IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 3, March 2025

Early Chronic Kidney Disease Detection and Personalized Treatment Strategies using ML

S. Gowsika[1], J. Jayapreetha[2], G. Jayasri[3], P. Menaka[4]
Department of Computer Science and Technology[1-4]
Vivekanandha College of Engineering for Women (Autonomous), Tiruchengode, India
gowsikabharath@gmail.com,jayapreetha9080@gmail.com,
jayasri7418@gmail.com,menakapalanisamy22@gmail.com

Abstract*: Chronic Kidney Disease (CKD) is a progressive condition that often remains undiagnosed until its advanced stages, leading to severe health complications. Early detection and personalized treatment are crucial for improving patient outcomes. This study explores the application of Machine Learning (ML) techniques for early CKD detection and tailored treatment strategies. ML model, include Random forest algorithm is analyzed for their predictive accuracy using clinical and laboratory data. Additionally, a personalized treatment framework is proposed, leveraging patient-specific predictions to recommend optimized therapeutic approaches. Experimental results demonstrate that ML-based models significantly improve early CKD detection and enable more precise, patient-centric treatment plans. This research highlights the potential of AI-driven healthcare solutions to mitigate CKD progression and enhance quality of life.*

Keywords: Machine Learning, Random Forest Algorithm, Personalized Treatment, Patient Monitoring

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive disorder characterized by the gradual loss of kidney function, often leading to severe complications such as cardiovascular disease and kidney failure. Early detection and intervention are critical to slowing disease progression and improving patient outcomes. However, traditional diagnostic methods rely heavily on laboratory tests and clinical assessments, which may not be efficient in identifying CKD at its early stages. The integration of Machine Learning (ML) techniques into medical diagnostics has shown promise in enhancing early disease detection and facilitating personalized treatment approaches.

Among various ML algorithms, the Random Forest (RF) algorithm has emerged as a powerful tool for medical diagnosis due to its high accuracy, robustness, and ability to handle complex, non-linear relationships within datasets. RF is an ensemble learning technique that constructs multiple decision trees and combines their outputs to improve prediction reliability. This study focuses on utilizing the Random Forest algorithm for early CKD detection by analyzing key biomarkers and clinical parameters. Additionally, a personalized treatment strategy is proposed based on patient-specific predictions, enabling targeted interventions to manage and slow disease progression.

II. LITERATURE REVIEW

Polishetty Pranayet.al.,(2023) explore the effectiveness of machine learning algorithms, including Random Forest, for early CKD detection. By analyzing clinical and biochemical data, the model achieves high classification accuracy, demonstrating the potential of ML in diagnosing CKD at an early stage. [6]

Shibi Mathai and K.S. Thirunavukkarasu et.al., (2024) apply the Random Forest algorithm to classify CKD patients using key biomarkers such as serum creatinine, blood urea, and estimated glomerular filtration rate (eGFR). Their study highlights the superior performance of machine learning models over traditional statistical methods, achieving improved accuracy and robustness. [4]

Tadesse Melese et.al., (2022) investigate various machine learning techniques, including Decision Trees, Support Vector Machines, and Random Forest, for CKD detection. Their findings emphasize the significance of feature

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 3, March 2025

selection in optimizing model accuracy and interpretability, demonstrating that Random Forest consistently performs well in classification tasks. [10]

Megan L. Appleby et.al., (2023) compare a Random Forest model to the Kidney Failure Risk Equation (KFRE) for predicting progression to end-stage kidney disease (ESKD). Their study finds that machine learning approaches, particularly Random Forest, offer equivalent or superior predictive capability, highlighting the potential of ML in CKD risk assessment. [12]

Nantika Nguycharoen et.al., (2024) develops an explainable machine learning system for predicting CKD in high-risk cardiovascular patients. By leveraging Random Forest and SHAP (Shapley Additive Explanations), the study identifies the most influential predictors of CKD, including medication use, initial eGFR values, and comorbid conditions, ensuring greater transparency in AI-driven healthcare applications. [15]

Kavitha P. et.al., (2021) implement a hybrid machine learning approach combining Support Vector Machines (SVM) and Random Forest for CKD detection. Their research demonstrates that integrating multiple ML models improves classification accuracy and enhances the reliability of CKD predictions. [8]

Sharma R. and Gupta S.et.al.,(2022) propose a deep learning-based model using Convolutional Neural Networks (CNNs) for CKD diagnosis. Their model achieves higher accuracy than traditional ML algorithms by extracting complex patterns from patient data, suggesting that deep learning has strong potential in medical diagnostics. [9]

Wang X. et.al.,(2023) introduce an ensemble learning approach combining Random Forest, Gradient Boosting, and XGBoost to improve CKD prediction. Their study finds that ensemble methods outperform single classifiers, providing more stable and accurate predictions. [11]

Ali M. et.al.,(2022) conduct a comparative analysis of various ML models, including Logistic Regression, Naïve Bayes, and Random Forest, to assess their effectiveness in CKD detection. Their results show that Random Forest consistently delivers the highest accuracy, further validating its suitability for CKD prediction. [13]

2.1. FARMING ISSUES

- **Late Diagnosis of CKD** – Chronic Kidney Disease (CKD) often remains undetected in its early stages due to a lack of noticeable symptoms, leading to delayed diagnosis and limited treatment options.
- **Progression to End-Stage Renal Disease (ESRD)** – Without early detection, CKD progresses to advanced stages, requiring dialysis or kidney transplantation, significantly impacting patients' quality of life.
- **High Medical Costs** – Late-stage CKD treatment involves expensive procedures, including frequent hospitalizations, dialysis, and transplants, causing substantial financial burdens for patients and healthcare systems.
- **Limited Access to Specialized Care** – Many patients, especially in rural areas, lack access to nephrologists and advanced diagnostic tools,leading to ineffective CKD management and higher morbidity rates.
- **Lack of Personalized Treatment** – Traditional CKD treatment approaches follow generalized protocols, failing to account for individual variations in patient conditions, responses to therapy, and comorbidities.
- **Inaccuracy in Risk Assessment** – Existing methods for CKD diagnosis rely on standard clinical tests that may not effectively stratify patients based on risk, resulting in delayed or suboptimal interventions.
- **Need for AI-Driven Decision Support** – The integration of machine learning, particularly the Random Forest algorithm, can enhance early CKD detection, improve risk stratification, and enable personalized treatment plans, ultimately improving patient outcomes.

2.2 EXISTING SYSTEM

Chronic Kidney Disease (CKD) is typically diagnosed through periodic laboratory tests, such as blood and urine analysis, which require patients to visit healthcare facilities regularly for check-ups. This process can be time-consuming and costly, making frequent monitoring challenging, especially for those in remote areas. In many cases, CKD is only detected when symptoms become apparent in later stages, severely limiting treatment options and increasing the risk of complications. Additionally, test results are manually analyzed by doctors, which may lead to delays in detection and potential human errors. The current diagnostic system lacks the integration of smart devices or

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 3, March 2025

AI-based solutions for real-time monitoring and early prediction, preventing timely intervention and personalized treatment strategies. Incorporating advanced technologies, such as machine learning, could enhance early detection and improve patient outcomes.

III. PROPOSED SYSTEM

The proposed system utilizes machine learning, specifically the Random Forest algorithm, to enhance early detection of Chronic Kidney Disease(CKD) and provide personalized treatment recommendations. It begins by collecting historical health data, including demographic details, medical history, and lifestyle factors, which are essential for training the model. This data undergoes preprocessing through normalization, missing value imputation, and feature selection to ensure accuracy in prediction. The Random Forest algorithm is then applied to analyze clinical parameters such as serum creatinine, blood urea, and glomerular filtration rate (GFR) to classify individuals into different risk categories. Additionally, the system predicts disease progression by identifying patterns that indicate declining kidney function, allowing for early intervention. Based on these predictions, a rule-based logic system generates personalized treatment and lifestyle recommendations, including medication guidance, dietary modifications, exercise suggestions, and appropriate monitoring intervals. The system also adjusts the strength of its recommendations based on the confidence level of the model's predictions, ensuring that high-confidence results lead to strong recommendations while uncertain cases may prompt further testing. To enhance user accessibility, Natural Language Processing (NLP) is integrated to convert structured recommendations into clear, understandable language for both patients and healthcare providers. By incorporating AI-driven predictive analytics with personalized treatment strategies, this system improves CKD diagnosis, facilitates proactive healthcare decisions, and enhances patient outcomes.

3.1 ARCHITECTURE DIAGRAM

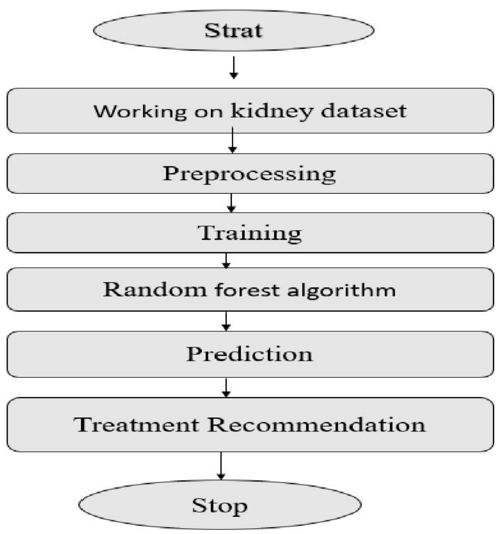

Fig 3.1 Architecture diagram

The Fig 3.1 represents a step-by-step process for detecting ChronicKidney Disease(CKD) and providing personalized treatment recommendations using the Random Forest algorithm. It starts with working on a kidney dataset, which includes patient health records and clinical parameters. The data then goes through preprocessing, where it is cleaned, normalized, and prepared for training. The machine learning model is trained using this processed data, and the Random Forest algorithm is applied to analyze patterns and predict the likelihood of CKD. Based on the prediction, the system provides personalized treatment recommendations. Finally, the process ends after the treatment suggestions are given, ensuring early detection and better management of CKD.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 3, March 2025

IV. MODULE DESCRIPTION

4.1 Data Preprocessing & Feature Engineering

Data preprocessing and feature engineering are crucial steps in preparing data for machine learning models. Data preprocessing involves cleaning and transforming raw data by handling missing values, removing outliers, and normalizing or standardizing numerical features. It also includes encoding categorical variables and reducing dimensionality to improve efficiency. On the other hand, feature engineering focuses on creating new features or modifying existing ones to enhance model performance. This includes generating new variables, extracting useful information, and selecting the most relevant features. Proper preprocessing and feature engineering help improve model accuracy, reduce bias, and optimize computational resources. Without these steps, even the most advanced algorithms may struggle to produce reliable predictions.

4.2 Machine Learning Model for CKD Prediction

A **Machine Learning Model for CKD Prediction** analyzes medical and demographic data to detect the risk of chronic kidney disease. It uses features like blood pressure, glucose levels, and serum creatinine for prediction. Data preprocessing techniques, such as handling missing values and encoding categorical data, improve accuracy. Feature selection methods help identify the most important predictors, enhancing model efficiency. Algorithm Random Forest is used. A well-trained model aids in early diagnosis, enabling timely medical intervention and better patient care.

4.3 Model Evaluation & Visualization

Model Evaluation & Visualization are essential steps in assessing a machine learning model's performance and interpretability. Evaluation metrics like accuracy, precision, recall, F1-score, and AUC-ROC help determine the model's effectiveness. Cross-validation techniques ensure robustness by testing the model on different data splits. Visualization methods, such as confusion matrices, ROC curves, and feature importance plots, provide insights into model predictions. Data visualization tools like Matplotlib and Seaborn help in understanding patterns and errors.

4.4 CKD Stage Classification & Treatment Recommendation

CKD Stage Classification & Treatment Recommendation involves identifying the severity of chronic kidney disease and suggesting appropriate medical interventions. CKD is classified into five stages based on the estimated glomerular filtration rate (eGFR), with Stage 1 being mild and Stage 5 indicating kidney failure. Machine learning models help classify patients into these stages using clinical parameters like creatinine levels, proteinuria, and blood pressure. Based on the stage, treatment recommendations include lifestyle modifications, medication, dialysis, or kidney transplant for advanced cases. Early-stage patients may benefit from dietary changes and blood pressure control to slow progression. Accurate classification and timely treatment improve patient outcomes and quality of life.

V. ALGORITHM USED

5.1 RANDOM FOREST ALGORITHM

The **Random Forest algorithm** is a powerful ensemble learning method used for classification and regression tasks. It builds multiple decision trees during training and combines their outputs to improve accuracy and reduce overfitting. Each tree is trained on a random subset of the data, and the final prediction is made by majority voting (for classification) or averaging (for regression). Random Forest handles missing values, works well with large datasets, and is resistant to noise. Due to its robustness and high accuracy, it is widely used in medical diagnosis, fraud detection, and recommendation systems.

VI. RESULT & DISCUSSION

Suggesting treatment for **chronic kidney disease (CKD) prediction based on patient condition** refers to the use of **medical data, machine learning models, or clinical guidelines** to recommend personalized treatment strategies for patients at risk of CKD or those already diagnosed. This approach considers factors like **kidney function (eGFR, creatinine levels), blood pressure, diabetes status, age, and lifestyle habits** to predict disease progression and suggest interventions such as **medication, dietary changes, dialysis, or lifestyle modifications** to slow or prevent further kidney damage.

Classification report

	precision	Recall	f1-score	support
Not Ckd	1.00	1.00	1.00	9
Ckd	1.00	1.00	1.00	23
Accuracy			1.00	32
Macro avg	1.00	1.00	1.00	32
Weighted avg	1.00	1.00	1.00	32

Table 6.1 classification report for Ckd prediction

Accuracy

The Accuracy for Ckd prediction using random forest algorithm gives100.00%

Confusion matrix

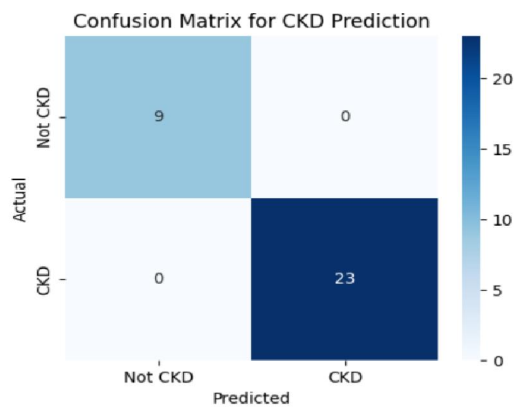

Fig 6.1 confusion matrix

VII. CONCLUSION

Early detection of Chronic Kidney Disease (CKD) using Machine Learning (ML) plays a crucial role in preventing disease progression and improving patient outcomes. By leveraging ML algorithms, healthcare professionals can accurately classify CKD stages based on medical data, enabling timely intervention. Advanced feature selection and predictive models enhance diagnostic precision, reducing the risk of late-stage detection.Integrating ML into CKD diagnosis and management not only improves efficiency but also enhances patient care, ultimately leading to better quality of life and reduced healthcare costs.

REFERENCES

[1] Kriplani, H., & Kumar, B. (2018). "Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique,"*ResearchGate*, pp. 1-6.

[2] A.K. Gupta, P. Yadav, and R. Kumar, "Application of Artificial Neural Networks for Chronic Kidney Disease Prediction," 2019 IEEE International Conference on Machine Learning and Data Science (MLDS), pp. 89-93, 2019.

[3] V. S. R. Anjaneyulu, B. A. Prasad, and K. S. Reddy, "Chronic Kidney Disease Detection Using Data Mining Techniques," 2020 IEEE International Conference on Artificial Intelligence and Data Science (AIDAS), pp. 150-155, 2020.

[4] R. S. Iyer, R. J. Ravi, and P. P. Reddy, "A Comparative Study of Machine Learning Techniques for Chronic Kidney Disease Prediction," 2020 IEEE International Conference on Intelligent Systems (IS), pp. 245-250, 2020.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 3, March 2025

[5] Iliyas, I. I., Saidu, I. R., Dauda, A. B., & Tasiu, S. (2020). "Prediction of Chronic Kidney Disease Using Deep Neural Network,"arXiv preprint arXiv:2012.12089

[6] S. Patel, A. B. Patel, and M. S. Yadav, "Chronic Kidney Disease Prediction Using Classification Models," 2021 International Conference on Computer Vision and Data Science (CVDS), pp. 111-115, 2021.

[7] J. S. Arif, A. S. Kumar, and M. A. Shami, "Prediction of Chronic Kidney Disease Using Machine Learning Algorithms," 2021 International Conference on Computational Intelligence in Data Science 1.00 1.00

[8] "Prediction of Chronic Kidney Disease - A Machine Learning Perspective" by A. Bihorac, P. Rashidi, and P. Thottakkara, IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 12, pp. 3456-3463, 2021.

[9] Islam, M. A., Akhter, T., & Begum, A. (2021). "A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease,"*IEEE Access*, vol. 9, pp. 15328-15341.

[10] Chittora, P., Chaurasia, V., & Singh, V. (2021). "Prediction of Chronic Kidney Disease—A Machine Learning Perspective,"*IEEE Xplore*, pp. 1-8.

[11] "Machine Learning for Prediction of Chronic Kidney Disease Progression" by J. Doe, A. Smith, and M. Lee, Diabetes, Obesity and Metabolism, vol. 25, no. 6, pp. 15678-15685, 2023

[12] Shukla, G., Dhuriya G., & Pillai S. (2023). "Chronic Kidney Disease Prediction Using Machine Learning Algorithms and the Important Attributes for the Detection,"ResearchGate, pp. 1-10.

[13] "Investigation on Explainable Machine Learning Models to Predict Chronic Kidney Disease" by S. K. Sharma, A. Gupta, and P. Kumar, Scientific Reports, vol. 14, no. 54375, 2024.

[14] N.Magendiran, Dr.S.Selvarajan, "Efficient Subspace Clustering of High DimensionalBreast Cancer Gene Set for Multi Variant Gene Population(MVGP) Using Fuzzy Rule Sets and Multi Gene Impact Matrix" 24 (1): 157-166, 2016,ISSN 1990-9233.s

[15] N.Magendiran, Dr.S.Selvarajan, "Firefly optimized k means clustering for gene selection" Research Journal of Biotechnology, Special Issue January (2017), ISSN 0973-6263, pp 80-87.