering, association, regression and their applicability and prediction efficiency specifically in health domain.

Keywords: Data Mining, Classification, Clustering, Healthcare

I. INTRODUCTION

Most crucial and challenging task in the field of medical sciences is getting the right diagnosis at right time. Getting the right diagnosis is big hurdle and frustrating task, it is specially challenging for someone with less commonly known disease. Many people don't get a diagnosis and some time misdiagnosed or have multiple, contradicting diagnoses from doctors. During earlier days, medical diagnosis was based upon information collected from patient history and physical examination of person and such processes also need diagnosis test for further diagnosis of patient. Nowadays some intelligent system has been developed by taking the help of such systems disease diagnosis can be done in earlier stages. Data Mining is a process of digging interest or meaningful information and knowledge from spam of data. Data mining has wide scope in various domain like in retail, finance, manufacturing, medicine, telecommunication, bioinformatics, web mining however out of which health care or medical domain has recently got wider scope as there is terabytes of data is available but still meaningful information is missing. The medical environment is data rich but weaker in information in terms of knowledge and robust tool to identify hidden patterns of knowledge in specific domain. Here data mining techniques has a great extent of applicability in the field of earlier disease prediction and prognosis and biomedical and health care.

Data mining has shown tremendous growth the last few years and already proven analytical capabilities in various research fields like,

- In Retail Domain like Market basket analysis, Customer relationship management (CRM)
- In Financial Domain like Credit scoring, fraud detection
- In Manufacturing Domain like Optimization, troubleshooting
- In Medicine Domain: Medical diagnosis
- In Telecommunication Domain like Quality of service optimization
- In Bioinformatics Domain like Motifs, alignment
- In Web mining Domain like Search engines

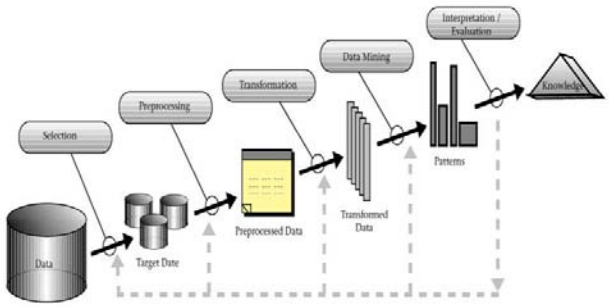


Fig. 1: An Overview of Steps Compose KDD Process [D. Fayyad et al., 1996] [41]

DM techniques have a proven set of scope towards disease diagnosis and revealing hidden biomedical and health care patterns. This leads to the development of intelligent systems and decision support systems in Healthcare domain to ensure better diagnosis, prediction of diseases. Through massive literature survey it is found that applicability of data mining techniques are more intended towards predicting heart diseases, lung cancer, and breast cancer and so on. The advantages of applying DM techniques in medical domain includes effective treatment identification and disease predictions at its early phases, Patients receive better and more affordable health care services, Generation of quicker reports and faster analysis, Increased operational efficiency and reduce operating cost whereas the disadvantages includes Heterogeneity of medical data volume and physicians explanations, Ethical, Legal and

social issues, Privacy and security of data, Inaccurate information.

II. KDD PROCESS

The knowledge discovery in database is referred as KDD process. It is a process of selecting the knowledgeable data or data of interest from different repositories of data. KDD has high level of applicability in data mining. The process includes selecting the relevant data, processing, transformation of the data into relevant information and finally extracting hidden information from spam of data. KDD process encompasses the following steps:

A. Data Cleaning
It is the process of selecting the relevant information by removal of noisy and inconsistent data.

B. Data Integration
In this process heterogeneous data sources are integrated.

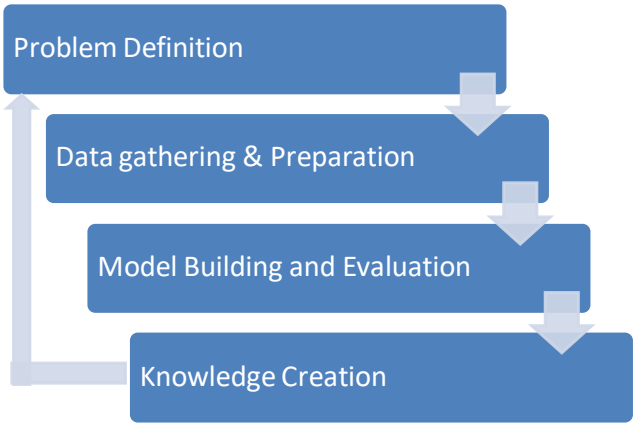
C. Data Selection
Data Selection involves retrieval of relevant data gathering for further analysis.

D. Data Transformation
Data collected from different repositories may have not appropriate for applying mining process initially. The data comes from heterogeneous sources so it should first converted or transformed in appropriate form so the data mining techniques can be implemented and patterns to be analyzed.

E. Data Mining
In this phase various data mining algorithm is applied in order to mine appropriate knowledge pattern.

F. Pattern Evolution
Pattern evolution includes finding the relevant patterns of information hidden in the data.

G. Knowledge Represntation
This phase represent knowledge in a more sophisticated way applied to a particular domain.



Fig, 2: Major steps involved in KDD process

III. TECHNIQUES USED IN DATA MINING

For the knowledge discovery and different pattern identification, the different algorithms and techniques of data mining are used. Majorly these techniques and algorithms can be categorized into classification and clustering, regression, decision tree, genetic algorithms and neural networks. A brief description of all above specified techniques is given below:

A. Association

Association is one of the techniques of data mining in which association function use to find a relation between co-occurrence of item in crowd of data. These rules or function use to predict occurrence on one item based on occurrence of other item. Association techniques are used to analyze and predict behavior of customer as well as play a significant role in analyzing shopping basket pattern and product clustering and catalog. Multilevel association, Multidimensional association, and Quantitative association all these rules come under Association. Association rules algorithms intend to production of rules with confidence values less than one.. Linear Regression, Multivariate Linear Regression, Nonlinear Regression, Multivariate Nonlinear Regression are major types of association rules.

B. Classification

classification techniques are supervised learning methods are most widely used tin data mining to classify the data in raw data, a value is assigned to each item in set of data to group them in a class ,these classifier models are mathematical techniques which are used to classify the data. This technique is more suitable for detecting any sort of discrepancy and risks analysis involved with a particular data set. Basically this approach uses decision tree like classification algorithms.

Classification process further can be categorized into two further process called learning and classification. In learning phase data is analyzed by classification algorithm first whereas in classification this test data is used to identify the accuracy of said classification rules if accuracy comes within acceptable bands than this rule can be applied on to newer data sets. The algorithms then encode these parameters into a model called a classifier. Classification techniques includes decision tree, Bayesian classification, neural networks, support vector machine, association based classification.

C. Clustering

Clustering is a data mining technique which is used to build the cluster of those objects having similarity in their behavior and cluster of the items which having dissimilarity by using the clustering techniques. Clustering is somehow different from classification as clustering is unsupervised learning methods in which cluster is form on base of similarity however in classification tags are assign to put items in a class. Clustering techniques are widely used in field of image processing, data analysis, pattern recognition, and market strategies. Clustering is used to identify similar classes of objects and dense and spare regions in object space along with correlations among the data attributes. Clustering process is an alternative of classification used for attribute subset selection and classification. The methods which come under clustering are Partition, hierarchical agglomerative, density based, grid based and model based.

D. Prediction

prediction is one of the data mining which is based on predict the future behavior of items and based on current and historical data to create a model of prediction .These techniques are used in market analysis. Regression techniques are used for predications. Regression is used to find the

relationship between independent and dependent variables. Independent variables are those which are already known and response variables are those which we want to predict. Some complex problems depend on complex interactions of multiple predicator variables. Some techniques like logistic regression, decision trees, or neural nets is quite helpful for future value forecasting and to solve complex problems. Even classification and regression models can be created by neural networks.

Prediction includes linear regression, multivariate linear regression, nonlinear regression, multivariate nonlinear regression. Figure 3 given below represents the summarized version of different algorithms applicable to different techniques in data mining.

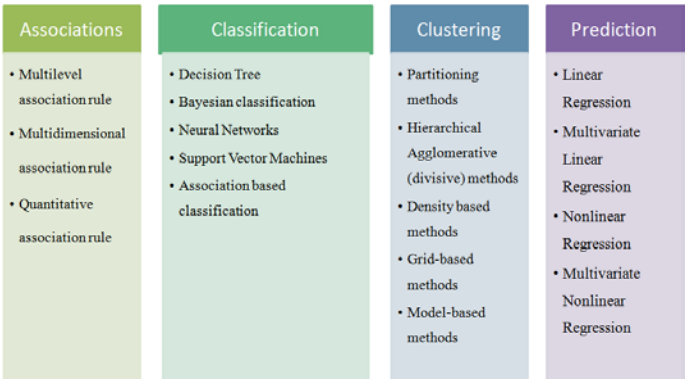


Fig. 3: Major techniques used in data mining

IV. LITERATURE SURVEY

Data Mining involves analyzing large data to discover trends and patterns that yields to business intelligence [31]. Data mining also plays a significant role in discovering knowledge related to various domains. Another dimension of data is the mining of medical data as Healthcare domain, which produces large amount of data about historical records of patients and their data of medical history, prior diseases, diagnosis and required information. With this approach the Healthcare domain, can improve their overall quality related to patient services and early detection and diagnosis of various applicable diseases [20]. The major application areas of Data mining are medical tests, prediction of surgical processes, and establishment of relationships between pathological with clinical data [21]. Apriori and FPGrowth are the most widely adopted techniques used by different researchers to find frequent used mining algorithms [1, 32]. Different algorithms has been studied and implemented in [1, 2, 4, 5, 6, 7 and 8]. Brameier and Banzhaf have discussed two different programming models like Neural networks and linier genetic programming for medical data mining [9]. Unsupervised neural network on medical data is well explained by Shalvi [1]. Adepele has introduced approach for drawing different association rules through medical image data [11]. Further Shim and Xu identified a novel classification method based on Bayesian model [13]. He implemented this model to classify liver disease further. M. Khan et al. used decision tree algorithm to perform analysis on medical images [15]. For cardio vascular disease most widely used classification algorithms has been defined by Cheng et al.[3].

A prediction method towards further improvements in classification techniques for analysis of medical data mining has been proposed by Podgorelec et al. [21]. Wang has implemented fuzzy cluster analysis for medical images [22]. Bethel et al. developed a learner mechanism regarding association rules based on parameters collected from past breast cancer patients [12]. Further in the similar direction Xue has applied Bayesian Network algorithm towards diagnosis of Coronary Heart Diseases [25]. Abraham et al. [26] discussed the techniques to increase the accuracy of classification of medical data by using Naive Bayesian classifier algorithm.

Syed Zahid Hassan has proposed an in integrated approach to predict diseases by combining K-means algorithm, Self Organizing Map and Naïve Bayes and Neural Network based classifiers.[27]. Berlingerio et al. studied Time Annotated Sequences (TAS) algorithm to mine data with temporal dimensions [28]. The generated patterns are helpful in accurate diagnosis. Abdullah applied apriori algorithm for medical data mining [10]. A decision tree can be generated by CART algorithm has implemented by Mohammad Saraee[18]. Karegowda proposed a model to classify diabetic database to further enhance classification accuracy [2], known as CFS, and GA. CHAO and WONG has developed a decision tree learning approach to interpret attributes in medical data classification with enhanced accuracy by comparing ITI algorithm. Balakrishnan proposed feature selection using SVM to classifying diabetes databases [31]. Role of Fuzzy Cognitive Maps to detect drugs and their health impacts has been explained by Froelich [29].

TANG and TSENG have studied three different classifiers for medical data mining to classify diabetic and cancer datasets [24]. Tu and shin has developed an intelligent system to predict diagnosis of heart diseases by the use of decision tree, bagging, and Naïve Bayes [20]. Significance of data mining algorithms like C4.5, Bayasian Network and Back Propogation was purposed by Further Su[14]. To predict Kidney disease Vijayarani explained the SVM and Naïve Bayes algorithms[42]. Naïve Bayes, Decision tree, SVM, J48, Multilayer Perceptron classification algorithm and Conjunctive Rules has explained by Lambodar and Narendra[44]. Similarly to predict Kidney functions some machine learning algorithms like random forest, kstar, naïve bayes,, J48 and AD tree has been analysed by P.Swathi and Panduranga Vital[46].Ramya has analyzed Back propagation Neural network, Random Forest, Radial Basis Function, in order to predict kidney function by using chronic kidney disease dataset [43].

Parul Sinha and Poonam Sinha [45] developed a decision support system to predict chronic kidney disease by using Support Vector machine and KNN. The summarized version related to the application of different data mining techniques in different disease identification or used as a data sets, is as under:

Table 1:

Technique	**Mining Problem Domain**	**References**
Association Rule	Patterns Identification	[11], [12]
J48, C4.5, C5, and CART	Decision Support for Heart Disease, Hypothyroid, Dengue	[14], [15], [16], [3], [18], [19], [1] ,[33], [34], [38], [39], [40]

K-means, SVM and Naïve Bayes	Classification for Dengue disease datasets	[27], [35], [37]
k-NN	Classification for Diabetes, Cancer datasets	[24]
Appriori	Association rule mining	[10], [23], [17], [4], [5], [6], [7], [8], [32]
Bayesian Ying Yang	Classification on Liver Disease datasets	[13]
Neural Networks	Patterns and Trends Extraction	[9], [1]
Outlier Prediction Technique	Classification	[21]
Fuzzy Cluster Analysis	Medical images	[22]
Classification Algorithm	Disease Classification for Cardio Vascular Diseases datasets	[3], [2], [24]
Bayesian Network Algorithm	Analysis of medical data for Coronary Heart Disease	[14], [25]
Naive Bayesian	Classification for Coronary Heart Disease datasets	[26], [36]
Genetic Algorithm	Classification on Diabetic Datasets	[9]
Time Series Technique	Disease diagnosis	[28], [29]
Clustering and Classification	Clustering and classification of biomedical databases	[30]
SVM	Classification for Diabetes datasets	[16], [31]
Fuzzy	Drugs and Health effects classification	[29]
SVM, ANN and ID3	Classification	[16]
Naïve Base, SVM	Kidney Disease	[42]
BPN, RBF, RF	Chronic Kidney Disease	[43]
Naïve Base, MLP, SVM, J48 Conjunction Rule, Decision Table	Chronic Kidney Disease	[44]
SVM, KNN	Chronic Kidney Disease	[45]
AD Trees, J48 KStar, Naïve Bays, Random Forest	Kidney Disease	[46]

V. GAP ANALYSIS

Authors have performed a regress study on available literature and found following gaps that needs to be resolved and identified as a key area of further research in this field,

- The available literature shows massive use of decision tree and Naive Bayes classification techniques in most of the models.
- Data Sets are primarily taken from the web repositories like UCI, NCBI etc.
- Very few researchers have contributed their own data collected from hospitals and performed mining on that data.
- Open source tools are being the first and recommended choice by most of the researchers.
- The most mined dataset is heart, taken from UCI repositories.
- Recent viral disease predictions were also been made like H1N1 and dengue etc.
- Very few researchers have been made their own contribution towards creating new data repositories collected from trusted resources.
- Performance efficiency of prediction model can also be enhanced by combining more than one algorithms/ techniques together.

VI. TOOLS FOR DATA MINING

Various tools are available in field of data mining to help the statistical and analytical studies some of them are as follows:

A. SPSS Clementine

SPSS Clementine is a tool for text analytical and data mining given by IBM. It is a powerful tool for versatile data analysis and to build a predictive model.

B. SAS

SAS is a statistical analysis system it is software, developed by SAS institute. It is used for advanced analytical problems. SAS is used for both predictive and descriptive modeling for huge volume of data.

C. MATLAB

MATLAB is most widely used and popular among different discipline, used by millions of users across industry and business purposes. Mathworks developed this programming language. MATLAB is used in computational mathematics.

D. Weka

Weka is an open source. Weka is most widely used machine learning programming tool which was developed by University of Waikato New Zealand for

Knowledge Analysis. it supports extensive number of data mining algorithms like classification, clustering, regression techniques to assumption on base of fixed number of attributes.

E. R

R is also called GUN S. R is a programming languages as well as free software which can run on various platforms. R is widely used to statistically explore data set to perform graphical display, analysis, classification and data manipulation.

F. Orange

Orange is component base machine learning tool for data mining used in analysis of data and visualization. Orange performs data analysis by visual programming. Orange has property to suggest most frequently used combination as it has power of remembrances of choice.

G. Tanagra

Tanagra free data mining machine learning software and quite user friendly for researchers and students to implement data mining methods by comparing the patterns of knowledge. It is used in real time studies.

VII. MAJOR DATASETS REPOSITORIES

UCI Machine leering public repository is most widely used repository for various domain it contain 351 datasets for learning and research purposes. It is collection of bench mark datasets for comparative assessment and to perform classification and regression techniques.

VIII. CONCLUDING REMARKS AND FUTURE SCOPE

In this paper authors have established the applicability of data mining techniques in the field of research in medical domain. Good amount of literature is available which shows major scope of data mining techniques in disease predication and identification. However data security, privacy, misuse, and accuracy algorithms are major issues, which still need to be addressed. The efficiencies of different data mining algorithms can be accessed on the basis most prevalent features available in dataset. Combining different algorithms to get more accurate prediction for weakly classified data is also the major requirement for this field yet to be explored. The future research work will include disease prediction and identification in a more accurate fashion by comparing and then combining to efficient algorithms on weakly or heterogeneous datasets by using different machine learning algorithms.

IX. REFERENCES

[1] Doron Shalvi and Nicholas DeClaris, “An Unsupervised Neural Network Approach to Medical Data Mining Techniques”, In IEEE proc of International Joint Conference on Neural Networks, pp. 171-176, Vol.1, ISBN: 0-7803-4859-1, 1998.

[2] Asha Gowda Karegowda M.A.Jayaram, “Cascading GA & CFS for Feature Subset selection in Medical Data Mining”, In proc of IEEE International Advance Computing Conference, pp.1428 – 1431, ISSN: 978-1-4244-2927-1, 2009

[3] Tsang-Hsiang Cheng, Chih-Ping Wei, Vincent S. Tseng, “Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches”, In IEEE proc of 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06), pp. 165 – 170, ISSN: 1063-7125, 2006

[4] Carson Kai-Sang Leung; Christopher L. Carmichael; Boyu Hao, “Efficient Mining of Frequent Patterns from Uncertain Data”, In proc of Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007), pp. 489-494, 2007, ISBN: 0-7695-3033-8.

[5] Shariq Bashir, Zahid Halim, A. Rauf Baig, “Mining Fault Tolerant Frequent Patterns using Pattern Growth Approach”, In Proc. of IEEE/ACS International Conference on Computer Systems and Applications, pp. 172-179, ISBN: 978-1-4244-1967-8, 2008.

[6] Sunil Joshi and R. C. Jain., “A Dynamic Approach for Frequent Pattern Mining Using Transposition of Database”, In proc of Second International Conference on Communication Software and Networks IEEE., p498-501. ISBN: 978-1-4244-5727-4, 2010.

[7] Thanh-Trung Nguyen, “An Improved Algorithm for Frequent Patterns Mining Problem”, In IEEE proc of International Symposium on Computer, Communication, Control and Automation (3CA), p503-507, 2010, ISBN: 978-1-4244-5568-3

[8] Xiaoyong Lin and Qunxiong Zhu, “Share-Inherit: A novel approach for mining frequent patterns”, In IEEE proc of 8th World Congress on Intelligent Control and Automation, pp. 2712-2717, 2010, ISBN: 978-1-4244-6712-9

[9] Markus Brameier and Wolfgang Banzhaf, “A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining”, IEEE Transactions on Evolutionary Computation IEEE.p1-10, 2001, pp. 17-26, Volume: 5, Issue: 1, 1089-778X

[10] Hai-Tao He and Shi-Ling Zhang., “A New method for Incremental Updating Frequent patterns mining”, In proc. of IEEE international confernce on Innovative Computing, Information and Control, 2007, pp. 561 - 561 ISBN 0-7695-2882-1

[11] Adepele Olukunle and Sylvanus Ehikioya, “A Fast Algorithm for Mining Association Rules in Medical Image Data”, In IEEE proc. of Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No.02CH37373) IEEE, pp. 1181 - 1187 vol.2, 2002, ISSN: 0840-7789.

[12] Cindy L. Bethel et al., “Mining for Implications in Medical Data”, In IEEE proc of 18th International Conference on Pattern Recognition (ICPR'06), pp. 1212 – 1215, vol. 1, 2006, ISBN: 0-7695-2521-0

[13] Jeong-Yon Shim, Lei Xu., “Medical Data Mining Model for Oriental Medicine via BYY Binary Independent Factor Analysis, In IEEE proc of Circuits and Systems, 2003. ISCAS '03, pp: 717-720, vol.5, ISBN: 0-7803-7761-3.

[14] Jenn-Lung Su, Guo-Zhen Wu, I-Pin Chao, “The Appraoch of Data Mining Methods for Medical Database”, In IEEE Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3824 - 3826 vol.4, 2001, ISSN: 1094-687X

[15] Islam Md., Rafiqul Chowdhury, Morshed Khan, Safwan, “Medical Image Classification Using an Efficient Data Mining Technique”, In proc of 7th Asia-Pacific Complex Systems Conference, pp. 34-42, 2004

[16] Yanwei Xing, Jie Wang and Zhihong Zhao, “Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease” In IEEE proc of International Conference on Convergence Information Technology (ICCIT 2007), pp. 868 – 872, ISBN: 0-7695-3038-9, 2007

[17] U. Abdullah et al., “Analysis of Effectiveness of Apriori Algorithm in Medical Billing Data Mining”, In IEEE proc of Emerging Technologies, 2008. ICET 2008, pp. 327 – 331, 2008, ISBN: 978-1-4244-2210-4

[18] Mohammad Saraee, George Koundourakis, Babis Theodoulidis, “Easyminer: Data Mining in Medical Databases”, In IEE

Colloquium on Intelligent Methods in Healthcare and Medical Applications, pp. 7/1 - 7/3, 1998

[19] Sam Chao, Fai Wong, "An Incremental Decision Tree Learninng Methodology Regarding Attributes in Medical Data Mining", In IEEE proceedings of the Eighth International Conference on Machine Learning and Cyb ernetics, Baoding, 12-15 July 2009, pp. 12-15, vol. 3, ISBN: 978-1-4244-3702-3

[20] My Chau Tu AND Dongil Shin, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms" In IEEE proc of Eighth International Conference on Dependable, Autonomic and Secure Computing, pp. 183-187, ISBN: 978-1-4244-5421-1, 2009

[21] Vili Podgorelec, Marjan Heriko, I. Rozman, "Improving Mining of Medical Data by Outliers Prediction" In IEEE Symposium on Computer-Based Medical Systems (CBMS'05), pp.91-96, ISBN: 0-7695-2355-2, 2005

[22] Shuyan Wang Mingquan Zhou Guohua Geng, "Application of Fuzzy Cluster Analysis for Medical Image Data Mining", In IEEE proc of IEEE International Conference Mechatronics and Automation, pp. 631 - 636 Vol. 2, ISBN: 0-7803-9044-X, 2005

[23] Cong-Rui Ji and Zhi-Hong Deng, "Mining Frequent Ordered Patterns without Candidate Generation", In proc. of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery - Volume 01 (FSKD '07), Vol. 1. IEEE Computer Society, Washington, DC, USA, pp. 402-406, 2007, ISBN:0-7695-2874-0

[24] Ping-Hung Tang; Ming-Hseng Tseng, "Medical Data Mining Using BGA and RGA for Weighting of Features in Fuzzy k-NN Classification", In proc of IEEE International Conference on Machine Learning and Cybernetics, pp. 3070 – 3075, vol. 5, ISBN: 978-1-4244-3702-3, 2009

[25] Weimin Xue, Yanan Sun, Yuchang Lu, "Research and Application of Data Mining in Traditional Chinese Medical Clinic Diagnosis", In proc of IEEE 8th international Conference on Signal Processing, Vol. 4, ISBN: 0-7803-9736-3, 2006

[26] Ranjit Abraham, Jay B.Simha, Iyengar, "A comparative analysis of discretization methods for Medical Datamining with Naïve Bayesian classifier", In proc of IEEE international conference on information Technology, pp. 235 - 236, 2006, ISBN: 0-7695-2635-7

[27] Syed Zahid Hassan and Brijesh Verma, "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases", In proc of IEEE Seventh International Conference on Intelligent Systems Design and Applications, pp. 503 – 510, ISBN: 0-7695-2976-3, 2007

[28] Michele Berlingeri, "Mining Clinical Data with a Temporal Dimension: a Case Study", In proc. of IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), pp. 429 – 436, ISBN: 0-7695-3031-1, 2007

[29] Wojciech Froelich, Alicja Wakulicz-Deja, "Mining Temporal Medical Data Using Adaptive Fuzzy Cognitive Maps", In proc of 2nd Conference on Human System Interactions, pp. 16-23, ISBN: 978-1-4244-3959-1, 2009

[30] Michael Barnathan, Jingjing Zhang, Vasileios, "A Wen-Accessible Framework for the Automated Storage and Texture Analysis of Biomedical Images", In proc. of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 257 – 259, ISBN: 978-1-4244-2002-5, 2007

[31] Sarojini Balakrishnan, "SVM Ranking with Backward Search for Feature Selection in Type II Diabetes Databases" In IEEE proc of International Conference on Systems, Man and Cybernetics, pp. 2628 – 2633, ISSN: 1062-922X

[32] Gitanjali J. et al., "APRIORI algorithm based medical data mining for frequent disease identification", IIJIT, Vol. 2, Issue 4 pp. 1-5, 2014, ISSN 2321-5976

[33] Mai Shouman et al., "Using Decision Tree for Diagnosing Heart Disease Patient", In Proc. of 9th Australasian Data Mining Conference, Ballarat, Austrailia, pp. 23-29, 2011

[34] Ganganjot Kaur, Amit Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes", IJCA, Vol. 98, No. 22, pp. 13-17, 2014, ISSN: 0975 - 8887

[35] P. Ramachandran et al., "Early Detection and Prevention of Cancer using Data Mining Techniques", IJCA, Vol. 97, No. 13, pp. 48-53, 2014, ISSN: 0975 - 8887

[36] G. Subbalakshami et al., "Decision Support in Heart Disease System using Naïve Bayes", IJCSE, Vol. 2 No. 2, pp. 170-176, 2011, ISSN : 0976-5166

[37] N. Subitha et al., "Diagnosis for Dengue Fever using Spatial Data Mining", IJCTT, Vol. 4, Issue 8, pp. 2646-2651, 2013, ISSN: 2231-2803

[38] Abhishek Taneja, "Heart Disease Prediction System using Data Mining Techniques", OJCST, Vol. 6, No. 4, pp. 457-466, ISSN: 0974-6471 ISSN: 0974-6471 ISSN: 0974-6471 ISSN: 0974-6471 ,2013

[39] Shivanee Pandey et al., "Diagnosis and Classification of Hypothyroid Disease using Data Mining Techniques", Vol. 2 Issue 6, pp. 3188-3192, ISSN: 2278-0181 ISSN: 2278-0181, 2013

[40] Daranee T. et al., "Data Mining of Dengue Infection using Decision Tree", In proc. of ACCIDS, Singapore, pp. 154-159, 2012

[41] Usama Fayyad et al. "From Data Mining to Knowledge Discovery in Databases, AI Magazine Volume 17 Number 3 pp. 37-54, 1996

[42] D. S. Vijayarani, Mr. S. Dhayanand, "Data Mining Classification Algorithms for Kidney Disease Prediction", International journal of Cybernetics and informatics (IJCI), pp 13-25 Vol. 4, No. 4, 2015.

[43] S. Ramya , Dr. N. Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms", International Journal of Innovative Research in Computer and Communication Engineering, pp. 812-820 Vol 4, issue 1, ISSN: 2320-9798, 2016.

[44] Lambodar Jeena, Narendra Ku. Kamila, "Distributed Data Mining Classification Algorithms for Prediction of Chronic Kidney Disease", International Journal of Engineering Research in management and Technology, pp. 110-118, Vol. 4, issue 11, ISSN : 2278-9359

[45] Paul Sinha, Poonam Sinha, "Comparative Study of Chronic Kidney Disease Prediction Using KNN and SVM", International Journal of Engineering Research and Technology, pp. 608-612, vol. 4 issue 12, ISSN : 2278-0181.

[46] P. Swathi baby, T. Panduranga Vital, "Statistical Analysis and Predicting Kidney Disease Using Machine Learning Algorithms", International Journal of Engineering Research and Technology, pp. 206-210, Vol. 4, Issue 07, ISSN : 2278-0181, 2015

