A REVIEW OF DATA MINING TECHNIQUES IN MEDICINE

Ionela-Cătălina ZAMFIR[1]
Ana-Maria Mihaela IORDACHE[2]

Abstract: *Data mining techniques found applications in many areas since their development. New techniques or combination of techniques are created continuously. Techniques like supervised or unsupervised learning are used for prediction different diseases and have the aim to identify (diagnosis) the disease and predict the incidence or make predictions about treatment and survival rate. This paper presents the new trends in applying data mining in healthcare area, taking into account the most relevant papers in this respect. By analyzing both applied and review articles, using text mining on keywords and abstracts, it is revealing the most used methodologies (Support Vector Machines, Artificial Neural Networks, K-Means Algorithm, Decision Trees, Logistic Regression) as well as the areas of interest (prediction of different diseases, like: breast cancer, lung cancer, heart diseases, diabetes, thyroid or kidney diseases).*

Keywords: *data mining, predictions, literature review, disease, healthcare, medicine*

JEL classification: I15, C38

1. Introduction

Data Mining techniques know a wide spread and applicability nowadays. Among the areas that use these techniques, the medical and economic fields have the most interesting applications. In economic area, the most studied issues are: the prediction of bankruptcy risk for companies, the prediction of fraud in insurance, fraud in financial statements, fraud in credit card transactions, or the customers' classification for targeted marketing campaign. In medical field, some of the most approached issues are: the identification of cancer (breast, lung, or other types), hearth diseases, diabetes, skin disease or the prediction of different diseases incidence.

Because "there are two primary goals for data mining prediction and description" (Sharma et.all., 2014), the pattern recognition techniques (both supervised and unsupervised) are the most used methods, as well as methods that extract essential knowledge from data, like dimension reduction techniques. The description goal

[1] Assistant professor, Phd, The Bucharest University of Economic Studies, Bucharest
[2] Lecturer, Phd, School of Computer Science for Business Management, Romanian-American University, Bucharest

suppose analyzing the raw data (huge amount of data) and extract important information that describe a specific case or model a group of facts with the same characteristics or rules. The prediction suppose analyzing a set of observations and variables that have similar features and describe the same phenomenon, classify observations into several classes of interest and extract a "rule" to fit other new observations into one of the classes.

This research is a review of the most relevant articles that describe the use of data mining techniques in medical field. The methodologies section describes briefly the analyzed papers, the results show the most important findings, while conclusions and further search summarize the research and show future directions.

2. Methodologies

Data Mining represents a multitude of techniques and methods that helps extracting "hidden" information or patters from big amount of data. In medicine, starting from patient medical history, symptoms or new medical investigation, all is data. Using modern technology, this data is recording in large databases. The idea of using this huge amount of data to develop a decision support system that helps doctors to diagnose, treat and follow patients is not new, as well as the usage of data mining techniques to accomplish this goal.

The focus of this paper is to analyze new articles (starting 2010, until 2018), in order to identify data mining techniques that are used for medical purpose and its applications. There were selected 76 most relevant studies from 2010 to 2018, some of them are literature review (that analyze other papers or review the possible data mining techniques and applications), but most of them are applicative articles.

In the applicative articles, in general is used a database that is formed either of patients medical information and details ([62], [52]), either of images like chest x-rays for detecting lung cancer ([41]). In the most databases used, the patient's diagnosis is known, as well as their treatment and medical future. What data mining techniques usually try to "solve" is creating a model based on available data that is capable to support doctors for diagnose new patients. In order to test these new models, several techniques are used, like substitution or n-fold cross validation (where n is equal to 10 in many articles). An accuracy degree, sensitivity and specificity are calculated for each technique or group of techniques, in order to give more credibility and robustness. In general, the accuracy of predicting the disease is over 80% and it depends on variable selection, case modeled and technique used.

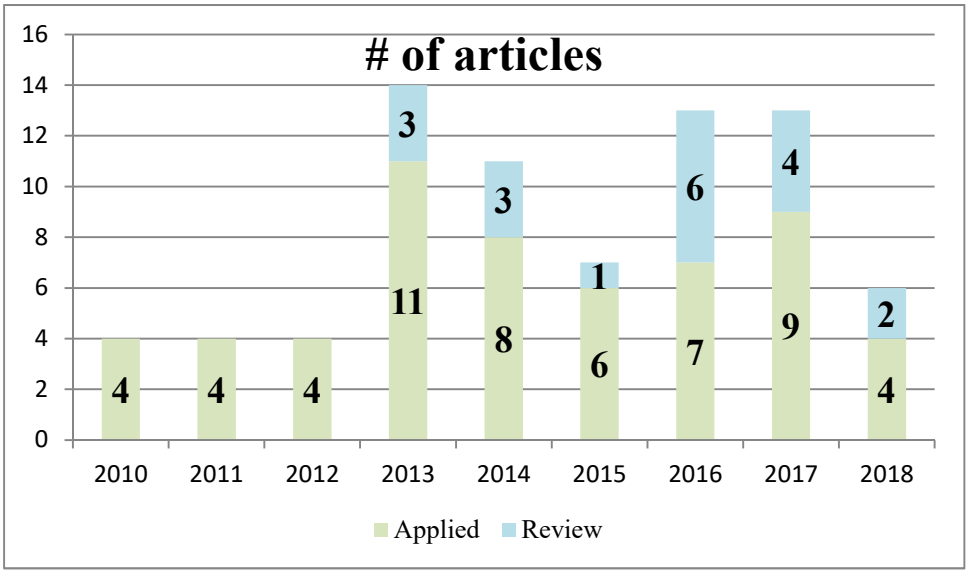

Figure 1. Articles database structure

The figure from above shows the distribution on years of the selected articles. There are 76 articles, 19 reviews (of literature or data mining techniques) and 57 applied papers from 2010 to 2018. Most of the articles (65%) are from the last 5 years.

Table 1. Papers description

Type of article	Subject	Examples
applied	cancer	breast cancer ([13], [20], [24], [25], [27], [29], [30], [35], [40], [62], [69], [70], [71], [74]), colon cancer ([36]), liver cancer ([63]), lung cancer ([41], [51], [57], [65], [68]), ovarian cancer ([33]), pancreatic cancer ([61], [75]), thyroid cancer ([76]), other type of cancer or multiple types ([17], [19], [37], [72])
	heart disease	studies like: [3], [5], [21], [34], [44], [47], [49], [52], [53], [56], [59], [67])
	diabetes	articles: [11], [14], [45], [64]
	multiple disease	studies like: [16], [42], [48], [54]
	specific disease	Parkinson's disease ([4]), thyroid disease ([12]), liver disease ([18]), kidney disease ([32], [43])
	other	studies like: [7], [8], [46]
review	cancer	lung cancer ([39])
	heart disease	studies like: [2], [6], [23], [31], [50], [55], [58], [60]
	multiple disease	[66]
	other	[1], [9], [10], [15], [22], [26], [28], [38], [73]

In the table from above is represented the main themes for selected articles. The subject that is approached often by researchers is the identification and prediction of cancer (that could be breast cancer, colon cancer, liver cancer, lung cancer, ovarian cancer, thyroid cancer or pancreatic cancer). Also, heart diseases, which are considered on top 5 most important causes of death on the entire world, represent an immense interest for researchers and doctors. The applied papers describe examples of data mining techniques, most of papers approaching several techniques that are compared from accuracy point of view (and training time), because each case have its own specificity and particularities, so one specific method or a combination of methods can be applied. The review papers discuss about literature review or techniques review in general, most of them without having a numerical example.

Also, there are some articles that have the main objective the proposal of a software application that is useful for disease prediction (web application that can predict a disease and have as input variables that represent the patient symptoms - [46]) or more specific for heart diseases and diabetes ([44]). On the other side, in articles like [68] and [76] it is presented models of survival prediction for patients with lung cancer and thyroid cancer. Authors used boosted Support Vector Machines, with an accuracy rate of 97% for lung cancer live expectancy prediction ([68]) and MLP and Logistic Regression with an average of over 80%-90% accuracy for thyroid cancer survival ([76]).

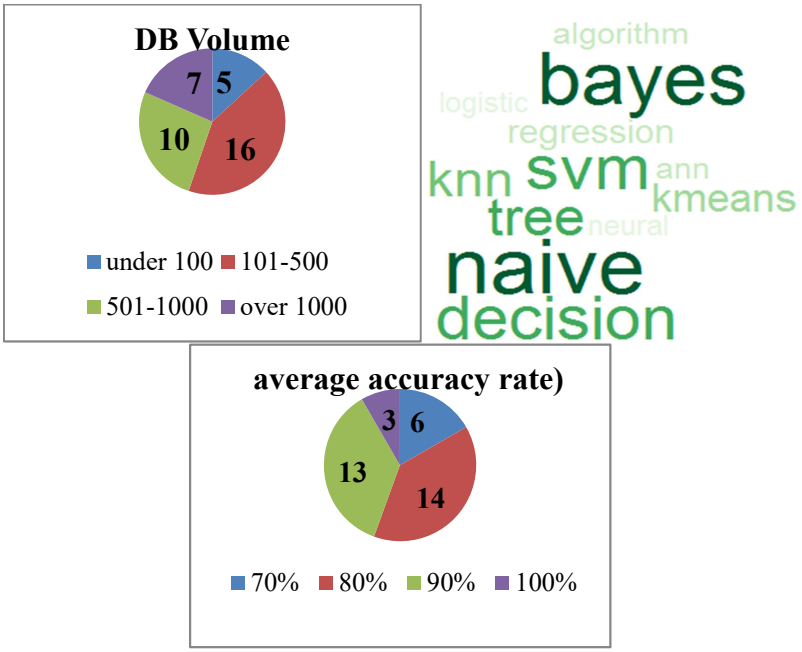

Figure 2. Databases volumes, methodologies used and accuracies rates (in average)

The databases size used in most applicative articles depends on the studied problem. Only 5 studies have datasets under 100 observations, while 7 studies have datasets over 1000. The methodologies most used (the middle figure from above was created in R, using text mining over methodologies that appear in applied studies) are: Decision Trees, Naive Bayes, K-Nearest-Neighbor, Support Vector Machines, K-Means, Logistic Regression, ANN (Artificial Neural Network). The bigger the word is in the middle figure from above, the higher frequency has in methodologies from the articles. On the other side, the accuracy degree is over 70% in most studies with applications, some reach to even 100% (the prediction of heart attack - [53], diagnostic diabetes - [64], breast, lung and skin cancer prediction - [72]), that means that the model is robust and reliable.

3. Results and discussions

Text mining represents (as data mining) extracting relevant information from text. Word cloud is a technique of text mining that uses the frequency of each word to create relevant graphs. This method could be used to extract (graphically) the most used words in different texts (after the texts were "cleaned" of punctuation, numbers), like: emails, reviews on a web site, comments to news. In this case, the technique was used to extract the relevant information from studies abstracts and keywords.

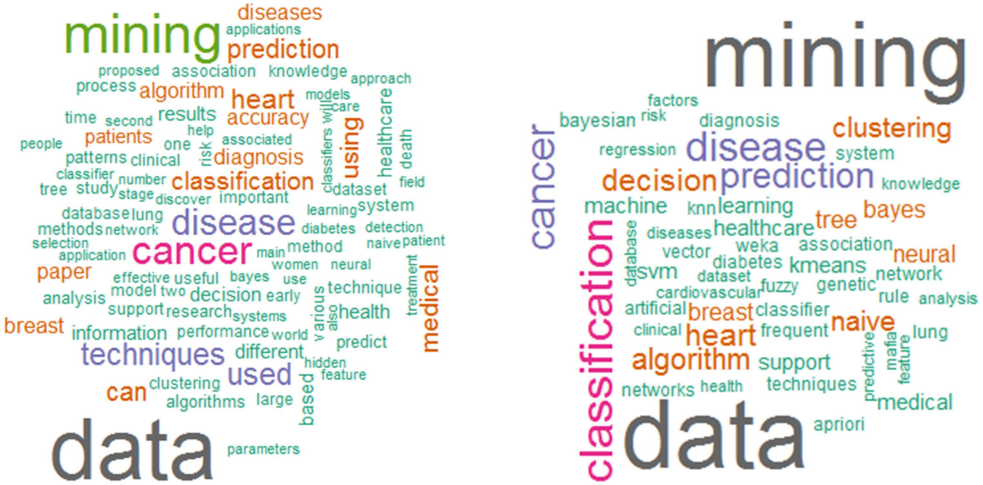

Figure 3. All abstracts (min freq=15) and all keywords (min freq=3)

The figure from above show the graphically representation of the most frequently used words in articles abstracts and keywords. In this graph, all articles (review and applied) were considered. For abstracts, only the words with a frequency higher than 15 appear in the graph, while keywords with frequency higher than 3 are printed. The bigger a word is represented, the higher frequency it has. Except "data

mining" and "techniques", words like "cancer", "heart", "breast", "medical", "patients" and "diagnosis" appear as the most approached issues.

Figure 4. Applied papers abstracts (min freq=15) and review papers abstracts (min freq=5)

The figure from above show the difference between the most used words in applied papers abstracts (with a frequency higher than 15) and the most used words from review articles (minimum frequency is 5). Except "data mining" words, in review papers, words like "techniques", "heart", "prediction" and "disease" are relevant, that show the concern of researchers about prediction techniques in diseases in

general and heart issues especially. In applied papers, "classification", "algorithm" and "accuracy" is relevant, because in most of the applications, an accuracy degree is computed as number of correct classified observations (obtained with methods like n-fold cross validation) divided by the number of total observations.

Figure 5. Articles from last 5 years: abstracts (min freq=10) and keywords (min freq=5)

In order to identify what are the new "trends" in applying data mining techniques for medical purpose, the articles from the past 5 years were selected from considered database of papers. The figure from above show the representation of most frequent words in abstracts (higher frequency than 10) and keywords (minimum appearances is 5) from 50 articles published between 2014 and 2018, both review and applied papers. It is interesting to notice that SVM and K-Means as methodologies does not appear in the keywords word cloud (right side of the figure). Also, issues like cancer and heart diseases remain the most researched areas.

4. Conclusions and further research

By analyzing 76 articles that contain description of data mining techniques and methods applied in healthcare area, as well as applications of these methods for predicting, identifying or follow different diseases cases, this research aim is to describe the latest concern about the use of methods that extract essential information from data. This article proposes a text mining technique to analyze the papers from an area of interest and see what words are common in most articles abstracts and keywords (with the highest frequency rate). This method could be used for many others areas of research, like emails spam detection, sentiment analysis or comments analysis, directions that represent further research.

References

[1] "Aarti Sharma et. all., 2014, Application of Data Mining – A Survey Paper, International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2023-2025.
[2] Ankita Shrivastava, Shivkumar Singh Tomar, 2016, A hybrid framework for heart disease prediction: review and analysis, International Journal of Advanced Technology and Engineering Exploration, Vol 3(15), 21-27, ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
[3] Dubey, A., Patel, R., Choure, K., 2014, An Efficient Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction, International Journal of Advanced Technology and Engineering Exploration, Vol. 1(1), 1-6, ISSN (Online): 2394-7454
[4] Tucker C. et. all., 2015, A data mining methodology for predicting early stage Parkinson's disease using non-invasive, high dimensional gait sensor data, IIE Trans Healthc Syst Eng., Vol. 5(4), 238-254
[5] K. Aravinthan, Dr. M.Vanitha, 2016, A Novel Method For Prediction Of Heart Disease Using Naïve Bayes, International Journal of Advanced Research Trends in Engineering and Technology, Vol. 3, Special Issue 20, , ISSN 2394-3777 (Print), ISSN 2394-3785 (Online)

[6] Amandeep Kaur, Jyoti Arora, 2018, Heart disease prediction using data mining techniques: a survey, International Journal of Advanced Research in Computer Science, Vol. 9, No. 2.

[7] Kamble, N., et.all., 2017, Smart Health Prediction System Using Data Mining, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 2, Issue 2, 1020-1025, ISSN : 2456-3307

[8] Ghiy, N., et.all., 2015, Forecasting Diseases by Classification and Clustering Techniques, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue 1, 18958-18961, ISSN(Online): 2319 - 8753, ISSN (Print) :2347 - 6710

[9] Kauser Ahmed P, 2017, Analysis of Data Mining Tools for Disease Prediction, Journal of Pharmaceutical Sciences and Research, Vol. 9(10), 1886-1888, ISSN: 0975-1459

[10] Nimna Jeewandara, Dinesh Asanka, 2017, Data mining techniques in prevention and diagnosis of non communicable diseases, International Journal of Research in Computer Applications and Robotics, Vol.5 Issue 11, Pg.: 11-17, ISSN 2320-7345

[11] G. Krishnaveni, T.Sudha, 2017, A novel technique to predict diabetic disease using data mining – classification techniques, International Journal of Advanced Scientific Technologies, Engineering and Management Sciences (IJASTEMS), Vol..3,Special Issue.1, , ISSN: 2454-356X

[12] Prerana, Parveen Sehgal, Khushboo Taneja, 2015, Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network, International Journal of Research in Management, Science & Technology, Vol. 3, No. 2, 75-80, E-ISSN: 2321-3264

[13] A.Kavipriya, B.Gomathy, 2013, Data Mining Applications in Medical Image Mining: An Analysis of Breast Cancer using Weighted Rule Mining and Classifiers, IOSR Journal of Computer Engineering (IOSRJCE), Vol. 8, Issue 4, 18-23, ISSN: 2278-0661, ISBN: 2278-8727

[14] P. Radha, Dr. B. Srinivasan, 2014, Predicting Diabetes by cosequencing the various Data Mining Classification Techniques, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1, Issue 6, 334-339, ISSN 2348 – 7968

[15] V.Kirubha, S.Manju Priya, 2016, Survey on Data Mining Algorithms in Disease Prediction, International Journal of Computer Trends and Technology, Vol.38(3), 124-128, ISSN:2231-2803

[16] V. Chaurasia, S. Pal, 2014, Performance analysis of data mining algorithms for diagnosis and prediction of heart and breast cancer disease, Review of Research, Vol. 3 (8), ISSN:-2249-894X

[17] Zakaria Suliman Zubi, Marim Aboajela Emsaed, 2010, Sequence Mining in DNA chips data for Diagnosing Cancer Patients, SELECTED TOPICS in APPLIED COMPUTER SCIENCE, 139-151, ISSN: 1792-4863

[18] Dalia Abd El Hamid Omran, et.all., 2016, Application of Data Mining Techniques to Explore Predictors of HCC in Egyptian Patients with HCV-related Chronic Liver Disease, Asian Pacific Journal of Cancer Prevention, Vol. 16, 381-385

[19] Joaquín Pérez-Ortega, et.all., 2010, Spatial Data Mining of a Population-Based Data Warehouse of Cancer in Mexico, International Journal of Combinatorial Optimization Problems and Informatics, Vol. 1, No. 1, pp. 61-67, ISSN: 2007-1558

[20] Wenjing Zhang, Donglai Ma, Wei Yao, 2014, Medical Diagnosis Data Mining Based on Improved Apriori Algorithm, JOURNAL OF NETWORKS, Vol. 9, No. 5, ISSN 1796-2056

[21] Boshra Bahrami, Mirsaeid Hosseini Shirvani, 2015, Prediction and Diagnosis of Heart Disease by Data Mining Techniques, Journal of Multidisciplinary Engineering Science and Technology (JMEST), Vol. 2 Issue 2, pp. 164-168, ISSN: 3159-0040

[22] R Shouval, et.all., 2013, Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT, Bone Marrow Transplantation, Vol. 49, 332-337

[23] R. Subha, K. Anandakumar, A. Bharathi, 2016, Study on Cardiovascular Disease Classification Using Machine Learning Approaches, International Journal of Applied Engineering Research, Vol. 11, Number 6, pp 4377-4380, ISSN 0973-4562

[24] Jaimini Majali, et.all., 2014, Data Mining Techniques For Diagnosis And Prognosis Of Breast Cancer, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5(5), 6487-6490, ISSN:0975-9646

[25] Ruijuan Hu, 2011, Medical Data Mining Based on Decision Tree Algorithm, Computer and Information Science, Vol. 4, No. 5, 14-19, ISSN 1913-8989, E-ISSN 1913-8997

[26] Parvez Ahmad, et.all., 2015, Techniques of Data Mining In Healthcare: A Review, International Journal of Computer Applications, Volume 120 – No.15, 38-50

[27] Aswini Kumar Mohanty, Saroj Kumar Lenka, 2010, Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer, Special Issue of IJCCT for International Conference [ICCT-2010], Vol. 2, Issue 2,3,4, 99-106

[28] Dr. D. P. Shukla, Shamsher Bahadur Patel, Ashish Kumar Sen, 2014, A Literature Review in Health Informatics Using Data Mining Techniques, International Journal of Software & Hardware Research in Engineering, Vol. 2, Issue 2, 123-129, ISSN: 2347-4890

[29] Tejal A. Patel, et.all, 2016, Correlating Mammographic and Pathologic Findings in Clinical Decision Support Using Natural Language Processing and Data Mining Methods, Cancer, 123(1), 114-121

[30] Vikas Chaurasia, Saurabh Pal, BB Tiwari, 2018, Prediction of benign and malignant breast cancer using data mining techniques, Journal of Algorithms & Computational Technology, Vol. 12(2), 119-126

[31] Beant Kaur, Williamjeet Singh, 2014, Review on Heart Disease Prediction System using Data Mining Techniques, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 Issue: 10, 3003-3008, ISSN: 2321-8169

[32] Lambodar Jena, Narendra Ku. Kamila, 2015, Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease, International Journal of Emerging Research in Management &Technology, Volume-4, Issue-11, 110-118, ISSN: 2278-9359

[33] Kavita Rawat, Kavita Burse, 2013, A Soft Computing Genetic-Neuro fuzzy Approach for Data Mining and Its Application to Medical Diagnosis, International Journal of Engineering and Advanced Technology (IJEAT), Volume-3, Issue-1, 409-411, ISSN: 2249 – 8958

[34] Nidhi Bhatla, Kiran Jyoti, 2012, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 8, , ISSN: 2278-0181

[35] K.R.Lakshmi, M.Veera Krishna, S.Prem Kumar, 2013, Performance comparison of data mining techniques for prediction and diagnosis of breast cancer disease survivability, Asian Journal Of Computer Science And Information Technology, Vol. 3(5), 81-87, ISSN 2249-5126

[36] D. S.V.G.K.Kaladhar et.all., 2013, The Elements of Statistical Learning in Colon Cancer Datasets: Data Mining, Inference and Prediction, Algorithms Research, Vol. 2(1), pp. 8-17

[37] Muhammad Shahbaz, et.all., 2012, Cancer diagnosis using data mining technology, Life Science Journal, Vol. 9(1), 308-313

[38] Isha Vashi, Prof. Shailendra Mishra, 2016, A Comparative Study of Classification Algorithms for Disease Prediction in Health Care, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 9, 16617 - 16621, ISSN(Online): 2320-9801, ISSN (Print): 2320-9798

[39] H Bharathi, TS Arulananth, 2017, A Review of Lung cancer Prediction System using Data Mining Techniques and Self Organizing Map (SOM), International Journal of Applied Engineering Research, Volume 12, Number 10, pp. 2190-2195, ISSN 0973-4562

[40] Emilien Gauthier, et.all., 2011, Breast cancer risk score: a data mining approach to improve readability, The International Conference on Data Mining, Las Vegas, United States, pp.15 - 21

[41] Z. Suliman Zubi, R. Asheibani Saad, 2014, Improves Treatment Programs of Lung Cancer Using Data Mining Techniques, Journal of Software Engineering and Applications, Vol. 7, 69-77

[42] Shelly Gupta, Dharminder Kumar, Anand Sharma, 2011, Performance analysis of various data Mining classification techniques on Healthcare data, International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 4, 155-169

[43] Jyoti Saini, R.C Gangwar, Mohit Marwaha, 2017, A novel detection for kidney disease Using improved support vector machine, International Journal of Latest Trends in Engineering and Technology, Vol.(8)Issue(1), pp.114-121, e-ISSN:2278-621X

[44] Ammar Aldallal, Amina Abdul Aziz Al-Moosa, 2017, Prediction of Heart Diseases and Diabetes Using Data Mining Techniques A Study of MOI Health Center, Kingdom of Bahrain, Journal of the Association for Information Science and Technology, Vol. 2017, issue 9, (Online) ISSN: 2330-1643

[45] Lee, Tian-Shyug; Dai, Wensheng; Huang, Bo-Lin; Lu, Chi-Jie, 2017, Data mining techniques for forecasting the medical resource consumption of patients with diabetic nephropathy, International Journal of Management, Economics and Social Sciences (IJMESS), Vol. 6, Iss. Special Issue, pp. 293-306, ISSN 2304-1366

[46] Pushkaraj R. Bhandari, et.all., 2016, A Survey on Predictive System for Medical Diagnosis with Expertise Analysis, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 2, 2026-2029, ISSN(Online): 2320-9801

[47] Sonam Nikhar, A.M. Karandikar, 2016, Prediction of Heart Disease Using Machine Learning Algorithms, International Journal of Advanced Engineering, Management and Science (IJAEMS), Vol-2, Issue-6, 617-621, ISSN : 2454-1311

[48] Rahul Deo Sah, Dr. Jitendra Sheetalani, 2017, Review of Medical Disease Symptoms Prediction Using Data Mining Technique, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 19, Issue 3, Ver. I, PP 59-70, e-ISSN: 2278-0661, p-ISSN: 2278-8727

[49] Shilna S, Navya EK, 2016, Heart disease forecasting system using k-mean clustering algorithm with PSO and other data mining methods, International Journal On Engineering Technology and Sciences, Vol. III, Issue IV, 51-55, ISSN(P): 2349-3968, ISSN (O): 2349-3976

[50] Susmitha K, B. Senthil Kumar, 2018, A Survey on Data Mining Techniques for Prediction of Heart Diseases, IOSR Journal of Engineering, Vol. 08, Issue 9, pp. 22-27, ISSN (e): 2250-3021

[51] Omar, Y., Tasleem, A., Pasquier, M., Sagahyroon, A., 2018, Lung Cancer Prognosis System using Data Mining Techniques, In Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018), Volume 5: HEALTHINF, pp. 361-368, ISBN: 978-989-758-281-3

[52] Noreen Akhtar, Muhammad Ramzan Talib, Nosheen Kanwal, 2018, Data Mining Techniques to Construct a Model: Cardiac Diseases, (IJACSA)

International Journal of Advanced Computer Science and Applications,, Vol. 9, No. 1, 532-536

[53] M. Mirmozaffari, A. Alinezhad, A. Gilanpour, 2017, Heart Disease Prediction with Data Mining Clustering Algorithms, Int'l Journal of Computing, Communications & Instrumentation Engg, Vol. 4,Issue 1, 16-19,ISSN 2349-1469 EISSN 2349-1477

[54] Shakuntala Jatav, Vivek Sharma, 2018, An algorithm for predictive data Mining approach in medical diagnosis, International Journal of Computer Science & Information Technology (IJCSIT), Vol 10, No 1, pp. 11-20

[55] Mudasir M Kirmani, 2017, Cardiovascular Disease Prediction Using Data Mining Techniques: A Review, Oriental Journal of Computer Science & Technology, Vol. 10, No. (2), 520-528, ISSN: 0974-6471

[56] K. Aparna, et.all., 2014, Disease Prediction in Data Mining Techniques, International Journal of Computer Science And Technology, Vol.5, Issue 2, 246-249, ISSN: 0976-8491 (Online)

[57] S. Durga, K. Kasturi, 2017, Lung Disease Prediction System Using Data Mining Techniques, Jour of Adv Research in Dynamical & Control Systems, Vol. 9, No. 5, 62-66, ISSN 1943-023X

[58] Ramandeep Kaur, et al, 2016, A Review - Heart Disease Forecasting Pattern using Various Data Mining Techniques, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.6, pg. 350-354, ISSN 2320–088X

[59] M.A.Nishara Banu, B Gomathy, 2013, Disease predicting system using data mining techniques, International Journal of Technical Research and Applications, Vol. 1, Issue 5, 41-45, e-ISSN: 2320-8163

[60] Sonam Nikhar, A. M. Karandikar, 2016, "Prediction of heart disease using data mining Techniques" - A Review, International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 02, 1291-1293, e-ISSN: 2395 -0056, p-ISSN: 2395-0072

[61] Xiaodong Feng, et.all., 2013, Assessing Pancreatic Cancer Risk Associated with Dipeptidyl Peptidase 4 Inhibitors: Data Mining of FDA Adverse Event Reporting System (FAERS), Journal of Pharmacovigilance, Volume 1, Issue 3, , ISSN: 2329-6887

[62] Subrata Kumar Mandal, 2017, Performance Analysis Of Data Mining Algorithms For Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree, International Journal Of Engineering And Computer Science, Vol. 6 Issue 2, 20388-20391, ISSN: 2319-7242

[63] K.-S. Leung, et.all., 2011, Data Mining on DNA Sequences of Hepatitis B Virus, IEEE/ACM Transactions on Computational Biology and Bioinformatics, VOL. 8, NO. 2, 428-440

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

[65] Ada, Rajneet Kaur, 2013, A Study of Detection of Lung Cancer Using Data Mining Classification Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 3, 131-134, ISSN: 2277 128X

[66] Mohammed Abdul Khaleel, et.all., 2013, A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 8, 149-153, ISSN: 2277 128X

[67] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, 2013, Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors, Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013), , 1227 - 1231

[68] Maciej Zieba, et.all., 2013, Boosted SVM for extracting rules from imbalanced data inapplication to prediction of the post-operative life expectancyin the lung cancer patients, Applied Soft Computing, 14 (2014), 99-108

[69] P. A. Idowu, et.all., 2015, Breast Cancer Risk Prediction Using Data Mining Classification Techniques, Transactions on Networks and Communications, Vol 3, No 2, pp.1-11,ISSN:2054-7420

[70] Ruijuan Hu, 2010, Medical Data Mining Based on Association Rules, Computer and Information Science, Vol. 3, No. 4, 104-108, ISSN 1913-8989, E-ISSN 1913-8997

[71] Hossein Ghayoumi Zadeh, et.all., 2012, Diagnosing Breast Cancer with the Aid of Fuzzy Logic Based on Data Mining of a Genetic Algorithm in Infrared Images, Middle East Journal of Cancer, Vol. 3(4), 119-129

[72] A.Priyanga, S.Prakasam, 2013, Effectiveness of Data Mining - based Cancer Prediction System (DMBCPS), International Journal of Computer Applications, Vol. 83 – No 10

[73] M. Durairaj, V. Ranjani, 2013, Data Mining Applications In Healthcare Sector: A Study, International Journal of Scientific & Technology Research, Vol. 2 (10), 29-35, ISSN 2277-8616

[74] S. Chidambaranathan, 2016, Breast Cancer Diagnosis Based on Feature Extraction by Hybrid of K-Means and Extreme Learning Machine Algorithms, ARPN Journal of Engineering and Applied Sciences, VOL. 11, NO. 7, 4581-4586, ISSN 1819-6608

[75] P. Revesz, C. Assi, 2013, Data mining the functional characterizations of proteins to predict their cancer-relatedness, Int'l Journal of Biology and Biomedical Engineering, Issue 1,Vol 7.

[76] M. Jajroudi, et.all., 2014, Prediction of Survival in Thyroid Cancer Using Data Mining Technique, Technology in Cancer Research and Treatment, Volume 13, Number 4, 353-359, ISSN 1533-0346"

[77] https://en.wikipedia.org/wiki/Examples_of_data_mining#Medical_data_mining