

RESEARCH

Open Access



Construction and validation of prognostic models in critically ill patients with sepsis-associated acute kidney injury: interpretable machine learning approach

Zhiyan Fan¹, Jiamei Jiang², Chen Xiao¹, Youlei Chen¹, Quan Xia¹, Juan Wang¹, Mengjuan Fang¹, Zesheng Wu¹ and Fanghui Chen^{1*}

Abstract

Background Acute kidney injury (AKI) is a common complication in critically ill patients with sepsis and is often associated with a poor prognosis. We aimed to construct and validate an interpretable prognostic prediction model for patients with sepsis-associated AKI (S-AKI) using machine learning (ML) methods.

Methods Data on the training cohort were collected from the Medical Information Mart for Intensive Care IV database version 2.2 to build the model, and data of patients were extracted from Hangzhou First People's Hospital Affiliated to Zhejiang University School of Medicine for external validation of model. Predictors of mortality were identified using Recursive Feature Elimination (RFE). Then, random forest, extreme gradient boosting (XGBoost), multilayer perceptron classifier, support vector classifier, and logistic regression were used to establish a prognosis prediction model for 7, 14, and 28 days after intensive care unit (ICU) admission, respectively. Prediction performance was assessed using the receiver operating characteristic (ROC) curve and decision curve analysis (DCA). SHapley Additive exPlanations (SHAP) were used to interpret the ML models.

Results In total, 2599 patients with S-AKI were included in the analysis. Forty variables were selected for the model development. According to the areas under the ROC curve (AUC) and DCA results for the training cohort, XGBoost model exhibited excellent performance with F1 Score of 0.847, 0.715, 0.765 and AUC (95% CI) of 0.91 (0.90, 0.92), 0.78 (0.76, 0.80), and 0.83 (0.81, 0.85) in 7 days, 14 days and 28 days group, respectively. It also demonstrated excellent discrimination in the external validation cohort. Its AUC (95% CI) was 0.81 (0.79, 0.83), 0.75 (0.73, 0.77), 0.79 (0.77, 0.81) in 7 days, 14 days and 28 days group, respectively. SHAP-based summary plot and force plot were used to interpret the XGBoost model globally and locally.

Conclusions ML is a reliable tool for predicting the prognosis of patients with S-AKI. SHAP methods were used to explain intrinsic information of the XGBoost model, which may prove clinically useful and help clinicians tailor precise management.

Keywords Acute kidney injury, Sepsis, Mortality, Critical illness, MIMIC-IV database, Prognosis, Machine learning, SHAP

*Correspondence:

Fanghui Chen

chenfanghui126@163.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Sepsis-associated acute kidney injury (S-AKI) is one of the most common diseases in hospitalized and critically ill patients; it is not only associated with an increased risk of chronic kidney disease but also with high morbidity and mortality rates [1–4]. Little is known about the epidemiology of S-AKI. Sepsis causes over 5.3 million deaths annually, with approximately 30% overall mortality, especially in the ICU [5, 6]. Extrapolating from the incidence rate in the United States, Adhikari et al. estimated 19 million sepsis cases worldwide per year. However, the true incidence rate is presumably much higher. As approximately one in three patients with sepsis will develop AKI [7], the annual global incidence of S-AKI may be approximately 6 million. Nevertheless, this number is lower than the estimates extrapolated from AKI incidence. The development of AKI in patients with sepsis is associated with increased mortality [8], resulting in a heavy burden on both patients and society.

However, the pathophysiological mechanisms underlying S-AKI remain poorly understood. What is certain is that these mechanisms are consistent with the organ injury associated with sepsis, including inflammation, microcirculatory dysfunction, and metabolic reprogramming.

Considering the high incidence and mortality rates, it is necessary to establish a reliable and efficient prognostic model for S-AKI. Several risk prediction models for AKI in critically ill patients have been widely studied and established [9–11]. da Hora Passos has developed a clinical score to predict early mortality in S-AKI, which merely centered on patients treated with continuous renal replacement therapy (CRRT) rather than critically ill patients, with a small sample size and lack of external validity [12]. Furthermore, the application of general severity scores in specific cohorts is controversial because of relatively unsatisfactory discrimination and calibration. Ohnuma demonstrated that most part of AKI-assessed scores published in the twenty-first century included general severity scores with a significantly low calibration ability [13]. Recently, Hu et al. proposed and validated a specific clinical model to predict the survival of critically ill patients with S-AKI [14]. However, the prediction model was based on traditional COX regression.

In recent years, diverse machine learning (ML) algorithms, a data analysis method that develops algorithms to forecast outcomes by “learning” from data, have been examined for the early revelation of S-AKI and it outperformed the traditional statistical methods, which require no assumptions regarding input variables and their relationships with the output. The advantage

of completely data-driven learning without reliance on rules-based programming is that ML constitutes a reasonable approach. Tseng et al. [15] revealed that a prediction model established by ML techniques confirmed risk factors following cardiac surgery, which enabled the optimization of postoperative interventions to reduce the postoperative complications following cardiac surgery. Dong et al. [16] developed an ML model to learn predisease patterns of physiological measurements and predict pediatric AKI up to 48 h in advance compared with presently established diagnostic guidelines. Furthermore, Zhang et al. [17] demonstrated that the XGBoost model could separate and sort patients into those who would and would not respond to fluid intake in the urine output better than the traditional logistic regression model. Yue [18] has developed an ML model for the early identification of critically ill patients with S-AKI and showed that the XGBoost model had the best predictive performance, which can be used to assist clinicians in identifying high-risk patients to minimize the mortality. These studies above suggest that ML algorithms can improve the development and validation of prediction models in critical care research. However, the primary outcome of all studies mentioned above was AKI detection rather than poor clinical outcomes, such as mortality due to AKI. Therefore, we aimed to develop a prognostic prediction model based on ML in critically ill patients with S-AKI. Furthermore, despite the promising performance of ML algorithms in previous studies, it is difficult to explain what features of the patient are responsible for the given prediction, owing to the “black-box” nature of ML algorithms. To date, the lack of interpretability has been a major obstacle to the implementation of ML models in the medical field [19]. To interpret the results of ML models, we combined an advanced ML algorithm with a method based on SHapley Additive exPlanations (SHAP). SHAP is a popular ML technique for obtaining insights into the complicated relationships between characteristics and predictions [20]. In addition to optimizing the predictive performance of mortality risk in critically ill patients with S-AKI, this study provides intuitive explanations that will help clinicians comprehensively understand the process of how the developed model makes a particular prediction and increase the opportunity for early interventions.

Accordingly, the purpose of this study was twofold: first, we aimed to determine the best-performing ML models in the prediction of short-term mortality in S-AKI patients; second, we planned to use an interpretable ML, by combining the SHAP value to examine risk factors and quantitatively visualize the relationships between risk factors and outcomes.

Methods

Training cohort

An open and free critical care database called the Medical Information Mart for Intensive Care IV database (MIMIC-IV) version 2.2 [21–23], which is the latest version that contains comprehensive clinical data of patients admitted to the Beth Israel Deaconess Medical Center between 2008 and 2019. MIMIC-IV, an update of the MIMIC-III, incorporates contemporary data and improves numerous aspects of MIMIC-III. It catalogs >200,000 emergency department admissions and >70,000 ICU stays. The clinical data in the database consist of demographic characteristics, vital signs, imaging examinations, laboratory test results, data dictionary, and documents containing codes of the International Classification of Diseases, Ninth and Tenth Revisions (ICD-9 and ICD-10, respectively) and records of hourly physiologic data from bedside monitors validated by ICU nurses; The health information obtained from the MIMIC-IV database was unidentified, so informed consent of patients was not required [21, 24]. An author (ZY Fan) was approved to extract data from the database for research purposes (Certification No. 46451755). This database was approved by the Institutional Review Boards (IRB) of the Massachusetts Institute of Technology (MIT).

Sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated host response to infection (sepsis 3.0) [5]. Organ dysfunction may be identified as an acute and infection-related change of at least two points in the sequential organ failure assessment (SOFA) score.

AKI is identified and sorted by the basis of the highest serum creatinine (SCr) level and urine output as stated by Kidney Disease Improving Global Outcomes (KDIGO) [25]. Definition as follows: increase in SCr to ≥ 1.5 times baseline must have occurred within the prior 7 days; or a ≥ 0.3 mg/dL increase in SCr occurred within 48 h; or urine output < 0.5 mL/kg/h for 6 h or more. If the preadmission SCr was not recorded, the first SCr value at admission was used as the baseline SCr. In this study, AKI was evaluated by the worst serum creatinine and urine volume within 72 h after the suspected diagnosis of sepsis.

External validation cohort

Patients were enrolled from Hangzhou First People's Hospital Affiliated to Zhejiang University School of Medicine (Zhejiang, China) between 2018 and 2022. Adult patients who had a diagnosis of S-AKI were included. The exclusion conditions were same as training cohort. This study was reviewed and approved by

the Ethics Committee of Hangzhou First People's Hospital Affiliated to Zhejiang University School of Medicine (KY2022124).

Data extraction

We first obtained raw data using Structured Query Language with Navicat Premium software (version 15.0.12). Structured Query Language was used to extract patient data, including sociodemographic characteristics, vital signs, laboratory parameters, complications, and microbiological information [26]. Patients in the database who met the following criteria were selected for the present study: [1] first ICU admission at first hospitalization; [2] ICU length of stay > 24 h; [3] age of > 18 years; [4] and met the diagnostic criteria for sepsis 3.0 and AKI development according to the criteria. ICD-9 (99591, 99592, and 78552) and ICD-10 (R65.20, R65.21) codes were used to identify patients with sepsis in the MIMIC-IV database. Data of these patients were used as the training cohort for model establishment. The data extraction procedure is illustrated in Fig. 1.

We extracted the following demographic data: age at admission, sex, ethnicity, weight, height, length of stay in the ICU, and hospital expire flag (the recording of in-hospital death in the database) at the first ICU admission. Next, the vital signs of the patients in the first 24 h of ICU stay, including mean arterial pressure (Meanbp), heart rate, temperature, respiratory rate, oxyhemoglobin saturation (SpO₂), urine output and then laboratory parameters in the first 24 h, including routine blood examination, liver and kidney function, blood glucose, and arterial blood gas (ABG), were collected. In addition, advanced life support recordings, such as mechanical ventilation and renal replacement therapy, were recorded. Comorbidities were identified using the Charlson table in materialized view.

To filter for missing data, the `missingno` module in Python 3.9.12 software was used. In Fig. 2, each column represents a clinical variable, and the white line represents the missing data. The denser white lines in each column, the more missing values there are for that variable. Detailed information regarding missing values is provided in Additional File 1. We removed variables missing $> 30\%$ of observations, such as height and serum albumin levels, to facilitate and ensure study accuracy. Missing values were imputed using multivariate imputation by chained equations [27]. The maximum, minimum, and mean values were used when incorporating the characteristics of vital signs and related laboratory parameters and were considered as independent features to be included in the study.

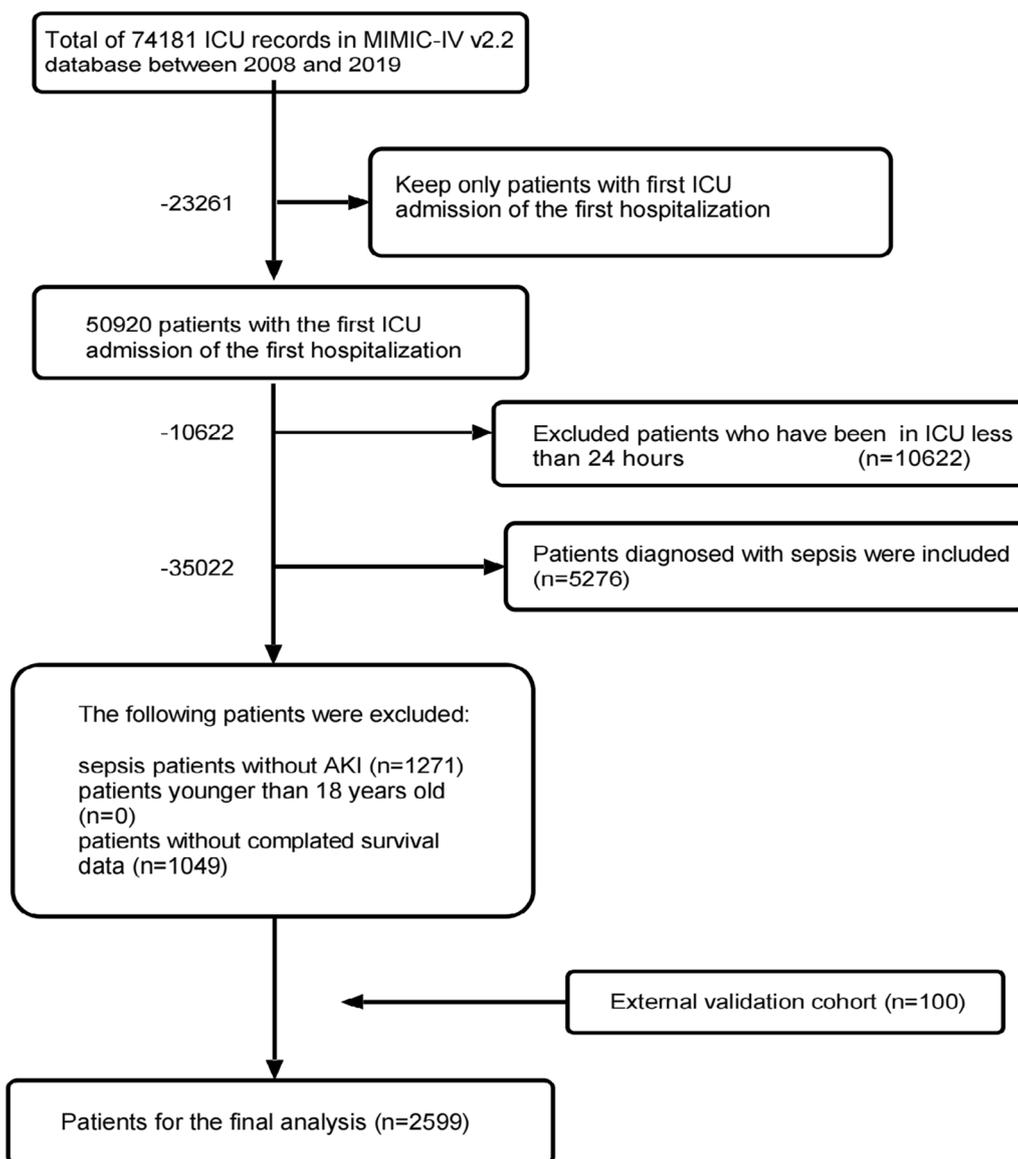


Fig. 1 Flowchart of screening

Statistical analyses

Normality testing was performed by Shapiro-Wilks test. Continuous variables with normal distributions are presented as the mean (SD, standard deviation) and compared with independent samples *t* tests. Non-normally distributed variables are expressed as the median (interquartile ranges), which were compared with Kruskal-Wallis test. Categorical variables were described as percentages and were compared using the chi-square test. Patients were categorized into “survival” and “non-survival” groups, according to their survival status within 7, 14, or 28 days. Specifically, these were 7-day survival and non-survival groups, 14-day survival and

non-survival groups, and 28-day survival and non-survival groups. Variables are displayed and compared in groups of 7 days in Table 1.

For ML models, scikit-learn Python library (version 1.2.1) and XGBoost (version 1.7.3) packages were used to create models and tune the hyperparameters in Python. During the model-building stage, the “Min-MaxScaler” method in the module of “sklearn.preprocessing” was used to scale the data of continuous variables, whereas “OneHotEncoder” method in the module of “sklearn.preprocessing” was used to encode the data of categorical variables. We randomly divided the training cohort patients and allocated 80% to the

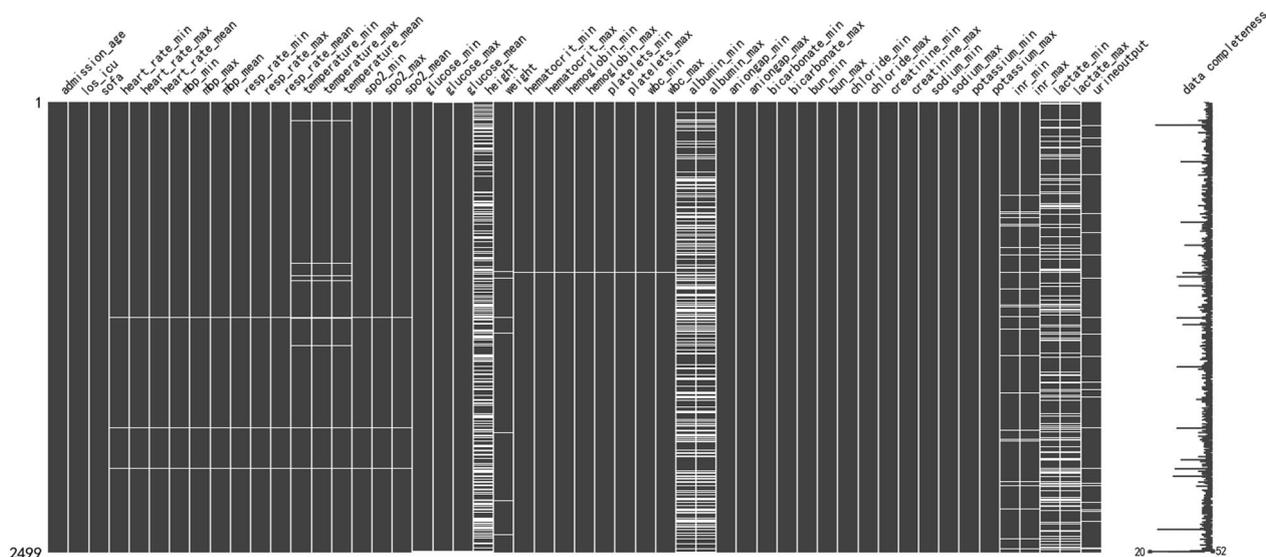


Fig. 2 Missing data distribution each column represents a clinical variable and the white line represents the missing data. The more white lines in each column, the more missing values for that variable

training set and 20% to the internal validation cohort. The training set was pretreated using the synthesizing minority oversampling technology (SMOTE) with the Tomek link (SMOTETomek) technique to balance positive and negative categories [28]. A recursive feature elimination (RFE) algorithm was used for the feature selection. The ML algorithms considered in this study included random forest (RF), support vector classifier (SVC), logistic regression (LR), XGBoost, and multilayer perceptron classifier (MLP). These were used to construct prediction models. Hyperparameter optimization and cross-validation through GridSearchCV were applied to prevent overfitting and increase model accuracy.

XGBoost is a tree ensemble technique based on the loss generated by weak decision tree-based learners. XGBoost was trained as the baseline model, followed by the training of the final model with optimized hyperparameters. The XGBoost model hyperparameters were tuned using the Scikit-learn GridSearchCV with tenfold cross-validation. The hyperparameters chosen for optimization were learning_rate, gamma, max_depth, subsample, min_child_weight, and n_estimators. The GridSearchCV method of scikit-learn with tenfold cross-validation also tunes the hyperparameters of the SVC, RF, MLP, and LR.

The prediction performance of the five models were assessed by ROC curves and DCA. What’s more, the accuracy, precision, recall, and F1 score of models were also evaluated. The external validation cohort was used to validate the performance of the five models mentioned above by same ways.

SHAP is a flexible method which can be used to explain individual predictions and for global interpretation. It has a substantial theoretical foundation in game theory and uses the concept of allocating optimal credits based on Shapley values to estimate the importance of features. SHAP force plots provide an intuitive visualization of how different features affect an individual prediction. One advantage of SHAP for global interpretation is that SHAP not only reveals about the importance of features but also their relationship with the output. Additionally, SHAP’s predictions are reasonably distributed among feature values. These factors are crucial in guaranteeing trust in the technique [29]. In our work, SHAP feature importance assessment were used for global interpretation of the developed baseline model (Fig. 6). SHAP was also used to come up with examples on how individual predictions can be explained locally (Fig. 7).

Results

A total of 2599 patients with S-AKI were included in this study, with 2499 included in the training cohort and 100 in the external validation cohort. Patients were categorized into “survival” and “non-survival” groups, according to their survival status within 7, 14, or 28 days. Variables are displayed and compared in groups of 7 days in Table 1.

Baseline characteristics

Table 1 shows the overall baseline characteristics, vital signs, and laboratory parameters of the training cohort based on the 7-day group. The overall mortality of

Table 1 The characteristics of training cohort patients when first ICU admission

Variables	Survival (n=1787)	Non-survival(n=712)	P-value
Age (year)	72 (62, 83)	75 (63, 84)	0.0565
Gender (%)			
Female	773 (43.3)	337 (47.3)	0.0675
Male	831 (57.3)	253 (58.4)	
Ethnicity (%)			
White	1209 (67.7)	470 (66.0)	<0.001
Black	176 (9.8)	42 (5.9)	
Yellow	60 (3.4)	22 (3.1)	
Others	342 (19.1)	178 (25.0)	
Weight (kg)	78.0 (64.9, 95.9)	76.8 (63.4, 94.0)	0.13
Admission type (%)			
Elective	191 (10.7)	75 (10.5)	0.722
Emergency	1095 (61.3)	426 (59.8)	
Urgent	501 (28.0)	211 (29.6)	
First Care unit (%)			
MICU	739 (41.3)	299 (42)	0.213
SICU	706 (39.5)	284 (39.8)	
CCU	80 (11.9)	64 (9.0)	
TSICU	161 (9.0)	49 (6.9)	
Others	101 (5.7)	16 (2.2)	
SOFA	9 (6, 12)	11.0 (8, 14)	<0.001
AKI stage (%)			
I	227 (12.7)	46 (6.5)	<0.001
II	627 (35.1)	158 (22.2)	
III	933 (52.2)	508 (71.3)	
Length of ICU stay (day)	6.54 (3.0, 12.1)	2.86 (1.8, 4.6)	<0.001
Comorbidity			
Congestive heart failure, n (%)	a720 (40.3)	242 (34.0)	0.00356
Chronic pulmonary disease, n (%)	a229 (12.8)	83 (11.7)	0.461
Diabetes complicated, n (%)	227 (12.7)	64 (9.0)	0.00862
Renal disease, n (%)	574 (32.1)	182 (14.8)	0.00123
Liver disease, n (%)	452 (25.3)	226 (31.7)	0.0012
Solid tumor, n (%)	385 (21.5)	179 (25.1)	0.0563
Metastatic cancer, n (%)	187 (10.5)	132 (18.5)	<0.001
AIDS, n (%)	16 (0.9)	4 (0.6)	
Vital signs ^a			
Heartrate_mean (min ⁻¹ ,mean SD)	90.5 ± 17.2	96.2 ± 18.1	<0.001
Heartrate_max(min ⁻¹ ,mean SD)	112 ± 23.1	117 ± 23.3	<0.001
Meanbp_mean (mmHg)	71.9 (66.9, 77.3)	69.9 (64.7, 75.3)	<0.001
Meanbp_min (mmHg)	53 (47, 59)	50.0 (41, 59)	<0.001
Resprate_mean (min ⁻¹)	20.4 (17.7, 23.7)	22.4 (19.4, 25.8)	<0.001
Resprate_max(min ⁻¹)	29.0 (25, 34)	31.0 (27, 36)	<0.001
Temperature_mean (°C)	36.8 (36.6, 37.2)	36.7 (36.4, 37.2)	0.809
Temperature_max (°C)	37.3 (37, 38)	37.2 (36.8, 38)	0.0148
SpO ₂ _mean (%)	97.0 (95.6, 98.4)	96.2 (94.4, 97.8)	<0.001
SpO ₂ _min (%)	92.0 (89, 94)	90.0 (86, 93)	<0.001
Urine output (mL)	1000 (556.5, 1625)	579 (223.8, 1102)	<0.001
Laboratory parameters ^b			
Aniongap_max (mmol/L)	18.0 (15, 21)	20.0 (17, 24)	<0.001

Table 1 (continued)

Variables	Survival (n=1787)	Non-survival(n=712)	P-value
Bicarbonate_min (mmol/L)	19.0 (16, 23)	17.0 (13, 20)	<0.001
Lactate_max (mmol/L)	2.4 (1.6, 4.2)	3.7 (2.2, 7)	<0.001
Glucose_mean (mg/dL)	135 (111.4, 170.7)	135 (104.5, 189.0)	0.804
Creatinine_max (mg/dL)	1.70 (1.1, 2.7)	2.0 (1.3, 3.1)	<0.001
Chloride_min (mmol/L)	101 (97, 106)	100 (96, 106)	0.0266
Chloride_max (mmol/L)	106 (101, 111)	106 (100, 111)	0.569
Hematocrit_min (%)	28.3 (24.2, 33)	28.5 (23.9, 33.4)	0.987
Hemoglobin_min (g/dL)	9.20 (7.9, 10.7)	9.15 (7.7, 10.8)	0.459
Platelet_min (10 ⁹ /L)	159 (98.5, 234)	138 (68, 213.25)	<0.001
Potassium_min (mmol/L)	3.80 (3.5, 4.3)	4.0 (3.5, 4.5)	<0.001
Potassium_max (mmol/L)	4.50 (4.1, 5.2)	4.80 (4.2, 5.5)	<0.001
Sodium_max (mmol/L)	140 (136, 143)	140 (136, 143)	0.837
INR_min	1.40 (1.2, 1.7)	1.50(1.2, 1.9)	<0.001
INR_max	1.50 (1.3, 2.1)	1.80 (1.4, 2.7)	<0.001
BUN_max (mmol/L)	29.0 (22, 56)	34.0 (28, 64)	<0.001
WBC_max (10 ⁹ /L)	35.0 (10.5, 21.9)	42.0 (10.3, 23.7)	<0.001
Advanced life support			
Renal replacement therapy (%)	125 (7.0)	70 (9.8)	0.0205
Mechanical ventilation (%)	1058 (59.2)	492 (69.1)	<0.001
Sources of infection, n (%)			
Blood	50 (2.8)	55 (7.7)	<0.001
Urine	201 (11.2)	55 (7.7)	0.00845
Sputum	196 (11.0)	28 (3.9)	<0.001
MRSA screen	36 (2.0)	9 (1.3)	0.245
Other	52 (2.9)	11 (1.5)	0.0587

Categorical data were showed as frequency (percentage). Continuous variables with normal distributions were presented as the mean (SD, standard deviation) and compared with independent samples t tests. Non-normally distributed variables are expressed as the median (interquartile ranges), which were compared with Kruskal-Wallis test

^a Vital signs data were calculated as mean/minimum/maximum value during the first 24h of ICU admission

^b The laboratory parameters recorded the mean/minimum/maximum value during the first 24h since ICU admission of each included patients

SD standard deviation, CCU coronary care unit, MICU medical intensive care unit, SICU surgical intensive care unit, TSICU trauma/surgical intensive care unit, SOFA sequential organ failure assessment, AIDS Acquired Immune Deficiency Syndrome, Meanbp mean arterial pressure, SpO₂ oxyhemoglobin saturation, INR international normalized ratio, BUN blood urea nitrogen, WBC white blood cell, MRSA methicillin-resistant *Staphylococcus aureus*

patients with S-AKI within 7 days was 28% (n=712) in the training cohort. In univariate analysis, age at admission; AKI stage; SOFA score; comorbidities such as congestive heart failure, diabetes complicated, and metastatic cancer; vital signs such as heart rate, respiratory rate, and SpO₂; indicators of ABG such as lactate_max, aniongap_max, and bicarbonate_max; blood routine indicators such as platelet count, white blood cell count, serum potassium level, mechanical ventilation, and positive blood culture and sputum culture were considered significant between the groups.

Features selected in models

This study used the RFE algorithm to select features from the data of training cohort. According to a specific feature ranking standard, RFE starts from a complete set

and then eliminates the least relevant feature one by one to select the most important features. Finally, the top of 40 important features were selected by RFE in the three groups, respectively. The order of feature importance was showed in Fig. 6 with SHAP method.

Model comparison

In the model development and validation stage, we first determined optimal hyperparameters of the XGBoost model for the 7-day group: learning_rate=0.6, gamma=0.9, max_depth=3, subsample=0.799, min_child_weight=1, and n_estimators=2000. The optimal hyperparameters of the XGBoost model for the 14-day group were learning_rate=0.5, gamma=0.6, max_depth=3, subsample=0.7, min_child_weight=1, and n_estimators=2000 and those for the 28-day group

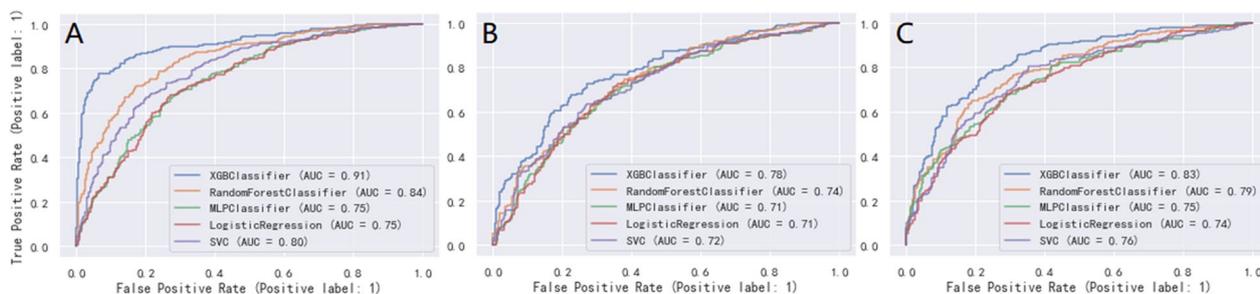


Fig 3 The ROC curves comparison of the five models with the three groups in training cohort. blue line = XGBoost model, organ line = random forest classifier model, green line = multilayer perceptron classifier model, red line = logistic regression model, purple line = support vector classifier model; **A:** 7-day group, **B:** 14-day group, **C:** 28-day group

were learning_rate=0.1, gamma=0.1, max_depth=5, subsample=0.799, min_child_weight=1, and n_estimators=2000. Detailed information regarding hyperparameters of other ML models is provided in Additional File 2. The final models were trained using optimized hyperparameters.

The five ML models (LR, RF, XGBoost, MLP and SVC) demonstrated good discriminative power with AUCs (95%CI) of 0.75 (0.73, 0.77), 0.84 (0.82, 0.86), 0.91 (0.90, 0.92), 0.75 (0.73, 0.77), 0.80 (0.78, 0.82) in the 7-day group, 0.71 (0.69, 0.73), 0.74 (0.72, 0.76), 0.78 (0.76, 0.80), 0.71 (0.69, 0.73), 0.72 (0.70, 0.74) in the 14-day group, and 0.74 (0.72, 0.76), 0.79 (0.77, 0.81), 0.83 (0.81, 0.85), 0.75 (0.73, 0.77), 0.76 (0.74, 0.78) in the 28-day group, respectively. ROC curve comparisons of the five models with the three groups in the training cohort are shown in Fig. 3. The XGBoost algorithm model showed the highest AUC in the 7-, 14-, and 28-day groups. Performance of the RF model was second only to XGBoost, and significantly better than that of the other three models. Results of the F1-score, accuracy, precision and recall of the three groups are shown in Fig. 4. Performance of the XGBoost classification model was better than that of the others in the three groups. According to the DCA results of the five prediction models (Fig. 5), the net benefit of XGBoost was significantly larger than that of the other models in all three groups.

Interpretability analysis

Initially, the global interpretability of baseline model was studied. The XGBoost model was regarded as the baseline model as it was found to be the best performing model. The feature importance estimates were based on overall samples of training cohort. The global importance of each feature we estimated in SHAP was used to understand the general impact of various features across all samples (see Fig. 6). The summary plot showed all of the 40 features in 7-, 14-, and 28- day groups.

The SHAP summary plot illustrated the entire distribution of each feature’s impact on the model output. The color allowed us to understand how changes in the value of a feature affected the change in outcome. Red represents a high feature value, whereas blue represents a low feature value. The further away a point is from the baseline SHAP value of zero, the stronger it effects the output. This way a features relationship with the SHAP value (and in turn the predicted output) can be better understood. In these three groups, the patient’s SOFA score when admitted to the ICU, indicators reflecting circulatory dysfunction (meanbp_max, lactate_min, and urine output), the mean value of SpO₂ within the first 24 h after ICU admission, AKI stage played a crucial role compared with the other risk factors (e.g., platelets, hematocrit, hemoglobin) of which the distributions of SHAP values by and large were crowded in the center. The direction of effects revealed that high SOFA with a long-right tail led to a high risk of death, whilst high urine output with a long-left tail were also significant and inversely related to predicted death.

Figure 7 illustrates how the SHAP method can be used to explain individual model predictions. Four examples were shown in the figure. It represented an intuitive way to guide the decisions of clinicians and patients and improve their understanding of how the developed model makes a particular prediction. The force plots start at the base value (the average of all predictions). Each predictor (and its corresponding Shapley value) is represented by an arrow which either increases (shown in red) or decreases (shown in blue) the model’s predicted value with respect to the base value. A predictor’s importance is shown by the size of its arrow, where a larger arrow represents a more important predictor. Feature values were listed at the bottom of the plot. Finally, the predicted output value of the model is illustrated by the point where the red and blue arrows meet. Figure 7 (A, B and C) showed the XGBoost model predicted values

for three individuals died within the admission of ICU after 7-day, 14-day and 28-day. In Fig. 7 (A), within the first 24 h of ICU admission, the value of urine output of 34.0 ml, the maximum value of International normalized ratio (INR) of 13.6, the maximum value of mean arterial pressure of 57.15 mmHg, age of 91 years old, the SOFA score of 16 when he admitted to ICU, greatly drove to the death of this patient. However, within the first 24 h of ICU admission, the minimum of heart rate (heart_rate_min) of 60.0/min inversely related with predicted death. Contrary to the other three, the patient's ultimate outcome was survival with the output value of -3.30 (Fig. 7 D).

External validation

We validated the external cohort enrolled from Hangzhou First People's Hospital Affiliated to Zhejiang University (Zhejiang, China) between 2018 and 2022. Comparison of variables between the training and external validation cohorts is shown in Additional File 3. Patients in the external validation cohort were older than those in the training cohort. Further, there were fewer patients with AKI stage III in the external validation cohort than in the training cohort. Compared with the training cohort, patients in the external validation cohort had lower body weight, higher SOFA scores, and higher mean arterial pressure and mean respiratory rate. In the external validation cohort, the AUCs (95%CI) of 0.70 (0.68, 0.72), 0.78 (0.76, 0.80), 0.81 (0.79, 0.83), 0.72 (0.70, 0.74), 0.64 (0.62, 0.66) in the 7-day group, 0.68 (0.66, 0.70), 0.69 (0.67, 0.71), 0.75 (0.73, 0.77), 0.59 (0.57, 0.61), 0.69 (0.67, 0.71) in the 14-day group, and 0.67 (0.65, 0.69), 0.67 (0.65, 0.69), 0.79 (0.77, 0.81), 0.73 (0.71, 0.75), 0.69 (0.67, 0.71) in the 28-day group were obtained with the LR, RF, XGBoost, MLP, and SVC models, respectively (Fig. 8). XGBoost showed the best performance among all models, especially compared with the SVC and MLP models, respectively. Figure 9 shows the comparison of F1 score, accuracy, precision, and recall among the five models. XGBoost had the best comprehensive performance.

Models display and application

In order to facilitate the application of our results to clinicians, related researchers, patients and their families, we have developed this prognostic prediction system, which can be assessed at the following websites: <https://hanmya-streamlit-pred-20230419streamlit40-model-tt9kpe.streamlit.app/> [30].

Discussion

We constructed and validated prediction ML models for prognosis prediction in critically ill patients with S-AKI and improved the interpretability of ML. This study, analyzed 78 features on demographic data, vital signs, and laboratory indicators in the first 24 h after critical care admission; microbiological culture; advanced life support data; and comorbidities using RFE. Forty features were selected to build ML models.

Age at admission; AKI stage III; vital signs within the first 24 h of ICU admission including respiratory rate (mean, min), temperature (mean, min), mean arterial pressure, SpO₂_min and SpO₂_mean, urine output and SOFA score; admitted to ICU, INR; lactate_min and bicarbonate_max of arterial blood; and serum creatinine_max were main variables (for details of selected features, see Fig. 6).

The XGBoost classifier model exhibited the best performance among the five ML classifiers; therefore, this model was used as the baseline model. Meanwhile, using SHAP values and plots, we demonstrated that the ML method could explain key features and establish a high-accuracy mortality prediction model in critically ill patients with S-AKI. The illustration of cumulative domain-specific feature importance and visualized interpretation of feature importance permit physicians to understand the fundamental features of XGBoost intuitively.

This study has made several contributions. First, we introduced the XGBoost algorithm, which has gained popularity in recent years, because of its fast computation, good generalization and high predictive performance [18, 31, 32]. Hyperparameter optimization based on GridSearchCV and SMOTETomek resampling techniques was also used.

Second, the DCA curve was plotted for the clinical application of the XGBoost classifier model and comparison with the other ML models. In the ROC curve comparisons among the five models, XGBoost displayed the best discrimination in the three groups, in the training and external validation cohorts (Figs. 3 and 8). A model is not always clinically useful, even with good discrimination [33]. Clinical intervention guided by the XGBoost model provided a greater net benefit in the training cohort when the threshold probability were 0.2–0.9 (Fig. 5A), 0.3–0.7 (Fig. 5B), and 0.2–0.8 (Fig. 5C). The DCA showed that the XGBoost model had the maximum benefit across the reasonable threshold probabilities, which means the XGBoost model is the optimal and other ML models inferior. In conclusion, the DCA showed that the prognosis prediction model based on XGBoost had a higher clinical application value and better clinical practicability.

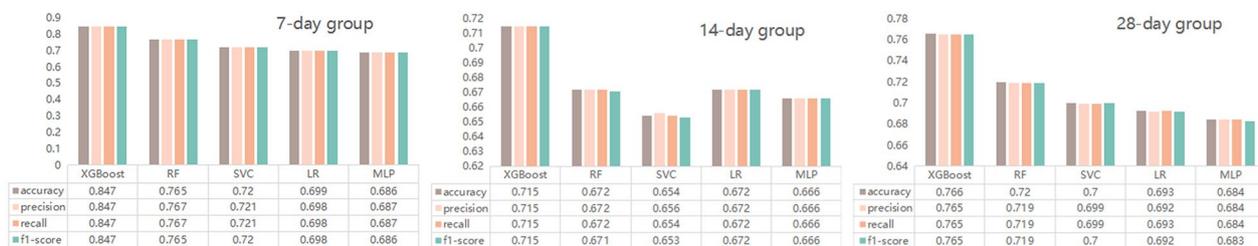


Fig 4 The comparison of performance in the five models with the three groups in training cohort. RF, Random forest model; XGBoost, Extreme Gradient Boosting; MLP, Multi-layer Perceptron classifier; SVC, Support vector Classifier; LR, Logistic regression

Third, one advantage of our study is that we used SHAP values to uncover the black box of ML. The SHAP summary plot illustrated the entire distribution of each feature’s impact on the model output. Sepsis is often

to quantify organ impairment by measuring the burden of organ malfunction in severely ill patients, which incorporates measures of cardiovascular, hemostatic, and renal dysfunction [5]. In recent years, the SOFA score has

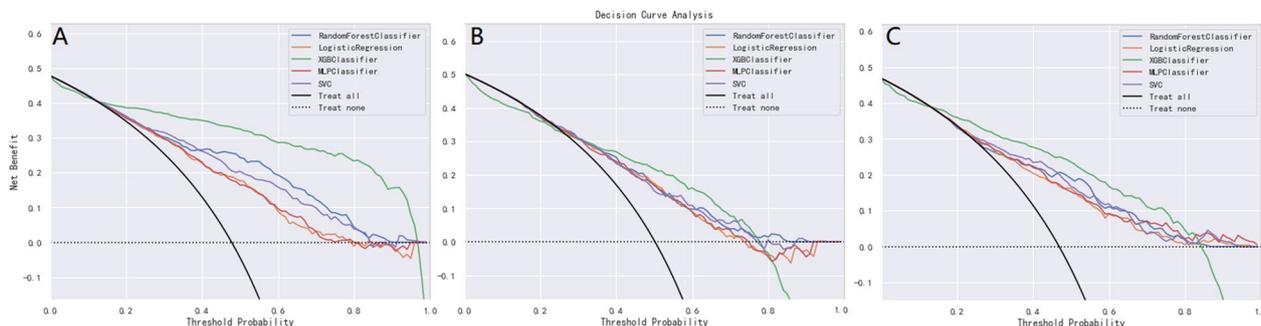


Fig 5 The DCA comparison of the five models with the three groups in training cohort. Decision curve analysis (DCA) of the five prediction models. The net benefit curves for the prognostic models are shown. X-axis indicates the threshold probability for critical care outcome and Y-axis indicates the net benefit. Green line = XGBoost model, blue line = random forest model, organ line = logistic regression model, red line = multilayer perceptron model, purple line = support vector classifier model; **A**: 7-day group, **B**: 14-day group, **C**: 28-day group

accompanied by hypotension and insufficient oxygen supply to the organs. As renal tubules receive a marginal oxygen supply and have high oxygen consumption under physiological circumstances, they are prone to hypoxia and consequent tubular necrosis, which has long been synonymous with AKI [34]. In the 7-day group, SOFA score, admitted to ICU, urine output, meanbp_mean, lactate_min, and SpO₂_mean were 5 of the 10 most important features. These indicators directly or indirectly reflect the hemodynamic status and tissue oxygenation of patients with S-AKI. In the summary plot (Fig. 6A), the deterioration of these indicators greatly contributed to the patient mortality within 7 days of ICU admission.

The composition of the top 20 most influential variables was roughly the same in the summary plot of the three groups and mainly consisted of AKI stage, circulatory status indicators, basic vital signs, age, and weight (Fig. 6 A, B, C). Unexpectedly, the SOFA score was the most important predictor of mortality in S-AKI patients in all three groups; it is a simple but effective rating index

been widely used in clinical practice for sepsis patients. Despite this, the prior models that predicted mortality in S-AKI patients did not use this crucial factor [14, 35].

Force plots visualized individual model prediction as a result of feature contribution. By demonstrating how the XGBoost model generates predictions for four representative individuals, this model provides an intuitive way to guide clinicians’ and patients’ decision-making and improves their understanding of how the model makes a particular prediction.

Fourth, in the present study, mortality predictors of S-AKI patients were examined and were found to be consistent with previous findings. Urine output and AKI stage were closely related to renal injury severity [36]. Urine output plays an important role in predicting mortality in S-AKI patients. This result has been confirmed in many related studies. Laranja et al. observed that patients with sepsis-related AKI have lower urine output than those with AKI induced by other factors or with chronic kidney disease [37]. Our findings were

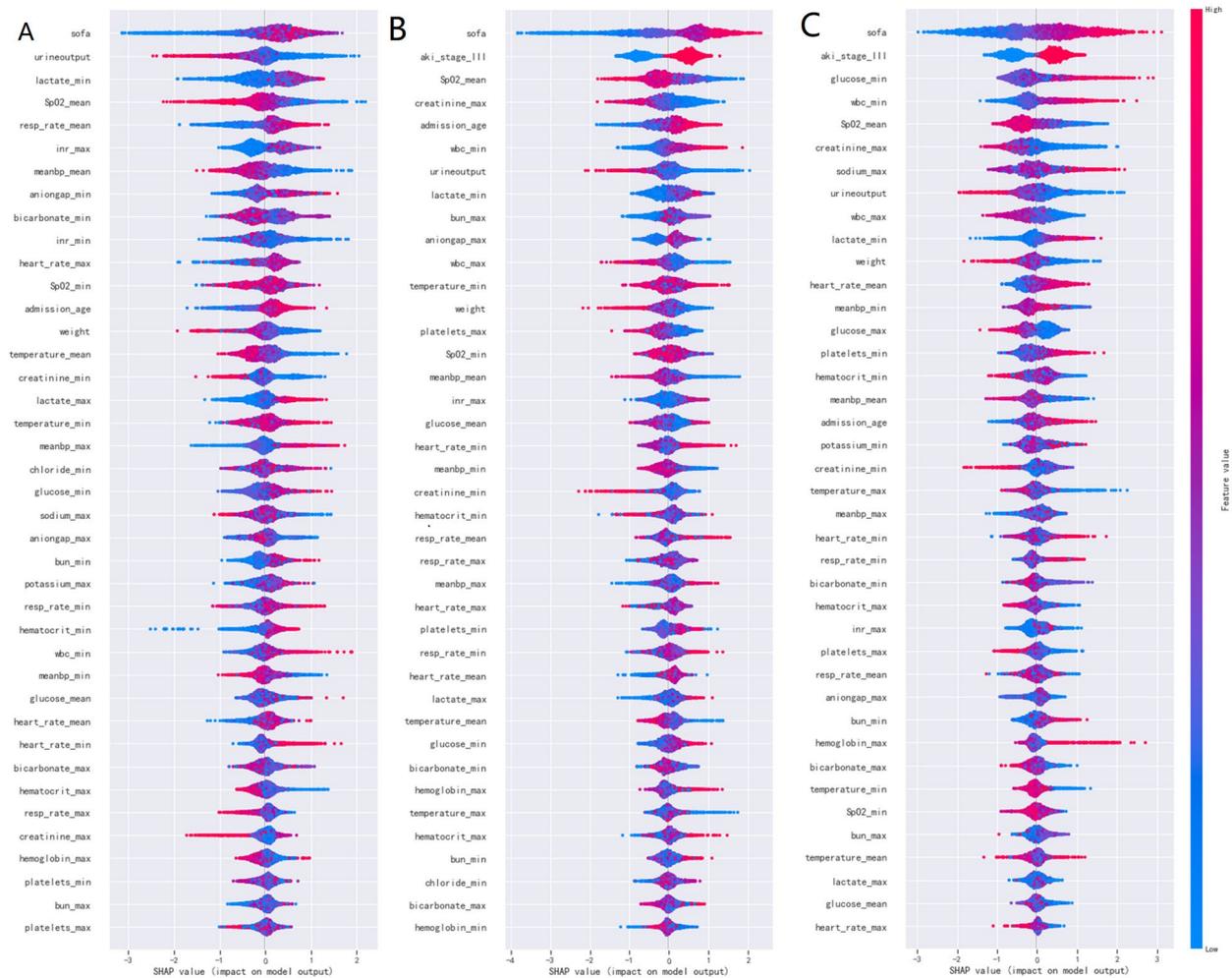


Fig 6 Feature importance analysis by SHAP method for XGBoost model SHAP summary plot of the 40 features of the XGBoost model. A dot is created for each feature attribution value for the model of each patient, and thus one patient is allocated one dot on the line for each feature. Dots are colored according to the values of features for the respective patient and accumulate vertically to depict density. Red represents a high feature value (in this case death), whereas blue represents a low feature value. The further away a point is from the baseline SHAP value of zero, the stronger it effects the output

consistent with those previous research findings that AKI staging positively correlated with higher mortality [38–41]. These results suggest that AKI occurrence and progress of AKI may lead to blood volume imbalance, fluid electrolyte disturbances, metabolite accumulation, and multiple-organ dysfunction aggravation, forming a vicious cycle in sepsis patients [42, 43].

Elevated lactate levels were closely correlated with poor prognosis in S-AKI patients. Hyperlactatemia was defined as a serum lactate level of > 2 mmol/L, whereas severe hyperlactatemia was defined as a serum lactate level of > 10 mmol/L [44]. In our study, patients in the non-survival group had higher lactate levels than those in the survival group. In other words, the mortality rate of S-AKI patients with severe hyperlactatemia was

considerably higher than that of patients with hyperlactatemia. In clinical practice, serum lactate level, as a sensitive indicator to diagnose hypoperfusion or hypoxia, has been shown to correlate with sepsis severity and prognosis [45–47]. Lactate and bicarbonate are both typical metabolic indicators. However, life-threatening S-AKI should not be dismissed in patients with normal lactate levels alone, and those with low bicarbonate levels, regardless of lactate levels, have high mortality rates and should also be considered for early, aggressive therapy [48].

Fifth, our models achieved promising predictive performance and demonstrated robustness and generalizability in the training and external validation cohorts. Predictors included in our models were collected from electronic

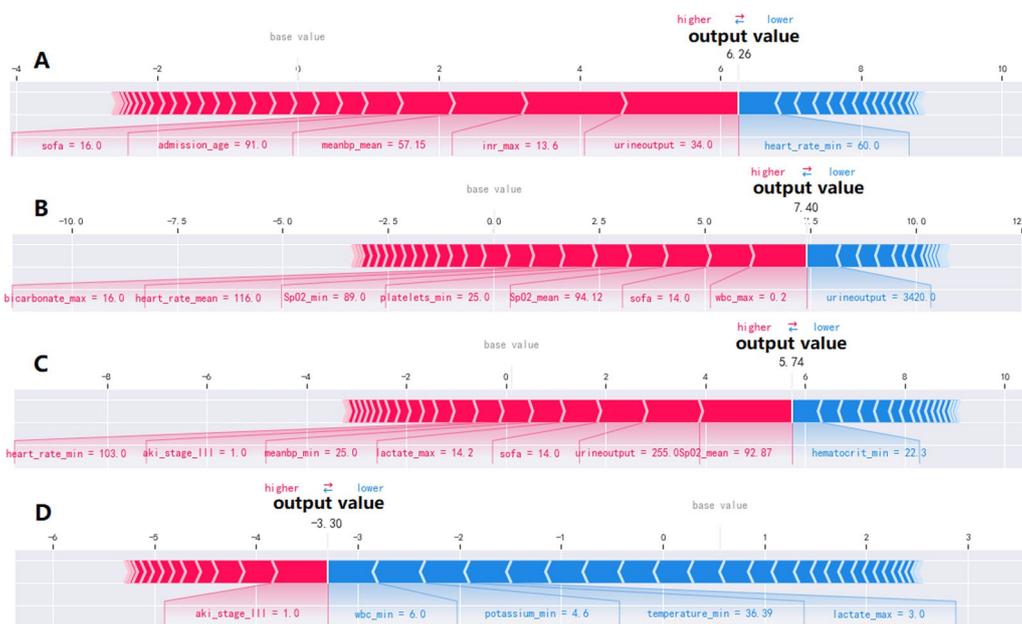


Fig 7 Force plots visualized individual model prediction as result of feature contributions. The base value represented the averaged predicted results. Feature values were listed at the plot bottom with feature names. Each group of features was ranked from center to both ends by the extent of their impact. **A:** A patient died within 7-day of ICU admission. “SOFA = 16” means that the score of SOFA of the patient was 16 when he admitted to ICU. “meanbp_mean = 57.15” represented that the maximum of mean arterial pressure of the patient was 57.15 mmHg in the first 24 h during ICU stay. “urine output = 34.0” means that the urine output of the patient was 34.0 ml in the first 24 h during ICU stay. **B:** A patient died within 14-day of ICU admission **C:** A patient died within 28-day of ICU admission. “aki_stage_III = 1.0” represented that the AKI stage of the patient was 3 when he admitted to ICU. **D:** A patient surviving after 28-day of ICU admission

medical records, and their values were seldom influenced by examiners. Only the most basic and commonly measured clinical data were used in our models, which can improve the generalizability of prediction models in ICUs. Our models were further validated in an external validation cohort that included 100 S-AKI patients from a critical care database. Training cohort data included were from Western countries, but our external validation cohort was from China, demonstrating that the model has applicability in different populations.

Predictors continuously change with changes in the associated pathological phenomenon as the disease

progresses. Unlike previous studies that used fixed predictors to predict in-hospital mortality in patients with severe S-AKI [49, 50], we selected different predictors for different survival time of 7, 14, and 28 days, which may represent different stages of the disease, with encouraging results. Rapid disease progression in critically ill patients may cause a delay in prediction using data from 24 h after ICU admission. In clinical practice, there is a growing need for better tools to assess progression and predict earlier which patients need treatment to halt disease progression. With further improvement in the MIMIC database and a better understanding of the

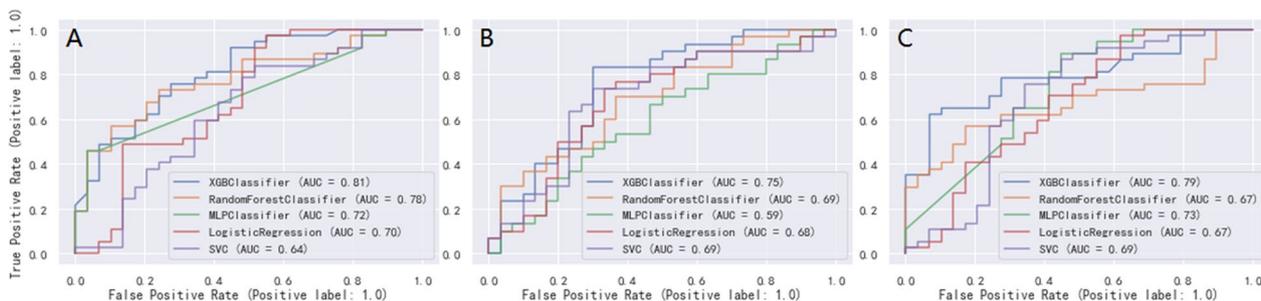


Fig 8 The ROC curves comparison of the five models with the three groups in external validation cohort. blue line = XGBoost model, organ line = random forest classifier model, green line = multilayer perceptron classifier model, red line = logistic regression model, purple line = support vector classifier model; **A:** 7-day group, **B:** 14-day group, **C:** 28-day group

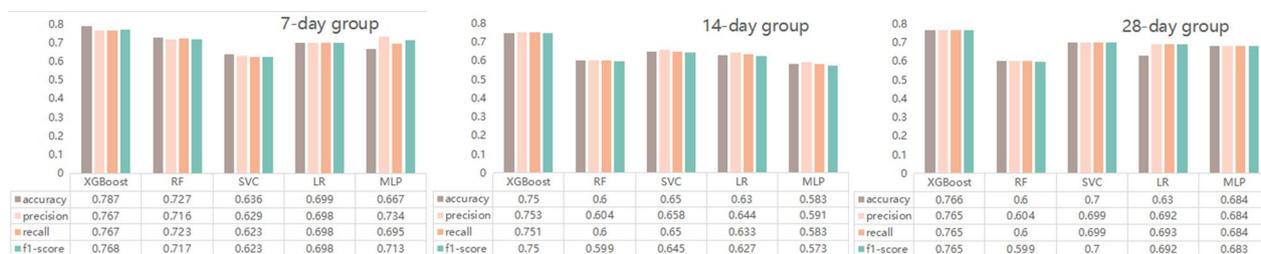


Fig 9 The comparison of performance in the five models with the three groups in external validation cohort

pathological mechanism of S-AKI and clinical interventions, combined with interpretable ML algorithms, a prediction system for the prognosis of patients with severe S-AKI at different stages may become a reality.

Our study was not without limitations. First, our training cohort was taken from the MIMIC-IV database, and the majority of the patients were from Western countries, which is quite different from our external validation cohort; second, we did not conduct a more comprehensive study of the database, which may have caused us to overlook some key variables, resulting in potential bias; and third, the retrospective and observational nature of this study may have led to selection bias. Nevertheless, our model still showed satisfactory performance for short-term mortality prediction in the external validation cohort.

Conclusions

In conclusion, ML methods are reliable tools for the prognosis prediction of patients with S-AKI. Global and local interpretability methods were combined to explain intrinsic information from the XGBoost model, which may prove clinically useful and help clinicians tailor precise management essential to maximize survival in patients with S-AKI.

Abbreviations

AKI	Acute kidney injury
S-AKI	Sepsis-associated acute kidney injury
MIMIC-IV	Medical information mart for intensive Care IV database
RFE	Recursive feature elimination
SMOTETomek	Synthesizing minority oversampling technology with Tomek link
RF	Random forest
XGBoost	Extreme gradient boosting
MLP	Multilayer perceptron classifier
SVC	Support vector classifier
LR	Logistic regression
ICU	Intensive care unit
ROC	Receiver operating characteristic
AUC	Areas under the ROC curves
RRT	Replacement therapy
IRB	The institutional review boards
MIT	The Massachusetts Institute of Technology

ICD-9	International classification of diseases and ninth revision
SOFA	Sequential organ failure assessment
SCr	Serum creatinine
KDIGO	Kidney disease improving global outcomes
Meanbp	Mean arterial pressure
SpO ₂	Oxyhemoglobin saturation
ABG	Arterial blood gas
INR	International normalized ratio
SHAP	SHapley additive exPlanations

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-023-04205-4>.

Additional file 1: Missing values information.

Additional file 2: Hyperparameters of ML models.

Additional file 3: The comparison of characteristics in training cohort and external validation cohort.

Acknowledgements

The authors thank all the researchers who created and managed the MIMIC-III database as well as the external validation cohort patients. Credit also goes to the National Institute of Biomedical Imaging and Bioengineering (NIBIB) alongside the MIT laboratory. The views expressed in this study are only those of the authors and have nothing to do with third parties.

Author contributions

ZF and JJ conceived the idea and designed the study; ZF, FC, and CX searched related literature; CX, YC, and QX contributed to data collection; ZF, JW, and MF reviewed medical records; ZF and JJ analyzed the data and built models; ZF and JJ wrote the main manuscript; FC and ZW prepared all figures; CX, MF, and YC assisted in data interpretation. All authors read and approved this manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The health information obtained from the MIMIC-IV database was unidentified, so informed consent of patients was not required. The data of external validation cohort were de-identified, and informed consent was not required. This study was reviewed and approved by the Ethics Committee of Hangzhou First People's Hospital Affiliated to Zhejiang University School of Medicine (KY2022124).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Emergency, Hangzhou First People's Hospital Affiliated to Zhejiang University School of Medicine, 310006 Hangzhou, Zhejiang, China.

²Department of Ultrasound, The First Affiliated Hospital Zhejiang University School of Medicine, 310003 Hangzhou, Zhejiang, China.

Received: 6 January 2023 Accepted: 15 May 2023

Published online: 22 June 2023

References

- Bouchard J, Acharya A, Cerda J, Maccariello ER, Madarasu RC, Tolwani AJ, Liang X, Fu P, Liu ZH, Mehta RL. A prospective international multicenter study of AKI in the intensive care unit. *Clin J Am Soc Nephrol*. 2015;10(8):1324–31.
- Cruz MG, Dantas JG, Levi TM, Rocha Mde S, de Souza SP, Boa-Sorte N, de Moura CG, Cruz CM. Septic versus non-septic acute kidney injury in critically ill patients: characteristics and clinical outcomes. *Rev Bras Ter Intensiva*. 2014;26(4):384–91. <https://doi.org/10.5935/0103-507X.2014059>.
- Hoste EA, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, Edipidis K, Forni LG, Gomersall CD, Govil D, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med*. 2015;41(8):1411–23. <https://doi.org/10.1007/s00134-015-3934-7>.
- Bagshaw SM, Uchino S, Bellomo R, Morimatsu H, Morgera S, Schetz M, Tan I, Bouman C, Macedo E, Gibney N, et al. Septic acute kidney injury in critically ill patients: clinical characteristics and outcomes. *Clin J Am Soc Nephrol*. 2007;2(3):431–9. <https://doi.org/10.2215/CJN.03681106>.
- Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Cooper-Smith CM, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. 2016;315(8):801–10. <https://doi.org/10.1001/jama.2016.0287>.
- Liu V, Escobar GJ, Greene JD, Soule J, Whippy A, Angus DC, Iwashyna TJ. Hospital deaths in patients with sepsis from 2 independent cohorts. *JAMA*. 2014;312(1):90–2. <https://doi.org/10.1001/jama.2014.5804>.
- Murugan R, Karajala-Subramanyam V, Lee M, Yende S, Kong L, Carter M, Angus DC, Kellum JA, Genetic IM, of Sepsis I. Acute kidney injury in non-severe pneumonia is associated with an increased immune response and lower survival. *Kidney Int*. 2010;77(6):527–35. <https://doi.org/10.1038/ki.2009.502>.
- Kellum JA, Chawla LS, Keener C, Singbartl K, Palevsky PM, Pike FL, Yealy DM, Huang DT, Angus DC, ProCess, et al. The Effects of Alternative Resuscitation Strategies on Acute Kidney Injury in Patients with Septic Shock. *Am J Respir Crit Care Med*. 2016;193(3):281–7. <https://doi.org/10.1164/rccm.201505-0995OC>.
- Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP. A clinical score to predict acute renal failure after cardiac surgery. *J A Soc Nephrol JASN*. 2005;16(1):162–8. <https://doi.org/10.1681/ASN.2004040331>.
- Chiofolo C, Chbat N, Ghosh E, Eshelman L, Kashani K. Automated continuous acute kidney injury prediction and surveillance: a random forest model. *Mayo Clin Proc*. 2019;94(5):783–92. <https://doi.org/10.1016/j.mayocp.2019.02.009>.
- RH Mehta JD Grab SM O'Brien CR Bridges JS Gammie CK Haan TB Ferguson ED Peterson Society of Thoracic Surgeons National Cardiac Surgery Database I. Bedside tool for predicting the risk of postoperative dialysis in patients undergoing cardiac surgery. *Circulation*. 2006;114(21):2208–16. <https://doi.org/10.1161/CIRCULATIONAHA.106.635573>.
- da HoraPassos R, Ramos JG, Mendonca EJ, Miranda EA, Dutra FR, Coelho MF, Pedroza AC, Correia LC, Batista PB, Macedo E, et al. A clinical score to predict mortality in septic acute kidney injury patients requiring continuous renal replacement therapy: the HELENICC score. *BMC Anesthesiol*. 2017;17(1):21. <https://doi.org/10.1186/s12871-017-0312-8>.
- Ohnuma T, Uchino S, Toki N, Takeda K, Namba Y, Katayama S, Kawarazaki H, Yasuda H, Izawa J, Uji M, et al. External validation for acute kidney injury severity scores: a multicenter retrospective study in 14 Japanese ICUs. *Am J Nephrol*. 2015;42(1):57–64. <https://doi.org/10.1159/000439118>.
- Hu H, Li L, Zhang Y, Sha T, Huang Q, Guo X, An S, Chen Z, Zeng Z. A prediction model for assessing prognosis in critically ill patients with sepsis-associated acute kidney injury. *Shock*. 2021;56(4):564–72. <https://doi.org/10.1097/SHK.0000000000001768>.
- Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, Chen KL, Yang CY, Lee OK. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020;24(1):478. <https://doi.org/10.1186/s13054-020-03179-9>.
- Dong J, Feng T, Thapa-Chhetry B, Cho BG, Shum T, Inwald DP, Newth CJL, Vaidya VU. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. *Crit Care*. 2021;25(1):288. <https://doi.org/10.1186/s13054-021-03724-0>.
- Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care*. 2019;23(1):12. <https://doi.org/10.1186/s13054-019-2411-z>.
- Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, Niu D, Wang Y, Tan W, Wu J. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J Transl Med*. 2022;20(1):215. <https://doi.org/10.1186/s12967-022-03364-0>.
- Calbitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. *JAMA*. 2017;318(6):517–8. <https://doi.org/10.1001/jama.2017.7797>.
- Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. 2017. <https://doi.org/10.48550/arXiv.1705.07874>
- Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23):E215–220. <https://doi.org/10.1161/01.cir.101.23.e215>.
- Lu X, Wang X, Gao Y, Yu S, Zhao L, Zhang Z, Zhu H, Li Y. Efficacy and safety of corticosteroids for septic shock in immunocompromised patients: a cohort study from MIMIC. *A J Emerg Med*. 2021;42:121–6. <https://doi.org/10.1016/j.ajem.2020.02.002>.
- Yang R, Huang T, Shen L, Feng A, Li L, Li S, Huang L, He N, Huang W, Liu H, et al. The use of antibiotics for ventilator-associated pneumonia in the MIMIC-IV database. *Front Pharmacol*. 2022;13:869499. <https://doi.org/10.3389/fphar.2022.869499>.
- Oweira H, Schmidt J, Mehrabi A, Kulaksiz H, Schneider P, Schob O, Giryas A, Abdel-Rahman O. Comparison of three prognostic models for predicting cancer-specific survival among patients with gastrointestinal stromal tumors. *Future Oncol*. 2018;14(4):379–89. <https://doi.org/10.2217/fon-2017-0450>.
- Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract*. 2012;120(4):c179–184. <https://doi.org/10.1159/000339789>.
- Yang J, Li Y, Liu Q, Li L, Feng A, Wang T, Zheng S, Xu A, Lyu J. Brief introduction of medical database and data mining technology in big data era. *J Evid Based Med*. 2020;13(1):57–69. <https://doi.org/10.1111/jebm.12373>.
- Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations. *J Stat Soft*. 2017. <https://doi.org/10.1198/jcgs.2011.10107>.
- Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explor Newsl*. 2004;6(1):20–9. <https://doi.org/10.1145/1007730.1007735>.
- Molnar C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). 2019. <https://christophm.github.io/interpretable-ml-book/>
- Fan ZY, Jiang JM. Prognostic Models in Critically Ill Patients with Sepsis-associated Acute Kidney Injury. 4, 2023, from <https://hanmuya-streamlit-pred-20230419streamlit40-model-tt9kpe.streamlit.app/>
- Deng X, Li M, Deng S, Wang L. Hybrid gene selection approach using XGBoost and multi-objective genetic algorithm for cancer classification.

- Med Biol Eng Comput. 2022;60(3):663–81. <https://doi.org/10.1007/s11517-021-02476-x>.
32. Hou N, Li M, He L, Xie B, Wang L, Zhang R, Yu Y, Sun X, Pan Z, Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med*. 2020;18(1):462. <https://doi.org/10.1186/s12967-020-02620-5>.
 33. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J*. 2021;21(10):1643–8. <https://doi.org/10.1016/j.spinee.2021.02.024>.
 34. Lankadeva YR, Okazaki N, Evans RG, Bellomo R, May CN. Renal medullary hypoxia: a new therapeutic target for septic acute kidney injury? *Semin Nephrol*. 2019;39(6):543–53. <https://doi.org/10.1016/j.semnephrol.2019.10.004>.
 35. Zhu Y, Zhang J, Wang G, Yao R, Ren C, Chen G, Jin X, Guo J, Liu S, Zheng H, et al. Machine learning prediction models for mechanically ventilated patients: analyses of the MIMIC-III database. *Front Med*. 2021;8:662340. <https://doi.org/10.3389/fmed.2021.662340>.
 36. Miller TR, Anderson RJ, Linas SL, Henrich WL, Berns AS, Gabow PA, Schrier RW. Urinary diagnostic indices in acute renal failure: a prospective study. *Ann Intern Med*. 1978;89(1):47–50. <https://doi.org/10.7326/0003-4819-89-1-47>.
 37. Pinheiro KHE, Azedo FA, Areco KCN, Laranja SMR. Risk factors and mortality in patients with sepsis, septic and non septic acute kidney injury in ICU. *J Bras Nefrol*. 2019;41(4):462–71. <https://doi.org/10.1590/2175-8239-JBN-2018-0240>.
 38. Jiang L, Zhu Y, Luo X, Wen Y, Du B, Wang M, Zhao Z, Yin Y, Zhu B, Xi X, et al. Epidemiology of acute kidney injury in intensive care units in Beijing: the multi-center BAKIT study. *BMC Nephrol*. 2019;20(1):468. <https://doi.org/10.1186/s12882-019-1660-z>.
 39. Charlton JR, Boohaker L, Askenazi D, Brophy PD, D'Angio C, Fuloria M, Gien J, Griffin R, Hingorani S, Ingraham S, et al. Incidence and risk factors of early onset neonatal AKI. *Clin J Am Soc Nephrol*. 2019;14(2):184–95. <https://doi.org/10.2215/CJN.03670318>.
 40. Cui X, Yu X, Wu X, Huang L, Tian Y, Huang X, Zhang Z, Cheng Z, Guo Q, Zhang Y, et al. Acute kidney injury in patients with the coronavirus disease 2019: a multicenter study. *Kidney Blood Press Res*. 2020;45(4):612–22. <https://doi.org/10.1159/000509517>.
 41. Murugan R, Kellum JA. Acute kidney injury: what's the prognosis? *Nat Rev Nephrol*. 2011;7(4):209–17. <https://doi.org/10.1038/nrneph.2011.13>.
 42. Poston JT, Koyner JL. Sepsis associated acute kidney injury. *BMJ*. 2019;364:k4891. <https://doi.org/10.1136/bmj.k4891>.
 43. Gaudry S, Hajage D, Benichou N, Chaibi K, Barbar S, Zarbock A, Lumlertgul N, Wald R, Bagshaw SM, Srisawat N, et al. Delayed versus early initiation of renal replacement therapy for severe acute kidney injury: a systematic review and individual patient data meta-analysis of randomised clinical trials. *Lancet*. 2020;395(10235):1506–15. [https://doi.org/10.1016/S0140-6736\(20\)30531-6](https://doi.org/10.1016/S0140-6736(20)30531-6).
 44. Haas SA, Lange T, Saugel B, Petzoldt M, Fuhrmann V, Metschke M, Kluge S. Severe hyperlactatemia, lactate clearance and mortality in unselected critically ill patients. *Intensive Care Med*. 2016;42(2):202–10. <https://doi.org/10.1007/s00134-015-4127-0>.
 45. Wan F, Du X, Liu H, He X, Zeng Y. Protective effect of anisodamine hydrobromide on lipopolysaccharide-induced acute kidney injury. 2020. *Biosci Rep*. <https://doi.org/10.1042/BSR20201812>.
 46. Wang L, Li Y, Wang X, Wang P, Essandoh K, Cui S, Huang W, Mu X, Liu Z, Wang Y, et al. GDF3 protects mice against sepsis-induced cardiac dysfunction and mortality by suppression of macrophage pro-inflammatory phenotype. *Cells*. 2020. <https://doi.org/10.3390/cells9010120>.
 47. Mikkelsen ME, Miliades AN, Gaieski DF, Goyal M, Fuchs BD, Shah CV, Bellamy SL, Christie JD. Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock. *Crit Care Med*. 2009;37(5):1670–7. <https://doi.org/10.1097/CCM.0b013e31819f9cf68>.
 48. Moroooka H, Kasugai D, Tanaka A, Ozaki M, Numaguchi A, Maruyama S. Prognostic impact of parameters of metabolic acidosis in critically ill children with acute kidney injury: a retrospective observational analysis using the PIC database. *Diagnostics*. 2020. <https://doi.org/10.3390/diagnos10110937>.
 49. Li X, Wu R, Zhao W, Shi R, Zhu Y, Wang Z, Pan H, Wang D. Machine learning algorithm to predict mortality in critically ill patients with sepsis-associated acute kidney injury. *Sci Rep*. 2023;13(1):5223. <https://doi.org/10.1038/s41598-023-32160-z>.
 50. Zhou H, Liu L, Zhao Q, Jin X, Peng Z, Wang W, Huang L, Xie Y, Xu H, Tao L, et al. Machine learning for the prediction of all-cause mortality in patients with sepsis-associated acute kidney injury during hospitalization. *Front Immunol*. 2023;14:1140755. <https://doi.org/10.3389/fimmu.2023.1140755>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

