Data Mining Techniques and Tools Used in Healthcare Databases

Elma Çela (Kolçe) [1], **Neki Frasheri** [2]

[1]Department of Computer Engineering, Polytechnic University of Tirana, Bul. "Dëshmorët e Kombit", Sheshi "NënëTereza", Nr. 4, Albania, elmakolce@yahoo.com

[2] Department of Computer Engineering, Polytechnic University of Tirana, Bul. "Dëshmorët e Kombit", Sheshi "NënëTereza", Nr. 4, Albania, nfrasheri@fti.edu.al

Abstract

This study presents an overview of the current research being carried out on different data mining techniques and tools used in diagnosis, prognosis and treatment of various diseases. 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. Healthcare databases have a huge amount of data but however, there is a lack of effective analysis tools to discover the hidden knowledge. The implementation of an efficient and accurate automated system, that can help physicians in their work, needs a comparative study of various techniques and tools available. The main aim of this study is to identify the most important application fields and the trends of the research in the healthcare domain, highlighting critical issues and summarizing the approaches in a set of learned lessons. The methodology used for this paper was through the survey of most recent publications and journals in the fields of computer science, engineering and health care. Analyses show that most of the researches have been done in studying classic data mining algorithms such as Decision Trees, Support Vector Machine, Artificial Neural Networks, Bayesian Networks and Naïve Bayes, Logistic Regression, Genetic Algorithms, Fuzzy Rules, Association Rules, showing acceptable levels of accuracy. Good efforts have also been done in developing novel and adapted data mining techniques for enhanced diagnosis and prognosis accuracy. However, using data mining techniques to identify treatment options for patients has received less attention. Results also show that most efforts have been done in the field of Cardiovascular diseases (Heart attack, Hypertension, Coronary Artery Disease, etc.) and Cancer diseases. The follow-up of this study will aim at dealing with data mining techniques that have wider spectra of application for groups of diseases.

Keywords:*Data Mining, Diagnosis, Prognosis, Healthcare Database*

1. INTRODUCTION

Data mining is defined as “a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database” by Fayyad [1]. 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 to suggest less expensive therapeutically equivalent alternatives. Efficient and accurate implementation of an automated system needs a comparative study of various techniques available. In this paper we aim to identify the most important application fields and the trends of the research in the healthcare domain, presenting an overview of the current research being carried out using the data mining techniques for the diagnosis, prognosis and treatment 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 identify the most important application fields and the trends of the research in the healthcare domain in chapter three, highlighting the data mining techniques used, with their accuracy levels and finally we show the conclusions of our work.

2. METHODOLOGY

The methodology used for this paper was through the survey of journals and publications in the fields of computer science, engineering and health care. European Journal of Scientific Research, International Journal on Computer Science and Engineering, Expert Systems with Applications, Data Science Journal are some of these journals. In order to obtain a general overview on the literature, book chapters, dissertations, working papers and conference papers are also included. The research is focused on most recent publications, with 2008 as the cut off year.

3. LITERATURE REVIEW

There are different kinds of studies for data mining techniques in healthcare databases, where we can identify application fields such as Diagnosis, Prognosis and Treatment of several diseases. There are also studies using data mining techniques to predict results of surgeries. By an overview of the literature, we can classify the used data mining techniques in: “Classical” and “New” techniques. We identify the following “Classical” algorithms: Decision Trees,Bayesian Networks and Naïve Bayes, Support Vector Machine, Artificial Neural Networks, Logistic Regression, Genetic Algorithms, Fuzzy Rules, Association Rules, etc.

3.1. Studies Using “Classical” Data Mining Algorithms

Classical data mining algorithms are widely used in disease diagnosis and prognosis. There are several **comparative studies for diagnosing heart diseases** as follow: *M.Kumari and S.Godara [3]* study DM classification methods such as RIPPER classifier, Decision Tree, Artificial neural networks (ANNs), and Support Vector Machine (SVM) used in cardiovascular disease prediction. *V.V.Jaya Rama Krishniah et al. [47]* use K-NN, K-Mean and Entropy based Mean Clustering (EMC) to Predict Heart Attacks, where EMC results most accurate. Also *C. S. Dangare and S.S. Apte [21]* apply three data mining classification techniques namely Decision trees, Naive Bayes and Neural Networks for Heart Disease Prediction. They also present Heart disease prediction system *[12]* using multilayer perceptron neural network along with back propagation algorithm with a resulting accuracy of nearly 100%. Also *B.D.C.N. Prasad et al. [15]* present a comparative study of machine learning algorithms such as Auto-associative memory neural networks, Bayesian networks, ID3 (Iterative Dichotomized 3) and C4.5 in diagnosing Asthma.While *S. Palaniappan and R.Awang [33]* also develop a prototype of Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Networks.

Decision Trees are used to detect diabetes by *K.Kavitha et al. [40]* using CART Decision Trees and *K. Rajesh et al. [46]* using C4.5 Decision Trees. Also *D.S.Kumar et al. [23]* propose the use of decision trees C4.5 algorithm, ID3 algorithm and CART (Classification and Regression Trees) algorithm to classify diabetes, hepatitis and heart diseases. They focus their research on medical diagnosis by learning pattern through the collected data of these diseases and develop intelligent medical decision support systems. **Decision Trees are also used to detect cancer and heart diseases** by *M.T.Khan et al.[24]*, who builds a prototype of Cancer/Heart Disease prediction model using two decision trees algorithms C4.5 and C5.0, showing that C5.0 produces rules in a very easy readable form but C4.5 generates the rule set in the form of a decision tree. While *K-M WANG et al.[49]* use decision trees and logistic regression for predicting breast cancer survivability, concluding that logistic regression performs better.

Bayesian Networks are also widely used in diagnosing heart diseases, as the following: *Chi-Ming Chu et al.[9]* build a Bayesian Expert System for clinically detecting Coronary Artery Disease. *N. AdityaSundar et al. [42]* build a web based system to diagnose patients with heart disease by using data mining techniques, namely Naïve Bayes and WAC (weighted associative classifier). Also *S.A.Pattekari and A. Parveen [45]* use Naive Bayes to predict heart diseases. While *H.P.LeiteFilho[38]* uses bayesian networks in diagnosis of genetic diseases, showing that Naïve Bayes presents the greatest accuracy, compared to decision - ID3, TAN e BAN tree models and neural networks. Also *M.Shahbaz et al. [25]* use data mining classification techniques such as Naïve Bayesian, K-Nearest Neighbors (KNN) and SVM, to build a decision support system to identify different types of cancer on the Genes dataset, from DNA Micro Arrays.

Comparative Studies of data mining techniques used for the prognosis of specific diseasesinclude: *S. Gupta et al. [13]* compare DM classification methods such as Decision Trees, SVM, Genetic Algorithms, Fuzzy Sets, Neural Networks and Rough Sets for Breast Cancer Diagnosis and Prognosis.*D.Delen [26]* uses logistic regression, decision trees, ANNs and SVMs in analyzing prostate cancer data. While*S.Floyd [42]* evaluates the use of ANNs, Bayesian Networks, and SVMs, and concludes that data mining techniques are capable ofimproved prognostic predictions of pancreatic cancer patient survival as compared with logistic regression.

Other **studies of data mining techniques for disease prognosis** are as follow:*A. O. Osofisan et al.[29]* use Artificial Neural Networks for prediction of kidney failure. They use two learning algorithms such as Back Propagation Pattern Mode and Back Propagation Delta-Delta Learning Algorithm to determine the features that are predictive of a patient life expectancy, detect the existence of renal failure in a patient, and to predict kidney dialysis survival.*Sh.G.Jacob and R.G.Ramani [48]* build an efficient classifier for Wisconsin Prognostic Breast Cancer (WPBC) data set from the UCI machine learning repository. This is achieved this by executing twenty classification algorithms, concluding that C4.5 algorithm is the best performing classification algorithm with 100% accuracy.While *T.A. Jilani et al.[28]* use binary regression to investigate factors that contribute significantly to enhancing the risk of acute coronary syndrome.

Studies for treatment of a disease include: *A.A.Aljumah et al. [10]* use regression analysis to predict which treatment contributes more to improvement in hypertension. *L. Duan et al.[18]* present a clinical recommender system, which utilizes a prefix-tree structure common in itemset mining to construct a ranked list of suggested care plan items based on previously-entered items. They use two different types of evaluation mechanisms, called random selection and greedy selection. *T.Sakthimurugan and S.Poonkuzhali [36]* propose a framework using KNN classifier to classify the semantic relations between disease and treatment. *D-Y Yeh et al. [37]* build a predictive model to enhance the accuracy of diagnosis and treatment of cerebrovascular disease, using three classification algorithms, decision tree, bayesian classifier and back propagation neural network, to construct classification models.

Data mining techniques are also used to predict results of surgeries:E.Dincer and N.Duru [27] present a prototype of a tool for analyzing laryngeal cancer operations using k-means clustering algorithm. A.Kika et al.[34] compares three machine learning algorithms, Part (a rule set algorithm), J48 (a decision trees algorithm), and Naïve Bayes to predict peripheral nerves surgery results and discover that J48 results the most accurate. While F.Soleimanianet. al. [43] use C4.5 Decision Trees algorithm to predict results of cesarian sections.

3.2. Studies Using Novel Data Mining Techniques

We can further classify the new techniques into the following categories: *Novel Techniques improving old ones and Hybrid Techniques.*

Novel Data Mining Techniques improving old ones are the following: *S.Chao and F.Wong [19]* propose **i+Learning** (Intelligent, Incremental and Interactive Learning) theory to complement the **traditional incremental decision tree learning algorithms** by concerning new available attributes in addition to the new incoming instances. *They also* present a model of practical data mining diagnostic workbench that intends to support real medical diagnosis by employing data mining and multi-agent technologies *[20]*. *Tyler H. McCormick et al. [17]* propose a statistical modeling technique, called the **Hierarchical Association Rule Model** (HARM), that predicts a patient's possible future symptoms given the patient's current and past history of reported symptoms. In [44]M.A.Jabbar et al. propose a new method to discover **association rules** in medical data to predict heart disease **using matrix based approach** to reduce the number of scans of database. Also **Binary Association Rules** are used by *IrshadUllah [22]* to identify relationships among items in medical databases. He uses binarization to transform numerical data to binary (0, 1) ones, and then he implements SI algorithm and Apriori algorithm for practical implementation purposes.

Novel techniques used for Breast Cancer are the following: A. S.Hesar [41] uses **Probabilistic Bayesian Classifier** (BPCs) for Breast Cancer Diagnosis and Prognosis on WDBC database. BPCs as a data mining tool encode relationships among a set of variables under uncertainty. In construction phase, several heuristic algorithms are evaluated, where Tabu is selected for graphical modeling of proposed BPC and Genie simulator (using Maximize Likelihood Estimation) is used to numerical modeling of BPN.*A.Shukla et al.[14]* present a novel approach to simulate Knowledge Based System for diagnosis of Breast Cancer using Soft Computing tools like Artificial Neural Networks (ANNs) and Neuro Fuzzy Systems.

Other novel approaches are the following: D.Patil et.al. [31] propose a system that predicts the current health status of the patient based on his/her historical and real time data by applying clustering algorithms viz**. K-means and D-stream**. They conclude that Density-based clustering algorithm i.e. the D-stream algorithm overcomes drawbacks of K-means algorithm. D-stream is parameter free and proves to give more accurate results than K-means.E. Savic et al.[16] suggest a general architecture of a system, which allows for combined **Sensor Nets and Data Mining** in medical applications and also two case studies, multiple sclerosis and retinal detachment are presented and analyzed.*M.L. Jimenez et al. [35]* design an Expert System for the diagnosis of health care problems. The Expert System Knowledge Base is composed by a set of production rules written in classic bi-valued logic and by a set of potential facts. The Expert System Inference Engine uses Gröbner Bases and Normal Form to obtain the diagnosis from the information stored in the Knowledge Base.

Hybrid data mininig Techniques are identified as follows: **Decision Trees can be further improved in hybridization with fuzzy model** as follows: M. G. Tsipouras et al. [32] present a fuzzy rule-based decision support system for diagnosing Coronary Artery Disease. They use **C4.5** decision trees and fuzzy model and the obtained results indicate that the proposed optimized fuzzy model DSS improves by 15% in terms of accuracy the results of the crisp rule-based classifier, while it compares well with other widely used classification schemes such as the ANFISand ANNs. Also *N.Bhatla and K.Jyoti [5]*apply Decision Tree and Naïve Bayes Classifiers using Fuzzy Logic to predict more accurately the presence of heart disease with reduced number of attributes. Originally, six attributes were involved in predicting the heart disease, which are further reduced to four attributes and subsequently automatically reduce the number of tests to be taken by a patient. Their proposed system predicts with 100% efficiency. Decision trees are also used by *S.H.Ha and S.H.Joo[11]*who suggest a hybrid data mining method for the medical classification of chest pain. They used an **Apriori algorithm** to identify lab tests and a **C5.0** algorithm to generate a classification rule base for the classification of chest pain, which can help physicians to make clinical decisions faster and more accurately.

In some cases applying **genetic algorithms** on classical algorithms also help improve their performance as in *[8] whereK.S.Kavithaet al.* model and design an evolutionary neural network for heart disease detection by applying hybridization to train the **neural network** usingGenetic algorithm and *[7]*where*J. Soni et al.* study **Decision Trees, Bayesian classification, KNN, Classification based on clustering, Neural Networks** in diagnosing Heart diseases and apply genetic algorithm to further improve Decision Trees and Bayesian Classification.

K-means clustering are usedby *P. Shantakumar et al. [2],* who design an intelligent and effective heart attack prediction system using DM and **Multi-layer Perceptron Neural Networks**. First, they extract significant patterns from the heart disease data warehouses. The preprocessed heart disease data warehouse is clustered with the **K-means clustering**. The frequent items were mined successfully with the aid of **MAFIA** algorithm. On basis of the calculated significant weightage, the frequent patterns comprising of value greater than a predefined threshold were selected for the prediction of heart attack. The MLPNN have been trained with the selected significant patterns by using Back-propagation training algorithm. K-means clustering are also used by *M.Shouman et al.[6]*who investigates integrating **k-means clustering with naïve bayes** in the diagnosis of heart disease patients, higlighting that initial centroid selection strongly affects k-means clustering results, so they systematically investigate applying different methods of initial centroid. Their results show that integrating k-means clustering and naïve bayes can enhance naïve bayes accuracy in the diagnosis of heart disease patients. *M.Shouman et al.[4]* also investigate applying KNN and integrating voting with KNN in the diagnosis of heart disease patients. Their results show that applying KNN achieved an accuracy of 97.4%, while **applying voting could not enhance the KNN accuracy in the diagnosis of heart disease.**

Anothernew hybrid technique is presented by *R. Parvathi and S. Palaniammalì [30]* that employ **image mining, to find spatial association rules**, namely the detection of associations between spatial objects, an effective method for discovering diseases from the symptoms.

4. RESULTS

As most efforts have been done in the field of Cardiovascular diseases (Heart attack, Hypertension, Coronary Artery Disease, etc.) and Cancer diseases, we identify the accuracy levels of classical algorithms for these two groups of diseases as follows:

In Table1, we show the Accuracy for Classical Algorithms, applied on Heart Disease Dataset all by Cleveland cardiovascular disease dataset from UCI repository, except the dataset in [47] which is not specified. A graphical representation of these data is also shown in Figure1. As we can see, almost all of them show good level of accuracy, where Artificial Neural Networks (ANN) with their Multilayer Perceptron Neural Network (MLPNN) model seem to be the best performing with 100% accuracy [12],[21] followed by Decision Trees[21]. However accuracy results may vary according to the number of attributes and the number of instances in the dataset.

Reference	Algorithm	Accuracy (%)	No. of Attributes	No. of Dataset Instances
3	ANN	80.06%	14	296
33	ANN	85.53%	15	909
12, 21	MLPNN	100.00%	15	573
12, 21	MLPNN	99.25%	13	573
23	DT CART	83.20%	14	473
3	DT	79.05%	14	296
21	DT	99.62%	15	573
21	DT	96.66%	13	575
23	DT C4.5	71.40%	14	473
23	DT ID3	64.80%	14	473
47	EMC	82.90%	14	3000
47	K-means	76.90%	14	3000
4	KNN	97.40%	13	573
47	KNN	45.67%	14	3000
21	Naive Bayes	90.74%	15	573
21	Naive Bayes	94.44%	13	574
33	Naive Bayes	86.53%	15	909
42	Naive Bayes	78.00%	15	2268
3	RIPPER	81.08%	14	296
3	SVM	84.12%	14	296
42	WAC	84.00%	15	2268

Table1: Algorithm Accuracy for Cardiovascular Disease diagnosis and prediction

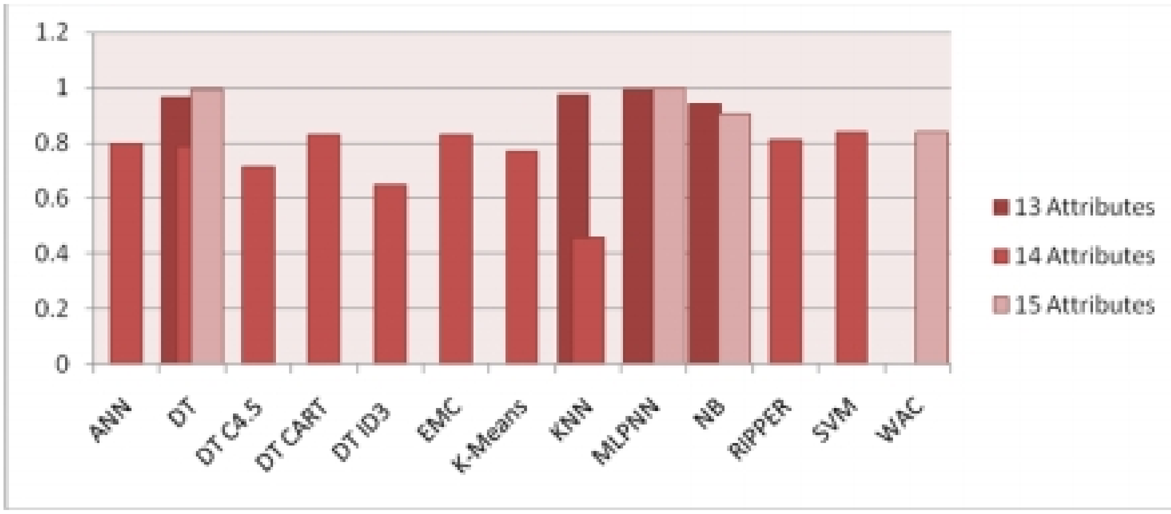

Figure 1: Algorithm Accuracy for Cardiovascular Diseases grouped by the number of attributes in the datasets

In Table2 and graphically in Figure2, we show the Accuracy levels for Classical Algorithms, applied on Cancer Disease Datasetswhere all of them show very good levels of accuracy, with C4.5 Decision Trees showing 100% accuracy on Breast Cancer.

Reference	Algorithm	Accuracy (%)	Disease
26	ANN	91.07%	Prostate Cancer
48	DT C4.5	100.00%	Breast Cancer
49	DT J48	91.07%	Breast Cancer
26	DT	90.00%	Prostate Cancer
25	KNN	90.72%	Leukemia
26	Logistic Regression	89.61%	Prostate Cancer
49	Logistic Regression	91.23%	Breast Cancer
25	Naïve Bayes	95.00%	Leukemia
26	SVM	92.85%	Prostate Cancer
25	SVM	90.27%	Leukemia

Table2: Algorithm Accuracy for Cancer Diseases diagnosis and prognosis

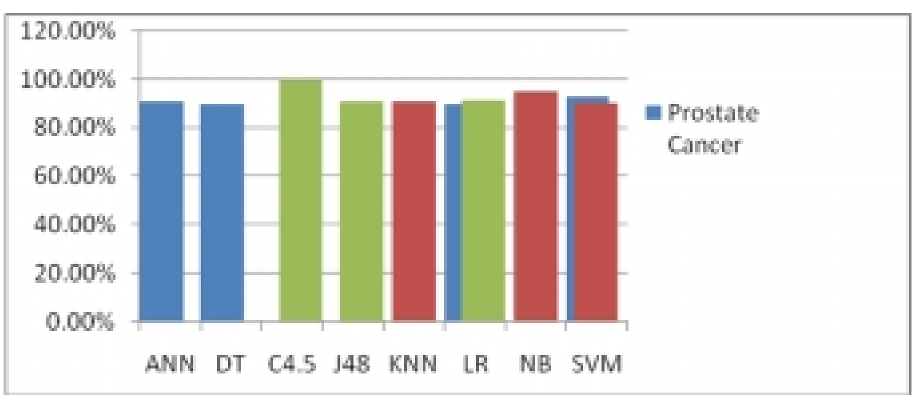

Figure 2: Algorithm Accuracy for Cancer Diseases classified by the type of cancer

5. LIST OF ABRREVIATIONS

ANN - Artificial Neural Networks
BN - Bayesian Networks
DT - Decision Trees
EMC - Entropy Based Mean Clustering
KM - K-Means Clustering
KNN - K-Nearest Neighbors
LR - Logistic Regression
MLPNN - Multilayer Perceptron Neural Network
NB - Naïve Bayes
SVM - Support Vector Machine
WAC - Weighted Associative Classifier

6. CONCLUSIONS

In this study we have presented an overview of the current research being carried out on different data mining techniques and tools used in diagnosis, prognosis and treatment of various diseases. Analyses show that most of the researches have been done in studying classic data mining algorithms such as Decision Trees,Bayesian Networks and Naïve Bayes, Support Vector Machine, Artificial Neural Networks, Logistic Regression, Genetic Algorithms, Fuzzy Rules, Association Rules, showing acceptable levels of accuracy.

Good efforts have also been done in developing novel and hybrid data mining techniques for enhanced diagnosis and prognosis accuracy. However, using data mining techniques to identify treatment options for patients has received less attention. Results also show that most efforts have been done in the field of Cardiovascular diseases (Heart attack, Hypertension, Coronary Artery Disease, etc.) and Cancer diseases.

The follow-up of this study will aim at dealing with data mining techniques that have wider spectra of application for groups of diseases.

7. REFERENCES

[1] Fayyad, U. M. ,Piatetsky-Shapiro, G., Smyth, P., Uthurusamy , R. G. R.: Advances in Knowledge Discovery and Data Mining. AAAI Press / The MIT Press (1996)

[2] ShantakumarB.Patil, Y.S.Kumaraswamy, Intelligent and EffectiveHeart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of ScientificResearchISSN1450-216X Vol.31 No.4 (2009), pp.642-656

[3] M.Kumari, S. Godara ,Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction, IJCST ISSN : 2229- 4333 Vol. 2, Issue 2, (2011)

[4] Mai Shouman, Tim Turner, and Rob Stocker, Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients, International Journal of Information and Education Technology, Vol. 2, No. 3 (2012) pp 220-223

[5] NidhiBhatlaKiranJyoti ,A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic, International Journal of Computer Applications (0975 – 8887) Volume 54– No.17, (2012) pp 16-21

[6] Mai Shouman, Tim Turner, Rob Stocker ,Integrating Naive Bayes And K-Means Clustering With Different Initial Centroid Selection Methods In The Diagnosis Of Heart Disease Patients, CS & IT-CSCP (2012) pp. 125–137

[7] J.Soni, U. Ansari, D. Sharma, S. Soni, Predictive Data Mining for Medical Diagnosis: An Overviewof Heart Disease Prediction (2011)
[8] K.S.Kavitha , K.V.Ramakrishnan , M. K. Singh, Modeling and design of evolutionary neural network for heart disease detection, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010, ISSN (Online): 1694-0814, pp 272-283
[9] Chi-Ming Chu, Wu-ChienChien, Ching-Huang Lai, Hans-Bernd Bludau, Huei-Jane Tschai, LuPai,Shih-Ming Hsieh, Nian-Fong Chu, Angus Klar, Reinhold Haux, Thomas Wetter, A Bayesian Expert System for Clinical Detecting Coronary Artery Disease, J Med Sci 2009; 29(4), pp 187-194
[10] A.A. Aljumah, M. G.Ahamad, M.K.Siddiqui , Predictive Analysis on Hypertension Treatment Using Data Mining Approach in Saudi Arabia, Intelligent Information Management, 3, (2011), pp 252-261
[11] S.H.Ha, S.H.Joo, A Hybrid Data Mining Method for the Medical Classification of Chest Pain, International Journal of Computer and Information Engineering 4:1, (2010), pp 33-38
[12] Chaitrali S. Dangare, Sulabha S. Apte, A Data Mining Approach For Prediction Of Heart Disease Using Neural Networks, International Journal Of Computer Engineering & Technology (IJCET) ISSN 0976 – 6367(Print) ISSN 0976 – 6375(Online) Volume3, Issue 3 (2012), pp. 30-40
[13] S.Gupta, D. Kumar, A.Sharma, Data Mining Classification Techniques Applied For Breast Cancer Diagnosis And Prognosis (2011)
[14] A.Shukla, R. Tiwari, P. Kaur, Knowledge Based Approach for Diagnosis of Breast Cancer , IEEE International Advance Computing Conference (IACC 2009)
[15] B.D.C.N. Prasad, P.E.S.N. K.Prasad, Y. Sagar, A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma) (2011)
[16] E. Savic, J.Potic, Z. Babovic, G. Rakocevic, V. Strineka, M. Dobrota, V. Milutinovic2, Sensor Nets and Data Mining in Medical Applications (2011)
[17] T.H. McCormick, C. Rudin, D.Madigan , A Hierarchical Model For Association Rule Mining Of Sequential Events: An Approach To Automated Medical Symptom Prediction ()
[18]L. Duan, W. N. Street & E. Xu (2011): Healthcare information systems: data mining methods in the creation of a clinical recommender system, Enterprise Information Systems, 5:2, 169-181
[19] S. CHAO, F.WONG, An Incremental Decision Tree Learning Methodology Regarding Attributes In Medical Data Mining (2009)
[20] S.Chao , F. Wong, A Multi-Agent Learning Paradigm for Medical Data Mining Diagnostic Workbench (2009)
[21] C. S. Dangare, S.S. Apte, Improved Study of Heart Disease Prediction SystemUsing Data Mining Classification Techniques (2012)
[22] I.Ullah, Data Mining Algorithms And Medical Sciences (2012)
[23]D.S.Kumar, G.Sathyadevi, S.Sivanesh, Decision Support System for Medical Diagnosis UsingDataMining (2011)
[24]Mohammad Taha Khan, Dr. ShamimulQamar and Laurent F. Massin ,A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012)
[25]M.Shahbaz et al.,Cancer Diagnosis Using Data Mining Technology, Life Science Journal. (2012) ;9(1) (ISSN:1097-8135), pp 308-313
[26]D.Delen, Analysis of cancer data: a data mining approach (2009)

[27]E.Dincer, N.Duru Prototype of a tool for analysing laryngeal cancer operations
[28]Acute Coronary Syndrome Prediction Using Data Mining Techniques- An Application, World Academy of Science, Engineering and Technology 59 (2009) pp.474-478
[29]A.O. Osofisan ,O.O. Adeyemo, B.A. Sawyerr, O. Eweje, Prediction of Kidney Failure Using Artificial Neural Networks (2011)
[30] R. Parvathi,S. Palaniammalì, An Improved Medical Diagnosing Technique Using Spatial Association Rules, European Journal of Scientific Research ISSN 1450-216X Vol.61 No.1 (2011), pp. 49-59
[31]D.Patil et.al. ,An Adaptive parameter free data mining approach for healthcare application, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 1 (2012)
[32] M.G. Tsipouras, T.P. Exarchos, D.I. Fotiadis,A.P. Kotsia, K.V. Vakalis, K.K. Naka, L. K. Michalis Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling (2008)
[33] S.Palaniappan, R. Awang, Intelligent Heart Disease Prediction System Using Data Mining Techniques (2008)
[34] A.Kika, B.Cico, R.Alimehmeti, Using Machine Learning for Preoperative Peripheral Nerve Surgical Prediction (2010)
[35]M. L.Jimenez , J. M. Santamarı, R. Barchino, L. Laita, L.M. Laita, L. A. Gonza´lez, A. Asenjo, Knowledge representation for diagnosis of care problems through an expert system: Model of the auto-care deficit situations, Expert Systems with Applications 34 (2008) 2847–2857
[36] T.Sakthimurugan, S.Poonkuzhali, An Effective Retrieval of Medical Records using Data Mining Techniques, International Journal Of Pharmaceutical Science And Health Care. ISSN: 2249-5738. 2(2) (2012), pp 72-78
[37]Duen-YianYeh, Ching-Hsue Cheng, Yen-Wen Chen,A predictive model for cerebro-vascular disease using data mining,Expert Systems with Applications 38 (2011) 8970–8977
[38]Hugo Pereira LeiteFilho ,Applicability of Data Mining Technique usingBayesians Network in Diagnosis of Genetic Diseases, (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 4, No.1, (2013) pp. 47-50
[40]Kavitha K, Sarojamma R M ,Monitoring of Diabetes with Data Mining via CART Method, International Journal of Emerging Technology and Advanced Engineering[IJETAE] ISSN 2250-2459, Vol.2, No. 11 (2012) pp 157-16
[41]AlirezaSadeghiHesar ,A Probabilistic Bayesian Classifier Approach for BreastCancer Diagnosis and Prognosis, Journal of Advances in Computer ResearchQuarterly ISSN: 2008-6148Vol. 3, No. 2 (2012), pp 35-44
[42]N. AdityaSundar, P. PushpaLatha, M. Rama Chandra,Performance Analysis Of Classification Data Mining Techniques Over Heart Disease Data Base, [IJESAT] International Journal Of Engineering Science & Advanced Technology, ISSN: 2250–3676, Vol.-2, Issue-3, pp470 – 478
[43]F.Soleimanianet. al. ,Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study, International Journal of Computer Applications (0975 – 8887)Vol.52, No. 6, (2012) pp 21-26
[44]M.A.Jabbar et al. ,Knowledge Discovery From Mining AssociationRules For Heart Disease Prediction, Journal of Theoretical and Applied Information Technology (2012) Vol. 41 No.2, ISSN: 1992-8645, pp 166-174

[45]S.A.Pattekari and A. Parveen ,Prediction System For Heart Disease Using Naive Bayes, International Journal of Advanced Computer and Mathematical SciencesISSN 2230-9624. Vol 3, Issue 3 (2012) pp 290-294

[46]K. Rajesh, V. Sangeetha ,Application of Data Mining Methods and Techniques for Diabetes Diagnosis, ISSN: 2277-3754International Journal of Engineering and Innovative Technology (IJEIT)Volume 2, Issue 3, (2012), pp 224-229

[47]V.V.Jaya Rama Krishniah et al. ,Predicting the Heart Attack Symptomsusing Biomedical Data Mining Techniques, The InternationalJournal of Computer Science & Applications (TIJCSA) ISSN – 2278-1080, Vol. 1 No.3 (2012), pp 10-18

[48]ShomonaGracia Jacob, R. GeethaRamani ,Efficient Classifier for Classification of Prognostic Breast Cancer Data through Data Mining Techniques, Vol IWCECS 2012, October 24-26, (2012) San Francisco, USA

[49] Kung-Min Wang et al.,Optimal Data Mining Method For Predicting Breast Cancer Survivability, ISME International 2012 ISSN 2185-5455, Vol. 3, No. 2, (2012) pp. 28-33