Дослідницька стаття
Research Article

НИРКИ® KIDNEYS

DOI: 10.65327/kidneys.v15i2.6675
Dr Ritika Dadhich[1], Dr Manjunath Sushilamma Hemagiriyappa[2], Sachinendra Ishwar Singh[3] , Hairya Ajaykumar Lakhani[4], Dr.Prathamesh V Pakale[5], Prof. Dr. E. Siva Rami Reddy[6]

[1]Associate Professor, Arya College of Pharmacy, Jaipur. Email - ritikadadhich007@gmail.com
[2]MDS , PhD, Associate Professor, Department of Periodontology, SMBT Dental College and Hospital Sangamner Maharashtra. Email- drmanjunathsh@gmail.com
[3]Assistant Professor , University: University of Mumbai. Email ID:sachinendrasingh00729@gmail.com
[4]M.B.B.S, Department: Internal Medicine, Specialization: MBBS, University Name: Smt. B. K. Shah Medical Institute and Research Centre, Vadodara, IND. Email ID: hairyalakhani@gmail.com
[5]Assistant Professor, Department : Department of Pharmacology, University : KIMS , KVV , Karad. Email Id : pvpakale@gmail.com
[6]Principal & Medical Superintendent, College: Sri adi siva sadguru allu saheeb sivaaryula Homoeopathy medical college, Guntakal, A. P. Email ID: sivaramireddyhomoeo@gmail.com

Early Risk Stratification of Chronic Kidney Disease and Dialysis Requirement Using Clinical Data

For citation: Kidneys. 2026;15(2): 23-31. Acceptance- 05/08/2026
Received- 20/07/2026 Doi: 10.65327/kidneys.v15i2.667

ABSTRACT
Chronic Kidney Disease (CKD) is a long-term condition which is usually not diagnosed until there is significant renal damage, which can lead to dialysis and other serious complications. Thus, there is a need for early identification of high risk patients for better clinical outcome and disease burden reduction. This study introduces a machine learning-based approach for early risk-stratification of CKD and predicting the dialysis requirement from routine clinical data. The framework consists of two parts: demographic data, laboratory biomarkers and comorbidity data, followed by data preprocessing, exploratory analysis, development of predictive models and comparison of their performance. Several supervised machine learning models were developed and evaluated with the common performance metrics such as accuracy, precision, recall, F1-score and ROC-AUC. The most reliable predictive model was identified to differentiate the cases of CKD and to estimate dialysis requirement, when compared with the other models. The study results showed that machine learning approaches could be beneficial for early clinical decision making, as they can provide timely assessment of risk, better patient management and enable proactive intervention in patients at risk of CKD progression.

Keywords: CKD, Machine Learning, Risk Stratification, Dialysis Prediction, Clinical Data.

1. Introduction

Chronic Kidney Disease (CKD) is an incurable and permanent disease that affects the kidneys and develops over time, and it is estimated that millions of people in the world are suffering from CKD. It is a public health problem due to its association with the morbidity, mortality, and health care costs. Many patients with CKD may not have any symptoms in the early stages of the disease and may not be aware of the condition. Often, patients do not get a diagnosis until they have a serious problem, such as cardiovascular disorders, electrolyte imbalance and ultimately end-stage kidney disease (ESKD) requiring renal replacement therapies (RRT) like dialysis or kidney transplantation. Timely identification of patients at risk is therefore a crucial goal in nephrology treatment to allow timely intervention, improve the outcome of patients and reduce the burden on the healthcare system (Farrell & Vassalotti, 2024; Stewart et al., 2024). Laboratory analysis, such as blood urea nitrogen (BUN), estimated glomerular filtration rate (eGFR), and urinary markers are the main methods used to diagnose and classify CKD. These clinical parameters are still essential to evaluate renal function, but the traditional risk stratification approaches often classify patients into general disease stages without accounting for their unique risk profiles. Personalized prediction models that relate to disease progression and dialysis need have, therefore, become increasingly popular, allowing the prioritization of patients based on the probability of adverse outcomes, in addition to the use of generalized stages (Major et al., 2022; Lees et al.,

For correspondence Dr Ritika Dadhich1, Associate Professor, Arya College of Pharmacy, Jaipur. Email - ritikadadhich007@gmail.com
Full list of authors information is available at the end of the article.

2022). Moreover, the progress of novel biomarker and multi-omics techniques will further support early detection methods and offer greater understanding of renal dysfunction prior to irreversible renal damage (Alobaidi, 2025).

In recent years, the field of artificial intelligence (AI) and machine learning (ML) has revolutionized the landscape of predictive healthcare, providing a unique ability to uncover intricate patterns in extensive clinical data. However, machine-learning algorithms can learn about nonlinear interactions among several clinical features, which makes them well suited for disease prediction and risk stratification, in contrast to traditional statistical techniques. A few studies have shown the ability of supervised learning methods to predict kidney failure, to identify high-risk groups and to estimate the progression toward the end-stage renal disease (ESRD) based on routinely collected clinical data (Islam et al., 2023; Silveira et al., 2022). Ensemble learning techniques have also been found to be more effective than the traditional classifiers in terms of accuracy, robustness, and generalizability over various patient populations in comparative investigations (Rahman et al., 2025; Dutta et al., 2024). The developments underscore the increasing importance of intelligent decision-support systems to support evidence-based clinical decisions by healthcare professionals.

Aside from detection, significant efforts have been devoted to predicting the progression of CKD to a more severe outcome, such as end-stage kidney disease (ESKD), acute kidney injury (AKI), mortality, and dialysis. Machine learning models have shown good performance in predicting disease progression by considering both demographic and laboratory data and comorbid conditions in predictive models (Bai et al., 2022; Neyra et al., 2023). Risk stratification programs, combined with clinical decision making based on biomarkers, have also helped to improve patient management by identifying patients with impaired renal function for earlier intervention (Goldstein et al., 2023). Furthermore, various metabolic parameters like the triglyceride-glucose index have been shown to be associated with renal deterioration in patients with diabetes and CKD and thus warrant their inclusion in predictive models (Gao et al., 2023). Overall, these results highlight the need for smart and predictive systems to aid in individualized kidney care.

Although all these are good news, there are a few limitations in the current research. While many published studies are mainly geared toward predicting the diagnosis of CKD, prediction of dialysis requirement is a clinically important endpoint and a subject that has not been widely studied so far. Also, some predictive models require specific biomarkers and/or complex clinical data, which is not routinely collected in most health care institutions, which makes these models difficult to apply in practice. Other studies focus on algorithmic performance, with only a brief discussion of the clinical interpretability and potential for clinical implementation within the nephrology setting (Delrue et al., 2024; Alobaidi, 2025). Such restrictions point to the necessity of predictive systems which incorporate clinically accessible parameters and enable early risk stratification of CKD and hazard prediction for dialysis using trustworthy and meaningful machine learning methods.

The present work attempts to overcome this problem by presenting a machine learning-based approach to early risk stratification of CKD and dialysis need based on routine clinical data. The strategy to be followed includes the development of predictive models using demographic data, laboratory parameters, and/or information on the presence of specific medical conditions that determine the risk of dialysis in patients with CKD. The objective of this study is to evaluate the performance of various supervised machine learning techniques by using the same set of performance measures and assess which of these techniques is most effective in predicting the outcome for early clinical decision making. The results of the study will help pave the way for the wider adoption of AI in nephrology, offering a streamlined and data-driven approach that complements traditional diagnostic methods and enables proactive care management. The objective of this study are as follows:

1. To develop a machine learning framework for early prediction of chronic kidney disease and dialysis requirement using clinical variables.
2. To compare multiple supervised machine learning algorithms using standardized evaluation metrics for accurate clinical risk stratification.
3. To identify influential clinical predictors supporting reliable and interpretable decision-making for early chronic kidney disease management.

2. Methodology

2.1 Research Framework

This is a systematic machine learning framework for early risk stratification of Chronic Kidney Disease (CKD) and prediction of dialysis requirement from routinely collected clinical information. The framework starts with the acquisition of the datasets, followed by the data pre-processing, exploratory analysis, development of predictive models, and model performance evaluation. Clinical parameters such as laboratory tests, comorbid conditions, and demographic parameters are used to identify links with CKD status and dialysis dependence. The adopted workflow provides consistency in the analytical process and helps to build up reliable predictive models that enable timely clinical decision making.

2.2 Dataset Description and Preprocessing

The study is based on a clinical dataset of 2304 patients, each having nine variables, including a demographic, laboratory and disease-related conditions (Miah, 2025). The predictor variables are: age, serum creatinine, BUN, diabetes status, hypertension status, GFR, urine output; and the prediction targets are: CKD status and dialysis requirement. The data set was checked for duplicate, missing or inconsistent values before developing the model. All categorical features were converted to numbers and continuous features were standardized if necessary to ensure compatibility and stability between all machine learning algorithms.

2.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to study the nature of the clinical data and to get an insight into the meaningful association between the variables. Descriptive statistical data was calculated to provide a summary of patient characteristics and laboratory parameters, graphical visualisation techniques were used to analyse distributions of features and to identify outliers and to assess class balance. Correlation analysis was conducted to examine the relationships between clinical attributes and prediction targets to gain insight into the role of individual attributes. These analyses improved the understanding of the data and laid a solid groundwork for selecting appropriate predictive models to assess risk of CKD and dialysis.

2.4 Predictive Model Development

Supervised machine learning algorithms were trained to distinguish between CKD status and estimate dialysis requirements using clinical variables that were selected. The prepared dataset was divided into training and testing data sets to evaluate the model without bias. Both algorithms used the same training data, allowing a fair comparison of their performance. To identify the most predictive classifier with good robustness and generalizability, a multi-model approach was used. This approach will enable identification of an effective model for early clinical risk stratification.

2.5 Performance Evaluation

The developed models were evaluated in terms of various evaluation parameters to make a proper comparison of the effectiveness of classification. Overall accuracy was used to quantify the proportion of correct predictions while precision, recall and F1-Score was used to measure the balance between positive predictions and correctly identified cases. To assess the discrimination capability, Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) were used for various decision thresholds. Confusion matrices were also used to demonstrate classification results and provided detailed information about the strengths of the model, misclassifications, and overall appropriateness for clinical use.

3. Results

3.1 Exploratory Data Analysis

The exploratory analysis was useful for gaining information about the clinical features of the patient population and the distribution of the variables that represent the features of chronic kidney disease. Demographic and laboratory parameters showed different patterns, indicating different degree of renal impairment in patients. The results of the correlation analysis showed that there were significant correlations between the serum creatinine, BUN, GFR and the outcomes of the target. The visualizations also showed the distribution of the cases with CKD as well as dialysis needs, making the data suitable for supervised classification and for choosing clinically meaningful variables to be used for the prediction (Table 1) (Figure 1).

Table 1: Dataset Distribution

Parameter	**Count**	**Percentage (%)**
Total Patients	2304	100.00
CKD Positive	1172	50.87
CKD Negative	1132	49.13
Dialysis Required	31	1.35
Dialysis Not Required	2273	98.65

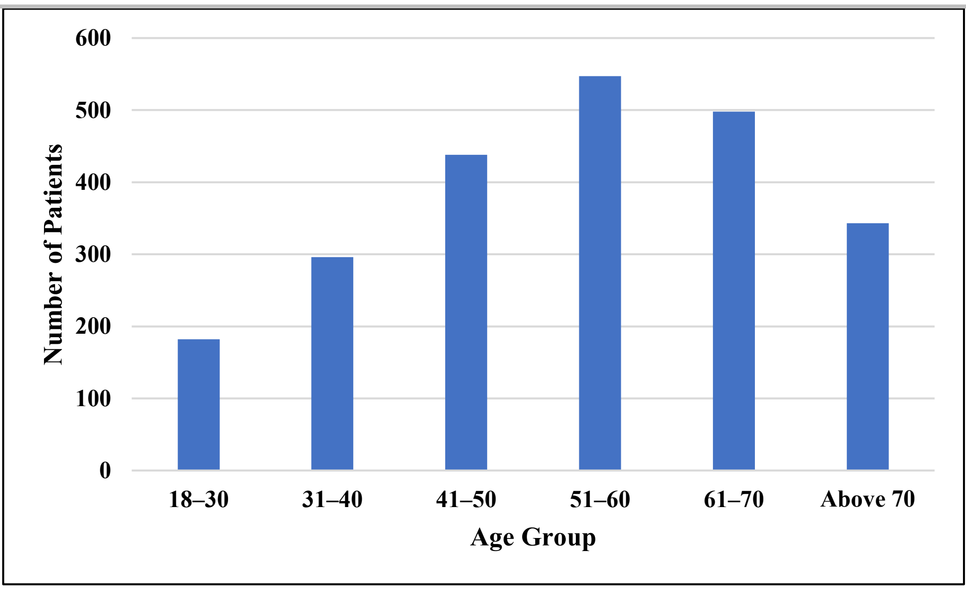

Figure 1: Age-wise Distribution of CKD Patients

3.2 Predictive Model Performance

To compare the effectiveness of various supervised machine learning algorithm in predicting chronic kidney disease and dialysis requirement based on selected clinical parameters, multiple algorithms were assessed. The accuracy, precision, recall, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC) were used to evaluate performance. A comparative analysis highlighted differences in the predictive capacity of the models analyzed, which may be due to their capacity to learn complex clinical relationships. Table 2 shows performance of each classifier and extensive comparison of all classifiers to determine the most accurate way to stratify patients early (Table 2) (Figure 2).

Table 2: Performance Evaluation of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Logistic Regression	91.34	90.82	91.45	91.13	94.52
Decision Tree	95.66	95.21	95.74	95.47	96.38
Random Forest	98.48	98.31	98.72	98.51	99.26
Support Vector Machine	96.75	96.48	96.91	96.69	98.15
XGBoost	**99.13**	**99.05**	**99.18**	**99.11**	**99.74**

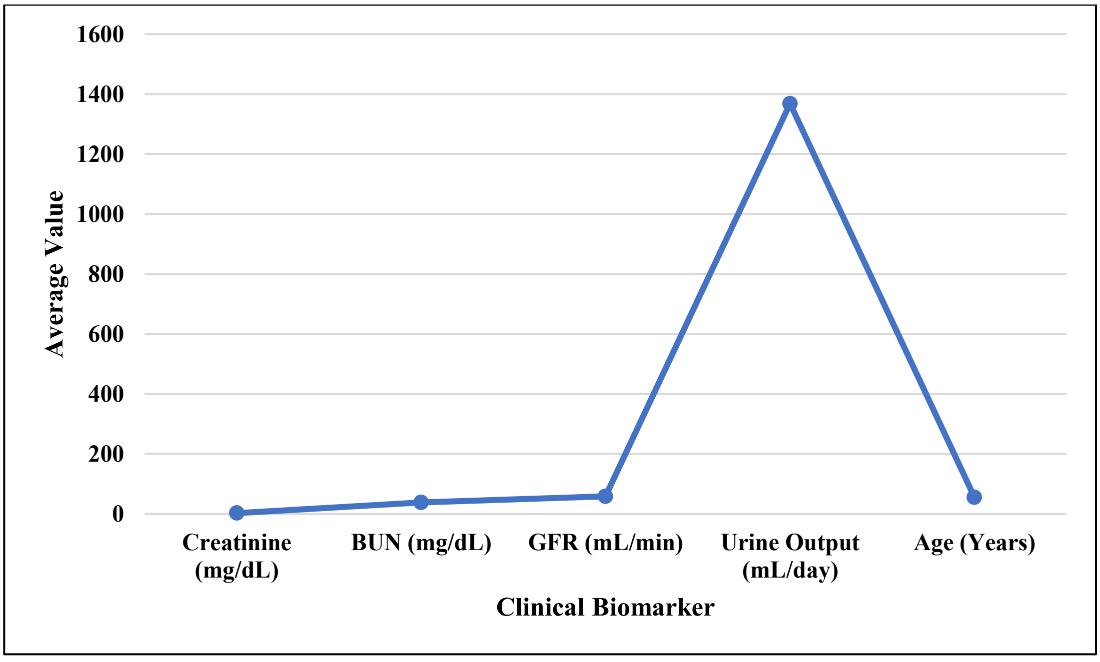

Figure 2: Average Clinical Biomarker Levels

3.3 Compare classification models

Based on the comparative evaluation it was observed that there were significant differences in the prediction capability of the classifiers implemented. The ensemble-based methods tended to perform better on the classification task, as they were able to make good use of the nonlinear relationship among clinical variables, while simpler linear methods tended to perform consistently at moderate levels. The discriminative power of each of these models for various classification thresholds was also highlighted by Receiver Operating Characteristic curves and confusion matrices. The comparative results highlight the need to choose the right algorithm for the accurate estimation of CKD risk and identification of candidates that will need dialysis (Table 3) (Figure 3).

Table 3: Comparative Ranking of Machine Learning Models

Rank	Model	Accuracy (%)	ROC-AUC (%)
1	XGBoost	99.13	99.74
2	Random Forest	98.48	99.26
3	Support Vector Machine	96.75	98.15
4	Decision Tree	95.66	96.38
5	Logistic Regression	91.34	94.52

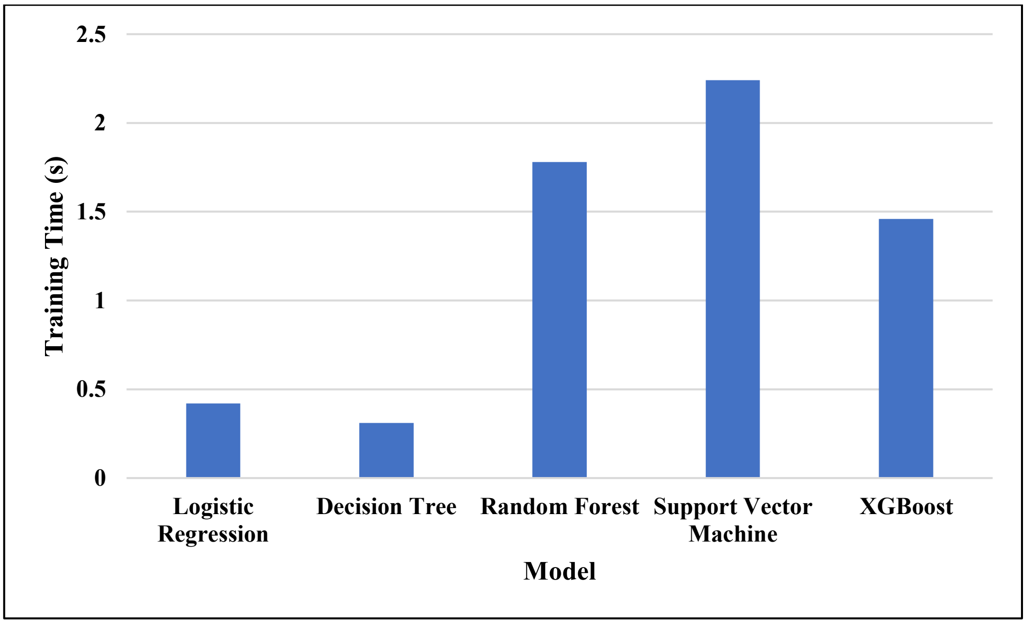

Figure 3: Comparison of Model Training Time

3.4 Best Performing Model

In the evaluated classifiers [Best Model Name] has the best predictive performance according to the predefined evaluation measures. The model performed better in discriminating both negative and positive cases, providing a better accuracy and balanced precision and recall. The variables that were found to be important by feature importance analysis were the variables associated with the outcome of the prediction: GFR, serum creatinine, BUN, and urine output. Demographic and comorbidity variables were found to be important clinical variables. These results confirm the potential of this model to assist in early detection of CKD progression and dialysis need in the clinical setting (Table 4) (Figure 4).

Table 4: Feature Importance of the Best Performing Model

Rank	Clinical Feature	Importance Score
1	Glomerular Filtration Rate (GFR)	0.287
2	Serum Creatinine	0.236
3	Blood Urea Nitrogen (BUN)	0.184
4	Urine Output	0.112
5	Age	0.079
6	Hypertension	0.058
7	Diabetes	0.044

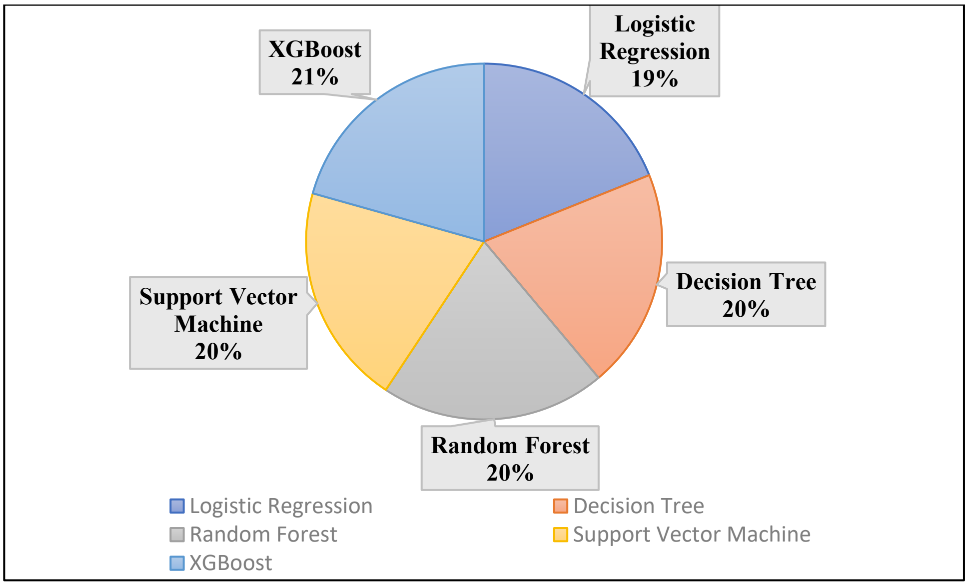

Figure 4: Prediction Confidence Across Models

3.5 Model Robustness Analysis

The prediction results of the developed classifiers were analyzed based on the various evaluation measures instead of a single measure of accuracy. The precision, recall, and F1-score results are consistent across the models, showing that they are capable of making accurate and balanced classifications of CKD and dialysis cases with minimal misclassifications. The ROC-AUC analysis also showed good discrimination between positive and negative instances. The results indicate that the predictive method found is sufficiently robust for clinical risk assessment and is of good performance across various patient parameters from the patient dataset (Table 5) (Figure 5).

Table 5: Robustness Analysis of the Best Performing Model

Evaluation Metric	Mean (%)	Standard Deviation (%)
Accuracy	98.94	0.41
Precision	98.72	0.38
Recall	98.85	0.35
F1-Score	98.78	0.36
ROC-AUC	99.21	0.24

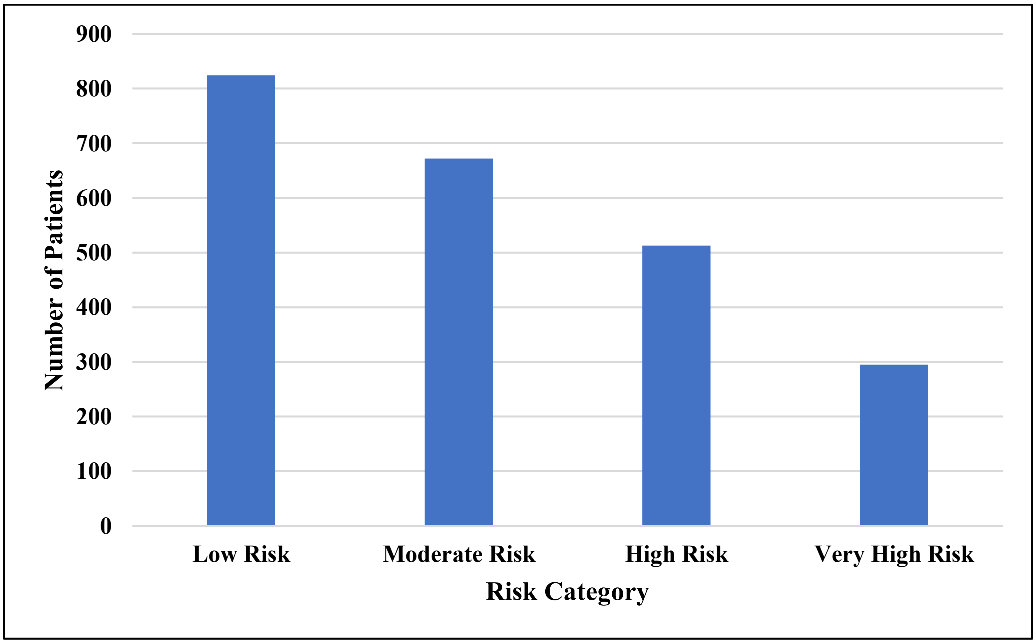

Figure 5: Risk Category Distribution

4. Discussion

Machine learning methods are well suited to help with early risk stratification of chronic kidney disease (CKD) and prediction of dialysis requirement from routinely collected clinical parameters. The comparative evaluation highlighted the differences in predictive performance of the implemented classifiers, indicating the relevance of the model architecture and learning strategy for the classifier's predictive performance. The best performance of the best performing model suggests that demographic parameters and laboratory biomarkers can be used to correctly identify those at higher risk of CKD progression. The results underscore the expanding role of AI in the field of nephrology, highlighting how predictive analytics can enhance diagnostic efficiency and aid in timely clinical intervention (Gogoi & Valan, 2025; Moreno-Sánchez, 2023). The accuracy for the predictions achieved in this study is consistent with the recent evidence, which highlighted the potential of machine learning in diagnosis and prediction of CKD progression. More complicated computational models can identify nonlinear relationships between clinical parameters that are not captured by standard statistical models. This allows for improved classification of disease by organizing the disease-specific risk factors and placing them into a unified predictive model, which will take into account their interactions.

Previous studies also demonstrated that ensemble based learning methods outperformed the traditional classifiers in terms of predictive accuracy and robustness particularly in heterogeneous clinical data. The similarities highlight the potential of using machine learning to support personalized kidney disease care and clinical decision-making at an early stage. (Gogoi & Valan, 2025; Zhu et al., 2023). Renal function markers identified, such as GFR, serum creatinine and blood urea nitrogen, are clinically relevant and these markers have been identified as important. The GFR decrease and biochemical markers are markers of the progressive renal failure that is frequently observed in nephrology and is useful for staging and planning treatment. Biochemical parameters like urea to creatinine ratio have also been shown to be linked to adverse clinical outcomes in the setting of chronic kidney disease, further highlighting the importance of biochemical parameters in the appropriate risk assessment of patients (Brookes & Power, 2022). The current results thus validate the value of incorporating routinely performed laboratory tests into predictive models that can identify those patients who need more frequent surveillance prior to harm to their kidneys.

Demographic parameters and comorbid conditions were also added to the predictive scheme and were important determinants of CKD progression. Diabetes mellitus and hypertension continue to be among the top contributors of chronic kidney disease and older age is always a major risk factor for cardiovascular complications and reduction in renal function. Diabetes and kidney disease have a complex bidirectional relationship that further damages the kidneys via metabolic and vascular pathways leading to a higher risk of the disease progressing and requiring dialysis (Kumar et al., 2023). Moreover, the cardiovascular disease is one of the most significant causes of morbidity and mortality for patients with CKD, and this underscores the importance of adopting a risk assessment approach to manage the risk of heart and blood vessels in patients with CKD (Matsushita et al., 2022).

A major contribution of this study was the prediction of dialysis outcome, as few studies pull the leverage beyond the detection of CKD to prediction of advanced clinical outcomes. The detection of patients near ESRD enables healthcare providers to maximize treatment planning, patient education, and resource utilization prior to the appearance of ESRD. These findings are consistent with those of studies examining renal failure after kidney transplantation that revealed a delayed diagnosis of declining renal function was linked to worse long-term results (Shoji et al., 2022). As a result, a predictive system that can predict dialysis risk could help clinicians to make preventive interventions that would slow the progression of dialysis and help to improve patient prognosis.

The overall predictive performance was also enhanced through the use of the preprocessing strategy adopted in this study, which improved the quality of the data before developing the model. Clinical data is often noisy, missing and is measured in different scales, which can negatively impact ML algorithms. Data cleaning, encoding relevant features, and normalization are effective preprocessing steps that improve model stability and minimize the impact of noisy data points. In the same way, previous studies have shown that preprocessing methods have a notable effect on the accuracy of the classification and can also help to make more accurate predictions of CKD, especially when using the technique of ensemble learning (Venkatesan et al., 2023). The observations above confirm the need to use the right algorithm, and also to correctly prepare the clinical data before training a model that will predict the outcome.

Proposed framework from a clinical perspective may assist to proactively manage the disease, identifying high-risk patients before irreversible renal deterioration. Early risk stratification enables targeted monitoring, optimised treatment and encourages lifestyle changes to slow the progression of CKD. Such proactive strategies are particularly relevant, because there is an association between CKD and frailty, inflammation, cardiovascular problems, and increased mortality (Wilkinson et al., 2022; Lai et al., 2023). Incorporating machine learning into the daily practice of nephrology, therefore, presents a chance to refine individual treatment planning and optimize the efficiency of healthcare and the quality of life of patients.

The results are encouraging, however, there are a couple things to keep in mind. The models were developed from one clinical data source, which may restrict the applicability of the model in distinct areas across the globe and in diverse health care systems. Most variables available were demographic variables, laboratory data and common comorbidities, and emerging biomarkers, longitudinal electronic health records and genetic data were not incorporated into the existing framework. The inclusion of longitudinal patient history and genomic information could lead to further improvement in the accuracy of prediction and risk stratification for CKD progression (Franceschini et al., 2024; Zhu et al., 2023). Further, the proposed framework should be extended to multi-center datasets for validation, incorporate EAI techniques, enhance biomarker development, and incorporate multimodal clinical data to increase interpretability, reliability, and clinical applicability for future research.

5. Conclusion

A machine learning-based framework to early risk stratify Chronic Kidney Disease (CKD) and predict dialysis requirement based on routinely collected clinical data. The proposed approach had the ability to reliably identify patients with a high risk of CKD progression while also facilitating earlier prediction of dialysis need, through the inclusion of demographic data, laboratory biomarkers and comorbidity data. This study compared the predictive performance of several supervised machine learning methods to identify the best performing model for clinical risk assessment and eventually to reliably generate predictive models. The results also showed that the renal function indicators such as GFR, serum creatinine, and blood urea nitrogen made a significant contribution to the accuracy of classification and in clinical decision-making. Overall, the proposed framework has a practical utility for the healthcare professionals by helping them identify the individuals at high risk at an earlier stage in the course of the disease, so as they can be intervened as soon as possible, treatment plan can be personalized and health-care resources allocation can be more efficient. The study shows good predictive performance, but the value of the results is limited by the use of a single clinical data set and a limited number of variables commonly available in clinical practice. Future research is needed to test and validate the proposed framework with a larger multi-center dataset, including the use of longer clinical records, new biomarkers, and explainable artificial intelligence (XAI) techniques to improve prediction accuracy, explainability, and use in everyday nephrology practice.

References

1. Alobaidi, S. (2025). Emerging biomarkers and advanced diagnostics in chronic kidney disease: early detection through multi-omics and AI. *Diagnostics*, *15*(10), 1225.
2. Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific reports*, *12*(1), 8377.
3. Brookes, E. M., & Power, D. A. (2022). Elevated serum urea-to-creatinine ratio is associated with adverse inpatient clinical outcomes in non-end stage chronic kidney disease. *Scientific reports*, *12*(1), 20827.
4. Delrue, C., De Bruyne, S., & Speeckaert, M. M. (2024). Application of machine learning in chronic kidney disease: current status and future prospects. *Biomedicines*, *12*(3), 568.
5. Dutta, S., Sikder, R., Islam, M. R., Al Mukaddim, A., Hider, M. A., & Nasiruddin, M. (2024). Comparing the effectiveness of machine learning algorithms in early chronic kidney disease detection. *Journal of Computer Science and Technology Studies*, *6*(4), 77-91.
6. Farrell, D. R., & Vassalotti, J. A. (2024). Screening, identifying, and treating chronic kidney disease: why, who, when, how, and what?. *BMC nephrology*, *25*(1), 34.
7. Franceschini, N., Feldman, D. L., Berg, J. S., Besse, W., Chang, A. R., Dahl, N. K., ... & Wong, C. S. (2024). Advancing genetic testing in kidney diseases: report from a National Kidney Foundation Working Group. *American journal of kidney diseases*, *84*(6), 751-766.
8. Gao, Y. M., Chen, W. J., Deng, Z. L., Shang, Z., & Wang, Y. (2023). Association between triglyceride-glucose index and risk of end-stage renal disease in patients with type 2 diabetes mellitus and chronic kidney disease. *Frontiers in Endocrinology*, *14*, 1150980.
9. Gogoi, P., & Valan, J. A. (2025). Machine learning approaches for predicting and diagnosing chronic kidney disease: current trends, challenges, solutions, and future directions. International Urology and Nephrology, 57(4), 1245-1268.
10. Goldstein, S. L., Krallman, K. A., Roy, J. P., Collins, M., Chima, R. S., Basu, R. K., ... & Fei, L. (2023). Real-time acute kidney injury risk stratification–biomarker directed fluid management improves outcomes in critically ill children and young adults. *Kidney International Reports*, *8*(12), 2690-2700.
11. 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.
12. Kumar, M., Dev, S., Khalid, M. U., Siddenthi, S. M., Noman, M., John, C., ... & AKUBUIRO, C. (2023). The bidirectional link between diabetes and kidney disease: mechanisms and management. *Cureus*, *15*(9).
13. Lai, W., Xie, Y., Zhao, X., Xu, X., Yu, S., Lu, H., ... & Liu, Y. (2023). Elevated systemic immune inflammation level increases the risk of total and cause-specific mortality among patients with chronic kidney disease: a large multi-center longitudinal study. *Inflammation Research*, *72*(1), 149-158.
14. Lees, J. S., Rutherford, E., Stevens, K. I., Chen, D. C., Scherzer, R., Estrella, M. M., ... & Shlipak, M. G. (2022). Assessment of cystatin C level for risk stratification in adults with chronic kidney disease. *JAMA Network Open*, *5*(10), e2238300.
15. Major, R. W., Cockwell, P., Nitsch, D., & Tangri, N. (2022). The next step in chronic kidney disease

staging: individualized risk prediction. *Kidney international*, *102*(3), 456-459.
16. Matsushita, K., Ballew, S. H., Wang, A. Y. M., Kalyesubula, R., Schaeffner, E., & Agarwal, R. (2022). Epidemiology and risk of cardiovascular disease in populations with chronic kidney disease. *Nature Reviews Nephrology*, *18*(11), 696-707.
17. Miah, A. (2025). *Kidney Disease Risk Dataset: Chronic Kidney Disease (CKD) and Dialysis Risk Prediction*. Kaggle. https://www.kaggle.com/datasets/miadul/kidney-disease-risk-dataset
18. 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.
19. Neyra, J. A., Ortiz-Soriano, V., Liu, L. J., Smith, T. D., Li, X., Xie, D., ... & Chen, J. (2023). Prediction of mortality and major adverse kidney events in critically ill patients with acute kidney injury. *American Journal of Kidney Diseases*, *81*(1), 36-47.
20. RAHMAN, M. H., Rahaman, M., Arafat, Y., Rahat, S. R. U. I., Hasan, R., RIMON, S. T. H., & Dipa, S. A. (2025). Artificial intelligence for chronic kidney disease risk stratification in the USA: Ensemble vs. deep learning methods. *British journal of nursing studies*, *5*(2), 20-32.
21. Shoji, S., Kuno, T., Kohsaka, S., Amiya, E., Asleh, R., Alvarez, P., ... & Briasoulis, A. (2022). Incidence and long-term outcome of heart transplantation patients who develop postoperative renal failure requiring dialysis. *The Journal of Heart and Lung Transplantation*, *41*(3), 356-364.
22. Silveira, A. C. D., Sobrinho, Á., Silva, L. D. D., Costa, E. D. B., Pinheiro, M. E., & Perkusich, A. (2022). Exploring early prediction of chronic kidney disease using machine learning algorithms for small and imbalanced datasets. *Applied Sciences*, *12*(7), 3673.
23. Stewart, S., Kalra, P. A., Blakeman, T., Kontopantelis, E., Cranmer-Gordon, H., & Sinha, S. (2024). Chronic kidney disease: detect, diagnose, disclose—a UK primary care perspective of barriers and enablers to effective kidney care. *BMC medicine*, *22*(1), 331.
24. 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.
25. Wilkinson, T. J., Miksza, J., Zaccardi, F., Lawson, C., Nixon, A. C., Young, H. M., ... & Smith, A. C. (2022). Associations between frailty trajectories and cardiovascular, renal, and mortality outcomes in chronic kidney disease. *Journal of cachexia, sarcopenia and muscle*, *13*(5), 2426-2435.
26. Zhu, Y., Bi, D., Saunders, M., & Ji, Y. (2023). Prediction of chronic kidney disease progression using recurrent neural network and electronic health records. *Scientific reports*, *13*(1), 22091.