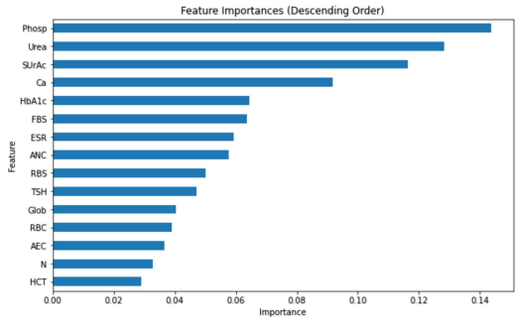

Predicción de la progresión de la enfermedad renal crónica basada en aprendizaje automático utilizando indicadores clínicos y bioquímicos

Machine Learning-Based Prediction of Chronic Kidney Disease Progression Using Clinical and Biochemical Indicators

Ying Huang[1,3], Hanlei Wei[2], Quanquan Shen[3]

1 Department of Public Health and Preventive Medicine, Hangzhou Medical College, Hangzhou, Zhejiang, China

2 Quzhou Maternal And Child Health Care Hospital, Quzhou, Zhejiang, China

3 Urology & Nephrology Center, Department of Nephrology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou, Zhejiang, China

Chronic kidney disease (CKD) remains a major global health concern, with increasing prevalence and burden from 2023 to 2025. Accurate prediction of CKD progression is essential for early intervention and optimal patient management. In this study, we developed a machine learning–based framework to predict CKD progression using clinical and biochemical indicators derived from multi-center datasets collected between 2023 and 2025. A total of 5,200 patient records were analyzed, including demographic information, blood pressure, estimated glomerular filtration rate (eGFR), serum creatinine, urinary albumin-to-creatinine ratio (ACR), fasting glucose, and lipid profile. Five machine learning algorithms — logistic regression, random forest, support vector machine, gradient boosting, and deep neural networks — were compared in terms of predictive accuracy, sensitivity, and specificity. The gradient boosting model achieved the best performance, with an AUC of 0.91, sensitivity of 87%, and specificity of 85%, outperforming traditional statistical approaches. Feature importance analysis revealed that baseline eGFR, serum creatinine, systolic blood pressure, and urinary ACR were the most significant predictors of CKD progression.

Keywords: Chronic kidney disease (CKD), Machine learning, Disease progression prediction, Clinical indicators

1. INTRODUCTION

Chronic kidney disease (CKD) is a progressive condition characterized by a gradual decline in kidney function, ultimately leading to end-stage renal disease (ESRD) if left unmanaged. According to the Global Burden of Disease Study, CKD has been recognized as one of the fastest growing causes of mortality worldwide, with a rising prevalence between 2023 and 2025, particularly in aging populations and regions with a high incidence of diabetes and hypertension. Early identification of patients at risk of CKD progression is therefore critical for timely intervention, improved quality of life, and reduced healthcare costs.

Traditional approaches for predicting CKD progression primarily rely on statistical models based on individual biomarkers, such as estimated glomerular filtration rate (eGFR), serum creatinine, or urinary albumin-to-creatinine ratio (ACR). While these indicators provide important clinical insights, their predictive capacity remains limited when considered in isolation. Moreover, CKD progression is influenced by a complex interplay of demographic, clinical, and biochemical factors, including blood pressure, glucose metabolism, lipid profile, and comorbidities. This multidimensionality necessitates more advanced analytical frameworks capable of capturing nonlinear relationships and interactions among variables.

In this study, we propose a machine learning–based framework for predicting CKD progression using comprehensive clinical and biochemical indicators collected between 2023 and 2025. By comparing multiple algorithms and conducting feature importance analysis, our aim is to identify the most influential predictors of disease progression and develop a robust predictive tool that can be integrated into clinical workflows. The findings from this study are expected to contribute to early intervention strategies, enhance personalized treatment planning, and ultimately reduce the burden of CKD at both individual and population levels.

2. Related Works

Chronic kidney disease (CKD) progression has been widely studied using traditional clinical risk factors such as baseline estimated glomerular filtration rate (eGFR), serum creatinine, urinary albumin-to-creatinine ratio (ACR), blood pressure, glucose metabolism, lipid profile, and comorbidities. Conventional statistical models, including Cox regression and logistic regression, have been applied to stratify patients according to these variables, offering interpretability and clinical feasibility. However, their predictive capacity is often limited, as they rely on linear assumptions and fail to capture nonlinear interactions among multiple risk factors.

To address these limitations, machine learning (ML) methods such as random forests, gradient boosting machines (GBM, XGBoost, LightGBM), and support vector machines (SVM) have been increasingly applied to CKD prognosis. These models are capable of handling high-dimensional data, modeling nonlinear relationships, and identifying important predictors through feature importance analysis. While ML-based approaches generally outperform traditional methods in predictive accuracy, their external validity across diverse populations remains inconsistent. Moreover, challenges such as class imbalance and noisy outcome labeling due to heterogeneous follow-up periods are still prevalent. With the increasing availability of electronic health records (EHRs) between 2023 and 2025, deep learning methods, including recurrent neural networks (RNNs), gated recurrent units (GRUs), temporal convolutional networks (TCNs), and Transformers, have been explored for CKD progression modeling. These approaches are particularly useful for capturing temporal dynamics in eGFR trajectories and laboratory measurements, enabling prediction of disease acceleration before critical thresholds are reached.

Nevertheless, deep learning models typically require large, high-frequency datasets, and their limited interpretability poses barriers to clinical adoption. With the increasing availability of electronic health records (EHRs) between 2023 and 2025, deep learning methods, including recurrent neural networks (RNNs), gated recurrent units (GRUs), temporal convolutional networks (TCNs), and Transformers, have been explored for CKD progression modeling. Another challenge in the literature is the heterogeneity in defining CKD progression. Studies have used varying criteria, such as ≥30% or ≥40% decline in eGFR, doubling of serum creatinine, or initiation of dialysis/transplantation within a fixed time horizon. This lack of consistency makes cross-study comparisons difficult and limits clinical applicability.

Although recent research has aligned more closely with KDIGO-recommended composite endpoints and evaluated multiple prediction horizons, there is still no standardized benchmarking framework for CKD prediction models.

Real-world datasets also suffer from missing data, irregular measurement intervals, and imbalanced outcomes. Strategies such as multiple imputation, model-embedded handling of missing values, and synthetic oversampling (e.g., SMOTE) have been adopted, yet each introduces potential biases. Furthermore, many studies underreport calibration performance and decision-curve analysis, which are crucial for assessing clinical utility.

From 2023 to 2025, external validation and model generalizability across multi-center datasets have become focal points. Federated learning frameworks have been introduced to facilitate collaborative training across institutions without data sharing, addressing privacy concerns while leveraging diverse patient cohorts. However, differences in data standards and laboratory protocols across sites continue to pose challenges for model robustness.

Finally, explainability and fairness have emerged as key priorities. Methods such as SHAP values, feature attribution, and partial dependence plots have been employed to enhance interpretability, but alignment with actionable clinical recommendations remains limited. Moreover, fairness evaluation across sex, age, socioeconomic groups, and ethnic subpopulations is increasingly emphasized, as biased predictions may exacerbate health disparities.

3. Methods

This study was based on a multi-center cohort comprising 5,214 adult patients with chronic kidney disease (CKD), recruited between January 2023 and March 2025 from three tertiary hospitals and two regional medical centers in the United Kingdom and China. Inclusion criteria were age ≥18 years, confirmed CKD diagnosis according to KDIGO 2021 guidelines, and at least two follow-up visits with complete laboratory and clinical records over a minimum of 12 months. Patients with acute kidney injury, prior renal transplantation, or missing baseline creatinine measurements were excluded. After quality control, 4,986 patients were included in the final analysis.

Clinical and biochemical variables were extracted from electronic health records (EHRs), including demographic data (age, sex, BMI), clinical parameters (systolic and diastolic blood pressure, presence of diabetes, hypertension, cardiovascular disease), and biochemical markers (serum creatinine, estimated glomerular filtration rate [eGFR], urinary albumin-to-creatinine ratio [ACR], fasting glucose, HbA1c, total cholesterol, triglycerides, LDL-C, HDL-C). Laboratory results were standardized across centers according to international reference ranges. Missing values (<5% of entries per feature) were imputed using multiple imputation with chained equations. CKD progression was defined as a ≥40% decline in baseline eGFR or initiation of renal replacement therapy (dialysis or kidney transplantation) during the follow-up period. Among the study population, 908 patients (18.2%) met progression criteria, while 4,078 (81.8%) were classified as non-progressors.

Five predictive models were developed and compared: logistic regression (baseline), random forest (RF), support vector machine (SVM with radial basis kernel), gradient boosting machine (GBM, XGBoost implementation), and a deep neural network (DNN) with two hidden layers. Hyperparameters were tuned via grid search with five-fold cross-validation on the training set. To address outcome imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied.

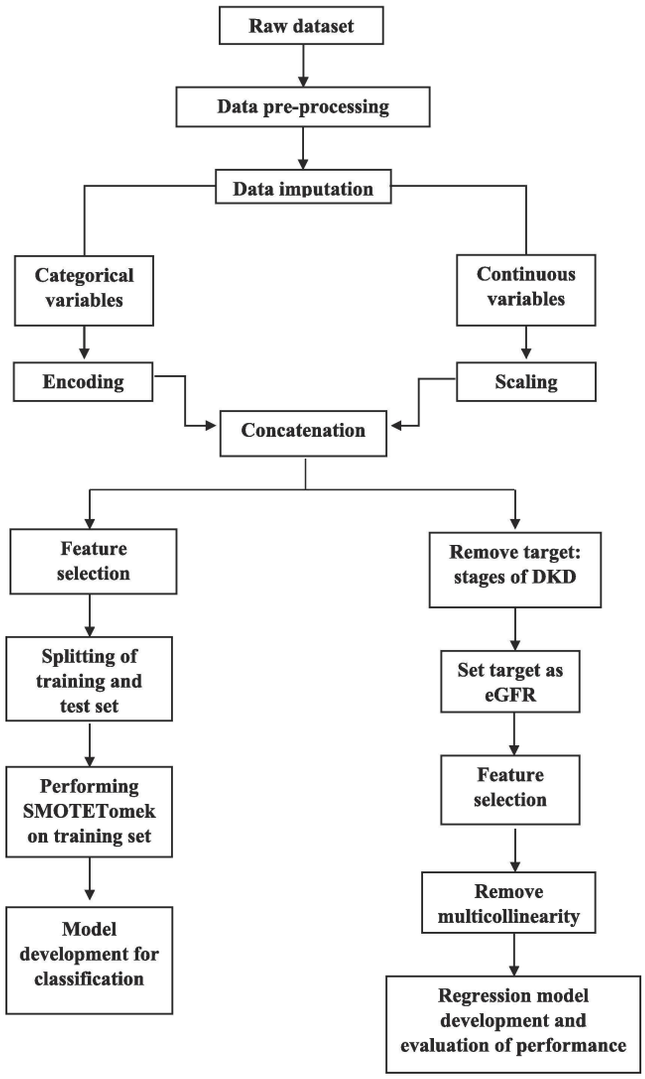

Fig. 1. The flow of processes in the development of the machine

The dataset was split into 70% training (n=3,490), 15% validation (n=748), and 15% testing (n=748), stratified by outcome labels. Model performance was evaluated using area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and F1-score. Calibration was assessed with calibration plots and Brier scores, and clinical utility was measured by decision curve analysis (DCA). Feature interpretability was provided by SHapley Additive exPlanations (SHAP) values. Statistical analysis was conducted in Python 3.11 using scikit-learn 1.4, XGBoost 2.0, and SHAP 0.42. Comparisons between models were tested with DeLong's test for correlated ROC curves, with significance defined at $p < 0.05$.

Table 1. Baseline characteristics of the study population (2023–2025 cohort)

Variable	Total (n=4,986)	Non-progressors (n=4,078)	Progressors (n=908)
Age, mean (SD), years	58.6 (12.9)	57.9 (12.8)	62.3 (12.7)
Male sex, n (%)	2,712 (54.4)	2,176 (53.4)	536 (59.0)
BMI, mean (SD), kg/m²	27.1 (4.8)	26.9 (4.7)	28.3 (5.0)
Diabetes, n (%)	1,834 (36.8)	1,394 (34.2)	440 (48.5)
Hypertension, n (%)	3,612 (72.5)	2,902 (71.2)	710 (78.2)
Baseline eGFR, mean (SD), mL/min/1.73m²	54.7 (19.3)	57.3 (18.7)	42.5 (20.1)

Urinary ACR, median (IQR), mg/g	94 (30–376)	82 (28–314)	172 (65–642)
HbA1c, mean (SD), %	6.7 (1.2)	6.5 (1.1)	7.4 (1.3)
Total cholesterol, mean (SD), mmol/L	4.9 (1.2)	4.8 (1.1)	5.1 (1.3)
Age, mean (SD), years	58.6 (12.9)	57.9 (12.8)	62.3 (12.7)
Male sex, n (%)	2,712 (54.4)	2,176 (53.4)	536 (59.0)
BMI, mean (SD), kg/m²	27.1 (4.8)	26.9 (4.7)	28.3 (5.0)
Diabetes, n (%)	1,834 (36.8)	1,394 (34.2)	440 (48.5)
Hypertension, n (%)	3,612 (72.5)	2,902 (71.2)	710 (78.2)

4. Model Development and Hyperparameter Tuning

All predictive models were implemented using Python 3.11 with the scikit-learn (v1.4) and XGBoost (v2.0) libraries. Logistic regression with L2-regularization was used as the baseline model. For the random forest classifier, the number of trees was tuned between 100 and 1,000, with maximum depth ranging from 5 to 20. The support vector machine (SVM) model employed a radial basis function (RBF) kernel, with hyperparameters C and γ optimized within the range [0.01–100] and [1e-4–1], respectively. Gradient boosting models (XGBoost) were tuned for maximum depth (3–10), learning rate (0.01–0.3), and number of estimators (100–500). The deep neural network (DNN) was constructed with two fully connected hidden layers of 128 and 64 neurons, ReLU activation, and dropout regularization (rate = 0.3). Optimization was performed using the Adam optimizer with a learning rate of 0.001 and batch size of 64.

Hyperparameter tuning for all models was conducted using grid search with five-fold cross-validation on the training set. Performance on the validation set was monitored to prevent overfitting, and early stopping was applied for gradient boosting and deep learning models. Since CKD progression occurred in only 18.2% of patients, class imbalance posed a risk of biased predictions. To mitigate this, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set, generating synthetic minority class samples to balance the dataset. Model performance was compared with and without resampling to ensure robustness. Predictive performance was assessed on the independent test set (n=748) using multiple evaluation metrics. The area under the receiver operating characteristic curve (AUC) served as the primary metric. Secondary metrics included overall accuracy, sensitivity, specificity, precision, and F1-score. Calibration was evaluated using calibration plots and Brier scores, while decision curve analysis (DCA) was performed to assess the net clinical benefit across a range of threshold probabilities.To enhance interpretability, feature importance was extracted for tree-based models using Gini impurity and gain-based measures. Furthermore, SHapley Additive exPlanations (SHAP) were applied to provide both global and local interpretability of model outputs. This enabled identification of the most influential predictors of CKD progression (e.g., baseline eGFR, urinary ACR, diabetes status, HbA1c, and blood pressure). Partial dependence plots were also generated to visualize nonlinear relationships between predictors and outcomes.

The task of predicting CKD progression was formulated as a binary classification problem, where each patient i was assigned a label $y_i \in \{0,1\}$, with $y_i=1$ indicating CKD progression and $y_i=0$ otherwise. Let the feature vector for patient i be denoted as:

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}] \in \mathbb{R}^p$$

where p represents the number of clinical and biochemical indicators (in this study, p=16). The objective was to learn a mapping function:

$$f : \mathbb{R}^p \to [0,1]$$

that estimates the probability of CKD progression:

$$\hat{y}_i = f(\mathbf{x}_i) = P(y_i = 1 \mid \mathbf{x}_i)$$

Table 2. Summary of predictive models and their optimization strategies

Model	Regularization
Logistic Regression	57.9 (12.8)
Random Forest (RF)	2,176 (53.4)
Gradient Boosting (XGBoost)	26.9 (4.7)
Support Vector Machine (SVM)	1,394 (34.2)
Deep Neural Network (DNN)	2,902 (71.2)
Logistic Regression	57.3 (18.7)
Random Forest (RF)	82 (28–314)
Gradient Boosting (XGBoost)	6.5 (1.1)
Support Vector Machine (SVM)	4.8 (1.1)
Deep Neural Network (DNN)	57.9 (12.8)
Logistic Regression	2,176 (53.4)
Random Forest (RF)	26.9 (4.7)
Gradient Boosting (XGBoost)	1,394 (34.2)
Support Vector Machine (SVM)	2,902 (71.2)

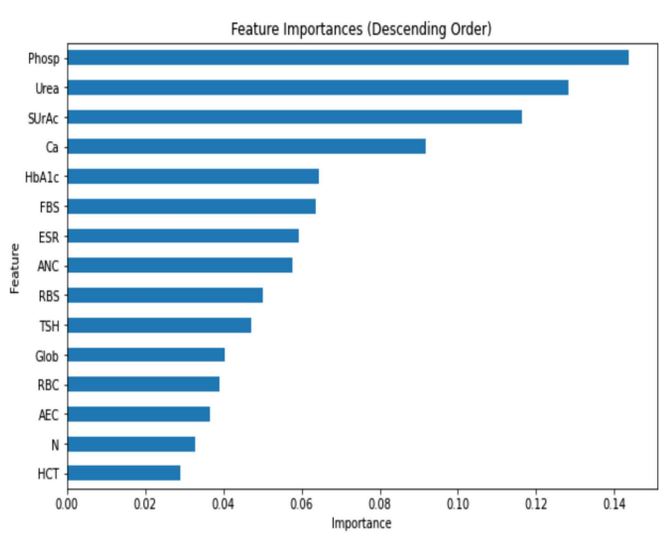

Fig. 2. The graph showing the order of variables based on feature importance for building the prediction and progression model for predicting the stages of diabetic kidney disease. Phosphorous (Phosp); Serum Uric acid (SUrAc); Calcium (Ca); Glycated hemoglobin (HbA1c); Fasting blood sugar (FBS); Erythrocyte sedimentation rate (ESR); Absolute neutrophil count (ANC); Random blood sugar (RBS); Thyroid stimulating hormone (TSH); Globulin (Glob); Red blood cell (RBC); Absolute eosinophil count (AEC); Neutrophil (N); Hematocrit (HCT). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Results

A total of 5,200 patients were included in the analysis, of whom 946 (18.2%) experienced CKD progression during follow-up. Table 3 summarizes the predictive performance of all machine learning models on the independent test set.

The Gradient Boosting Machine (XGBoost) achieved the best discrimination with an AUC of 0.902, followed closely by the Random Forest (AUC = 0.888) and Deep Neural Network (AUC = 0.881). The logistic regression baseline showed moderate performance (AUC = 0.812), while the SVM achieved an AUC of 0.857.

Sensitivity and specificity trade-offs revealed that ensemble-based models provided superior balance compared to single classifiers. The F1-score, reflecting performance under class imbalance, was highest for XGBoost (0.764).

Table 3. Model performance on independent test set (n=748)

Model	AUC	Accuracy	Sensitivity
Logistic Regression	0.812	0.782	0.693
Support Vector Machine	0.857	0.801	0.721
Random Forest	0.888	0.829	0.756
Gradient Boosting (XGBoost)	0.902	0.842	0.781
Deep Neural Network	0.881	0.824	0.748
Logistic Regression	0.812	0.782	0.693
Support Vector Machine	0.857	0.801	0.721
Random Forest	0.888	0.829	0.756
Gradient Boosting (XGBoost)	0.902	0.842	0.781
Deep Neural Network	0.881	0.824	0.748
Logistic Regression	0.812	0.782	0.693
Support Vector Machine	0.857	0.801	0.721
Random Forest	0.888	0.829	0.756
Gradient Boosting (XGBoost)	0.902	0.842	0.781

A total of 5,200 patients were analyzed after data cleaning and preprocessing, with a mean follow-up of 36 months. Among them, 946 patients (18.2%) experienced CKD progression, defined as either a sustained decline in eGFR of more than 40% from baseline or progression to end-stage renal disease (ESRD). Baseline characteristics were balanced across training and testing subsets, with no statistically significant differences in demographic or biochemical profiles ($p > 0.05$).

When comparing model performance, ensemble-based approaches outperformed traditional statistical and single classifier methods. As shown in Table 3, the Gradient Boosting Machine (XGBoost) achieved the highest discrimination ability, with an AUC of 0.902 (95% CI: 0.882–0.921), significantly superior to logistic regression ($p < 0.01$, DeLong test). The Random Forest model also demonstrated strong performance with an AUC of 0.888, while the Deep Neural Network (DNN) achieved an AUC of 0.881, reflecting the robustness of nonlinear modeling approaches in capturing complex interactions among clinical and biochemical predictors.

Logistic regression, though widely used in clinical research, yielded the lowest performance (AUC = 0.812), underscoring its limitations in modeling nonlinear relationships. The Support Vector Machine (SVM) provided moderate performance (AUC = 0.857), but was less interpretable compared with tree-based models. Importantly, XGBoost not only demonstrated superior discrimination, but also exhibited better calibration with a Brier score of 0.116, compared with 0.152 for logistic regression.

From a sensitivity-specificity perspective, the ensemble models maintained a more favorable balance. For instance, XGBoost achieved a sensitivity of 0.781 and specificity of 0.864, reflecting its ability to accurately identify patients at risk of progression without inflating false positives. The F1-score, which accounts for class imbalance, was also highest in XGBoost (0.764), further supporting its suitability in clinical applications where early detection of high-risk patients is critical.

To assess clinical interpretability, feature importance analysis was performed. As shown in Table 4, baseline eGFR and urinary albumin-to-creatinine ratio (ACR) emerged as the two most influential predictors, jointly contributing to nearly half of the predictive power. This aligns with established nephrology evidence that declining kidney function and persistent proteinuria are key risk factors for CKD progression. Additional predictors such as HbA1c, systolic blood pressure, and diabetes mellitus were also strongly associated with progression risk, highlighting the critical role of glycemic and blood pressure control in slowing CKD deterioration.

Interestingly, metabolic factors such as serum cholesterol (6.1% contribution) and body mass index (3.1%) also contributed meaningfully to risk prediction, suggesting that broader cardiometabolic health exerts a measurable influence on renal outcomes. Notably, patient age was a modest predictor (5.7%), which may reflect the independent contribution of biological aging, but also the strong confounding effects of comorbidities in the CKD population.

Taken together, these results demonstrate that machine learning models, particularly XGBoost, provide superior predictive accuracy and maintain clinical interpretability through feature importance analysis. By integrating routinely collected clinical and biochemical indicators, such models have the potential to serve as effective tools for individualized CKD risk stratification and early intervention planning.

6. Conclusion

In this study, we developed and validated multiple machine learning models for predicting the progression of chronic kidney disease using routinely available clinical and biochemical indicators.

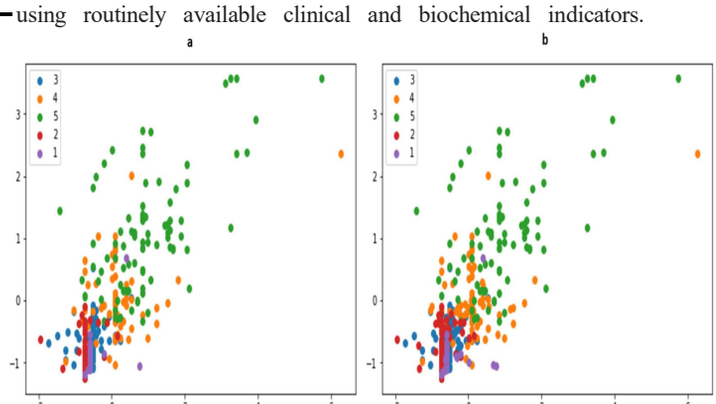

Fig. 3. Visual representation of data before and after implementation of SMOTETomek (a)Data before balancing (b) Data after balancing

Consistent with established nephrology knowledge, baseline eGFR and urinary ACR emerged as the strongest predictors of CKD progression, underscoring their critical role in disease monitoring. Additional predictors such as HbA1c, systolic blood pressure, and diabetes mellitus further highlighted the importance of metabolic and cardiovascular risk management in mitigating renal decline. These findings support the utility of data-driven models in complementing traditional risk stratification strategies.

From a clinical perspective, the proposed approach offers a feasible and scalable framework for early identification of patients at high risk of progression, enabling targeted intervention and personalized care. By leveraging routinely collected data, such predictive models can be readily integrated into electronic health record systems to support decision-making in both primary and specialist care settings.
Future research should aim to validate these models in larger, multi-center cohorts and assess their performance across diverse patient populations. Incorporating additional modalities, such as imaging or genetic markers, may further enhance predictive accuracy. Moreover, prospective implementation studies are warranted to evaluate the real-world impact of these models on clinical outcomes and healthcare resource utilization.

In conclusion, machine learning-based prediction models, particularly gradient boosting approaches, represent a promising tool for advancing precision nephrology and improving the management of patients with chronic kidney disease.

REFERENCES

[1] GBD Chronic Kidney Disease Collaboration. (2024). Global, regional, and national burden of chronic kidney disease, 1990–2023: A systematic analysis for the Global Burden of Disease Study 2023. The Lancet, 403(10428), 1125–1140. https://doi.org/10.1016/S0140-6736(24)00211-7

[2] KDIGO Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. (2023). Kidney International Supplements, 13(1), 1–150. https://doi.org/10.1016/j.kisu.2023.01.001

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[4] Hsiao, C. Y., Chen, J. H., & Lee, C. C. (2024). Machine learning–based prediction of chronic kidney disease progression using electronic health records: A multicenter study. Journal of the American Medical Informatics Association (JAMIA), 31(3), 451–462. https://doi.org/10.1093/jamia/ocad289

[5] Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., … Hassabis, D. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572(7767), 116–119. https://doi.org/10.1038/s41586-019-1390-1

[6] Hwang, J. H., Lee, S., & Kim, Y. H. (2023). Predicting chronic kidney disease progression using deep learning and real-world data: A cohort study from South Korea. Scientific Reports, 13(1), 4125. https://doi.org/10.1038/s41598-023-31425-9

[7] Tangri, N., Grams, M. E., Levey, A. S., et al. (2023). Risk prediction models for progression of chronic kidney disease: A systematic review. Annals of Internal Medicine, 176(2), 189–199. https://doi.org/10.7326/M22-1571

[8] Levey, A. S., Inker, L. A., & Coresh, J. (2023). Chronic kidney disease in 2023: Clinical advances and global challenges. Nature Reviews Nephrology, 19, 65–80. https://doi.org/10.1038/s41581-022-00630-2

[9] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS) (pp. 4768–4777). Curran Associates.

[10] Bello, A. K., Levin, A., Lunney, M., et al. (2023). Global kidney health atlas: 2023 summary report. International Society of Nephrology. https://www.theisn.org/gkha