Employability Of The Machine Learning Tools And Techniques In The Early Detection And Diagnosis Of Chronic Kidney Disease

Vanshika Goel

Vishwakarma University, Pune, Maharashtra, India

DOI:10.37648/ijrmst.v18i01.004

[1]*Received: 21 May 2024; Accepted: 28 July 2024; Published:18 August 2024*

ABSTRACT

Chronic kidney disease is a significant medical condition that requires ongoing monitoring and early detection to prevent negative outcomes. This paper presents a novel research of using machine learning techniques on real-time clinical datasets for early CKD detection and progression tracking. In order to produce accurate insights into the onset and course of chronic kidney disease (CKD), predictive models are constructed utilizing a wide variety of clinical tests and patient data. The suggested approach combines test results with medical histories to evaluate longitudinally collected data in an efficient manner. Through the use of ensemble techniques, this study enhances the efficacy of machine learning algorithms for early CKD detection and progression monitoring. By integrating several clinical data sources, these methods enhance interpretability and accuracy, enabling medical professionals to maximize patient outcomes and treatment.

INTRODUCTION

The kidneys, which resemble beans and are situated behind the spine, carry out a number of vital tasks that are essential for general health. Urine is produced as a result of their complex role in removing waste products and extra fluid from the bloodstream. In addition, kidneys help produce red blood cells and regulate blood pressure, potassium, and salt levels. Although there have been great advancements in healthcare, a critical health issue that affects 11% to 15% of people is becoming more common. CKD is fragile and sometimes disregarded in its early phases. Heart disease is more common in people with chronic kidney disease [1]. Early stages of chronic kidney disease (CKD) often have few symptoms and must be identified by tests like blood and urine tests. Preventive measures and improved therapy are made possible by early detection of chronic kidney disease (CKD), which may reduce the need for dialysis or transplantation. It has been demonstrated that early diagnosis by primary care physicians and nephrology-specialized nurses reduces the progression of illness. A progressive loss of renal function over time characterizes chronic kidney disease (CKD), a disorder that progresses and eventually results in cardiovascular disease, high mortality rates, and End-Stage Renal Disease (ESRD). When kidneys are injured and unable to carry out these vital functions—blood filtration, waste elimination, and fluid and electrolyte regulation—problems arise [2]. Early detection, protection, and treatment of chronic kidney disease (CKD) are essential due to the disease's dynamic and hidden nature in its early stages and individual diversity. Since CKD is characterized by staging, accurate prediction of its progression is essential since it affects therapy choices and decisions [3]. It has identified a number of CKD risk factors that fall under the category of initiating or perpetuating variables. Nephron loss is brought on by initiating variables like diabetes and old age, whereas disease development is accelerated by perpetuating variables like proteinuria. New biomarkers that could be used to forecast the course of CKD include urine connective tissue growth factor [4].

---

[1] *How to cite the article:* Goel V. (August 2024); Employability Of The Machine Learning Tools And Techniques In The Early Detection And Diagnosis Of Chronic Kidney Disease; *International Journal of Research in Medical Sciences and Technology;* Vol 18, 27-33, DOI: http://doi.org/10.37648/ijrmst.v18i01.004

Reducing the cost of dialysis-related healthcare and delaying the course of CKD are both dependent on effective risk factor management [5].

Risk factors for the onset and progression of chronic kidney disease (CKD) include diabetes, hypertension, glomerulonephritis, polycystic kidney disease, age, smoking, obesity, cardiovascular disease, high cholesterol, chronic illnesses, urinary tract infections, and sensitivity to nephrotoxic substances and medications [6]. The stages of CKD range from moderate to end-stage, and proper management calls for appropriate therapy [7]. Real-time data is essential for tracking the early diagnosis and progression of chronic kidney disease (CKD), as it offers valuable insights into the onset and course of the illness [8].

By utilizing the power of real-time data analytics on an extensive dataset that contains the precise characteristics that affect early prognosis and illness progression, this effort aims to improve CKD management strategies. A significant advancement in healthcare analytics has been made with the collection and utilization of a dataset that includes precise parameters crucial for CKD development monitoring and prediction. By utilizing real-time data insights, researchers and medical practitioners can discover predictive biomarkers, enhance current prognostic models, and gain a better understanding of chronic kidney disease (CKD) in order to better serve patients.

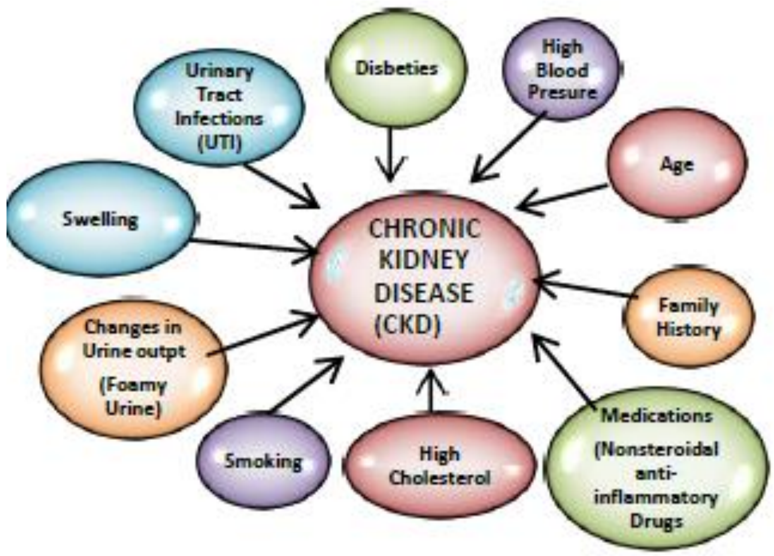

Figure 1: Common Causes of Long-Term Diabetic Failure

MACHINE LEARNING'S PART IN CKD

Medical data processing, diagnosis procedures, and patient care delivery are all being revolutionized by machine learning. Massive datasets including genetic information, electronic health records, and medical imaging are scanned by algorithms to find patterns and insights that are missed by human observers [9]. By offering early disease detection, customized treatment plans, and enhanced clinical judgment, these algorithms optimize resource utilization and patient outcomes [10]. The application of machine learning is becoming more and more significant in the early diagnosis and treatment of Chronic Kidney Disease (CKD), offering novel approaches to prognosis, diagnosis, and individualized care [10]. Machine learning use statistical models and methodologies to examine intricate datasets and detect latent patterns. Risk prediction, multimodal data integration, predictive modeling, medical plan optimization, and intervention result determination are important aspects of machine learning in chronic kidney disease (CKD) [11].

SURVEY OF THE LITERATURE

Machine learning (ML) tools have shown considerable promise in the early diagnosis and forecasting of chronic renal failure. Researchers looked into a number of machine learning techniques and algorithms to develop efficient diagnostic models that would enhance patient outcomes and enable early intervention. Using the XGBoost algorithm, the study's ensemble machine learning approach for data preprocessing and classification produced 98% accuracy in the early diagnosis of CKD [12]. suggested a hybrid machine learning model that addresses overfitting issues and achieves 100% accuracy by combining gradient boosting, naive Bayes, decision tree, and random forest techniques [13]. Scalability concerns are still an issue, though. Using logistic regression (LR) and support vector machine (SVM), a novel approach known as snake optimization feature selection strategy in conjunction with optimal learning methodologies produced results of 97.23% accuracy and 99.9% sensitivity [14]. Nonetheless, the performance of the model can be impacted by the dataset's irregular class distributions. focused approach on clinically applicable machine learning techniques, employing key clinical test factors and SHAP values to achieve 99% accuracy with a random forest (RF) classifier. Beyond the RF classifier, nevertheless, the study's scope was constrained [15].

Using AdaBoost classifiers, an explainable AI model for early CKD detection obtained 97.5% accuracy. However, the study's reliance on a limited set of qualities highlights the need for a larger feature set and led to the creation of a unique model [16]. A further study that looked at seven machine learning techniques empirically found that the CHIRP algorithm had the best accuracy, at 97.75%, for CKD prediction. The study's little dataset, therefore, might not fairly reflect data from the actual world [17].

Using the CKD-15 dataset, an intelligent diagnosis system based on ensemble approaches obtained 96% accuracy. Additional validation using clinical benchmark data is necessary in order to make suitable comparisons [18]. With 99.75% accuracy utilizing RF, a variety of machine learning algorithms were employed to predict CKD based on significant factors like hemoglobin, albumin, and specific gravity. Relationships, however, may impede the removal of features [19]. A decision tree algorithm was used to identify chronic kidney disease (CKD); the random forest method scored 78.25% accuracy, while the J48 strategy scored 85.5%.The latter, however, was more impacted by category values and slower for real-time forecasting [20]. A 99.8% accuracy rate was attained by pairing an affordable AdaBoost with a filter-based attribute selection technique. However, the research revealed unequal distributions of classes [21].

A study that looked into feature optimization techniques for the diagnosis of chronic kidney disease (CKD) used linear discriminant analysis (LDA) and attained 99.5% accuracy. But just one UCI dataset and a small dataset were employed in the study [22]. In one study, a researcher constructed highly accurate diagnostic prediction models, such as logistic regression (LR), with an accuracy rate of 98.5%. Nevertheless, the features and sample size had an impact on the model's performance [23]. A hybrid feature selection technique using machine learning algorithms was provided by the author in a recent study. The strategy revealed an amazing 98% accuracy with extra trees classifiers in the detection of Chronic Kidney Disease (CKD). To ensure the technique's reliability and robustness, they stressed the significance of testing it on a larger cohort of CKD patients. Concurrently, [24] a study using fuzzy logic-based systems to identify chronic kidney disease (CKD), achieving a remarkable 93.75% accuracy rate through adaptive neural fuzzy algorithms. Notwithstanding their success, the researchers noted that just a small number of potential future factors and measurements had been explored, pointing to areas in need of more research and development [25].

In a similar vein, a model developed using support vector machine classifiers for early-stage CKD prediction obtained 98.8% accuracy. They did, however, point out that the size of the dataset may have an impact on predictions, emphasizing the need for bigger and more varied datasets to enhance model performance and generalizability [26].

In order to predict the risk of CKD, one study used undersampling and cost-sensitive loss functions, depending on ensemble models that combined the output from eight classifiers. They attempted to generalize the model, but encountered difficulties because of an unbalanced target variable [27]. This highlights the need to rectify data imbalance in order to make predictions that are more accurate. Additionally, a study utilizing rough K-means clustering achieves an exceptional 100% accuracy and precision, demonstrating soft clustering techniques for chronic illness prediction [28].

SUGGESTED WORK

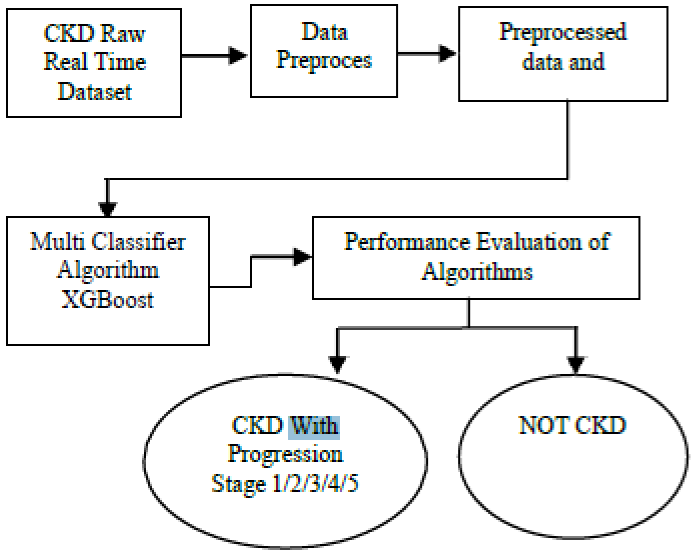

Figure 2: Suggested Approach

- Data Collection and Preprocessing: To build an accurate dataset, the study approach's initial step is to gather various clinical data from numerous sources, including medical records. The dataset needs to be cleaned up by eliminating outliers, inconsistencies, and missing values in order to improve the quality of the data. To guarantee clarity, minimize and normalize features.
- Feature Extraction and Selection: To identify and incorporate crucial elements for early diagnosis and Chronic Kidney Disease (CKD) development, a two-step process will be employed for feature extraction and selection. While taking historical factors into account to capture the disease's course, relevant features will be found using a combination of statistics and subject experience. By identifying the most crucial features for usage by next machine learning algorithms, the goal is to reduce the size of the dataset.
- Model Development and Training: The process for creating and instructing models will again be done in two steps. First, appropriate machine learning techniques (e.g., Gradient Boosting, Random Forests, XGBoost) for CKD progression models and early prediction models must be selected. Next, the entire dataset will be utilized to produce training and validation groups, with progression models being periodically taken into account. The training phase then starts, during which the selected algorithms identify features in the training set. The models are assessed on the validation set and optimized through iterative processes to ensure the best possible result. The development of robust models for progression and early CKD detection is made possible by this two-step approach.
- Model Evaluation: To assess how well machine learning models function, the data should be split into two categories: training and testing. The accuracy and precision of the models' forecasts are evaluated using metrics like these. One technique used to ensure that the models perform effectively under various conditions is k-fold cross-validation. Updated information is provided by incorporating real-time data from clinical test results, and the models are constantly changing as more data is added. Simple descriptions must be used in order to understand how the models make decisions. Close collaboration with healthcare professionals ensures that the models make sense in practical medical situations. The models' stability and effectiveness are confirmed by running them through several sets of data tests and consulting with specialists.

K-fold cross-validation equation

Performance = k 1∑ i=1 k Performance$_i$.

where:

- k is the number of folds (subsets)
- Performance$_i$ is the performance metric obtained from the i – th iteration of cross-validation.

$$\textbf{Accuracy} = \frac{TP + TN}{FP + FN + TP + TN}$$

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

$$\textbf{Recall} = \frac{TP}{TP + FN}$$

$$\textbf{F1 Score} = 2 \text{ X } \frac{\mathit{Precision\ x\ Recall}}{\mathit{Precision} + \mathit{Recall}}$$

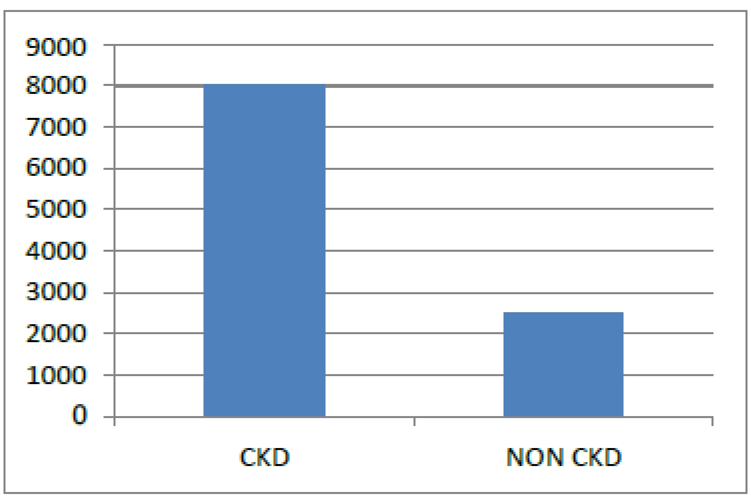

Figure 3: Number of diagnoses for renal disease

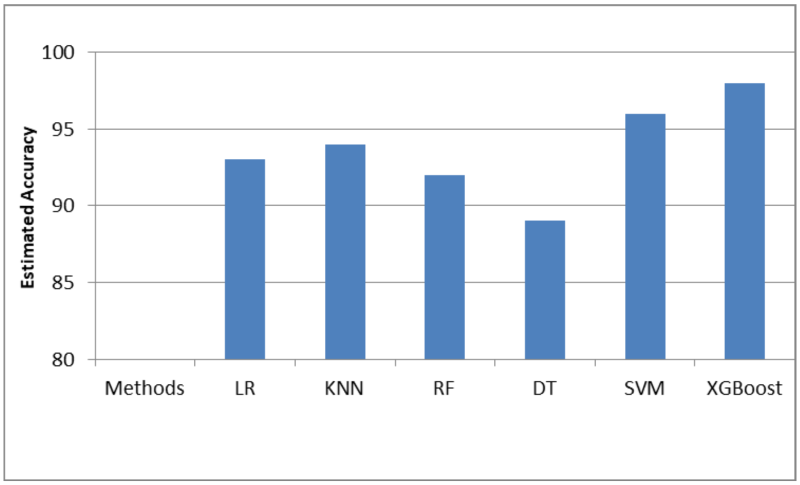

Figure 4: Various ML techniques' performance

OUTCOMES AND COMMENTS

XGBoost fared better than LR, DT, RF, SVM, and KNN on real-time datasets. Its exceptional performance indicates that it can handle complex data with ease. XGBoost's consistent superiority validates its reputation as a robust and dependable algorithm in contemporary machine learning. The results highlight how crucial it is to use the right algorithm for predictive modeling, with XGBoost emerging as the best option for attaining high accuracy across a range of applications. These results illustrate the applicability and usefulness of XGBoost in real-world scenarios by highlighting the significance of applying sophisticated strategies like boosting algorithms when working with complex datasets.

CONCLUSION

The purpose of this research is to employ novel approaches to improve the early detection and tracking of chronic kidney disease. Recursive Feature Elimination (RFE) and an ensemble technique are used in the study to find significant attributes required for a successful diagnosis of CKD. The goal of the research project is to predict the course of CKD with unparalleled precision by applying k-fold cross-validation and the XGBoost algorithm on a well curated real-time dataset. The study predicted notable advancements in the prompt prediction and monitoring of CKD using the power of real-time data sets, promising improved patient outcomes and higher efficiency in the medical sector. With the goal of lowering the global burden of CKD, this study paves the way for future investigations in personalized medicine and predictive analytics.

REFERENCES

1. Langston, C. (2008). Managing fluid and electrolyte disorders in renal failure. Veterinary Clinics of North America: Small Animal Practice, 38(3), 677-697.
2. Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. Scientific reports, 9(1), 9583.
3. Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. Informatics in Medicine Unlocked, 15, 100178.
4. Tsai, W. C., Wu, H. Y et al, (2016). Risk factors for development and progression of chronic kidney disease: a systematic review and exploratory meta- analysis. Medicine, 95(11).
5. De Nicola, L., Chiodini, P. et al, & SIN- TABLE CKD Study Group. (2011). Prognosis of CKD patients receiving outpatient nephrology care in Italy. Clinical journal of the American Society of Nephrology: CJASN, 6(10), 2421.
6. Ortiz, A., Wanner, C. et al, (2022). Chronic kidney disease as cardiovascular risk factor in routine clinical practice: a position statement by the Council of the European Renal Association. European journal of preventive cardiology, 29(17), 2211-2215.

7. Provenzano, M., Puchades, M. J., Garofalo, C., Jongs, N., D'Marco, L., Andreucci, M., & Gorriz, J. L. (2022). Albuminuria-lowering effect of dapagliflozin, eplerenone, and their combination in patients with chronic kidney disease: a randomized crossover clinical trial. Journal of the American Society of Nephrology, 33(8), 1569-1580.
8. Fraser, S. D., & Blakeman, T. (2016). Chronic kidney disease: identification and management in primary care. Pragmatic and observational research,21-32.
9. Park, J. I., Baek, H., Kim, B. R., & Jung, H. H. (2017). Comparison of urine dipstick and albumin: creatinine ratio for chronic kidney disease screening: A population-based study. PloS one, 12(2), e0171106.
10. Song, P., Xu, X., Zhao, Y., Gu, M., Chen, X., Zhang, H., ... & Guo, Q. (2022). Different stages of chronic kidney disease are associated with physical performance in adults over 60 years. Frontiers in Public Health, 10, 963913.
11. Gaitonde, D. Y., Cook, D. L., & Rivera, I. M. (2017). Chronic kidney disease: detection and evaluation. American family physician, 96(12), 776- 783.
12. Venkatesan, V. K., Ramakrishna, M. T., Izonin, I., Tkachenko, R., & Havryliuk, M. (2023). Efficient Data Preprocessing with Ensemble Machine Learning Technique for the Early Detection of Chronic Kidney Disease. Applied Sciences, 13(5), 2885.
13. Khalid, H., Khan, A., Zahid Khan, M., Mehmood, G., & Shuaib Qureshi, M. (2023). Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease. Computational Intelligence and Neuroscience.
14. Ismail, W. N. (2023). Snake-Efficient Feature Selection-Based Framework for Precise Early Detection of Chronic Kidney Disease. Diagnostics, 13(15), 2501.
15. Rashed-Al-Mahfuz, M., Haque et al. (2021). Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (CKD) for use in low-cost diagnostic screening. IEEE Journal of Translational Engineering in Health and Medicine, 9, 1-11.
16. Moreno-Sánchez, P. A. (2023). Data-Driven Early Diagnosis of Chronic Kidney Disease: Development and Evaluation of an Explainable AI Model. IEEE Access, 11, 38359-38369.
17. Khan, B., Naseem, R., Muhammad, F., Abbas, G., & Kim, S. (2020). An empirical evaluation of machine learning techniques for chronic kidney disease prophecy. IEEE Access, 8, 55012-55022.
18. Md Ashafuddula, N. I., Islam, B., & Islam, R. (2023). An Intelligent Diagnostic System to Analyze Early-Stage Chronic Kidney Disease for Clinical Application. Applied Computational Intelligence and Soft Computing, 2023.
19. Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. Journal of Pathology Informatics, 14, 100189.
20. Ilyas, H., Ali, S., Ponum, M., Hasan, O., Mahmood, M. T., Iftikhar, M., & Malik, M. H. (2021). Chronic kidney disease diagnosis using decision tree algorithms. BMC nephrology, 22(1), 1-11.
21. Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. Bioengineering, 9(8), 350.
22. Hossain, M. M., Swarna, R. A., Mostafiz, R., Shaha, P., Pinky, L. Y., Rahman, M. M., ... & Iqbal, M. S. (2022). Analysis of the performance of feature optimization techniques for the diagnosis of machine learning-based chronic kidney disease. Machine Learning with Applications, 9, 100330.
23. Poonia, R. C., Gupta, M. K., Abunadi, I., Albraikan, A. A., Al-Wesabi, F. N., & Hamza, M. (2022). Intelligent diagnostic prediction and classification models for detection of kidney disease. In Healthcare (Vol. 10, No. 2, p. 371). MDPI.
24. Dey, S. K., Uddin, K. M. M., Babu, H. M. H., Rahman, M. M., Howlader, A., & Uddin, K. A. (2022). Chi2-MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease. Intelligent Systems with Applications, 16, 200144.
25. Ahmed, T. I., Bhola, J., Shabaz, M., Singla, J., Rakhra, M., More, S., & Samori, I. A. (2022). Fuzzy logic-based systems for the diagnosis of chronic kidney disease. BioMed Research International, 2022.
26. Arumugam, V., & Priya, S. B. (2022). Selecting Dominant Features for the Prediction of Early-Stage Chronic Kidney Disease. Intelligent Automation & Soft Computing, 31(2).
27. Wang, W., Chakraborty, G. et al, (2020). Predicting the risk of chronic kidney disease (ckd) using machine learning algorithm. Applied Sciences, 11(1), 202.
28. Aldhyani, T. H., Alshebami, A. S., & Alzahrani, M. Y. (2020). Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms. Journal of healthcare engineering, 2020.