# Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study

Junlong Hu,[a] Jing Xu,[a] Min Li,[b] Zhen Jiang,[c] Jie Mao,[a] Lian Feng,[a] Kexin Miao,[a] Huiwen Li,[a] Jiao Chen,[d] Zhenjiang Bai,[d] Xiaozhong Li,[a] Guoping Lu,[e] and Yanhong Li[a,f,*]

[a]Department of Nephrology and Immunology, Children's Hospital of Soochow University, Suzhou, Jiangsu province, China
[b]Pediatric Intensive Care Unit, Anhui Provincial Children's Hospital, Hefei, Anhui province, China
[c]Pediatric Intensive Care Unit, Xuzhou Children's Hospital, Xuzhou, Jiangsu province, China
[d]Pediatric Intensive Care Unit, Children's Hospital of Soochow University, Suzhou, Jiangsu province, China
[e]Pediatric Intensive Care Unit, Children's Hospital of Fudan University, Shanghai, China
[f]Institute of Pediatric Research, Children's Hospital of Soochow University, Suzhou, Jiangsu province, China

eClinicalMedicine
2024;68: 102409

Published Online xxx
https://doi.org/10.1016/j.eclinm.2023.102409

## Summary

**Background** Acute kidney injury (AKI) is a common and serious organ dysfunction in critically ill children. Early identification and prediction of AKI are of great significance. However, current AKI criteria are insufficiently sensitive and specific, and AKI heterogeneity limits the clinical value of AKI biomarkers. This study aimed to establish and validate an explainable prediction model based on the machine learning (ML) approach for AKI, and assess its prognostic implications in children admitted to the pediatric intensive care unit (PICU).

**Methods** This multicenter prospective study in China was conducted on critically ill children for the derivation and validation of the prediction model. The derivation cohort, consisting of 957 children admitted to four independent PICUs from September 2020 to January 2021, was separated for training and internal validation, and an external data set of 866 children admitted from February 2021 to February 2022 was employed for external validation. AKI was defined based on serum creatinine and urine output using the Kidney Disease: Improving Global Outcome (KDIGO) criteria. With 33 medical characteristics easily obtained or evaluated during the first 24 h after PICU admission, 11 ML algorithms were used to construct prediction models. Several evaluation indexes, including the area under the receiver-operating-characteristic curve (AUC), were used to compare the predictive performance. The SHapley Additive exPlanation method was used to rank the feature importance and explain the final model. A probability threshold for the final model was identified for AKI prediction and subgrouping. Clinical outcomes were evaluated in various subgroups determined by a combination of the final model and KDIGO criteria.

**Findings** The random forest (RF) model performed best in discriminative ability among the 11 ML models. After reducing features according to feature importance rank, an explainable final RF model was established with 8 features. The final model could accurately predict AKI in both internal (AUC = 0.929) and external (AUC = 0.910) validations, and has been translated into a convenient tool to facilitate its utility in clinical settings. Critically ill children with a probability exceeding or equal to the threshold in the final model had a higher risk of death and multiple organ dysfunctions, regardless of whether they met the KDIGO criteria for AKI.

**Interpretation** Our explainable ML model was not only successfully developed to accurately predict AKI but was also highly relevant to adverse outcomes in individual children at an early stage of PICU admission, and it mitigated the concern of the "black-box" issue with an undirect interpretation of the ML technique.

**Funding** The National Natural Science Foundation of China, Jiangsu Province Science and Technology Support Program, Key talent of women's and children's health of Jiangsu Province, and Postgraduate Research & Practice Innovation Program of Jiangsu Province.



*Corresponding author. Department of Nephrology and Immunology, Institute of Pediatric Research, Children's Hospital of Soochow University, Suzhou, Jiangsu province, China.
*E-mail address:* lyh072006@hotmail.com (Y. Li).

Keywords: Acute kidney injury; Adverse outcome; Critically ill children; Machine learning; Prediction model; SHAP

## Research in context

### Evidence before this study

Acute kidney injury (AKI) is a serious organ dysfunction in critically ill children. However, the currently used AKI criteria based on serum creatinine and urine output are not sensitive or specific enough. We searched PubMed for articles published up to August 31, 2023, using the keywords "(AKI OR acute kidney injury) and (prediction model)", with no language restriction. Several explainable prediction models for AKI were developed in adult intensive care unit (ICU) patients. However, no study has established an explainable prediction model for AKI occurring during the pediatric ICU (PICU) stay and investigated its prognostic implications in critically ill children.

### Added value of this study

In this study, we developed an explainable AKI prediction model using the random forest (RF) algorithm for PICU cohorts. This final explainable model performed well both in internal and external validations and had been translated into a convenient application to facilitate its utility for clinicians. The optimal probability cutoff value was employed as a threshold, and children with a probability exceeding or equal to the threshold had a higher risk of organ complications and mortality regardless of whether they met the diagnostic criteria for AKI.

### Implications of all the available evidence

This is the first prospective, multicenter study to investigate and compare 11 machine learning models for comprehensive AKI prediction analyses in PICU cohorts. Our final explainable RF model was established through a feature reduction, and it performed well in predicting AKI in any stage for critically ill children during their whole PICU stay. This model incorporated 8 variables that are easily obtained or evaluated in the first 24 h after PICU admission and reflect the children's conditions. The SHapley Additive explanation (SHAP) approach was used to explain this model via a global explanation that describes the overall functionality of a model and a local explanation that details how a certain prediction is made for an individual child by inputting the individualized data. Furthermore, children at high risk for poor outcomes were identified through a subgrouping based on the combination of the final model and the criteria of Kidney Disease: Improving Global Outcome (KDIGO). The above findings clarified the universal clinical application and advantages of our final model.

## Introduction

Acute kidney injury (AKI) is characterized by an abrupt decrease in kidney function and is a common complication in intensive care unit (ICU) scenarios.[1,2] Notably, AKI is associated with an increase in the length of ICU stays and ICU mortality.[1] Hence, early and accurate identification of patients at high risk for AKI is critical for initiating prompt therapeutic measures to potentially improve prognosis in clinical settings.

Currently, an increase in serum creatinine (SCr) and a drop in urine output (UO) are widely used in the diagnosis of AKI based on the criteria of Kidney Disease: Improving Global Outcome (KDIGO); however, they are not sensitive or specific enough for the kidney injury.[2,3] Instead of detecting AKI before it occurs, SCr only begins to rise after kidney injury has already set in, which may limit the timing of intervention for AKI. Although numerous studies focused on AKI have been published recently, the search for a single biomarker seems to have been unsuccessful due to the complex pathophysiology of AKI.[4] The heterogeneity of ICU populations, including those in pediatric ICU (PICU), may also limit the value of a single biomarker.

The machine learning (ML) approaches derived from electronic medical records (EMR) have gained the attention and recognition of clinicians in recent years.[5–7] The widespread utilization of EMR in hospitals has allowed for a more accurate and convenient collection of clinical data for patients. Many ML techniques are currently being employed in the development of AKI prediction models, and most have shown good predictive values, including those conducted on critically ill adult[7–13] and pediatric[14–16] patients. However, although the ML approach is powerful due to the complexity of the model, it is still limited by the difficulty of stating a direct interpretation, as a so-called "black-box".[17] Several studies have ranked the feature importance of ML models for AKI prediction in critically ill adult patients,[7–12] but only two of those explained the model at the individual adult level.[10,11] The Acute Dialysis Quality Initiative (ADQI) group also advised that the prediction model should present information about clinical variables influencing the risk of AKI.[18] To overcome these issues of the "black-box", the SHapley Additive exPlanation (SHAP) method was utilized to explain the ML models and visualize individual variable predictions.[19] The SHAP method is a unified approach for explaining the outputs of ML models in earlier studies, including in AKI studies.[20,21] Regardless, scarce research has focused on AKI in critically ill pediatric patients, with the SHAP method being used to explain prediction models.

In this study, we aimed to develop and validate explainable ML models for early and accurate prediction

of AKI in critically ill children in a multicenter prospective cohort, elucidate feature importance and explain the model via the SHAP method, and determine the prognostic implications of the final model in the subgrouping of critically ill children.

## Methods

### Study population

The prospective multicenter cohort study in China was conducted in critically ill children for the derivation and validation of the prediction model. The derivation cohort consisted of critically ill children admitted to the PICUs of four independent tertiary hospitals (Children's Hospital of Soochow University, Children's Hospital of Fudan University, Anhui Provincial Children's Hospital, and Xuzhou Children's Hospital) from September 2020 to January 2021. The criteria for PICU admission were strictly adopted in accordance with the guidelines for developing admission and discharge policies for the pediatric intensive care unit.[22] Children who were aged 1 month to 18 years and met the criteria for PICU admission were considered for inclusion in all four PICUs. Children with chronic kidney disease (CKD) or those with serious clinical and laboratory data missing were excluded. This study was performed in accordance with the Declaration of Helsinki, with approvals of the Institutional Review Board of Children's Hospital of Soochow University (2020KS009), Children's Hospital of Fudan University [(2020) 404], Anhui Provincial Children's Hospital (EYLL-2020-023), and Xuzhou Children's Hospital (2020-1-3). Written informed consent was obtained from each participating individual's guardian.

### Data collection and processing

We utilized demographic characteristics, vital sign measurements, and laboratory data collected within the first 24 h after PICU admission to identify features and construct prediction models, as shown in Supplementary Table S1. All data were obtained from the EMR system. Given that the UO recorded by the nursing record system on the first day after PICU admission was from the time of admission until 7 a.m. the following morning, we chose the UO recorded during this period to construct prediction models. The severity of the illness was assessed using the Pediatric Risk of Mortality III (PRISM III) score, which was calculated based on physiological parameters recorded on the first 24 h of PICU admission.[23]

Since multi-collinearity among features may affect prediction accuracy, one feature that was less correlated with outcome was eliminated from the data set when two features were highly correlated (correlation coefficient > 0.6) in Spearman's correlation analyses, as shown in Supplementary Fig. S1. Although age and height are also important features for children, body weight was finally selected for model development; a detailed explanation is displayed in Supplementary Appendix S1. In addition, features with over 25% missing values were excluded in the following analyses to minimize the bias resulting from missing data. Finally, 33 features, including body weight, body mass index (BMI), sex, PRISM III score, minimum and maximum mean arterial pressure (MAP) and temperature (Temp), minimum heart rate (HR), estimated glomerular filtration rate (eGFR), UO, the initial value of pH, bicarbonate ($HCO_3^-$), activated partial thromboplastin time (APTT), international normalized ratio (INR), creatine kinase-myocardial band (CK-MB), total bilirubin (TBil), alanine aminotransferase (ALT), lactate dehydrogenase (LDH), blood urea nitrogen (BUN), albumin, glucose, potassium (K), sodium (Na), chlorine (Cl), calcium (Ca), white blood cell (WBC), C-reactive protein (CRP) and procalcitonin (PCT), the minimum value of partial pressure of oxygen/fraction of inspiration oxygen ($PaO_2/FiO_2$) and platelet (PLT), and the maximum value of partial pressure of carbon dioxide ($PaCO_2$) and lactate (Lac) during the first 24 h following PICU admission, were utilized to develop the prediction models.

In addition, the renal angina index (RAI), as a predictive indicator for AKI,[24,25] was determined using the data obtained during the first 24 h of PICU admission, and the predictive value of the RAI was subsequently compared to that of the prediction model.

### External validation

An external data set consisting of critically ill children admitted to the PICU of Children's Hospital of Soochow University from February 2021 to February 2022 was employed for the external validation. The inclusion and exclusion criteria were identical to those of the derivation cohort.

### Definition of AKI

The diagnosis and stage of AKI developed during the PICU stay were defined based on SCr and UO in accordance with the KDIGO criteria.[2] The baseline SCr was defined as the lowest SCr level within 3 months prior to PICU admission.[1] For children without a known baseline SCr level, we assumed a normal eGFR of 120 mL/min/1.73 $m^2$, which was selected to be consistent with previous studies of pediatric AKI,[1,26,27] and back-calculated an expected baseline SCr by the modified Schwartz estimating equation [SCr (mg/dl) = 0.413 × height (cm)/eGFR].[26–28] AKI stage 1 was defined as mild AKI, and AKI stages 2 and 3 were defined as severe AKI. Persistent AKI was defined as AKI that lasted beyond 48 h of onset.[29]

### Model development and comparison

The data from the derivation cohort comprising four separate tertiary hospitals were divided, with 70%

utilized for training and 30% for validation (internal validation), in order to avoid problems with overfitting. In addition, an external data set was used for testing (external validation).

The total of 33 features mentioned above were used to develop the prediction models. Missing data were handled by the median imputation method,[30,31] and the proportions of missing data per variable are shown in Supplementary Table S1. Eleven ML models, namely, adaptive boosting (AdaBoost), artificial neutral network (ANN), decision tree (DT), extra tree (ET), gradient boosting machine (GBM), K-nearest neighbor (KNN), light gradient boosting machine (LightGBM), logistic regression (LR), random forest (RF), support vector machine (SVM), and eXtreme gradient boosting (XGboost) were used to predict AKI in critically ill children. In order to optimize the prediction model, grid search combined with manual fine tuning was applied to obtain the final hyperparameters.

Several commonly used evaluation indexes, such as the area under the receiver-operating-characteristic (ROC) curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 score, were used to evaluated the reliability of these models. In addition, five-fold and ten-fold cross validations were conducted in the derivation cohort for the validation of the prediction model.

### Feature selection and model explanation

It is challenging to get the correct interpretation of a ML model. The SHAP method is an approach that could rank the importance of input features and explain the results of the prediction model, and it is implemented to overcome the "black-box" issue.[19]

SHAP value-assisted feature selection was used to restrict the prediction model from 33 to 3 features in accordance with feature importance rank; thereby, the final model with the best predictive ability in the process of reducing features was chosen for further analysis. The nonparametric method of Delong[32] was used to compare the difference between AUCs using MedCalc Version 19.6 (https://www.medcalc.org), and the features of the chosen ML model were gradually reduced until the AUC was dramatically decreased.

The SHAP method offered global and local explanations for the model explanation. The global explanation could give consistent and accurate attribution values for each feature within a model to show the associations between input features and AKI. The local explanation could demonstrate a specific prediction for an individual child by inputting the specific data.

### Webpage deployment tool based on streamlit framework

To facilitate the utility of the model in clinical settings, the final prediction model was implemented into a web application established based on the Streamlit Python-based framework. When the values of corresponding features from the final model are provided, the application can return the probability of AKI and the force plot for the individual child.

### Statistical analysis

Data analyses were conducted using Python version 3.6.5 (https://www.python.org) and SPSS Statistical Software Version 23.0 (https://www.ibm.com/spss). Continuous variables with skewed distributions were presented as median with interquartile range and compared using the Mann–Whitney U test or Kruskal–Wallis H test. Categorical variables were presented as numbers with percentages and compared using the Chi-square test or Fisher's exact test. The analysis of covariance (ANCOVA) was used to adjust for the confounder. The AUCs were used to evaluate the predictive power, and the optimal cutoff value was established by maximizing the Youden index (sensitivity + specificity-1). Decision curve analysis (DCA) and precision-recall (P–R) curve analysis were conducted by R version 4.1.0 (https://www.r-project.org). A two-tailed P value $< 0.05$ was considered statistically significant.

### Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## Results

### Patient characteristics

This prospective study involved 957 critically ill children in the derivation cohort for the identification of the prediction model. Of the 964 children admitted to the four PICUs during the study period, 4 with CKD and 3 with serious data missing were excluded. These 957 children were allocated into separate training and internal validation sets. In addition, among 872 children admitted to the PICU in the external validation cohort, 3 with CKD and 3 with serious data missing were excluded. The external cohort consisting of 866 children was used as the external validation set. The comparison of demographic and clinical variables among the training, internal validation, and external validation sets is shown in Supplementary Table S2. Details of the study design are displayed in Fig. 1.

Among the 957 children in the derivation cohort, 284 (29.7%) developed AKI during the PICU stay, including 108 with AKI stage 1, 82 with AKI stage 2, and 94 with AKI stage 3. Of the 284 AKI children, 201 (70.8%) developed AKI on the first day, 271 (95.4%) within the first week, and 278 (97.9%) within the first two weeks. Moreover, among AKI children, 226 (79.6%) met the criteria of SCr alone, 41 (14.4%) met the criteria of UO

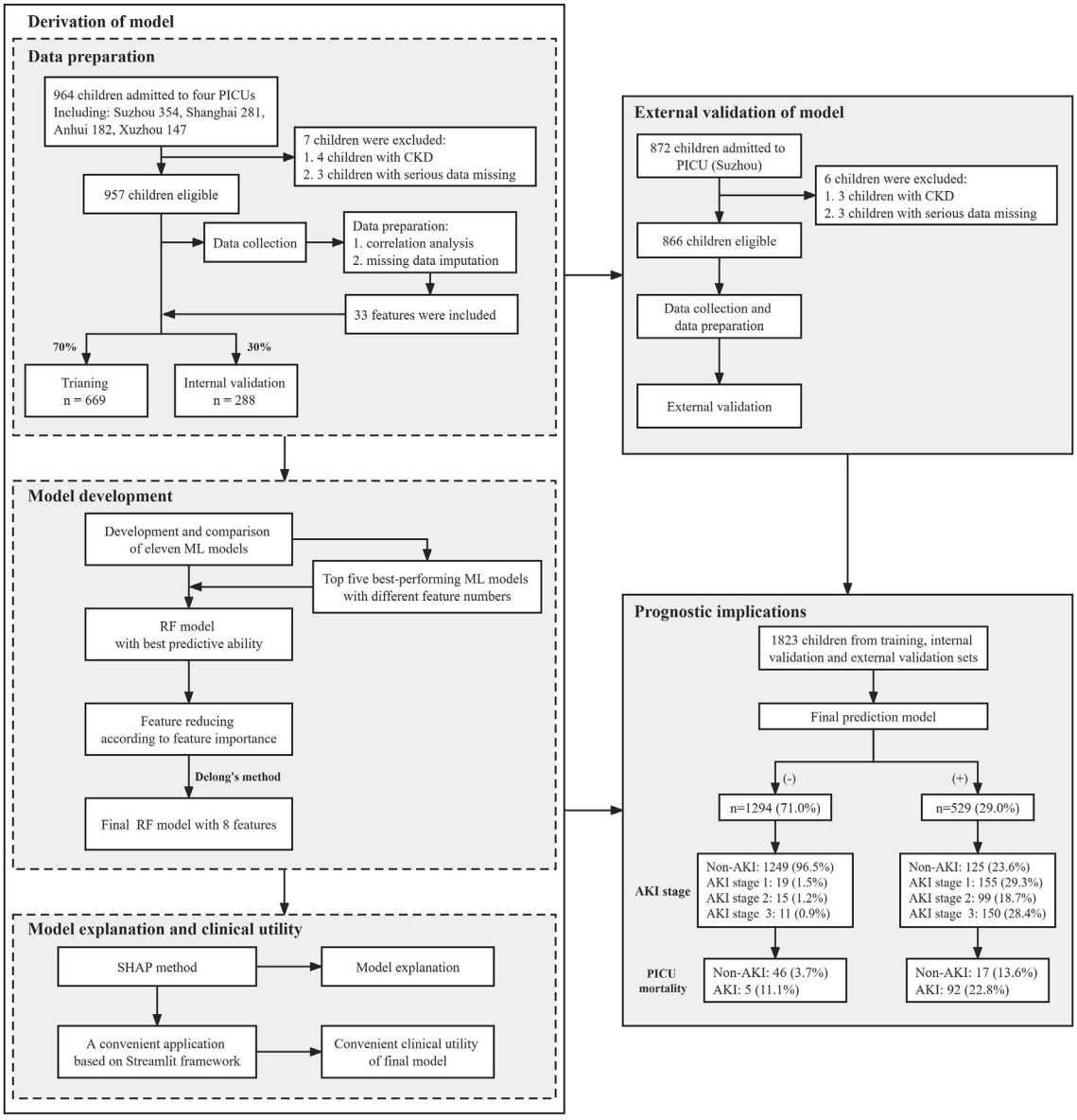


***Fig. 1:* Flow chart of the study design.** AKI: acute kidney injury; CKD: chronic kidney disease; ML: machine learning; PICU: pediatric intensive care unit; RF: random forest; SHAP: SHapley Additive explanation.

alone, and 17 (6.0%) met the criteria of SCr and UO simultaneously. The demographic and clinical characteristics between non-AKI and AKI are listed in Table 1. The major reasons of the 957 children for PICU admission were respiratory disease (24.8%), followed by neurological disease (21.9%) and hematologic and oncologic disease (15.2%).

### Model development and performance comparison

The data collected in the first 24 h after PICU admission were used to generate 11 ML models to predict AKI developed during the PICU stay in critically ill children. Among the 11 models, RF model (AUC = 0.940) had the best predictive effect for AKI, followed by LightGBM model (AUC = 0.936) and GBM model (AUC = 0.922). The discriminative performances of these 11 models are listed in Supplementary Table S3, and the ROC curves and the SHAP summary plots of the top 20 features for the top five best-performing ML models are presented in Fig. 2A and Supplementary Fig. S2A–E, respectively. During the process of reducing features based on the feature important rank, the changes in AUCs for these five models showed that the RF model maintained nearly the best predictive ability among these five

| Variables | Non-AKI (n = 673) | AKI (n = 284) | P value |
|---|---|---|---|
| Age, month | 34.0 [12.0–84.0] | 22.0 [4.0–106.8] | 0.041 |
| Body weight, kg | 14.0 [9.2–24.0] | 11.3 [6.5–25.0] | 0.006 |
| BMI, kg/m$^2$ | 16.3 [14.9–18.8] | 16.5 [14.5–19.6] | 0.742 |
| Male, n (%) | 408 (60.6) | 164 (57.7) | 0.407 |
| PRISM III score | 3.0 [2.0–7.0] | 8.0 [3.0–13.0] | <0.001 |
| MAP_min, mmHg | 68.3 [61.7–76.7] | 64.3 [54.0–74.6] | <0.001 |
| MAP_max, mmHg | 82.7 [73.0–92.7] | 81.3 [67.4–90.0] | 0.006 |
| HR_min, beats/min | 109.0 [90.0–126.0] | 116.0 [92.0–133.0] | 0.007 |
| Temp_min, °C | 36.6 [36.4–36.9] | 36.6 [36.3–37.0] | 0.932 |
| Temp_max, °C | 37.5 [37.0–38.5] | 37.7 [37.0–38.7] | 0.030 |
| $PaCO_2$_max, mmHg | 36.5 [30.6–41.3] | 37.4 [29.9–45.9] | 0.020 |
| $PaO_2/FIO_2$, mmHg | 452.4 [278.8–624.0] | 342.0 [192.9–568.0] | <0.001 |
| pH | 7.4 [7.4–7.5] | 7.4 [7.3–7.5] | 0.169 |
| $HCO_3^-$, mmol/L | 21.6 [18.9–24.2] | 21.7 [17.4–24.7] | 0.580 |
| APTT, seconds | 31.6 [28.0–36.6] | 34.7 [28.6–42.3] | <0.001 |
| INR | 1.1 [1.0–1.2] | 1.1 [1.0–1.4] | <0.001 |
| CK-MB, ng/mL | 17.0 [3.8–29.0] | 17.0 [4.1–36.8] | 0.090 |
| TBil, μmol/L | 7.0 [4.8–10.9] | 10.4 [5.7–19.4] | <0.001 |
| ALT, U/L | 18.0 [12.0–31.0] | 24.9 [15.0–70.9] | <0.001 |
| LDH, U/L | 326.8 [255.1–422.3] | 359.0 [281.6–693.8] | <0.001 |
| Lac_max, mmol/L | 1.8 [1.2–2.7] | 2.3 [1.4–3.6] | <0.001 |
| SCr, μmol/L | 25.0 [19.0–33.7] | 35.0 [24.3–55.3] | <0.001 |
| BUN, mmol/L | 3.8 [2.8–4.9] | 5.0 [3.3–7.9] | <0.001 |
| Albumin, g/L | 40.9 [36.8–44.0] | 38.7 [34.0–43.1] | <0.001 |
| Glucose, mmol/L | 5.9 [5.2–7.3] | 6.1 [5.2–7.6] | 0.300 |
| K, mmol/L | 3.8 [3.5–4.2] | 3.9 [3.5–4.4] | 0.142 |
| Na, mmol/L | 137.0 [135.0–139.0] | 137.0 [133.0–140.0] | 0.499 |
| Cl, mmol/L | 105.1 [103.0–108.0] | 105.0 [102.0–108.2] | 0.621 |
| Ca, mmol/L | 1.2 [1.1–1.2] | 1.1 [1.0–1.2] | <0.001 |
| WBC, 10$^9$/L | 10.5 [6.9–14.6] | 10.2 [6.5–14.9] | 0.721 |
| PLT_min, 10$^9$/L | 280.0 [187.5–367.0] | 240.0 [114.5–341.0] | <0.001 |
| CRP, mg/L | 6.0 [0.6–13.0] | 8.0 [1.3–50.8] | <0.001 |
| PCT, ng/mL | 0.2 [0.1–0.4] | 0.3 [0.2–2.0] | <0.001 |
| Urine output, mL/kg/h | 2.1 [1.3–3.1] | 2.1 [1.0–3.6] | 0.697 |
| eGFR[a], mL/min/1.73 m$^2$ | 130.9 [110.3–159.8] | 83.0 [63.6–127.2] | <0.001 |
| MODS, n (%) | 76 (11.3) | 124 (43.7) | <0.001 |
| Shock/DIC, n (%) | 52 (7.7) | 76 (26.8) | <0.001 |
| MV, n (%) | 170 (25.3) | 127 (44.7) | <0.001 |
| RRT, n (%) | 2 (0.3) | 61 (21.5) | <0.001 |
| PICU length of stay, hours | 102.1 [48.7–195.2] | 144.2 [70.6–288.9] | <0.001 |
| PICU mortality, n (%) | 30 (4.5) | 49 (17.3) | <0.001 |

Continuous values were presented as median [interquartile range]. Categorical values were presented as number (percentage). Max and min represented the maximum and minimum values during the first 24 h after PICU admission, respectively. AKI: acute kidney injury; BMI: body mass index; DIC: disseminated intravascular coagulation; MODS: multiple organ dysfunction syndrome; MV: mechanical ventilation; PICU: pediatric intensive care unit; PRISM III: Pediatric Risk of Mortality III, RRT: renal replacement therapy. [a]The eGFR was calculated based on the first available serum creatinine during the first 24 h after PICU admission.

***Table 1:*** **Comparison of demographic and clinical characteristics and outcomes between non-AKI and AKI in the derivation cohort.**

models, as shown in Fig. 2B. Thus, it can be observed that, of the five models mentioned above, the RF model fared best in terms of AKI prediction. The performance of the RF model with varied numbers of features is displayed in Fig. 2C and Supplementary Table S4. The sensitivity, specificity, PPV, NPV, accuracy, and F1 score were calculated at the optimal cutoff value that maximized the Youden index.

### Identification of the final model

The final model was identified during the feature reduction of the RF model. As displayed in Fig. 2C and Supplementary Fig. S3A, the 33-feature model was significantly better than 3-feature model ($\triangle$AUC = 0.036, P = 0.004) and 7-feature model ($\triangle$AUC = 0.020, P = 0.042), respectively, but not significantly better than the 8-feature model ($\triangle$AUC = 0.011, P = 0.161) in predicting AKI developed during the PICU stay. The 8-feature model had a good net benefit and a high threshold probability, comparable to the 33-feature model. Meanwhile, the area under the P–R curve of the 8-feature model was only marginally lower than that of the 33-feature model, indicating that both models have similar and high clinical utility, as shown in Supplementary Fig. S3B–G. Hence, we focused on the 8-feature RF model, including body weight, PRISM III score, MAP_min, APTT, TBil, eGFR, BUN, and UO, as the final model for further analysis. The final RF model achieved an AUC of 0.929 with a sensitivity of 0.886, a specificity of 0.866, a PPV of 0.714, a NPV of 0.953, an accuracy of 0.872, and an F1 score of 0.791 for predicting AKI in critically ill children. Since a large proportion of AKI cases developed on the first day after PICU admission, the predictive performance of the final model for AKI that occurred at different timing points was further explored in Supplementary Table S5. The final model achieved AUCs of 0.977 and 0.927 for AKI on day 1 and days 2–7 of the PICU stay, respectively.

To validate the appropriate sample size for this study and the robustness of this model to site variation, cross validations were further performed. As shown in Supplementary Fig. S4A and B, the final model achieved mean AUCs of 0.909 ± 0.031 and 0.912 ± 0.042 in the five-fold and ten-fold cross validations, respectively. In addition, the final model displayed mean AUCs of 0.929 ± 0.061 in Suzhou cohort (n = 350), 0.905 ± 0.093 in Shanghai cohort (n = 278), 0.897 ± 0.062 in Anhui cohort (n = 182), and 0.874 ± 0.104 in Xuzhou cohort (n = 147) in the ten-fold cross validation.

The predictive values of SCr, BUN, UO, and eGFR, which all reflect renal function, as well as RAI and PRISM III, were further investigated and compared with the 8-feature final model. As illustrated in Supplementary Fig. S5A, SCr ($\Delta$AUC = 0.250, P < 0.001), BUN ($\Delta$AUC = 0.225, P < 0.001), UO ($\Delta$AUC = 0.474, P < 0.001), eGFR ($\Delta$AUC = 0.143, P < 0.001), RAI ($\Delta$AUC = 0.162, P < 0.001), and the PRISM III ($\Delta$AUC = 0.231, P < 0.001) performed worse in the internal validation than the final model, respectively. The DCA curves also revealed that the final model had greater clinical utility than SCr, BUN, UO, eGFR,

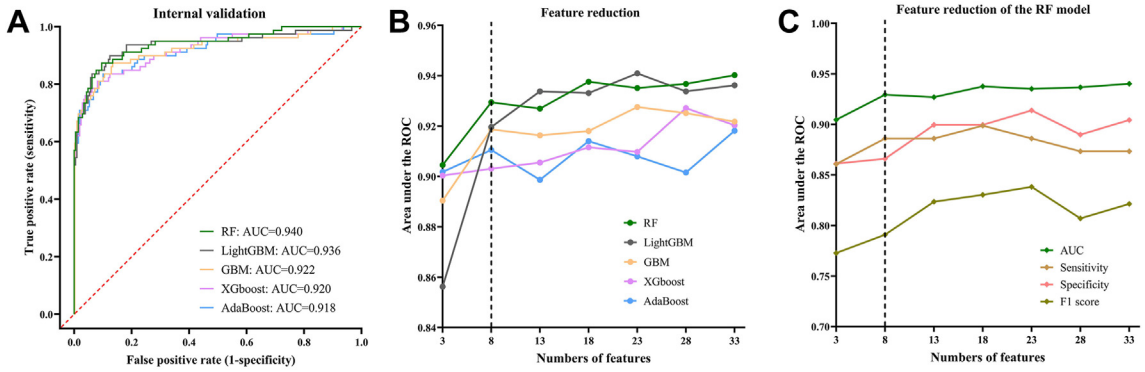


***Fig. 2:*** **Performance of ML models to predict AKI. (A)** ROC curves of the top five best-performing ML models. **(B)** AUCs of the top five best-performing ML models with varied numbers of features. **(C)** AUC, sensitivity, specificity, and F1 score of the RF model with varied numbers of features. AdaBoost: adaptive boosting; AKI: acute kidney injury; AUC: area under the ROC curve; GBM: gradient boosting machine; LightGBM: light gradient boosting machine; ML: machine learning; RF: random forest; ROC: receiver-operating-characteristic; XGboost: eXtreme gradient boosting.

RAI, and the PRISM III score, respectively, as presented in Supplementary Fig. S5B.

### External validation of the final model

For the external validation, the final model gave an AUC of 0.910, which was similar to that in the internal validation (ΔAUC = 0.019, P = 0.477), indicating that the final model showed great performance both in internal and external validations.

Moreover, as the MAP values in the final model were not the first measured values obtained after PICU admission, we selected the initial values of MAP and replaced them in the external validation data set to investigate whether AKI can be accurately predicted. In the secondary external validation, the final model displayed an AUC of 0.910, showing an excellent ability to identify critically ill children at high risk for AKI in the external validation cohort.

### Model explanation

Since it is difficult for clinicians to accept a prediction model that is not directly explainable and interpretable, the SHAP method is utilized to interpret the output of the final model by calculating the contribution of each variable to the prediction. This explainable method provided two types of explanations: global explanation of the model at the feature level and local explanation at the individual level. Global explanation described the overall functionality of the model. As shown in SHAP summary plots (Fig. 3A and B), the contributions of the feature to the model were evaluated using the average SHAP values and exhibited in descending order. Additionally, the SHAP dependence plot can facilitate understanding how a single feature affects the output of the prediction model. The real values versus the SHAP values of these 8 features are shown in Fig. 3C, and SHAP values that are higher than zero correspond to a positive class prediction in the model, in other words, a higher risk of AKI. For instance, children with an eGFR ≤86.3 mL/min/1.73 $m^2$ or a PRISM III score ≥9 scores had SHAP values higher than zero, which pushed the decision towards the "AKI" class. In addition, a low actual value ≤57.7 mmHg or a high actual value ≥96.3 mmHg of MAP_min pushed the decision towards the "AKI" class, as well as a low actual value ≤7.5 kg or a high actual value ≥54 kg of body weight.

In addition, local explanation analyzed how a certain prediction was made for a specific individual by incorporating the individualized input data. Fig. 4A–C showed a child who did not develop AKI during the PICU stay. Fig. 4A represented this child towards the "AKI" class with a probability of 5.6%, and Fig. 4B represented this child towards the "non-AKI" class with a probability of 94.4%, according to the prediction model. The actual measured values of features were also displayed in the waterfall plot, as shown in Fig. 4A and B. As observed, the values of eGFR, PRISM III score, BUN, UO, MAP_min, APTT, and body weight pushed the decision towards the "non-AKI" class, but TBil did not. If the actual values for most features were normal, such as eGFR, the risk of developing AKI would be low. In contrast, the TBil, with an actual value outside the normal range, may increase the risk for AKI in this child, even though the overall prediction pushed this case into the "non-AKI" class.

A similar phenomenon was observed for a child developing AKI during the PICU stay in Fig. 4D–F. The features pushing or pulling the decision toward the "AKI" class and their actual measured values were displayed in Fig. 4D and E. The decision for this case leaned towards "AKI" with a probability of 92.7% and "non-AKI" with a probability of 7.3%. Furthermore, a force plot of interpretation for children in the internal

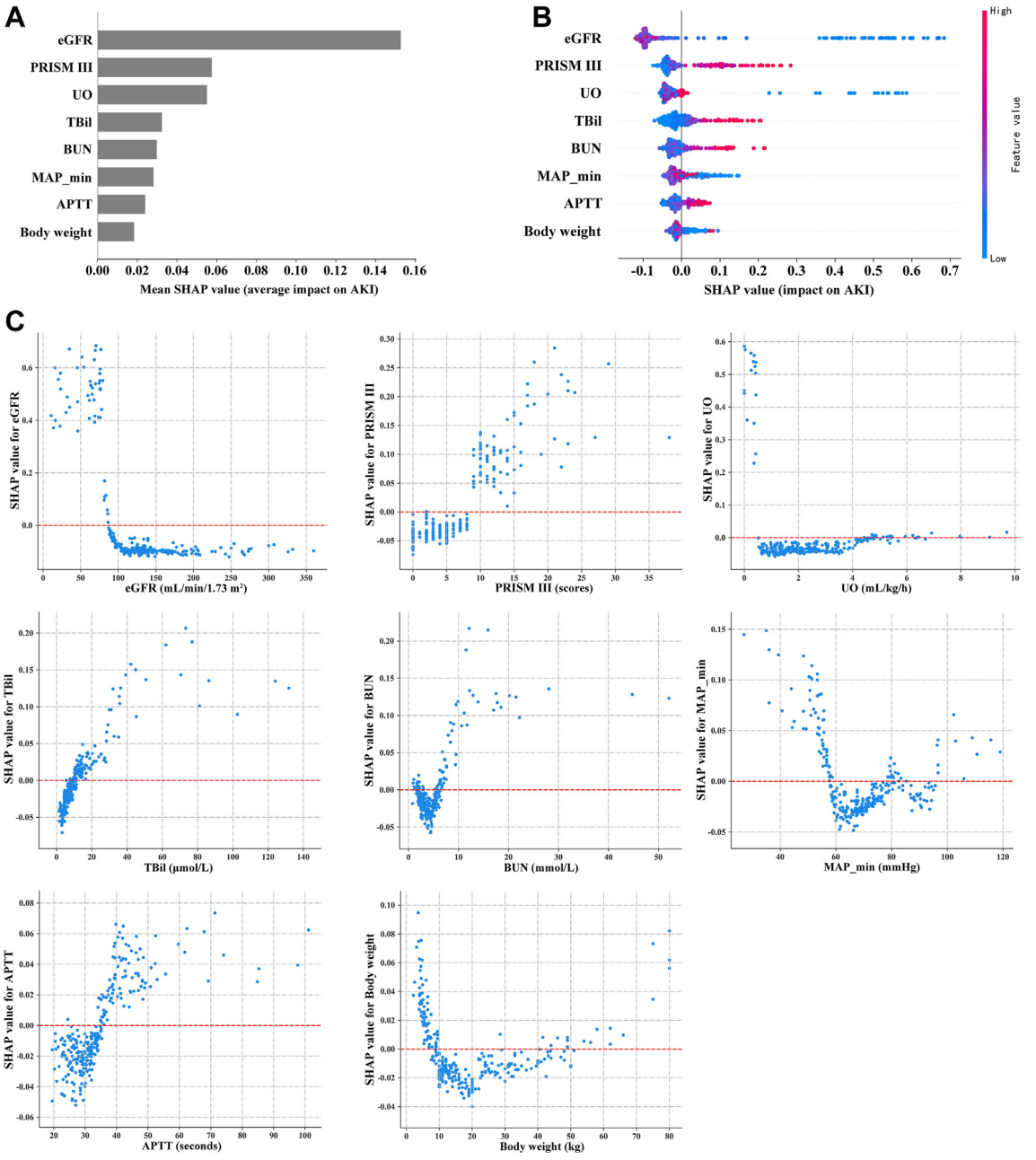


*Fig. 3:* **Global model explanation by the SHAP method. (A)** SHAP summary bar plot. **(B)** SHAP summary dot plot. The probability of AKI development increases with the SHAP value of a feature. A dot is made for SHAP value in the model for each single patient, so each patient has one dot on the line for each feature. The colors of the dots demonstrate the actual values of the features for each patient, as red means a higher feature value and blue means a lower feature value. The dots are stacked vertically to show density. **(C)** SHAP dependence plot. Each dependence plot shows how a single feature affects the output of the prediction model, and each dot represents a single patient. For example, both a low actual value ≤57.7 mmHg or a high actual value ≥96.3 mmHg of MAP_min push the decision towards the "AKI" class. SHAP values are represented by the y-axis, and actual values are represented by the x-axis. The SHAP values for specific features exceeding zero push the decision towards the "AKI" class. AKI: acute kidney injury; APTT: activated partial thromboplastin time; BUN: blood urea nitrogen; eGFR: estimated glomerular filtration rate; MAP: mean arterial pressure; PRISM III: pediatric risk of mortality III; SHAP: SHapley Additive explanation; TBil: total bilirubin; UO: urine output.

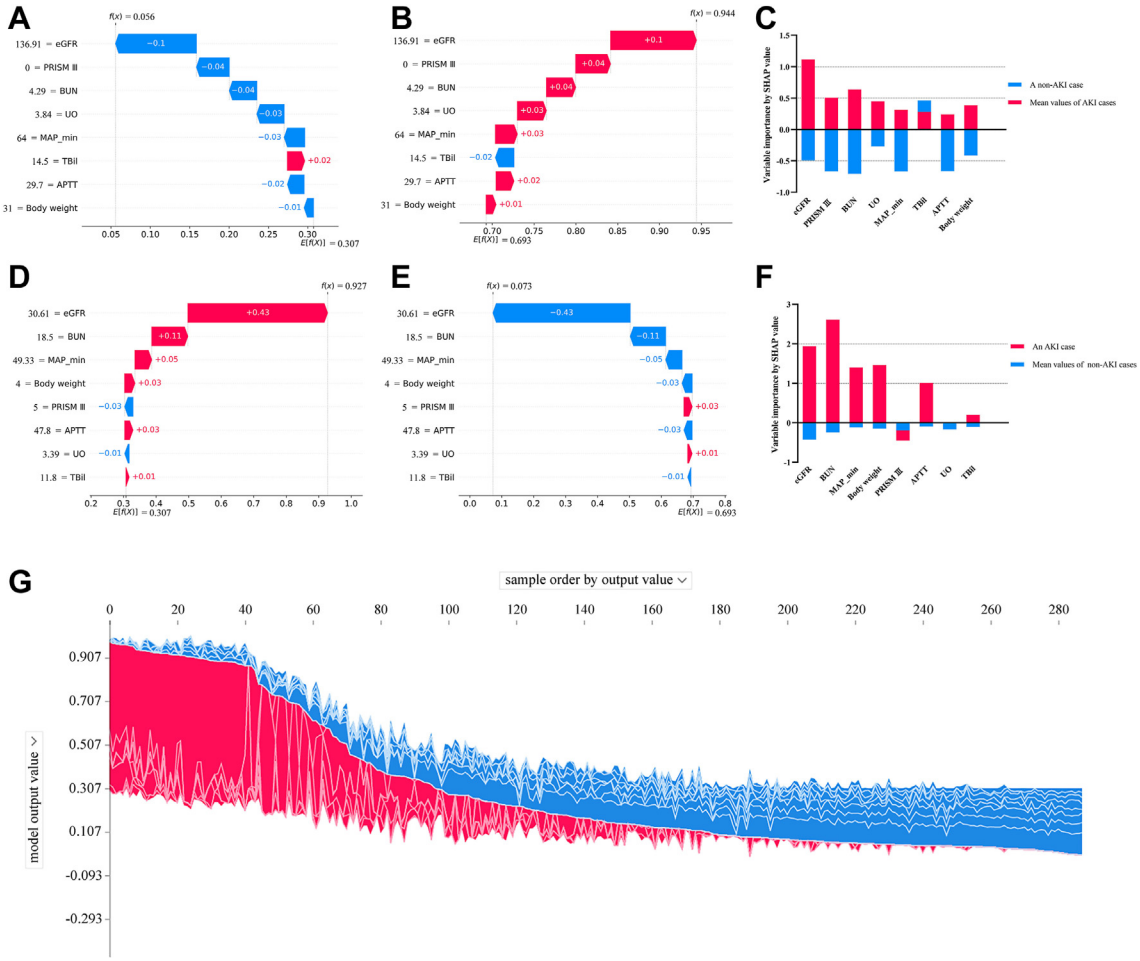


*Fig. 4:* **Local model explanation by the SHAP method. (A–F)** Waterfall plot and evolution of risks contributed by each feature for individual child at low **(A–C)** or high **(D–F)** risk of developing AKI: **A** and **D** represented the individual child towards the "AKI" class, and **B** and **E** represented the individual child towards the "non-AKI" class. In **C** and **F**, all the input features were standardized to zero-mean and unit-variance. **(G)** Force plot for the internal validation set. Each patient was represented by the x-axis, while the features' contributions were represented by the y-axis: an increased red part for each individual patient represented a greater probability towards the decision of "AKI". AKI: acute kidney injury; APTT: activated partial thromboplastin time; BUN: blood urea nitrogen; eGFR: estimated glomerular filtration rate; MAP: mean arterial pressure; PRISM III: pediatric risk of mortality III; SHAP: SHapley Additive explanation; TBil: total bilirubin; UO: urine output.

validation cohort is illustrated in Fig. 4G. The x-axis represents each patient, and the y-axis represents the contributions of the features. An increased red part for each individual patient represents a greater probability towards the decision of "AKI".

## Convenient application for clinical utility

The final prediction model was implemented into the web application to facilitate its utility in clinical scenarios, as shown in Fig. 5. When the actual values of the 8 features required for the model are entered, this application will automatically predict the risk of AKI for an individual child. Additionally, a force plot for the individual child will also be displayed to indicate the features that contribute to the decision of AKI: the blue features on the right are the features pushing the prediction towards "non-AKI", while the red features on the left are pushing the prediction towards "AKI". The web application is accessible online at https://prediction-model-for-aki.streamlit.app.

## Prognostic implications

The investigation of prognostic implications employed all cases, including 1823 critically ill children from the training, internal validation, and external validation cohorts. Supplementary Table S6 lists several probability cutoff values that had a Youden index exceeding 0.8 in the training cohort, as well as their sensitivity, specificity, PPV, and NPV for predicting AKI, severe AKI, and persistent AKI in 1823 critically ill children.

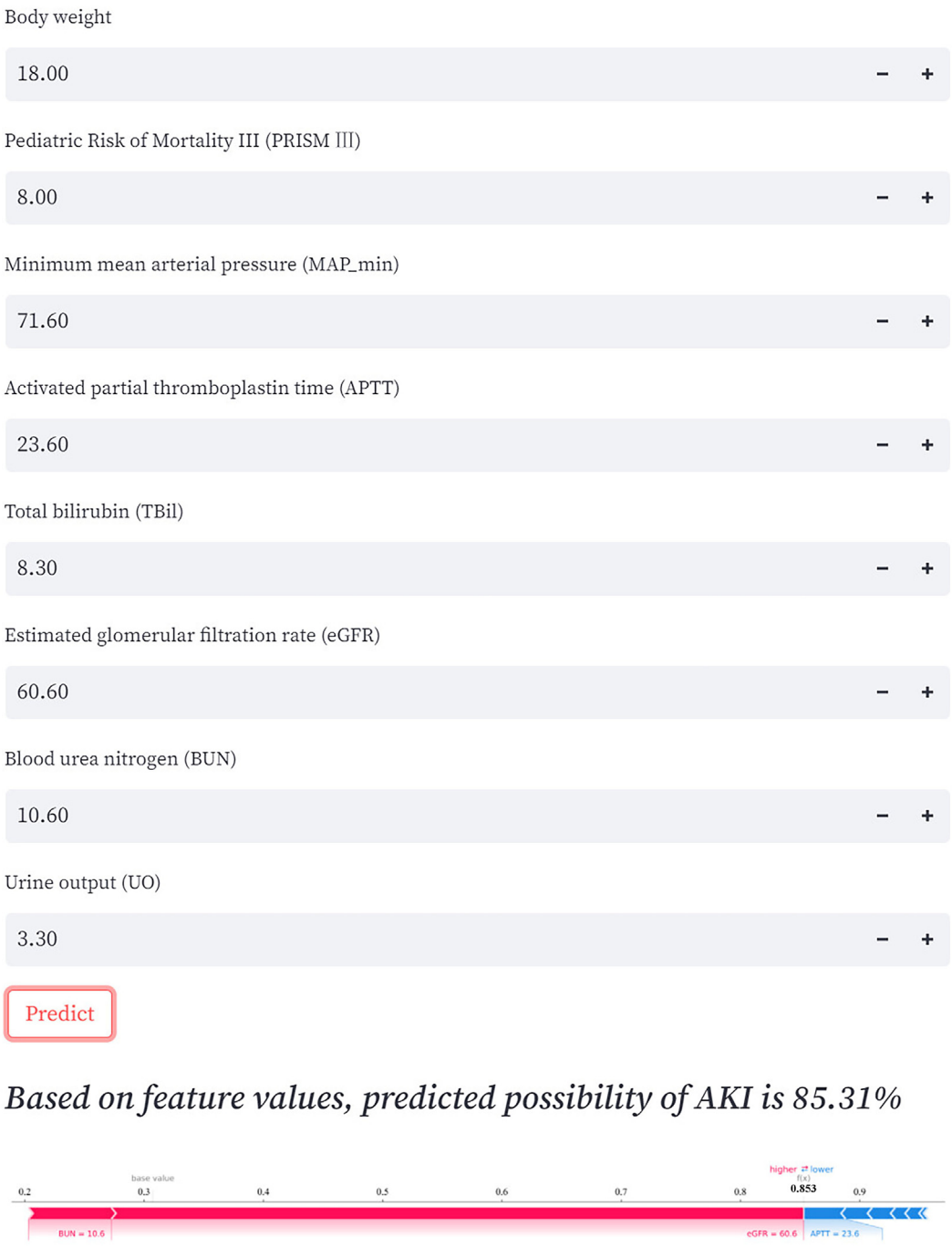


***Fig. 5:* Convenient application for clinical utility.** The convenient application of the final RF model with 8 features is available for AKI prediction. When entering actual values of the 8 features, this application automatically displays the probability of 85.31%. Meanwhile, the force plot for individual child indicates the features that contribute to the decision of "AKI": the blue features on the right are the features pushing the prediction towards the "non-AKI" class, while the red features on the left are pushing the prediction towards the "AKI" class. AKI: acute kidney injury; RF: random forest.

Supplementary Fig. S6 shows the performance of the final model in predicting AKI, severe AKI, and persistent AKI during the whole PICU stay as well as the first week of the PICU stay. The optimal probability cutoff value of 30.8%, which maximized sensitivity and specificity, for predicting AKI was used as the clinical subgrouping threshold to define "Prediction model (+)" and "Prediction model (–)". Critically ill children with a probability equal to or greater than 30.8%, defined as "Prediction model (+)", had a higher incidence of AKI (76.4% vs. 3.5%), multiple organ dysfunction syndrome (MODS) (46.5% vs. 14.1%), and shock/disseminated intravascular coagulation (DIC) (27.0% vs. 3.9%), as well as a longer length of PICU stay (117.9 h vs. 74.8 h) and an increased PICU mortality rate (20.6% vs. 3.9%) than those with a probability less than 30.8%, defined as "Prediction model (–)".

To further investigate the prognostic implications of AKI identified by this explainable model, children were classified into subgroups based on the combination of the final model and KDIGO criteria: subgroup 1 [AKI(–)/Prediction model (–)], subgroup 2 [AKI(–)/Prediction model (+)], subgroup 3 [AKI(+)/Prediction model (–)], and subgroup 4 [AKI(+)/Prediction model (+)], as shown in Fig. 6.

The demographic and clinical characteristics of these four subgroups are presented in Table 2. As illustrated in Fig. 6A and B, children in subgroups 2 and 4 had significantly higher mortality than those in subgroup 1, respectively. Meanwhile, children in subgroup 2 had higher illness severity assessed by the PRISM III score and were more likely to be complicated by MODS and shock/DIC, as well as requiring mechanical ventilation, compared to those in subgroup 1. Children in subgroup 4 had increased scores of the illness severity and were more likely to develop shock/DIC than those in subgroup 3, as shown in Table 2 and Fig. 6B. Although children in subgroup 2, defined as AKI(–)/Prediction model (+), did not develop KDIGO-defined AKI, they had higher illness severity scores and a comparable incidence of mortality and MODS compared to children in subgroup 3, defined as AKI(+)/Prediction model (–).

Moreover, broad differences in the distributions of clinical variables across subgroups are depicted in Fig. 6C. Some variables reflecting cardiovascular, hepatic, renal functions, coagulation, and inflammatory reactions were worse in subgroup 2 than subgroup 1. Similar differences were observed between children in subgroups 3 and 4.

## Discussion

This is the first prospective, multicenter study, to our knowledge, to investigate and compare 11 ML models for comprehensive AKI prediction analyses in PICU cohorts. We identified a set of predictive risk factors and constructed a prediction model for children admitted to the PICU using ML algorithms alongside clinical and laboratory data easily extracted from the EMR system.

To date, numerous studies have concentrated on AKI prediction,[33,34] but a single analyte seems to be difficult to put into clinical practice due to the heterogeneous underlying pathophysiology of AKI.[4] The ML technique is a powerful computational method to deal with complex and extensive data because it can handle highly variable data sets and understand the complex relationship between variables in a way that is flexible and can be trained. Combining EMR data, which are easily available and more accurate for clinicians and researchers to gain clinical data, with sophisticated ML algorithms can facilitate the development of clinical prediction models.[35] Among 11 ML models, the RF model had the best AUC value with a good net benefit and a high threshold probability in feature reduction. RF is an ensemble classifier that combines a set of decision trees by majority voting and is widely used as a classification model.[36] Several studies have proven that the RF method has excellent predictive value in the field of medicine.[37–39] In this study, we employed the RF algorithm to develop a final model with 8 features. These features can be obtained or evaluated easily within the first 24 h of PICU admission, making this model promising as an early discriminative tool for AKI in critically ill children during the PICU stay, even for those who have not met the criteria for AKI at the moment of data collection.

Due to the lack of guidelines or consensus for selecting features for the prediction model, how many features should be included in the model remains elusive. Although more features may provide more information for the prediction model, including a large number of features may limit the clinical use of the model, and including non-causal features may reduce the accuracy of the prediction.[40] The SHAP method was employed to assist feature selection. Our final model, established as a simple and convenient ML prediction model, might be easily used to facilitate clinical decision-making in PICU populations.

The final model we developed had a superior ability to predict AKI in pediatric cohorts compared to traditional single markers. The major characteristics of AKI are a rise in SCr and a decline in UO during a brief period of time.[2] However, low and varying SCr levels are characteristic in young children.[41,42] We used eGFR instead of SCr in the development of model due to the fact that eGFR gives more information than just SCr alone and reflects the physical conditions of specific children of different ages. Due to the critical significance of eGFR and UO in the AKI diagnosis, including these two features was beneficial for strengthening the predictive ability of the final model. Nevertheless, comparing the predictive power of the final model to that of the eGFR, SCr, BUN, and UO, we observed that the final model outperformed each of the traditional

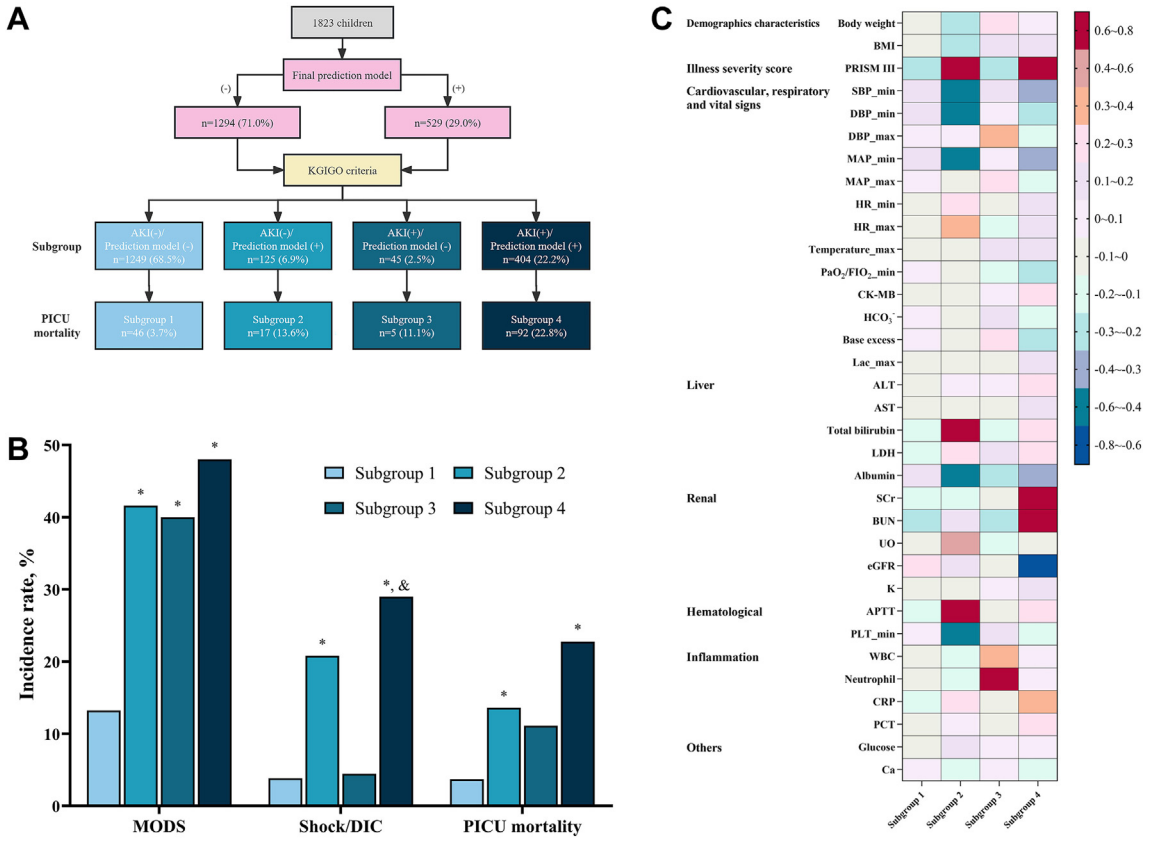


***Fig. 6:*** **Comparison of the incidence rate and clinical variables according to the final model and KDIGO criteria. (A)** Flow chart describing 1823 children subgrouped according to the final model and KDIGO criteria. AKI was defined based on serum creatinine and urine output in accordance with the KDIGO criteria. Prediction model (−) represented children with a probability <30.8%, and prediction model (+) represented children with a probability ≥30.8% in the final model. **(B)** MODS ($P < 0.001$), Shock/DIC ($P < 0.001$), and PICU mortality ($P < 0.001$) among subgroups. **(C)** Heatmap of the distribution of clinical and laboratory variables. Variables with a statistical difference among four subgroups ($P < 0.05$) were included. Each variable was standardized to zero-mean and unit-variance. The color gradient was used to show differences in mean values, with red for higher values and blue for lower values. Max and min represented the maximum and minimum values during the first 24 h after PICU admission, respectively. *$P < 0.05$ vs. subgroup 1, $^{\#}P < 0.05$ vs. subgroup 2, $^{\&}P < 0.05$ vs. subgroup 3.

markers, respectively. In addition, the final model performed better than the RAI and the PRISM III score, which reflect the risk and injury state of the kidney and the illness severity, respectively, demonstrating its excellent ability to predict AKI. It's interesting that the final model also consisted of PRISM III score, TBil, MAP, APTT, and body weight. Although there is no evidence demonstrating their ability to independently predict AKI, these variables are associated with an increased risk of different types of AKI.[43–48] The PRISM III score increases among critically ill children with AKI diagnosed using various criteria.[43,44] Although renal perfusion is heavily reliant on blood pressure and a low MAP level may result in renal hypotension and predispose individuals to AKI,[49] hospitalized children complicated with hypertension or hypotension both have an increased risk of AKI,[50] which supports the U-shape relationship between MAP and AKI observed in this study. Despite the underlying mechanism between body weight and AKI remaining elucidated, the correlation of body weight with AKI has been described previously.[44,48,51] Therefore, these clinical variables could contribute to the final model, and the combination of them may be superior to a single marker in predicting AKI.

Our final model performed well both in internal and external validations, with an AUC of 0.929 and 0.910, respectively. As previously reported, several studies have conducted ML models for the prediction of AKI. However, these researchers specifically focused on AKI in an adult population within a specific clinical setting, such as AKI following cardiac surgery, liver transplantation, or sepsis.[20,21,52] As far as we know, the cause and course of AKI exhibit extensive heterogeneity, and the immature kidney function of children differs from that of adults. Thus, establishing prediction models for

| Variables | Non-AKI n = 1374 | | AKI n = 449 | | P value |
|---|---|---|---|---|---|
| | AKI(−)/Prediction model (−) | AKI(−)/Prediction model (+) | AKI(+)/Prediction model (−) | AKI(+)/Prediction model (+) | |
| Number, n (%) | 1249 (68.5) | 125 (6.9) | 45 (2.5) | 404 (22.2) | |
| Age, month | 41.0 [14.0–100.0] | 38.0 [3.8–87.0] | 79.0 [20.0–123.5] | 39.0 [6.0–120.0] | 0.087 |
| Body weight, kg | 15.0 [10.0–27.0] | 13.0 [6.0–24.0][a] | 20.0 [12.0–40.5][b] | 14.0 [7.4–33.0] | 0.003 |
| BMI, kg/m$^2$ | 16.5 [14.9–18.9] | 16.0 [14.2–17.7][a] | 18.1 [15.4–20.0][b] | 16.6 [14.5–19.4] | 0.005 |
| Male, n (%) | 728 (58.3) | 74 (59.2) | 30 (66.7) | 237 (58.7) | 0.735 |
| PRISM III score | 4.0 [2.0–7.0] | 11.0 [5.0–14.0][a] | 4.0 [1.0–7.0][b] | 9.0 [3.0–15.0][a,c] | <0.001 |
| MODS, n (%) | 165 (13.2) | 52 (41.6)[a] | 18 (40.0)[a] | 194 (48.0)[a] | <0.001 |
| Shock/DIC, n (%) | 48 (3.8) | 26 (20.8)[a] | 2 (4.4) | 117 (29.0)[a,c] | <0.001 |
| MV, n (%) | 210 (16.8) | 34 (27.2)[a] | 19 (42.2)[a] | 172 (42.6)[a,b] | <0.001 |
| RRT, n (%) | 12 (1.0) | 0 (0.0) | 10 (22.2)[a,b] | 96 (23.8)[a,b] | <0.001 |
| Length of PICU stay, hours | 73.2 [43.7–142.1] | 113.3 [64.2–219.4][a] | 217.1 [99.2–332.0][a] | 121.3 [59.6–255.2][a,c] | <0.001 |
| PICU mortality, n (%) | 46 (3.7) | 17 (13.6)[a] | 5 (11.1) | 92 (22.8)[a] | <0.001 |

All 1823 critically ill children from derivation and external validation cohorts were included in this table. Continuous values were presented as median [interquartile range]. Categorical values were presented as number (percentage). AKI developed during the PICU stay was defined based on serum creatinine and urine output in accordance with the KDIGO criteria. Prediction model (−) represented children with a probability <30.8%, and prediction model (+) represented children with a probability ≥30.8% in the final model. AKI: acute kidney injury; BMI: body mass index; DIC: disseminated intravascular coagulation; KDIGO: Kidney Disease: Improving Global Outcome; MODS: multiple organ dysfunction syndrome; MV: mechanical ventilation; PICU: pediatric intensive care unit; PRISM III: Pediatric Risk of Mortality III; RRT: renal replacement therapy. [a]P < 0.05 vs. AKI(−)/Prediction model (−). [b]P < 0.05 vs. AKI(−)/Prediction model (+). [c]P < 0.05 vs. AKI(+)/Prediction model (−).

***Table 2:*** **Comparisons of demographic and clinical characteristics and outcomes among subgroups based on the final model and KDIGO criteria.**

pediatric patients in a heterogeneous PICU cohort is necessary. Although a few retrospective studies were conducted to identify AKI prediction models in hospitalized children,[14–16,53] these models performed less accurately than our final model[14,15,53] or only predicted AKI stages 2 and 3 with a time window of 24–48 h before AKI occurrence.[16] Our final model was established through a comparison of 11 ML models and a feature reduction, and it performed well in predicting AKI in any stage for critically ill children during their whole PICU stay. The SHAP method was employed to explain this model, enabling a better understanding and facilitating its utility for clinicians. Furthermore, children at high risk for adverse outcomes were identified based on the combination of the final model and KDIGO criteria. The aforementioned findings clarified the universal clinical application and advantages of our final model.

The ML technique has been described as a "black-box" with little explanation about how predictions are derived. This may result in clinicians refusing to use it because they are hesitant to make medical decisions based on opaque information. This brought up another advantage of this study: we utilized the SHAP approach to explain the "black-box" of ML models. The SHAP method could explain this model via a global explanation that describes the overall functionality of a model and a local explanation that details how a certain prediction is made for an individual child by inputting the individualized data. Moreover, with a convenient tool based on the Streamlit framework, this prediction model can be used on the webpage and shared with more clinicians.

The other major finding in this study was to utilize the ML model and KDIGO criteria to identify clinical subgroups and evaluate their prognostic implications in critically ill children. Due to the heterogeneous etiology and complicated pathophysiology of AKI, it is a challenge for clinicians to manage AKI patients in clinical settings. The subgrouping for AKI may be used to identify patients at high risk and improve the diagnostic accuracy of AKI.[54] In the present study, we used the RF model in a special and innovative way to define clinical subgroups, and the probability of 30.8%, as the optimal cutoff value that maximized sensitivity and specificity, was used as the threshold for subgrouping. AKI children predicted only by our final model exhibited higher levels of illness severity and a comparable incidence of multiple organ dysfunctions and death compared to AKI children predicted only by the KDIGO criteria. Moreover, among critically ill children with and without KDIGO-defined AKI, those with a probability exceeding or equal to the threshold had a higher illness severity and an increased risk of death and organ complications. Considering the association with poor outcomes, subgrouping critically ill children had potential clinical significance and provided important information for clinicians. Our findings verified that clinicians should concentrate more on critically ill children predicted by the final prediction model due to the fact that they might be more susceptible to organ complications no matter whether they meet the KDIGO criteria for AKI. Consequently, not all standardly staged AKI is associated with worse outcomes within PICU patients, and the combination of prediction model and KDIGO criteria is superior to traditional markers alone in predicting AKI severity and poor outcomes.

We should acknowledge several limitations of this study. First, as a prevalent issue in hospitalized patients, baseline SCr values were unavailable for most of the children in this study.[26,42] We estimated the baseline SCr

using the modified Schwartz estimating equation, which has been employed in many pediatric AKI studies, to minimize errors as much as possible.[26,27] Second, we built the AKI prediction model in heterogeneous PICU cohorts, regardless of the causes of AKI. However, AKI is complex in its pathophysiological mechanisms and accompanied by diverse etiologies. The question of whether this model performs well in predicting various AKI types remains unanswered. Third, this model was constructed based on Chinese populations, and its generalizability to global populations was not clear. However, this is a multicenter study with a heterogeneous pediatric population, which could provide adequate evidence for our results to be generalizable. Further assessment is required for the generalizability of this model. Fourth, although "big data" is required for ML techniques to construct prediction models, no standard is available for computing sample sizes for the development of machine learning-based prediction models. Nevertheless, well-performed cross and external validations, indicating an appropriate sample size, with a prospective multicenter study design, provided adequate power for exploring the prediction model of AKI in critically ill children. Fifth, it seemed that the results of AKI cases occurring on the first day of PICU stay could be a large driver of the overall study results, which may be a bias of this study towards positive results. Nevertheless, this is a common phenomenon in clinical settings, with most AKI cases identified on the first day of the ICU or PICU stay, as previously reported.[55,56] The final model still performed well for predicting AKI that occurred on days 2–7 of the PICU stay. Additionally, this final model could only predict the occurrence of AKI but not the timing of AKI. Further study is required to investigate the prediction of the timing of AKI occurring, especially in a time window of 24 h or 48 h before AKI occurrence. Sixth, we did not include the disease category in the model. Critically ill children admitted to the PICU frequently complicate more than one disease, and there are plenty of disease categories in clinical settings. These result in the difficulty of incorporating the disease category into the prediction model.

In conclusion, we successfully developed an explainable ML model to predict AKI in critically ill children based on clinical data easily extracted from the EMR system. The final RF model had an excellent ability to predict AKI in both internal and external validations. The optimal probability cutoff value was employed as a threshold, and critically ill children with a probability greater than or equal to the threshold had an increased risk of organ complications and PICU mortality and should receive more attention in clinical settings. Further randomized and controlled studies are required to figure out whether individualized and prompt therapeutic measures according to the final prediction model could improve patient outcomes in PICU cohorts.

### Contributors

JH performed the data analyses, established the machine learning models, and drafted the manuscript. JX, ML, ZJ, JM, LF, KM, and HL participated in data collection. JC, ZB, XL, and GL participated in the design of the study and coordination. YL had primary responsibility for study design, data analyses, data interpretation, and writing the manuscript. JH and YL have accessed and verified the data. All authors read and approved the final manuscript.

### Data sharing statement

The data analyzed and the codes used during the current study are available from the corresponding author on reasonable request.

### Declaration of interests

The authors declare that they have no competing interests.

### Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (81971432), Jiangsu Province Science and Technology Support Program (BE2020660), Key talent of women's and children's health of Jiangsu Province (FRC201738), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX23_1666). The funders had no role in study design, data collection, preparation of the manuscript, and decision to publish.

### Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.102409.

### References

1 Kaddourah A, Basu RK, Bagshaw SM, Goldstein SL, Aware Investigators. Epidemiology of acute kidney injury in critically ill children and young adults. *N Engl J Med*. 2017;376:11–20.
2 Palevsky PM, Liu KD, Brophy PD, et al. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for acute kidney injury. *Am J Kidney Dis*. 2013;61:649–672.
3 Ronco C, Bellomo R, Kellum JA. Acute kidney injury. *Lancet*. 2019;394:1949–1964.
4 De Vlieger G, Forni L, Schneider A. New diagnostics for AKI in critically ill patients: what to expect in the future. *Intensive Care Med*. 2022;48:1632–1634.
5 Koyner JL, Adhikari R, Edelson DP, Churpek MM. Development of a multicenter ward-based AKI prediction model. *Clin J Am Soc Nephrol*. 2016;11:1935–1943.
6 Churpek MM, Carey KA, Edelson DP, et al. Internal and external validation of a machine learning risk score for acute kidney injury. *JAMA Netw Open*. 2020;3:e2012892.
7 Gao W, Wang J, Zhou L, et al. Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. *Comput Biol Med*. 2021;140:105097.
8 Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med*. 2018;46:1070–1077.
9 Alfieri F, Ancona A, Tripepi G, et al. Continuous and early prediction of future moderate and severe Acute Kidney Injury in critically ill patients: development and multi-centric, multi-national external validation of a machine-learning model. *PLoS One*. 2023;18:e0287398.
10 Gong K, Lee HK, Yu K, Xie X, Li J. A prediction and interpretation framework of acute kidney injury in critical care. *J Biomed Inform*. 2021;113:103653.
11 Jiang X, Hu Y, Guo S, Du C, Cheng X. Prediction of persistent acute kidney injury in postoperative intensive care unit patients using integrated machine learning: a retrospective cohort study. *Sci Rep*. 2022;12:17134.
12 Shawwa K, Ghosh E, Lanius S, Schwager E, Eshelman L, Kashani KB. Predicting acute kidney injury in critically ill patients using comorbid conditions utilizing machine learning. *Clin Kidney J*. 2021;14:1428–1435.
13 Flechet M, Güiza F, Schetz M, et al. AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum

neutrophil gelatinase-associated lipocalin. *Intensive Care Med*. 2017;43:764–773.
14 Sanchez-Pinto LN, Khemani RG. Development of a prediction model of early acute kidney injury in critically ill children using electronic health record data. *Pediatr Crit Care Med*. 2016;17:508–515.
15 Wang L, McGregor TL, Jones DP, et al. Electronic health record-based predictive models for acute kidney injury screening in pediatric inpatients. *Pediatr Res*. 2017;82:465–473.
16 Dong J, Feng T, Thapa-Chhetry B, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. *Crit Care*. 2021;25:288.
17 Azodi CB, Tang J, Shiu SH. Opening the black box: interpretable machine learning for geneticists. *Trends Genet*. 2020;36:442–455.
18 Sutherland SM, Chawla LS, Kane-Gill SL, et al. Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15(th) ADQI Consensus Conference. *Can J Kidney Health Dis*. 2016;3:11.
19 Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *31st conference on neural information processing systems (NIPS 2017)*. 2017. Long Beach, CA, USA.
20 Tseng PY, Chen YT, Wang CH, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020;24:478.
21 Zhang Y, Yang D, Liu Z, et al. An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation. *J Transl Med*. 2021;19:321.
22 American Academy of Pediatrics, Committee on Hospital Care and Section of Critical Care, Society of Critical Care Medicine, Pediatric Section Admission Criteria Task Force. Guidelines for developing admission and discharge policies for the pediatric intensive care unit. *Pediatrics*. 1999;103:840–842.
23 Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med*. 1996;24:743–752.
24 Basu RK, Zappitelli M, Brunner L, et al. Derivation and validation of the renal angina index to improve the prediction of acute kidney injury in critically ill children. *Kidney Int*. 2014;85:659–667.
25 Basu RK, Kaddourah A, Goldstein SL, AWARE Study Investigators. Assessment of a renal angina index for prediction of severe acute kidney injury in critically ill children: a multicentre, multinational, prospective observational study. *Lancet Child Adolesc Health*. 2018;2:112–120.
26 Hursh BE, Ronsley R, Islam N, Mammen C, Panagiotopoulos C. Acute kidney injury in children with type 1 diabetes hospitalized for diabetic ketoacidosis. *JAMA Pediatr*. 2017;171:e170020.
27 Myers SR, Glaser NS, Trainor JL, et al. Frequency and risk factors of acute kidney injury during diabetic ketoacidosis in children and association with neurocognitive outcomes. *JAMA Netw Open*. 2020;3:e2025481.
28 Schwartz GJ, Muñoz A, Schneider MF, et al. New equations to estimate GFR in children with CKD. *J Am Soc Nephrol*. 2009;20:629–637.
29 Chawla LS, Bellomo R, Bihorac A, et al. Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup. *Nat Rev Nephrol*. 2017;13:241–257.
30 Ikemura N, Spertus JA, Kimura T, et al. Baseline and post-procedural health status outcomes in contemporary patients with atrial fibrillation who underwent catheter ablation: a report from the Japanese outpatient registry. *J Am Heart Assoc*. 2021;10:e019983.
31 Berkelmans GFN, Read SH, Gudbjörnsdottir S, et al. Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice. *J Clin Epidemiol*. 2022;145:70–80.
32 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
33 Pickkers P, Darmon M, Hoste E, et al. Acute kidney injury in the critically ill: an updated review on pathophysiology and management. *Intensive Care Med*. 2021;47:835–850.
34 Poston JT, Koyner JL. Sepsis associated acute kidney injury. *BMJ*. 2019;364:k4891.
35 Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. *Cell*. 2020;181:92–101.
36 Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
37 Kang MW, Kim J, Kim DK, et al. Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. *Crit Care*. 2020;24:42.
38 Zhang X, Yue P, Zhang J, et al. A novel machine learning model and a public online prediction platform for prediction of post-ERCP-cholecystitis (PEC). *EClinicalMedicine*. 2022;48:101431.
39 Potash E, Ghani R, Walsh J, et al. Validation of a machine learning model to predict childhood lead poisoning. *JAMA Netw Open*. 2020;3:e2012734.
40 Li Y, Sperrin M, Ashcroft DM, van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar. *BMJ*. 2020;371:m3919.
41 Xu X, Nie S, Zhang A, et al. A new criterion for pediatric AKI based on the reference change value of serum creatinine. *J Am Soc Nephrol*. 2018;29:2432–2442.
42 Kuai Y, Li M, Chen J, et al. Comparison of diagnostic criteria for acute kidney injury in critically ill children: a multicenter cohort study. *Crit Care*. 2022;26:207.
43 Palmieri T, Lavrentieva A, Greenhalgh D. An assessment of acute kidney injury with modified RIFLE criteria in pediatric patients with severe burns. *Intensive Care Med*. 2009;35:2125–2129.
44 Hu J, Zhou Y, Huang H, et al. Prediction of urinary dickkopf-3 for AKI, sepsis-associated AKI, and PICU mortality in children. *Pediatr Res*. 2023;93:1651–1658.
45 Choi HK, Song YG, Han SH, et al. Clinical features and outcomes of acute kidney injury among patients with acute hepatitis A. *J Clin Virol*. 2011;52:192–197.
46 He L, Liang S, Liang Y, et al. Defining a postoperative mean arterial pressure threshold in association with acute kidney injury after cardiac surgery: a prospective observational study. *Intern Emerg Med*. 2023;18:439–448.
47 Xu Z, Zhang Y, Zhang C, Xiong F, Zhang J, Xiong J. Clinical features and outcomes of COVID-19 patients with acute kidney injury and acute kidney injury on chronic kidney disease. *Aging Dis*. 2022;13:884–898.
48 Hilmi IA, Damian D, Al-Khafaji A, et al. Acute kidney injury following orthotopic liver transplantation: incidence, risk factors, and effects on patient and graft outcomes. *Br J Anaesth*. 2015;114:919–926.
49 Vanmassenhove J, Kielstein J, Jörres A, Biesen WV. Management of patients at risk of acute kidney injury. *Lancet*. 2017;389:2139–2151.
50 Nugent JT, Ghazi L, Yamamoto Y, Bakhoum C, Wilson FP, Greenberg JH. Hypertension, blood pressure variability, and acute kidney injury in hospitalized children. *J Am Heart Assoc*. 2023;12:e029059.
51 Schetz M, De Jong A, Deane AM, et al. Obesity in the critically ill: a narrative review. *Intensive Care Med*. 2019;45:757–769.
52 Yue S, Li S, Huang X, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J Transl Med*. 2022;20:215.
53 Sandokji I, Yamamoto Y, Biswas A, et al. A time-updated, parsimonious model to predict AKI in hospitalized children. *J Am Soc Nephrol*. 2020;31:1348–1357.
54 Ostermann M, Zarbock A, Goldstein S, et al. Recommendations on acute kidney injury biomarkers from the acute disease quality initiative consensus conference: a consensus statement. *JAMA Netw Open*. 2020;3:e2019209.
55 Matsa R, Ashley E, Sharma V, Walden AP, Keating L. Plasma and urine neutrophil gelatinase-associated lipocalin in the diagnosis of new onset acute kidney injury in critically ill patients. *Crit Care*. 2014;18:R137.
56 Ferreira MCR, Lima EQ. Impact of the development of acute kidney injury on patients admitted to the pediatric intensive care unit. *J Pediatr*. 2020;96:576–581.