European Journal of Medical Research

RESEARCH

Open Access

Construction and validation of risk prediction models for renal replacement therapy in patients with acute pancreatitis

Fei Zuo[1,4], Lei Zhong[2,4], Jie Min[2,4], Jinyu Zhang[3,4] and Longping Yao[2,4*]

Abstract

Background Renal replacement therapy (RRT) plays a crucial role in managing acute pancreatitis (AP). This study aimed to develop and evaluate predictive models for determining the need for RRT among patients with AP in the intensive care unit (ICU).

Methods A retrospective selection of patients with AP was made from the Medical Information Mart for Intensive Care IV (MIMIC-IV, version V2.0). The cohort was randomly divided into a training set (447 patients) and a validation set (150 patients). The least absolute shrinkage and selection operator (LASSO) regression cross-validation method was utilized to identify key features for model construction. Using these features, four machine learning (ML) algorithms were developed. The optimal model was visualized and clarified using SHapley Additive exPlanations (SHAP) and presented as a nomogram.

Results The mean age of the cohort was 59.17 years, with an average Acute Physiology and Chronic Health Evaluation II (APACHE II) score of 17.55. Acute kidney injury (AKI) was observed in 52.43% of patients with AP, and 9.05% required RRT. After feature selection, four of 41 clinical factors were ultimately chosen for use in model construction. The Lasso-Logistic Regression (Lasso-LR) model showed a high discriminative ability to predict RRT risk in patients with AP, with an area under the receiver operating characteristic (AUROC) of 0.955 (95% CI 0.924–0.987) in the training set. In the validation set, it maintained its discriminative performance, achieving an AUROC of 0.985 (95% CI 0.970–1.000). Calibration curves indicated an excellent fit in both sets (Brier scores: 0.039 and 0.032, respectively), suggesting high consistency. Decision curve analysis (DCA) highlighted the Lasso-LR model's significant clinical utility in predicting RRT likelihood in patients with AP.

Conclusions Developed via the LASSO regression cross-validation method, the Lasso-LR model significantly excels in predicting the requirement for RRT in patients with AP, demonstrating its potential for clinical application.

Keywords Acute pancreatitis, Renal replacement therapy, Intensive care unit, Predictive model, MIMIC-IV database

*Correspondence:
Longping Yao
zsmj19901004@163.com
Full list of author information is available at the end of the article

Introduction

Acute pancreatitis (AP) is a frequently encountered clinical emergency, marked by the premature activation of pancreatic enzymes, self-digestion of the pancreas, and the ensuing local and systemic inflammatory responses [1]. Over the past few decades, the global incidence rate of AP has varied between 4.9 to 73.9 cases per 100,000 individuals each year [2, 3], with the annual incidence demonstrating a rising trend [4, 5]. Although most patients with AP have a favorable prognosis, about 20–30% of them evolve into severe acute pancreatitis (SAP) [6], which carries a mortality rate of 15–30% [1, 7]. The kidneys, being among the organs most frequently compromised in SAP, often indicate an escalation in the severity of the condition and a decline in prognosis [8]. The early liberation of a substantial volume of inflammatory mediators in patients with AP can instigate a systemic inflammatory response syndrome and the infiltration of vital organs by inflammatory cells, thereby worsening organ damage and potentially leading to multiple organ dysfunction syndrome (MODS) [9].

To date, a myriad of therapeutic strategies targeting the pathogenesis of AP have been established [10]. In addition to pharmacotherapy, surgical, and interventional therapies, renal replacement therapy (RRT) plays an integral role in the treatment regimen. RRT has progressed beyond mere kidney function substitution, expanding its role to multiple organ support. It may confer benefits by regulating the internal environment and electrolyte balance, reducing toxins and inflammatory mediators, decreasing intra-abdominal pressure, mitigating edema, and thereby further protecting organ function, all of which are advantageous in treating patients with AP [11]. Despite these advancements, no unified international guidelines or consensus exist regarding the optimal timing and methodology for implementing RRT in patients with AP. Clinical decisions are often tailored to the patient's unique circumstances, necessitating an assessment of vital signs, laboratory test results, organ function, and overall prognosis. Within this framework, effective machine learning (ML) predictive models can offer substantial decision-making support to clinicians [3]. Nevertheless, the current predictive models for AP largely focus on assessing severity [12, 13], mortality [14], and complications [15, 16], as well as determining the appropriate timing for surgery [17]. There remains a significant gap in research regarding clinical risk prediction models specifically for RRT in AP patients.

To bridge this knowledge gap, we performed a retrospective analysis using the Medical Information Mart for Intensive Care IV (MIMIC IV, version V2.0) database to explore the risk factors that influence the requirement for RRT in patients with AP. Specifically, we employed four ML algorithms to create and validate early warning models for AP patients admitted to the intensive care unit (ICU) who might need RRT. Additionally, we developed and validated a user-friendly prognostic nomogram. The objective of this model is to lay the groundwork for timely RRT interventions for these patients, thus aiding physicians in making more informed RRT medical decisions while considering the advantages, disadvantages, and the scarcity of medical resources.

Methods

Data source

The study's data were obtained from the MIMIC-IV database, developed by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology (MIT). This open-access and free resource encompasses de-identified clinical information of patients admitted to the ICU at Beth Israel Deaconess Medical Center (BIDMC) from 2008 to 2019 [18]. The database utilization for this research has received approval from the Institutional Review Boards of BIDMC and MIT. Three of the authors have successfully completed the CITI Program's certification process, granting them access to the database. A rigorous de-identification process of patient information within the MIMIC database ensures patient privacy, precluding the need for additional informed consent. This investigation was conducted as a retrospective analysis and was aligned with the principles of the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [19].

Study population

The inclusion criteria for this study were as follows: (1) Patients with a primary diagnosis of AP in the MIMIC-IV database according to the International Classification of Diseases, Ninth Edition (ICD-9 code 577.0) and Tenth Edition (ICD-10 code K85%). (2) For patients with multiple admissions to the ICU, only the first admission record is considered. The exclusion criteria included: (1) age < 18 years; (2) patients not admitted to the ICU; (3) patients with a clinical data missing rate exceeding 20%. Ultimately, a total of 597 patients diagnosed with AP were included. The study cohort was subsequently divided, in a random manner, into a training set of 447 patients and a validation set of 150 patients, maintaining a training-to-validation ratio of 3:1.

Variable extraction

This investigation employed PostgreSQL software (version V14.5-1) and Navicat Premium 15 to retrieve and process data from the MIMIC-IV database. The dataset covered demographic details, disease severity scores,

laboratory results, comorbid conditions, and therapeutic intervention records. Our predictive model focused exclusively on clinical and lab data captured within the first 24 h of ICU admission, identifying comorbidities with ICD-9/10 coding.

For data extraction, we aligned with established research, prioritizing clinical relevance and data availability. Demographic information such as age and sex at the time of hospital admission, vital signs including mean arterial pressure (MAP), and severity scores from the Sequential Organ Failure Assessment (SOFA) and the Acute Physiology and Chronic Health Evaluation II (APACHE II) were collated. Laboratory data included white blood cell (WBC) count, hemoglobin levels, platelet count, red cell distribution width (RDW), and concentrations of alanine aminotransferase (ALT), aspartate aminotransferase (AST), bilirubin, anion gap, bicarbonate, glucose, creatinine, blood urea nitrogen (BUN), prothrombin time, and electrolytes such as calcium, sodium, chloride, potassium, magnesium, and phosphorus.

We also documented comorbidities including obesity, hypertension, diabetes, cerebral infarction, cirrhosis, atrial fibrillation, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), malignant tumor, septic shock, acute kidney injury (AKI), chronic kidney disease (CKD), depression, delirium, and oliguria. Therapeutic measures in the dataset involved the administration of norepinephrine, the application of mechanical ventilation (MV), and RRT.

Statistical analysis

In this study, data management and analysis were performed using Stata software (version 14.0) and R (version 4.2.3), setting the threshold for statistical significance at $P < 0.05$. For variables with less than 20% missing data, multiple imputation was performed using the 'mice' package in the R programming language to estimate the missing values. Continuous variables conforming to a normal distribution were presented as the mean ± standard deviation (SD) and were compared using the independent samples t-test. Continuous variables not adhering to a normal distribution were described by the median and interquartile range and were analyzed using the Mann–Whitney U test. Categorical variables were represented by frequency and percentage and were examined for differences using the Chi-square test or Fisher's exact test, as appropriate.

To reduce the potential for overfitting, the Least Absolute Shrinkage and Selection Operator (LASSO) regression model, in conjunction with tenfold cross-validation, was employed for key feature selection in this study. Additionally, fivefold cross-validation and grid search techniques were combined to refine hyperparameters within the training dataset, thus improving the model's assessment performance. Following the selection of features, four distinct prediction models were constructed for subsequent analysis and validation: Lasso-Logistic Regression (Lasso-LR) model, Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM).

The validation dataset was utilized to conduct a comprehensive assessment of the predictive models. The discriminative ability was quantified by the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUROC). Model accuracy was determined through calibration curve analysis, and clinical utility was evaluated by decision curve analysis (DCA). Among the models, the Lasso-LR model demonstrated superior performance, integrating these metrics. SHapley Additive exPlanations (SHAP) were subsequently employed to enhance model interpretability. The findings of the Lasso-LR model were finally illustrated through a nomogram, designed to aid comprehension and application by clinical decision-makers.

Results

Patient characteristics

Figure 1 presents the composition of the study cohort, comprising 597 patients with AP, stratified into a training set (447 individuals) and a validation set (150 individuals). The cohort had a mean age of 59.17 years (SD = 17.94) and an average APACHE II score of 17.55 (SD = 8.05). AKI was present in 52.43% of the patients (313 individuals), and RRT was necessitated in 9.05% (54 individuals). Demographic and clinical variables between the training and validation sets were compared, as demonstrated in Table 1. Apart from a pronounced incidence of delirium in the validation group, the comparison revealed no significant differences in other variables between the groups.

Feature selection

Employing the glmnet package in R, a LASSO regression analysis with tenfold cross-validation was utilized to predict the necessity for RRT in patients with AP, considering 41 distinct features. As depicted in Fig. 2, the optimal lambda (λ) value of 0.055 was determined, leading to the identification of four principal predictive factors: the SOFA score, serum creatinine levels, the presence of oliguria on the first day, and the administration of norepinephrine during hospitalization.

Models' construction and validation

The four identified features were incorporated into various predictive models, including Lasso-LR, RF, XGBoost, and SVM. A fivefold cross-validation approach was

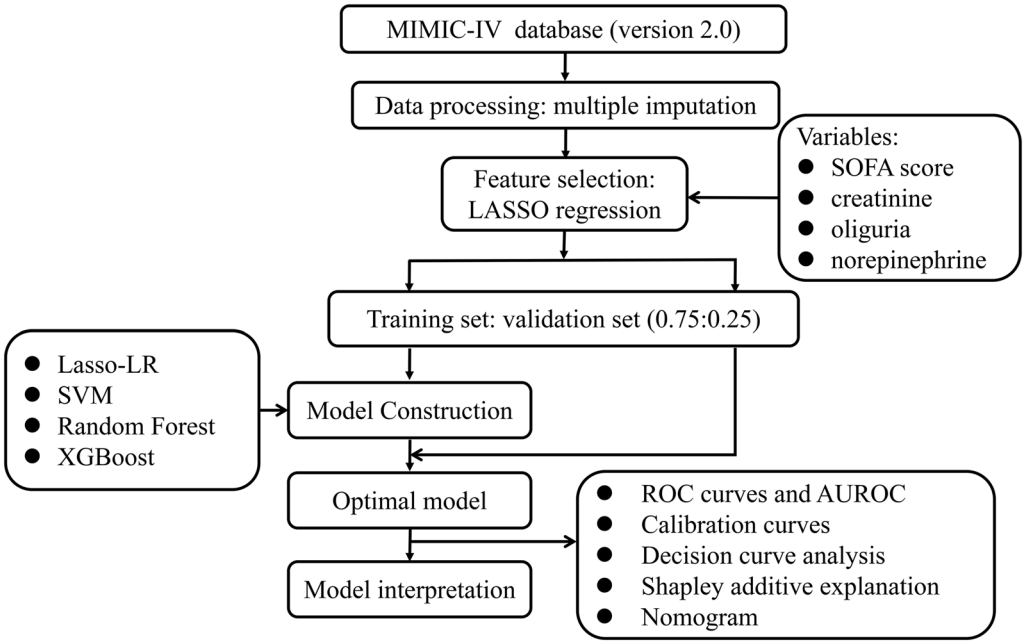

Fig. 1 Flow diagram of this study. *MIMIC-IV* Medical Information Mart for Intensive Care IV, *ROC* Receiver Operating Characteristic, *AUROC* Area Under the Receiver Operating Characteristic Curve, *Lasso-LR* Lasso-Logistic Regression, *XGBoost* eXtreme Gradient Boosting, *SVM* Support Vector Machine

utilized for parameter refinement, and the models underwent several iterations of training to ascertain the optimal configuration. The detailed performance metrics of the machine learning models across the training and validation sets are presented in Table 2.

All models demonstrated good AUROC performance, particularly in the training set, where both the Lasso-LR and SVM achieved an AUROC value of 0.955, indicating similar performance. In terms of sensitivity, all models performed comparably in the training set, with the RF model exhibiting the highest sensitivity at 90%, while the other models had a sensitivity of 87.5%. Regarding specificity, Lasso-LR and SVM achieved the highest specificity at 95.3% and 94.8%, respectively, reflecting their efficiency in identifying negative samples. In terms of accuracy, Lasso-LR and SVM reached accuracy rates of 94.6% and 94.2%, respectively, showcasing strong predictive capability.

In the validation set, the Lasso-LR model continued to exhibit excellent performance, achieving an AUROC value of 0.985 (95% CI: 0.970–1.000), as shown in Fig. 3. The sensitivity for Lasso-LR was 85.7%, specificity was 95.6%, accuracy was 94.7%, and the Kappa value was 0.721, indicating good classification consistency. The SVM's performance was similar to that of Lasso-LR, with an AUROC value of 0.984 (95% CI: 0.967–1.000), maintaining high levels of sensitivity and specificity. The RF model demonstrated a sensitivity of 100% in the validation set, indicating it captured all positive samples; however, its specificity was slightly lower at 93.4%. The XGBoost model had an AUROC value of 0.964 (95% CI: 0.935–0.993), with both sensitivity and specificity at 85.7% and 95.6%, respectively. All models maintained accuracy rates above 94% in the validation set, indicating their good generalization ability on unseen data.

The calibration curve in Supplementary Fig. 1 confirmed the predictive accuracy of the models, with Lasso-LR showing superior calibration accuracy and a Brier score of 0.032. The DCA presented in Supplementary Fig. 2 indicated that, except for the XGBoost model, all other models demonstrated strong clinical utility.

In summary, the Lasso-LR model exhibited superior performance in AUROC, sensitivity, specificity, and accuracy, alongside favorable results in its calibration curve and DCA. Therefore, Lasso-LR was identified as the best predictive model.

Model interpretations

The SHAP method provides a comprehensive and transparent approach for interpreting predictive models, ensuring the accurate and consistent quantification of each feature's contribution to prediction outcomes as SHAP values. Higher SHAP values indicate a stronger positive correlation with the likelihood of requiring RRT. Figure 4 illustrates the impacts of four distinct features on prediction outcomes within the Lasso-LR model. The

Table 1 The baseline characteristics of acute pancreatitis

Characteristic	Total (n = 597)	Training set (n = 447)	Validation set (n = 150)	$t/Z/\chi^2$	*P* value
Age (years)	59.17 ± 17.94	59.40 ± 17.80	58.47 ± 18.37	0.550	0.582
Gender, n (%)					
Female	238 (39.87)	177 (39.60)	61 (40.67)	0.054	0.817
Male	359 (60.13)	270 (60.40)	89 (59.33)		
SOFA score	4.00 (2.00, 7.00)	4.00 (2.00, 7.00)	4.00 (2.00, 7.00)	0.114	0.909
APACHE II score	17.55 ± 8.05	17.42 ± 7.86	17.95 ± 8.60	− 0.698	0.485
MAP (mmHg)	89.72 ± 19.76	89.65 ± 19.37	89.93 ± 20.93	− 0.148	0.882
Laboratory tests					
WBC ($\times 10^9$/L)	12.60 (8.80, 17.60)	12.60 (8.80, 17.40)	12.65 (8.90 18.50)	− 0.502	0.616
Hemoglobin (g/L)	119.68 ± 25.56	119.74 ± 25.79	119.50 ± 24.95	0.100	0.921
Platelet ($\times 10^9$/L)	207.00 (151.00, 289.00)	205.00 (148.00, 282.00)	210.00 (156.00, 307.00)	− 0.722	0.471
RDW (%)	14.73 ± 1.83	14.67 ± 1.74	14.90 ± 2.07	− 1.349	0.178
ALT (U/L)	54.00 (25.00, 158.00)	50.00 (25.00, 151.00)	62.00 (27.00, 174.00)	− 0.855	0.393
AST (U/L)	71.00 (35.00, 170.00)	76.00 (34.00, 177.00)	63.00 (36.00, 146.00)	0.548	0.584
Bilirubin (umol/L)	17.10 (10.26, 41.04)	15.39 (8.55, 41.04)	18.81 (10.26, 32.75)	− 0.996	0.319
Anion gap (mmol/L)	16.45 ± 5.79	16.54 ± 5.85	16.16 ± 5.65	0.697	0.486
Bicarbonate (mmol/L)	21.55 ± 5.30	21.53 ± 5.56	21.61 ± 4.45	− 0.153	0.879
Glucose (mmol/L)	7.17(5.72, 9.83)	7.22 (5.72, 9.83)	7.03 (5.50, 9.89)	0.442	0.659
Creatinine (umol/L)	88.40 (61.88, 132.60)	88.40 (61.88, 132.60)	79.56 (61.88, 123.76)	1.287	0.198
BUN (mmol/L)	6.05 (3.92, 9.97)	6.05 (3.92, 10.32)	6.05 (3.56, 9.97)	0.655	0.513
Prothrombin time (s)	13.70 (12.50, 15.30)	13.80 (12.60, 15.40)	13.30 (12.30, 15.10)	1.596	0.111
Calcium (mmol/L)	1.99 ± 0.27	1.98 ± 0.29	2.02 ± 0.21	− 1.479	0.140
Sodium (mmol/L)	137.84 ± 5.34	137.92 ± 5.56	137.60 ± 4.64	0.633	0.527
Chlorine (mmol/L)	103.17 ± 7.18	103.16 ± 7.30	103.21 ± 6.84	− 0.084	0.933
Potassium (mmol/L)	4.20 ± 0.85	4.19 ± 0.84	4.26 ± 0.89	− 0.865	0.388
Magnesium (mmol/L)	0.76 ± 0.17	0.77 ± 0.17	0.76 ± 0.17	0.560	0.575
Phosphate (mmol/L)	1.05 ± 0.47	1.03 ± 0.45	1.10 ± 0.53	− 1.688	0.092
Comorbidities, n (%)					
Obesity	171 (28.64)	125 (27.96)	46 (30.67)	0.401	0.526
Hypertension	294 (49.25)	220 (49.22)	74 (49.33)	0.001	0.980
Diabetes	173 (28.98)	125 (27.96)	48 (32.00)	0.889	0.346
Cerebral infarction	32 (5.36)	21 (4.70)	11 (7.33)	1.538	0.215
Cirrhosis	34 (5.70)	24 (5.37)	10 (6.67)	0.352	0.553
Atrial fibrillation	99 (16.58)	78 (17.45)	21 (14.00)	0.966	0.326
CHF	80 (13.40)	58 (12.98)	22 (14.67)	0.277	0.599
COPD	48 (8.04)	35 (7.83)	13 (8.67)	0.106	0.744
Malignant tumor	37 (6.20)	25 (5.59)	12 (8.00)	1.119	0.290
AKI	313 (52.43)	230 (51.45)	83 (55.33)	0.678	0.410
Septic shock	62 (10.39)	48 (10.74)	14 (9.33)	0.238	0.626
CKD	75 (12.56)	56 (12.53)	19 (12.67)	0.002	0.965
Depression	91 (15.24)	64 (14.32)	27 (18.00)	1.179	0.278
Delirium	60 (10.05)	37 (8.28)	23 (15.33)	6.185	0.013
Oliguria	54 (9.05)	40 (8.95)	14 (9.33)	0.020	0.887
Treatment, n (%)					
Norepinephrine	94 (15.75)	77 (17.23)	17 (11.33)	2.940	0.086
Mechanical ventilation	242 (40.54)	188 (42.06)	54 (36.00)	1.710	0.191
RRT	54 (9.05)	40 (8.95)	14 (9.33)	0.020	0.887

SOFA Sequential Organ Failure Assessment, *APACHE II* Acute Physiology and Chronic Health Evaluation II, *MAP* Mean Arterial Pressure, *WBC* White Blood Cell, *RDW* Red Cell Distribution Width, *ALT* Alanine Aminotransferase, *AST* Aspartate Aminotransferase, *BUN* Blood Urea Nitrogen, *CHF* Congestive Heart Failure, *COPD* Chronic Obstructive Pulmonary Disease, *AKI* Acute Kidney Injury, *CKD* Chronic Kidney Disease, *RRT* Renal Replacement Therapy

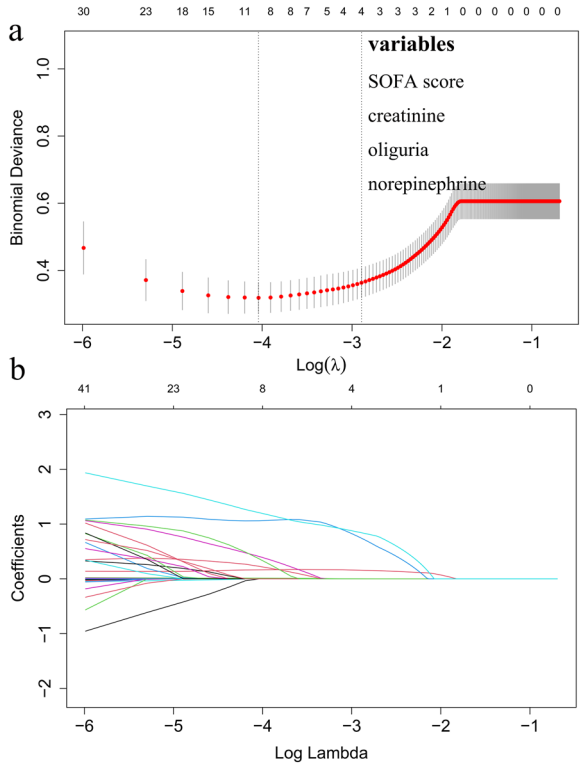

Fig. 2 Clinical features were selected based on LASSO regression with cross-validation. **a** Turing parameter (λ) selection in the LASSO model using tenfold cross-validation; **b** LASSO coefficient profiles of the 41 candidate variables. *LASSO* Least Absolute Shrinkage and Selection Operator, *SOFA* Sequential Organ Failure Assessment

SOFA score, serum creatinine levels, administration of norepinephrine, and the presence of oliguria on the first day of admission are shown to be positively correlated with SHAP values, indicating that increases in these factors are associated with a higher risk of RRT in patients with AP.

Evaluation of the optimal model

By presenting complex logistic regression or Cox regression models through nomograms, their results are simplified and visually displayed, thereby enhancing clinical utility. In this study, the Lasso-LR model was converted into an easily interpretable static nomogram using R language, as depicted in Fig. 5. Its performance was subsequently evaluated.

The Lasso-LR model demonstrated a strong discriminative ability to predict the risk of RRT in patients with AP, achieving an AUROC of 0.955 (95% CI: 0.924–0.987) in the training set. In the validation set, the model sustained its efficient discriminative capability, with an AUROC of 0.985 (95% CI: 0.970–1.000). For further details, please refer to Supplementary Fig. 3. The calibration curves for the Lasso-LR model revealed an excellent fit in both sets (Brier scores: 0.039 and 0.032, respectively), indicating high consistency. Additional information can be found in Supplementary Fig. 4. DCA results revealed that the model performed well when the threshold probability ranged from 0 to 88% in the training cohort, while the threshold probability range for the validation cohort extended from 0 to 100%. Within this range, using the model to predict the risk of RRT provided greater net benefits compared to the "treat all" or "treat none" strategies, suggesting significant clinical applicability. For more details, please refer to Supplementary Fig. 5.

Discussion

Renal Replacement Therapy (RRT) is pivotal in managing acute pancreatitis (AP); however, its efficacy hinges on timely and accurate execution. The complex demands of RRT's operation and management can lead not only to complications such as hemorrhage, thrombosis, and infections but also to treatment delays due to overly conservative approaches [11, 20]. Thus, the

Table 2 Performance comparison of four different algorithmic models in the training set and the validation set

Models	AUROC (95%CI)	Sensitivity (%)	Specificity (%)	Accuracy	Kappa
Training set					
Lasso-LR	0.955 (0.924–0.987)	0.875	0.953	0.946	0.715
Support vector machines	0.955 (0.922–0.988)	0.875	0.948	0.942	0.698
Random forest	0.944 (0.901–0.987)	0.900	0.924	0.922	0.632
XGBoost	0.939 (0.896–0.981)	0.875	0.914	0.911	0.590
Validation set					
Lasso-LR	0.985 (0.970–1.000)	0.857	0.956	0.947	0.721
Support vector machines	0.984 (0.967–1.000)	0.857	0.956	0.947	0.721
Random forest	0.977 (0.956–0.998)	1.000	0.934	0.940	0.725
XGBoost	0.964 (0.935–0.993)	0.857	0.956	0.947	0.721

AUROC Area Under the Receiver Operating Characteristic Curve, *CI* Confidence Interval, *Lasso-LR* Least Absolute Shrinkage and Selection Operator-Logistic Regression, *XGBoost* eXtreme Gradient Boosting

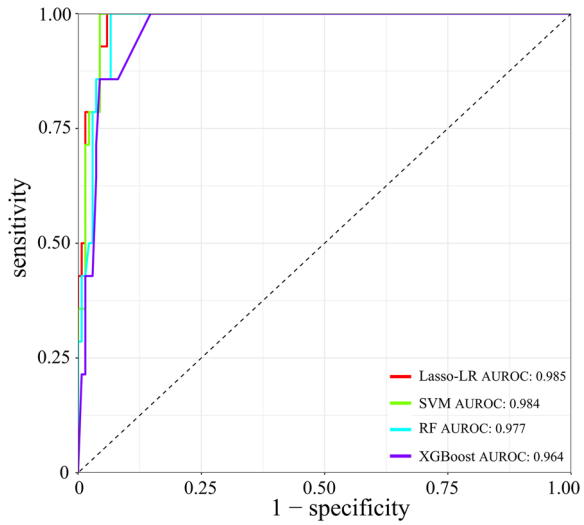

Fig. 3 ROC curves and AUROCs of four machine learning models in the validation set. *ROC* Receiver Operating Characteristic, *AUROC* Area Under the Receiver Operating Characteristic Curve, *Lasso-LR* Lasso-Logistic Regression, *SVM* Support Vector Machine, *RF* Random Forest, *XGBoost* eXtreme Gradient Boosting

accurate determination of the optimal timing for initiating RRT is crucial, especially as decision-making in non-emergency situations becomes more complex [21]. The application of ML algorithms presents a novel approach to enhancing medical decision-making by identifying key indicators and developing precise risk prediction models. This study utilized the MIMIC-IV database to develop several ML-based models for evaluating the RRT risk in patients with AP, incorporating critical parameters like the SOFA score, creatinine levels, urine output, and norepinephrine administration. The Lasso-LR model achieved the highest AUROC in both the training and validation cohorts, demonstrating an excellent balance between sensitivity and specificity. The sensitivity reached 85.7%, indicating that the model effectively identifies the majority of patients who genuinely require RRT, thus providing a basis for timely intervention. The specificity reached 95.6%, indicating that the model excels in excluding patients who do not require RRT, thereby reducing the likelihood of misdiagnosis and unnecessary medical interventions. The accuracy was 94.7%, further confirming the reliability of the Lasso-LR model in classifying AP patients as needing or not needing RRT. Additionally, the Lasso-LR model exhibits strong calibration and clinical relevance. Consequently, the Lasso-LR model has proven to be a reliable tool for identifying AP patients who may require RRT, thus providing invaluable assistance in clinical decision-making and significantly impacting the treatment and management of this patient population.

In our study, through the constructed nomogram models, it was found that patients with AP are more likely to require RRT when presenting with specific characteristics: high SOFA scores and creatinine levels, oliguria upon first-day admission, and norepinephrine need during hospitalization. The SOFA score, indicative of organ dysfunction severity, markedly heightens RRT necessity. Recent studies have shown that the SOFA score outperformed other indices like APACHE II, BISAP, and Ranson in predicting AP severity, ICU admission rates, and mortality, with higher statistical significance [22]. Additionally, studies have indicated that an increase in the SOFA score is independently associated with a surge in adverse outcomes [23], and serves as a crucial predictor for intra-abdominal hypertension (IAH) in AP cases [24]. IAH may aggravate AP-induced AKI by various mechanisms: it compromises renal blood flow, escalates inflammatory responses, and increases cytotoxic effects, thus raising the risk of RRT [25]. In our predictive model, the SOFA score was incorporated as a significant independent variable.

Our predictive model incorporates high creatinine levels and oliguria on the first day of admission, recognized indicators of AKI. Research has indicated that substantial changes in renal function occur early in AP, particularly in SAP [26]. Elevated serum creatinine levels often point to renal dysfunction, a result of inflammation, endotoxemia, and hemodynamic instability. In the ML-based predictive model for AP patients with AKI, the role of serum creatinine is especially critical [27, 28]. Moreover, creatinine levels at admission are not only independent predictors of poor outcomes in AP patients [29] but also correlate with an increased risk of surgical complications [30]. In our nomogram, oliguria, defined as a daily urine output of less than 400 mL, serves as a warning sign for initiating RRT, reflecting early damage to renal concentrating function [31]. Despite some limitations of the Kidney Disease Improving Global Outcomes (KDIGO) criteria based on serum creatinine and urine output [32], it remains irreplaceable in the clinical treatment and research of AKI until new guidelines are introduced. The significance of these parameters in predicting AKI is well-recognized in the field of ML. As demonstrated by Shao et al. [33], incorporating such parameters into ML models can markedly enhance predictive accuracy and clinical applicability. Elevated creatinine and oliguria necessitate a proactive approach to identify causes, investigate reversible factors, and implement targeted interventions, including infection control, hemodynamic stabilization, and fluid resuscitation.

Norepinephrine, a potent vasopressor widely utilized in managing shock, particularly in severe septic shock stemming from AP, reflects the severity of the patient's

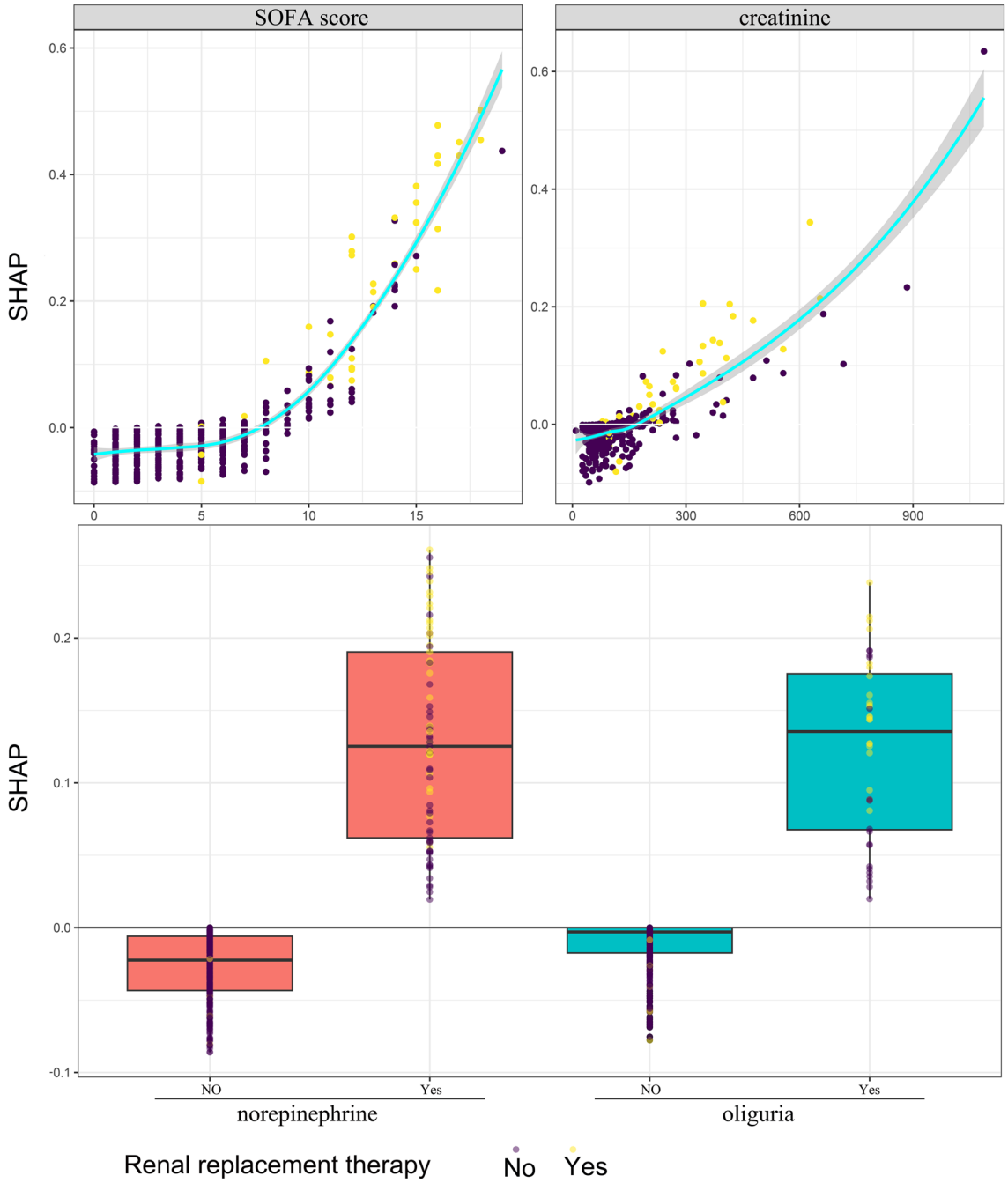

Fig. 4 The influence of four distinct features on the predicted outcomes. The horizontal axis displays the values of continuous variables or the presence/absence of categorical variables, whereas the vertical axis denotes the SHAP values. Patient attributions to outcomes are visually depicted through colored dots: yellow dots suggest a requirement for RRT, while black dots signify no such need. *SOFA* Sequential Organ Failure Assessment, *SHAP* SHapley Additive exPlanations, *RRT* Renal Replacement Treatment

condition and shock status through its required dosage and intensity [34, 35]. Some perspectives argue that vasopressors can mitigate the risk of AKI by enhancing glomerular perfusion pressure, which may also reduce the need for RRT [36]. Conversely, more research suggests that norepinephrine use is a predictive indicator for AKI development in patients with AP [37] and is correlated with increased mortality rates [38]. This contradiction may stem from the fact that norepinephrine usage signifies severe circulatory failure in patients, a stage

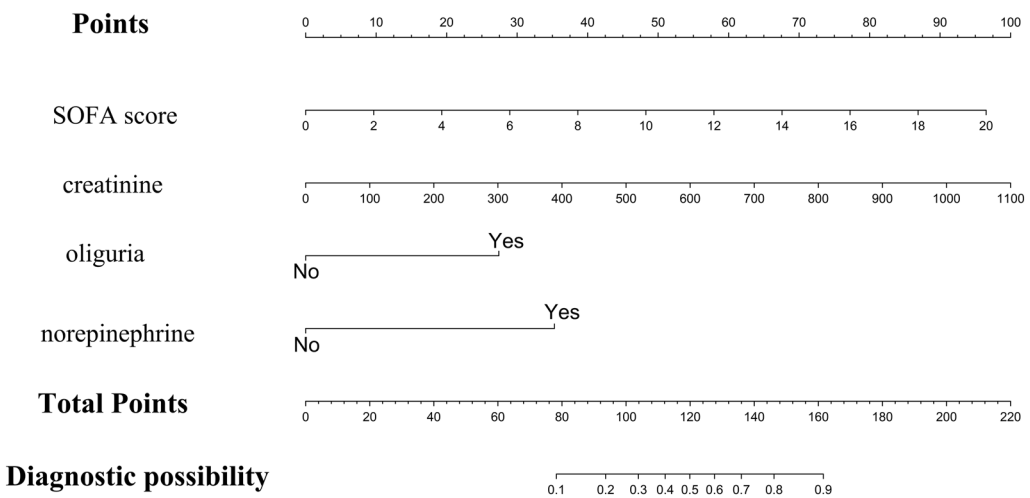

Fig. 5 Nomogram for predicting renal replacement treatment in patients with acute pancreatitis. *SOFA* Sequential Organ Failure Assessment

where renal function is already significantly impaired, which demands early and aggressive supportive care. In our study, the use of norepinephrine during hospitalization was incorporated into the model as an independent predictor. This highlights the severity of the condition in patients receiving such treatment and the challenges associated with renal recovery, thereby indicating an increased need for RRT. These findings further confirm that assessing the necessity and intensity of vasopressor use is crucial for predicting renal recovery and determining the need for RRT in managing patients with AP.

This study indeed presents several significant advantages. Firstly, a large public healthcare database is utilized for the first time to develop a predictive model for RRT in patients with AP. This approach allows the study to be based on an extensive clinical dataset, enhancing the model's generalizability and reliability. Secondly, multiple ML algorithms are employed, and cross-validation is used for model evaluation and optimization, ensuring the model's robustness and accuracy. The use of SHAP values and nomogram increases the model's transparency and interpretability. Lastly, the metrics adopted by the model are primarily based on data within the first 24 h after patient admission, highlighting the model's early prediction capability. Early identification of high-risk patients with AP and providing them with timely intervention and management are crucial for improving outcomes.

This study has certain limitations. Firstly, as a single-center retrospective study limited by geographic and demographic factors, its generalizability across diverse populations may be compromised. Secondly, the reliance on the MIMIC-IV database constrains the scope of accessible and analyzable features. Although 41 features were considered, the complex and variable etiology of AP suggests the existence of significant predictors not included in this study. Finally, while internal validation of the model has been conducted, the absence of external validation could hinder its practical application in clinical settings. Future research, by conducting large-sample prospective studies on datasets from different regions and populations, could enhance the generalizability and clinical utility of the model.

Conclusions

The Lasso-LR model, developed through the LASSO regression cross-validation method, demonstrates a significant advantage in predicting the necessity of RRT for patients with AP. This model demonstrates potential clinical application, offering crucial decision support to physicians. To further confirm the broad applicability and accuracy of the study results, future research needs to be validated on a wider range of multi-center cohorts.

Abbreviations

RRT	Renal replacement treatment
AP	Acute pancreatitis
ICU	Intensive care unit
MIMIC-IV	Medical Information Mart for Intensive Care IV
LASSO	Least absolute shrinkage and selection operator
ML	Machine learning
SHAP	SHapley Additive exPlanations
APACHE II	Acute Physiology and Chronic Health Evaluation II
AKI	Acute kidney injury
SOFA	Sequential Organ Failure Assessment
Lasso-LR	Lasso-Logistic Regression
AUROC	Area under the receiver operating characteristic
DCA	Decision curve analysis
SAP	Severe acute pancreatitis
MODS	Multiple organ dysfunction syndrome
MIT	Massachusetts Institute of Technology
BIDMC	Beth Israel Deaconess Medical Center
TRIPOD	Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis
MAP	Mean arterial pressure

WBC	White blood cell
RDW	Red cell distribution width
ALT	Alanine aminotransferase
AST	Aspartate transaminase
BUN	Blood urea nitrogen
CHF	Congestive heart failure
COPD	Chronic obstructive pulmonary disease
CKD	Chronic kidney disease
MV	Mechanical ventilation
SD	Standard deviation
RF	Random Forest
XGBoost	EXtreme Gradient Boosting
SVM	Support Vector Machine
ROC	Receiver operating characteristic
KDIGO	Kidney Disease Improving Global Outcomes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40001-025-02345-5.

Supplementary Material 1. Fig. 1 Calibration curves of four machine learning models in the validation set. Abbreviations: Lasso-LR, Lasso-Logistic Regression; SVM, Support Vector Machine; RF, Random Forest; XGBoost, eXtreme Gradient Boosting

Supplementary Material 2. Fig. 2 Decision curve analysis of four machine learning models in the validation set. Abbreviations: Lasso-LR, Lasso-Logistic Regression; SVM, Support Vector Machine; RF, Random Forest; XGBoost, eXtreme Gradient Boosting

Supplementary Material 3. Fig. 3 ROC curves and AUROCs of the Lasso-LR model in the training set and validation set. Abbreviations: ROC, Receiver Operating Characteristic; Lasso-LR, Least Absolute Shrinkage and Selection Operator-Logistic Regression

Supplementary Material 4. Fig. 4 Calibration curves of the Lasso-LR model in the training set and validation set. Abbreviations: ROC, Receiver Operating Characteristic; Lasso-LR, Least Absolute Shrinkage and Selection Operator-Logistic Regression

Supplementary Material 5. Fig. 5 Decision curve analysis of the Lasso-LR model in the training set and validation set. The green horizontal line indicates patients who do not require renal replacement therapy, while the red diagonal line represents patients who do require RRT. The blue solid line illustrates the risk curve for RRT. In the DCA, within the range of threshold probabilities, the Lasso-LR model demonstrates greater net benefits than either full treatment or no treatment. Abbreviations: DCA, Decision Curve Analysis; Lasso-LR, Least Absolute Shrinkage and Selection Operator-Logistic Regression; RRT, RRT, Renal Replacement Treatment

Acknowledgements
None.

Author contributions
Conception and design: Fei Zuo, Lei Zhong and Longping Yao. Administrative support: Longping Yao. Provision of study materials or patients: Lei Zhong and Jie Min. Collection and assembly of data: Lei Zhong, Jie Min and Jinyu Zhang. Data analysis and interpretation: Lei Zhong and Jie Min. Manuscript writing: Fei Zuo, Jie Min, Jinyu Zhang and Longping Yao. Final approval of manuscript: All authors.

Funding
This study was funded by the Science and Technology Program of Huzhou (No: 2021GYB57).

Data availability
Publicly available datasets were analyzed in this study. This data can be found here: https://physionet.org/content/mimiciv/2.0/. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The dataset in this study was obtained from MIMIC-IV v2.0. We had completed the CITI Program course known as Human Research and Data or Specimens Only Research to apply for permission to access the database (Record ID: 51774135, 53446653, 61758806). The individual information of the patients included in this database was anonymous, and ethical review and informed consent were waived. All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Gastroenterology, Huzhou Central Hospital, Affiliated Central Hospital of Huzhou University, Huzhou 313000, China. [2]Department of Intensive Care Unit, Huzhou Central Hospital, Affiliated Central Hospital of Huzhou University, Huzhou 313000, China. [3]Department of General Surgery, Huzhou Central Hospital, Affiliated Central Hospital of Huzhou University, Huzhou 313000, China. [4]Huzhou Central Hospital, Fifth School of Clinical Medicine of Zhejiang Chinese Medical University, Huzhou 313000, China.

Received: 15 April 2024 Accepted: 29 January 2025
Published online: 04 February 2025

References

1. Leppäniemi A, Tolonen M, Tarasconi A, et al. 2019 WSES guidelines for the management of severe acute pancreatitis. World J Emerg Surg. 2019;14:27.
2. Petrov MS, Yadav D. Global epidemiology and holistic prevention of pancreatitis. Nat Rev Gastroenterol Hepatol. 2019;16(3):175–84.
3. Zhou Y, Ge YT, Shi XL, et al. Machine learning predictive models for acute pancreatitis: a systematic review. Int J Med Inform. 2022;157: 104641.
4. Iannuzzi JP, King JA, Leong JH, Quan J, Windsor JW, Tanyingoh D, Coward S, Forbes N, Heitman SJ, Shaheen AA, et al. Global incidence of acute pancreatitis is increasing over time: a systematic review and meta-analysis. Gastroenterology. 2022;162:122–34. https://doi.org/10.1053/j.gastro.2021.09.043.
5. Guidelines for diagnosis and treatment of acute pancreatitis in China. Zhonghua Wai Ke Za Zhi. 2021. 59(7): 578–587.
6. Xu F, Chen X, Li C, et al. Prediction of multiple organ failure complicated by moderately severe or severe acute pancreatitis based on machine learning: a multicenter cohort study. Mediat Inflamm. 2021;2021:5525118.
7. Banks PA, Bollen TL, Dervenis C, et al. Classification of acute pancreatitis–2012: revision of the Atlanta classification and definitions by international consensus. Gut. 2013;62(1):102–11.
8. Nassar TI, Qunibi WY. AKI associated with acute pancreatitis. Clin J Am Soc Nephrol. 2019;14(7):1106–15.
9. Garg PK, Singh VP. Organ failure due to systemic injury in acute pancreatitis. Gastroenterology. 2019;156(7):2008–23.
10. Mederos MA, Reber HA, Girgis MD. Acute pancreatitis: a review. JAMA. 2021;325(4):382–90.
11. Zhang X, Cao Y, Pan CK, et al. Effect of initiation of renal replacement therapy on mortality in acute pancreatitis patients. Medicine (Baltimore). 2020;99(47): e23413.
12. Li B, Wu W, Liu A, et al. Establishment and validation of a nomogram prediction model for the severe acute pancreatitis. J Inflamm Res. 2023;16:2831–43.
13. Zhou Y, Han F, Shi XL, et al. Prediction of the severity of acute pancreatitis using machine learning models. Postgrad Med. 2022;134(7):703–10.
14. Ding N, Guo C, Li C, Zhou Y, Chai X. An artificial neural networks model for early predicting in-hospital mortality in acute pancreatitis in MIMIC-III. Biomed Res Int. 2021;2021:6638919.

15. Liu F, Yao J, Liu C, Shou S. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis. BMC Surg. 2023;23(1):267.
16. Lin S, Lu W, Wang T, et al. Predictive model of acute kidney injury in critically ill patients with acute pancreatitis: a machine learning approach using the MIMIC-IV database. Ren Fail. 2024;46(1):2303395.
17. Lan L, Guo Q, Zhang Z, et al. Classification of infected necrotizing pancreatitis for surgery within or beyond 4 weeks using machine learning. Front Bioeng Biotechnol. 2020;8:541.
18. Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101:E215-220.
19. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63.
20. Gautam SC, Lim J, Jaar BG. Complications associated with continuous RRT. Kidney360. 2022;3(11):1980–90.
21. Wald R, Beaubien-Souligny W, Chanchlani R, et al. Delivering optimal renal replacement therapy to critically ill patients with acute kidney injury. Intensive Care Med. 2022;48(10):1368–81.
22. Teng T, Tan J, Baey S, et al. Sequential organ failure assessment score is superior to other prognostic indices in acute pancreatitis. World J Crit Care Med. 2021;10(6):355–68.
23. Para O, Caruso L, Savo MT, et al. The challenge of prognostic markers in acute pancreatitis: internist's point of view. J Genet Eng Biotechnol. 2021;19(1):77.
24. Stojanović M, Đurić M, Nenadić I, Dimić N, Bojić S, Stevanović P. Evaluation of intra-abdominal hypertension parameters in patients with acute pancreatitis. Life (Basel). 2023. https://doi.org/10.3390/life13061227.
25. Zarnescu NO, Dumitrascu I, Zarnescu EC, Costea R. Abdominal compartment syndrome in acute pancreatitis: a narrative review. Diagnostics (Basel). 2022. https://doi.org/10.3390/diagnostics13010001.
26. Dumnicka P, Mazur-Laskowska M, Ceranowicz P, et al. Acute changes in serum creatinine and kinetic glomerular filtration rate estimation in early phase of acute pancreatitis. J Clin Med. 2022. https://doi.org/10.3390/jcm11206159.
27. Yang Y, Xiao W, Liu X, Zhang Y, Jin X, Li X. Machine learning-assisted ensemble analysis for the prediction of acute pancreatitis with acute kidney injury. Int J Gen Med. 2022;15:5061–72.
28. Zhang R, Yin M, Jiang A, Zhang S, Xu X, Liu L. Automated machine learning for early prediction of acute kidney injury in acute pancreatitis. BMC Med Inform Decis Mak. 2024;24(1):16.
29. Czapári D, Váradi A, Farkas N, et al. Detailed characteristics of post-discharge mortality in acute pancreatitis. Gastroenterology. 2023;165(3):682–95.
30. Zhifeng Z, Rongli X, Li L, et al. Risk factors for reoperation after debridement of acute pancreatitis. J Surg Res. 2020;251:63–70.
31. Yamao Y, Oami T, Yamabe J, Takahashi N, Nakada TA. Machine-learning model for predicting oliguria in critically ill patients. Sci Rep. 2024;14(1):1054.
32. Vijayan A. Tackling AKI: prevention, timing of dialysis and follow-up. Nat Rev Nephrol. 2021;17(2):87–8.
33. Shao J, Liu F, Ji S, Song C, Ma Y, Shen M, et al. Development, external validation, and visualization of machine learning models for predicting occurrence of acute kidney injury after cardiac surgery. Rev Cardiovasc Med. 2023;24(8):229.
34. Ammar MA, Ammar AA, Wieruszewski PM, et al. Timing of vasoactive agents and corticosteroid initiation in septic shock. Ann Intensive Care. 2022;12(1):47.
35. Kotani Y, Di Gioia A, Landoni G, Belletti A, Khanna AK. An updated "norepinephrine equivalent" score in intensive care as a marker of shock severity. Crit Care. 2023;27(1):29.
36. 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(8):835–50.
37. Wu S, Zhou Q, Cai Y, Duan X. Development and validation of a prediction model for the early occurrence of acute kidney injury in patients with acute pancreatitis. Ren Fail. 2023;45(1):2194436.
38. Fischer AJ, Andreottola F, Lenz P, Lebiedz P. Acute pancreatitis in intensive care medicine : which risk score is useful? Med Klin Intensivmed Notfmed. 2017;112(8):717–23.

Publisher's Note