International Journal of
Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: http://dx.doi.org/10.7324/IJASRE.2017.32486

Vol.3 (8) Sep - 2017

Data Mining for diagnosis in Healthcare Sector-A Review

Umar Sidiq[1] and Dr.Rafi Ahmad Khan[2]

Research Scholar[1]Mewar University, Rajastan, India.

Assistant professor[2] Kashmir University

Jammu & Kashmir, India.

ABSTRACT

Data mining is a powerful tool suitable for uncovering new trends in healthcare organization. Data mining plays a major role in predicting and diagnosing various ailments in various health care sectors. It uses a variety of techniques to find hidden patterns and these patterns can be used by physicians in evaluating the effectiveness of medical treatments.The current study explores the utility of various Data Mining techniques such as ANN, Decision Tree, K-NN, SVM, Bayesian Methods in healthcare domain and also highlights various applications in healthcare. The main focus of this study will be to carry out the survey of existing data mining techniques used in healthcare sectors.

Keywords:*Data mining, ANN, Decision Tree, K-NN, SVM, Bayesian Method.*

1. INTRODUCTION

Data mining is a powerful tool suitable for fetching previously unknown patterns and useful information from voluminous dataset relating to human ailments or diseases. Now a day's Information technologies are being increasingly implemented in healthcare organizations in order to respond to the needs of doctors in their daily decision making activities. There are various data mining techniques found in the medical related areas such as Medical device industry, Pharmaceutical Industry and Hospital Management. To find the useful and hidden knowledge from the database is the purpose behind the data mining applications. Data mining techniques applied in healthcare industry play a major role in prediction and diagnosis of the diseases. Data mining is a fast growing technology, is being generally used in biomedical sciences and research. Modern medicine generates vast information stored in the medical database. For example, medical data may contain ECG, MRI, blood pressure, blood sugar, cholesterol levels, etc., and also physician's analysis. Extracting useful knowledge and providing scientific decision-making for the treatment, diagnosis and prediction of disease from the database increasingly becomes necessary and are better positioned to meet their long-term needs [1] [2].Han et al. defined the data mining as the process of discover interesting knowledge from large complex data stored either in data warehouses, databases or other information repositories [3]. Witten et al. defined data mining as the process of extracting implicit, previously unknown and potentially useful information from data [4]. Hand et al. defined data mining as the analysis of observational data set to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [5]. Healthcare industry today generates large complex data about patients, hospitals resources, disease diagnosis and prediction, electronic patient records, medical devices etc. This large complex data is a key resource to be processed and analyzed for knowledge extraction that enables support for decision making and cost-savings. Data mining brings a set of tools and techniques that can be applied to the processed data to discover unknown patterns that provide healthcare professionals an additional source of knowledge for making decisions .As with the use of data mining tools the huge complex or voluminous healthcare data are being collected and made available for a variety of areas, such as doctors who use patterns by measuring clinical indicators, quality indicators, customer satisfaction and economic indicators, performance of physicians from multiple

perspectives to optimize use of resources, cost efficiency and decision making based on evidence, identifying high-risk patients and intervene proactively, optimize healthcare, etc [6].

The below sections of this paper are planned as follows. Section-2 provides techniques of data mining used in healthcare sector.Section-3 reviews literature pertaining to data mining and applications of data mining techniques in healthcare. Section-4 shows various applications of data mining used in healthcare sector. Section-5 presents the conclusion of the paper. And in final section-6 references are mentioned.

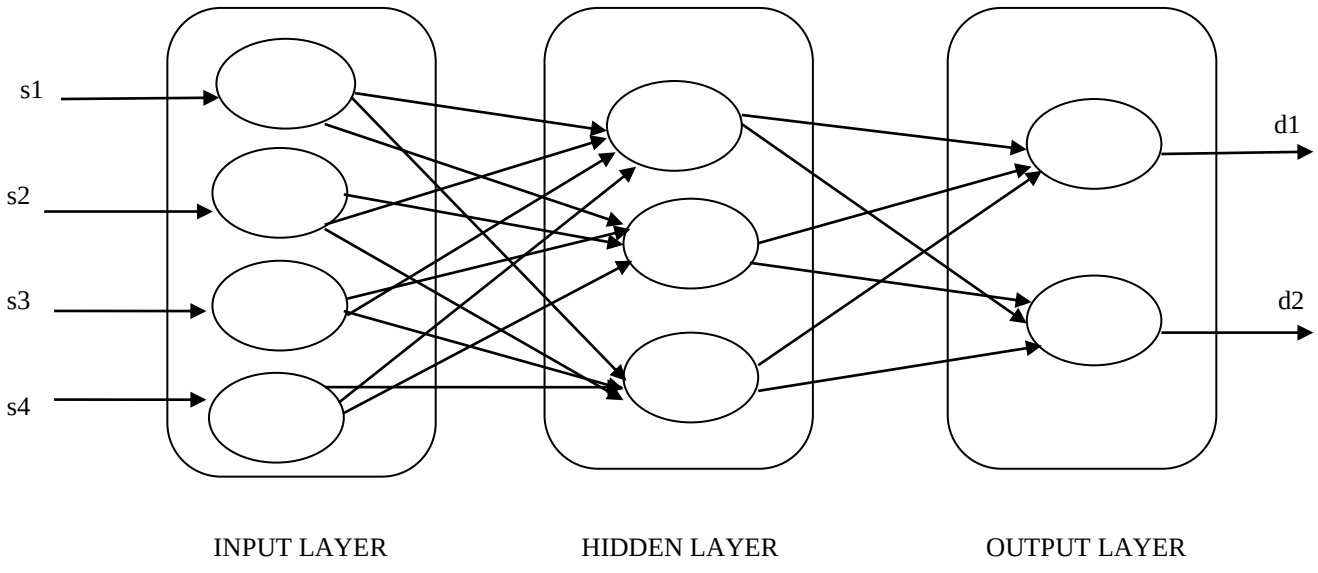
2. TECHNIQUES OF DATA MINING USED IN HEALTH CARE SECTOR

Classification is one of the most commonly used methods of Data Mining in Healthcare organization/sectors. Different data mining classification techniques have been used to help healthcare professionals for prediction, diagnosis, detection of various diseases such as thyroid, heart, diabetes, cancer diseases etc and also in Treatment effectiveness, Management of healthcare, Detection of fraud and abuse, Customer relationship management etc. The most common classification data mining techniques used in healthcare are: neural network, decision tree, nearest neighbor algorithms, support vector machine, Bayesian Methods**.**

2.1 Artificial neural networks

Neural networks (NN) are those systems modeled based on the human brain working. In Neural networks, basic elements are neurons or nodes. These neurons are interconnected and within the network they worked together in parallel in order to produce the output functions. Neural network is the most widely used classification algorithm in various biomedicine and healthcare fields [7] [8] [9]. For example, NN has been widely used as the algorithm supporting the diagnosis of diseases including cancers [10] [11] [12] [13] [14]and predict outcomes [15] [16] [17].

Figure1. NEURAL NETWORK DIAGRAM

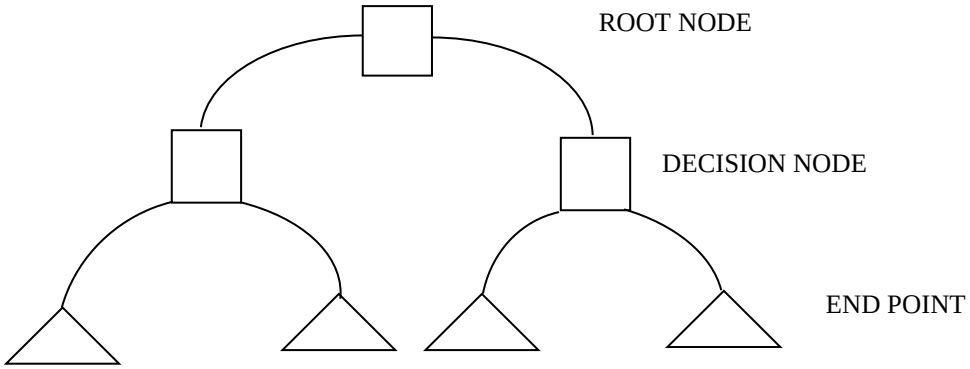


2.2 Decision tree

Decision Tree is usually used by various researchers in healthcare field. Decision tree is among the classification technique that solves large complex problems by providing rules in an understandable form [18]. It is a knowledge representation structure consisting of nodes and branches organized in a tree shaped such that, every internal non-leaf node is labeled with values of the attributes and the branches coming out from an internal node are labeled with values of the attributes in that node. Every node is

labeled with a class (a value of the goal attribute). Tree based models which include classification and regression trees, are the common implementation of induction modeling [19]. Decision tree models are best suited for data mining as they are easy to interpret, inexpensive to construct, easy to integrate with database system and they have comparable or better accuracy in many applications. By Using Decision Tree, decision makers can choose best traversal and alternative from root to leaf indicates unique class separation based on maximum information gain [20].

Figure 2.DECISION TREE DIAGRAM



2.3 K-Nearest Neighbor (K-NN)

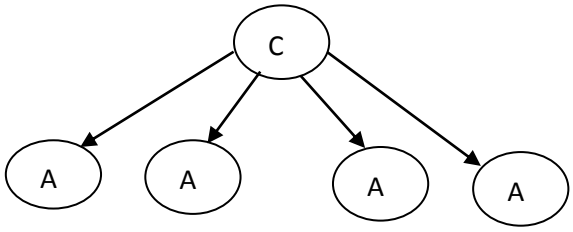
K-Nearest Neighbor (K-NN) is widely used data mining technique in healthcare domain and have a simplest classifier known as K-NN classifier that discovers the unknown data point using the previously known data points (nearest neighbor) and classified data points according to the voting system [21]. The objects are classified by the majority votes of its neighbors. Sometimes it is also called Memory-based classification because at run time training data must always be stored in memory. Euclidean distance is calculated when the differences between the attributes are taken in case of continuous attributes. But it has a drawback, when large values bear down the smaller ones. Continuous attributes must be normalized in order to take over this major problem so that they have same influence on the distance measure between distances [22].

2.4 Support Vector Machine (SVM) Vapnik *et al.* [23] [24]was the first one who gave the notion of SVM. Among all the available algorithms it provides very accurate results. It is gaining popularity because it can be easily extended to problems related to multiclass though it was mainly developed for problems related to binary classification [25]. In order to be useful for various effective and efficient tasks it is capable of creating single as well as multiple hyper planes in high dimensional space. The main aim of creating hyper plane by SVM in order to separate the data points. There are two ways of implementing SVM. The first technique employs mathematical programming and second technique involves kernel functions. With the help of training datasets non linear functions can be easily mapped to high dimensional space. Such situation can only be possible when we use kernel functions. Gaussian, polynomial, sigmoid etc. are some examples of kernel functions. For the classification of data points hyper plane is used. The primary task of hyper plane is to maximize the separation between data points. Support vectors are used in order to construct the hyper plane.

2.5 Bayesian Method

For probabilistic learning method Bayesian classification is used. With the help of classification algorithm we can easily obtained it [26]. Bayes theorem of statistics plays huge role in it, but in medical domain, attributes such as patient symptoms and their health conditions are correlated with each other but Naive Bayes Classifier assumes that all attributes are independent with each other. This is the key drawback with Naive Bayes Classifier. If attributes are independent with each other then Naïve Bayesian classifier has shown high accuracy. In healthcare field they play very important roles. Hence, researchers across the world used

them there are main two advantages of BBN. First One is it helps to makes computation process very easy. Second one is that for huge datasets it has better speed and accuracy. Bayesian Belief Network is widely used by many researchers in healthcare domain. Liu et al. develop a decision support system using BBN for analyzing risks that are associated with health [27]. Curiac *et al.*, used BBN in making significant decision regarding patient health suffering from psychiatric disease to analyze the psychiatric patient data and performed experiment on real data obtain from Lugoj Municipal Hospital [28].The structural model is represented as a directed graph where the nodes represent attributes and arcs represent attribute dependency. A Representation of a Bayesian Classifier Structure [29].



3. LITERATURE REVIEW

Breast cancer is one of the deadly and dangerous diseases in women. Potter *et al*., had performed experiment on the breast cancer data set simulated in WEKA tool and then analyzed the performance of different classifier using 10-fold cross validation method [30].

Huang *et al.*, constructed a hybrid SVM-based model in diagnosis of breast cancer, because in Taiwan women especially young women suffered from breast cancer. In this research several types of DNA viruses are studied. These DNA viruses are EBV (Epstein-Barr virus), CMV (cytomegalovirus), HSV-1 (herpes simplex virus type-1), HPV (human papilloma virus), and HHV-8 (human hepesvirus-8). On the basis of experimental results, either {HSV-1,HHV-8} or {HSV-1, HHV-8,CMV}can achieved the identical high accuracy. The main aim of the study was to obtain the bioinformatics about the breast cancer and DNA viruses. Apart from SVM-based model, also a diagnosis model called LDA (Linear discriminate analysis) was constructed in this research. After comparing the accuracies of both SVM and LDA, the SVM outperformed over LDA [31].

Chang *et al.*, used an integrated decision tree model for characterize the skin diseases in adults and children. The main focus of this research was to analyze the results of five experiments on the six main skin diseases. The main aim of this research was to construct the best predictive model in dermatology by combining decision tree and neural network classification methods. On the basis of experimental result, it has been found that neural network has achieved 92.62% accuracy in prediction of skin diseases [32].

Das *et al.*, proposed a intelligent medical decision support system based on SAS software for the diagnosis of heart diseases. In order to construct the proposed system, neural networks method was mainly used. In this research, the data was taken from Cleveland heart disease database. On the basis of experiments, it has been found that neural networks have achieved 89.01% accuracy [33].

Gunasundari *et al.*, used ANN to identify lung diseases. This research work examine the chest Computed Tomography (CT) and extract significant lung tissue feature to reduce the data size from the Chest CT and then extracted textual attributes were given to neural network as input to identify the various lung diseases [34].

Soni *et al.*, proposed the associative classification approach for better analyzing the healthcare data. The proposed approach was the combined approach that integrated the association rules as well as classification rules. This integrated approach was useful for discovering rules in the database and then using these rules to construct an efficient classifier. In this research, experiments on the data of heart patients were performed in order to find out that accuracy of associative classifiers was better than accuracy of traditional classifiers. Apart from this, the research also generated the rules using weighted associative classifier [35].

K. Srinivas et al., discusses mainly examine the potential use of data mining classification techniques such as Rule Based, Naïve Bayes, Decision tree and Artificial Neural Network to the voluminous healthcare data. Using sex, age, blood sugar and blood pressure medical profiles it can predict the possibility of patients getting a heart disease [36].

Bakar *et al.*, proposed predictive models by using multiple rule based classifiers for the better early detection of dengue disease. In this research, the multiple rule based classifiers which were used in the proposed models were decision tree, rough set classifier, naïve bayes, and associative classifier. In order to predict the early detection of dengue disease the several classifiers were investigated in this research. The classifiers were investigated individually as well as combinatory in order to study their performance. On the basis of experimental results, it has been found that the accuracy of multiple classifiers showed better was better accuracy than single classifier [37].

Jena *et al.*, used K-NN and Linear Discriminate Analysis (LDA) for classification of chronic disease in order to generate early warning system. This research work used K-NN to analyze the relationship between cardiovascular disease and hypertension and the risk factors of various chronic diseases in order to construct an early warning system to reduce the complication occurrence of these diseases [38].

Arvind Sharma et al. discussed Data mining can have important benefits to the blood bank sector. J48 algorithm have been implemented in WEKA tool to perform this research work. Classification rules performed well in the classification of blood donors and achieved accuracy rate 89.9% [39].

Shweta Kharya discussed various data mining approaches that have been utilized for breast cancer diagnosis and prognosis. Decision tree is found the best predictor with 93.62% Accuracy on benchmark dataset as well as on SEER data set [40].

Abdi *et al.*, constructed a PSO based SVM model for discovering erythemato-squamous diseases which consists two stages. In the first stage optimal feature were extracted using association rule and in second phase the PSO was used to discovered best kernel parameters for SVM in order to improve the accuracy of classifier model [41].

Zuoa *et al.*, introduced an adaptive Fuzzy K-NN approach for detection of Parkinson disease [42].

Rusdah *et al.*, reviewed the various data mining methods for tuberculosis diagnosis. In this work, it has been found that the support vector machines (SVM) outperformed over other methods for diagnosis of tuberculosis [43].

Peng Z *et al.*, This study aimed to identify biomarkers for estimating the overall and survival of prostate cancer (PCA) patients at the time of their diagnosis. In the research, a total of 641 ESC gene predictors (ESCGPs) were identified by using microarray data sets. For estimating the survival a k-nearest neighbor (K-NN) algorithm was used to estimate the overall survival [44].

4. Applications of data mining in healthcare sector

Following are the several applications of Data Mining in healthcare sector. HianChyeKoh and Gerald Tan mainly discusses data mining and its applications with major areas like Treatment effectiveness, Management of healthcare, Detection of fraud and abuse, Customer relationship management [45].

a) Effectiveness of Treatment:

By Using Data Mining, physicians and patients can easily compare causes, symptoms, and courses of treatments. They can examine the effectiveness of available treatments and find out which technique is better and effective [46].

Some diagnostic and laboratory procedures are invasive, painful and costly to patients. An example of this is conducting a biopsy in women to detect cervical cancer. Thangavel et al. used K-means clustering algorithm to analyze cervical cancer patients and found that clustering outperformed over other existing medical opinion. They found a set of interesting attributes that could be used by doctors as additional support on whether or not to recommend a biopsy for a patient suspected of having the cervical cancer [47].

b) Healthcare management: Data mining applications can be developed to find out better and track chronic diseases states and high-risk patients, design suitable interventions, and reduce the different number of hospital admissions and claims to aid healthcare management.

c) Customer relationship management:

Data Mining helps the healthcare organization to understand the quality, behavior, preferences, patterns and needs of their customer in order to make better relation with them. Hallick has suggested that Data mining techniques provide the information to patient regarding various diseases and their prevention so we can say that CRM in healthcare can help to promote disease education, prevention, and wellness services [48].

d) Hospital Ranking: Different data mining methods are used to analyze the various hospital details in order to determine their ranks [49]. The Ranking of the hospitals are done on the basis of their capability to handle the high risk patients.

e) Improved Patient care: Large amount of data is collected with the advancement in electronic patient record. Patient data which it store in digitized form improve the healthcare system quality. Kolar has identified that healthcare organizations used data mining techniques for patient grouping [50].

f) Fraud and abuse detection: Healthcare insurer develops a model by using data mining techniques to detect the abuse and fraud in the medical claims. This model is helpful for identifying the inappropriate prescriptions, fake or irregular patterns in medical claims made by patients, physicians, hospitals etc.

Kou et al. [51] presents survey of knowledge discovery and neural network used in credit card fraud detection and insurance claims. By extension, these techniques could also be used to detect anomalous patterns in health insurance claims.

Johnson *et al.* proposed a multistage methodology in order to better detection of fraud committed by patients as well as by providers for healthcare insurance companies. The proposed methodology helped a lot in reducing significant costs for insurance companies. The proposed methodology was made up of various stages including risk determination stage. The risk threshold was determined by using a decision tree based method. The proposed methodology was compared with unsupervised and supervised neural network techniques. After comparison, it has been found that the proposed methodology outperforms over unsupervised and supervised neural network techniques in order to detect the fraud. Apart from this, proposed methodology plays a significant role in validated the legitimated claims by obtaining the information from insurance claim forms [52].

5. Conclusion

The objective of our work is to provide a study of different data mining techniques that can be employed in healthcare sectors. Various data mining classification techniques are defined in this work which has emerged in recent years for efficient and effective in healthcare domain. The various applications of data mining in healthcare have also been discussed which can help the healthcare researchers to do better decision making.

References

[1] Sandhya Joshi, P. Deepa Shenoy, Venugopal K R, and L M Patnaik, "Data Analysis and Classification of Various Stages of Dementia Adopting Rough Sets Theory," *International Journal on Information processing*, vol. 4, no. 1, pp. 86-99, 2010.

[2] Benko A and Wilson B, "Online decision support gives plans an edge," *Managed Healthcare Executive*, vol. 13, no. 5, pp. 20-25.

[3] Han J. and Kamber M., *Data Mining: Concepts and Techniques*. San Francisco, USA: Morgan Kaufmann, 2001.

[4] Ian H. Witten and Eibe Frank, *Data mining: Practical machine learning tools and techniques with Java implementations*. CA,

San Francisco, USA: Morgan Kaufmann, 2000.

[5] Hand D. J., Mannila H., and Smyth P, *Principles of data mining*. Boston,MA, USA: MIT Press, 2001.

[6] Ishtake S. H and Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques," *International Journal of Healthcare & Biomedical Research*, vol. 1, no. 3, pp. 94-101, 2013.

[7] Bellazzi R. and Zupan B., "Predictive data mining in clinical medicine: current issues and guidelines," *International Journal Of Medical Informatics*, vol. 77, no. 2, pp. 81-97, 2008.

[8] Übeyli E. D., "Comparison of different classification algorithms in clinical- decision making," *Expert Systems*, vol. 24, no. 1, pp. 17-31, february 2007.

[9] Harleen Kaur and Siri K.Wasan, "Empirical study on applications of data mining techniques in healthcare," *Journal Of Computer Science*, vol. 2, no. 2, pp. 194-200, 2006.

[10] Romeo M., Burden F., Quinn M., and Wood B. and McNaughton D., "Infrared microspectroscopy and artificial neural networks in the diagnosis of cervical cancer," *Cellular and Molecular Biology(Noisy-le-Grand France)*, vol. 44, no. 1, pp. 179-187, 1998.

[11] Ball G. et al., "An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers," *Bioinformatics*, vol. 18, no. 3, pp. 395–404, 2002.

[12] Sergey Aleynikov and Evangelia Micheli-Tzanakou, "Classification of retinal damage by a neural network based system," *Journal of Medical Systems*, vol. 22, no. 3, pp. 129-136, 1998.

[13] Potter R., "Comparison of classification algorithms applied to breast cancer diagnosis and prognosis advances in data mining," in *7th Industrial Conference ICDM 2007*, Leipzig Germany, july 2007, pp. 40–49.

[14] Igor Kononenko, Ivan Bratko, and Matjaz Kukar, "Application of machine learning to medical diagnosis. Machine Learning and Data Mining: Methods and Applications," *John Wiley and Sons Ltd*, pp. 389-408, 1997.

[15] Sharma A. and Roy R.J., "Design of a recognition system to predict movement during anesthesia," *IEEE Trans. Biomedical Engineering* , vol. 44, no. 6, pp. 505–511, 1997.

[16] Einstein A. J., Wu H. S., Sanchez M., and Gil J., "Fractal characterization of chromatin appearance for diagnosis in breast cytology," *Journal Of Pathology*, vol. 185, no. 4, pp. 366-381, August 1998.

[17] Brickley M., Shepherd J. P., and Armstrong R. A., "Neural networks: a new technique for development of decision support systems in dentistry," *Journal Of Dentistry*, vol. 26, no. 4, pp. 305-309, May 1998.

[18] López-Vallverdú JA, Riano D, and Bohada J, *Improving medical decision trees by combining relevant health-care criteria.*: Expert Systems with Applications, 2012, vol. 39.

[19] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kauffmann Publishers, 2001.

[20] Apte and S.M. Weiss, "Data Mining with Decision Trees and Decision Rules," *T.J. Watson Research Center,http://www.research.ibm.com/dar/papers/pdf/fgcsaptewe issue_with_cover.pdf*, 1997.

[21] E ALPAYDIN, "Voting over Multiple Condensed Nearest Neighbors," *Artificial Intelligence Review*, 1997.

[22] Max Bramer and M.A.Bramer, *Principles of data mining*. Springer, 2007.

[23] V. Vapnik, *Statistical Learning Theory*. Wiley, 1998.

[24] V. Vapnik, *The support vector method of function estimation*., 1998.

[25] N. Chistianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines, and other kernel-based learning methods*.: Cambridge University Press, 2000.

[26] C. McGregor, C. Christina, and J. Andrew, "A process mining driven framework for clinical guideline improvement in critical care," in *Learning from Medical Data Streams 13th Conference on Artificial Intelligence in Medicine (LEMEDS). http://ceur-ws.org*, vol. 765, p. 1012.

[27] K. F. R. Liu and C. F. Lu, "BBN-Based Decision Support for Health Risk Analysis," in *Fifth International Joint Conference on INC, IMS and IDC*, 2009.

[28] D. I. Curiac, G. Vasile, O. Banias, C. Volosencu, and A. Albu, "Bayesian Network Model for Diagnosis of Psychiatric Diseases," in *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, Cavtat, Croatia*, 2009.

[29] L. Jiang and et al, "A novel bayes model: hidden naive bayes," *IEEE Trans. on Knowl. and Data Eng.*, vol. 21, pp. 1361-1371, 2009.

[30] R. Potter, "Comparison of classification algorithms applied to breast cancer diagnosis and prognosis," in *advances in data mining, 7th Industrial Conference, ICDM*, Leipzig, Germany, 2007, pp. 40-49.

[31] L. Huang, H. C. Liao, and M. C. Chen, "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," *Expert Systems with Applications*, vol. 34, pp. 578-587, 2008.

[32] L. Chang and C. H. Chen, "Applying decision tree and neural network to increase quality of dermatologic diagnosis," *Expert Systems with Applications, Elsevier*, vol. 36, pp. 4035-4041, 2009.

[33] R. Das, I. Turkoglub, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, pp. 7675-7680, 2009.

[34] S. Gunasundari and S. Baskar, "Application of Artificial Neural Network in identification of Lung Diseases," in *Nature & Biologically Inspired Computing, World Congress on IEEE*, 2009.

[35] Sunita Soni and O.P.Vyas, "Using Associative Classifiers for Predictive Analysis in Health Care Data Mining," *International Journal of Computer Applications*, vol. 4, no. 5, july 2010.

[36] K. Srinivas, B. Kavitha Rani, and Dr. A. Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," *International Journal on Computer Science and Engineering*, 2010.

[37] A. Bakar, Z. Kefli, S. Abdullah, and M. Sahani, "Predictive Models for Dengue Outbreak Using Multiple Rulebase Classifiers," in *International Conference on Electrical Engineering and Informatics*, Bandung, Indonesia, 2011, pp. 17-19.

[38] H. Jena, C. C. Wang, B. C. Jiangc, Y. H. Chub, and M. S. Chen, "Application of classification techniques on development an early-warning systemfor chronic illnesses," *Expert Systems with Applications*, vol. 39, pp. 8852-8858, 2012.

[39] Arvind Sharma and P.C. Gupta, "Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool," *International Journal of Communication and Computer Technologies*, vol. 01, no. 02, september 2012.

[40] Shweta Kharya, "Using Data Mining Techniques ForDiagnosis And Prognosis Of Cancer Disease," *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, vol. 2, April 2012.

[41] M. J. Abdi and D. Giveki, "Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 603-608, 2013.

[42] W. L. Zuoa, Z. Y. Wanga, T. Liua, and H. L. Chenc, "Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomedical Signal Processing and Control*, pp. 364-373, 2013.

[43] Rusdah and Edi Winarko, "Review on Data Mining Methods for Tuberculosis Diagnosis," in *Information Systems International Conference (ISICO)*, 2013.

[44] Peng Z et al, "An expression signature at diagnosis to estimate prostate cancer patients' overall survival Prostate Cancer and Prostatic Disease," january 2014.

[45] HianChyeKoh and Gerald Tan, "Data Mining Applications in Healthcare," *journal of Healthcare Information Management*, vol. 19, no. 2.

[46] k.Kincade, "Data mining: digging for healthcare gold," *Insurance & Technology*, vol. 23, no. 2, 1998.

[47] Kuttiannan Thangavel, P.Palanichamy Jaganathan, and P.O.Easmi, "Data Mining Approach to Cervical Cancer Patients Analysis Using Clustering Technique," *Asian Journal of Information Technology*, vol. 5, no. 4, pp. 413-417, 2006.

[48] Hallick JN, "Analytics and the data warehouse," *Health Management Technology*, vol. 22, no. 6, pp. 24-25, 2001.

[49] Mary K.Obenshian and MAT, "Application of Data Mining Techniques to Healthcare Data," *Infection Control and Hospital Epidemiology*, August 2004.

[50] H. R. Kolar, "Caring for healthcare," *Health Management Technology*, vol. 22, no. 4, pp. 46-47, 2001.

[51] Yufeng Kou, Chang-Tien, Sirirat Sirwongwattana, and Yo-Ping Huang, "Survey of fraud detection techniques," in *IEEE International Conference on Networking, Sensing and Control*, Taipei,Taiwan, 2004, pp. 749-754.

[52] Marina Evrim Johnson and Nagen Nagarur, "Multi-stage methodology to detect health insurance claim fraud," *Health Care Management Science*, vol. 19, no. 3, pp. 249-260, september 2016.