International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Data-Driven Early Diagnosis of Chronic Kidney Disease - A Review

Abisha Nazrin M. A[1], Nasreen T. B[2], Safna K. A[3], Shamila Nazeer[4], Theresa Jose[5]

[1,2,3,4] 4th Year Student, [5]Assistant Professor

Department of Computer Science and Engineering, Ilahia College of Engineering and Technology, Mulavoor- 686673, India

ABSTRACT:

Chronic Kidney Disease (CKD) is a global health concern that is becoming more prevalent and has the potential to cause premature death if not detected early. This not only leads to increased healthcare expenses but also poses a significant burden on individuals and society as a whole. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as potential tools for early diagnosis and prevention of further kidney damage in CKD. However, healthcare professionals may be hesitant to adopt AI models if they cannot comprehend the reasoning behind the predictions they generate. To address this concern, Explainable AI (XAI) has been developed to provide understandable explanations for the output of AI models. In the context of CKD, an explainable prediction model has been developed and evaluated. This model not only aids in early diagnosis but also provides valuable information on how different clinical features contribute to the prediction. The development of this model involved the use of an optimization framework that strikes a balance between accuracy and explainability. By doing so, it offers a platform for users to interact with doctors and facilitates the creation of a balanced diet plan for CKD patients. The key contribution of this approach lies in its ability to provide a data-driven explanation of the specific clinical features that play a role in the early diagnosis of CKD. By focusing on a small number of selected features, the cost of early CKD diagnosis is reduced, making it a promising solution for developing countries.

Keywords: AI, CKD, ML, XAI.

1. Introduction

Chronic kidney disease (CKD) is a global public health concern that has seen a growing prevalence, impacting more than 800 million individuals in 2017. This condition carries a significant threat of premature mortality, resulting in 1.2 million deaths from CKD in the same year. CKD is part of a small group of non-communicable diseases that have witnessed an increase in associated fatalities over the last two decades, placing a substantial burden on healthcare systems, particularly in low-middle-income countries where inadequate renal replacement therapy contributes to a high mortality rate. The primary causes of CKD typically involve diabetes and hypertension, both of which are chronic non-communicable diseases that frequently coexist with cardiovascular conditions, which are the leading cause of early morbidity and mortality in CKD patients. Regrettably, CKD often manifests without early symptoms, and by the time it is detected through laboratory testing that measures the estimated glomerular filtration rate (eGFR), the kidney has already lost 25 percent of its function and irreversible damage has commenced towards end-stage kidney disease. During this stage, patients may experience symptoms such as swelling in the legs, extreme fatigue, weakness, shortness of breath, loss of appetite, or confusion. To prevent further deterioration, it is vital to manage underlying risk factors such as hypertension, obesity, heart disease, and age. If these risk factors are not effectively controlled, hemodialysis or kidney transplantation may become necessary to mitigate the exponentially increasing risk of death. Therefore, early diagnosis of CKD based on risk factors, coupled with regular monitoring, allows for the implementation of preventive treatments and therapeutic interventions that can decelerate the progression of kidney damage and prolong patients' lives. Furthermore, the identification of high-risk groups for CKD at an early stage has become a central focus in kidney disease management strategies.

2. Abbreviations And Acronyms

CKD - Chronic Kidney Disease, CHIRP- Composite Hypercube on Iterated Random Projection, DBF- Deep Belief Network, DFD - Dataflow Diagram, DT -Decision Tree, eGFR- Estimated Glomerular Filtration Rate, IoMT- Internet of medical things, LR- Logistic Regression, MAE -Mean absolute error, RAE -Relative Absolute Error, RF -Random Forest, RMSE -Root Mean Squared Error, ROI -Return on Investment, RRSE -Root Relative Squared Error, SVM -Support Vector Machine, UCI -University of California Irvine Machine Learning Repository.

3. Methodologies

Asif Salekin and John Stankovic [1] have proposed a technique that involves developing a machine-learning classifier to detect Chronic Kidney Disease (CKD) based on 24 predictive parameters. By testing this method on a dataset of 400 individuals, they were able to achieve a significant 56% improvement in mean square error when compared to the existing CKD-EPI equation. Additionally, the method uncovered new predictive attributes and conducted a thorough analysis of the tradeoff between cost and accuracy for a novel CKD detection approach.

The study conducted by Sobrinho, Queiroz, Silva, Barros Costa, and Pinheiro [2] explores the potential of machine learning techniques in assisting with the early detection of CKD in developing nations. By conducting a systematic literature review and an experiment using machine learning techniques and the k-fold cross-validation method, the researchers were able to perform qualitative and quantitative comparative analyses. These analyses provide insights into the feasibility of using machine learning techniques to identify CKD risk, particularly in low-income and hard-to-reach settings in developing countries that often face unique challenges like inadequate primary healthcare. The study's findings suggest that the J48 decision tree is a suitable machine-learning technique for screening purposes in such developing countries.

In their research paper, Bilal Khan, Rashid Naseem, Fazal Muhammed, and Ghulam Abbas [3] propose an empirical analysis of machine learning (ML) techniques to classify a dataset of kidney patients as either CKD or NOTCKD. They employ seven ML techniques, namely NBTree, J48, Support Vector Machine, Logistic Regression, Multi-layer Perceptron, Naïve Bayes, and Composite Hypercube on Iterated Random Projection (CHIRP), and evaluate them using various assessment measures such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), recall, precision, F-measure, and accuracy. The experimental results reveal that the MAE values are as follows: 0.0419 for NB, 0.035 for LR, 0.265 for MLP, 0.0229 for J48, 0.015 for SVM, 0.0158 for NBTree, and 0.0025 for CHIRP.

Ahmed Alaiad, Hassan Najadat, Belal Mohsen, and Khaled Balhaf [4] have introduced a method that aims to develop an efficient prediction system for Chronic Kidney Disease (CKD). The primary objective of this system is to reduce medical errors and healthcare costs. To achieve this, the researchers utilized classification and association rule mining techniques on the WEKA and SPSS platforms. They employed five classification algorithms, namely naive Bayes, decision tree, support vector machine, K-nearest neighbor, and JRip, to construct an effective system for predicting and diagnosing CKD and its causes. Additionally, the A priori algorithm was used to identify strong relationship rules among attributes. For their experiments, real medical datasets obtained from hospitals and patient monitoring systems were utilized. The results of these experiments were promising, with a high accuracy rate of 98.50% achieved by the K-nearest neighbor (KNN) classifier and a 96.00% accuracy rate for the classifier based on the association rule (JRip). These findings demonstrate that integrating classification algorithms and association rule mining can significantly improve the accuracy of CKD prediction. Overall, this approach has the potential to enhance healthcare outcomes and holds both theoretical and practical implications for the medical field and healthcare industry.

The research conducted by Syed Ali, Hafis Syed Muhammed, and Fahad Ahmed [5] introduces a novel approach to group-based feature selection through a cost-sensitive ensemble feature ranking method. This method offers two potential solutions for group-based feature selection, each catering to different objectives. By combining these solutions, a consolidated approach is achieved. This study is significant as it is one of the pioneering works that successfully apply cost-sensitive ensemble feature ranking to non-overlapping groups, resulting in solutions that are both low-cost and highly accurate. To validate the effectiveness of the proposed approach, extensive experiments were conducted using seven widely recognized classification algorithms and eight comparative feature selection methods. The results demonstrate that the approach yields a cost-effective and accurate solution for addressing the CKD (Chronic Kidney Disease) problem. By incorporating cost considerations into the solution formulation's objective space, the applicability of automated CKD systems can be significantly improved. Ultimately, a trade-off solution can be attained, which is both cost-effective and accurate enough to serve as a reliable CKD screening system.

The proposed system by P Chittora and Sandeep Chaurasya [6] focuses on predicting Chronic Kidney Disease using a dataset obtained from the UCI repository. The study utilizes seven different classifier algorithms, which include artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with penalty L1 and L2, and random tree. Furthermore, feature selection techniques were applied to the dataset. The results for each classifier were calculated based on various feature selection methods, such as full features, correlation-based feature selection, Wrapper method feature selection, Least absolute shrinkage and selection operator regression, synthetic minority over-sampling technique with Least absolute shrinkage and selection operator regression selected features, and synthetic minority over-sampling technique with full features.

L Antony, Sami Azam, and Abhijith Reddy [7] have proposed the creation of an intelligent system that can categorize patients as either 'CKD' or 'Non-CKD', which would assist doctors in managing multiple patients and expediting diagnoses. This system could be implemented in regional clinics with limited medical experts, allowing for early diagnosis of patients in remote areas. The main goal of this study is to implement and compare the performance of various unsupervised algorithms, to identify the most effective combinations that yield higher accuracy and detection rates. In this research, five unsupervised algorithms, namely K-Means Clustering, DB-Scan, I-Forest, and Autoencoder, have been utilized, along with the integration of different feature selection methods. The integration of feature reduction methods with the K-Means Clustering algorithm has resulted in an overall accuracy.

In the system presented by S.M.M Elkholy [8], an intelligent classification and prediction model is proposed. This model utilizes the Deep Belief Network (DBN) as a classification algorithm to accurately predict kidney-related diseases. The performance evaluation of the model demonstrates an impressive accuracy rate of 98.5% and a sensitivity rate of 87.5% when compared to existing models. Despite advancements in surgical care and treatment, chronic kidney disease (CKD) continues to be a significant health concern. Researchers worldwide have displayed a keen interest in developing high-performance methods for the diagnosis, treatment, and prevention of CKD, given its increasing prevalence.

G Nandhini [9] employed a dataset on chronic kidney disease from the UCI Machine Learning repository, consisting of 25 features. The analysis involved the utilization of three machine learning classifiers, namely Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM). To enhance the model's performance, a bagging ensemble method was applied. The machine learning classifiers were trained using the clusters derived from the chronic kidney disease dataset. The Kidney Disease Collection was then summarized based on category and non-linear features. The highest accuracy of 95.92% was achieved with the decision tree classifier. However, after implementing the bagging ensemble method, the accuracy further improved to 97.23%.

Shamima Akter, Ahsan Habib, and Ashiqul Islam [10] have put forward a variety of clinical features of CKD that were utilized in their study. They implemented seven state-of-the-art deep learning algorithms (ANN, LSTM, GRU, Bidirectional LSTM, Bidirectional GRU, MLP, and Simple RNN) for the prediction and classification of CKD. These algorithms were applied using artificial intelligence techniques, extracting and evaluating features from pre-processed and fitted CKD datasets using five different approaches. To assess the performance of the proposed algorithms in predicting CKD, measurements such as accuracy, precision, recall, loss, and validation loss were taken. Additionally, computation time, prediction ratio, AUC, and statistical significance were analyzed to compare the performances of the models. Among the algorithms used for classifying CKD, ANN, Simple RNN, and MLP demonstrated high accuracy rates of 99%, 96%, and 97% respectively. These models also showed a good prediction ratio and reduced time. The deep learning models used in this study outperformed traditional data classification techniques, highlighting their superior predictive ability. Furthermore, the study suggested integrating the best-performing deep learning models into the Internet of Medical Things (IoMT).

4. Conclusion

In conclusion, the early identification of CKD is of utmost importance to prevent its progression and enhance patient outcomes. The Data-Driven Early Diagnosis of Chronic Kidney Disease is an approach based on machine learning that aims to predict CKD and offer personalized support to users. To achieve this objective, the system employs various machine learning algorithms and data analysis techniques. It surpasses the mere prediction of CKD by providing users with customized diet plans that cater to their specific needs and health conditions. This is made possible through a user-friendly interface that allows easy access to CKD risk assessment, facilitates communication with doctors for personalized consultations, and enables exploration of tailored diet plans. This user-centric design promotes engagement, facilitates proactive health management, and empowers users to actively participate in the management of their health and the reduction of CKD risk. By enabling early detection and timely interventions, this technology can contribute to improved patient outcomes, reduced healthcare expenses, and an enhanced quality of life for individuals with CKD.

5. References

[1] Asif Salekin; John Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes" Published on 2016 IEEE International Conference on Healthcare Informatics (ICHI).

[2] Alvaro Sobrinho, Andressa C. M. Da S. Queiroz, Leandro Dias Da Silva, (Member, Ieee), Evandro De Barros Costa, (Member, IEEE), Maria Eliete Pinheiro, And Angelo Perkusich, (Member, Ieee), "Computer-Aided Diagnosis of Chronic Kidney Disease In Developing Countries: A Comparative Analysis Of Machine Learning Techniques", Published on February 3, 2020.

[3] Bilal Khan, Rashid Naseem 2, Fazal Muhammad, Ghulam Abbas, (Senior Member, IEEE), Sunghwan Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy", Published on March 18, 2020.

[4] Ahmad Alaiad, Hassan Najadat, Belal Mohsen, Khaled Balhaf, "Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease", Published on March 11, 2020.

[5] Syed Imran Ali, Hafiz Syed Muhammad Bilal, Musarrat Hussain, Jamil Hussain, Fahad Ahmed Satti1, Maqbool Hussain, Gwang Hoon Park, (Senior Member, IEEE), Taechoong Chung, And Sungyoung Lee, (Member, IEEE) "Ensemble Feature Ranking for Cost-Based Non- Overlapping Groups: A Case Study of Chronic Kidney Disease Diagnosis in Developing Countries", Published on November 26, 2020.

[6] P. Chittora, "Prediction of Chronic Kidney Disease-A Machine Learning Perspective" Published on January 22, 2021.

[7] Linta Antony, Sami Azam, (Member, Ieee), Eva Ignatious, Ryana Quadir, Abhijith Reddy Beeravolu Mirjam Jonkman, (Member, IEEE) and Friso De Boer, "A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction" Published on August 30, 2021.

[8]S.M.M Elkholy, "Early Prediction of Chronic Kidney Disease Using Deep Belief Network", Published on September 7, 2021.

[9]G Nandhini, J Aravind, "Chronic Kidney Disease Prediction Using Machine Learning Techniques", Published on October 29, 2021.

[10] Shamima Akter, "Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease", Published on November 16, 2021.