ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management, Vol-4, Issue-6, 2020.

A Survey on Kidney Disease Risk Prediction Using Machine Learning Algorithm

Anita Kumari[1], Chinmay Bhatt[2], Varsha Namdeo[3]

Department of Computer Science & Engineering[1,2,3]

SRK University, Bhopal, (M.P.), India[1,2,3]

ABSTRACT

The According the 2010 global burden of disease study, Chronic Kidney Diseases (CKD) was ranked 18th in the list of causes of total no. of deaths worldwide. 10% of the population worldwide is affected by CKD. The prediction of CKD can become a boon for the population to predict the health. Various method and techniques are undergoing the research phase for developing the most accurate CKD prediction system. Using Machine Learning techniques is the most promising one in this area due to its computing function and Machine Learning rules. Existing Systems are working well in predicting the accurate result but still more attributes of data and complicity of health parameter make the root layer for the innovation of new approaches. This study focuses on a novel approach for improving the prediction of CKD. In recent time Neural network system has discovered its use in disease diagnoses, which is depended upon prediction from symptoms data set. Chronic kidney disease detection system using neural network is shown here. This system of neural network accepts disease-symptoms as input and it is trained according to various training algorithms. After neural network is trained using back propagation algorithms, this trained neural network system is used for detection of kidney disease in the human body.

Keywords: Machine learning, Chronic kidney disease, Classification, K-Nearest Neighbor, Support vector machine**.**

Introduction

Chronic kidney disease is a worldwide public health problem with an increasing incidence, prevalence, and high cost. Approximately 2.5-11.2% of the adult population across Europe, Asia, North America, and Australia are reported to have chronic kidney disease [1], where in the USA alone it has affected more than 27 million individuals [2]. According to The National Kidney Foundation about 59% of all American are at risk of developing kidney disease in their lifetime [3].
The increase of CKD is partially explained by the increasing prevalence of diabetes mellitus and hypertension which are the leading risk factors for CKD. CKD promotes hypertension and dyslipidemia, which, in turn, can contribute to the progression of renal failure.

Recent studies suggest that some of these adverse outcomes can be prevented or delayed by early detection and treatment [4]. Awareness of CKD among patients is gradually increasing, but still low. According to the 2003-2004 National Health and Nutrition Examination Survey, less than 5 percent of patients with stage 1 or 2 CKD and less than 10 percent with stage 3 reported having been diagnosed with CKD; only 45 percent of patients with stage 4 were aware of their condition [5].

Since there is a relatively small number of practicing nephrologists, nephrologists cannot exclusively manage all patients with CKD. The

burden of CKD management thus falls largely on primary care providers (PCPs). A recent study [6] has shown that awareness of CKD by all types of PCPs is unacceptably low and knowledge of CKD management is particularly poor among family practitioners, especially among those with more than 10 years in clinical practice and who spend more than 50% of their time practicing clinical medicine. Hence an accurate, convenient, and automated CKD detection method is important for clinical practice.

In this paper we reviewed machine leaning solution to detect CKD and explore 24 parameters related to kidney disease. The various dataset used for evaluation and suffers from noisy and missing data. We need a robust classifier that can deal with these issues. Hence, we evaluate methods with different classifiers.

II Related Work

In 2015, Konstantina Kourou et.al [7] proposed a study of Machine learning applications in cancer prognosis and prediction. In this paper, they have presented a review of various recent ML approaches that are applied for the prediction of cancer detection. Here they have presented review of newly published content for the work done so far in cancer detection.

In 2015 P.Swathi Baby et. al [8] proposed a project to diagnosis and prediction system based on predictive mining. Here kidney disease data set is used and analysed using Weka and Orange software. Here the Machine learning algorithms such as AD Trees, J48, K star , Naïve Bayes, Random forest are used for the performance study of each algorithm which gives the Statistical analysis and predicting kidney diseases using the algorithms. Their observation shows that the best algorithms K-Star and Random Forest for the used Dataset ,where Build the models are less time(0 sec and 0.6 sec) and the ROC values are 1.

In 2014 K.R.Lakshmi et.al [9] proposed performance evaluation of three data mining techniques for predicting kidney dialysis survivability. In this research, various data mining techniques (Artificial Neural Networks, Decision tree and Logical Regression) are used to extract knowledge about the interaction between these variables and patient survival. A performance comparison of three data mining techniques is applied for extracting knowledge. The concepts introduced in this research have been engaged and tested using a data collected at different dialysis sites. The outcomes are reported. Finally, ANN is suggested for Kidney dialysis to get better results with accuracy and performance.

Shital Shah et. al. [10] projected a research on predicting survival of kidney dialysis patients using data mining techniques. In this research, a data mining approach is used to extract knowledge about the interaction between these variables and patient survival. Two different data mining algorithms are employed for extracting knowledge in the form of decision rules. Data mining is performed on the individual visits of the "most invariant" patients as they form "signatures" for their decision categories. It concludes that the overall classification accuracy for all data mining algorithms was significantly higher using the individual visit data set over the aggregate data set. The prediction accuracy of individual visit based rule sets increased over the aggregate based rule sets.

In 2015,Mr. S Dayanand [11] proposed the research work to predict kidney diseases by using Support Vector Machine (SVM) and Artificial Neural Network (ANN). The aim of this work is to compare the performance of these two algorithms on the basis of its accuracy and execution time. From the experimental results it is observed that the performance of the ANN is better than the other algorithm. Chronic kidney disease (CKD) is defined by the presence of structural or functional abnormalities of the kidney with or without an accompanying reduction in glomerular filtration rate (GFR). Persons with CKD may have one or more of the following: pathologic abnormalities, markers of kidney damage (i.e., imaging abnormalities and abnormalities in serum or urine,

including proteinuria and abnormal urinary sediment), or GFR less than 60 mL per minute per 1.73m2 for at least three months. Glomerular filtration rate (GFR) is one of the commonly used indexes for early detection of CKD. A five-stage classification system for the disorder has been established by the US National Kidney Foundation's Kidney Disease Outcomes Quality Initiative and adopted internationally by the Kidney Disease: Improving Global Outcomes (KDIGO) initiative to guide identification of cases and facilitate management [14], [15], [16], where glomerular filtration rate (GFR) is the estimator for CKD. Estimation of GFR varies by age, sex, and body size. GFR is approximately 120 to 130 mL per minute per 1.73 m2 in young adults, and decreases by an average of 1 mL per minute per 1.73 m2 per year after 30 years of age [17]. A GFR less than 60 mL per minute per 1.73 m2 represents a loss of at least one-half of normal kidney function; below this level, there is an increased prevalence of CKD complications.

Earlier studies focused on plasma creatinine (Pcr) and creatinine clearance as markers of GFR, but Pcr usually does not increase until GFR has decreased by 50% or more, and many patients with normal Pcr levels frequently have lower GFR [18]. Creatinine clearance is also used to estimate the GFR. But, it overestimates true GFR [19] since creatinine is filtered and secreted by the proximal tubules. Generation of creatinine is determined by muscle mass and diet, whereas tubular secretion could be decreased by the use of medications such as trimethoprim and cimetidine (Tagamet). The serum creatinine level is an insensitive marker of GFR early in the course of CKD. A 33% decrease in GFR may raise the creatinine level from 0.8 to only 1.2 mgperdL(70.72 to 106.08molperL). If the prior creatinine level is not known, this decrease in GFR may go unrecognized. When estimated GFR is suspected to be inaccurate, for example, in patients with severe malnutrition or paraplegia-a 24-hour urine collection should be performed to evaluate creatinine clearance.
Currently, there are three equations commonly used to estimate GFR on the basis of creatinine concentration in serum and demographic features: the Cockcroft-Gault equation [20], the Modification of Diet in Renal Disease (MDRD) equation, [12] and the more accurate Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) [13] formula. Equations 1 and 2 show the MDRD and CKD-EPI equations expressed as single equations where Scr is serum creatinine in mg/dL and α, k are constant values depending on the gender of the patient.

III Proposed Work

The existing prediction system for chronic kidney disease is fine with some limitations. Below is the table shown, describing the worked done for prediction and detection of various kidney diseases. A new CKD prediction system is still the need. A decision support system for chronic kidney disease is still the need for early prediction, as not much work is done for the same.

IV Conclusion

As we have already seen the applications of data mining and machine learning in medical sector. In this paper, a new decision support system is implemented for prediction of CKD. Although ML classifiers worked efficiently in the prediction of other diseases too. In this paper, Chronic Kidney Disease is predicted using two different classifiers and a comparative study of their performance is done. From the analysis we found that, out of two classifiers SVM and KNN, KNN classifier performed better than the other. The rate of prediction of CKD is improved.

There are other possible evolutionary techniques that may be used to improve results of the proposed classifiers. In this paper, SVM and KNN are applied to detect CKD. We can also evaluate and compare the performance of the used classifiers with other existing classifiers. CKD early detection helps in timely treatment of the patients suffering from the disease and also to avoid the disease from getting worse. Early prediction of the disease and timely treatment are the need for medical sector. New classifiers can be used and their performance can be evaluated to

find better solutions of the objective function in future work.

References

[1] Q.-L. Zhang and D. Rothenbacher, "Prevalence of chronic kidney disease in population-based studies: systematic review," BMC public health, vol. 8, no. 1, p. 117, 2008.

[2] M. Baumgarten and T. Gehr, "Chronic kidney disease: detection and evaluation," American family physician, vol. 84, no. 10, p. 1138, 2011.

[3] V. A. Moyer, "Screening for chronic kidney disease: Us preventive services task force recommendation statement," Annals of internal medicine, vol. 157, no. 8, pp. 567–570, 2012.

[4] G. Remuzzi, P. Ruggenenti, and N. Perico, "Chronic renal diseases: renoprotective benefits of renin–angiotensin system inhibition," Annals of internal medicine, vol. 136, no. 8, pp. 604–615, 2002.

[5] L. C. Plantinga, L. E. Boulware, J. Coresh, L. A. Stevens, E. R. Miller, R. Saran, K. L. Messer, A. S. Levey, and N. R. Powe, "Patient awareness of chronic kidney disease: trends and predictors," Archives of internal medicine, vol. 168, no. 20, pp. 2268–2275, 2008.

[6] L. C. Plantinga, D. S. Tuot, and N. R. Powe, "Awareness of chronic kidney disease among patients and providers," Advances in chronic kidney disease, vol. 17, no. 3, pp. 225–236, 2010.

[7] Konstantina Kourou et.al, "Machine learning applications in cancer prognosis and prediction" Computational and structural bitechnology Journal, Elsevier.

[8] P.Swathi Baby, T. Panduranga Vital ,"Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-018, Vol. 4 Issue 07, July-2015,206-210.

[9] K.R.Lakshmi1, Y.Nagesh2 and M.VeeraKrishna3, "Performance Comparision 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.

[10] AndrewKusiak, Bradley Dixonb, Shital Shaha, (2005) Predicting survival time for kidney dialysis patients: a data mining approach, Elsevier Publication, Computers in Biology and Medicine 35, page no 311–327

[11] Dr. S. Vijayarani, Mr.S.Dhayanand, "Kidney Disease Prediction Using SVM and ANN Algorithms" IJCBR , ISSN (online): 2229-6166,Volume 6 Issue 2 March 2015

[12] A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers, and D. Roth, "A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation," Annals of internal medicine, vol. 130, no. 6, pp. 461–470, 1999.

[13] A. S. Levey, L. A. Stevens, C. H. Schmid, Y. L. Zhang, A. F. Castro, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene et al., "A new equation to estimate glomerular filtration rate," Annals of internal medicine, vol. 150, no. 9, pp. 604–612, 2009.

[14] A. S. Levey, J. Coresh, E. Balk, A. T. Kausz, A. Levin, M. W. Steffes, R. J. Hogg, R. D. Perrone, J. Lau, and G. Eknoyan, "National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification," Annals of internal medicine, vol. 139, no. 2, pp. 137–147, 2003.

[15] S. Moe, T. Drueke, J. Cunningham, W. Goodman, K. Martin, K. Olgaard, S. Ott, S. Sprague, N. Lameire, and G. Eknoyan, "Kidney disease: Improving global outcomes (kdigo). definition and classification of chronic kidney disease: a position statement from kidney disease:

Improving global outcomes (kdigo)," Kidney Int, vol. 67, pp. 2089–2100, 2005. 269

[16] A. Levin, B. Hemmelgarn, B. Culleton, S. Tobe, P. McFarlane, M. Ruzicka, K. Burns, B. Manns, C. White, F. Madore et al., "Guidelines for the management of chronic kidney disease," Canadian Medical Association Journal, vol. 179, no. 11, pp. 1154–1162, 2008.

[17] J. D. Kopple, "National kidney foundation k/doqi clinical practice guidelines for nutrition in chronic renal failure," American journal of kidney diseases, vol. 37, no. 1, pp. S66–S70, 2001.

[18] R. D. Perrone, N. E. Madias, and A. S. Levey, "Serum creatinine as an index of renal function: new insights into old concepts." Clinical chemistry, vol. 38, no. 10, pp. 1933–1953, 1992.

[19] S. Giovannetti and G. Barsotti, "In defense of creatinine clearance," Nephron, vol. 59, no. 1, pp. 11–14, 1991.

[20] D. W. Cockcroft and M. H. Gault, "Prediction of creatinine clearance from serum creatinine," Nephron, vol. 16, no. 1, pp. 31–41, 1976.