scientific reports

Check for updates

OPEN

Construction of a machine learning-based interpretable prediction model for acute kidney injury in hospitalized patients

Xiang Yu[1], WanLing Wang[2], RiLiGe Wu[2], XinYan Gong[1], YuWei Ji[1] & Zhe Feng[1]✉

In this observational study, we used data from 59,936 hospitalized adults to construct a model. For the models constructed with all 53 variables, all five models achieved acceptable performance with the validation cohort, with the extreme gradient boosting (XGBoost) model showing the best predictive efficacy and stability (area under the curve (AUC), 0.9301). For the simpler models constructed with 39 significant variables screened by the random forest recursive feature elimination method, the XGBoost model also had the best performance (AUC, 0.9357). All the models showed significant net returns according to decision analysis curves, and the XGBoost model achieved the optimal results. In addition, the Shapley additive explanation (SHAP) importance matrices revealed that uric acid, colloidal solution, first creatinine value on admission, pulse and albumin represented the top five most important variables for both modeling strategies. With the external validation cohort based on 4022 hospitalized patients, the performance of all models declined, among which the Support vector machine (SVM) model showed the best predictive efficacy (AUC, 0.8230 and 0.8329), followed by the XGBoost model (0.8124 and 0.8316). Thus, our model can predict the occurrence and risk of acute kidney injury (AKI) up to 48 h in advance, enabling clinicians to assess the risk of AKI in hospitalized patients more accurately and intuitively and to develop necessary AKI management strategies.

Acute kidney injury (AKI) is a complex syndrome that is caused by a variety of factors[1–4]. AKI is diagnosed using SCr measurements and/or urine output, which are considered the gold standards for diagnosing AKI. However, the lagging effect of SCr and the inaccuracy of urine volume recordings both limit the efficacy of predicting AKI to some extent, and biomarker studies, which have been gradually emphasized in recent years, have been limited. To date, only matrix metalloproteinase inhibitor-2 (Timp-2) and recombinant human insulin-like growth factor-binding protein-7 (IGFBP7) have been approved by the FDA, and their clinical availability is still questionable[5–7].

Artificial intelligence (AI) techniques, including machine learning methods, have been used to develop risk prediction models, and AI techniques have several advantages, especially when dealing with large-scale data from electronic health records. Since AKI events can be determined through mathematical calculations based on SCr levels, the development of AKI risk prediction models based on AI techniques is promising. Some studies have reported algorithms developed on the basis of electronic health records, and prior findings have confirmed that these algorithmic models can predict AKI earlier than the SCr model can[8–10]. In particular, the results of Tomasev et al.'s[11] team have advanced the current understanding in this field.

In this study, we aimed to use machine learning algorithms to construct an interpretable real-time AKI prediction model based on large-sample data and AI technology that can be used for hospitalized patients in various wards in Chinese general hospitals, such as our hospital, so that clinicians can more accurately and intuitively assess the risk of AKI for hospitalized patients and develop necessary management strategies.

[1]First Medical Center of Chinese PLA General Hospital, Department of Nephrology, First Medical Center of Chinese PLA General Hospital, State Key Laboratory of Kidney Diseases, National Clinical Research Center for Kidney Diseases, Beijing Key Laboratory of Medical Devices and Integrated Traditional Chinese and Western Drug Development for Severe Kidney Diseases, Beijing Key Laboratory of Digital Intelligent TCM for the Preventionand Treatment of Pan-vascular Diseases, Key Disciplines of National Administration of Traditional Chinese Medicine(zyyzdxk-2023310), Beijing 100853, China. [2]Medical Innovation Research Division, Chinese PLA General Hospital, Beijing 100853, China. ✉email: zhezhe_4025@126.com

Results

A total of 59,936 patients hospitalized at the First Medical Center of PLA General Hospital were included in the modeling study, with 29,968 patients in the AKI group and 29,968 patients in the non-AKI group. A total of 4,022 hospitalized patients at the Fifth Medical Center of the PLA General Hospital were included in an external validation cohort, including 854 AKI patients and 3,168 non-AKI patients. In both databases, there were significant differences between patients in the AKI group and those in the non-AKI group (Table S1 and Fig. S1).

Screening for significant variables

The Pearson's correlations for different characteristics revealed strong correlations among the five groups for nine variables among the 58 preselected variables: prothrombin activity versus prothrombin time versus international normalized ratio, percentage of lymphocytes versus percentage of neutrophils, total versus direct bilirubin, hemoglobin, and erythrocyte pressure volume (Fig. 1a). The variables retained after variables with lower correlation variables were excluded included the international normalized ratio, percentage of neutrophils, hemoglobin and direct bilirubin. In addition, the best diagnostic accuracy in the cross-validation test was achieved with a model using 39 features based on RF-RFE (Fig. 1b). The 39 significant variables that were identified included 20 laboratory test variables (CRP, D-dimer, UA, SCr, Alb, Hb, LDH, ALT, DBil, WBC, AST, Fib, TT, ALP, NA, K, INR, NE, PLT, and CK), 8 therapeutic medication variables (colloidal solutions, vasoactive drugs, immunosuppressants, glucocorticoids, nonsteroidal anti-inflammatory drugs, proton pump inhibitors, diuretics, and statins), 4 vital sign variables (systolic blood pressure, temperature, diastolic blood pressure, and pulse rate), 2 baseline characteristic variables (age and mode of admission), 4 special investigation or treatment variables (enhanced CT examination, mechanical ventilation, coronary angiography, and angioembolization), and 1 underlying disease variable (tumor).

Evaluation of model performance

Among the models built using all 53 variables or 39 significant variables, the best predictive efficacy and stability in the internal validation were achieved by the XGBoost model, with area under the curve (AUC) values of 0.9301 and 0.9357, respectively. The SVM model showed the best predictive efficacy with AUC values of 0.8230 and 0.8329 in the external validation, followed by the XGBoost model with AUC values of 0.8124 and 0.8316, respectively. Compared with the former modeling strategy, none of the latter models constructed with the 39 variables showed significant changes in performance, with fluctuating AUC values ranging from 0.0018 to 0.0083 in the internal validation and from −0.0076 to 0.0244 in the external validation (Table 1). Figure 2 and S2 show the receiver operating characteristic (ROC) curves and confusion plots for multiple odels based on the two modeling strategies with the internal and external validation groups.

In terms of calibration, the Brier scores of all the models with the internal and external validation groups were close to 0.2, with the lowest Brier scores for the XGBoost model model (Table 2). Furthermore, Fig. 3 shows the calibration curves of multiple forecasting models obtained in the validation study. With the internal validation group, the deviations in the model calibration curves were all below the reference line (diagonal), indicating that these models underestimated or overestimated lower AKI risks, whereas the models were slightly less effective when tested with the external validation group.

The decision analysis curve (DCA) shows that increasing threshold probabilities and decreasing net gains were observed for all models; however, all the models still achieved significant net gains compared with the hypothetical all-treated and no-treatment models with wide risk thresholds, with the net gains ranging from

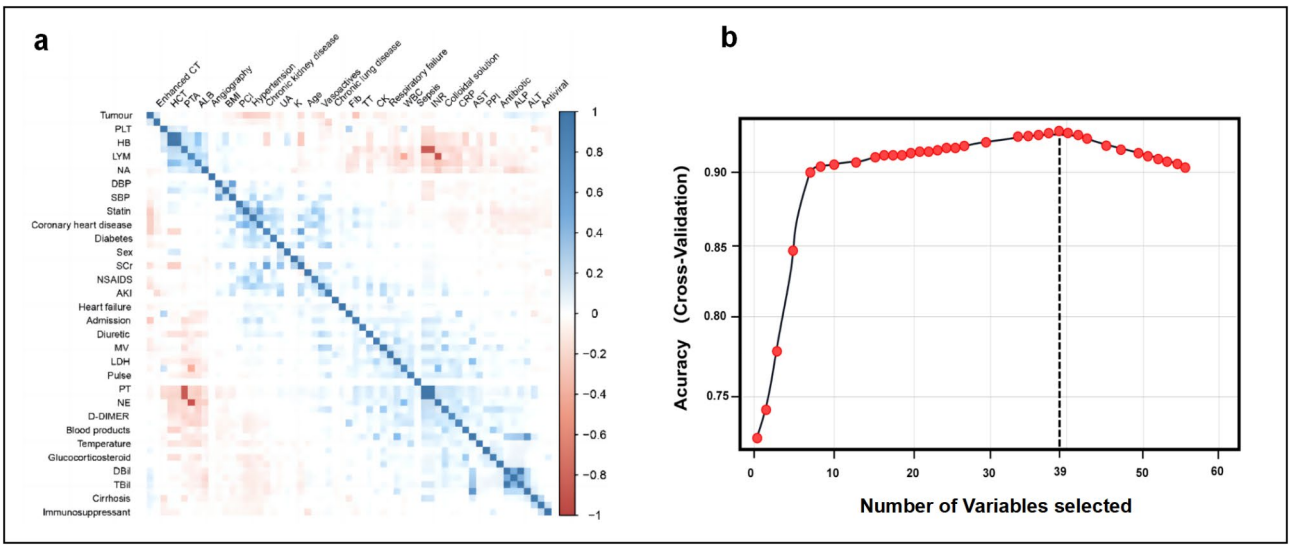

Fig. 1. (**a**) Plot of Pearson's correlations for different characteristics. (**b**) Visualization of optimal model variable combination quantity based on RF-RFE.

Model	The first medical center of the PLA general hospital						The fifth medical center of the PLA general hospital					
	AUC	Accuracy	Sensitivity	Specificity	F1	MCC	AUC	Accuracy	Sensitivity	Specificity	F1	MCC
Performance based on all variables												
Decision tree (DT)	0.8157	0.7911	0.7934	0.7804	0.7923	0.5811	0.7533	0.7545	0.6902	0.6945	0.5773	0.3604
Random forest (RF)	0.9115	0.8485	0.8591	0.8690	0.8678	0.7335	0.7901	0.7985	0.7018	0.6982	0.6963	0.5344
Logistic regression (LR)	0.8603	0.8220	0.7952	0.8275	0.8094	0.6246	0.8034	0.7649	0.6715	0.8475	0.6606	0.6115
Support vector machine (SVM)	0.9029	0.8631	0.8830	0.8581	0.8725	0.7467	0.8230	0.7615	0.8356	0.4848	0.7490	0.6668
Extreme gradient boosting (XGBoost)	0.9301	0.8742	0.8710	0.8788	0.8756	0.7542	0.8124	0.7706	0.7264	0.5713	0.7046	0.7389
Performance based on important variables												
Decision tree (DT)	0.8240	0.7967	0.7942	0.7886	0.7927	0.5829	0.7777	0.7515	0.6934	0.6985	0.3739	0.3634
Random forest (RF)	0.9143	0.8599	0.8661	0.8707	0.8694	0.7347	0.7932	0.7749	0.7112	0.6650	0.4841	0.6150
Logistic regression (LR)	0.8621	0.8024	0.7939	0.8276	0.8100	0.6123	0.7958	0.7482	0.6972	0.8110	0.4578	0.6912
Support vector machine (SVM)	0.9051	0.8633	0.8768	0.8637	0.8736	0.7426	0.8329	0.7654	0.6812	0.6569	0.4659	0.6866
Extreme gradient boosting (XGBoost)	0.9357	0.8683	0.8611	0.8777	0.8716	0.7455	0.8316	0.7886	0.7210	0.7112	0.4574	0.6811

Table 1. Evaluation of model performance. The model constructed based on all variables includes all 53 variables. The model constructed based on important variables includes 39 variables selected through random forest recursive feature elimination method. *AUC* area under the curve, *MCC* Mathews correlation coefficient.

approximately 5–90%. For both types of modeling strategies, the net gains of the four models followed the same order, and all of the XGBoost models yielded acceptable results (Fig. 4).

Model interpretation

In an internal validation study based on data collected at the First Medical Center, the SHAP importance matrix results revealed that uric acid, colloidal solution, first creatinine value on admission, pulse and albumin were the top five most important variables for both modeling strategies (Fig. 5). The SHAP summary plot revealed that the risk of AKI occurrence for hospitalized patients changed as the variable values changed, whereas the predictive trends of the variables for AKI events were consistent for both modeling strategies (Fig. 6). The SHAP dependency plots revealed the correlations between changes in the normalized values of the top ten most important variables and AKI occurrence (Fig. S3).

Discussion

In this study, we constructed a robust AKI prediction model for hospitalized patients based on multiple machine learning algorithms that can accurately predict the occurrence of AKI up to 48 h in advance. This model was constructed on the basis of 39 important variables, which were selected to balance the minimum number of features with the best performance according to a multidimensional performance evaluation. Moreover, in internal and external validation studies, the XGBoost model and the RF model obtained the best performance metrics in terms of discrimination, calibration, and clinical utility.

Overall, this study has several advantages over previous research. First, our prediction model was developed on the basis of various patient data and is applicable to almost all hospitalized patients except those with emergency admissions. Moreover, this model may be compatible in different scenarios, including for various disease characteristics, levels of care, patient demographic characteristics, and AKI severity levels.

Second, in some previous studies, open-source data such as the Medical Information Mart for Intensive Care (MIMIC-III/IV) and Amsterdam UMC databases were used for analysis and model development. These databases mainly include ICU case information, and such data have advantages such as being open, integrated, and information rich[12–15]. However, considering model generalizability, ethnographic factors, and the national context of developing countries, we did not utilize such data in the present study. We included Chinese patients because this choice was consistent with one of the objectives of our study, which were to analyze the characteristics of AKI onset in the Chinese population, accurately identify the risk factors for the deterioration of various conditions for hospitalized patients, and establish a modeling approach applicable to the Chinese population while reducing the number of relevant interfering factors.

Third, in this study, we utilized both static and dynamic variable modeling to consider the time-invariant features of patient susceptibility while also incorporating the effects of time-varying features such as laboratory test results and acute exposures. Laboratory test results such as routine blood counts and blood biochemistry are real-time and objective reflections of patient pathophysiological status, whereas acute exposures such as surgery and the postadmission use of nephrotoxic medication represent the impact of immediate diagnostic and therapeutic interventions. This approach differs from the methodology used in most previous studies. For example, in the coronary arteriography-associated AKI prediction model developed by Mehran et al.[16], the included variables were determined on the basis of underlying conditions such as hypertension, heart failure, age, anemia, and diabetes mellitus. Kheterpal et al.[17] also used a large number of susceptibility variables, such as CKD, ascites, heart failure, and age, in their study of risk indices for AKI after general surgery. However, the idea that the inclusion of prehospital susceptibility factors improves the predictive performance of the model is currently being questioned by other researchers, such as Cheng et al.[18], who concluded in 2018 that the addition of prehospital data did not improve model predictions for AKI in the general population of hospitalized patients;

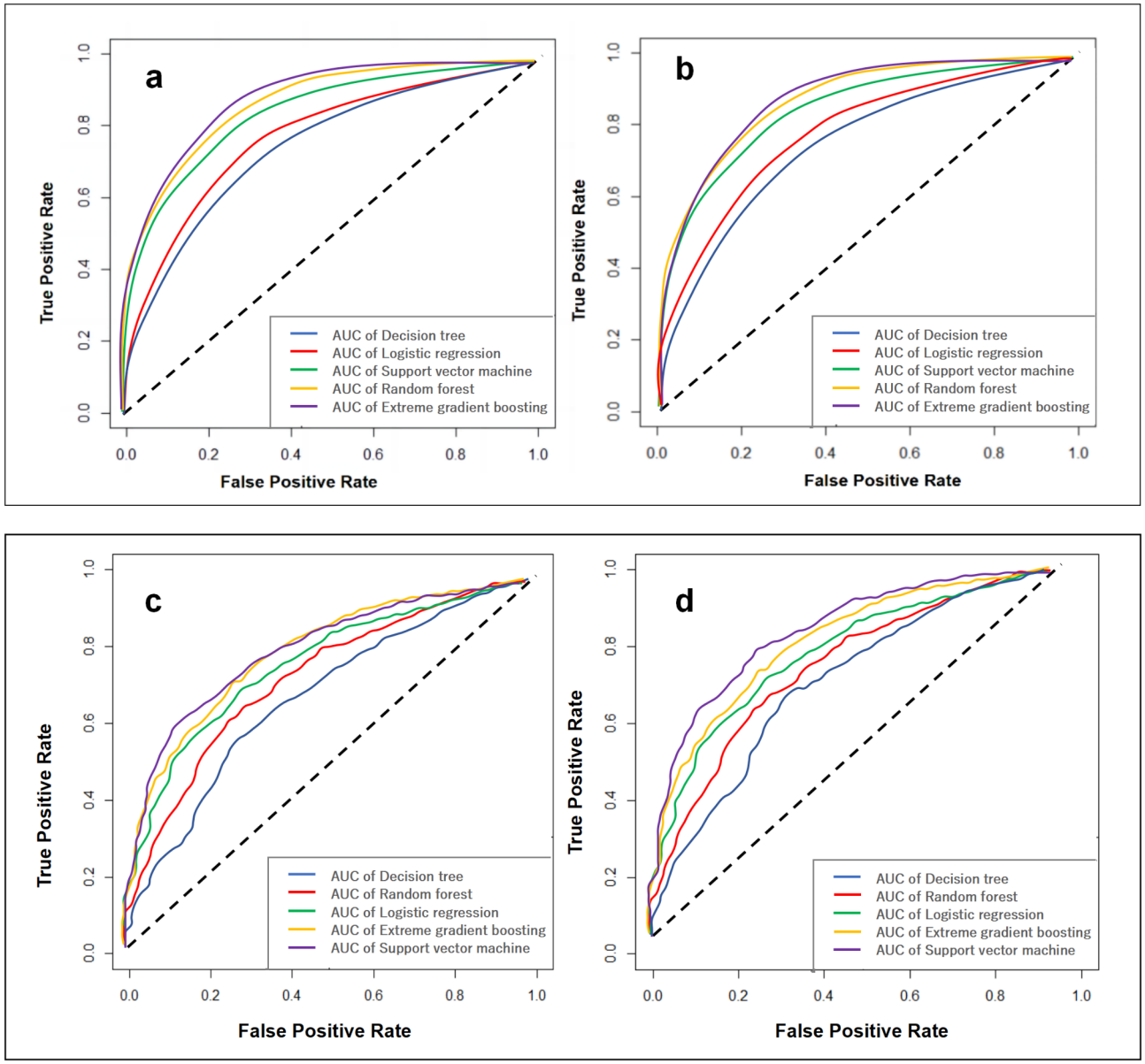

Fig. 2. The receiver operating characteristic (ROC) curves of different models. (**a**) Models constructed using all variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**b**) models constructed using important variables on the basis of the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**c**) models constructed using all variables in the external validation cohort based on data collected at the Fifth Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the external validation cohort based on data collected at the Fifth Medical Center of PLA General Hospital.

this result may have been due to model overfitting of time-varying clinical variables relative to static variables such as demographic factors, admission diagnosis, and comorbidities. To address this issue, although we did not conduct additional controlled tests to determine the effect of the presence or absence of prehospital data on the performance of the prediction model, we performed interpretative analyses and compared the variable importance weights using SHAP values. The results suggested that comorbid tumors and sex contributed more to model predictions than other variables and that these contributions were even more prominent than those of dynamic indicators such as certain laboratory test values. The findings also suggested that the model could not predict AKI among general hospitalized patients. Therefore, it is reasonable to suggest that the inclusion of time-invariant static variables was meaningful in our research.

Fourth, in this study, we collected conventional and easily available variable data in the form of time windows to form a package of "point" events, which were input into the model to calculate specific prediction values, and visualized binary results ("yes" or "no") were output according to the threshold settings to obtain real-time predictions of AKI events and help monitor and manage AKI by providing risk score trajectories and dynamic

Model	The first medical center of the PLA general hospital		The fifth medical center of the PLA general hospital	
	Model 1 ($n=53$)	Model 2 ($n=39$)	Model 1 ($n=53$)	Model 2 ($n=39$)
Decision tree (DT)	0.1562	0.1583	0.2445	0.2415
Random forest (RF)	0.1124	0.1101	0.2088	0.2130
Logistic regression (LR)	0.1332	0.1378	0.1688	0.1769
Support vector machine (SVM)	0.0940	0.0924	0.2281	0.2230
Extreme gradient boosting (XGBoost)	0.0894	0.0919	0.1647	0.1510

Table 2. Comparison of Brier scores for models. Model 1 refers to a model constructed by all 53 variables, model 2 refers to a model constructed by 39 important variables selected through random forest recursive feature elimination method.

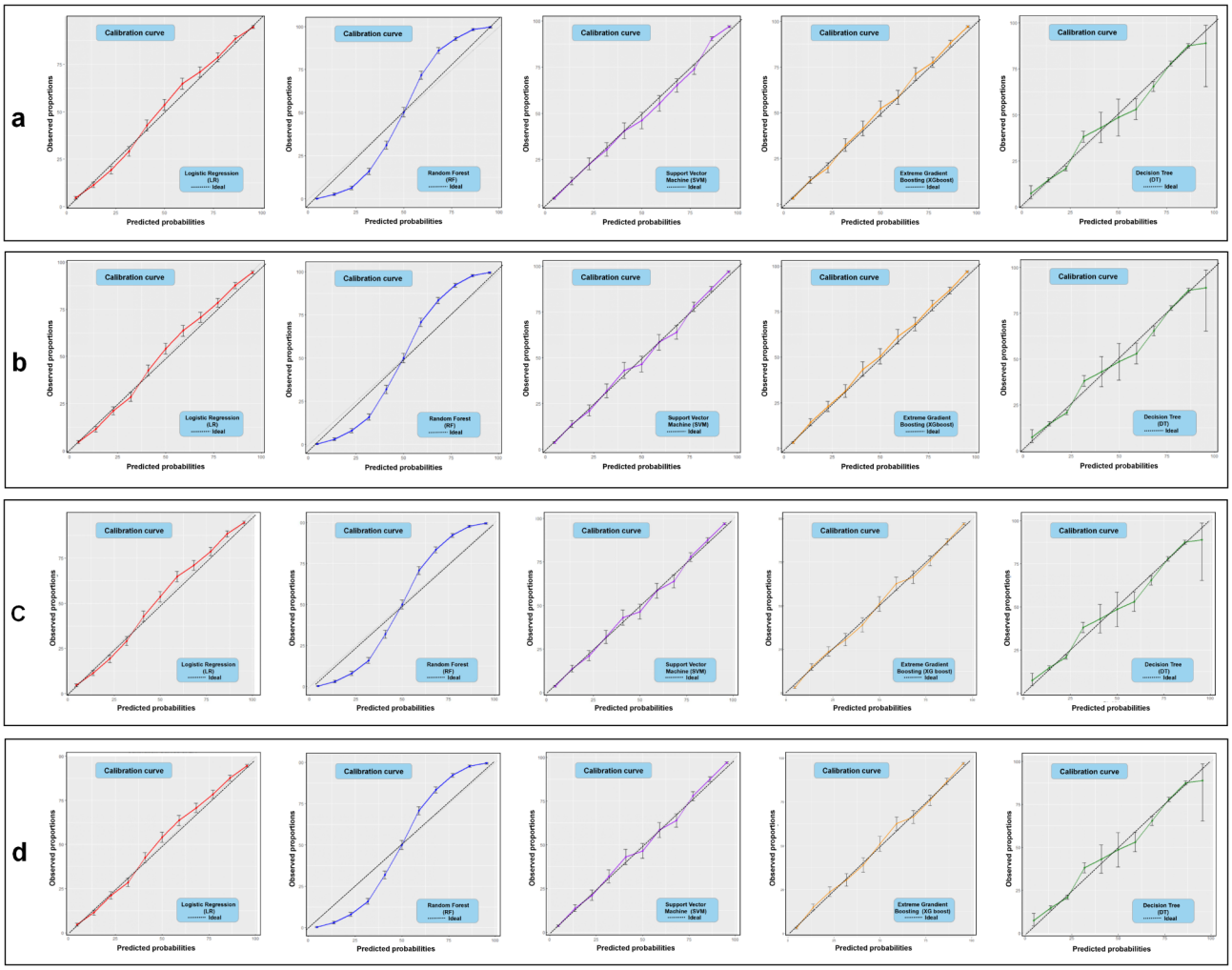

Fig. 3. Calibration curves of different models. (**a**) Models constructed using all variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**c**) models constructed using all variables in the external validation cohort based on data collected at the Fifth Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the external validation cohort based on data collected at the Fifth Medical Center of PLA General Hospital.

trends, which are also essential for model development. In most previous studies, predictions were obtained at specific times, such as 24–48 h after admission, or after medical interventions, whereas the processes involved in hospitalization are continuous, and AKI can occur at any time. One study revealed that 39.4% of AKIs occurred 5 days after admission and that 15.7% of AKIs occurred 10 days after admission. This issue has been reported

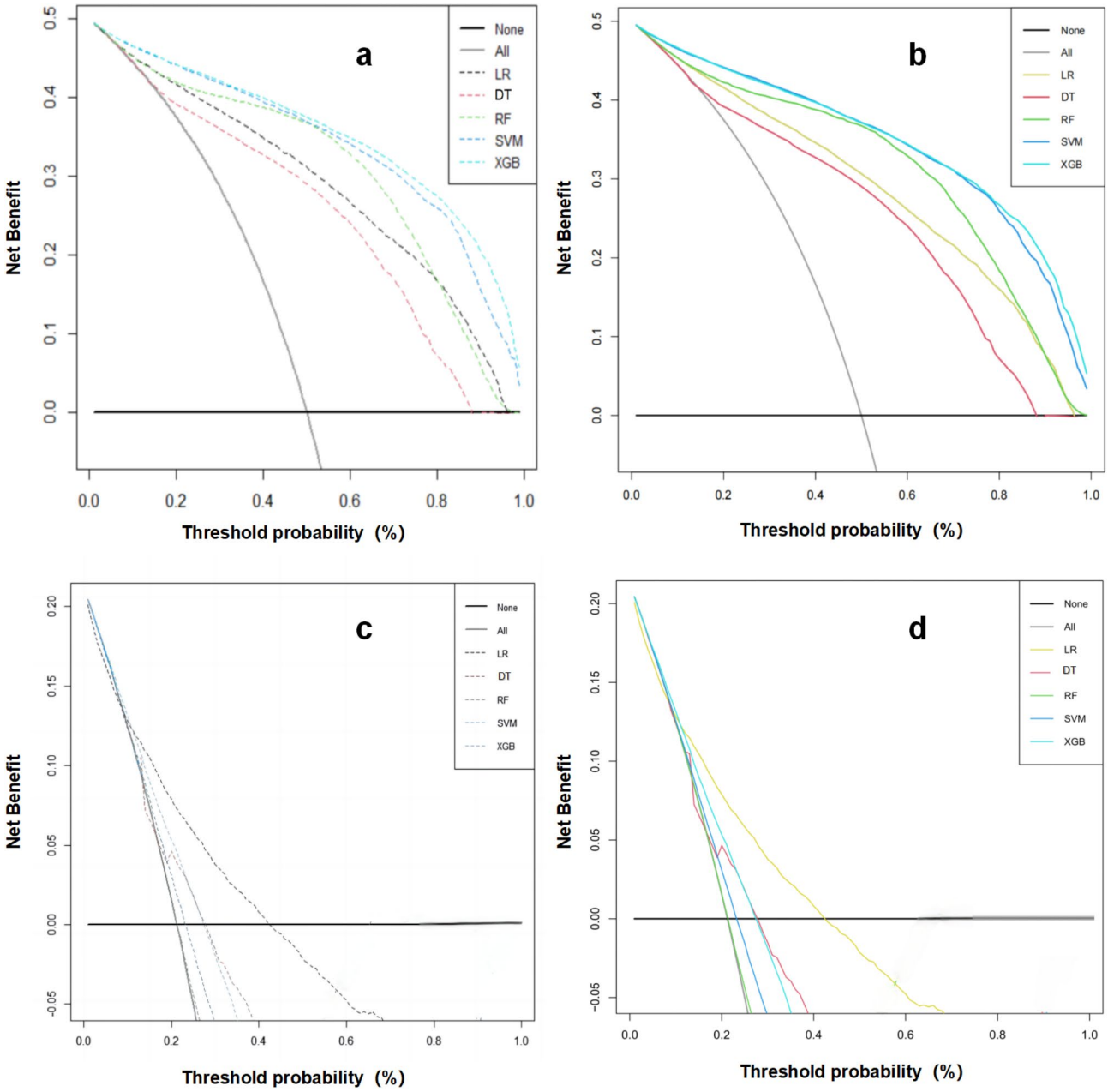

Fig. 4. Decision analysis curves of different models. (**a**) Models constructed using all variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**c**) models constructed using all variables in the external validation cohort based on data collected at the Fifth Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the external validation cohort based on data collected at the Fifth Medical Center of PLA General Hospital.

elsewhere, such as by He et al.[19], who proposed a framework for predicting the daily risk of AKI on the basis of a 24-hour interval, but we note that that study differs from the present work in terms of the prediction time. Moreover, the model assessment by He et al.[19] was separate, i.e., separate AUC values were produced for the first day and the second day after admission, and the model lacks a comprehensive rating index to support the comparative analyses performed in similar studies. Another study was conducted by Chiofolo et al.[20], who used the RF algorithm for real-time AKI risk scoring, obtaining an AUC value of 0.88 for predicting AKI events 6 h in advance based on the validation cohort. However, this approach had several limitations; for example, only patients of one ethnicity (Caucasian) were included in the analysis, limiting model generalizability; moreover, to set the gold standard, an AUC value of 0.88 was used, which was determined based on the validation cohort. Additionally, an AKI detection tool was developed in one study but not tested with humans, resulting in biased

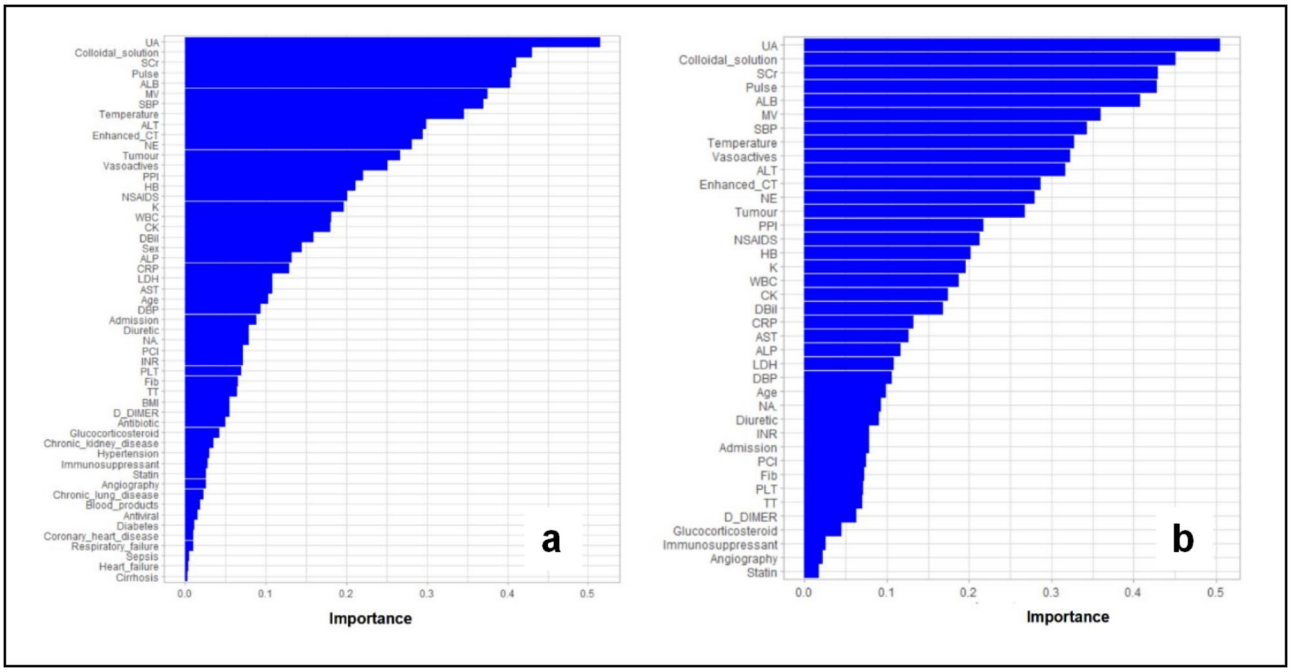

Fig. 5. SHAP importance matrix plots. (**a**) Models constructed using all variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital.

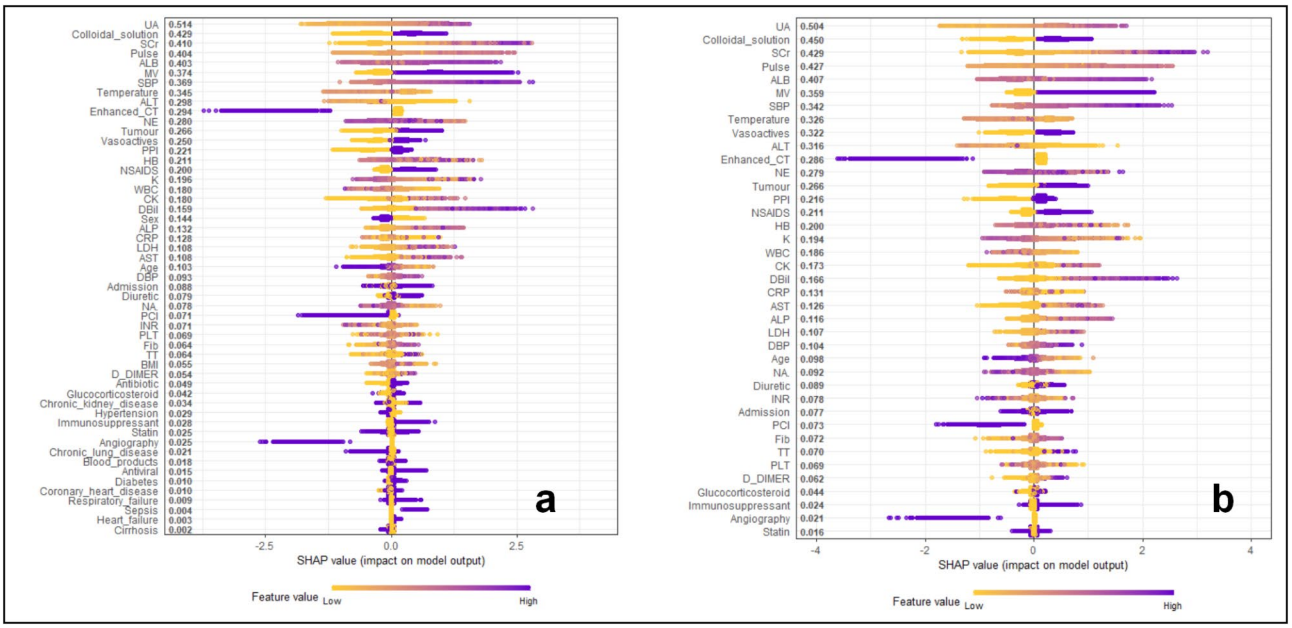

Fig. 6. SHAP summary plots. (**a**) Models constructed using all variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital; (**b**) models constructed using important variables in the internal validation cohort based on data collected at the First Medical Center of PLA General Hospital.

predictions[21]. The results of this study were not as favorable as those of Tomasev et al.[11]. While Tomasev et al.'s recurrent neural network model demonstrated high discriminatory power for real-time AKI prediction, which provides considerable insight for parallel studies, the sex of the patients in the dataset was imbalanced (6% female), and the complex data processing process somewhat limits its external validation.

Fifth, in this study, we chose five algorithms for modeling, and the results showed that even the RF algorithm, which was the lowest ranked algorithm in terms of performance, had an AUC value of more than 0.8368. Similarly, the AUC value of the traditional logistic regression model was more than 0.86 on the basis of 750,000 cases with 200 million data points in a study by Tomasev et al.[11] We agree with Tomasev et al.[11] that "data may be more important than algorithms", and the differences in model performance may be due to the structure of the data, the amount of data, the characteristics of the variables and the adaptability of the algorithms, e.g., deep neural networks are more suitable with nontabular data (images, audio, text)[22]. However, XGBoost performs better with tabular data[23]. In previous studies, one algorithm was often used for modeling and analysis, which is effective in reducing the computational workload; however, with this approach, the advantages of other algorithms in information mining and data learning may be neglected. In this study, we used five machine learning algorithms and considered various evaluation metrics to compare the effectiveness of different model parameters and selected the algorithm that achieved the best prediction effect and stability to construct the prediction model. The results suggest that the XGBoost model has the best prediction accuracy and efficiency, which is consistent with the results of several previous studies, possibly because this algorithm is more adaptable to the combination of categorical and continuous variables in the dataset[24]. Additionally, it is worth noting that in the external validation cohort, the model did not demonstrate high performance consistent with the internal validation cohort. This may be related to the characteristics of this population, as the diagnosis of tumors, infections, and hematological diseases is predominant in this group. However, infection related variables represented by antibiotic orders did not show high predictive importance during the modeling process. Therefore, for the external validation cohort, some important information was inevitably wasted.

Sixth, we introduced the DCA curve to evaluate the clinical practicability of the model. This method has been widely adopted in recent research and is considered to have high practicality for model evaluation[25,26]. The results show that within the threshold range of 5–90% of the model can be set, after the model outputs the results, appropriate clinical measures, including timely and effective intervention or continued observation, can produce significant clinical benefits for patients with predicted AKI or non-AKI. This means that our model is helpful in the clinical treatment process, and the support for clinical decision is scientific. In addition, we used propensity matching in our modeling to address the issue of imbalanced positive and negative samples in the data. The external validation data is real data, and the model results can also demonstrate our model's generalization ability, stability, or robustness.

Limitations of this study

This study had several limitations. First, in this study cohort, some patients did not have their baseline SCr levels measured, and standard creatinine values were used as a substitute. In fact, owing to preadmission renal injury, the admission SCr value may have been pathologically elevated; therefore, the use of the standard creatinine value as a baseline reference means that the actual incidence of AKI may have been underestimated. Second, a history of previous AKI was not included in the comorbidity variables; some studies have shown that a history of AKI may be associated with the recurrence of AKI. Third, urine volume criteria were not adopted in the diagnosis of AKI. Although some studies have shown that the addition of urine volume criteria improves the rate of AKI diagnosis while reducing diagnostic delays, it is difficult to monitor the urine volume in clinical practice except in the monitoring room. Fourth, for the external validation, only data from one medical center were included, which may have resulted in insufficient data to study the applicability and generalizability of the model. Furthermore, the volume of urine is affected by the amount of fluid infused and the use of diuretics, which may interfere with diagnoses for patients who already had oliguria but exhibited elevated SCr levels.

Conclusion

In this study, we developed an AKI prediction model by using machine learning algorithms considering many variables, including demographic data, laboratory test results, previous comorbidities, special medical orders and invasive operations, such as surgery. This model is applicable to a wide range of patients, is interpretable, has better predictive efficacy than previous models, and is capable of predicting the occurrence and risk of AKI up to 48 h earlier than when only SCr levels are used. In the future, more prediction models will be developed in the context of smart health care, and the use of AI technology will provide more opportunities for updating decision-making models in medical contexts. In the future, we will continue to optimize the parameters of the prediction model and focus on promoting its applicability, especially in areas that are underdeveloped in terms of medical technology, to provide greater technical support and thereby improve the AKI prognosis for the target population of patients. Additionally, we will actively explore and expand new applications of AI technology and machine learning algorithms in AKI management.

Methods

Study population

We screened selected patients admitted to the First Medical Center of the Chinese People's Liberation Army (PLA) General Hospital between January 1, 2011, and December 31, 2019. In this retrospective study, some patients admitted to the Fifth Medical Center of the PLA General Hospital during the same period were screened for inclusion in the external validation study. The inclusion criteria were as follows: (i) age > 18 years; (ii) hospitalization time > 48 h; (iii) complete case data with < 10% missing information; and iv) ≥ 2

blood creatinine measurements during hospitalization. The exclusion criteria were as follows: (i) creatinine value > 353.5 μmol/L measured for the first time after admission (to prevent inclusion of patients with impaired renal function at baseline); (ii) patients who were diagnosed with AKI prior to hospital admission (admission diagnosis or outpatient diagnosis of AKI); and (iii) patients with maintenance HD, peritoneal dialysis, or renal transplantation.

The study protocol was approved by the institutional review board of the General Hospital of the People's Liberation Army (PLA) (No. S2022-260-01). Informed consent was waived because data analyses were performed retrospectively using anonymized data from the General Hospital of the People's Liberation Army (PLA) database. All procedures in this study were performed in accordance with the Declaration of Helsinki.

Data collection

We relied on the Big Data Center of the PLA General Hospital for data extraction, and the extracted predictor variables were selected on the basis of previously published literature and clinical experience and reviewed by a group of clinicians and data engineers to narrow the dimensions of the available variables, determine the validity of the variables and assess the reliability of the collected data. The data included baseline patient information, underlying conditions prior to hospital admission, comorbidities, vital signs, laboratory test data, medication orders, and other specific interventions (Table S2), and the data were automatically calibrated in the system on the basis of the collection time.

Definition of AKI

AKI was defined according to the 2020 KDIGO Clinical Practice Guidelines as follows: blood creatinine ≥ 26.5 μmol/L (0.3 mg/dl) within 48 h, blood creatinine ≥ 1.5 times the basal level within 7 d, or urine output < 0.5 mL/kg/h for 6 h[27]. Since the majority of cases in this study could not be traced back to valid prehospital data to determine the baseline creatinine value, AKI was defined by utilizing the expanded AKI judgment criteria adopted in previous studies[28]. To define AKI, the blood creatinine values of hospitalized patients were sorted chronologically, and the time point of any blood creatinine test for the patient was set as t. The mean value of the blood creatinine in the week before the time point was used as the baseline creatinine value, and any creatinine value in the week after t was compared with the selected baseline creatinine value. The earliest date on which the change in the SCr value met the KDIGO criteria was defined as the date of AKI onset. Urine volume measurements were excluded from this study because of their uncertainty.

Variable preprocessing

Pearson's correlation test between variables was performed to initially screen and eliminate variables that were highly correlated with each other and much less correlated with the predicted outcomes[29]. The continuous variables in the database were subsequently standardized, i.e., extreme values outside the 1–99 percentile were removed to exclude the interference of extreme values due to entry errors, and then the data were normalized to form a new assignment cohort with statistics close to the standard normal distribution[30]. Similarly, the categorical variables were converted into binary variables on the basis of domain knowledge, i.e., "yes" or "no".

We eliminate variables and samples with a missing rate of more than 20% and use the random Forest chain equation (MissForest) interpolation method to fill in the missing values. MissForest makes use of the nonlinear fitting ability of random forest algorithm and the advantages of ensemble learning, and iteratively trains the model to predict and fill in missing data to maintain data integrity and availability.

Model training

In the process of modeling based on the database collected from patients at the First Medical Center of the PLA General Hospital, we used a downsampling method to randomly select some patients from the non-AKI group so that the ratios of both samples to the number of cases were maintained at 1:1. Since we had already limited the missing rate of the extracted data to be no more than 10%, substantial information was not missing for any patient, and we prevented effects due to randomness and data representativeness through various types of padding.

The downsampled processed data were randomly divided into a training group (70%) and a validation group (30%), and the baseline characteristics of the two groups were compared. For the modeling analysis, we fixed the time window for data collection, setting the upper and lower thresholds of the window as the day of admission and two days before the occurrence of AKI, respectively; such data included all static variables and all real-time dynamic data including in the modeling process. The values of the repeated-measures dynamic variables within this time window were adopted as the highest or lowest, and the static variables were converted to binary variables, i.e., "yes (1)" or "no (0)". For model validation, to illustrate the changes in longitudinal data over time and to evaluate the prediction capability of the model, we introduced a discrete-time analysis model, i.e., we set 12 h as the time frame, considered the variables in the previous time frame and dynamically updated and added the latest variable information to predict the likelihood of the occurrence of AKI in the next 48 h (four time windows). This approach allowed analysis models to be created using binary classifiers, and missing values for predictor variables such as vital signs or laboratory indicators within the time frame were filled in using a carry-over interpolation approach, i.e., using the most recently available measurements within the previous time frame (Fig. S1).

Five machine learning methods were used to build models, and the training group was further randomly divided into ten equal-sized groups for tenfold cross-validation, which reduced overfitting to a certain extent, allowed as much valid information as possible to be obtained from the limited data, and provided a more stable and reliable method for evaluating the model's performance during the model development process[31]. In

addition, the recursive feature elimination in random forests (RF-RFE) method was used to identify important variables; these variables and the best algorithm were then used to build the final model.

Model performance evaluation and interpretation

As recommended by the TRIPOD Statement of Methodological Guidelines for Clinical Predictive Modeling[32], the evaluation and validation of predictive models require simultaneous reporting of model discrimination, calibration, and clinical utility metrics to assess all three aspects of predictive modeling on the basis of the combined dependent and independent variables. To evaluate model discriminability, the area under the curve (AUC), sensitivity, specificity, accuracy, F1 score and Mathews correlation coefficient (MCC) were employed; to evaluate model calibration, calibration curves and Brier scores were used; and to evaluate clinical utility; decision curve analysis (DCA) was used.

In this study, we constructed SHAP importance matrices to visualize the importance of each predictor variable in the prediction model, SHAP summary plots to show the impact of changes in the values of each variable on the overall model prediction results, and SHAP dependency plots to represent the impact of changes in the values of individual variables on the prediction results[33,34].

Data availability

The data that support the findings of this study are available from HMedical Innovation Research Division, Chinese PLA General Hospital, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the corresponding author (Zhe Feng) upon reasonable request and with permission of Medical Innovation Research Division.

Received: 6 November 2024; Accepted: 13 February 2025
Published online: 18 March 2025

References

1. Turgut, F., Awad, A. S. & Abdel-Rahman, E. M. Acute kidney injury: medical causes and pathogenesis. *J. Clin. Med.* **12**, 375 (2023).
2. Kellum, J. A. et al. Acute kidney injury. *Nat. Rev. Dis. Primers* **7**, 52 (2021).
3. Ostermann, M., Liu, K. & Kashani, K. Fluid management in acute kidney injury. *Chest* **156**, 594–603 (2019).
4. Bellomo, R., Kellum, J. A. & Ronco, C. Acute kidney injury. *Lancet* **380**, 756–766 (2012).
5. Stanski, N. L. et al. Precision management of acute kidney injury in the intensive care unit: current state of the art. *Intensive Care Med.* **49**, 1049–1061 (2023).
6. Tai, Q. et al. The accuracy of urinary TIMP-2 and IGFBP7 for the diagnosis of cardiac surgery-associated acute kidney injury: a systematic review and meta-analysis. *J. Intensive Care Med.* **35**, 1013–1025 (2020).
7. Liu, C. et al. The diagnostic accuracy of urinary [TIMP-2]·[IGFBP7] for acute kidney injury in adults: a PRISMA-compliant meta-analysis. *Medicine (Baltimore)* **96**, 27 (2017).
8. Ibrahim, N. E. et al. A clinical, proteomics, and artificial intelligence-driven model to predict acute kidney injury in patients undergoing coronary angiography. *Clin. Cardiol.* **42**, 292–298 (2019).
9. Lee, H. C. et al. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. *J. Clin. Med.* **7**, 322 (2018).
10. Tseng, P. Y. et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit. Care* **25**, 478 (2021).
11. Tomašev, N. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* **572**, 116–119 (2019).
12. Sun, M. et al. Early prediction of acute kidney injury in critical care setting using clinical notes and structured multivariate physiological measurements. *Stud. Health Technol. Inf.* **264**, 368–372 (2019).
13. Alfieri, F. et al. A deep-learning model to continuously predict severe acute kidney injury based on urine output changes in critically ill patients. *J. Nephrol.* **34**, 1875–1886 (2021).
14. Zhang, X. et al. Machine learning for the prediction of acute kidney injury in critical care patients with acute cerebrovascular disease. *Ren. Fail.* **44**, 43–53 (2022).
15. Li, Y. et al. Early prediction of acute kidney injury in critical care setting using clinical notes. *Proc. (IEEE Int. Conf. Bioinf. Biomed.)* **2018**, 683–686 (2018).
16. Mehran, R. et al. A contemporary simple risk score for prediction of contrast-associated acute kidney injury after percutaneous coronary intervention: derivation and validation from an observational registry. *Lancet* **398**, 1974–1983 (2021).
17. Kheterpal, S. et al. Development and validation of an acute kidney injury risk index for patients undergoing general surgery: results from a national data set. *Anesthesiology* **110**, 505–515 (2009).
18. Cheng, P., Waitman, L. R., Hu, Y. & Liu, M. Predicting inpatient acute kidney injury over different time horizons: how early and accurate? *AMIA Annu. Symp. Proc.* 565–574 (2017). (2017).
19. He, J. et al. Multi-perspective predictive modeling for acute kidney injury in general hospital populations using electronic medical records. *JAMIA Open* **2**, 115–122 (2019).
20. 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.* **94**, 783–792 (2019).
21. Kriegeskorte, N. & Golan, T. Neural network models and deep learning. *Curr. Biol.* **29**, R231–R236 (2019).
22. Song, X. et al. XGBoost-based feature learning method for mining COVID-19 novel diagnostic markers. *Front. Public Health* **10**, 926069 (2022).
23. Park, J., Kim, J., Ryu, D. & Choi, H. Y. Factors related to steroid treatment responsiveness in thyroid eye disease patients and application of SHAP for feature analysis with XGBoost. *Front. Endocrinol. (Lausanne)* **14**, 1079628 (2023).
24. Ostermann, M. et al. Controversies in acute kidney injury: conclusions from a kidney disease: improving global outcomes (KDIGO) conference. *Kidney Int.* **98**, 294–309 (2020).
25. Fan, Z. et al. Construction and validation of prognostic models in critically ill patients with sepsis-associated acute kidney injury: interpretable machine learning approach. *J. Transl. Med.* **21**, 406 (2023).
26. Xinsai, L. et al. Prediction model of acute kidney injury after different types of acute aortic dissection based on machine learning. *Front. Cardiovasc. Med.* **9**, 984772 (2022).
26. Ostermann, M. et al. Controversies in acute kidney injury: conclusions from a kidney disease: Improving global outcomes (KDIGO) conference. *Kidney Int.* **98**, 294–309 (2020).

28. Xu, X. et al. Epidemiology and clinical correlates of AKI in Chinese hospitalized adults. *Clin. J. Am. Soc. Nephrol.* **10**, 1510–1518 (2015).
29. Pripp, A. H. Pearson's or Spearman's correlation coefficients. *Tidsskr. Laegeforen.* **138** https://doi.org/10.4045/tidsskr.18.0042 (2018).
30. Demircioğlu, A. The effect of feature normalization methods in radiomics. *Insights Imaging* **15**, 2 (2024).
31. Kernbach, J. M. & Staartjes, V. E. Foundations of machine learning-based clinical prediction modeling: part II-generalization and overfitting. *Acta Neurochir. Suppl.* **134**, 15–21 (2022).
32. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, g7594 (2015).
33. Dickinson, Q., Meyer, J. G. & Positional SHAP (PoSHAP) for interpretation of machine learning models trained from biological sequences. *PLoS Comput. Biol.* **18**, e1009736 (2022).
34. Yang, Y., Yuan, Y., Han, Z. & Liu, G. Interpretability analysis for thermal sensation machine learning models: an exploration based on the SHAP approach. *Indoor Air* **32**, e12984 (2022).

Acknowledgements

We thank all authors for their participation in this study and for their kind assistance in collecting data and samples.

Author contributions

Conceptualization, X.Y. and Z.F.; methodology, X.Y. and Z.F.; software, WL.W and RLG.W.validation, X.Y., Z.F; formal analysis, X.Y.; investigation, Z.F.; resources, X.Y. and WL.W.; data curation, X.Y., Z.F. and WL.W.; writing—original draft preparation, X.Y and XY.G.; writing—review and editing, X.Y and YW.J.; visualization, X.Y.; supervision, Z.F.; project administration, Z.F.;funding acquisition, Z.F.; X.Y. and WL.W. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Natural Science Foundation of China (82170686) and the Grant for GYC (22KJLJ001).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-90459-5.

Correspondence and requests for materials should be addressed to Z.F.

Reprints and permissions information is available at www.nature.com/reprints.

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