Grenze International Journal of Engineering and Technology, June Issue

A Survey on Various Machine Learning Classifiers for the Diagnosis of Chronic Kidney Disease

Smeet D. Thakur[1], Abhishek A. Gulhane[2], Shrikant N. Sarda[3] and Himanshu D. Kale[4]
[1-4]Prof.Ram Meghe Institute of Technology & Research, Badnera-Amravati, India,444701
Email: smeet87@gmail.com, aagabhi@gmail.com, shrikantnsarda@gmail.com, himkale@gmail.com

***Abstract*—Chronic Kidney Diseases (CKD) are among one the significant health emergencies around the world that are causing the death of many individuals consistently every year. Major causes of CKD are diabetes, hypertension, and stroke. The need for quick and precise techniques for the identification of CKD has created a demand for machine learning strategies in the healthcare sector. This survey helps in investigating preventive measures for CKD through early analysis utilizing techniques of machine learning. This paper surveys various studies that have been previously performed for the early identification of CKD using various machine learning classifiers. This paper surveys the comparison of various machine learning classifiers presented based on sensitivity, accuracy rate, and specificity used to identify suitable classifiers for the early diagnosis of CKD in patients suffering from it. In this paper, there is a comparative analysis of previous studies performed on Decision Tree, Logistic Regression (LR), K-Nearest Neighbour (K-NN), Random Forest (RF), Support Vector Machine (SVM), Gradient boosting, Artificial Neural Network (ANN), and Radial Basis Function (RBF) for the diagnosis of CKD.**

***Keywords*— CKD, Classifiers, Dialysis, Kidney disease, Machine learning.**

I. INTRODUCTION

People who are in their sixties have a higher risk of having kidney disease. When the symptoms of kidney problems remain constantly for more than ninety days it can be considered CKD. CKD is among the top ten significant reasons behind the death of individuals on the planet. Hypertension, diabetes, and obesity are considered major reasons behind CKD.

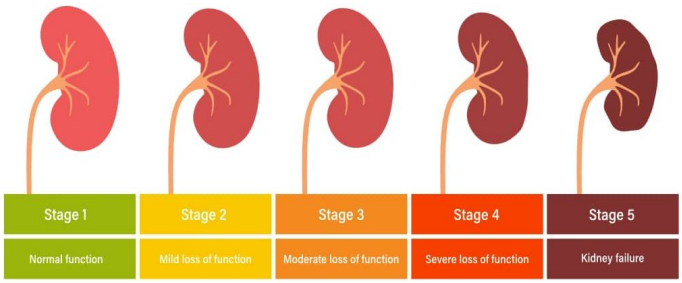

Figure 1: Stages of CKD [6].

Grenze ID: 01.GIJET.11.2.228_20

The application of machine learning techniques helps in the analysis of lab records and other data on patients which is very helpful in the early discovery of CKD. Machine learning classifiers low-level information can be converted into high-level information that can be very helpful in disease identification. This conversion of information can assist practitioners with early diagnoses of CKD [2].
Figure 1 describes stage by stage effect of CKD. According to the Center for Prevention and Control [7], around thirty million individuals worldwide are living with CKD, and more than ninety-five percent of those do not even know that they are suffering from it. After that, it is making decisions based on their knowledge and experience sometimes that may not be satisfactory for the condition of the patient [8].

II. Role of Machine learning in CKD

It comes to chronic kidney disease (CKD), monitoring, therapy, and early diagnosis are all vital. CKD is a hidden illness in its early stages, it is necessary to predict the disease's development with reasonable accuracy. The stage of CKD dictates the medical therapy required at that point. Aside from that, it is critical to find out the presence of CKD because it provides very few indicators. Early identification of CKD is the foundation that can be very helpful in finding treatment and drugs. Medical care is a critical application area for cognitive intelligence systems [13].

A. K-NN

It is used in solving problems that require classification and regression. It is mostly employed for categorization difficulties. A specific training stage is not involved in this method instead of that it uses entire data for training purposes. As it lacks a model, it can use the entire dataset without having to learn anything new.

B. Random tree

It is a supervised classifier that is used in machine learning. It generates many learners who are all different. The tree is created with the use of a stochastic process. It is a classification approach that operates in the same way as a decision tree with the exception that a random subset of characteristics is used for each split.

C. Naive Bayes technique

This method is derived from the Bayes theorem. This training technique is used in both modelings and solving of the dataset which is both quick and efficient. This approach uses predictors, and it is particularly useful for categorizing data sets because it uses non-complicated iterative parameters and is hence suitable for large datasets.

III. Proposed System

Ensemble learning is a machine learning technique that aggregates two or more learners (e.g. regression models, neural networks) in order to produce better predictions for CKD. In other words, an ensemble model combines several individual models to produce more accurate predictions than a single model alone for diagnosis of chronic Kidney Disease.

IV. Literature Review

Various research on machine learning investigates and assesses chronic kidney disease utilizing a variety of strategies for early detection.
Abdel-Fattah et al., (2022) [24] examined that chronic kidney disease (CKD) has permeated across society and has become a pervasive sickness. It is associated with several major hazards, including heart disease, heightened risk,
Singh et al., (2022) [26] suggested that among the most common CKD include diabetes and elevated blood pressure (CKD). As a result, they deserve the right of the neural network to be established by first specifying the parameters and then running multiple trials on them.
Moreno-Sanchez, P. A. et al. (2021) [29] intended that chronic kidney disease (CKD) is now a worldwide concern with a huge financial burden on healthcare systems in which delayed diagnosis leads to early death.
Jongbo, O. A., et al. (2020) [31] studied that millions of people die of chronic kidney disease, which is mostly caused by a combination of bad lifestyle choices and genetic predispositions.
Bhattacharya, et al. (2018) [37] suggested that meta-classification (parallel classification) was used to distinguish related records belonging to the leading part class from those belonging to all combined minority classes.

Wickramasinghe et al. (2017) [39] stated that the use of an adequate diet plan was able to treat chronic kidney disease (CKD) and prescribe diet programs to various individuals based on their classification approach
Patil (2016) [48] investigated the detection accuracy of several data mining strategies. These techniques include LR, multilayer perception, artificial neural network, decision tree, K-NN, Naive Bayes, and Sequence-based minimum optimization.
Sinha and Sinha (2015) [50] studied the performance of the SVM and the K-NN classifier in the diagnosis of chronic kidney disease. The accuracy rates for both systems were 73 and 78 percent, respectively.

V. COMPARATIVE ANALYSIS

It is the primary premise of this comparison study to compare three prior types of research on the diagnosis of CKD using machine learning classifiers, which offers a prediction model based on machine learning to determine the correct therapies for saving lives. Comparative analysis of three previously done studies is performed to identify the performance of machine learning classifiers by assessing their accuracy, sensitivity, and specificity.

A. Accuracy

Table 1 describes the comparative study of the models that are used by various authors for the diagnosis of CKD from which Gradient boosting has a 99 percent accuracy rate.

TABLE I: COMPARISON OF ACCURACY OF VARIOUS CLASSIFIERS

Author	Technique	Accuracy rate (%)
Charleonnan et al. [21]	Logistic Regression Decision Tree KNN	96.55 94.8 98.1
Almasoud and Ward [20]	Support Vector Machine Random forest Gradient boosting	97.5 98.5 99.0
Ramya and Radha [46]	Neural Network Radial Basis Function Random Forest	80.4 85.3 78.6

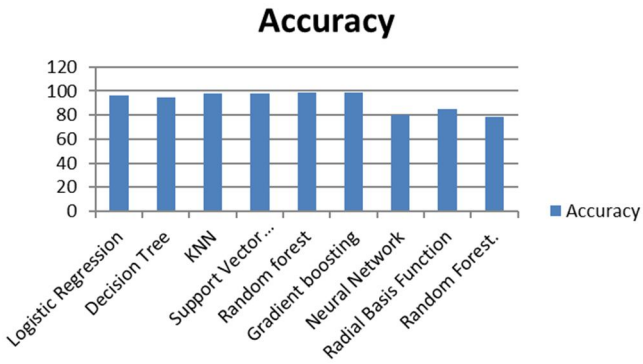

Figure 2: Graphical representation of the accuracy of various classifiers

VI. CONCLUSION

This research is based on an examination of the use of several machine learning algorithms to predict chronic kidney disease (CKD). Globally, the prevalence of chronic kidney disease is alarmingly high, and it is widely acknowledged to be a serious hazard to human life. This study reviewed and analyzed the strategies to increase the classification accuracy of the system to facilitate the effective diagnosis of chronic kidney disease. This topic is directly related to human lives, so the utmost sincerity is required with maximum accuracy. This study has the potential to be extremely beneficial in the accurate prediction of disease, which is in benefits both doctors and patients. This strategy has the potential to lower the cost of disease detection. As a result, the essential interventions can be carried out at the most opportune point in the treatment process. The gradient boosting classifier has the highest accuracy rate of ninety-nine percent, the support vector machine has the highest sensitivity rate of 0.99. Gradient boosting and support vector machines have the highest specificity rate of 0.99 percent.

REFERENCES

[1] Dulhare UN & Ayesha M, “Extraction of action rules for chronic kidney disease using Naïve Bayes classifier”, IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), (2016), pp.1-5.

[2] Kumar G, Kumar K & Sachdeva M, “The use of artificial intelligence-based techniques for intrusion detection: a review”, Artificial Intelligence Review, Vol.34, No.4, (2010), pp.369-387.

[3] Z. Chen, X. Zhang, and Z. Zhang, “Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models,” International Urology and Nephrology, Vol. 48, no. 12, pp. 2069–2075, 2016.

[4] Glomerular Filtration Rate (GFR), National Kidney Foundation, New York, NY, USA, 2020, https://www.kidney.org/atoz/content/gfr.

[5] T. H. Aldhyani, A. S. Alshebami, and M. Y. Alzahrani, “Soft computing model to predict chronic diseases,” Information Science and Engineering, vol. 36, no. 2, pp. 365–376, 2020.

[6] https://www.miskawaanhealth.com/understanding-chronic-kidney-disease/.

[7] Centers for Disease Control and Prevention (2017). Chronic kidney disease surveillance systems.

[8] A. Levey, J. Coresh, Chronic kidney disease, Lancet 379 (9811) (2012) 165–180 pp.

[9] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, in University of Illinois„2nd ed., Elsevier, San Francisco, 2009, pp. 285–306.

[10] W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, ``Estimated glomerular filtration rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice," Oman Med. J., vol. 27, no. 4, pp. 339_340, 2012.

[11] A. S. Levey, D. Cattran, A. Friedman, W. G. Miller, J. Sedor, K. Tuttle, B. Kasiske, and T. Hostetter, Proteinuria as a surrogate outcome in CKD: Report of a scientific workshop sponsored by the national kidney foundation and the US food and drug administration," Amer. J. Kidney Diseases, Vol. 54, no. 2, pp. 205_226, Aug. 2009.

[12] S. Gerogianni, ``Concerns of patients on dialysis: A research study, ‘Health Sci. J., vol. 8, no. 4, pp. 423_437, 2014.

[13] B. Zupan, A. J. Halter, and M. Bohanec, ``Qualitative model approach to computer-assisted reasoning in physiology," in Proc. Intell. Data Anal.Med. Pharmacol. (IDAMAP), Brighton, U.K., 2018, pp. 1_7.

[14] Zixian Wang, Jae Won Chung, Xilin Jiang, Yantong Cui, Mining Wang, and Anqi Zheng, “Machine Learning-Based Prediction System for Chronic Kidney Disease Using Associative Classification Technique” in International Journal of Engineering & Technology, 7 (4.36) (2018) 1161-1167, pp. 1.

[15] Pankaj Chittora, Chaurasia Sandeep, Prasun Chakrabarti, and Zbigniew Leonowicz, “Prediction of Chronic Kidney Disease - A Machine Learning Perspective”, IEEE Access · January 2021.