



# A machine learning-based approach for predicting renal function recovery in general ward patients with acute kidney injury

Nam-Jun Cho<sup>1,\*</sup>, Inyong Jeong<sup>2,\*</sup>, Yeongmin Kim<sup>2</sup>, Dong Ok Kim<sup>2</sup>, Se-Jin Ahn<sup>2</sup>, Sang-Hee Kang<sup>3</sup>, Hyo-Wook Gil<sup>1</sup>, Hwamin Lee<sup>2</sup>

<sup>1</sup>Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Republic of Korea

<sup>2</sup>Department of Medical Informatics, Korea University College of Medicine, Seoul, Republic of Korea

<sup>3</sup>Department of Surgery, Korea University Guro Hospital, Seoul, Republic of Korea

**Background:** Acute kidney injury (AKI) is a significant challenge in healthcare. While there are considerable researches dedicated to AKI patients, a crucial factor in their renal function recovery, is often overlooked. Thus, our study aims to address this issue through the development of a machine learning model to predict restoration of kidney function in patients with AKI.

**Methods:** Our study encompassed data from 350,345 cases, derived from three hospitals. AKI was classified in accordance with the Kidney Disease: Improving Global Outcomes. Criteria for recovery were established as either a 33% decrease in serum creatinine levels at AKI onset, which was initially employed for the diagnosis of AKI. We employed various machine learning models, selecting 43 pertinent features for analysis.

**Results:** Our analysis contained 7,041 and 2,929 patients' data from internal cohort and external cohort respectively. The Categorical Boosting Model demonstrated significant predictive accuracy, as evidenced by an internal area under the receiver operating characteristic (AUROC) of 0.7860, and an external AUROC score of 0.7316, thereby confirming its robustness in predictive performance. SHapley Additive exPlanations (SHAP) values were employed to explain key factors impacting recovery of renal function in AKI patients.

**Conclusion:** This study presented a machine learning approach for predicting renal function recovery in patients with AKI. The model performance was assessed across distinct hospital settings, which revealed its efficacy. Although the model exhibited favorable outcomes, the necessity for further enhancements and the incorporation of more diverse datasets is imperative for its application in real-world.

**Keywords:** Acute kidney injury, Hospital records, Machine learning, Recovery of function, Creatinine

## Introduction

Despite treatment advancements in recent years, acute

kidney injury (AKI) remains a significant concern in the medical field. AKI independently contributes to the escalation of healthcare costs and prolongation of hospitalization

**Received:** December 4, 2023; **Revised:** February 27, 2024; **Accepted:** April 23, 2024

**Correspondence:** Hwamin Lee

Department of Medical Informatics, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea.

E-mail: hwamin@korea.ac.kr

ORCID: <https://orcid.org/0000-0002-6482-3511>

\*Nam-Jun Cho and Inyong Jeong contributed equally to this study as co-first authors.

© 2024 by The Korean Society of Nephrology

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial and No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits unrestricted non-commercial use, distribution of the material without any modifications, and reproduction in any medium, provided the original works properly cited.

periods, while also elevating the incidence of in-hospital complications and mortality rates [1–3]. This condition is recognized as one of the most prevalent diseases, exhibiting incidence rates of 10% to 15% in general hospital admissions and escalating to 50% to 60% within intensive care unit (ICU) settings. The duration of renal function recovery is increasingly recognized as a pivotal factor in predicting patient outcomes [4]. Studies have shown a correlation between prolonged AKI and heightened risks of complications and mortality [5,6].

Nevertheless, the crucial role of renal function recovery in the prognosis of patients with AKI has been largely overlooked [7]. Consequently, this research area experiences a significant dearth of research, necessitating the need for new investigations. Prior efforts to predict renal function recovery have been hindered by limitations, such as small sample sizes [8–10], exclusive focus on ICU patients, and the absence of an all-encompassing definition for renal function recovery, which impedes the application of these research outcomes in clinical settings [11–15]. Therefore, this study aimed to fill this gap by developing a machine learning-based approach that includes validation in external settings, aimed to predict renal function recovery in patients with AKI, with a particular focus on patients in general wards.

## Methods

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. The Institutional Review Boards (IRBs) of Soonchunhyang University Cheonan Hospital and Korea University Anam Hospital and Guro Hospital approved the study protocol (No. 2019-10-023, 2023AN0145, and 2023GR0425). The need for informed consent was waived by the IRB as the current study was a retrospective review of anonymized clinical data.

### Study population

For model development, patient datasets were extracted from the Korea University Anam Hospital (Hospital A) and Guro Hospital (Hospital B) from January 1, 2015, to December 31, 2021. These datasets were used as an internal cohort. Additional datasets were extracted from Soonchunhyang University Cheonan Hospital (Hospital C),

specifically from patients who were admitted to the general wards from March 1, 2016, to March 31, 2021 for an external cohort. We only considered admissions for individuals aged 19 years and older. Using these data, we constructed a retrospective cohort and applied the following exclusion criteria.

- 1) Patients with a hospital stay of less than 24 hours.
- 2) Patients with no blood pressure measurement recorded within 24 hours of admission or with fewer than two measurements during their hospital stay.
- 3) Patients without serum creatinine (Cr) measurement or estimated glomerular filtration rate (eGFR) of less than 60 mL/min/1.73 m<sup>2</sup> on the first day.

Patients with an eGFR of less than 60 mL/min/1.73 m<sup>2</sup> were excluded from this research due to the unavailability of comprehensive blood test results prior to hospitalization, which precluded accurate assessment of baseline renal function in these individuals. In other words, the eGFR at the time of hospital admission could represent the patient's baseline renal function or the eGFR following the onset of AKI. To exclude such ambiguity and focus exclusively on in-hospital AKI, only patients with an eGFR of 60 mL/min/1.73 m<sup>2</sup> or above upon admission were included in this study, indicating relatively preserved renal function.

### Acute kidney injury definition

AKI was defined according to the KDIGO (Kidney Disease: Improving Global Outcomes) guidelines. AKI was diagnosed based on the following criteria.

- 1) An increase in Cr by 0.3 mg/dL (or 26.5 μmol/L) or more within 48 hours. The baseline value is defined as the lowest Cr value measured in the preceding 2 days.
- 2) A Cr level that is 1.5 times or more and the baseline. The baseline value is defined as the lowest Cr value measured in the preceding 7 days.

Baseline Cr was determined as the lowest Cr value measured in the preceding 2 or 7 days. If no Cr value was measured for definition 2, the most recent measurement within the last 180 days was used as the baseline. Urinary output criteria were not utilized for AKI diagnosis because of the predominance of missing data in the patient records.

### Acute kidney injury recovery definition

To define AKI recovery, we referred to previous studies [4,16] and used the following criteria; AKI recovery was de-

terminated based on the following two criteria for Cr:

- 1) A decrease of 33% or more compared to the Cr at the time of AKI onset.
- 2) A decrease below the baseline used for AKI diagnosis.

We established a recovery period of 7 days to evaluate AKI recovery. If recovery does not occur within this period, the patient was categorized as suffering from acute kidney disease rather instead of AKI [17].

### Cohort organization and outcome labeling

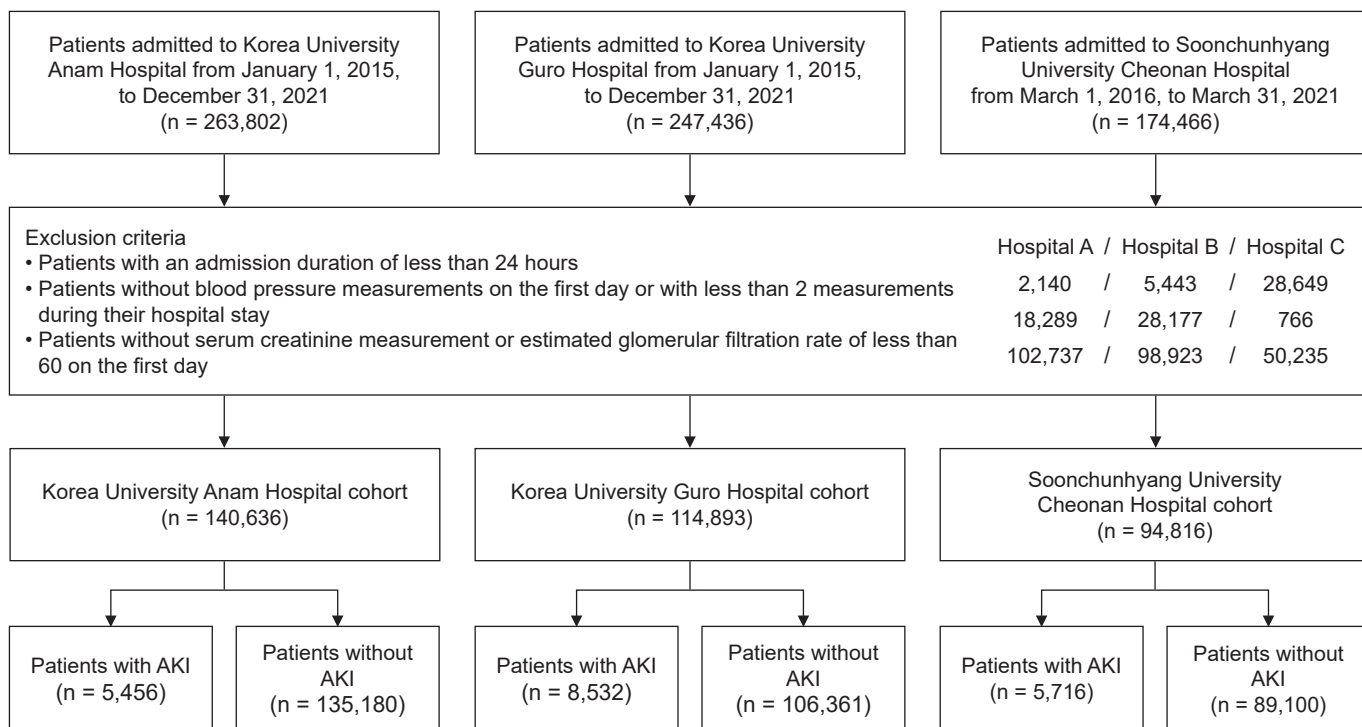
After applying these exclusion criteria, we established a research cohort consisting of 140,636 cases from Hospital A and 114,893 cases from Hospital B. Among these data, 5,456 and 8,532 cases of AKI cases during hospitalization were selected from Hospitals A and B, respectively. 94,816 cases of research cohort from hospital C were collected and 5,716 AKI cases during hospitalization were selected. The process of dataset construction is depicted in Fig. 1, and the distribution of AKI cases according to hospitalization period is presented in Supplementary Fig. 1 (available online). The data from Hospitals A and B were utilized for training and internal validation, and the data from Hospital C was

utilized for external validation.

For labeling purposes, patients who satisfied AKI recovery criteria were labeled with 1, others were labeled with 0. Patients were excluded from this study if, following the onset of AKI, Cr levels were not measured or, even if measured, recovery could not be definitively determined within 7 days. Examples regarding this process are shown in Supplementary Fig. 2 (available online).

### Statistical analysis

For continuous variables, the median and interquartile values were provided when a normal distribution could not be assumed; otherwise, the mean and standard deviation (SD) were presented. Categorical variables were represented by the number and percentage of patients. To assess the differences between recovery and non-recovery, t tests were performed for continuous variables showing a normal distribution, the Mann-Whitney U tests for non-normally distributed continuous variables, and the chi-square tests for categorical variables, all at a significance level of 0.05.



**Figure 1. The data composition process.**  
AKI, acute kidney injury.

## Data preprocessing

Data collected from the electronic health records included measurements, measurement times, and specific variables, resulting in numerous missing values. To address this issue, we adopted a method of summarizing the data at 24-hour intervals. This approach has been validated for its efficacy in prior studies, providing benefits in terms of simplicity of deployment and utilization of the developed model. By summarizing the data at 24-hour intervals, we maintained consistency and facilitated data analysis.

Variables with multiple measurements within 24 hours, were summarized as maximum, mean, minimum, and the number of measurements, for vital signs data. The laboratory test results were based on recent measurements. Additionally, variables such as the prescription of nephrotoxic drugs (e.g., nephrotoxic antibiotics, nonsteroidal anti-inflammatory drugs [NSAIDs], and cytotoxic chemotherapeutic agents), vascular imaging studies, surgery with general anesthesia, contrast-enhanced computed tomography (CT), and transfer to the ICU, were considered influential within 7 days of the corresponding measurement times. Next, eGFR was calculated using the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) 2021 Cr equation [18]. We also added a variable named “Increased amount of Cr” by subtracting the baseline Cr from Cr at the time of AKI occurrence. Finally, the “BUN/Cr ratio” represents the value obtained by dividing blood urea nitrogen (BUN) by Cr. Subsequently, robust scaling was applied to all the continuous variables. For categorical variables, one-hot encoding was applied.

### Feature selection

Approximately 120 features were extracted from the electronic health records, covering basic patient information, vital signs, laboratory test results, and other relevant factors. Following a comprehensive literature review and consultation with domain experts, a two-step feature selection process was undertaken. Initially, the LASSO (Least Absolute Shrinkage and Selection Operator) was utilized to examine the regression coefficients and SHapley Additive exPlanations (SHAP) values for all features. Simultaneously, the outcomes were assessed using a stepwise method with logistic regression as the criterion. Subsequently, based on correlation coefficients and missing value ratios, the fea-

ture set was refined to 43 features. Details of the selected 43 features are provided in [Supplementary Table 1](#) (available online).

### Outlier and missing value handling

To address outliers, the data distribution for each feature was thoroughly examined, and individual patient records were reviewed. Detection thresholds for certain numerical variables were established in collaboration with clinical experts, utilizing histograms and quantile-quantile plots.

To address missing values in the data, two methods were employed: imputation with preceding values, and the use of missing value indicators. For variables with available preceding values, the missing values were replaced with the preceding values to maintain data continuity. Additionally, for variables with low missing value rates (<20%), the multiple imputation by chained equations (MICE) method was used to handle missing values from the three hospitals. MICE is widely used to generate imputations that closely resemble true distributions when the rate of missing values is low [19]. The choice between the missing indicator method and MICE was determined based on the extent of missing values for each variable. The missing indicator method was used for variables with missing value rates exceeding 20%, where missing values were marked as ‘unknown.’ This approach allowed us to identify and distinguish the patterns of missing values within the variables. [Supplementary Table 2](#) (available online) presents the missing value rates before and after handling missing values with preceding values in each hospital. Additionally, [Supplementary Table 3](#) (available online) provides an explanation of the application of the missing indicator method, where missing values are marked as ‘unknown.’

### Machine learning models

Deep learning models generally exhibit superior performance compared to traditional machine learning models, but due to the amount of data required for training and high time and space complexities [20–22], traditional machine learning models were selected in this study. Various machine learning models were used in this study, including logistic regression (LR), random forest (RF), eXtreme Gradient Boosting (XGB), Light Gradient Boosting Model (LGBM), and Categorical Boosting (CAT). CAT was utilized

as an effective model for handling categorical variables, negating the need for separate one-hot encoding procedure for such variables [23].

For model training purposes, the dataset was randomly divided into 10 groups, ensuring an equal outcome ratio across them. The last group (group 10) was utilized for internal validation, while the remaining nine groups were used for cross-validation to conduct hyperparameter tuning and model selection. During the nine cycles of iteration, seven groups served for training purposes, one group was used to set early stopping criteria, and the remaining group was employed for performance evaluation. Grid search was applied for hyperparameter tuning, and the range of hyperparameters adjusted according to the model is detailed in [Supplementary Table 4](#) (available online). The processes of early stopping, hyperparameter tuning, and model selection were all based on the area under the receiver operating characteristic curve (AUROC). A graphical representation of the training process can be found in [Supplementary Fig. 3](#) (available online).

### Model evaluation and interpretation

For the evaluation metrics, accuracy, precision, recall, specificity, F1 score, AUROC, and area under the precision-recall curve (AUPRC) were utilized. Internal validation was conducted using group 10 from the internal cohort, which has not been employed in the training phase, while external validation utilized all data from the external cohort at Soonchunhyang University Cheonan Hospital. The evaluation of metrics such as accuracy, precision, recall, specificity, and F1 score, which vary according to the threshold, involved adjusting the threshold in increments of 0.01 from 0.01 to 0.99, and the results were presented in [Supplementary Fig. 4](#) (available online). The performance metrics of this manuscript are reported based on a threshold of 0.5. Evaluations were carried out on both internal and external validation datasets.

An in-depth sub-cohort analysis was conducted to gain a deeper understanding of early renal function recovery using the developed model. The criteria for the sub-cohort included patients with an eGFR less than 60 mL/min/1.73 m<sup>2</sup> and those with an eGFR of 60 mL/min/1.73 m<sup>2</sup> or above at the time of AKI onset, female and male patients, patients aged 65 years and over versus those under 65 years, and the

use of cytotoxic chemotherapeutic agents. To enhance the interpretability of the models, SHAP values were provided for each analysis.

## Results

### Baseline characteristics

The final analysis included 7,041 patients from the internal cohort and 2,929 patients from the external cohort. The labeling ratios for recovery in these hospitals were 46.9% and 51.8%, respectively. Among recovered patients, the mean time to recovery post-AKI was 5.05 days (SD, 2.26) in the internal cohort and 4.75 days (SD, 2.20) in the external cohort. The distribution of recovery in patients with AKI is illustrated in [Supplementary Fig. 5](#) (available online). The baseline characteristics of the recovery and non-recovery groups from each hospital are presented in [Supplementary Table 5](#) (available online). Although Cr was not utilized in the actual model training, it is included in the table. [Supplementary Table 6](#) (available online) shows the result after applying the missing indicator.

Numerous variables exhibited statistically significant differences between the renal function recovery group and the non-recovery group. Among continuous variables, vital signs such as blood pressure, heart rate, respiratory rate, along with white blood cell count, BUN, increased amount of Cr, glucose, blood sugar test, uric acid, phosphorus, chloride, and urine specific gravity (SG) were higher in the recovery group compared to the non-recovery group in both hospitals. Conversely, platelet, eGFR, alkaline phosphatase, pH, and total CO<sub>2</sub> levels were higher in the non-recovery group than in the recovery group across both hospitals. For categorical variables, the proportion of patients using nephrotoxic antibiotics in the non-recovery group was higher, while the incidence of contrast-enhanced CT and general anesthesia showed the opposite trend. Even when applying a missing indicator, variables that showed statistically significant differences among continuous variables largely maintained their disparities. Additionally, C-reactive protein and pro-brain natriuretic peptide, which did not exhibit statistically significant differences before applying the missing indicator, showed differences afterward. Some variables demonstrated statistical differences in only one of either the internal or external cohort. Total

cholesterol showed statistical differences in both cohorts, but with the non-recovery group showing higher levels in the internal cohort, whereas in the external cohort, the recovery group had higher levels. These differences between hospitals could potentially impact the model’s generalizability and suggest that analyzing a larger patient population is necessary to enhance generalization performance.

**Model performance**

Table 1 presents the performance outcomes of internal and external validation, evaluating five machine learning models: LR, RF, XGB, LGBM, and CAT. The CAT model demonstrated strong predictive capability with the highest AUROC of 0.7816. The RF model also performed well, with an AUROC of 0.7727, whereas the XGB model exhibited commendable performance with an AUROC of 0.7543. LR showed a moderate performance, with an AUROC of 0.7402. The LGBM displayed relatively lower predictive power, with an AUROC of 0.7254. The external validation results indicated a general decrease in performance across all models. The performance degradation in external validation, as measured by AUROC ranged from a minimum decrease of 0.0168 to a maximum decrease of 0.0456. The CAT model consistently outperformed the other models in terms of internal validation. The RF model demonstrated the highest performance in external validation. However, external validation data were not considered in the model selection process, and the CAT model was ultimately chosen. Supplementary Fig. 6 (available online) shows the

AUROC and AUPRC curves for internal and external validation.

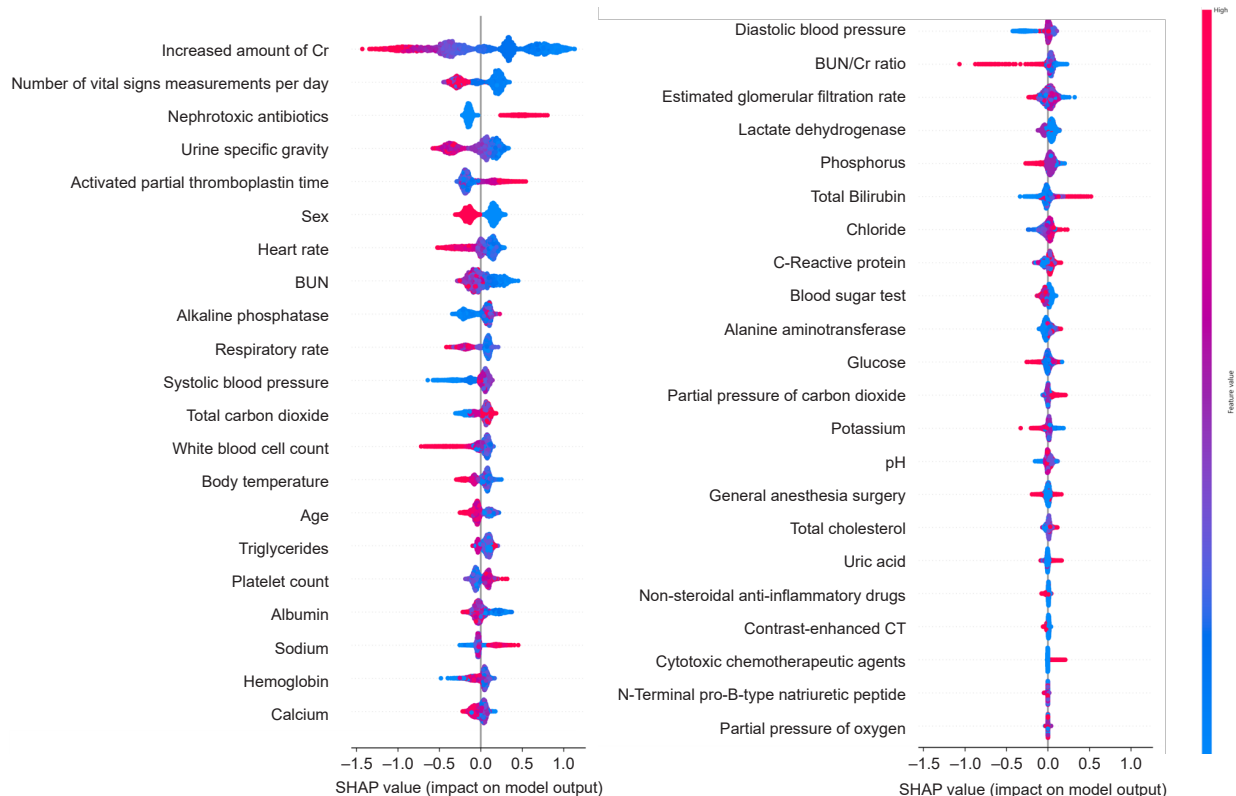
**Model evaluation and interpretation**

The SHAP values for the model are depicted in Fig. 2. In the figure, the red color indicates higher values of the respective feature, while the blue color signifies lower values. The position of the dots placed towards the right signifies a greater contribution by the feature value to the model’s prediction of non-recovery for the patient. For categorical variables, red dots represent a value of 1, and blue dots signify a value of 0. The variables that the model considered important include increased amount of Cr, the use of anti-SG, activated partial thromboplastin time, sex, heart rate, BUN, alkaline phosphatase, respiratory rate, systolic blood pressure, total carbon dioxide, white blood cell count, body temperature, age, triglycerides, platelet count, and albumin (Alb). When AKI occurs, all patients tend to have increased Cr. In cases where there was a greater change in Cr, it appeared to have an extreme impact, resulting in recovery. Patients with a high BUN/Cr ratio or high SG typically raise suspicions of dehydration. Patients who were dehydrated tended to have a more favorable prognosis for recovery. Variables such as high Alb, BUN, heart rate, and body temperature are positively associated with better recovery. The use of NSAIDs contributes to improved renal function recovery, whereas nephrotoxic antibiotics or cytotoxic chemotherapeutic agents hinder renal recovery.

**Table 1.** Comparative analysis of internal and external validation outcomes

Validation	Model	Accuracy	Precision	Recall	F1	AUROC	AUPRC
Internal	LR	0.6805	0.6881	0.7258	0.7065	0.7402	0.7633
	RF	0.7165	0.7354	0.7258	0.7306	0.7727	0.7932
	XGB	0.6473	0.6260	0.8303	0.7138	0.7543	0.7711
	LGBM	0.6321	0.6119	0.8355	0.7064	0.7254	0.7511
	CAT	0.7206	0.7453	0.7180	0.7314	0.7816	0.7962
External	LR	0.6268	0.5766	0.8499	0.6871	0.7234	0.7027
	RF	0.6726	0.6359	0.7507	0.6885	0.7394	0.7231
	XGB	0.6238	0.5725	0.8669	0.6896	0.7327	0.7164
	LGBM	0.6193	0.5747	0.8095	0.6722	0.7081	0.6683
	CAT	0.6685	0.6275	0.7684	0.6909	0.7360	0.7152

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; CAT, Categorical Boosting; LGBM, Light Gradient Boosting Model; LR, logistic regression; RF, random forest; XGB, eXtreme Gradient Boosting.



**Figure 2. SHapley Additive exPlanations (SHAP) value.**  
 BUN, blood urea nitrogen; Cr, creatinine; CT, computed tomography.

**Table 2. Subgroup analysis of internal and external validation outcomes**

Validation	Group	Accuracy	Precision	Recall	F1	AUROC	AUPRC
Internal	eGFR $\geq 60^a$	0.7298	0.7507	0.7833	0.7667	0.7846	0.8183
	eGFR $< 60^a$	0.6863	0.6875	0.3667	0.4783	0.7486	0.6284
	Male	0.7192	0.7538	0.6901	0.7206	0.7783	0.8003
	Female	0.7224	0.7356	0.7529	0.7442	0.7844	0.7966
	Age $\geq 65$ yr	0.7304	0.7366	0.7173	0.7268	0.7783	0.7698
	Age $< 65$ yr	0.7097	0.7541	0.7188	0.7360	0.7838	0.8247
	Yes <sup>b</sup>	0.7036	0.7279	0.7133	0.7205	0.7547	0.7786
	No <sup>b</sup>	0.7791	0.8133	0.7349	0.7722	0.8714	0.8824
External	eGFR $\geq 60^a$	0.6719	0.6410	0.8754	0.7401	0.7435	0.7503
	eGFR $< 60^a$	0.6597	0.5137	0.3357	0.4060	0.6323	0.4794
	Male	0.6647	0.6195	0.7484	0.6778	0.7265	0.7014
	Female	0.6735	0.6373	0.7933	0.7068	0.7467	0.7319
	Age $\geq 65$ yr	0.6567	0.6107	0.7647	0.6791	0.7231	0.6972
	Age $< 65$ yr	0.6809	0.6455	0.7722	0.7032	0.7496	0.7332
	Yes <sup>b</sup>	0.6700	0.6500	0.7696	0.7048	0.7285	0.7277
	No <sup>b</sup>	0.6629	0.5307	0.7621	0.6257	0.7607	0.6664

AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; eGFR, estimated glomerular filtration rate. <sup>a</sup>The unit is mL/min/1.73 m<sup>2</sup>. <sup>b</sup>It means general anesthesia surgery.

## Subgroup analysis

**Table 2** presents the results of subgroup analyses based on four criteria at the onset of AKI: eGFR levels, sex, age, and whether the patient underwent general anesthesia. Only internal and external validations were conducted for the model previously trained. While the models' performance did not significantly differ by sex and age, it exhibited higher efficiency in groups with an eGFR of 60 mL/min/1.73 m<sup>2</sup> or above at the onset of AKI and in those who did not undergo surgery under general anesthesia compared to their counterparts. These findings suggest that the model developed in this study may perform better in patients with relatively good renal function and those who have not received general anesthesia, indicating its potential superiority in milder cases.

## Discussion

In this study, we introduced a method to predict recovery after AKI. The model achieved a moderate performance level with an AUROC of 0.7816 (internal) and 0.7360 (external). Despite its significant relevance to patient prognosis, AKI recovery remains unclear, making predictions challenging for clinicians, the model that demonstrated the highest performance is CAT. CAT is known as a tree-based ensemble model specialized in handling categorical variables through mechanisms such as categorical feature combination [23]. It is speculated that the increase in the number of categorical variables, due to the use of the missing indicator, contributed to CAT's superior performance. We utilized SHAP values to identify the factors contributing to AKI recovery and quantify their significance. As expected, patients with low Alb or high blood pressure, showed a tendency toward poorer renal function recovery. Conversely, patients with elevated white blood cell counts and high body temperatures tended to have better renal function recovery. For patients where AKI might be associated with infection, resolution of the causative infection also led to AKI recovery. Although the use of NSAIDs is related to better renal recovery, the use of chemotherapeutic agents or nephrotoxic antibiotics tends to hinder recovery. The use of NSAIDs can be easily discontinued in AKI patients; however, agents such as chemotherapy may be challenging to cease depending on the patient's condition and are also

presumed to act as negative factors, potentially causing more severe damage. Furthermore, in dehydrated patients with a high BUN/Cr ratio or urine SG, a more favorable recovery trend was observed as the state of dehydration improved, high BUN had a positive impact on recovery. This suggests that AKI caused by prerenal factors tends to lead to a relatively good recovery.

It was observed that patients who underwent surgery under general anesthesia exhibited more substantial recovery in statistical analysis. However, this trend was not evident in the SHAP analysis. While the tendency to predict recovery or non-recovery was not distinct for patients who had undergone general anesthesia, the absolute value of SHAP values was larger compared to those who had not received general anesthesia. The outcomes following surgery under general anesthesia could vary greatly depending on the cause and type of surgery, and this complexity seems to be reflected in the model. For patients who did not undergo surgery, the model demonstrated a very high performance with a score of 0.8714. Future research may benefit from distinguishing characteristics based on the purpose and type of surgery rather than categorizing all patients uniformly based on the administration of general anesthesia.

Interpreting the model through SHAP values offers various advantages over relying solely on statistics. By quantifying the factors contributing to AKI recovery in each patient, this approach provides insights into the crucial elements influencing AKI recovery. Predicting whether and when patients with AKI recover in real-life situations is challenging. The model proposed in this study could serve as an effective tool to assist clinicians in situations where clinical judgment alone may be insufficient.

Another objective of this study was to evaluate the clinical applicability of the machine learning model. To achieve this, we conducted various studies on renal function recovery using data from three hospitals. We observed a trend of decreased external validation performance compared to internal validation, which was attributed to differences in patient populations, disease severity, and hospitalization patterns among institutions. Additionally, the inability to precisely match the features used in the model and potential overfitting of the training data may have contributed to this trend [24]. Therefore, limited data and patient numbers may interfere with the model's ability to learn generalized patterns.

It is important to note that almost all medical artificial in-



telligence studies have been retrospective. Therefore, several biases (e.g., biases stemming from data loss, label definitions based on operational definitions, and selection of Cr baseline or reference values) occurred during patient selection and cohort formation owing to the exclusion of a substantial number of patients. These factors are critical and contribute to the uncertainty in the general performance of the developed model. Additionally, the process of organizing data on a daily basis was intended for the convenience of model development and application but might have introduced biases into the model. This underscores the need for model calibration when applied to clinical settings and the importance of including diverse patient populations with sufficient sample sizes in multi-institutional studies [25].

This study is subject to certain limitations. First, the lack of consensus on AKI recovery criteria necessitated reliance on existing research findings to establish a definition. Moreover, recovery status was evaluated using the initial Cr level at the time of AKI onset as a fixed reference point, an approach that may not fully account for the clinical context of peak Cr levels when evaluating Cr stages. Further, limitations in data utilization precluded the incorporation of additional clinical evidence related to AKI beyond Cr levels [16]. As a result, this criterion was omitted from the AKI definition, potentially leading to unidentified AKI cases and introducing bias. Secondly, the absence of data on dialysis or kidney transplantation was noted, and this was addressed by excluding patients with an eGFR of less than 60 mL/min/1.73 m<sup>2</sup>, an issue that warrants attention in future research. Thirdly, the study did not account for the differences in interventions post-AKI occurrence. The nature of interventions following AKI can significantly influence patient prognosis, and considering this through additional data collection could be highly significant [26]. Lastly, while the model's performance was commendable, its adequacy for seamless real-world application is acknowledged to be limited. Future efforts should aim at enhancing performance by securing more diverse and extensive datasets and refining the machine learning methodology for predicting renal function recovery in patients with AKI.

In conclusion, our study introduced a machine learning-based approach for predicting recovery after AKI, revealing key factors influencing renal function recovery. This approach will be helpful in aiding clinical decision-making and further future research.

## Conflicts of interest

All authors have no conflicts of interest to declare.

## Funding

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Basic Science Research Program of the National Research Foundation (NRF-2021R1A2C1009290), the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2023-RS-2022-00156439) supervised by the IITP (Institute of Information and Communications Technology Planning and Evaluation), a Korea University Grant (K2210721) and Soonchunhyang University Research Fund.

## Data sharing statement

The data set used in this study is not publicly available. However, the data of this study can be obtained upon reasonable request from the corresponding author. The code to generate the result of this study can be accessed at [https://github.com/5454dls/AKI\\_recovery](https://github.com/5454dls/AKI_recovery), upon reasonable request.

## Authors' contributions

Conceptualization: NJC, IJ  
 Data curation: NJC, SHK, HWG  
 Formal analysis: NJC, IJ, YK, HWG  
 Funding acquisition, Project administration, Resources: HL  
 Investigation: IJ, YK, DOK  
 Methodology: IJ  
 Supervision: HWG, HL  
 Validation: SJA, SHK  
 Writing–original draft: All authors  
 Writing–review & editing: All authors  
 All authors read and approved the final manuscript.

## ORCID

Nam-Jun Cho, <https://orcid.org/0000-0002-9053-0499>  
 Inyong Jeong, <https://orcid.org/0009-0009-1504-0463>  
 Yeongmin Kim, <https://orcid.org/0009-0001-3234-0101>  
 Dong Ok Kim, <https://orcid.org/0000-0001-7510-8380>

Se-Jin Ahn, <https://orcid.org/0009-0003-5379-3673>  
 Sang-Hee Kang, <https://orcid.org/0000-0002-6097-8831>  
 Hyo-Wook Gil, <https://orcid.org/0000-0003-2550-2739>  
 Hwamin Lee, <https://orcid.org/0000-0002-6482-3511>

## References

1. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–119.
2. Rewa O, Bagshaw SM. Acute kidney injury-epidemiology, outcomes and economics. *Nat Rev Nephrol* 2014;10:193–207.
3. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical decision support for in-hospital AKI. *J Am Soc Nephrol* 2018;29:654–660.
4. Hoste EA, Bagshaw SM, Bellomo R, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med* 2015;41:1411–1423.
5. Kellum JA, Sileanu FE, Bihorac A, Hoste EA, Chawla LS. Recovery after acute kidney injury. *Am J Respir Crit Care Med* 2017;195:784–791.
6. Haredasht FN, Vanhoutte L, Vens C, Pottel H, Viaene L, De Corte W. Validated risk prediction models for outcomes of acute kidney injury: a systematic review. *BMC Nephrol* 2023;24:133.
7. Forni LG, Darmon M, Ostermann M, et al. Renal recovery after acute kidney injury. *Intensive Care Med* 2017;43:855–866.
8. Hickson LJ, Chaudhary S, Williams AW, et al. Predictors of out-patient kidney function recovery among patients who initiate hemodialysis in the hospital. *Am J Kidney Dis* 2015;65:592–602.
9. Craven AM, Hawley CM, McDonald SP, Rosman JB, Brown FG, Johnson DW. Predictors of renal recovery in Australian and New Zealand end-stage renal failure patients treated with peritoneal dialysis. *Perit Dial Int* 2007;27:184–191.
10. Srisawat N, Wen X, Lee M, et al. Urinary biomarkers and renal recovery in critically ill patients with renal support. *Clin J Am Soc Nephrol* 2011;6:1815–1823.
11. Zhao X, Lu Y, Li S, et al. Predicting renal function recovery and short-term reversibility among acute kidney injury patients in the ICU: comparison of machine learning methods and conventional regression. *Ren Fail* 2022;44:1326–1337.
12. Liu CL, Tain YL, Lin YC, Hsu CN. Prediction and clinically important factors of acute kidney injury non-recovery. *Front Med (Lausanne)* 2021;8:789874.
13. He J, Lin J, Duan M. Application of machine learning to predict acute kidney disease in patients with sepsis associated acute kidney injury. *Front Med (Lausanne)* 2021;8:792974.
14. Luo XQ, Yan P, Zhang NY, et al. Machine learning for early discrimination between transient and persistent acute kidney injury in critically ill patients with sepsis. *Sci Rep* 2021;11:20269.
15. Huang CY, Güüza F, De Vlieger G, et al. Development and validation of clinical prediction models for acute kidney injury recovery at hospital discharge in critically ill adults. *J Clin Monit Comput* 2023;37:113–125.
16. Chawla LS, Bellomo R, Bihorac A, et al. Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup. *Nat Rev Nephrol* 2017;13:241–257.
17. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. *Lancet* 2012;380:756–766.
18. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med* 2021;385:1737–1749.
19. Mera-Gaona M, Neumann U, Vargas-Canas R, López DM. Evaluating the impact of multivariate imputation by MICE in feature selection. *PLoS One* 2021;16:e0254720.
20. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;1:206–215.
21. Qi GJ, Luo J. Small data challenges in big data era: a survey of recent progress on unsupervised and semi-supervised methods. *IEEE Trans Pattern Anal Mach Intell* 2022;44:2168–2187.
22. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8:53.
23. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features [Preprint]. arXiv 2017 [updated 2019 Jan 20; cited 2023 Dec 3]. Available from: <https://doi.org/10.48550/arXiv.1706.09516>
24. Sagawa S, Koh PW, Hashimoto TB, Liang P. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization [Preprint]. arXiv 2019 [updated 2020 Apr 2; cited 2023 Dec 3]. Available from: <https://doi.org/10.48550/arXiv.1911.08731>
25. Lee SW, Lee HC, Suh J, et al. Multi-center validation of machine learning model for preoperative prediction of postoperative mortality. *NPJ Digit Med* 2022;5:91.
26. Balasubramanian G, Al-Aly Z, Moiz A, et al. Early nephrologist involvement in hospital-acquired acute kidney injury: a pilot study. *Am J Kidney Dis* 2011;57:228–234.