

A Survey on Chronic Kidney Disease Detection Using Novel Methods

Jayalakshmi V¹, Lipsa Nayak² and ³K.Dharmarajan

¹Department of Computer Applications, ²Research Scholar, ³Department of Information Technology

School of computing sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India

¹Jayasekar1996@yahoo.co.in, ²info.lipsa@gmail.com and ³dharmak07@gmail.com

Abstract— Chronic kidney disease (CKD) also called as chronic renal failure (CRF), is a medical term that refers to decreased renal function due to damaged—at risk through low, moderate, and severe chronic kidney failure. In the United States, the number of incidences related to kidney failure is increasing. Medical field has heterogeneous data like, figures, text and images in order to provide useful medical information. The information obtained from these medical records helps the doctors to predict or detect the pattern of the disease. This information in turn can be used to help in the survival of the predicted patients. The severity of the disease can also be predicted using this medical information. The paper aims to analyze and understand some of the techniques that are used in predicting kidney disease. Chronic Kidney Disease (CKD) classification is performed using classification techniques like Artificial Neural Network(ANN), Naive Bayes, SVM, MBPN, ada boost classifier, LDA, KNN, etc. Medical industry mining data is used to identify the data sets in kidney disease using the MATLAB tool.

Keywords: Classification, Kidney Disease, Prediction, KNN, ANN.

I. INTRODUCTION

Kidney disease has become a kind of a common disease with serious problems. Slow kidney damage occurs usually in case of diabetes, high blood pressure, etc. This is termed as chronic kidney disease. When a person is sick it can so happen that the kidney gets injured with sudden changes in the cause, or by consuming certain medications, and this is called as acute kidney injury. This usually takes place with a person having normal kidneys or to a

person who has kidney problems. The main factors identified as cause for the disease are: High blood pressure, Diabetes, Cardiovascular (heart and blood vessel) disease, hereditary records on kidney failure. Chronic kidney disease (CKD) is a medical condition where the kidneys are damaged and blood cannot be filtered. In the end-stage of the disease the renal

disease (CKD), renal function is severely damaged. Slow kidney damage occurs usually in case of diabetes, high blood pressure, etc. This is termed as chronic kidney disease. When a person is sick it can so happen that the kidney gets injured with sudden changes in the cause, or by consuming certain medications, and this is called as acute kidney injury. This usually takes place with a person having normal kidneys or to a person who has kidney problems. Around 10% of the population in the world suffers chronic kidney disease (CKD), leading to millions of deaths every year. Chronic kidney disease ranks number 27 in 1990 as world's leading cause of death. In the year 2010 it ranks in the 18th place. Some of the risk factors that can be defined are obesity, smoking, hypercholesterolemia, Obstructive renal disease, diabetes (type 1 and 2), obstruction of the bladder by BPH, cirrhosis, atherosclerosis and hepatic failure, narrowing of the arterial supply of the kidneys, bladder cancer, kidney cancer, kidney stones and kidney infections.

II. DATA MINING TECHNIQUES

A. Naive Bayes

A Naive Bayes classifier is a probabilistic classifier that works on the concept of Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. It can also be defined as "independent feature model". This approach works well even with restricted individuality assumption infrequently clutches true in real world applications. Hence it is termed as Naive and yet the algorithm inclines to perform well and can learn rapidly in various supervised classification problems.

B. Support Vector Machine (SVM)

Support vector machine is a machine learning technique that works on statistical learning theory. It draws a discrete hyperplane in the descriptor space of the training data and the

compounds are classified based on the side of hyperplane situated.

C. K Nearest Neighbor

It is a distance based algorithm that is applicable when all the attribute values are continuous. It can be modified according to categorical attributes. This helps to estimate the classification of an unnoticed instance through classification of the instance or instances that are closest to it. Even if there are more instances in the training set it applies the same principle to classify the k nearest neighbour or most nearest one known as k- Nearest Neighbour.

D. Decision Tree Based Algorithm

This approach works well in cases where a tree is constructed to model the classification process when the classification becomes complicated. This tree is applied on each tuple in the database to generate classification.

E. K-Means Clustering

This is a centroid-based approach where the number of partitions k is considered as the input and creates k clusters of the input database consist of n objects or records by optimizing the rule of clustering. The end clusters with high similarity are kept as intra-cluster and the low similarity are kept as inter-cluster. A random set of k objects are chosen as the cluster centers in order to compute the mean value that represents the cluster mean or centre.

F. Neural Networks

A neural network creates an interconnected group of artificial neurons in order to process information and computation is done using associated weights. These weights are updated or adjusted during prediction of input records.

III. LITERATURE REVIEW

Kübra Eroğlu et.al 2016[1] proposed this study to make a comparison of the performance of the different classification methods and ensemble algorithms that are used for detection of chronic kidney disease. This study involves six different basic classifier namely: k nearest neighbor (KNN), naive bayes, support vector machines (SVM), random trees, J48, decision tables and three different ensemble algorithm namely: bagging, AdaBoost, random subspace are used in the study. Classification results were derived using three different performance evaluation criteria (kappa, accuracy and the area under the ROC curve (AUC)). The result

says that J48 basis algorithm for use with random subspace and bagging ensemble algorithms and random tree basis algorithm for use with bagging ensemble algorithm has provided 100% classification success.

Veenita Kunwar et.al 2016[2] Data mining is vastly being investigated in diagnostic results. Huge amount of un-mined data derived from healthcare industry is used to discover sensible information in order to diagnose and help in decision making. Data mining helps in extracting hidden information from huge dataset. It also helps in categorizing data, validate them and derive unique patterns in them. Several data mining techniques like classification, clustering, regression, association analysis, etc are used in health care data mining. The objective of this paper is for the prediction of Chronic Kidney Disease (CKD) through classification techniques like Artificial Neural Network (ANN) and Naive Bayes. Rapidminer tool was used for experimental purpose and the results show that Naive Bayes obtains more accurate results than the Artificial Neural Network.

In K.R.Lakshmi et.al 2014 [3] proposed a study that records the performance evaluation of three data mining techniques that are used for predicting kidney dialysis survivability. This research involves various data mining techniques like Decision tree, Artificial Neural Networks and Logical Regression to extract knowledge based on the interaction between these variables and predict patient survival. A performance comparison based on these three data mining techniques is performed using the data collected at different dialysis sites. The outcomes are reported and finally it is concluded that ANN is advisable to be used for Kidney dialysis in order to derive better results with performance and accuracy.

Naganna Chetty et.al 2015[4], Chronic Kidney Disease (CKD) has become a common problem to the public in recent days. CKD refers to kidney damage and is measured in terms of GFR (Glomerular Filtration Rate). Researchers from health care as well as the academicians are working on the CKD problem and design an efficient model in order to predict and classify the CKD patient in the initial stage of CKD. This can help health care persons to take the necessary treatment to prevent or cure CKD. There are several classification models built using different classification algorithms. Some of the methods are; bestfirst search method and Wrappersubset attribute evaluator to predict and classify the CKD and non CKD patients. These models are implemented on the CKD dataset that was downloaded from the UCI repository. These models show better performance in classifying CKD and non CKD cases.

Zeinab Sedighi et.al 2015[5] Chronic_Kidney_Disease is a UCI data sets. It contains 400 instances that include some missing value. This missing data can really cause severe problems for analyzing data set. But there are several ways to handle this. We have the option to either estimate, discard or replace their values. Since we lose power of analysis in case of discarding the missing values, so we use imputation approaches here. An imputation procedure fills the missing values with the estimated ones. This paper uses knearest neighbour measured in terms of GFR (Glomerular Filtration Rate). Researchers from health care as well as the academicians are working on the CKD problem and design an efficient model in order to predict and classify the CKD patient in the initial stage of CKD. This can help health care persons to take the necessary treatment to prevent or cure CKD. There are several classification models built using different classification algorithms. Some of the methods are; bestfirst search method and Wrappersubset attribute evaluator to predict and classify the CKD and non CKD patients. These models are implemented on the CKD dataset that was downloaded from the UCI repository. These models show better performance in classifying CKD and non CKD cases model to estimate the missing data. KNN finds the most similar or approximate instances by searching through all the data set in order to find the best substitutes for missing values.

Ani R et.al 2016[6] There are different stages of Chronic Renal Failure (CRF) that starts with loss of renal functions and then gradually leading to complete failure of all kidney functions. It is fatal at its end stage if the kidney is not replaced with the artificial filtering mechanism. Hence it is important to predict the disease at the early stage in order to save the human life. The objective of this research is to develop and create a clinical decision support system that uses machine learning techniques [17]. Classification techniques like probability based Naive Bayes, neural network based back propagation (BPN), LDA classifier, tree based decision tree, lazy learner K Nearest Neighbor (KNN) and Random subspace classification algorithms are analyzed. The accuracy of each algorithm found is 78%, 81.5%, 90%, 93% 76% and 94% respectively on a dataset collected from UCI repository that holds 25 attributes along with 400 instances. The result shows that the algorithm with the best results is implemented in the Clinical Decision Support System.

Six algorithms were considered for the classification approaches and a comparison was made to select the best among them in predicting the Chronic Renal Failure. The random subspace

classification showed a better accuracy in the prediction of CRF. 10 fold cross validation method is considered in this study. The random subspace classification method with KNN as base classifier demonstrated the best results so far.

Miguel A et.al 2014[7], demonstrated the possibility of a distributed approach for the management of alarms that are related to the monitoring of CKD patients who are part of the eNefro project. In order to satisfy this purpose and from the methodological point of view, these alarms and prioritization issues are concerned methodologically according to the definitions provided by the ISO IEC/CD 60601. Apart from the technological perspective, this approach is based on the publisher/subscriber model and is developed according to the DDS specification. This double perspective approach makes it possible for a proper management of the HD alarms described in eNefro. The proof of concept proposed states the most common actors (nephrologist, carer/nurse and supplier) and scenarios (hospital, home and day hospital) for HD that pertains to the so-called DDS domain. This approach demonstrated the potential of DDS as a communication infrastructure for the personalization and prioritization of alarms for the purpose of optimal delivery of information, in situations like eNefro.

Merve Doğruyol Başar et.al 2016[8], Chronic kidney disease is predictable using several automatic diagnosis systems. This study is based on the diagnosis of chronic kidney disease with the help of Adaboost ensemble learning algorithm. Decision tree based classifiers are used in this diagnosis approach. The classification performance is evaluated with the attributes such as: mean absolute error (MAE), kappa, root mean squared error (RMSE) and area under curve (AUC) criteria. The result shows that adaboost ensemble learning algorithm provides better classification performance than individual classification.

Anu Chaudhary et al 2014 [9] proposed a prediction system using k-means algorithm and A-priori for kidney failure and heart disease prediction. The survey used A-priori and k-mean algorithm to predict kidney failure patient considering 42 attributes. Data was analyzed using machine learning tools like distribution and attribute statistics, followed by A-priori and k-means algorithms. Data was evaluated using Receiver Operating Characteristic (ROC) plot and calibration plots.

Lakshmi.K.R et al 2014 [10] analyzed Decision tree, Artificial Neural Networks and Logical Regression supervised machine learning algorithms. These algorithms are used for Kidney dialysis. For

classification process they used a data mining algorithms. The 10 fold cross validation was implemented to evaluate the classified data proceeded by the comparison of those data. The final results show that ANN performed better than the Logical Regression algorithms and the Decision tree.

Jaime Arturo Jojoa et.al 2017[11] Presents a study that combines the factor that are used to identify the presence of chronic kidney disease (CKD). Two major factors namely prognosis, is a new way to approach the study in a much more comprehensive concept of the disease, which, by to have quantifiable prognostic risks factors, it helps in establishing a better and objective therapeutic strategy.

Daniel E et.al 2007[12] the objective of this study was to review the burden of CKD and its co morbidities on physicians, patients and payers so as to discuss the potential benefits to individual patients and society of identifying and treating earlier stages of CKD.

S.Vijayarani et al. 2015 [13] The study was to predict kidney disease using vector-based vector machines (SVM) and artificial neural networks (ANN). The aim of this study is to compare the performance of the two algorithms based on accuracy and run-time. From the experimental results it shows that the yield of RNA is better than other algorithms.

S. Ramya et al.2016[14] This work was proposed to reduce the diagnostic time and also to improve the accuracy of the diagnosis using classification algorithm. The proposed work considers the classification of the different stages of chronic kidney disease severity. This work was carried out in different algorithms as back propagation neural networks, random forests and radial basis functions. Experimental results prove that the radial basis function of the algorithm gives the results of other sorting algorithms and produces an accuracy of 85.3%.

Basma Boukenze et al. 2016[15] The development of large data sets in the health system is outlined and used in a collection of medical data using three learning algorithms. The aim of this work is to predict kidney disease using multiple machine learning algorithms, C4.5, SVM and Bayesian Networks (BNs), and selected efficient.

Milandeep Arora et al 2016.[16] presented a knowledge flow window tool Weka's data mining, it not only gives the statistical evaluation of the results but also provides a data flow diagram. It is finally concluded that J48 is the best algorithm to achieve accurate and early detection of chronic kidney

disease in the three algorithms used in the early stages.

IV. CONCLUSION

The aim of this study was to observe and analyze the results obtained by applying these algorithms in the medical field in order to predict disease. Data mining definitely yields good results when applied appropriately along with tools and techniques that can enhance its performance in diagnosing the disease. Kidney disease is predictable using multiple classifications. The level of chronic kidney disease can also be predicted using the algorithmic classifier. Though this study involves a number of parameters, KD is limited. But some new results were derived unless each finding of a relationship between the other studies in the relevant factor prior knowledge KD was confirmed. In future this algorithm can be designed to take symptoms as input in order to predict the disease based on old patients record. Hence in future this algorithm can be modified by adding the SMS/Email module where doctors can receives an SMS or Email regarding the Id and password of the patient. We can also improve it further by adding the query module to the application where doctor and admin of the application can interact with each other. Various feature selection methods can also be implemented to further enhance the classification results.

References

- [1] Kübra Eroğlu, Tuğba Palabaş, "The Impact on the Classification Performance of the Combined Use of Different Classification Methods and Different Ensemble Algorithms in Chronic Kidney Disease Detection".
- [2] Veenita Kunwar 1 Khushboo Chandel2 A. Sai Sabitha3 Abhay Bansal4" CHRONIC KIDNEY DISEASE ANALYSIS USING DATA MINING CLASSIFICATION TECHNIQUES", 978-1-4673-8203-8/16/\$31.00 c 2016 IEEE.
- [3] K.R.Lakshmi1, Y.Nagesh2 and M.VeeraKrishna3, "Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability", International Journal of Advances in Engineering & Technology, Mar. 2014, Vol. 7, Issue 1, pp. 242-254
- [4] Naganna Chetty, Kunwar Singh Vaisla, Sithu D Sudarsan, "Role of Attributes Selection in Classification of Chronic Kidney Disease Patients", 978-1-4673-9354-6/15/\$31.00 ©2015 IEEE.
- [5] Zeinab Sedighi, Hossein Ebrahimpour-Komleh, Seyed Jalaleddin Mousavirad, "Feature selection effects on kidney disease analysis", 978-1-4673-9762-9/15/\$31.00 ©2015 IEEE.
- [6] Ani R1, Greeshma Sasi1, Resmi Sankar U1, O.S Deepa, "Decision Support system for diagnosis and prediction of Chronic Renal Failure using Random Subspace Classification", 2016 Intl.

Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.

[7] Miguel A. Estudillo-Valderrama, Alejandro Talaminos-Barroso, Laura M. Roa, "A Distributed Approach to Alarm Management in Chronic Kidney Disease", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 18, NO. 6, NOVEMBER 2014.

[8] Merve Doğruyol Başar1 , Pelin Sarı1 , Niyazi Kılıç2 , Aydın Akan, "Detection of Chronic Kidney Disease by Using Adaboost Ensemble Learning Approach", 978-1-5090-1679-2/16/\$31.00 ©2016 IEEE.

[9] Anu Chaudhary, Puneet Garg,(2014) Detecting and Diagnosing a Disease by Patient Monitoring System, International Journal of Mechanical Engineering And Information Technology, Vol. 2 Issue 6 //June //Page No: 493-499.

[10] Lakshmi. K.R, Nagesh. Y and VeeraKrishna. M, (2014) Performance Comparison Of Three Data Mining Techniques For Predicting Kidney Dialysis Survivability, International Journal of Advances in Engineering & Technology, Mar., Vol. 7, Issue 1, pg no. 242-254.

[11] Jaime Arturo Jojoa1 *, Carolina Enríquez Rivera2 , Edison Muñoz Delgado3 , Harold Mauricio Casas4 , Giovana Marcella Rosas5 , Carol Yovanna Rosero6 and Franco Andrés Montenegro, "New ABC Chronic Kidney Disease Classification", International Journal of Nephrology and Kidney Failure, 23 Sep 2017.

[12] Daniel E. Weiner, MD, MS, "Causes and Consequences of Chronic Kidney Disease: Implications for Managed Health Care", Vol. 13, No. 3 April 2007 JMCP.

[13] . S.Vijayarani, S.Dhayanand "Kidney disease Prediction Using SVM and ANN Algorithms", International Journal of Computing and Business Research (IJCBR) ISSN (Online): 2229-6166 Vol. 6, Issue 2, March 2015.

[14]. S. Ramya, N. Radha "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms", International Journal of Innovative Research in Computer and Communication Engineering. Vol. 4, Issue 1, January 2016.

[15]. Basma Boukenze, Hajar Mousannif and Abdelkrim Haqiq "Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease", International Journal of Database Management Systems (IJDBMS) Vol.8, No.3, June 2016.

[16]. Milandeep Arora and Er. Ajay Sharma "Chronic Kidney Disease Detection by Analyzing Medical Datasets in Weka", International Journal of Computer Application (2250-1797) Volume 6– No.4, July- August 2016.

[17] Dharmarajan, K., and M. A. Dorairangaswamy. "Discovering Student E-Learning Preferred Navigation Paths Using Selection Page and Time Preference Algorithm." International Journal of Emerging Technologies in Learning (IJET) 12.10 (2017): 202-211.

