Review on Data Mining Techniques for Prediction of Chronic Kidney Disease

Pallavi Sharma[#1], Gurmanik Kaur[*2]
[#]*Ph.D. Scholar & Department of ECE& SBBSU*
[*]*Assistant Professor & Department of EE & SBBSU*
Padhiana, Punjab, India

Abstract

Data mining is the process of extracting hidden interesting patterns from massive database. Medical domain contains heterogeneous data that can be mined properly to provide a variety of useful information for the physicians to detect a disease, predict the survivability of the patients after disease, severity of diseases etc. The main aim of this paper is to analyse the application of data mining in medical domain for prediction of chronic kidney disease. In the healthcare area chronic kidney disease can be very well predicted using data mining techniques.

Keywords — *Data mining, medical data, chronic kidney disease, disease prediction.*

I. INTRODUCTION

Data mining deals with the extraction of useful information from huge amounts of data. Many other terms are being used to understand data mining, such as mining of knowledge from databases, knowledge extraction, data analysis, and data archaeology. Basically, data mining is a crucial step in the process of knowledge discovery in databases (KDD) [1].

Medical data mining is one of the applications of data mining that plays a vital role in human's lives. It has become famous in healthcare management to predict, to detect or to find the hidden patterns and information in health data. It provides a new platform for emerging and discovering the different direction in healthcare organization. In the healthcare management, various categories of data are available such as graphics, signals, web and text etc. Data mining tools and techniques can be applied to processed data for aiding health professionals in appropriate decision making and thus improving the management of patients [2-4].

Chronic kidney disease is increasing its grip over the world at an alarming rate, it's a very dangerous and life threatening disease. It is the condition in which the kidney gets damaged and toxic wastes are not filtered in the body [5, 6]. In the 2015 Global Burden of Disease Study, kidney disease was the 12th most common cause of death, accounting for 1.1 million deaths worldwide. Overall CKD mortality has increased by 31.7% over the last 10 years, making it one of the fastest rising major causes of death, alongside diabetes and dementia. In the same study, CKD ranked as the 17th leading cause of global years lost of life, an 18.4% increase since 2005, and the third largest increase of any major cause of death. This is in stark contrast to other non-communicable diseases, for example cardiovascular disease and chronic obstructive pulmonary disease, where global years lost of life fell during the same time period (-10.2% and -3.0%, respectively). Hence, an accurate diagnosis is immediately required so that precautions can be taken beforehand in time that may prove beneficial to patients suffering from such diseases [7].

II. LITERATURE REVIEW

Sirage Zeynu and Shruti Patil [8] presented machine learning and data mining techniques for prediction of disease to diagnose CKD. To predict CKD, construct two important models: feature selection method and ensemble model. It has been observed that ensemble model achieved better accuracy (99%).

Rajni Garg and Vikas Mongia [9] five data mining algorithms were implemented on chronic kidney disease dataset. The accuracy of Naïve Bayes classification algorithm is being measured highest. The results showed that after feature reduction most of the algorithms improved their accuracy (98.75%) and computation.

S. Gopika and Dr.M.Vanitha [10] proposed new algorithm Improved Hybrid Fuzzy C-Means (IHFCM) which is an improvisation of FCM with Euclidean distances for prediction of kidney disease in patients. Results of clustering shows that FCM based clustering algorithms achieve greater accuracy (96%) than most existing algorithms.

Basma Boukenze et al. [11] discussed various machine learning algorithms to predict kidney disease. In the end, those algorithms were compared and evaluated the most efficient one on the basis of multiple criteria. It has been achieved that C4.5 has the best rates.

Pavithra N et al. [12] proposed a symbolic fuzzy C-means clustering algorithm with fuzzy data in the structure. The system was used to predict and diagnose patients with renal dysfunction. The FCM clustering algorithm was applied to the location of the disease in kidney disease patient files.

Veenita Kunwar et al. [13] discussed prediction and diagnosis on Chronic Kidney Disease using data

mining classifiers such as Artificial Neural Network (ANN) and Naive Bayes. The tool named as Rapid miner is used to compare the performances of both mining classifiers. The results obtained that Naive Bayes shows better accuracy (100%).

Basma Boukenze et al. [14] presented three learning algorithms on a set of medical data and predicted kidney disease by using multiple machine learning algorithms that are Support Vector Machine (SVM), Decision Tree (C4.5), and Bayesian Network (BN) and chose the most efficient one.

Neha Sharma and Rohit Kumar Verma [15] detected and predicted kidney diseases as a prelude to proper treatment to patients. The system was used for detection in patients with kidney disease and the results of their IF-THEN rules predicted the presence of a disease. Generally results based on mathematics tend to have higher accuracy.

Veenita Kunwar et al. [16] proposed naive Bayesian classification and artificial neural network (ANN) for prediction of chronic kidney disease. The results obtained that naive Bayesian produced accurate results than artificial neural networks. It was also observed that classification algorithms were widely used for investigation and identification of CKDs.

Parul Sinha and Poonam Sinha [17] proposed a new decision support system to predict chronic kidney disease and compared the performance of SVM and k-NN classifier on the basis of parameters. It has been showed that k-NN have better performance (78.75% of accuracy).

Swathi Baby P et al. [18] demonstrated that data mining methods could be effectively used in medical field. K-means algorithm can find out number of clusters in large data sets. Their study analysed tree AD, J48, star K, Bayesian sensible, random forest and tree - based ADT naive Bayesian on J48 Kidney Disease Data Set and noted that the techniques offer statistical analysis on the use of algorithms to predict kidney diseases in patients.

Vijayarani and Dhayanand [19] discussed a classification process that is used to classify four types of kidney diseases. Comparisons of Support Vector Machine (SVM) and Naïve Bayes classification algorithms are prepared on the basis of performance factors, classification, accuracy and execution time. As a result, the SVM achieves better classification performance. Hence it is considered as the best classifier when compared with Naïve Bayes classifier algorithm.

Lambodar Jena et al. [20] implemented algorithms using WEKA data mining tool to analyse accuracy and compared. Multilayer perceptron (MLP) algorithm achieved better classification accuracy (99.75%) and prediction performance to predict chronic kidney disease (CKD).

Vijayarani and Dhayanand [21] presented a work to predict kidney disease by classifying four types of kidney diseases: Acute Nephritic Syndrome, Chronic Kidney disease, Acute Renal Failure and Chronic Glomerulonephritis, using Support Vector Machine (SVM) and Artificial Neural Network (ANN), and then comparing the performance of those two algorithms on the basis of accuracy and execution time. The results prove that the performance of the ANN is better than the SVM algorithm.

Shakil Ahmed et al. [22] presented a diagnosis system based on fuzzy logic to report the healthiness of a patient's kidney. Using the selected input variables, fuzzy expert system achieved accurate (86.7%) results.

Tommaso Di Noia et al. [23] described a decision support system exploiting an ensemble of ANNs to predict end stage kidney disease (ESKD). The model has been built by training a set of neural networks via a dataset containing patient's information collected. The developed tool has been made available both as an online Web application and as an Android mobile app.

Lakshmi K.R. et al. [24] three data mining techniques (Artificial Neural Networks, Decision tree and Logical Regression) are used to elicit knowledge about the interaction between variables and patient survival. A performance comparison of three data mining techniques is employed for extracting knowledge from data collected at different dialysis sites. ANN is suggested for Kidney dialysis to get better accuracy and performance.

Ruey Kei Chiu and Renee Yu-Jing [25] developed a best-fitting neural network model which can be employed to detect various severity levels of CKD based on the two different sets of influence factors. Three fundamental neural network models selected for detection and comparison include back-propagation network (BPN), generalized feed forward neural networks (GRNN), and modular neural network (MNN). The comparison among network models was based on the CKD detection performance measured by the computation accuracy, sensitivity and specificity in model training and testing.

Julia Hippisley-Cox and Carol Coupland [26] designed two new risk prediction algorithms to predict risk of moderate severe CKD and end stage kidney failure. These new algorithms have the potential to identify high risk patients who might benefit from more detailed assessment and closer monitoring to reduce their risk.

Susan Snyder, M.D. and Bernadette Pendergraph, M.D. [27] suggested that progression to kidney failure can be delayed or prevented by controlling blood sugar levels and blood pressure and by treating proteinuria. Chronic kidney disease affects approximately 19 million adult Americans, and its incidence is increasing rapidly. Diabetes and hypertension are the primary causes in most cases of chronic kidney disease.

III.CONCLUSION

As conclusion, the application of data mining techniques for predictive analysis is very important in

the health care field because it gives us the power to face diseases earlier and therefore save people's lives through the anticipation of cure. This paper was aimed to analyse the application of data mining in CKD prediction. It is observed data mining provides good results in CKD diagnosis when appropriate tools and techniques applied. Anticipating diseases still remains a major challenge in medical field and pushes us to increase our efforts in developing more data mining algorithms to exploit information intelligently and extract the best knowledge from it.

ACKNOWLEGDEMENT

The authors would like to express their gratitude to the management of Sant Baba Bhag Singh University, Padhiana, and Punjab, India for making this study possible. Special thanks to the anonymous reviewers for their widespread and helpful feedback.

REFERENCES

[1] Neelamadhab Padhy, Dr. Pragnyaban Mishra and Rasmita Panigrahi (2012), "The Survey of Data Mining Applications and Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2(3), pp 43-58.

[2] J.C.Prather, D. F. Lobach, L. K. Goodwin, J. W. Hales, M. L. Hage, and W. E. Hammond (1997), "Medical data mining: knowledge discovery in a clinical data warehouse", Proceedings of the AMIA Annual Fall Symposium, pp 101–105.

[3] Neesha Jothi, Nur Aini Abdul Rashid, Wahidah Husain (2015), "Data mining in Healthcare: A Review", Procedia Computer Science, 72, pp 306-313.

[4] Subhash Chandra Pandey (2016), "Data mining techniques for medical data: A review", International Conference on Signal Processing, Communication, Power and Embedded System, Paralakhemundi, India, pp 972-982.

[5] Levey A. S. and Coresh, J. (2012), "Chronic kidney disease", Lancet, 379, pp 165–180.

[6] Wang H, Naghavi M, Allen C (2016), "GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015", Lancet, 388, pp 1459–544.

[7] Sirage Zeynu and Shruti Patil (2018), "Prediction of Chronic Kidney Disease using Data Mining Feature Selection and Ensemble Method", International Journal of Data Mining in Genomics & Proteomics, 9(1), pp 1-9

[8] Rajni Garg and Vikas Mongia (2018), "A Comparative Study of Different Classification Algorithms on Kidney Disease Prediction", International Journal for Research in Applied Science & Engineering Technology, 6(2), pp 741-746.

[9] S.Gopika and Dr.M.Vanitha (2017), "Efficiency of Data Mining Techniques for Predicting Kidney Disease", International Journal of Engineering and Technology, 9(5), pp 3586-3591

[10] Basma Boukenze, Abdelkrim Haqiq and Hajar Mousannif (2017), "Predicting Chronic Kidney Failure Disease using Data Mining Techniques", Advances in Ubiquitous Networking, Springer, pp 701-712.

[11] Pavithra N and Dr. R. Shanmugavadivu (2017), "Efficient Early Risk Factor Analysis of Kidney Disorder using Data mining Technique", International Journal of Innovative Research in Computer and Communication Engineering, 5(2), pp 1690-1698.

[12] Veenita Kunwar, Khushboo Chandel, A.Sai Sabitha and Abhay Bansal (2016), "Chronic Kidney Disease Analysis Using Data Mining Classification Techniques", 6th International Conference on Cloud System and Big Data Engineering, pp 300-305

[13] Basma Boukenze, Hajar Mousannif and Abdelkrim Haqiq (2016), "Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease", International Journal of Database Management Systems, 8(3), pp 1-9.

[14] Neha Sharma and Rohit Kumar Verma (2016), "Prediction of Kidney Disease by using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, 6(9), pp 66-70

[15] Veenita Kunwar, Khushboo Chandel, A. Sai Sabitha, and Abhay Bansal (2016), "Chronic Kidney Disease Analysis using Data Mining Classification Techniques", IEEE on Cloud System and big data Engineering , pp 300-305.

[16] Parul Sinha and Poonam Sinha (2015), "Comparative study of Chronic kidney disease prediction using KNN and SVM", International Journal of Engineering Research & Technology, 4(12), pp 608-612

[17] Swathi Baby P and Panduranga Vital T (2015), "Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms", International Journal of Engineering Research & Technology, 4(7), pp 206-210.

[18] Vijayarani, S. and Dhayanand, S. (2015), "Data mining classification algorithms for kidney disease prediction", Int. J. Cybern. Inf. (IJCI), 4(4), pp 13–25.

[19] Lambodar Jena and Narendra K. Kamila (2015), "Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease", International Journal of Emerging Research in Management & Technology, 4, (11), pp 110-118

[20] Vijayarani, S., Dhayanand, S. (2015), "Kidney disease prediction using SVM and ANN algorithms", International Journal of Computing and Business Research, 6(2).

[21] Shakil Ahmed, Md. Tanzir Kabir, Navid Tanzeem Mahmood and Rashedur M Rahman (2014), "Diagnosis of kidney disease using fuzzy expert system", IEEE conference on software, knowledge, Information management and Applications.

[22] Tommaso Di Noia, Vito Claudio Ostuni, Francesco Pesce, Giulio Binetti, David Naso, Francesco Paolo Schena and Eugenio Di Sciascio (2013), "An end stage kidney disease predictor based on an artificial neural networks ensemble", Elsevier Publication, Expert Systems with Applications , 4438-4445.

[23] Lakshmi, K.R., Nagesh, Y., Veera Krishna, M. (2013), "Performance comparison of three data mining techniques for predicting kidney dialysis survivability", International Journal of Advances in Engineering & Technology, 7(1), 242–254.

[24] Ruey Kei Chiu and Renee Yu-Jing (2011)," Constructing Models for Chronic Kidney Disease Detection and Risk Estimation", Proceedings of 22nd IEEE International Symposium on Intelligent Control, Singapore, 166-171.

[25] Julia Hippisley-Cox* and Carol Coupland (2010), "Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the Qkidney scores", BMC family practice, 1-13.

[26] Susan Snyder, M.D. and Bernadette Pendergraph, M.D. (2005), "Detection and Evaluation of Chronic Kidney Disease", American Family Physician, 72(9), 1723-1732.