Machine Learning-Based Prediction of 30-Day Hospital Readmission in Patients with Chronic Kidney Disease: A Retrospective Cohort Study

Houshang Sanadgol,[1] Sedigheh Barati,[1*#] Nasibeh Rady Raz[2,3*#]

[1]Division of Nephrology, Department of Internal Medicine, Hasheminejad Kidney Center, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
[2]Department of Artificial Intelligence in Medicine, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran
[3]Breast Cancer Research Center, Iran University of Medical Sciences (IUMS), Tehran, Iran

[#]Sedigheh Barati and Nasibeh Rady Raz equally contributed to this work.

Keywords. Chronic kidney disease; Hospital readmission; Machine learning; XGBoost; Predictive modeling; Nephrology; Catheter infection; eGFR

Introduction. Chronic Kidney Disease (CKD) is associated with high 30-day hospital readmission rates due to progressive renal dysfunction, multiple comorbidities, and complications related to dialysis and catheter use. Artificial Intelligence (AI) and Machine Learning (ML) offer promising tools for early identification of high-risk patients. To develop and evaluate ML models for predicting 30-day hospital readmission among CKD patients and identify key clinical and laboratory predictors related to readmission risk.
Methods. This retrospective study analyzed 277 hospitalized patients with CKD at Hasheminejad Kidney Center (2019–2022). Forty-four demographic, clinical, and laboratory features were included. Preprocessing included handling missing data, normalization, outlier removal, categorical encoding, and oversampling. Six ML models, including eXtreme Gradient Boosting (XGBoost), Random Forest, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Logistic Regression, were trained using a 70/30 train-test split with cross-validation. Feature selection employed SHAP values, mutual information, F-values, SVM, and chi-squared tests.
Results. XGBoost outperformed other models (accuracy > 90%). The strongest predictors were estimated Glomerular Filtration Rate (eGFR), Blood Urea Nitrogen (BUN) and creatinine levels, age, presence of diabetes mellitus and hypertension, catheter-related infection, and triglycerides, intact Parathyroid hormone (iPTH), and albumin levels. Catheter infection emerged as a modifiable, high-impact predictor. The SHAP values analysis confirmed strong contributions of kidney function markers, inflammatory indicators, and metabolic variables to re-admission risk.
Conclusion. ML-based prediction models, particularly XGBoost, demonstrated high accuracy in identifying CKD patients at risk of 30-day readmission. Integration of these models into clinical workflows may improve early intervention, reduce hospital readmissions, and support evidence-based nephrology care.

IJKD 2026;20:69-76
www.ijkd.org

INTRODUCTION

Chronic Kidney Disease (CKD) affects more than 850 million individuals globally, constituting a significant cause of morbidity, mortality, and healthcare expenditure. With a global prevalence of 11–13%, CKD is most frequently associated with diabetes mellitus and hypertension.[1,2] Due to the progressive nature of renal dysfunction, multiple comorbidities, recurrent infections, and dialysis-related complications, CKD patients experience

disproportionately high rates of unplanned hospital readmissions. Thirty-day readmission rates often exceed 15–30%, reflecting significant clinical instability.[3,4]

The integration of AI into healthcare has significantly enhanced the potential to improve patient care and predict clinical outcomes.[5-7] Specifically, the application of machine learning (ML) and deep learning (DL) algorithms, particularly recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, has proven effective for processing temporal clinical data and forecasting future events from historical patient records.[8] Furthermore, the Binary Grey Wolf Optimization method is used for feature subset selection to improve the accuracy of CKD prediction.[9] Also, ML algorithms and the Internet of Things (IoT) are used for alarming and predicting the requirement for blood transfusion, considering patient variability.[10]

This study aimed to develop ML-based predictive models for hospital readmission in patients with CKD, evaluate their performance, and identify the strongest predictors of readmission risk.

MATERIALS AND METHODS

Study Design and Setting

A retrospective cohort study was conducted at Hasheminejad Kidney Center in Tehran, Iran. Eligible patients were adults (> 18 years old) with CKD who were hospitalized and subsequently readmitted within 30 days between 2019 and 2022. The minimum sample size was determined using G*Power version 3.1.9.2, applying a chi-square goodness-of-fit test with the following parameters: alpha = 0.05, power = 0.8, and effect size = 0.3, yielding a sample size of 100 participants.

Participants

A total of 285 patients were screened; after excluding those with incomplete records, 277 were included. Data were anonymized and approved by the institutional ethics committee.

Data Collection

A total of 44 variables were extracted from medical records:

Demographic and clinical variables

- Age, sex
- CKD stage (I–V)
- Diabetes mellitus
- Hypertension
- Heart failure
- Catheter presence, catheter infection, catheter dysfunction
- Glomerulonephritis, polycystic kidney disease
- Dialysis treatment
- Urinary tract infection (UTI) and respiratory infection
- Surgical and transplant-related data

Laboratory variables

- Renal markers: Estimated Glomerular Filtration Rate (eGFR), Blood Urea Nitrogen (BUN), and creatinine levels
- Hematologic indices: Hemoglobin (Hb) level, White Blood Count (WBC), platelets
- Inflammatory markers: C-reactive protein (CRP), erythrocyte sedimentation rate (ESR)
- Electrolytes: serum sodium (Na), potassium (K), calcium (Ca), and phosphate (P) levels.
- Albumin, intact Parathyroid Hormone (iPTH), lipid profile (cholesterol, triglycerides)
- Urine RBC, WBC, protein

Preprocessing

Variables with > 90% missing data were removed. Then, median imputation was performed for missing continuous variables. Moreover, the encoding process for categorical variables, including gender, CKD stages, and 13 comorbidities, was done. Standardization was applied to normalize the distribution and prevent dominant features with large values. Outlier detection and removal were performed. To mitigate potential learning bias arising from data limitations and class imbalance, data augmentation via oversampling was employed. It is noted that imputation, standardization, outlier removal, and oversampling were all applied after the train-test split to avoid data leakage.

Feature Selection

Five machine learning methods were applied to identify significant factors affecting readmissions in patients with CKD and to address overfitting in imbalanced data, including SHAP (SHAPley Additive exPlanations) values, Mutual Information, F-statistics, Support Vector Machine (SVM) ranking, and the Chi-Squared test. Note that feature selection is performed after the train-test split to avoid data leakage.

Machine Learning Models

Readmission prediction was performed using six ML algorithms commonly employed in predictive models for their diverse learning approaches: the eXtreme Gradient Boosting (XGBoost), Random Forest, Decision Tree, AdaBoost, Multilayer Perceptron (MLP), and Logistic Regression. XGBoost is a supervised learning boosting algorithm that uses gradient-based optimization. This ML algorithm is practical, given its properties of quickness, efficiency, and scalability with large datasets. Additionally, the Random Forest is an ML algorithm that uses multiple decision trees, collectively referred to as a "forest," to make predictions.[11] Moreover, a decision tree is a recursive partitioning-based decision system that uses a tree-like structure for decision-making. Also, AdaBoost is a boosting technique that sequentially combines several weak classifiers to produce a strong classifier. Furthermore, the MLP is an artificial neural network with multiple layers of neurons arranged in a hierarchical structure. Finally, logistic regression is a supervised ML algorithm used for classification problems.

Each algorithm's performance was evaluated by its ability to predict readmission, while accounting for clinical parameters such as catheter use, CKD stage, diabetes mellitus, hypertension, polycystic kidney disease, UTI, and respiratory infection. For implementation, Python 3 was run on an Intel® Xeon® CPU operating at 2.30 GHz, with 12 GB of RAM and 100 GB of disk space. The data were split into training and test sets in a 70%:30% ratio to facilitate model evaluation.

Evaluation Metrics

Model evaluation was performed using accuracy, Precision, Recall, F1-score, ROC-AUC, 5-fold cross-validation, and nested cross-validation. The accuracy of the selected variables was calculated using the formula Accuracy = (true positives + true negatives) / (true positives + true negatives + false positives + false negatives). Table 1 and Figure 1 show the results.

RESULTS

Patient Characteristics

The cohort was first statistically analyzed as follows:

- Mean age: 52 years
- 60% male
- 63% older than 60 years

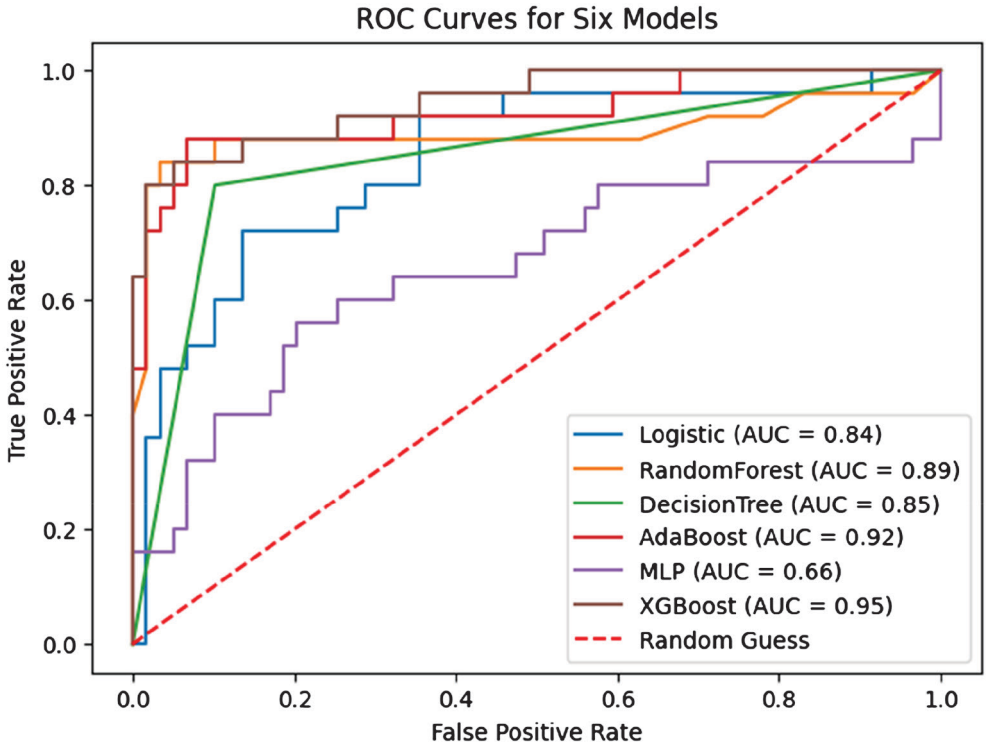

Figure 1. The Comparison of the ROC of the ML Algorithms.

Table 1. Comparing CKD Patient Readmission Prediction Model Performance Evaluation using ML Algorithms for each Patient Case. [Chronic kidney disease (CKD); multi-layer perception (MLP); hypertension (HTN); urinary tract infection (UTI).]

Patient Case	XGBoost				Random Forest				Decision Tree				AdaBoost				MLP				Logistic Regression			
	Accuracy	precision	recall	f1-score	Accuracy	precision	recall	f1-score	Accuracy	precision	recall	f1-score	Accuracy	precision	recall	f1-score	Accuracy	precision	recall	f1-score	Accuracy	precision	recall	f1-score
Catheter	0.94	0.98	0.92	0.94	0.92	0.97	0.87	0.91	0.85	0.88	0.82	0.85	0.92	0.97	0.87	0.91	0.94	0.97	0.92	0.94	0.92	0.97	0.87	0.91
CKD Stage	0.99	0.98	0.98	0.98	0.90	0.94	0.90	0.91	0.99	0.99	0.99	0.99	0.24	0.99	0.24	0.35	0.59	0.62	0.59	0.60	0.86	0.85	0.86	0.85
Diabetic	0.70	0.81	0.75	0.78	0.77	0.93	0.78	0.85	0.69	0.74	0.79	0.76	0.74	0.82	0.80	0.81	0.64	0.77	0.72	0.75	0.68	0.79	0.75	0.77
HTN	0.60	0.86	0.61	0.72	0.60	0.74	0.64	0.69	0.57	0.70	0.62	0.66	0.55	0.64	0.62	0.63	0.56	0.56	0.65	0.60	0.61	0.66	0.67	0.67
Polycystic kidney	0.95	1.00	0.95	0.98	0.95	1.00	0.95	0.98	0.88	0.93	0.95	0.94	0.92	0.96	0.95	0.96	0.95	1.00	0.95	0.98	0.94	0.99	0.95	0.97
UTI	0.94	0.99	0.94	0.96	0.94	1.00	0.94	0.97	0.88	0.92	0.95	0.94	0.93	0.99	0.94	0.96	0.94	1.00	0.94	0.97	0.93	0.97	0.95	0.96
Respiratory infection	0.90	1.00	0.89	0.94	0.89	1.00	0.89	0.94	0.81	0.91	0.88	0.89	0.88	0.99	0.89	0.94	0.89	1.00	0.89	0.94	0.88	0.99	0.89	0.94

- 49.5% at CKD stage V
- 39% diabetic
- 41% hypertensive
- 19% catheter infection
- 29% undergoing dialysis

Key Predictors

To analyze relationships among variables, correlation coefficients and *P*-values were first calculated. Results revealed correlations and predictive effects between patients' previous and current admission values for Cr, Hb, BUN, PLT, and WBC. Additionally, significant correlations were observed between Cr and BUN, iPTH and Cr, CRP and ESR, urine WBC and RBC, and urine RBC and protein. During feature engineering, when variables show high correlations (0.8-1.0), one of the two is eliminated to prevent overfitting.

Feature importance was assessed using machine learning algorithms, including SHAP values, mutual information, F-values, SVMs, and Chi-squared tests, to identify the most effective predictors of readmission. Feature importance algorithms identified the most influential parameters associated with readmission among patients with CKD across stages. When XGBoost is used as the primary predictive model, eGFR, BUN, and creatinine are the most important predictors. Specifically, eGFR was the most predictive factor for readmission in patients with CKD. Across all feature-selection methods, the top predictors were:

- eGFR (primary predictor)
- BUN and creatinine levels
- Age
- Diabetes mellitus and hypertension
- Catheter infection and dysfunction
- Albumin and iPTH levels
- Triglyceride levels

Model Performance

Among artificial intelligence algorithms used to predict the most effective variables associated with hospital readmission in patients with CKD, XGBoost and RF generally outperformed other algorithms. For readmissions with diabetes mellitus and HTN, the model requires more specific parameters to achieve better performance.

Additionally, the model's performance at the "CKD stage" is compared with that of other features, including catheter, diabetes mellitus, HTN, polycystic kidney, and UTI. Therefore, the separate results of each of the five CKD stages are not the aim of the research. Thus, the average results across the five CKD stages are reported in Table 1. However, the results of model performance analysis by CKD stage indicate that the model performs better in patients with late-stage CKD, as the dataset contains more in patients with late-stage than those with early-stage CKD.

Furthermore, 5-fold cross-validation (CV) and nested 5-fold CV were used to evaluate the generalization and performance of the proposed model. Moreover, the Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) values of the algorithms were compared in Figure 1. The model curve, which is closer to the top-left corner and has a higher AUC value, performs better. Therefore, the results indicate that XGBoost outperforms the other models.

In conclusion, XGBoost performed best:

- Accuracy > 90% (Table 1)
- High ROC-AUC (Figure 1)
- Best generalizability in nested CV (Figure 2)

Random Forest and MLP showed moderate-to-strong performance. Logistic Regression and AdaBoost performed modestly. Decision Trees showed the weakest performance.

SHAP Values Analysis

Model interpretability was evaluated using SHAP values in Figure 3 for the readmission model for patients with catheters. The Y-axis shows the feature's global importance, and the X-axis shows the SHAP value for a prediction. A feature with greater impact on the model is shown in red, and a feature with lower impact is shown in blue.

SHAP values plot revealed:

- Increasing risk with catheter infection, low eGFR, high BUN/Cr, advanced age, diabetes mellitus, inflammation (CRP/ESR), and metabolic abnormalities.
- Higher albumin and higher eGFR were protective.

DISCUSSION

Chronic kidney disease represents a significant global health challenge, accompanied with high morbidity and substantial economic burden. Hospital readmissions among CKD patients are frequent and costly, arising from a complex

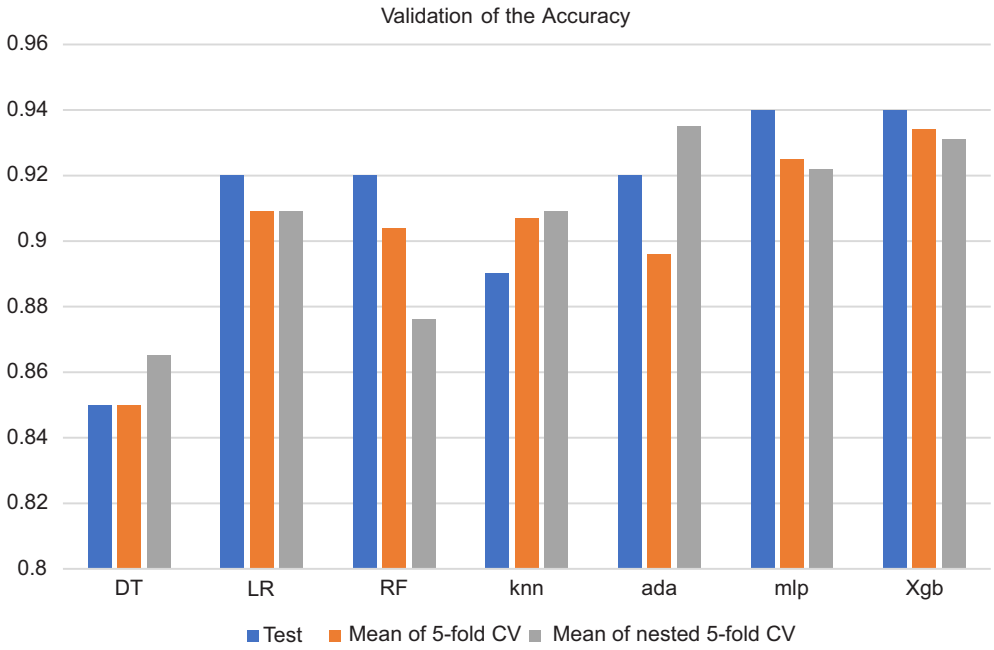

Figure 2. Comparison of the Accuracy of the Proposed Method using the mean of 5-fold CV and nested 5-fold CV.

interplay of clinical, systemic, and social factors.[3,11] The current study focused on hospitalized patients with chronic kidney disease. It employed advanced artificial intelligence algorithms to identify factors associated with readmission and to assess the predictive performance of these models.

The findings underscored the critical roles of catheter use, eGFR, BUN, Cr, diabetes, age, triglycerides (TG), and catheter infections in predicting hospital readmission rates. Additionally, it identified chronic conditions such as diabetes, HTN, and heart failure as the most prevalent comorbidities among these patients.

Moreover, the findings revealed a significant association between inflammatory markers (CRP and ESR) and hematological parameters (WBC, Hb, and PLT) in relation to kidney injury and CKD progression. Markers such as CRP and ESR reflect systemic inflammation, which exacerbates the decline in kidney function and increases the risk of complications. In the current study, feature-importance algorithms identified critical predictors of hospital readmission, including catheter use, eGFR, BUN, Cr, and TG levels, diabetes, increasing age, and catheter-related infections. Diabetes mellitus, hypertension, and heart failure were strongly associated with readmission, whereas BUN and Cr levels showed the strongest correlations. Predicting CKD progression in diabetic patients remains a challenge, with studies highlighting the impact of older age, HTN, heart disease, elevated serum Cr, and reduced eGFR on CKD risk.[12]

Overall, the current study demonstrates the effectiveness of the XGBoost algorithm in accurately predicting hospital readmissions among patients with CKD. The results emphasize the critical role of key clinical parameters, specifically catheter use, eGFR, BUN, Cr, and TG levels, diabetes, increasing age, and catheter-related infections as essential predictors of readmission risk. This highlights the potential to integrate these clinical indicators with advanced artificial intelligence-based prediction models to improve the identification of at-risk patients significantly. Such integration can empower healthcare providers to implement targeted preventive interventions, thereby improving patient outcomes and reducing the burden of hospital readmissions in this vulnerable population.

Currently, physicians make subjective decisions about readmission based on their skills and experience. The proposed AI-based readmission system leverages electronic health record (EHR) data and enables real-time predictions that align with clinical workflows; therefore, it is particularly beneficial for young doctors with limited experience. Also, in complex cases, the AI-based system thoroughly considers all available data to predict readmission. Overall, this study

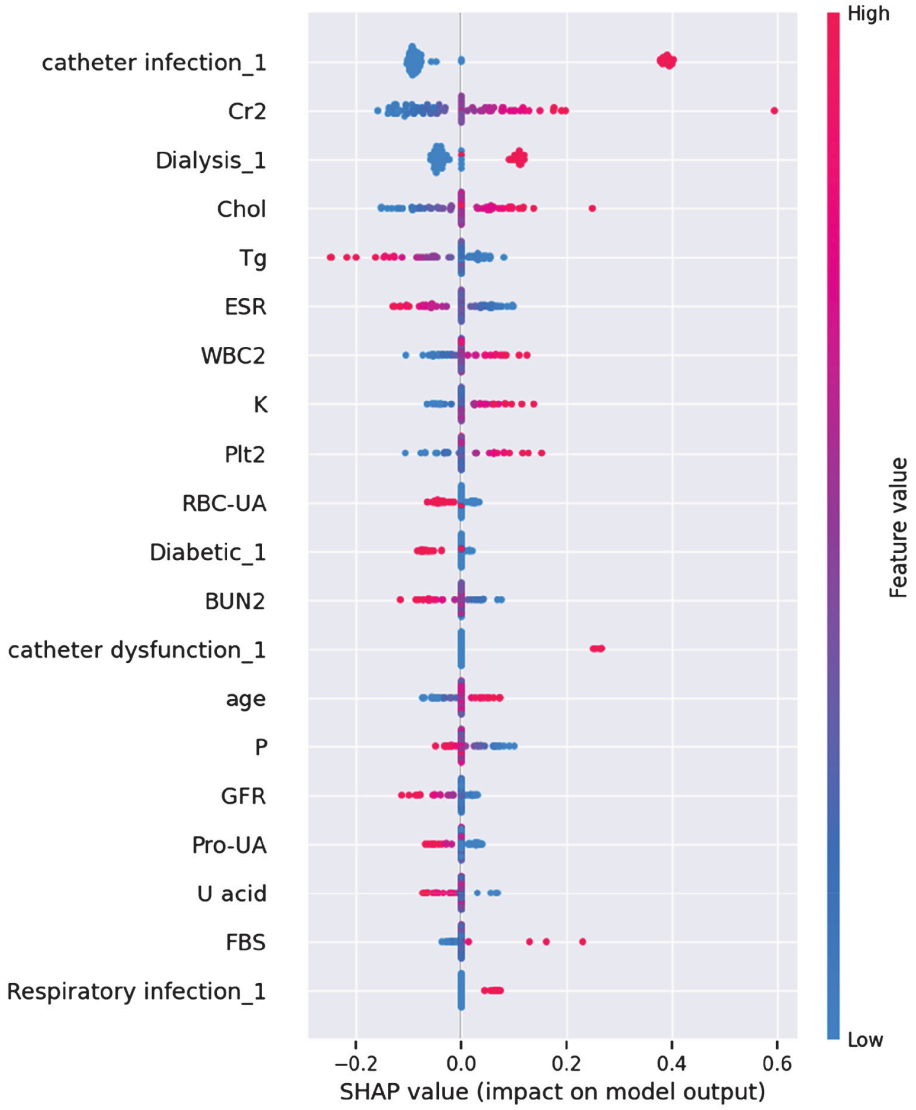

Figure 3. SHAP values.

provides valuable insights into the application of AI in CKD management, suggesting pathways for future research and clinical practice to optimize care for these patients.

Study Limitations

- Single-center retrospective design
- Limited sample size
- Absence of socioeconomic and medication-related variables
- Lack of external validation on other populations

CONCLUSION

Hospital readmissions in chronic kidney disease (CKD) impose a substantial burden on healthcare systems, underscoring the need for reliable risk stratification. In this study, we developed and validated a predictive model for 30-day readmissions using XGBoost, with performance assessed through standard and nested 5-fold cross-validation and compared with other machine learning approaches, where XGBoost showed the best performance. Model interpretability was assessed using SHAP analysis. The findings underscore the importance of key clinical parameters, specifically catheter use, eGFR, BUN, Cr, and triglyceride levels, diabetes, age, and catheter infections, as critical predictors of readmission risk. This suggests that integrating these clinical indicators with advanced AI-based prediction models can significantly enhance the

identification of patients at risk. Further work should focus on prospective validation and clinical implementation to support targeted interventions and reduce readmissions in this population.

HIGHLIGHTS

- ML can accurately predict 30-day hospital readmission in CKD patients.
- XGBoost achieved the highest predictive performance (> 90% accuracy).
- eGFR, BUN, and creatinine were the strongest predictors.
- Catheter infection is a major modifiable risk factor.
- ML integration can improve early intervention and reduce readmissions.

REFERENCES

1. Francis A, Harhay MN, Ong ACM, Tummalapalli SL, Ortiz A, Fogo AB, et al. Chronic kidney disease and the global public health agenda: an international consensus. Nat Rev Nephrol. 2024 Jul;20(7):473–85. doi:10.1038/s41581-024-00820-6 PubMed PMID: 38570631.
2. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. Kidney Int Suppl. 2022 Apr;12(1):7–11. doi:10.1016/j.kisu.2021.11.003 PubMed PMID: 35529086; PubMed Central PMCID: PMC9073222.
3. Surasura A, Siva Krishna Pavan Kumar BG, Chinamanagonda S, Durga D, Gubbala S. Understanding Hospital Readmissions: Insights, Patterns, and Interventions for Improvement in Chronic Kidney Disease. Cureus. 2024 May;16(5):e59524. doi:10.7759/cureus.59524 PubMed PMID: 38827006; PubMed Central PMCID: PMC11144055.
4. Hickson LJ, Thorsteinsdottir B, Ramar P, Reinalda MS, Crowson CS, Williams AW, et al. Hospital Readmission among New Dialysis Patients Associated with Young Age and Poor Functional Status. Nephron. 2018;139(1):1–12. doi:10.1159/000485985 PubMed PMID: 29402792; PubMed Central PMCID: PMC5920698.
5. Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. Bioengineering. 2024 Mar 29;11(4):337. doi:10.3390/bioengineering11040337 PubMed PMID: 38671759; PubMed Central PMCID: PMC11047988.
6. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022 Apr 8;12(1):5979. doi:10.1038/s41598-022-09954-8 PubMed PMID: 35395867; PubMed Central PMCID: PMC8993826.
7. Faraji N, Aali S, Motiei M, Sadat Mansouri S. Letter to the Editor: ChatGPT's contribution to cancer study. J Cancer Res Clin Oncol. 2023 Nov;149(14):13495–500. doi:10.1007/s00432-023-05183-2 PubMed PMID: 37462773; PubMed Central PMCID: PMC11797205.
8. Cissoko MBH, Castelain V, Lachiche N. Multi-Way adaptive Time Aware LSTM for irregularly collected sequential ICU data. Expert Syst Appl. 2025 Feb 1;261:125548. doi:10.1016/j.eswa.2024.125548
9. BilalA, Alzahrani A, Almuhaimeed A, Khan AH, Ahmad Z, Long H. Advanced CKD detectionthrough optimized metaheuristic modeling in healthcare informatics. Sci Rep.2024 Jun 1;14(1):12601. doi:10.1038/s41598-024-63292-5
10. Rady Raz N, Anoushirvani AA, Rahimian N, Ghoerishi M, Alibeik N, Sajadi Rad M. Nano fuzzy alarming system for blood transfusion requirement detection in cancer using deep learning. Sci Rep. 2024 Jul 10;14(1):15958. doi:10.1038/s41598-024-66607-8 PubMed PMID: 38987580; PubMed Central PMCID: PMC11237028.
11. Jha V, Al-Ghamdi SMG, Li G, Wu MS, Stafylas P, Retat L, et al. Global Economic Burden Associated with Chronic Kidney Disease: A Pragmatic Review of Medical Costs for the Inside CKD Research Programme. Adv Ther. 2023;40(10):4405–20. doi:10.1007/s12325-023-02608-9 PubMed PMID: 37493856; PubMed Central PMCID: PMC10499937.
12. Jairoun AA, Ping CC, Ibrahim B. Predictors of chronic kidney disease survival in type 2 diabetes: a 12-year retrospective cohort study utilizing estimated glomerular filtration rate. Sci Rep. 2024 Apr 19;14(1):9014. doi:10.1038/s41598-024-58574-x PubMed PMID: 38641627; PubMed Central PMCID: PMC11031608.

*Correspondence to:
Sedigheh Baratir,
Division of Nephrology, Department of Internal Medicine, Hasheminejad Kidney Center, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
E-mail: s.barati50@gmail.com

Nasibeh Rady Raz,
Department of Artificial Intelligence in Medicine, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran
E-mail: radyraz.n@iums.ac.ir

Received December 2025
Revised February 2026
Accepted February 2026