Big Data and Information Analytics

https://www.aimspress.com/journal/BDIA

BDIA, 9: 302–327.
DOI:10.3934/bdia.2025014
Received: 27 October 2025
Revised: 18 November 2025
Accepted: 20 November 2025
Published: 03 December 2025

Research article

Time-series analysis of acute kidney injury based on machine learning

Hui Chang[1], **Qiannan Hou**[2] **and Yueli Chen**[2,*]

[1] Department of Critical Care Medicine, Luoyang Orthopedic-Traumatological Hospital of Henan Province (Henan Provincial Orthopedic Hospital), Zhengzhou 450016, China

[2] School of Methematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

* **Correspondence:** Email: chenyueli@zzu.edu.cn.

Abstract: Acute kidney injury (AKI) is a common and critical condition in intensive care unit (ICU) settings, characterized by rapid onset, high rates of underdiagnosis, and poor clinical outcomes. While existing early prediction models predominantly focus on AKI stage 1, the acute dialysis quality initiative (ADQI) emphasizes that early identification of advanced stages (2 and 3) is of greater clinical significance due to their increased severity and substantially higher mortality rates. In this study, we develop and validate a dynamic early prediction model for AKI stages 2 and 3 using multivariate time-series data extracted from the MIMIC-III database. To enhance the model's robustness, we perform comprehensive data preprocessing, including resampling to address temporal irregularities and multiple imputation to handle missing values. The sequence lengths are standardized through truncation, and the baseline feature differences between patient groups are analyzed. First, we evaluate several machine learning models, including logistic regression, for direct prediction within 96 hours after admission to the ICU, with CatBoost demonstrating the highest performance. Subsequently, we propose a four-layer bidirectional long short-term memory (LSTM) network for dynamic risk assessment, which leverages temporal dependencies in clinical data and significantly outperforms static prediction approaches. Furthermore, we incorporate an interpretability analysis to identify key predictive variables, thus supporting clinical decision-making and allowing timely interventions for high-risk patients. Our results demonstrate the potential of deep learning models to improve the early detection of severe AKI, with implications for optimizing ICU management strategies.

Keywords: acute kidney injury; machine learning; direct prediction; dynamic prediction; interpretability analysis

1. Introduction

Acute kidney injury (AKI) is a complex clinical syndrome precipitated by diverse etiologies, including nephrotoxic drug exposure, direct renal ischemia, systemic infection, and perioperative complications [1–3]. As a critical component of multiorgan dysfunction, AKI not only impairs renal function but also destabilizes the respiratory, cardiovascular, hematologic, and nervous systems through systemic inflammatory cascades. In current practice, an AKI diagnosis predominantly relies on laboratory markers—most notably serum creatinine (Scr)—whose dynamic changes define the kidney disease: improving global outcomes (KDIGO) staging criteria [4]. However, Scr is a delayed biomarker, typically rising 48–72 hours after the onset of renal injury [5], and its utility is further limited by suboptimal monitoring frequency. For example, in China, fewer than 25.3% of hospitalized patients receive more than two Scr measurements during admission [6], thus contributing to a high rate of missed or delayed diagnoses. Given the narrow therapeutic window for intervention, such delays are associated with significantly increased in-hospital mortality [5], thus underscoring the urgent need for proactive, predictive approaches to AKI detection.

The KDIGO 2012 criteria remain the global standard for AKI diagnoses and staging, thereby integrating changes in Scr and urine output to classify severity into stages 1, 2, and 3 [4]. While this framework enables risk stratification, both biomarkers have critical limitations. As noted, Scr lacks sensitivity in early injury. Urine output, though more responsive, is often inconsistently recorded in electronic health records due to challenges in continuous, accurate measurements [7]. This diagnostic latency undermines timely clinical action and stands in stark contrast to the International Society of Nephrology's "0 by 25" initiative, which aims to eliminate preventable AKI-related deaths by 2025 [4]. With AKI incidence rising and therapeutic options remaining largely supportive, early prediction has emerged as a pivotal strategy to improve outcomes.

Importantly, the prognostic implications of AKI are not uniform across stages. A growing body of evidence demonstrates that mortality, organ failure burden, and healthcare utilization exponentially escalate with advancing AKI severity [8–13]. This has prompted the acute dialysis quality initiative (ADQI) to advocate for a paradigm shift—from detecting AKI to predicting and managing moderate-to-severe disease (KDIGO stages 2–3), which carries substantially higher morbidity and mortality [7, 14]. Despite this, clinical awareness and staging precision remain suboptimal, with many cases misclassified or underappreciated as stage 1. This gap is mirrored in the artificial intelligence (AI) literature: while numerous machine learning models have been developed for general AKI risk prediction, few specifically focus on the early identification of stages 2–3, and even fewer support dynamic, real-time risk assessments.

The temporal nature of AKI—characterized by variable onset and evolving physiology—demands models capable of capturing complex longitudinal patterns in high-dimensional clinical data. Conventional statistical and machine learning methods often struggle with such tasks due to limitations in modeling long-term dependencies, handling irregular sampling, and mitigating gradient instability in sequential learning [15]. The long short-term memory (LSTM) networks, a class of recurrent neural networks (RNNs), offer a powerful alternative. Equipped with gating mechanisms, LSTMs can selectively retain or discard information over time, thus enabling robust modeling of both short-term fluctuations and long-term trends in physiological trajectories.

In this study, we leverage the MIMIC-III critical care database to develop and validate a deep

learning–based early prediction model for AKI stages 2–3 using a four-layer bidirectional LSTM architecture. Our approach emphasizes dynamic risk stratification from the earliest hours of intensive care unit (ICU) admission, thus enabling continuous monitoring and timely interventions. Beyond the predictive performance, we incorporate interpretability techniques to identify salient clinical features, thus enhancing the model transparency and clinical utility. By focusing on advanced AKI stages and integrating temporal dynamics, this work aims to bridge a critical gap between AI innovation and actionable clinical decision support in critical care nephrology.

2. Materials and method

2.1. Data extraction

2.1.1. Variable selection

Based on the KDIGO criteria and the 2016 ADQI consensus conference [7, 16, 17], clinically common examination indicators and patient demographic information were selected as predictors for AKI stages 2 and 3. Compared to using a single serum creatinine level for AKI staging, the method employed in this study better aligns with the ADQI consensus requirements for sensitivity and specificity. The specific classifications of variables are as follows:

1) Demographic variables (3): age, race, gender;
2) Treatment-related variables (3): mechanical ventilation, vasopressor use, and sedative use [10, 18, 19];
3) Laboratory variables (22): anion gap, albumin, band neutrophils, bicarbonate, total bilirubin, serum creatinine, chloride, blood glucose, hematocrit, hemoglobin, lactate, platelet count, serum potassium, activated partial thromboplastin time (aPTT), international normalized ratio (INR), prothrombin time (PT), serum sodium, blood urea nitrogen (BUN), white blood cell count (WBC), 24-hour urine output, 12-hour urine output, and 6-hour urine output;
4) Vital signs variables (6): mean arterial pressure, heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate, and temperature.

2.2. Variable extraction

This study focuses on time-series data. However, the selected demographic characteristics are static variables whose values remain constant over time. In contrast, dynamic variables are collected at varying sampling frequencies in clinical settings. Therefore, the study variables are categorized and stored in separate tables: the demographic variable table, laboratory variable table, vital sign variable table, intervention measure variable table, and AKI staging label table. The following section describes the process of reconstructing these tables from the MIMIC-III database. Figure 1 illustrates the data extraction flowchart.

1) The demographic variable table was created to store core patient characteristics, including age, gender, and ethnicity, sourced from the MIMIC-III database. Data were extracted using PostgreSQL from four relevant tables—`noteevents`, `icustays`, `admissions`, and `patients`—within the 26-table MIMIC-III schema. The joined query resulted in a total of 56,443 ICU stay identifiers (`icustay_id`), 52,835 hospital admission identifiers (`hadm_id`),

and 42,022 distinct patient identifiers (subject_id). The observed discrepancy in identifier counts stems from the structure of clinical care: each hospital admission is assigned a unique hadm_id, and each ICU stay within an admission receives a separate icustay_id. In contrast, the subject_id serves as a persistent patient-level identifier, thus remaining constant across all hospital and ICU encounters.

2) Generation of the laboratory variable table. Time-series data for 19 laboratory variables were extracted from the icustays and chartevents tables in the MIMIC-III database using their corresponding item identifiers (ITEMID). Subsequently we calculated the missing value percentage for each variable and excluded those with more than 80% missing data, specifically albumin, bilirubin, banded neutrophils, and lactate. This curation process yielded a final dataset of 1,522,865 laboratory records.
3) Generation of the vital signs variable table. Time-series data for seven vital signs variables were extracted from the icustays and chartevents tables in the MIMIC-III database using their corresponding item identifiers (ITEMID). We assessed data completeness by calculating the missing value percentage for each variable. Variables exceeding the 80% missingness threshold, which included temperature, were excluded from the final dataset. This curation process resulted in a total of 9,079,442 vital sign records available for analysis.
4) Generation of the AKI staging label table. The AKI staging label table was constructed to store key clinical parameters for patients in the MIMIC-III database, including ICU stay identifiers (icustay_id), timestamps (charttime), serum creatinine levels, urine output measurements over 6-hour (uo_rt_6hr), 12-hour (uo_rt_12hr), and 24-hour (uo_rt_24hr) intervals, and the corresponding AKI stage labels. The staging system follows the KDIGO criteria, where stage 0 indicates no AKI, and stages 1–3 represent the increasing severity of acute kidney injury. Serum creatinine (ITEMID: 50912) and urine output (corresponding to 26 distinct ITEMIDs) data were extracted from the icustays, labevents, outputevents, and chartevents tables using PostgreSQL. AKI staging was programmatically determined according to the KDIGO guidelines, with baseline serum creatinine defined as the lowest value recorded within the preceding seven-day period. In cases where the staging based on serum creatinine and urine output criteria conflicted, the more severe stage was retained. The final curated dataset comprised 3,737,147 AKI staging records suitable for the subsequent analysis.
5) Generation of the intervention variable table. Data for three key therapeutic interventions—mechanical ventilation, vasopressor administration, and sedative use—were extracted from five source tables in the MIMIC-III database: icustays, labevents, outputevents, chartevents, and inputevents. The extraction process yielded a total of 10,483,277 intervention records, thus forming a comprehensive dataset for the subsequent analytical modeling.

Subsequently, the five aforementioned tables were merged to create a unified dataset. The demographic, laboratory, vital signs, and intervention variable tables were joined with the AKI staging label table using patient identifiers (subject_id, hadm_id, icustay_id) and temporal markers (charttime) to form the final composite dataset. Then, the merged dataset was refined according to the following exclusion criteria:

1) Patients under 18 years of age;
2) Hospital length of stay less than 24 hours;

3) Fewer than two recorded serum creatinine measurements during the ICU stay;
4) Patients with maximum AKI stage below 2;
5) Pre-existing chronic kidney disease;
6) Initiation of renal replacement therapy within 24 hours of ICU admission; and
7) History of kidney transplantation.

After applying these exclusion criteria, a total of 29,628 patients with an AKI stage of 2 or 3 who met the eligibility criteria were included.

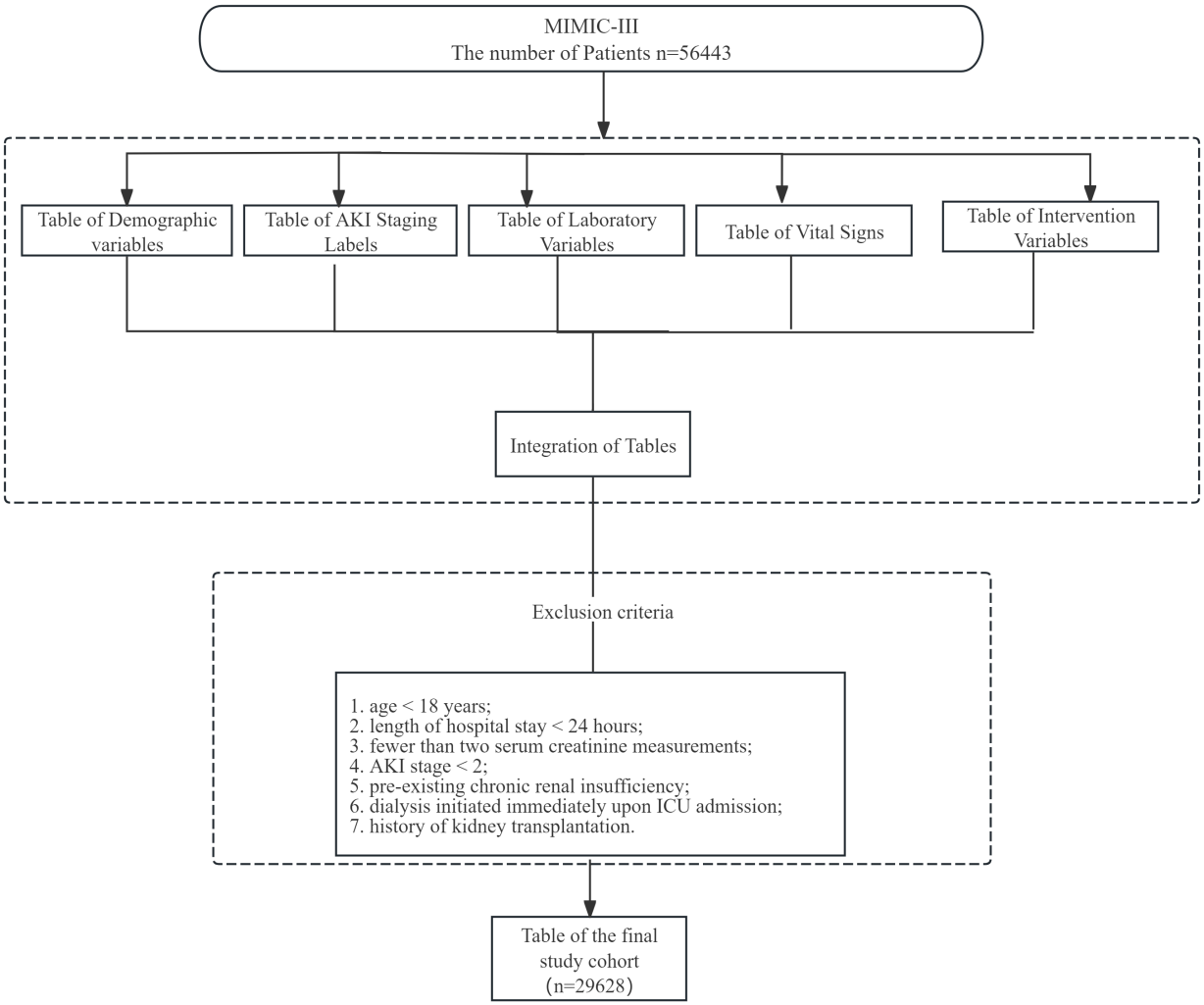


Figure 1. Flowchart of data extraction.

2.3. Data preprocessing

Medical time series data are frequently characterized by irregular sampling frequencies, extensive missingness, and considerable heterogeneity in measurement frequency and value ranges across patients. Furthermore, the AKI stage for a given patient may exhibit dynamic progression over time.

Given the critical impact of data quality on the model's performance, the extracted time-series data for AKI stage 2 and stage 3 will undergo comprehensive preprocessing prior to modeling.

2.3.1. Resampling

This study analyzed time-series data from 29,628 patients diagnosed with stage 2 or 3 AKI. The raw dataset exhibited inherent irregular sampling, with substantially varying measurement frequencies across laboratory values, vital signs, therapeutic interventions, and AKI staging labels. For instance, while vital signs in the ICU are typically recorded every 15 minutes, AKI staging according to KDIGO criteria relies on 6-hour urine output assessments. Since such irregular temporal patterns are incompatible with standard machine learning algorithms that require uniformly spaced observations, we converted all data to a regular time series using a 6-hour resampling window.

To address inter-patient heterogeneity, resampling was performed at the individual level. Through a groupby operation based on ICU stay identifiers (`icustay_id`), each patient's records were independently processed. Within each 6-hour window, the dynamic variables were aggregated by computing their mean values, while the static variables were represented by their maximum values. Missing measurements within a given interval were encoded as NaN (not a number), which is consistent with the approach described in [18]. The specific resampling procedure comprised the following steps:

1) Raw data extraction: for each patient, extract all recorded clinical variables following the ICU admission, organized by `icustay_id`.
2) Temporal resampling: apply a 6-hour fixed window to align measurements. Compute mean values for dynamic variables and maximum values for the static variables within each interval.
3) Missing value handling: assign NaN to variables with no available measurements within a resampling window.
4) Iterative processing: repeat steps 1–3 for all 29,628 patients to generate a unified, regularized time-series dataset.

2.3.2. Missing data imputation

Medical time-series data often exhibit a high prevalence of missing values due to the incorporation of the time dimension. Commonly used methods for handling missing data include the following: directly using variables with missing values; deleting variables containing missing values; imputing missing values; and employing machine learning algorithms to predict missing values [20]. According to the characteristics of the label variable and dynamic variables in this study, the following two imputation methods were adopted:

1) For the missing values in the AKI staging label, forward filling was applied. AKI staging is a categorical label value, which requires careful handling during imputation. Since a patient's AKI status in the ICU dynamically changes and the current AKI stage is often correlated with previous states, forward filling was used to preserve data integrity by leveraging historical information.
2) For missing values in the dynamic variables, Multiple Imputation was employed. This approach accounts for uncertainty in the missing data by simulating their random distribution and drawing imputed values from the generated distribution. Specifically, four multiply imputed datasets were

created with a random seed set to 1991. Using the multiple imputation by chained equations (MICE) algorithm, three iterations were performed to obtain the final imputed datasets.

As shown in Table 1, the percentage of missing values in the dynamic variables increased after table merging prior to resampling. However, after 6-hour resampling, the missing percentage decreased for all 18 laboratory variables and AKI staging, which indicates that resampling the dynamic variables with different measurement frequencies can reduce missingness in medical time-series data. Ultimately, a complete medical time-series dataset was obtained through systematic imputation of both the AKI staging labels and the dynamic variables.

Table 1. Missing rate of dynamic variables.

Variable	Single table	Pre-resampling	Post-resampling
anion gap	52.2%	93.9%	69.99%
bicarbonate	51.44%	93.80%	69.52%
bun	51.44%	93.69%	68.17%
chloride	46.98%	93.23%	68.17%
creatinine	51.28%	66.05%	47.40%
glucose	38.47%	91.99%	65.90%
hematocrit	39.49%	92.21%	65.73%
hemoglobin	48.61%	93.37%	70.38%
platelets	53.03%	93.95%	70.78%
potassium ion concentration	33.52%	91.48%	63.16%
sodium ion concentration	44.50%	92.91%	67.19%
oxygen saturation	33.05%	48.32%	62.07%
6-hour urine output	12.58%	69.10%	59.79%
24-hour urine output	12.58%	69.10%	59.79%
12-hour urine output	12.58%	69.10%	59.79%
white blood cell count	54.84%	94.18%	71.57%
international normalized ratio	69.76%	96.07%	81.19%
prothrombin time	69.77%	96.07%	81.19%
activated partial thromboplastin time	67.84%	95.82%	80.05%
mean arterial pressure	36.21%	50.82%	61.86%
diastolic blood pressure	36.39%	50.96%	62.00%
systolic blood pressure	36.37%	50.95%	61.99%
heart rate	13.44%	47.05%	61.68%
respiratory rate	30.93%	46.74%	61.68%
aki staging	00.00%	69.10%	29.26%

2.3.3. Data squence length truncation

Research by Brown [21] on patient similarity indicated that patients with similar clinical characteristics often share comparable treatment pathways and prognosis. Therefore, the information from similar patients can provide important references for the outcome prediction and disease risk

assessment. Additionally, the consistency in ICU length of stay reflects, to some extent, the similarity among patients. The strategic truncation of sequence lengths can facilitate the identification of patient cohorts with comparable clinical profiles. For instance, patients with short ICU stays typically significantly differ from those with long stays in terms of AKI severity, etiology complexity, and complication diversity. Consequently, patients with similar ICU stay durations exhibit inherent clinical homogeneity.

$$\text{Sequence Length} = \text{ICU Length of Stay (days)} \times 4 \tag{2.1}$$

This study begins with a comprehensive analysis of sequence lengths in the AKI time-series data. As defined in Eq (2.1), each unit of sequence length corresponds to a 6-hour interval of ICU stay. As shown in Table 2, among the 29,628 patients with AKI stages 2 and 3, the distribution of sequence lengths without truncation exhibits substantial variability: 25% of patients have a sequence length of 3, 50% have a length of 6, and 75% reach a length of 13. The mean sequence length is 10.75 (standard deviation: 15.23), thus reflecting a wide heterogeneity in ICU duration across the individuals.

Two critical challenges arise from the use of untruncated sequences:

1) Presence of outlying clinical trajectories: approximately 25% of patients exhibit prolonged ICU stays, which corresponds to extended sequence lengths. As illustrated in Figure 2, an inflection point emerges at a sequence length of approximately 20 (i.e., 120 hours or 5 days), beyond which the number of patients plateaus. This suggests that only a small fraction of patients remain in the ICU for more than five consecutive days. Given that the sequence length serves as a proxy for clinical course complexity, these long-stay patients often present with more severe comorbidities, complex treatment regimens, and intensified care requirements—features that significantly diverge from the typical AKI stage 2 and 3 populations. Consequently, they represent statistical outliers whose inclusion may bias the model's training and impair generalizability.

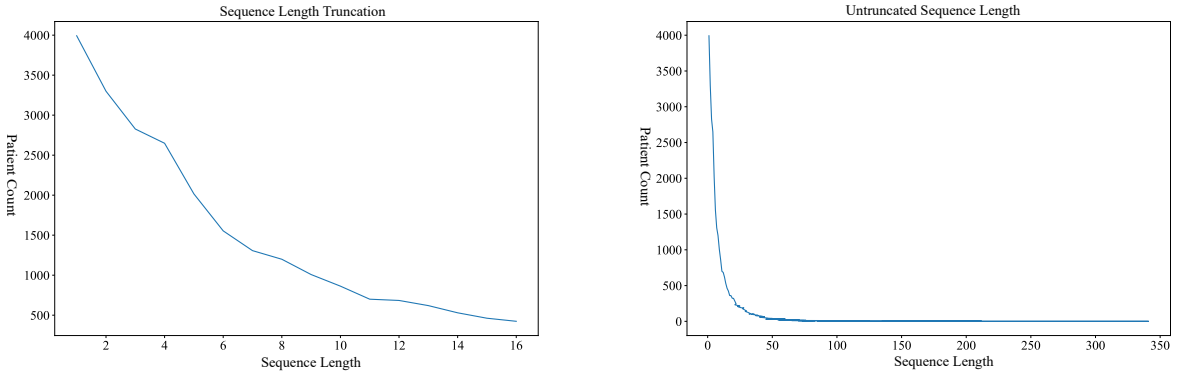
2) Input dimensionality and model performance constraints: our modeling framework requires a patient-specific time-series analysis, yet the inherent variability in ICU length of stay leads to non-uniform sequence lengths. However, deep learning models, however, necessitate fixed-length input sequences. Setting the truncation threshold excessively high would require extensive zero-padding for the majority of shorter sequences, thus introducing spurious noise and potentially degrading the model's performance due to dilution of informative temporal patterns.

To balance the clinical interpretability, data utility, and computational efficiency, we determined an optimal truncation length based on both empirical observation and clinical relevance. The inflection point observed at a sequence length of 20 in Figure 2 indicates diminishing returns in information content beyond this point. Furthermore, given that the third quartile of the raw distribution is 13 and that clinical assessments are typically framed in daily intervals, we set the truncation length to 16—corresponding to 96 hours (4 days)—to preserve the majority of clinically meaningful trajectories while minimizing outlier influence.

After truncating all sequences to a maximum length of 16, the standard deviation of sequence lengths markedly decreased from 15.23 to 5.41, thus indicating a substantially more homogeneous and stable distribution. In the truncated dataset, the first quartile falls at a length of 2 (25% of patients have $\leq$ 12 hours of recorded data), and the third quartile at 8 (75% of patients have $\leq$ 48 hours), which further supports the concentration of clinically representative episodes within the first few days of AKI onset.

Table 2. Sequence length comparison.

–	Patients count	Mean	Standard deviation	Min	25%	50%	75%	Max
Original length	29628	10.75	15.23	1	3	6	13	341
Truncated length	24134	5.41	4.06	1	2	4	8	16



(a) **(b)**

Figure 2. (a) Distribution of patient counts across truncated sequence lengths; (b) Distribution of patient counts across original (untruncated) sequence lengths.

From Figure 2, it can be observed that when the AKI data has not undergone truncation, a notable inflection point occurs around the sequence length of 20, which indicates that the data after this inflection point contains relatively limited effective information. Further data analysis reveals that the upper quartile of the untruncated sequence lengths is 13. Considering that patient research in the medical field is typically conducted on a daily basis, after comprehensive deliberation, the truncated sequence length is ultimately determined to be 16, which corresponds to an actual time span of 4 days. After truncating the data of 29,628 patients in AKI stages 2 and 3 to a sequence length of 16, the data analysis shows that the standard deviation of the truncated data is 5.41, which represents a significant reduction compared to the pre-truncation standard deviation of 10.75. Since the standard deviation is a crucial metric to measure data dispersion, with a smaller standard deviation implying a more concentrated data distribution, it can be inferred that truncating the sequence to a length of 16 effectively improves the stationarity of the medical time-series data distribution. Moreover, the lower quartile of the truncated data is 2, meaning that when the untruncated sequence lengths are arranged in ascending order, 25% of the patients have sequence lengths less than 2. In contrast, the upper quartile is 8, which indicates that 75% of the patients have sequence lengths less than 8.

2.3.4. Normalization

Normalization is a technique that maps data with large disparities in scale to the interval [0, 1] through mathematical transformation. This process mitigates the negative impact of significant differences in the magnitudes of input and output data on the prediction accuracy of a network. Common normalization methods include min-max scaling, neural network-based normalization, and

L2 normalization. This study employed the min-max scaling method, which is implemented using formula (2.2), where max and min represent the maximum and minimum values of each feature, respectively.

$$x_{normalization} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2.2}$$

Figure 3 illustrates the data preprocessing pipeline. First, the time-series data for AKI stages 2 and 3 were resampled at 6-hour intervals to convert irregular time series into regular ones. Subsequently, discrete variables in the dataset were subjected to one-hot encoding. Specifically, gender was encoded into two categories (male, female), and race was encoded into seven categories (White, Black, Hispanic, Asian, Native American, Unknown, and Other). Following this, the dynamic variables for AKI stages 2 and 3 underwent missing value imputation, sequence truncation, and normalization to obtain comparable patient data. Furthermore, the patients were grouped by their `icustay_id`, and the sequences for each patient with an AKI stage of 2 or 3 were input to account for inter-patient heterogeneity. The final preprocessed cohort comprised 24,134 unique patients with an AKI stage of 2 or 3.

2.3.5. Exploratory data analysis

This subsection aims to employ an exploratory data analysis to gain a deeper understanding of the dataset's fundamental characteristics and distributions, assess its quality, and preliminarily investigate potential relationships among variables, thereby providing a crucial foundation for the subsequent modeling work.

First, to validate the positive correlation between the stages of AKI and patient mortality, this study analyzed temporal medical data from 24,134 ICU patients in the MIMIC-III database. The results indicate that higher stages of AKI are associated with increased patient mortality, which underscores the critical importance of an accurate stage of AKI to guide stratified treatments and reduce the mortality rates. As shown in Figure 4, the mortality rate for patients with stage 3 AKI was as high as 34.66%, which is approximately 18.72 percentage points higher than that for stage 2 AKI (15.94%).

Second, a comparative analysis of demographic variables was conducted on 24,134 patients diagnosed with either stage 2 or stage 3 AKI. The overall cohort was predominantly male, middle-aged to elderly, and of a White ethnicity. Specifically, male patients accounted for 56.54% of the total, exceeding the proportion of female patients by 13.14 percentage points, while White patients constituted the majority at 72.81%. A comparative analysis between the stages revealed that the proportions of females, individuals over 80 years old, and White patients were lower in stage 3 than in stage 2, thus preliminarily suggesting that these demographic factors may not be risk factors for progression from stage 2 to stage 3 AKI. Furthermore, the data highlights the age distribution of this patient population, with the 61–80 age group comprising a significant 45.31% as shown in Table 3, thus indicating that nearly half of the patients with stage 2 or 3 AKI were elderly.

The final dataset comprised 130,616 records from 24,134 patients (stage 2: 22,109 patients; stage 3: 11,457 patients). For the comparative analysis of the 23 dynamic variables between these two AKI stages, we initially calculated the mean values and defined the inter-group "difference" (stage 3 mean minus stage 2 mean). However, this "difference" metric is scale-dependent and thus not directly comparable across different variables. Therefore, to objectively quantify the magnitude of change in

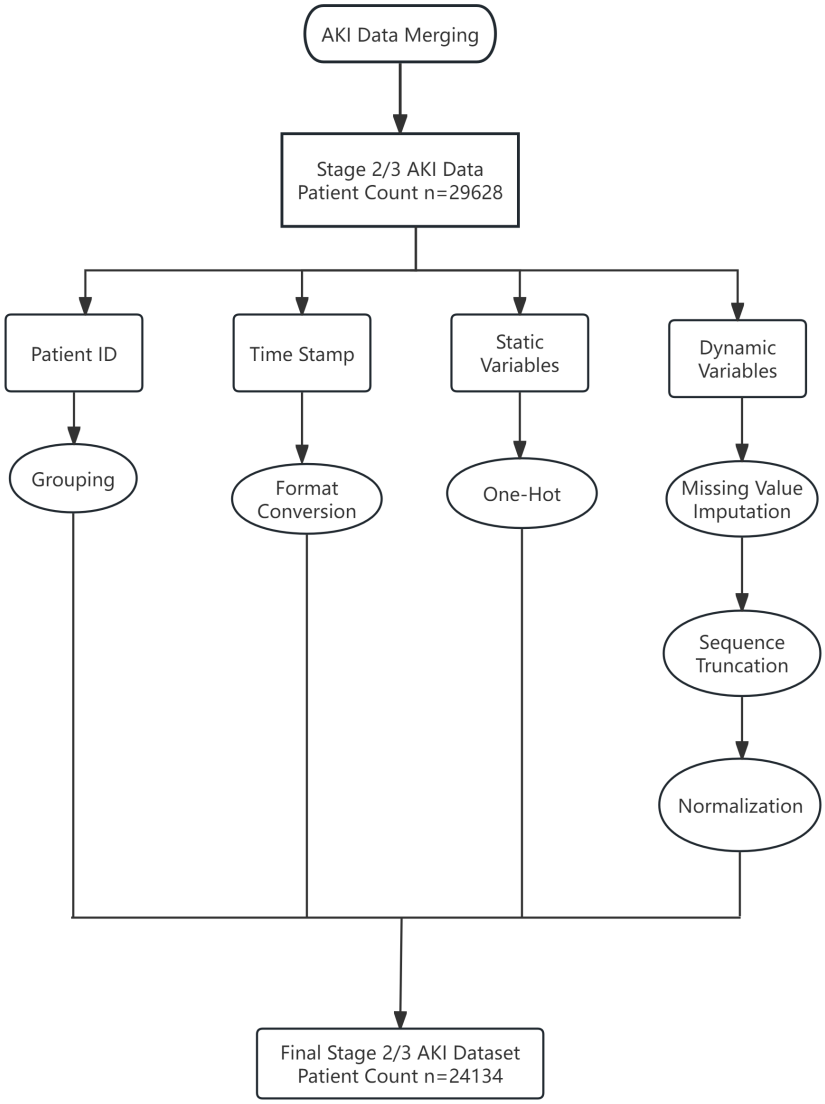


Figure 3. Flowchart of data preprocessing operations.

variables with AKI progression, we introduced a scale-free "statistical index", defined by the following formula:

$$D_{2,3} = |\frac{\bar{X}_3 - \bar{X}_2}{\bar{X}_2}| \times 100\% \tag{2.3}$$

The magnitude of the $D_{2,3}$ value was used to gauge a variable's sensitivity to the progression from AKI stage 2 to stage 3, with a larger value indicating a higher sensitivity. As shown in Table 4, the value for serum creatinine (157.66%) was substantially higher than that of other variables, followed by the 24-hour urine output (40.13%), blood urea nitrogen (27.89%), 12-hour urine output (13.81%), 6-hour urine output (13.11%), and anion gap (9.58%). This indicates that serum creatinine, urine output (across different time intervals), blood urea nitrogen, and anion gap are sensitive variables for AKI progression.

Table 3. Comparison of patient characteristics by AKI stage after multiple imputation.

Variable	AKI stage 2 (n = 12,677)	AKI stage 3 (n = 11,457)	Total (n = 24,134)
Gender			
Male	7150 (56.40%)	6496 (56.70%)	**13,646 (56.54%)**
Female	5527 (43.60%)	4961 (43.30%)	10,488 (43.46%)
Age group, years			
18–40	784 (6.18%)	770 (6.72%)	1554 (6.44%)
41–60	3187 (25.14%)	3035 (26.49%)	6222 (25.78%)
61–80	5695 (44.92%)	5239 (45.73%)	**10,934 (45.31%)**
>80	3011 (23.75%)	2413 (21.06%)	5424 (22.47%)
Race			
Asian	208 (1.64%)	267 (2.33%)	475 (1.97%)
Black	950 (7.49%)	1222 (10.67%)	2172 (9.00%)
Hispanic	315 (2.48%)	371 (3.24%)	686 (2.84%)
Native American	7 (0.06%)	8 (0.07%)	15 (0.06%)
Other	332 (2.62%)	285 (2.49%)	617 (2.56%)
Unknown	1460 (11.52%)	1137 (9.92%)	2597 (10.76%)
White	9405 (74.19%)	8167 (71.28%)	**17,572 (72.81%)**

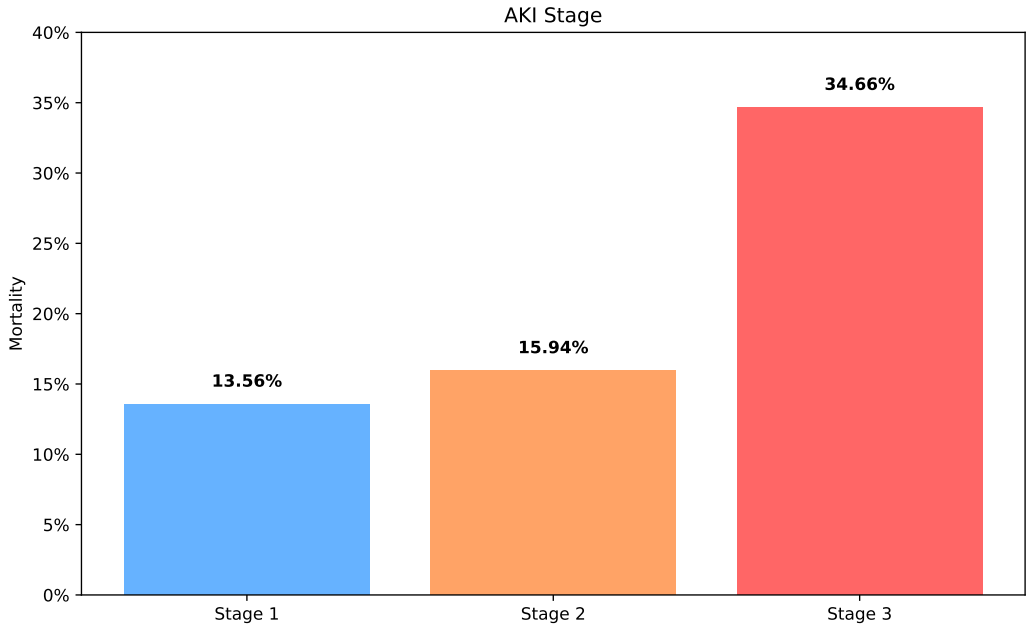


Figure 4. Mortality rate by AKI stage.

3. Results

3.1. Development of a machine learning model for direct staging of AKI

Current research on medical data often relies on small-sample cohorts and frequently fails to adequately account for the temporal dynamics of the data and the heterogeneity among patients. To

Table 4. Comparison of dynamic variables between AKI stage 2 and stage 3.

Variable	AKI stage 2 (N = 91,782)	AKI stage 3 (N = 38,834)	Difference	$D_{2,3}$
Laboratory values (Mean ± SD)				
Anion Gap	13.30 ± 3.72	14.57 ± 4.44	**−1.27**	**9.58%**
Bicarbonate	24.83 ± 5.07	24.30 ± 5.25	0.53	2.13%
Blood urea nitrogen (BUN)	26.29 ± 20.23	33.63 ± 25.23	**−7.33**	**27.89%**
Chloride	104.29 ± 6.59	103.47 ± 6.79	0.82	0.78%
Serum Creatinine	0.36 ± 0.91	0.94 ± 1.76	−0.57	157.66%
Glucose	134.17 ± 54.84	134.24 ± 60.16	−0.06	0.05%
Hematocrit	29.91 ± 4.29	30.01 ± 4.49	−0.10	0.34%
Hemoglobin	10.06 ± 1.49	10.05 ± 1.54	0.00	0.04%
Platelet Count	200.71 ± 129.11	197.68 ± 130.76	**3.03**	1.51%
Potassium	4.13 ± 0.58	4.20 ± 0.63	−0.07	1.80%
Sodium	138.26 ± 5.17	138.02 ± 5.28	0.24	0.17%
Oxygen Saturation	96.83 ± 3.06	96.70 ± 4.05	0.13	0.13%
6-hour Urine Output (mL/kg/h)	0.56 ± 0.66	0.49 ± 0.93	0.07	**13.11%**
24-hour Urine Output (mL/kg/h)	0.59 ± 0.48	0.35 ± 0.70	0.24	**40.13%**
12-hour Urine Output (mL/kg/h)	0.50 ± 0.48	0.43 ± 1.11	0.07	**13.81%**
White Blood Cell Count ($10^9/L$)	11.76 ± 9.28	12.01 ± 9.50	−0.25	2.11%
Coagulation Profile (Mean ± SD)				
International Normalized Ratio (INR)	1.63 ± 1.01	1.68 ± 1.08	−0.05	2.85%
Prothrombin Time (PT), s	16.80 ± 6.82	17.11 ± 7.30	−0.31	1.86%
Partial Thromboplastin Time (PTT), s	52.06 ± 27.99	50.98 ± 27.88	**1.07**	2.06%
Vital Signs (Mean ± SD)				
Mean Arterial Pressure (MAP), mmHg	77.63 ± 12.48	77.00 ± 13.47	0.63	0.81%
Diastolic Blood Pressure, mmHg	59.32 ± 11.42	58.90 ± 12.03	0.42	0.70%
Systolic Blood Pressure, mmHg	119.63 ± 18.85	118.44 ± 20.60	**1.19**	0.99%
Heart Rate, beats/min	85.49 ± 16.37	86.45 ± 17.00	−0.96	1.12%
Respiratory Rate, breaths/min	19.83 ± 4.86	19.97 ± 5.03	−0.14	0.69%

address these limitations, this study developed direct staging prediction models for AKI stage 2 and stage 3, thereby utilizing 37 time-series features obtained from the preprocessing steps detailed in the previous section. We employed three classical machine learning algorithms—logistic regression, random forest, and CatBoost—and complemented the modeling with an interpretability analysis. This approach aims to enhance the credibility and practical utility of the model outputs in clinical decision-making contexts.

3.1.1. Model training and evaluation

The initial cohort comprised 24,134 patients diagnosed with AKI stage 2 or stage 3, and yielded a total of 130,616 time-series records. A critical challenge emerged in adapting these data for binary classifications using conventional machine learning models—specifically logistic regression, random

forest, and CatBoost. These algorithms assume a one-to-one correspondence between the samples and the static labels; however, the AKI stage is inherently dynamic, and often changes over the course of the ICU admission. To ensure label consistency and model compatibility, each patient was assigned a single, unique outcome label based on their clinical trajectory over a fixed observation window. As a result, the dataset was collapsed from 130,616 longitudinal records to 78,616 analyzable instances, with one label per patient.

Prior to modeling, a standardized labeling strategy was implemented. The time-series data were resampled to a fixed length of 16 (corresponding to 96 hours, or 4 days), thus aligning with the truncation length determined in prior preprocessing steps. Patients who progressed to AKI stage 3 at any point within this 96-hour window were labeled as "1" (progressors); those who remained at stage 2 throughout were labeled as "0" (non-progressors). This approach defines the prediction horizon as 96 hours prior to either the onset of stage 3 (for progressors) or the end of monitoring (for non-progressors), thus enabling early risk stratification before clinical deterioration.

Then, the dataset was partitioned into training and testing sets using a $9:1$ ratio, stratified by unique `icustay_id` to prevent data leakage and ensure patient-level independence. The training set included 21,720 patients (70,904 time-point records and 21,720 labels), while the test set contained 2414 patients (7712 records and 2414 labels). Given that the raw time-series lengths varied across individuals and standard implementations of the selected models cannot process variable-length sequences, zero-padding was applied using the `pad` function to extend all sequences to a uniform length. This step preserved the temporal structure while ensuring computational compatibility.

For the model's input, each patient's full 96-hour time series was treated as a single sample. First, the two-dimensional tabular data (patients $\times$time points $\times$features) were transformed into a three-dimensional NumPy array (see Table 5 for format details), then reshaped into a two-dimensional matrix suitable for conventional classifiers. The final input dimensions were $21{,}720 \times 592$ for the training set and 2414×592 for the testing set, where 592 corresponds to the flattened feature space (37 features $\times$ 16 time points).

All experiments were conducted in Python 3.8.8 on a Windows 10 platform equipped with an Intel(R) Core(TM) i5-7200U CPU @ 2.50 GHz (dual-core, up to 2.70 GHz) and 8 GB of RAM. Key libraries included Pandas (v1.2.4) for data manipulation, NumPy (v1.22.3) for numerical operations, Matplotlib (v3.3.4) and Seaborn (v0.11.1) for visualization, and Scikit-learn (v0.24.1) for baseline model implementation.

To develop the direct staging prediction model for AKI progression, we evaluated three established machine learning classifiers: Logistic Regression, Random Forest, and CatBoost. The hyperparameters were set to default values unless explicitly optimized. A targeted hyperparameter tuning procedure was only performed for the subset of parameters listed in Table 6, thereby using either a grid search or a randomized search with cross-validation on the training set to avoid overfitting.

To evaluate the model's performance, we first trained the logistic regression, random forest, and CatBoost models on the training set. Then, their performance was assessed on the test set using the `sklearn.metrics` module, which provided accuracy, precision, recall, F1-score, Brier score, and AUC values (all metrics rounded to two decimal places, as shown in Table 7).

In the initial comparison between models, the logistic regression and random forest each demonstrated distinct strengths: the former achieved a higher precision, while the latter exhibited a

Table 5. Data structure transformation for direct prediction modeling.

Dataset name	Data structure type	Array shape	Dimensions
3D Array			
	3D Training Set	`X_train.shape`	(21,720, 16, 37)
		`y_train.shape`	(21,720,)
	3D Test Set	`X_test.shape`	(2414, 16, 37)
		`y_test.shape`	(2414,)
2D Array			
	2D Training Set	`X_train.shape`	(21,720, 592)
		`y_train.shape`	(21,720,)
	2D Test Set	`X_test.shape`	(2414, 592)
		`y_test.shape`	(2414,)

Table 6. Hyperparameter selection for the three models.

Model	Parameter	Value
Logistic Regression		
	`penalty`	L1
	`C`	0.1
	`fit_intercept`	True
	`solver`	saga
	`max_iter`	1000
	`class_weight`	balanced
	`random_state`	42
Random Forest		
	`max_features`	37
	`max_depth`	5
CatBoost		
	`iterations`	107

Table 7. Comparison of evaluation metrics for the three models.

Model	Accuracy	Precision	Recall	F1-score	Brier score
Logistic Regression	0.73	0.76	0.70	0.73	0.19
Random Forest	0.80	0.71	0.84	0.77	0.17
CatBoost	**0.86**	**0.84**	**0.87**	**0.85**	**0.10**

superior recall. To make a balanced assessment, we introduced the F1-score as a comprehensive metric. The results indicated that the random forest attained a higher F1-score than logistic regression, suggesting better overall predictive performance.

When evaluating the model's calibration performance, we considered the Brier score, where a lower value indicates better-calibrated probabilistic predictions. CatBoost achieved the lowest Brier score among the three models. Finally, a comprehensive review of the five key metrics visualized in Figure 5 clearly showed that CatBoost ranked first in all four classification performance metrics: precision,

precision, recall, and F1 score, while also achieving the optimal calibration score. Therefore, we preliminarily conclude that CatBoost demonstrates the best overall performance for the direct staging prediction of AKI.

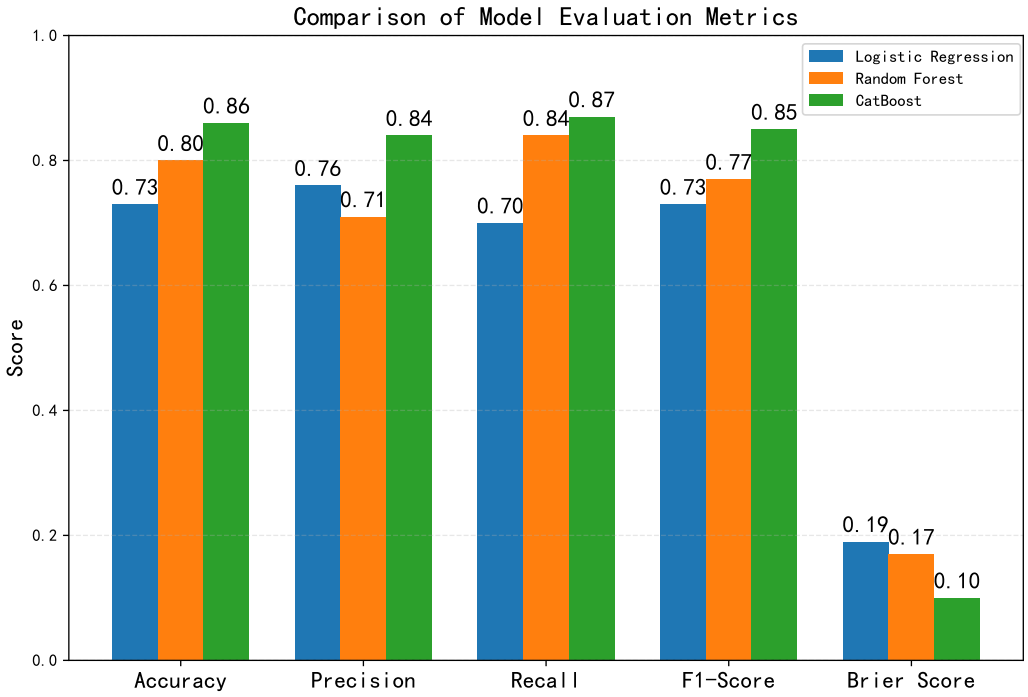


Figure 5. Comparison of evaluation metrics for the three models.

To further assess the classification capability, receiver operating characteristic (ROC) curves were plotted. The ROC curve illustrates the relationship between the false positive rate ($FPR = FP/(FP + TN)$) on the x-axis and the true positive rate ($TPR = TP/(TP + FN)$) on the y-axis across different classification thresholds. The area under the curve (AUC) quantifies the overall classification performance, with values closer to 1 indicating a better model efficacy. Figure 6 compares the ROC curves of the three models in one plot. The random forest achieved an AUC of 0.86, which indicates a good performance, whereas CatBoost reached an AUC of 0.94, thus significantly outperforming the other models and further validating its superior capability in AKI staging prediction.

3.2. LSTM-based model for dynamic AKI stage prediction

While machine learning-based direct staging models can provide a diagnostic reference for AKI stage 2 and 3, their static nature falls short of meeting the need for dynamic prediction in clinical practice. Given that the exact onset time of AKI is often uncertain, an ideal prediction system should be capable of continuous risk assessments throughout the hospitalization period to guide timely medical interventions. To address this gap, the present chapter employs an LSTM network to develop a dynamic prediction model. Furthermore, integrated gradients are incorporated to enhance the model's interpretability, with the ultimate goal of more effectively bridging the model outputs into the clinical decision-making workflow.

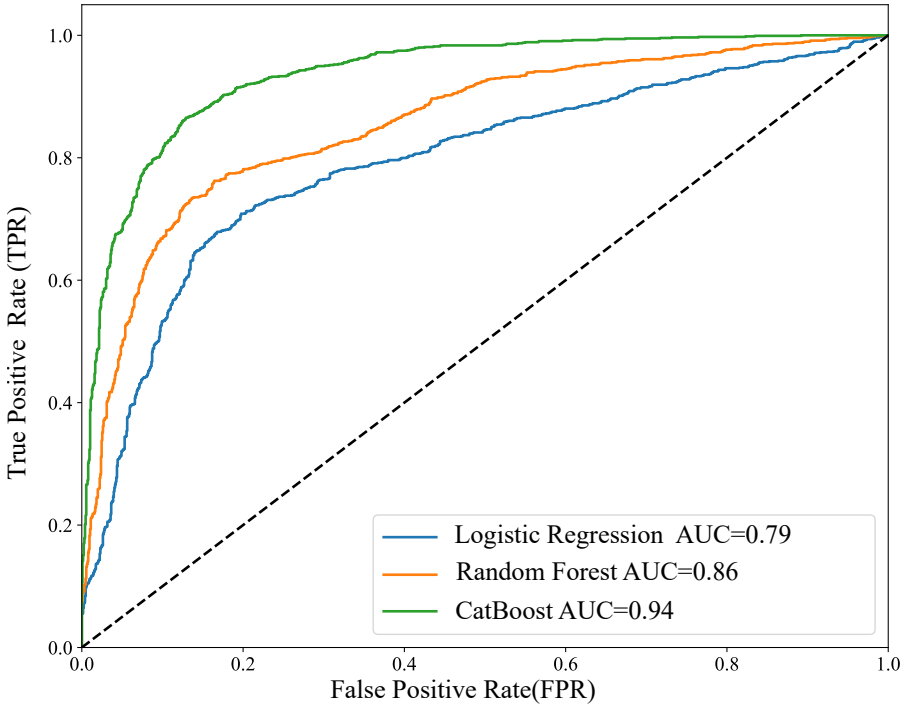


Figure 6. Comparison of evaluation metrics for the three models.

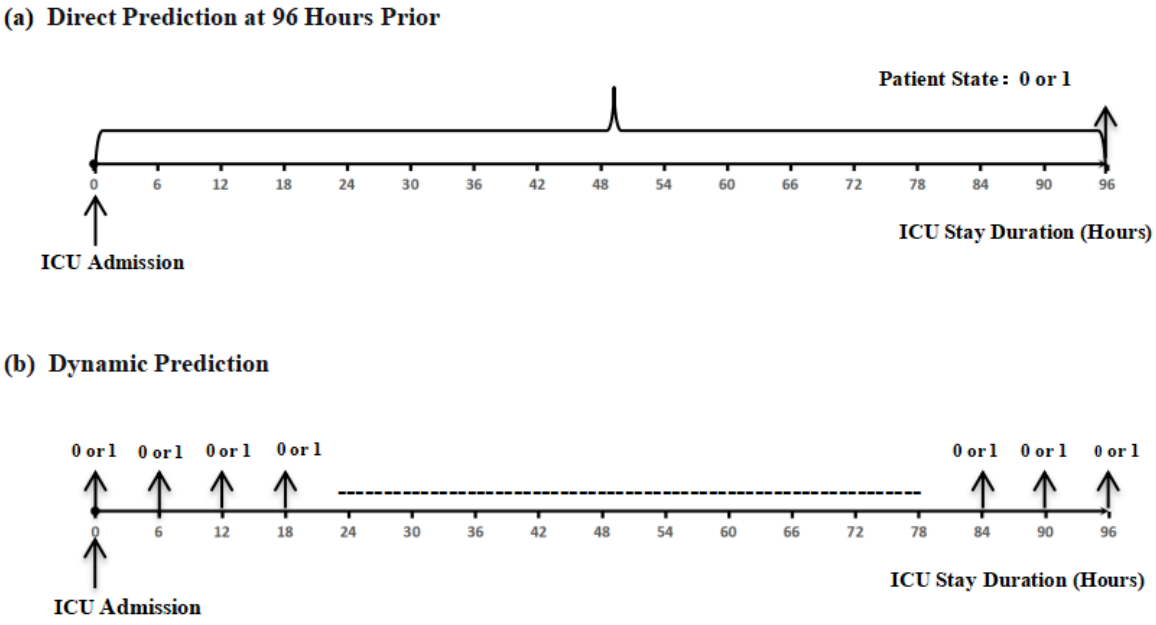


Figure 7. Direct prediction at 96 hours versus dynamic prediction.

Following the data preprocessing steps outlined in Section 2.3, a total of 24,134 patients diagnosed with an AKI stage of 2 or 3 were included, corresponding to 130,616 time-series records. The dataset was partitioned into training, testing, and validation sets in an 8 : 1 : 1 ratio based on unique patient `icustay_id` values. Specifically, the training set contained 19,307 patients with 104,575 records, the testing set included 2414 patients with 12,864 records, and the validation set consisted of 2413 patients with 13,177 records.

Figure 7 illustrates the structural differences between the 96-hour direct prediction and dynamic prediction approaches. In the direct prediction framework, only the AKI stage label at the 96-hour endpoint is retained, and the model is input with all feature variables from the preceding 96-hour period. In contrast, the dynamic prediction approach retains the state labels and corresponding feature values at each resampling time point within the 96-hour window, thereby constructing complete temporal samples. To model this sequential data, we employed an LSTM network, which captures temporal dependencies in the patient's historical feature evolution and state transitions to enable dynamic AKI stage predictions.

In deep learning, a batch refers to a subset of samples used in a single model iteration. In this study, a custom batching function was implemented to structure the training, testing, and validation sets by separating each sample into feature tensors and corresponding labels, which serve as model inputs and supervised learning targets, respectively. The function accepts `batch_size` as a parameter to control the number of samples per batch. Specifically, the training set was configured with a `batch_size` of 5 to facilitate mini-batch gradient updates, while the validation and test sets were assigned `batch_size` values of 2413 and 2414, respectively, thus effectively processing each of these sets as a single batch to enable a comprehensive evaluation of the model's performance.

As indicated by the analysis of the AKI staging labels in Section 2.1, a dynamic transition exists between stage 2 and stage 3 AKI: patients may progress from stage 2 to stage 3 due to worsening conditions, or regress from stage 3 to stage 2 following effective clinical interventions. Therefore, a patient's current staging status depends not only on historical states but also on the future progression. To capture such bidirectional temporal dependencies, this study employs a Bidirectional long short-term memory network (BiLSTM) to develop a dynamic AKI staging prediction model. The structure of the four-layer BiLSTM is illustrated in Figure 8, with the functions of each layer described as follows:

1) Input layer: this layer accepts data processed by the custom batching function, with a dimension of (5, size, 37). Here, size denotes the length of the patient's time series (i.e., the number of 6-hour resampled points), and 37 represents the feature dimension.
2) Multidimensional time series linear layer: this layer maps high-dimensional sparse input features into a low-dimensional dense representation, thus preparing the data for subsequent processing in the bidirectional LSTM layer.
3) Bidirectional LSTM layer: this layer consists of two LSTM units which operate in opposite directions, thereby modeling the sequence from forward and backward perspectives. The output at each time step is formed by concatenating the hidden states from both directions, thereby enabling the simultaneous capture of past and future contextual information and enhancing the modeling capability for temporal dynamics.
4) Fully connected layer: this layer integrates the bidirectional outputs from the four BiLSTM layers to further extract high-level feature representations.
5) Output layer: Using the rectified linear unit (ReLU) activation function, this layer estimates the

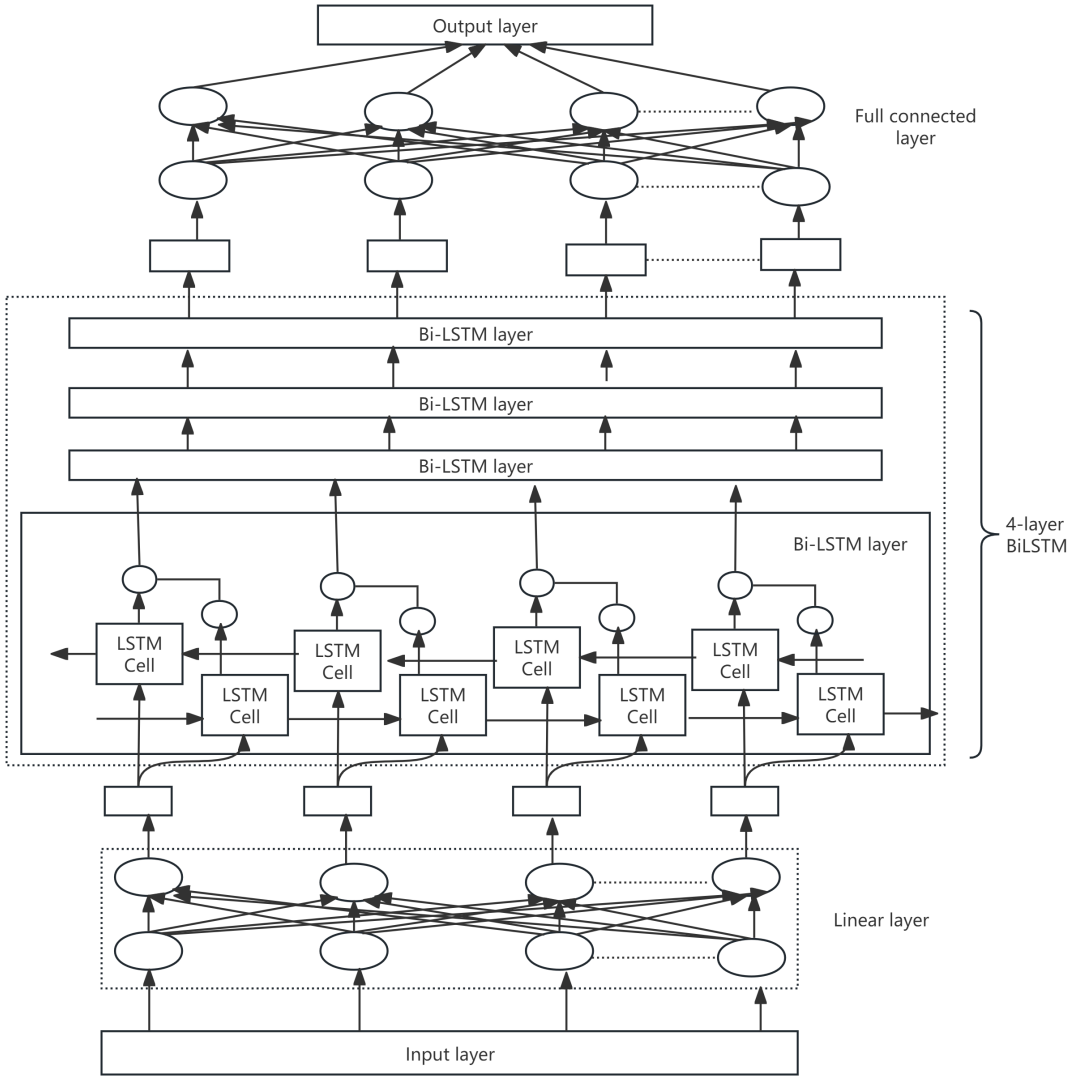
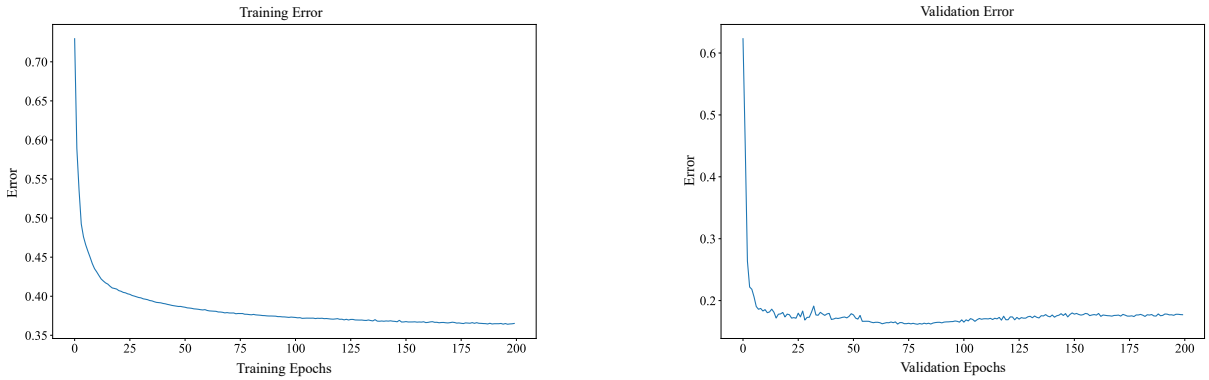


Figure 8. Structure of the 4-layer BiLSTM model.

probability of a patient belonging to AKI stage 3 (labeled as 1) or stage 2 (labeled as 0), thereby performing the binary classification task.

The model was implemented using the PyTorch 1.11.0 framework and run in a CPU environment. During training, the BCEWithLogitsLoss function was used as the loss criterion, and the Adam optimizer was applied for parameter optimization. The model parameters were iteratively updated via backpropagation and gradient descent algorithms to ultimately derive an optimal model for AKI stage prediction. After hyperparameter tuning, the model was configured with the following key hyperparameters: `features=37`, `bi_directional = True`, `n_epochs = 200`, `emb_size = round(features / 1)`, `number_layers= 4`, `input_size = 37`, `output_size = 1`.

Figure 9 displays the loss curves of the four-layer BiLSTM model over 200 training epochs. As



(a) **(b)**

Figure 9. **(a)** Training loss versus epoch during model optimization. **(b)** Validation loss versus epoch on the held-out validation set.

shown in this Figure, both training and validation errors consistently decreased with increasing epochs, thus indicating a stable training process without overfitting and demonstrating favorable convergence behavior.

Due to the extensive zero-padding applied to the test set, the resulting class imbalance limits the reliability of the confusion matrix in reflecting the true model performance. Therefore, the LSTM-based dynamic prediction model was evaluated using accuracy, AUC, and Brier scores, and its performance was compared with the direct prediction models (logistic regression, random forest, and CatBoost) described in Section 3.1.

Table 8. Performance comparison between direct and dynamic prediction models.

Model	Prediction type	Accuracy	AUC	Brier score
Logistic regression	Direct prediction	0.73	0.79	0.19
Random forest	Direct prediction	0.80	0.86	0.17
CatBoost	Direct prediction	0.86	0.94	0.10
LSTM	Dynamic prediction	**0.91**	**0.98**	**0.04**

As shown in Table 8, the proposed four-layer bidirectional LSTM model achieved an AUC of 0.98, a Brier score of 0.04, and an accuracy of 0.91 on the test set. Its AUC surpassed those of the logistic regression (0.79), random forest (0.86), and CatBoost (0.94), thus indicating that the LSTM model not only enables the dynamic prediction of AKI progression but also excels in discriminating between stage 2 and stage 3 AKI.

The four-layer bidirectional LSTM model helps mitigate the high rate of underdiagnosis often associated with conventional AKI staging methods. By providing continuous, dynamic risk assessments over time, the model can assist clinicians in promptly detecting changes in the AKI stage and adjusting treatment strategies in a timely manner, thus potentially contributing to improved patient outcomes.

4. Discussion

Although machine learning models demonstrate good accuracies in predicting AKI stages, their "black-box" nature limits their reliability in clinical applications. This study employed a feature importance analysis to identify key risk factors for the progression from AKI stage 2 to stage 3, thus providing evidence for clinical decision-making.

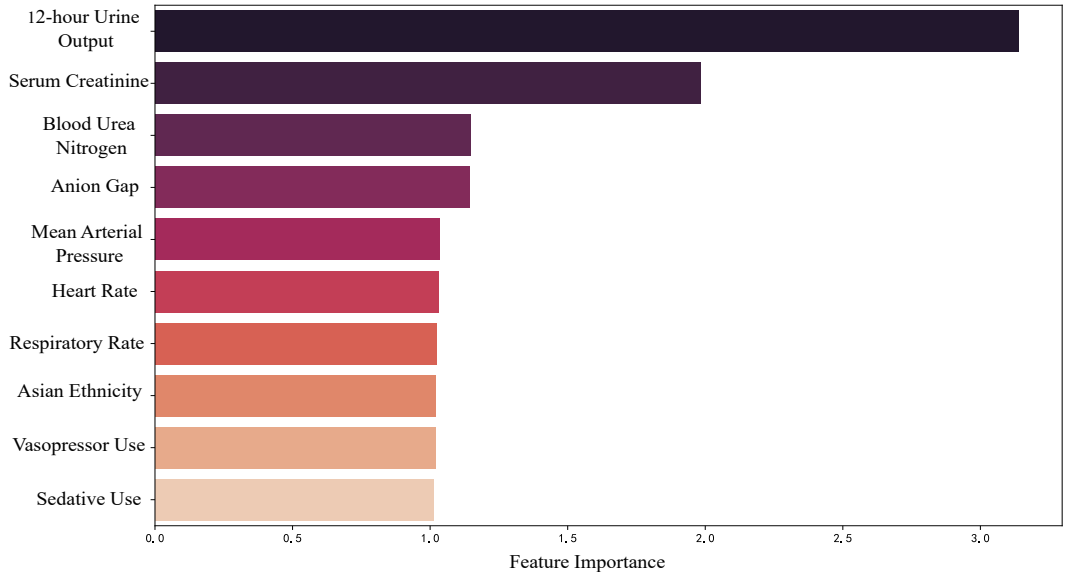


Figure 10. Top 10 features based on logistic regression feature importance analysis.

In the direct prediction models, three algorithms—logistic regression, random forest, and CatBoost—consistently identified urine output (6-hour, 12-hour, 24-hour), serum creatinine, blood urea nitrogen (BUN), and anion gap as the most important predictive variables, as shown in Figures 10–12, respectively. Notably, contrary to conventional understanding, age did not exhibit significant influence in these models, which suggests that individual differences in physiological reserve may be more predictive than chronological age once AKI progresses to a more severe stage.

The interpretability analysis of the dynamic prediction model (LSTM) further confirmed the central role of urine output, with 24-hour and 12-hour urine volumes emerging as the most prominent features in Figure 13. In addition, the model identified novel risk factors such as race (white), bicarbonate, coagulation markers, and hemoglobin, thus offering new insights into the mechanisms which underlie the progression of AKI.

A comprehensive analysis indicated that BUN and anion gap, traditionally regarded as auxiliary indicators in AKI evaluations, hold significant value in predicting clinical deterioration. The BUN/creatinine ratio reflects the state of muscle catabolism, while the anion gap indicates the degree of acid-base imbalance. Together, they suggest that AKI progression is accompanied by aggravated metabolic dysregulation.

Based on these findings, we recommend enhancing the monitoring frequency of four readily

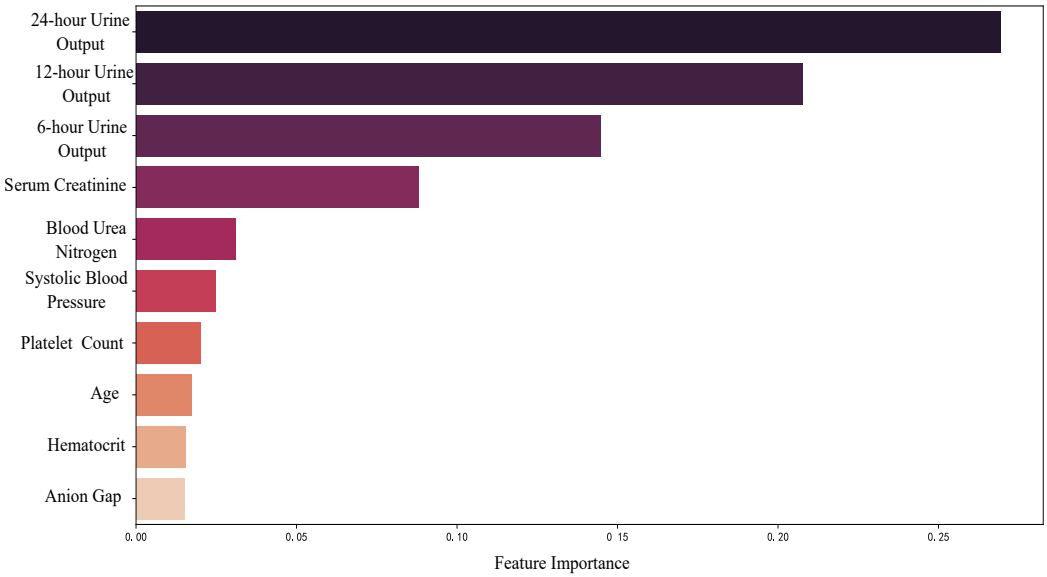


Figure 11. Top 10 features based on random forest feature importance analysis.

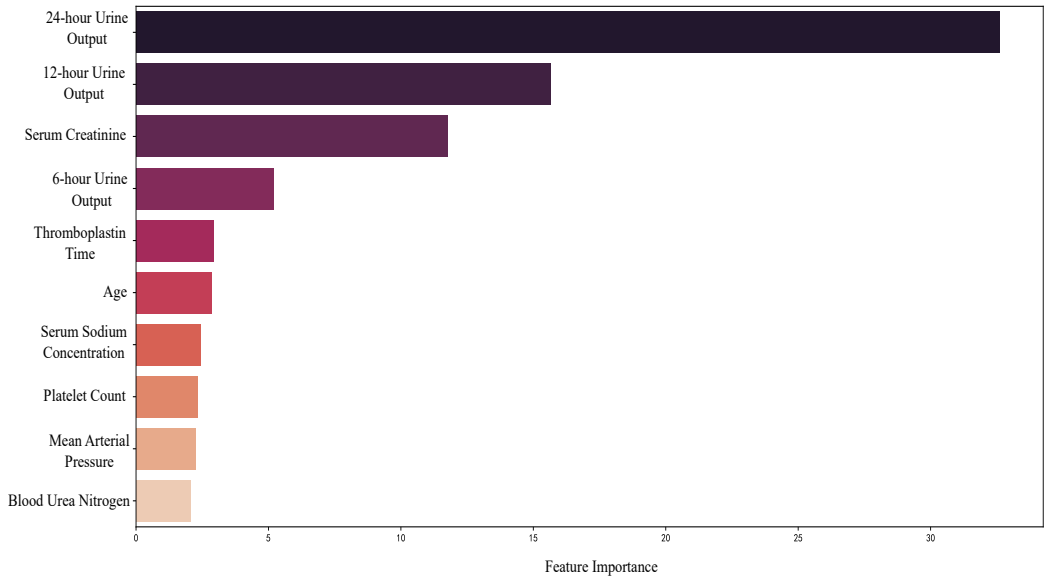


Figure 12. Top 10 features based on CatBoost feature importance analysis.

available indicators—urine output, serum creatinine, BUN, and anion gap—in clinical practice. Implementing staged management for patients at different AKI stages could facilitate early detection of clinical deterioration and help reduce the incidence of severe AKI.

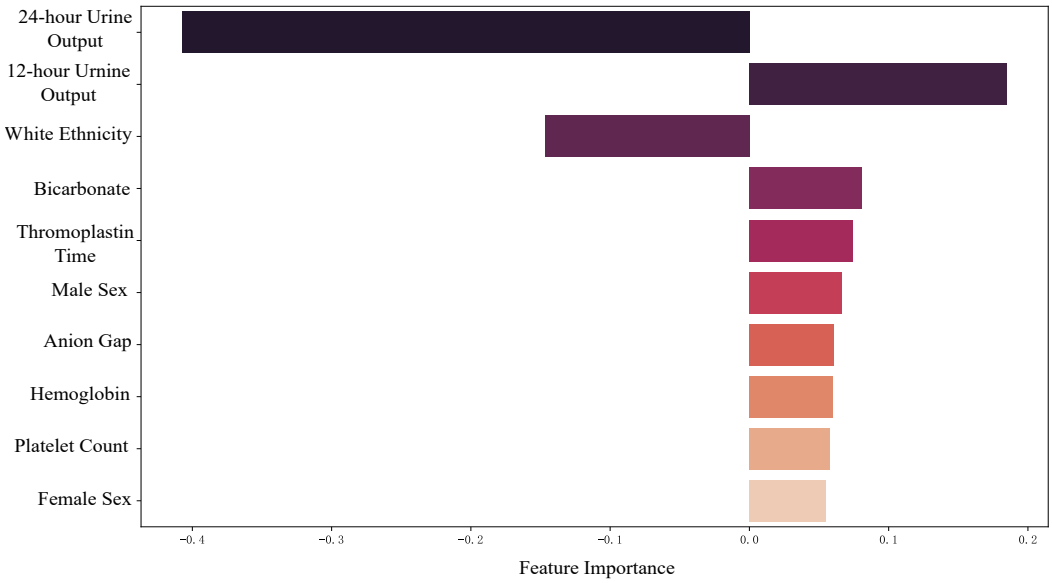


Figure 13. Top 10 Features Based on LSTM Feature Importance Analysis

5. Conclusions

The AKI is a prevalent and life-threatening complication among critically ill patients in the ICU, strongly associated with increased morbidity, the need for renal replacement therapy (RRT), and elevated mortality. The 15th ADQI International Consensus underscores the critical importance of early prediction and accurate staging of AKI—particularly stage 2 and stage 3—in the era of big data and precision medicine. Patients who progress to higher AKI stages face disproportionately worse outcomes, including higher RRT utilization and a significantly increased risk of death. Despite the KDIGO guidelines advocating stage-specific management strategies, clinical practice often fails to fully implement this framework. Many cases are either underdiagnosed or broadly categorized as stage 1, which leads to delayed recognition, suboptimal treatments, and underappreciation of the prognostic and research value of advanced AKI stages.

To address the persistent challenges of underdiagnosis, imprecise staging, and delayed detection, we developed a personalized, data-driven model for direct staging and the dynamic progression prediction of AKI stage 2 and stage 3 using high-resolution time-series data from electronic health records. Our approach enables continuous, real-time risk assessments, aiming to improve the diagnostic accuracy and support timely, stage-appropriate clinical interventions. By facilitating an earlier identification of patients at risk of deterioration, this model has the potential to reduce the incidence of stage 3 AKI, mitigate acute kidney damage, and ultimately improve outcomes in critically ill populations.

The key contributions of this study are fourfold:

(1) We established a dynamic labeling scheme aligned with the KDIGO criteria, thereby integrating both serum creatinine and urine output trajectories over time, and systematically curated a comprehensive set of demographic and clinical time-series features;

(2) We implemented a robust preprocessing pipeline—including 6-hour resampling, multiple imputation, and normalization—to address missing data and irregular sampling while preserving inter-patient heterogeneity and intra-patient temporal dynamics;

(3) In a comparative evaluation, a Bi-LSTM network demonstrated a superior performance in dynamic AKI progression prediction, thereby achieving an AUC of 0.98, and outperformed conventional machine learning models such as logistic regression, random forest, and CatBoost;

(4) Through an interpretability analysis using integrated gradients, we identified key predictors of progression from stage 2 to stage 3 AKI, thus revealing that the continuous monitoring of urine output provides earlier and more sensitive warning signals than serum creatinine changes alone.

Future work will extend this framework by incorporating the patient's comorbidities (e.g., sepsis, diabetes) to assess their influence on AKI trajectory, investigating the temporal association between nephrotoxic medication exposure and AKI progression; and developing a real-time electronic alert system integrated with the predictive model. Such a system could enable automated monitoring and prompt clinical interventions, thus enhancing the translation of AI-driven insights into routine critical care practice.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under the Grant Nos. 11702250 and the Key Scientific Research Projects of Colleges and Universities in Henan Province under Grant No.22A520010.

Conflict of interest

The authors declare that they have no conflict of interest with respect to the research, authorship, and publication of the article.

References

1. Beker BM, Corleto MG, Fieiras C, Musso CG, (2018) Novel acute kidney injury biomarkers: Their characteristics, utility and concerns. *Int Urol Nephrol* 50: 705–713. https://doi.org/10.1007/s11255-017-1781-x

2. Perazella MA, Rosner MH, (2022) Drug-induced acute kidney injury. *Clin J Am Soc Nephrol* 17: 1220–1233. https://doi.org/10.2215/CJN.11290821

3. Li JR, (2014) Acute kidney injury in the intensive care unit. *Chin J Integr Tradit West Med Intensive Criti Care* 3: 238–240. https://doi.org/10.3969/j.issn.1008-9691.2014.03.023

4. Mehta RL, Cerdá J, Burdmann EA, Tonelli M, García-García G, Jha V, et al. (2015) International Society of Nephrology's 0 by 25 initiative for acute kidney injury (zero preventable deaths by 2025): A human rights case for nephrology. *Lancet* 385: 2616–2643. https://doi.org/10.1016/S0140-6736(15)60126-X

5. Luo XY, Zhou JC, Xi XM, (2017) Diagnosis of acute kidney injury: Serum creatinine or urine output?. *Chin J Crit Care Med* 3: 9–13. https://doi.org/10.3877/cma.j.issn.2096-1537.2017.01.004

6. Dong CX, Li Y, (2017) Epidemiology of AKI: Incidence, patient mortality, and renal mortality of AKI. *Chin J Blood Purif* 16: 8–10. https://doi.org/10.3969/j.issn.1671-4091.2017.01.003

7. Hoste EAJ, Kashani K, Gibney N, Wilson FP, Ronco C, Goldstein SL, et al. (2016) Impact of electronic-alerting of acute kidney injury: Workgroup statements from the 15th ADQI Consensus Conference. *Can J Kidney Health Dis* 3: 101. https://doi.org/10.1186/s40697-016-0101-1

8. Dong GY, Qin JP, An YZ, Kang Y, Yu XY, Zhao MY, et al. (2020) Exploration of the necessity for further refinement of staging and typing in the KDIGO-AKIScr criteria for adult ICU patients: A retrospective analysis of a multicenter prospective study. *Chin Crit Care Med* 32: 313–318. https://doi.org/10.3760/cma.j.cn121430-20200218-00192

9. Li T, Yang Y, Huang J, Chen R, Wu Y, Li Z, et al. (2022) Machine learning to predict post-operative acute kidney injury stage 3 after heart transplantation. *BMC Cardiovasc Disord* 22: 288. https://doi.org/10.1186/s12872-022-02721-7

10. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. (2019) A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572: 116–119. https://doi.org/10.1038/s41586-019-1390-1

11. Song YQ, Cai GY, Guo LX, Pan Q, (2022) Prognostic characteristics and new strategies for prevention and treatment of severe acute kidney injury. *J Clin Internal Med* 39: 361–364. https://doi.org/10.3969/j.issn.1001-9057.2022.06.001

12. Duff S, Murray PT, (2020) Defining early recovery of acute kidney injury. *Clin J Am Soc Nephrol* 15: 1358–1360. https://doi.org/10.2215/CJN.13381019

13. Jiang L, Zhu Y, Luo X, Wen Y, Du B, Wang M, et al. (2019) Epidemiology of acute kidney injury in intensive care units in Beijing: The multi-center BAKIT study. *BMC Nephrol* 20: 459–468. https://doi.org/10.1186/s12882-019-1660-z

14. Sutherland SM, Chawla LS, Kane-Gill SL, Hsu RK, Kramer AA, Goldstein SL, et al. (2016) Utilizing electronic health records to predict acute kidney injury risk and outcomes: Workgroup statements from the 15th ADQI Consensus Conference. *Can J Kidney Health Dis* 3: 99. https://doi.org/10.1186/s40697-016-0099-4

15. Lngkvist M, Karlsson L, Loutfi A, (2014) A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit Lett* 42: 11–24. https://doi.org/10.1016/j.patrec.2014.01.008

16. Bellomo R, Ronco C, Kellum JA, Mehta RL, Palevsky P, the ADQI workgroup, (2004) Acute renal failure—definition, outcome measures, animal models, fluid therapy and information technology needs: The Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care* 8: R204. https://doi.org/10.1186/cc2872

17. Khwaja A, (2012) KDIGO clinical practice guideline for acute kidney injury. *Nephron Clin Pract* 120: c179–c184. https://doi.org/10.1159/000339789

18. Vagliano I, Lvova O, Schut MC, (2021) Interpretable and continuous prediction of acute kidney injury in the intensive care, In: *Public Health and Informatics*, Netherlands: IOS Press, 103–107. https://doi.org/10.3233/SHTI210129

19. Le S, Allen A, Calvert J, Palevsky PM, Braden G, Patel S, et al. (2021) Convolutional neural network model for intensive care unit acute kidney injury prediction. *Kidney Int Rep* 6: 1289–1298. https://doi.org/10.1016/j.ekir.2021.02.031

20. Harrington P, (2013) *Machine Learning in Action*, Beijing: Posts & Telecommunications Press, 85–86.

21. Brown SA, (2016) Patient similarity: Emerging concepts in systems and precision medicine. *Front Physiol* 7: 561. https://doi.org/10.3389/fphys.2016.00561

AIMS Press