e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

HYBRID MODEL FOR PREDICTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

Dr. M. Jayamanmadha Rao*1, K. Gowthami*2, S. Naveen*3, K. Raju*4, K. Harini*5

*1Guide, Professor, Dept. Of ECE, Aitam, Tekkali, Ap, India.

*2,3,4,5B.Tech UG Students, Department Of ECE, Aitam, Tekkali, A.P, India.

DOI : https://www.doi.org/10.56726/IRJMETS52140

ABSTRACT

This study focuses on using machine learning techniques to detect chronic kidney disease (CKD) early, as timely intervention significantly impacts patient outcomes. CKD is often misdiagnosed until it worsens, leading to poorer treatment results. By leveraging advanced machine learning algorithms like Random Forest, Gradient Boosting, and others, the study aims to develop accurate predictive models for early CKD detection. Automation in diagnostics is crucial for expediting and simplifying the process, enhancing accessibility, and overcoming manual limitations. Ultimately, this research aims to improve patient outcomes, reduce healthcare costs, and enhance the quality of life for CKD patients through early detection and tailored interventions. The study emphasizes the silent nature of early-stage CKD and the need for data-driven approaches to evaluate risk factors. By comparing various machine learning classifiers and emphasizing automation, it aims to enhance diagnostic accuracy and accessibility. Ultimately, this research aims to revolutionize early CKD detection, leading to better patient outcomes and a higher standard of living.

Keywords: Chronic Kidney Disease, Machine Learning, Extra Trees Classifier, Decision Tree Classifier, Gradient Boosting Classifier, Random Forest Classifier, K-Nearest Neighbors.

I. INTRODUCTION

Chronic kidney disease (CKD), which is defined by a progressive loss of kidney function over time, is a major global health concern. Chronic kidney disease (CKD) is a silent epidemic that frequently goes undiagnosed until it reaches late stages, at which point there are few treatment choices available and a significant increase in the risk of consequences, such as end-stage renal disease (ESRD) and cardiovascular events. Since CKD is thought to be the cause of 1.7 million fatalities each year, early detection and treatment are essential to reducing the disease's negative effects on both general health and personal wellbeing.

There are five phases of chronic kidney disease (CKD), which go from mild kidney impairment (Stage 1) to total renal failure (Stage 5). The presence of kidney damage, such as proteinuria, and the glomerular filtration rate (GFR), a measure of renal function, define the severity of chronic kidney disease (CKD). People with Stage 3 or 4 CKD frequently experience moderate to severe renal impairment, underscoring the need of early detection and treatment in slowing the illness's progression.

It's critical to recognize the symptoms and indicators of CKD in order to seek medical attention quickly. Advanced chronic kidney disease (CKD) can cause symptoms like weariness, edema, altered urine patterns, and unexplained weight loss, whereas early-stage CKD may not cause any symptoms at all. These symptoms are nonspecific and may coexist with other illnesses, which emphasizes the necessity of a thorough diagnostic assessment by a medical expert.

Healthcare professionals are essential in the identification and treatment of chronic kidney disease (CKD), stressing the value of routine examinations and screening exams for those who are at-risk. Even in the absence of symptoms, those with risk factors such diabetes, hypertension, obesity, and a family history of kidney disease are advised to get screened for chronic kidney disease (CKD) on a regular basis. Early detection of chronic kidney disease (CKD) enables prompt initiation of pharmaceutical therapies, lifestyle adjustments, and appropriate referral to nephrology specialists.

Clinical examination, laboratory testing, and imaging studies are commonly used in the diagnostic diagnosis of chronic kidney disease (CKD). To evaluate kidney function and identify indicators of kidney disease, routine blood and urine tests are necessary. These procedures include serum creatinine, estimated GFR (eGFR), and

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

urine albumin-to-creatinine ratio (ACR). Additionally, kidney shape can be assessed and potential anomalies can be identified using imaging modalities like ultrasound.

Growing interest has been seen in utilizing automation and machine learning to improve CKD risk assessment and diagnosis in light of recent developments in these fields. Algorithms for machine learning have the capacity to evaluate intricate data sets, pinpoint prognostic indicators, and create precise models for the early diagnosis of chronic kidney disease.

With an emphasis on accuracy and automation, this study looks into the usefulness of machine learning prediction models for early CKD diagnosis. Through a combination of clinical data integration and machine learning classifier evaluation, our goal is to create reliable algorithms that can both detect patients at risk of chronic kidney disease (CKD) development and enable prompt therapies. We want to leverage machine learning to improve the delivery of individualized care to patients with chronic kidney disease (CKD) and address its associated issues. To achieve this, we are collaborating with data scientists, healthcare professionals, and technological developers.

II. PROBLEM STATEMENT

Chronic kidney disease (CKD) poses a significant public health challenge due to its silent progression in the early stages, often leading to delayed diagnosis and treatment initiation. One critical consequence of this delayed diagnosis is that approximately 50% of individuals do not receive treatment during the long-term asymptomatic phase of the disease. As a result, CKD continues to advance unnoticed until symptoms become evident or complications arise, at which point the disease may have already reached advanced stages.

The impact of CKD on global health cannot be overstated. It is currently the leading cause of death worldwide, contributing to approximately 1.7 million deaths annually. This staggering statistic underscores the urgent need for effective strategies to improve CKD detection and management.

In the realm of healthcare, diagnosing illnesses typically involves a comprehensive approach. Healthcare professionals conduct physical examinations and review the patient's medical history to gather pertinent information. However, due to CKD's asymptomatic nature in its early stages, relying solely on these traditional diagnostic methods may lead to missed opportunities for early detection.

To overcome this challenge, it is essential to recognize that early CKD diagnosis requires a thorough examination of both typical and uncommon signs and symptoms. While fatigue, edema, and altered urine patterns may manifest in advanced stages, they are nonspecific and may overlap with symptoms of other conditions. Therefore, healthcare providers must maintain a high index of suspicion and consider CKD in their differential diagnosis, particularly in patients with risk factors such as diabetes, hypertension, or a family history of kidney disease.

Achieving an earlier diagnosis of CKD is paramount as there is currently no complete cure for the disease. However, early detection can lead to more successful treatment outcomes by enabling timely interventions to slow disease progression and manage complications effectively. Treatment modalities may include lifestyle modifications, pharmacological therapies, and, in some cases, renal replacement therapy such as dialysis or kidney transplantation.

Furthermore, CKD significantly impacts patients' quality of life. As the disease progresses, individuals may experience debilitating symptoms, diminished physical function, and psychological distress. Moreover, CKD is often accompanied by comorbidities such as cardiovascular disease, further exacerbating the burden on patients' well-being.

Addressing the challenges associated with CKD requires a multifaceted approach that emphasizes the importance of early detection and intervention. By enhancing awareness, implementing screening programs for at-risk populations, and leveraging advancements in diagnostic technologies, healthcare systems can improve outcomes for CKD patients and mitigate the global burden of this devastating disease.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

III. METHODOLOGY

The below figure shows the work process of the project and the individual tasks.

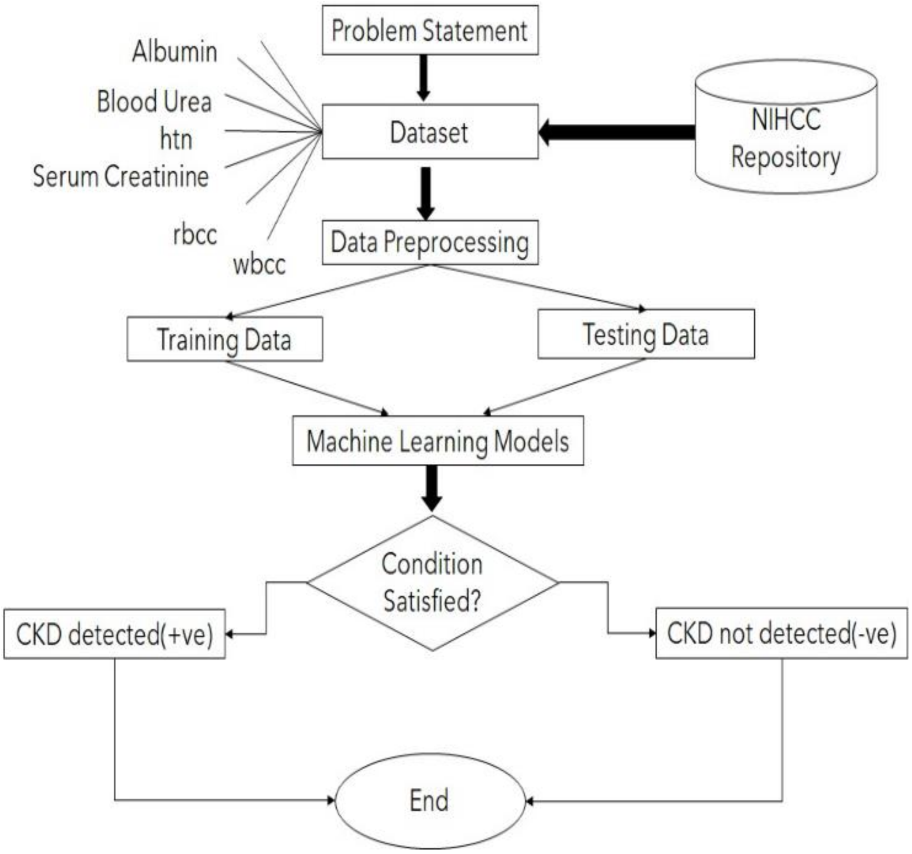

Figure 1: Proposed Method's Process Flow Diagram

Dataset

Because the variables in this dataset are directly related to kidney function and general health, they are essential for predicting chronic kidney disease (CKD). Because blood pressure (Bp) is linked to renal dysfunction and the advancement of disease, it is the major indication. The levels of albumin (Al), which indicate renal impairment and proteinuria, are important. Serum creatinine (Sc) and blood urea (Bu) are vital indicators of renal filtration efficiency, and potassium (Pot) and sodium (Sod) levels show electrolyte abnormalities, which are common in chronic kidney disease (CKD). Hemoglobin (Hemo) values shed light on anemia, a common CKD consequence.

Abnormalities in Red Blood Cell Count (Rbcc) and White Blood Cell Count (Wbcc) may indicate underlying renal illness or its consequences. A major comorbidity of CKD patients is hypertension (Htn), which emphasizes the significance of blood pressure monitoring. When combined, these characteristics provide a thorough awareness of anemia status, electrolyte balance, renal function, and related comorbidities, enabling precise CKD prediction and early identification.

Data Preprocessing

A number of procedures are involved in data preprocessing to guarantee that the dataset is suitable and of high quality for machine learning analysis. To handle missing numbers, outliers, and inconsistencies, the dataset is first cleaned. For example, mean, median, or interpolation methods can be used to impute missing data. To avoid distorting the study, outliers are found and either eliminated or handled properly. In order to scale features to a consistent range and aid in model convergence and performance optimization, data normalization or standardization can also be used. Numerical representations of categorical variables are created by methods such as one-hot encoding. Following preprocessing, the dataset is divided, usually in an 80-20 or 70-30 ratio, into training and testing sets.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

The testing data assesses the machine learning model's performance on unknown data to determine its robustness and generalization ability, whereas the training data is used to train the model. It is also possible to use cross-validation methods like k-fold validation to reduce overfitting and further validate model performance. The model is then fine-tuned using hyperparameter optimization to maximize performance measures including recall, accuracy, precision, and F1-score. This improves the model's ability to predict the condition of chronic kidney disease based on the chosen features.

Training & Evaluation

To train and evaluate the performance of five machine learning algorithms—Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors, Gradient Boosting, and Extra Tree Classifier—the dataset is split into training and testing subsets. Blood pressure, albumin levels, blood urea, serum creatinine, sodium, potassium, hemoglobin, white blood cell count, red blood cell count, and hypertension status are among the attributes that make up the training data that each model is trained on.

The models are tested on the testing dataset after training to see how well they predict the presence of chronic kidney disease (CKD). The algorithm that performs the best for CKD prediction is the one that yields the maximum accuracy on the testing data. This method makes it possible to compare various machine learning approaches, which helps choose the best model for early CKD identification.

IV. RESULTS

	Model	Score
4	Extra Trees Classifier	0.97
2	Random Forest Classifier	0.96
1	Decision Tree Classifier	0.95
3	Gradient Boosting Classifier	0.95
0	KNN	0.64

Figure 2: Accuracy of the algorithms

The above figure gives detailed view of accuracy of each individual algorithm. As shown Extra Tree Classifier algorithm achieved the highest accuracy(97%);followed by Random Forest Classifier(96%),Decision Tree Classifier(95%),Gradient Boosting Classifier(95%) and lastly KNN(94%).

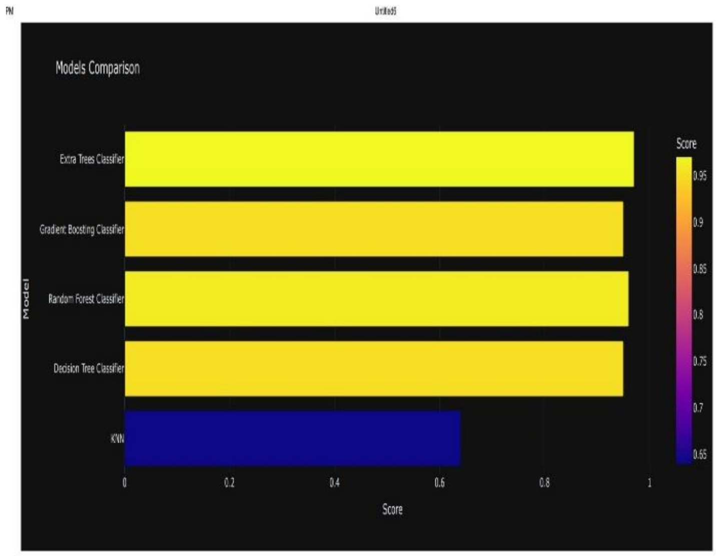

Figure 3: Comparison of the models

V. CONCLUSION

The study's main finding was that machine learning algorithms can be used to predict the stages of chronic kidney disease (CKD) using medical information. With the greatest accuracy of any studied method, Extra Tree Classifier scored 97%, closely followed by Gradient Boosting (95%), Random Forest (96%), and Decision Tree

(95%). The accuracy of K-Nearest Neighbors (KNN) was 64%, which was lower. The results demonstrate how well ensemble learning techniques like Random Forest and Extra Tree work for predicting CKD stage.

VI. FUTURE SCOPES

1. Integration of Advanced Feature Engineering approaches: To improve the prediction performance of the models, future research may investigate the integration of advanced feature engineering approaches. To increase the precision of CKD stage prediction, this may entail extracting further data from medical records, such as novel biomarkers or genetic markers.
2. Using Deep Learning Models: For the diagnosis of chronic kidney disease (CKD), deep learning models like recurrent neural networks (RNNs) and convolutional neural networks (CNNs) may be investigated. These models may provide better accuracy and generalization capabilities for CKD stage prediction, and they have demonstrated promise in a number of medical applications.
3. Creation of a User-Friendly Web-Based Diagnostic Tool: Creating a web-based diagnostic tool that combines the top-performing machine learning model (such Extra Tree Classifier) for CKD stage prediction could be a future approach. With the use of this instrument, medical personnel may be able to obtain precise CKD diagnosis in real time, allowing for prompt patient treatment and action. It might also provide patients with informational materials and assistance so they can comprehend their illness and available treatments.

VII. REFERENCES

[1] Hira Khalid, Ajab Khan, Muhammad Zahid Khan, Gulzar Mehmood. “Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease”. Retrieved from https://doi.org/10.1155/2023/9266889

[2] Anandanadarajah Nishanth; Tharmarajah Thiruvaran. “Identifying Important Attributes for Early Detection of Chronic Kidney Disease”.Retrieved from https://ieeexplore.ieee.org/document/8239857

[3] Md. Ariful Islam,Md. Ziaul Hasan Majumder, and Md. Alomgeer Hussein. “Chronic kidney disease prediction based on machine learning algorithms”. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9874070/

[4] Hamida Ilyas , Sajid Ali , Mahvish Ponum , Osman Hasan , Muhammad Tahir Mahmood .“Chronic kidney disease diagnosis using decision tree algorithms”. Retrieved from https://pubmed.ncbi.nlm.nih.gov/34372817/

[5] Md. Johir Raihan, Md. Al-Masrur Khan, Seong-Hoon Kee .“Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP”. Retrieved from https://www.nature.com/articles/s41598-023-33525-0#Abs1

[6] Amogh Babu K A, Priyanka K, Raghavendra Babu T M. “Chronic Kidney Disease Prediction Based On Naive Bayes Technique”.Retrieved from https://www.irjet.net/archives/V6/i9/IRJET-V6I9251.pdf

[7] G.-S. Fu, Y. Levin-Schwartz, Q.-H. Lin, and D. Zhang, “Machine learning for medical imaging,” Journal of healthcare engineering, vol. 2019, pp. 1-2, 2019.

[8] R. Devika, S. V. Avilala, and V. Subramaniyaswamy, “Comparative study of classifier for chronic kidney disease prediction using naive Bayes, KNN and random forest,” in Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, March 2019.

[9] A.S.A. Rabby, R. Mamata, M. A. Laboni, Ohidujjaman, and S. Abujar, “Machine learning applied to kidney disease prediction: comparison study,” in Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, July 2019.

[10] M. Nishat, F. Faisal, R. Dip et al., “A comprehensive analysis on detecting chronic kidney disease by employing machine learning algorithms,” EAI Endorsed Transactions on Pervasive Health and Technology, vol. 7, Article ID 170671, 2018.

[11] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, “A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease,” in Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, July 2017.

[12] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Intelligent heart disease prediction system using random forest and evolutionary approach," Journal of network and innovative computing, vol. 4, pp. 175–184, 2016.

[13] S. Ramya and N. Radha, "Diagnosis of chronic kidney disease using machine learning algorithms," International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, no. 1, 2016.

[14] M. Kumar, "Prediction of chronic kidney disease using random forest machine learning algorithm," International Journal of Computer Science and Mobile Computing, vol. 5, pp. 24–33, 2016.

[15] M. S. Basarslan and F. Kayaalp, "Performance analysis of fuzzy rough set-based and correlation-based attribute selection methods on detection of chronic kidney disease with various classifiers," in Proceedings of the 2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT), IEEE, Istanbul, Turkey, April 2019.