

Developing a predictive model for septic shock risk in acute pancreatitis patients using interpretable machine learning algorithms

Binglin Song¹ , Ping Liu² , Kangrui Fu²  and Chun Liu²

Abstract

Background: Septic shock is a severe complication of acute pancreatitis (AP), often associated with poor prognosis. This study aims to analyze the clinical characteristics of patients with acute pancreatitis and develop an interpretable early prediction model for septic shock in these patients using machine learning (ML). The model is intended to assist emergency physicians in resource allocation and medical decision making.

Methods: Data were collected from the MIMIC-IV 3.0 database. The dataset was divided into a training set and a test set in a 7:3 ratio. Feature selection was performed using LASSO (Least Absolute Shrinkage and Selection Operator) regression. Subsequently, 10 ML models were developed: Random Forest, Logistic Regression, Gradient Boosting Machine, Neural Network, Extreme Gradient Boosting (XGBoost), K-Nearest Neighbor, Adaptive Boosting, Light Gradient Boosting Machine, Category Boosting, and Support Vector Machine. To enhance and optimize model interpretability, Shapley Additive Explanations (SHAP) were employed.

Results: A total of 1032 patients with AP were included in this study, from which 31 variables were selected for model development. By comparing the area under the receiver operating characteristic curve and decision curve analysis results between the training and test sets, the XGBoost model demonstrated a significant advantage over other models. SHAP analysis revealed that white blood cell count, total bilirubin (bilirubin total), and bicarbonate (HCO_3^-) levels were the three most critical risk factors for the development of septic shock in patients with AP.

Conclusion: ML approaches exhibited promising performance in predicting septic shock in patients with AP. These models may aid in guiding treatment decisions for patients with AP in the emergency department.

Keywords

Acute pancreatitis, septic shock, Shapley Additive Explanations (SHAP), machine learning

Received: 12 February 2025; accepted: 14 May 2025

Background

Acute pancreatitis (AP) is a common emergency condition, with approximately 20% of patients progressing to severe acute pancreatitis (SAP), which is associated with multiple complications.^{1,2} Sepsis-induced shock is one of the most frequent and fatal complications in patients with SAP, with an incidence of approximately 10% and a very poor prognosis. Early identification and intervention of septic shock are critical for the prognosis of patients with AP.^{3,5} Several risk scoring systems have been developed to assess the severity of AP. Commonly used scores include the Ranson criteria, the Acute Physiology and Chronic Health

Evaluation II (APACHE II), C-reactive protein (CRP) levels, the Computed Tomography Severity Index (CTSI) and its modified version (mCTSI), the Harmless Acute Pancreatitis Score (HAPS), and the Bedside Index for

¹Clinical Medical College, North Sichuan Medical College, Nanchong, China

²Emergency Department, Dazhou Central Hospital, Dazhou, China

Corresponding author:

Chun Liu, Emergency Department, Dazhou Central Hospital, No.156, Dongda Road, Tongchuan District, Dazhou 635000, Sichuan Province, China. Email: 1501213610@qq.com



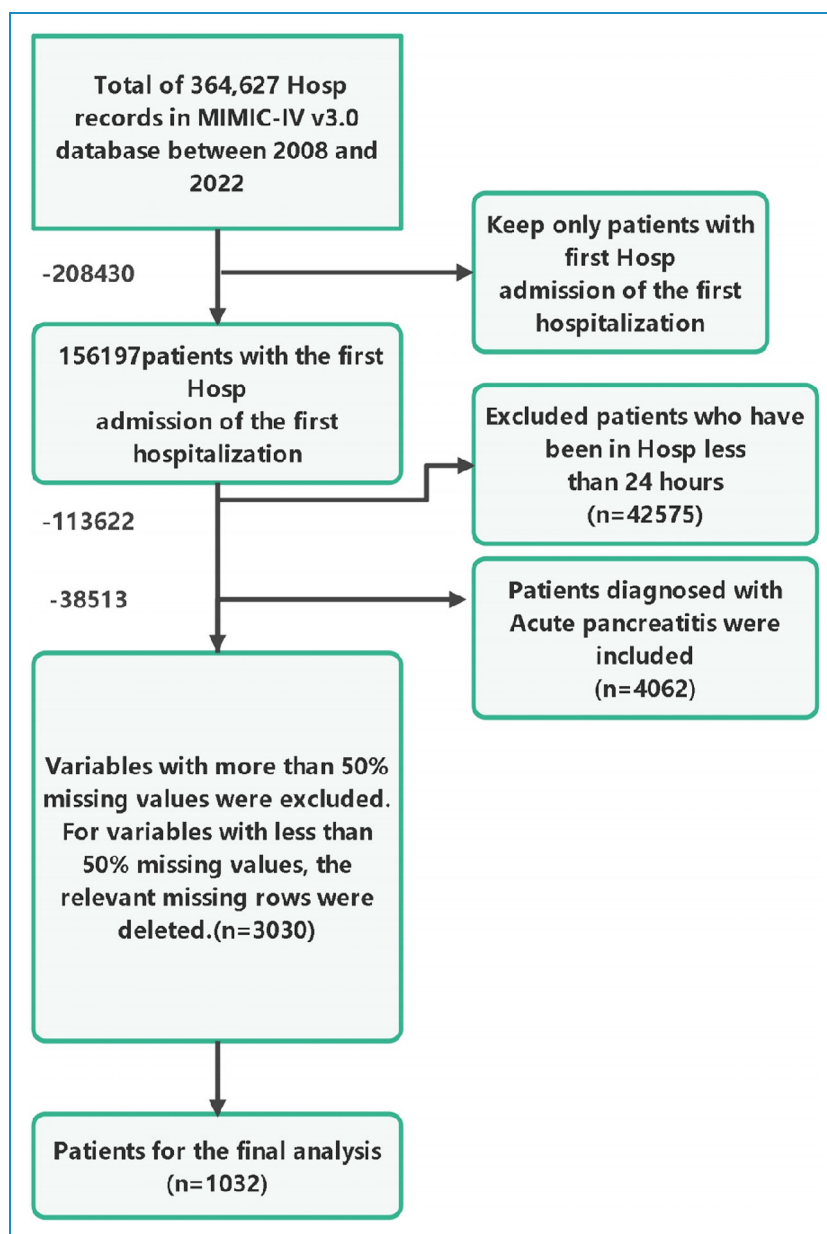


Figure 1. Flow chart of data screening.

Severity in Acute Pancreatitis (BISAP) score.^{6–8} These systems provide an overall assessment of the prognosis of patients with AP but do not specifically analyze the causal relationship between AP and septic shock. Emergency department physicians face a large volume of complex acute cases, and early identification of patients with AP at risk of septic shock is crucial for enhancing clinical decision-making efficiency and optimizing resource allocation, ultimately improving patient outcomes.^{9,11}

Machine learning (ML) is a computational algorithm that learns from data and identifies underlying patterns, capable of uncovering complex relationships within large datasets,

making it of significant clinical value.^{12,13} However, ML models are often criticized for their “black-box” nature, as the process from input to output remains largely uninterpretable, which hinders their clinical application.¹⁴ Shapley Additive Explanations (SHAP), a post hoc interpretability method, addresses this limitation by attributing SHAP values to ML models, thereby offering an effective way to explain and visualize their decision-making process.¹⁵ However, no studies have yet applied SHAP to ML models predicting septic shock in patients with pancreatitis. This study leverages the MIMIC-IV database to develop an interpretable ML model to predict the risk of septic shock in patients with AP, aiding emergency department physicians in the early identification

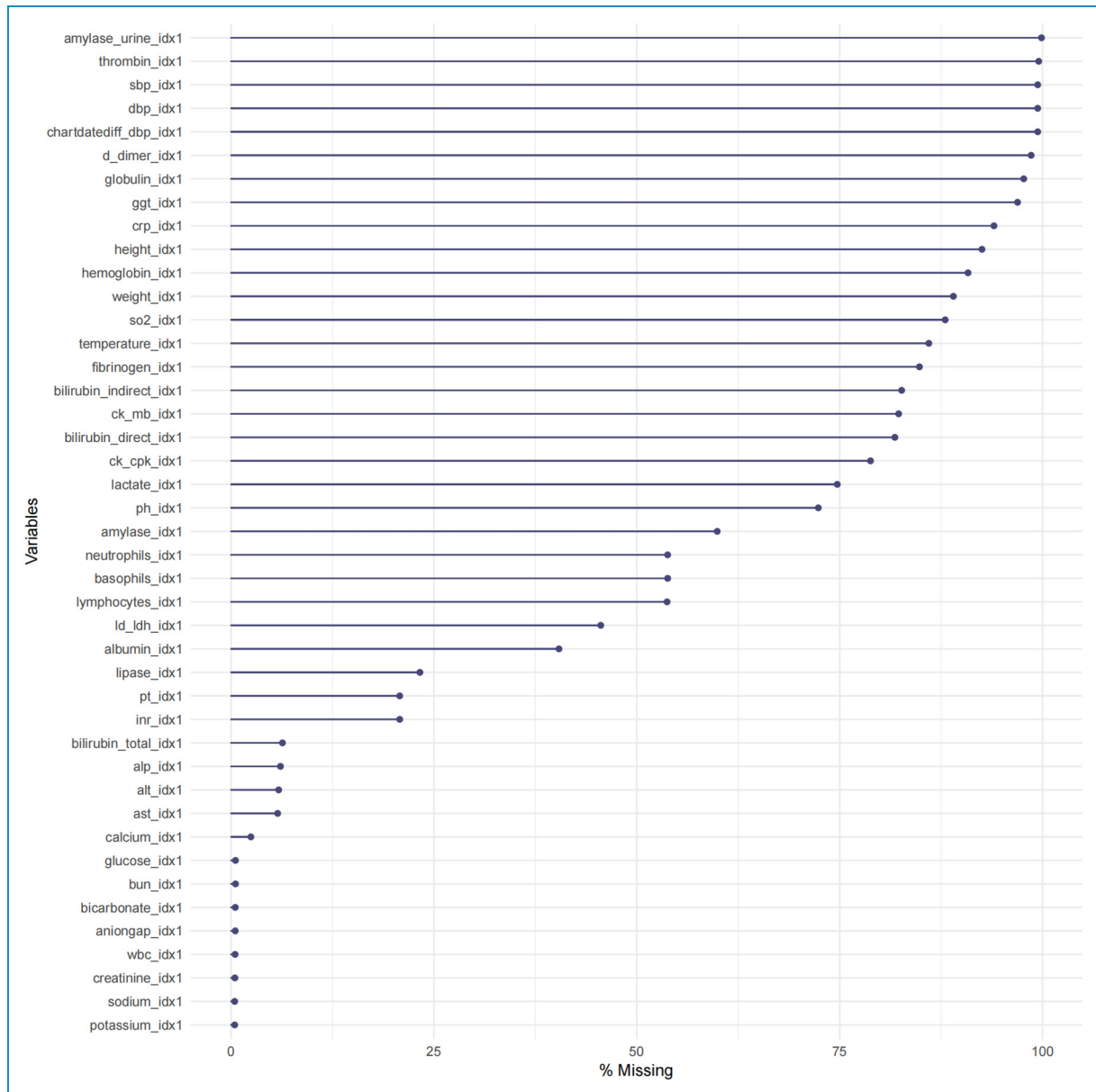


Figure 2. Missing value percentage per variable prior to data preprocessing. Variables such as `amylase_urine_idx1`, `thrombin_idx1`, and `sbp_idx1` exhibited the highest missingness (>50%) and were excluded from analysis. Clinical variables with low missingness such as `potassium_idx1`, `sodium_idx1`, and `creatinine_idx1` were retained.

of high-risk patients with AP and informing subsequent treatment decisions.

Methods

Data source

This study is a retrospective analysis based on the MIMIC-IV database. The MIMIC-IV (Medical Information Mart for Intensive Care, version 3.0) is an open-access, freely available critical care database that contains comprehensive clinical data from patients admitted to medical centers between

2008 and 2022. The database includes detailed information on 300,000 emergency department (HOSP) patients and 90,000 intensive care unit (ICU) patients, covering demographic characteristics, vital signs, imaging results, laboratory findings, and other clinical data. Additionally, it includes codes from the International Classification of Diseases, Ninth and Tenth Editions (ICD-9 and ICD-10). The MIMIC-IV database is a publicly available dataset comprising de-identified patient information; therefore, no additional ethical approval is required for its use.¹⁶ Data extraction for this study was approved by the Institutional Review Board (IRB) of the Massachusetts

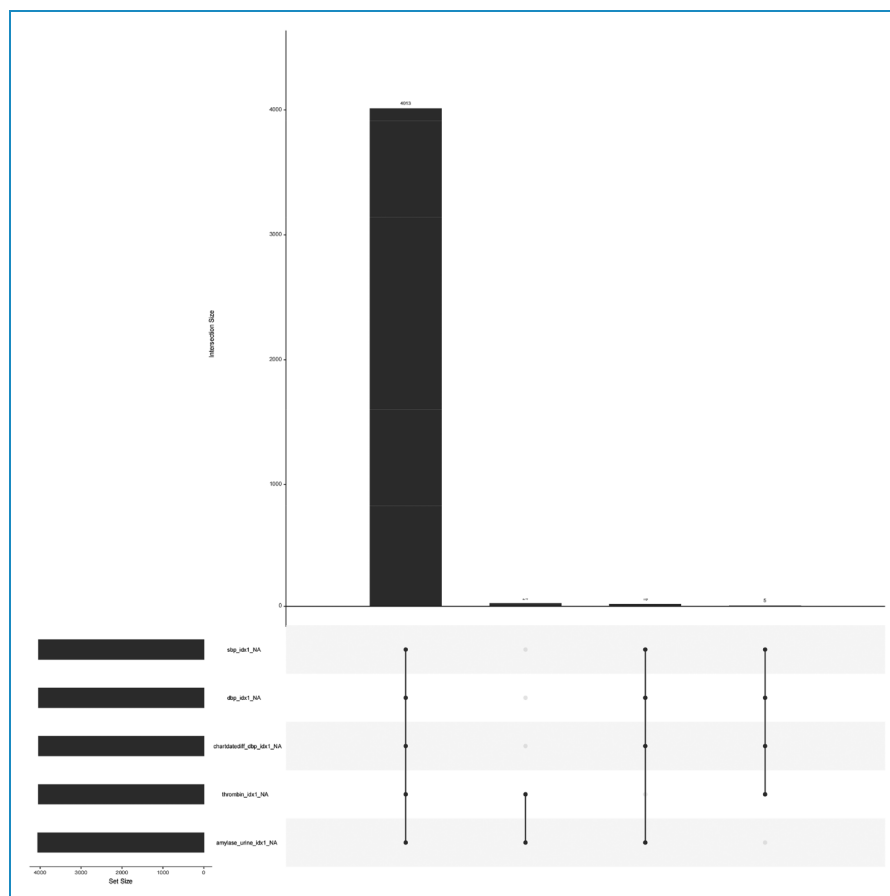


Figure 3. Upset plot depicting the co-occurrence patterns of missingness among the top five variables with the highest missing rates. Each vertical bar represents the number of observations sharing a specific pattern of missing values (shown by the connected black dots in the lower matrix). For example, the tallest bar indicates that over 4000 records were missing only sbp_idx1. This visualization confirms that missing values are not highly clustered across multiple variables or within specific patient subgroups, supporting the appropriateness of our listwise deletion strategy.

Institute of Technology (MIT) (Certification No: 65828043, Named: Zhihao Li).

Study population

The inclusion criteria for this study were as follows: (1) age ≥ 18 years; (2) diagnosis of acute pancreatitis based on the 2012 revised Atlanta classification criteria¹⁷; and (3) diagnosis of sepsis as per the Sepsis-3 definition.¹⁸ Patients were excluded if they met any of the following criteria: (1) chronic pancreatitis; (2) acute pancreatitis resulting from trauma; (3) pregnancy or peripartum period; (4) patients who had already been administered vasopressors at the time of admission to maintain blood pressure; and (5) patients with incomplete clinical data or lacking follow-up information.

Data collection

We first extracted and processed the raw data using Structured Query Language (SQL) and Navicat Premium (version

15.0.12). SQL scripts were employed to retrieve patient data from the **hosp** database, which included demographic characteristics, vital signs, laboratory parameters, comorbidities, and microbiological data. The demographic features extracted comprised sex, admission age, race, weight, height, length of hospital stay, and ICU stay. Vital sign data included heart rate, respiratory rate, systolic blood pressure, mean arterial pressure, and oxygen saturation. Comorbidities such as hypertension, congestive heart failure, chronic lung disease, diabetes, renal disease, severe liver disease, and acute kidney injury (AKI) staging were also recorded.

Laboratory parameters included the following:

- White blood cell count (WBC), neutrophil percentage (NEUT%), platelet count (PLT), hemoglobin (HGB), hematocrit (HCT), high-sensitivity CRP, blood urea nitrogen (BUN), uric acid (UA), triglycerides (TG), cholesterol (CHOL), albumin (ALB), potassium (K+), calcium (Ca2+), sodium (Na), chloride (Cl⁻), bicarbonate (HCO₃⁻), amylase (AMY), lipase,

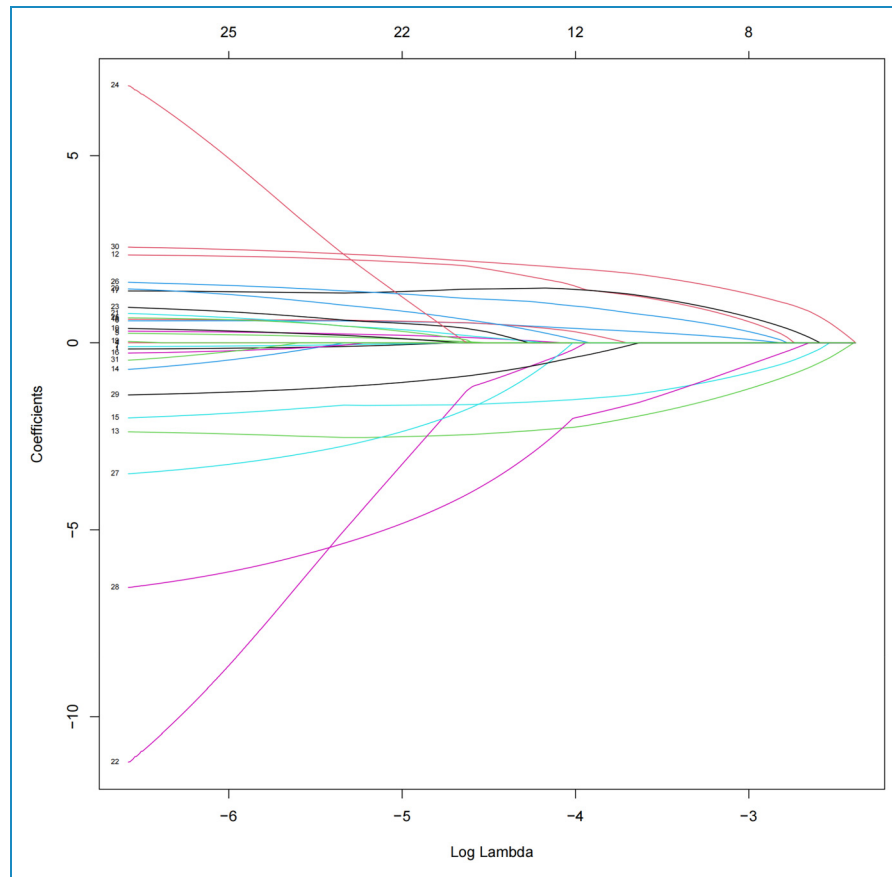


Figure 4. Path plot of LASSO regression coefficients. LASSO: Least Absolute Shrinkage and Selection Operator.

total bilirubin (TBIL), direct bilirubin (DBIL), indirect bilirubin (UBIL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (γ -GGT), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), prothrombin time (PT), activated partial thromboplastin time (APTT), d-dimer (DDHS), PaO₂/FiO₂ ratio, lactate (Lac), base excess (BE), and arterial-venous CO₂ tension gap (Gap).

Additionally, surgical details such as pancreatic resection and drainage procedures were extracted. All the variables retrieved served as the foundation for constructing the predictive model of septic shock in acute pancreatitis patients and were used in subsequent analyses.

Data preprocessing

- All data processing was conducted using R (version 4.4.1). We performed missing data analysis, as certain variables in the hosp database had significant missing values. To retain important variables, variables with >50% missingness were excluded. For the remaining variables, listwise deletion was

performed to ensure complete-case inputs. Figure 1 illustrates the process of data screening. Figure 2 summarizes the variable-wise missingness rates, highlighting that some variables such as *amylase_urine_idx1* and *thrombin_idx1* exceeded 50% missingness and were therefore excluded. To further explore the structure of missing data, Figure 3 presents an UpSet plot showing the co-occurrence of missing values among the top five variables with the highest missing rates. This visualization confirms that most missing values occurred independently rather than in the same samples, supporting the appropriateness of the listwise deletion strategy. After preprocessing, 1032 patients and 31 variables were used for model development (Figures 1–3).

Statistical methods

For normally distributed continuous variables, data are presented as mean $\pm$ standard deviation ($\bar{x} \pm s$). Non-normally distributed continuous variables are expressed as median and interquartile range (M(Q1, Q3)), with group differences compared using the Mann-Whitney U test. Categorical

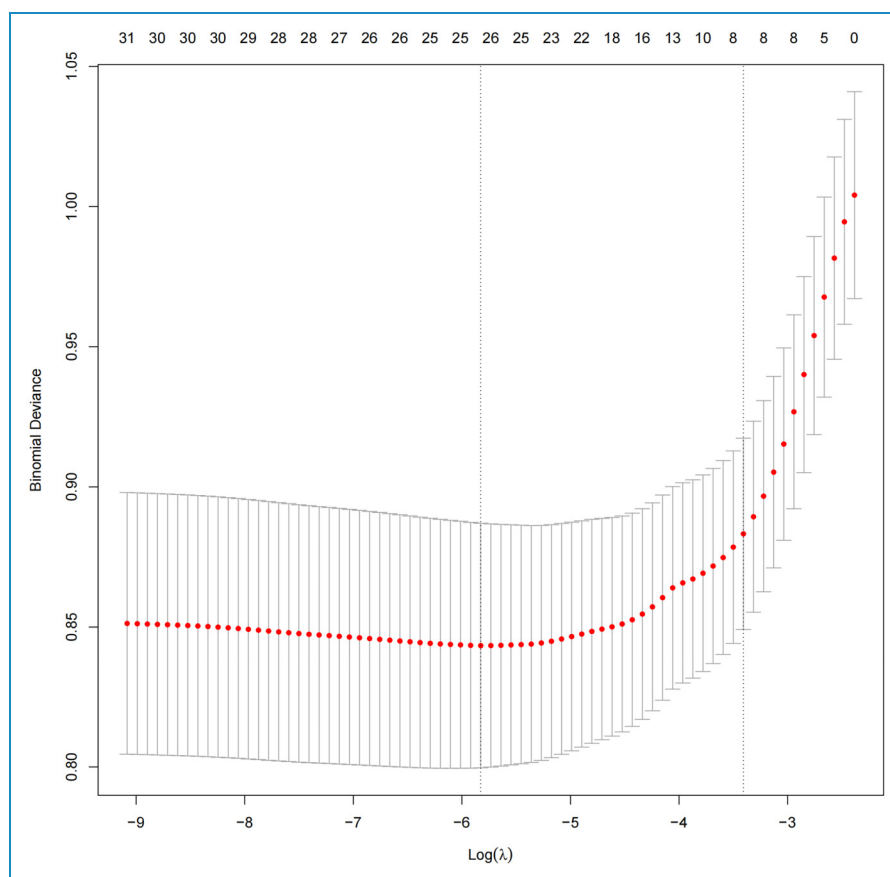


Figure 5. LASSO regression cross-validation plot. LASSO: Least Absolute Shrinkage and Selection Operator.

variables are presented as frequency (percentage), with Chi-square tests or Fisher's exact test used for comparison. Ordinal variables are analyzed using the rank-sum test.

Cohort division and feature selection

The dataset was randomly split into a training cohort (70%, $n=722$) and a validation cohort (30%, $n=310$). To mitigate collinearity in the data analysis, Lasso regression was employed to select variables with a p -value < 0.05 from univariate analysis.¹⁹ As the penalty term (λ) increased, the regression coefficients of the model's predictors were progressively shrunk toward zero. Ultimately, eight variables with non-zero coefficients were selected, including: white blood cell count (wbc), total bilirubin, bicarbonate, blood urea nitrogen (bun), prothrombin time (pt), lymphocyte count (lymphocytes), albumin, and severe liver disease (Figures 4 and 5).

Machine learning models and evaluation

Ten machine learning (ML) models were used to predict the risk of septic shock: Random Forest (RF), Logistic Regression (LR), Gradient Boosting Machine (GBM),

Neural Network (NN), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Adaptive Boosting (AdaBoost), Light Gradient Boosting Machine (LightGBM), Category Boosting (CatBoost), and Support Vector Machine (SVM). To assess model performance, 200 random resampling iterations were conducted in both the training and test datasets. Model evaluation was performed using the area under the receiver operating characteristic curve (AUC-ROC) and decision curve analysis (DCA).

The model's efficiency and effectiveness in managing septic shock among acute pancreatitis patients were further evaluated on both the test and validation sets using metrics including Accuracy, Precision, Recall, and F1-Score.

SHAP interpretability analysis

To enhance the interpretability of our model, we employed the SHAP method. SHAP is based on the concept of Shapley values from cooperative game theory, and it uses additive models to assess the contribution of each feature to the model's predictions. This method provides a deeper understanding of model outputs, making it particularly useful for explaining complex ML models.²⁰ We utilized the shapviz package to calculate and visualize the SHAP values

Table 1. Baseline characteristics of patients with acute pancreatitis with and without septic shock.

Variables	Total (n = 1031)	0 (n = 824)	1 (n = 207)	p
Gender, n (%)				0.61
0	472 (46)	381 (46)	91 (44)	
1	559 (54)	443 (54)	116 (56)	
los_hospital, Median (Q1,Q3)	14.11 (7.37, 25.76)	12.68 (6.76, 21.84)	22.69 (13.59, 42.55)	< 0.001
Age, median (Q1,Q3)	59.31 (46.36, 72.42)	59 (45.7, 73.16)	61.31 (51.75, 70.51)	0.169
hosp_outcome_7d, n (%)				0.011
0	1014 (98)	815 (99)	199 (96)	
1	17 (2)	9 (1)	8 (4)	
hosp_outcome_28d, n (%)				< 0.001
0	913 (89)	773 (94)	140 (68)	
1	118 (11)	51 (6)	67 (32)	
hosp_outcome_90d, n (%)				< 0.001
0	823 (80)	714 (87)	109 (53)	
1	208 (20)	110 (13)	98 (47)	
chronic_pulmonary_disease, n (%)				0.503
0	822 (80)	653 (79)	169 (82)	
1	209 (20)	171 (21)	38 (18)	
peptic_ulcer_disease, n (%)				< 0.001
0	972 (94)	788 (96)	184 (89)	
1	59 (6)	36 (4)	23 (11)	
mild_liver_disease, n (%)				0.095
0	738 (72)	600 (73)	138 (67)	
1	293 (28)	224 (27)	69 (33)	
Diabetes, n (%)				0.273
0	722 (70)	584 (71)	138 (67)	
1	309 (30)	240 (29)	69 (33)	
renal_disease, n (%)				0.017
0	821 (80)	669 (81)	152 (73)	
1	210 (20)	155 (19)	55 (27)	

(continued)

Table 1. Continued.

Variables	Total (n = 1031)	0 (n = 824)	1 (n = 207)	p
malignant_cancer, n (%)				0.816
0	899 (87)	717 (87)	182 (88)	
1	132 (13)	107 (13)	25 (12)	
severe_liver_disease, n (%)				< 0.001
0	914 (89)	752 (91)	162 (78)	
1	117 (11)	72 (9)	45 (22)	
Hypertension, n (%)				0.033
0	618 (60)	480 (58)	138 (67)	
1	413 (40)	344 (42)	69 (33)	
Surgery, n (%)				0.004
0	961 (93)	778 (94)	183 (88)	
1	70 (7)	46 (6)	24 (12)	
norepinephrine_hosp_used, n (%)				< 0.001
0	701 (68)	674 (82)	27 (13)	
1	330 (32)	150 (18)	180 (87)	
dobutamine_hosp_used, n (%)				0.018
0	1016 (99)	816 (99)	200 (97)	
1	15 (1)	8 (1)	7 (3)	
dopamine_hosp_used, n (%)				< 0.001
0	1004 (97)	811 (98)	193 (93)	
1	27 (3)	13 (2)	14 (7)	
congestive_heart_failure, n (%)				0.097
0	836 (81)	677 (82)	159 (77)	
1	195 (19)	147 (18)	48 (23)	
hospital_expire_flag, n (%)				< 0.001
0	890 (86)	767 (93)	123 (59)	
1	141 (14)	57 (7)	84 (41)	
inr_idx1, median (Q1,Q3)	1.3 (1.1, 1.5)	1.3 (1.1, 1.5)	1.4 (1.2, 1.95)	< 0.001
pt_idx1, median (Q1,Q3)	14.2 (12.7, 16.8)	14 (12.57, 16)	15.8 (13.65, 21.15)	< 0.001

(continued)

Table 1. Continued.

Variables	Total (n = 1031)	0 (n = 824)	1 (n = 207)	p
albumin_idx1, median (Q1,Q3)	3 (2.6, 3.5)	3.1 (2.7, 3.5)	2.7 (2.3, 3.2)	< 0.001
aniongap_idx1, median (Q1,Q3)	15 (12, 17)	14 (12, 17)	16 (13, 19)	< 0.001
bicarbonate_idx1, median (Q1,Q3)	23 (19, 25)	23 (20, 26)	20 (17, 23)	< 0.001
calcium_idx1, median (Q1,Q3)	8.3 (7.7, 8.8)	8.3 (7.7, 8.8)	8 (7.4, 8.65)	< 0.001
bun_idx1, median (Q1,Q3)	18 (11, 31.5)	16 (10, 28)	26 (17, 42.5)	< 0.001
potassium_idx1, median (Q1,Q3)	4 (3.6, 4.4)	3.9 (3.6, 4.3)	4.1 (3.8, 4.6)	< 0.001
sodium_idx1, median (Q1,Q3)	138 (135, 141)	138 (135, 141)	137 (134, 140)	0.004
glucose_idx1, median (Q1,Q3)	118 (95, 156)	115 (94, 151.25)	132 (99, 169.5)	< 0.001
creatinine_idx1, median (Q1,Q3)	1 (0.7, 1.7)	0.9 (0.7, 1.42)	1.5 (0.9, 2.35)	< 0.001
alt_idx1, median (Q1,Q3)	48 (23, 144)	47 (23, 144)	51 (24, 143)	0.645
alp_idx1, median (Q1,Q3)	106 (70, 181.5)	105 (70, 174)	112 (71.5, 219.5)	0.1
ast_idx1, median (Q1,Q3)	65 (31.5, 166)	58 (30, 151.25)	82 (39.5, 209)	0.001
ld_ldh_idx1, median (Q1,Q3)	308 (211, 487)	294 (204.75, 474)	364 (238, 574.5)	< 0.001
bilirubin_total_idx1, median (Q1,Q3)	1.1 (0.5, 3)	1 (0.5, 2.52)	1.8 (0.6, 5.5)	< 0.001
neutrophils_idx1, median (Q1,Q3)	80.4 (70.45, 86.7)	79.65 (69.88, 86.7)	82.6 (74.95, 86.8)	0.02
lymphocytes_idx1, median (Q1,Q3)	9 (5, 15.6)	10 (6, 17.3)	6 (3.75, 10)	< 0.001
basophils_idx1, median (Q1,Q3)	0.2 (0, 0.4)	0.2 (0, 0.4)	0 (0, 0.2)	< 0.001
wbc_idx1, median (Q1,Q3)	11.3 (7.3, 16.85)	10.7 (7.07, 15.65)	14.7 (9.1, 21.7)	< 0.001
lipase_idx1, median (Q1,Q3)	200 (67.5, 712.5)	213.5 (72.75, 724.25)	168 (61, 633)	0.123

Note: Data are presented as number (%) for categorical variables and median (Q1, Q3) for continuous variables. Group 0 indicates patients without septic shock; group 1 indicates patients with septic shock. p-Values are derived from chi-square test for categorical variables and Wilcoxon rank-sum test for continuous variables.

of the XGBoost model, thereby identifying the key factors associated with septic shock in acute pancreatitis patients.

In this study, we followed the SHAP guidelines (<https://github.com/slundberg/shap>) to interpret and visualize the XGBoost model. The SHAP histograms, SHAP bee swarm plots, and waterfall plots were generated. A higher absolute value of SHAP indicates a greater influence of the feature on the prediction outcome. As an example, we visualized the XGBoost model using SHAP histograms and SHAP waterfall plots for two patients: one who developed septic shock and one who did not.

Clinical and demographic characteristics

This study included 1032 patients who met the inclusion criteria (median age: 59 years, interquartile range (IQR):

46–72 years), with 559 males (54%). Hypertension was the most prevalent comorbidity (40%), followed by diabetes (30%). Among these patients, those who developed septic shock had significantly higher mortality rates at 7, 28, and 90 days compared to those who did not develop septic shock (Table 1).

Model development and validation

Using LASSO regression, we selected 8 important variables from 31 clinical features associated with the occurrence of sepsis. Based on these variables, we developed 10 ML models for prediction. After excluding overfitted models, we performed a comprehensive evaluation using ROC curves and DCA. The XGBoost model demonstrated the best

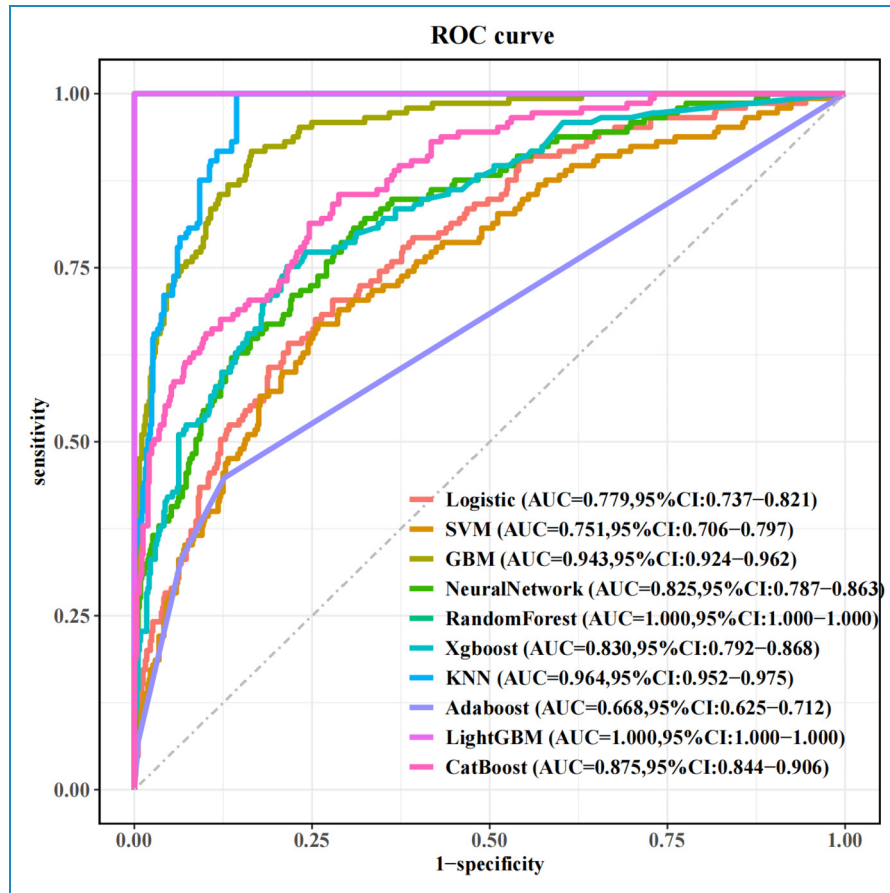


Figure 6. ROC curve of the training set. ROC: receiver operating characteristic.

performance, with an AUC of 0.83 on the test set and 0.78 on the validation set, significantly outperforming the other models. Therefore, XGBoost was selected as the final predictive model (Figures 6–9; Table 2).

In addition, this study employed additional performance metrics to evaluate model effectiveness, as detailed in Tables 3 and 4.

SHAP analysis and model interpretability

SHAP histograms and SHAP beeswarm plots were used to visualize the contributions of multiple features in the XGBoost model for predicting septic shock. The plots revealed that WBC showed a large positive SHAP value at higher levels, indicating a substantial contribution to the model's prediction. Specifically, higher WBC values were associated with an increased risk of septic shock. Additionally, higher values of total bilirubin and bicarbonate also corresponded to stronger positive SHAP values, further supporting the model's prediction of event occurrence. Conversely, features such as BUN and PT had smaller SHAP values, suggesting their limited contribution to the model. Notably, when the values of these features

were elevated, their SHAP values approached zero, indicating minimal impact on the prediction of septic shock (Figures 10 and 11).

We also demonstrated the interaction between different variables. The SHAP scatter plot visualized the interaction between WBC and albumin in the XGBoost model's prediction of septic shock. The plot showed that samples with low WBC and high albumin levels corresponded to SHAP values near zero, indicating that these features contributed minimally to the model's prediction, possibly reflecting a lower risk of septic shock. In contrast, when WBC was elevated and albumin levels were low, the SHAP values increased significantly, suggesting that this combination of features played a larger role in the model's prediction, highlighting the association between elevated WBC and low albumin levels with the occurrence of severe conditions such as septic shock (Figure 12).

The visualization results of the SHAP histogram and SHAP beeswarm plot demonstrate the performance of the XGBoost model. For a patient who developed septic shock, the SHAP value analysis reveals that the WBC had a SHAP value of +0.537, total bilirubin (bilirubin) had a SHAP value of +0.233, BUN had a SHAP value of +0.0696,

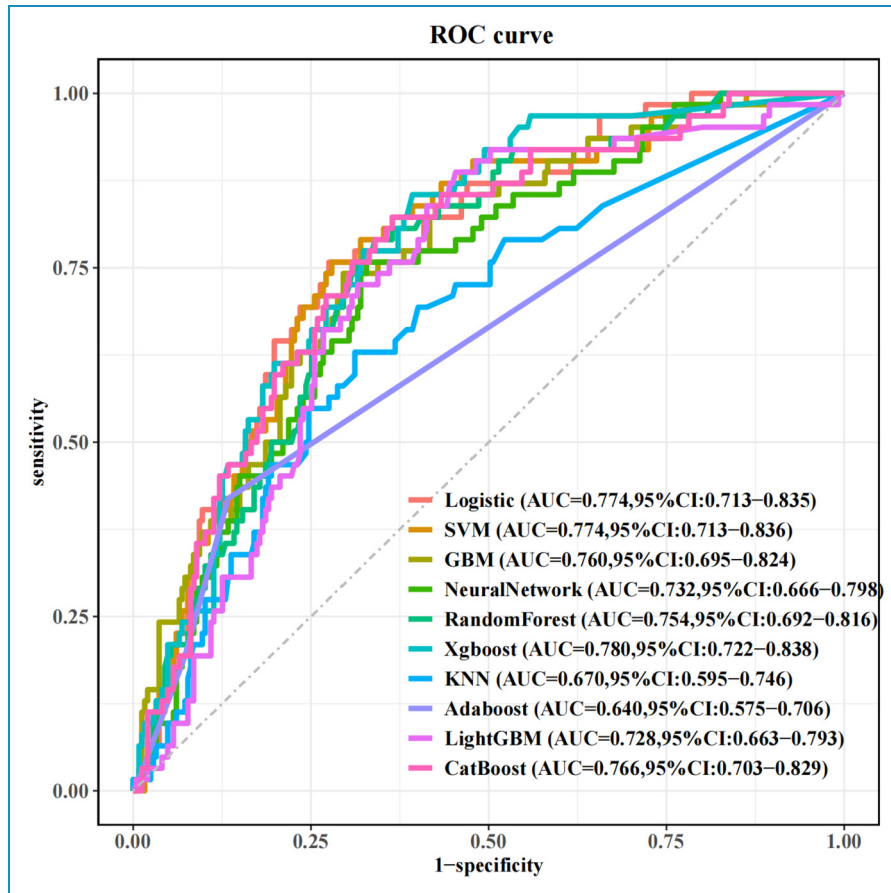


Figure 7. ROC curve of the test set. ROC: receiver operating characteristic.

and bicarbonate (HCO_3) had a SHAP value of +0.0661. Ultimately, the XGBoost model predicted a septic shock risk of 1 for this patient (Figure 13).

In contrast, for a patient who did not experience septic shock, the SHAP value analysis indicated that the SHAP values for WBC, total bilirubin (bilirubin), and PT were all -0.00197 . As a result, the XGBoost model predicted a risk $f(x)$ close to 0, demonstrating a strong predictive tendency of the model toward the absence of the event (Figure 14).

Discussion

This study extracts large-scale data from public databases and employs LASSO regression combined with 10 common ML algorithms to construct a predictive model for identifying the risk of septic shock in patients with acute pancreatitis. Additionally, the SHAP method was used to interpret and visualize the XGBoost model, further enhancing its clinical applicability. This approach aids clinicians in early identification of high-risk patients with AP, providing valuable support for subsequent medical decision making.

The findings indicate that the mortality rate in patients with AP complicated by septic shock is significantly higher than that in patients without septic shock. Early identification and intervention are thus critical for improving prognosis. Using the XGBoost model, the study identifies several key factors associated with septic shock in patients with AP, including WBC, total bilirubin (bilirubin), bicarbonate levels (bicarbonate), BUN, PT, lymphocyte count (lymphocytes), albumin (albumin), and severe liver disease (severe liver disease). SHAP histograms rank these eight features in order of their influence on the model: WBC, bilirubin, bicarbonate, BUN, PT, lymphocytes, albumin, and severe liver disease.

WBC is a critical marker of systemic inflammation. A significant increase in WBC suggests the presence of severe infection and systemic inflammatory response syndrome (SIRS), which are crucial risk factors for sepsis and septic shock.²¹ An elevated WBC count exacerbates the inflammatory response in the pancreas and surrounding tissues, damaging pancreatic microcirculation and ultimately leading to septic shock.²² SHAP values demonstrate that WBC is the most influential feature in predicting the risk of septic shock in AP.

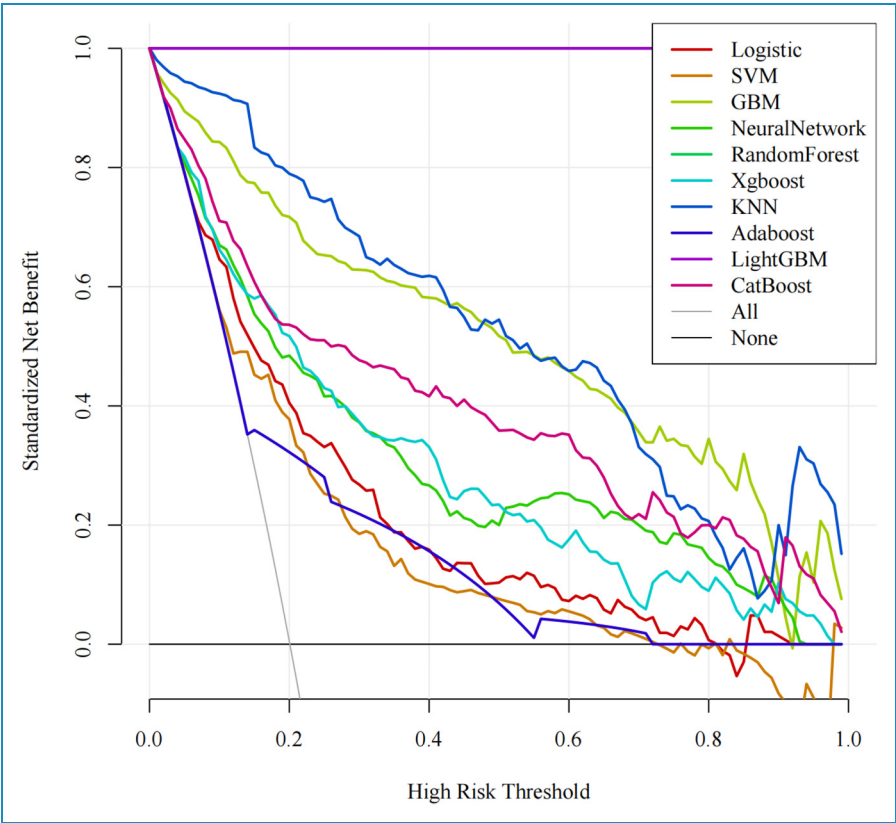


Figure 8. DCA curve of the training set. DCA: decision curve analysis.

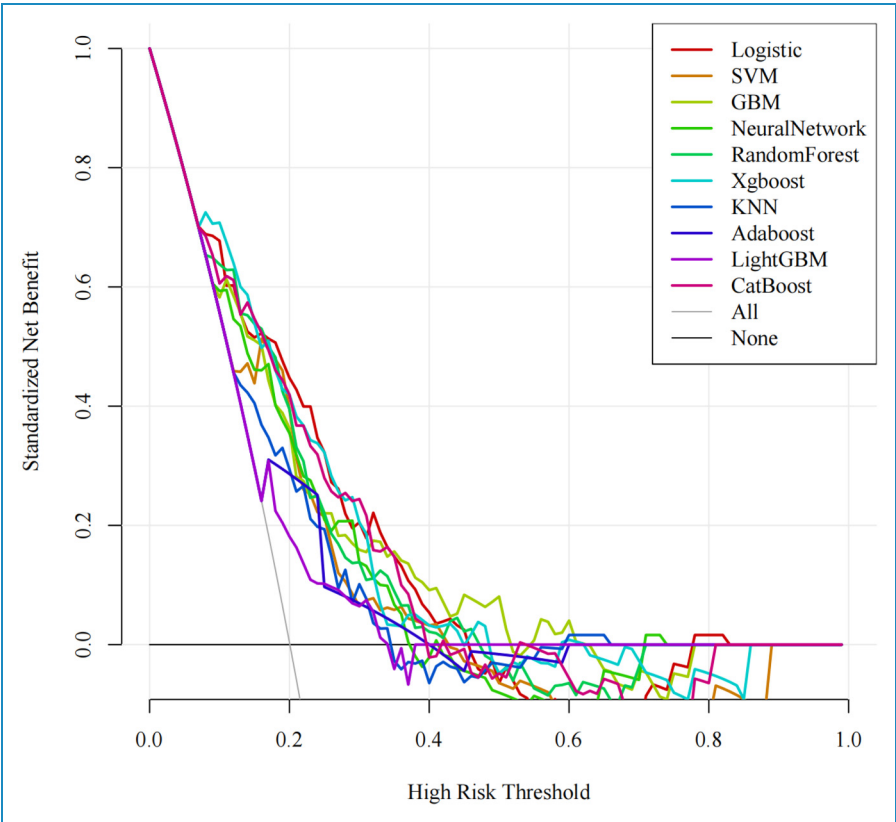


Figure 9. DCA curve of the test set. DCA: decision curve analysis.

Table 2. Discriminative performance (AUC $\pm$ SD) of different ML models in predicting septic shock in patients with acute pancreatitis.

ML model	Training set	Set of tests
Logistic	0.779 $\pm$ 0.042	0.774 $\pm$ 0.033
SVM	0.751 $\pm$ 0.043	0.774 $\pm$ 0.034
GBM	0.943 $\pm$ 0.018	0.760 $\pm$ 0.029
Neural Network	0.825 $\pm$ 0.036	0.732 $\pm$ 0.036
Random Forest	1.000 $\pm$ 0.000	0.754 $\pm$ 0.033
XGBoost	0.830 $\pm$ 0.036	0.780 $\pm$ 0.038
KNN	0.964 $\pm$ 0.009	0.670 $\pm$ 0.038
AdaBoost	0.668 $\pm$ 0.047	0.640 $\pm$ 0.031
LightGBM	1.000 $\pm$ 0.000	0.728 $\pm$ 0.038
CatBoost	0.875 $\pm$ 0.031	0.766 $\pm$ 0.031

Note: Performance was assessed using the AUC with SD in both training and testing sets. Ten ML models were evaluated, including traditional algorithms (e.g. Logistic Regression, SVM) and ensemble methods (e.g., XGBoost, CatBoost, LightGBM). AdaBoost: Adaptive Boosting; AUC: area under the receiver operating characteristic curve; CatBoost: Category Boosting; KNN: K-Nearest Neighbor; LightGBM: Light Gradient Boosting Machine; ML: machine learning; SD: standard deviation; SVM: Support Vector Machine; XGBoost: Extreme Gradient Boosting.

Table 3. Performance metrics of machine learning models on the training set.

Model	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1
NA	NA	NA	NA	NA	NA	NA
Logistic	0.226812051	0.755	0.641	0.783	0.427	0.512
SVM	0.195472222	0.727	0.669	0.742	0.394	0.496
GBM	0.212578475	0.852	0.917	0.835	0.583	0.713
Neural Network	0.178378867	0.715	0.807	0.692	0.397	0.532
Random Forest	0.5	1	1	1	1	1
XGBoost	0.497234672	0.778	0.752	0.785	0.468	0.577
KNN	0.236070832	0.885	1	0.856	0.636	0.777
AdaBoost	0.133247364	0.788	0.448	0.873	0.471	0.459
LightGBM	0.419310291	1	1	1	1	1
CatBoost	0.549922991	0.766	0.814	0.754	0.454	0.583

Note: The table reports threshold, accuracy, sensitivity, specificity, precision, and F1-score for each machine learning model trained to predict septic shock in acute pancreatitis patients. AdaBoost: Adaptive Boosting; CatBoost: Category Boosting; KNN: K-Nearest Neighbor; LightGBM: Light Gradient Boosting Machine; SVM: Support Vector Machine; XGBoost: Extreme Gradient Boosting.

An increase in total bilirubin is often associated with liver dysfunction or microcirculatory disturbances, both of which can worsen the course of septic shock. SHAP analysis positions bilirubin

as the second most important factor influencing the progression of acute pancreatitis.²³ The contribution of severe liver disease in the SHAP analysis further corroborates this finding.

Table 4. Performance metrics of machine learning models on the test set.

Model	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1
NA	NA	NA	NA	NA	NA	NA
Logistic	0.199074144	0.731	0.758	0.725	0.409	0.531
SVM	0.195127017	0.728	0.758	0.721	0.405	0.528
GBM	0.164909183	0.712	0.742	0.704	0.387	0.508
Neural Network	0.192524147	0.689	0.758	0.672	0.367	0.495
Random Forest	0.215	0.696	0.774	0.676	0.375	0.505
XGBoost	0.49674949	0.657	0.855	0.607	0.353	0.5
KNN	0.159921339	0.676	0.629	0.688	0.336	0.438
AdaBoost	0.133247364	0.777	0.419	0.866	0.441	0.43
LightGBM	0.007242025	0.615	0.887	0.547	0.329	0.48
CatBoost	0.540824952	0.673	0.823	0.636	0.362	0.502

Note: This table presents the performance of each machine learning model in predicting septic shock among patients with acute pancreatitis in the test set. Metrics include optimal threshold, accuracy, sensitivity, specificity, precision, and F1-score. XGBoost demonstrated the best overall sensitivity (0.855), supporting its utility in identifying high-risk patients. AdaBoost: Adaptive Boosting; CatBoost: Category Boosting; KNN: K-Nearest Neighbor; LightGBM: Light Gradient Boosting Machine; SVM: Support Vector Machine; XGBoost: Extreme Gradient Boosting.

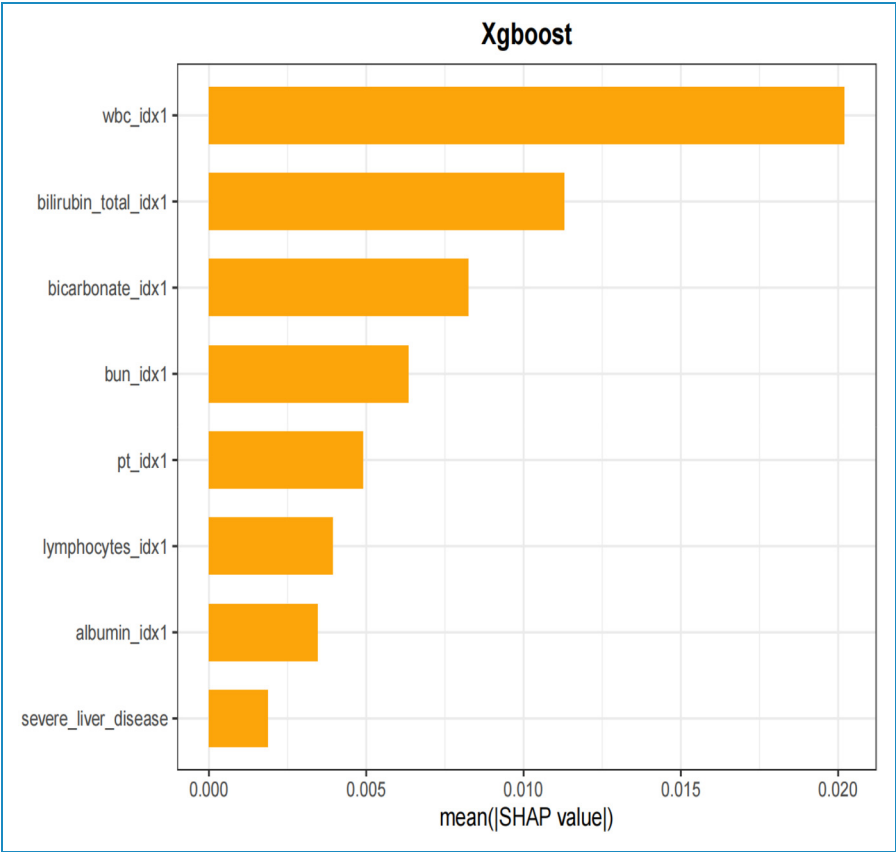


Figure 10. Histogram of SHAP for predicting. SHAP: Shapley Additive Explanations.

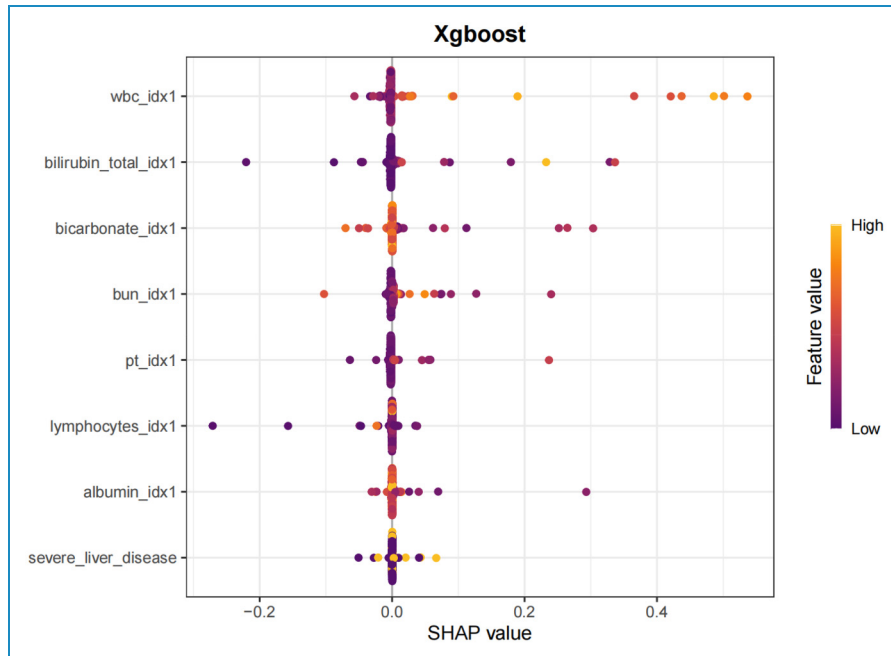


Figure 11. SHAP swarm plot for predicting characteristic variables of septic shock. SHAP: Shapley Additive Explanations.

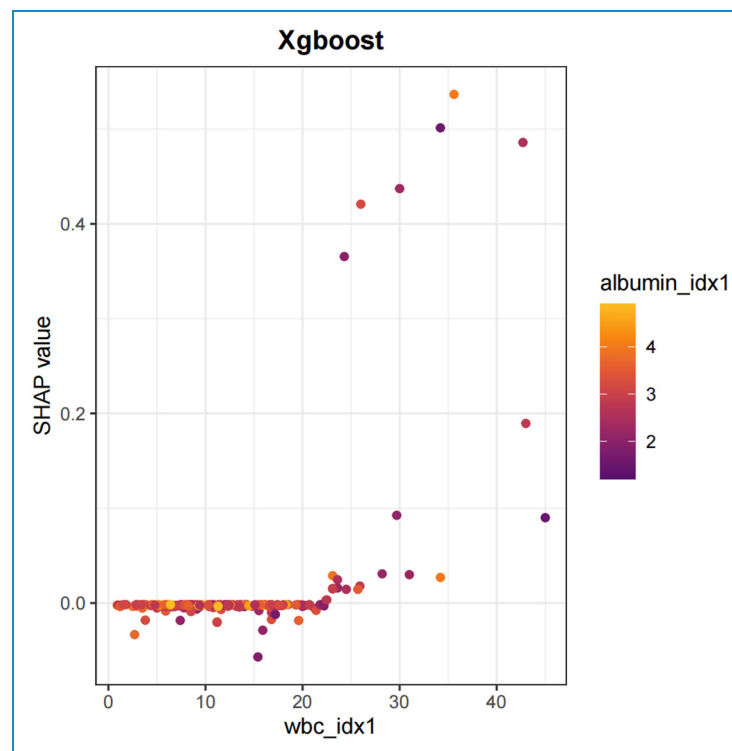


Figure 12. Scatter plot characteristic variables of septic shock.

Reduced bicarbonate levels are a hallmark of metabolic acidosis, commonly observed in patients with septic shock. Metabolic acidosis impairs cellular metabolism and may

exacerbate endothelial injury and oxygen delivery issues, thereby accelerating disease progression.^{24,25} The SHAP analysis confirms a strong positive correlation between

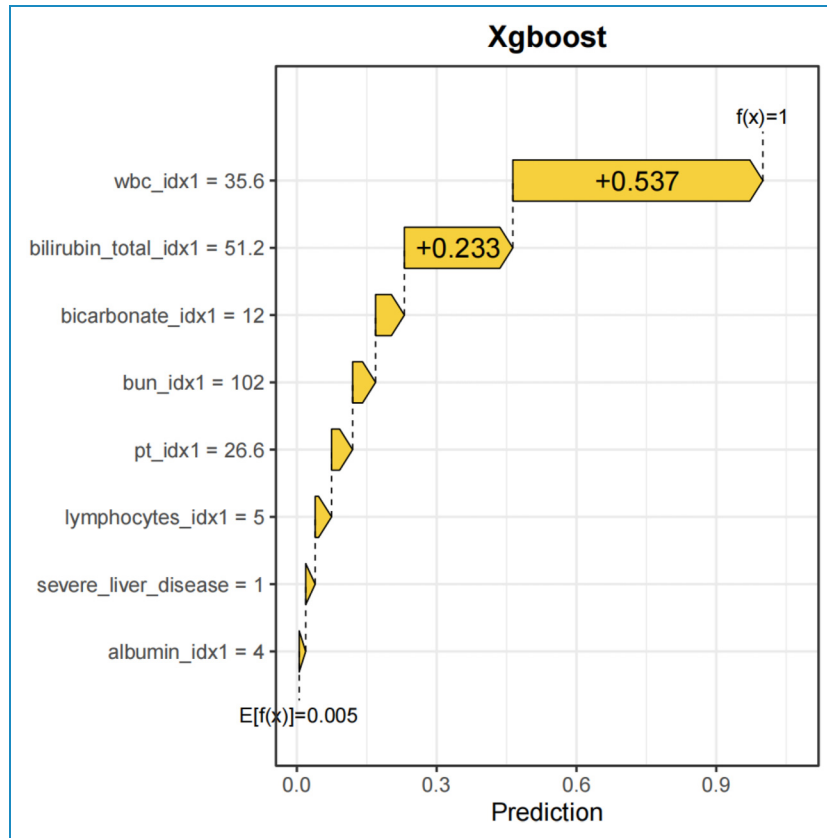


Figure 13. Visualization results of one XGBoost model with septic shock. XGBoost: Extreme Gradient Boosting.

low bicarbonate levels and the high risk of septic shock in patients with AP.²⁶

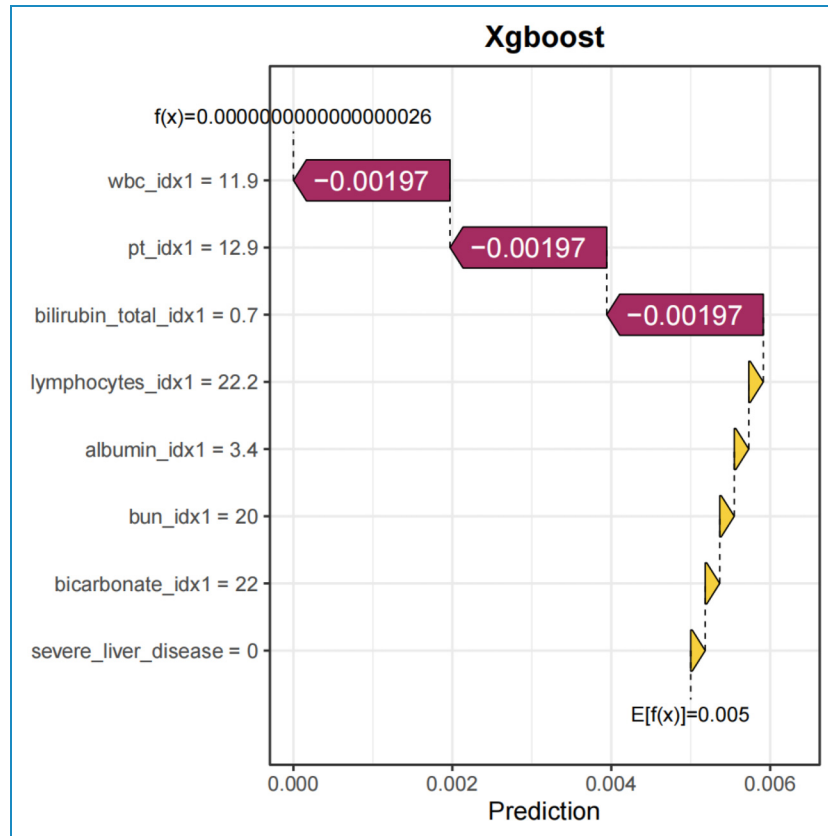
Elevated BUN reflects renal impairment or inadequate renal perfusion in patients with AP.²⁷ In the context of septic shock, microcirculatory dysfunction reduces glomerular filtration rate, leading to further accumulation of BUN and an increased risk of septic shock.²⁸ Prolonged PT is indicative of coagulation dysfunction, often seen in sepsis and septic shock, reflecting systemic inflammation, microthrombus formation, and endothelial damage. SHAP value analysis underscores the importance of both BUN and PT in predicting septic shock risk.²⁹ A reduction in lymphocyte count is indicative of sepsis-related immunosuppression. Lower lymphocyte levels are associated with impaired pathogen clearance, increasing the likelihood of infection spread and the risk of septic shock.³⁰ Similarly, albumin levels, linked to nutritional status and inflammatory burden, are inversely related to the development of septic shock. Low albumin levels worsen microcirculatory dysfunction and tissue hypoxia, potentially promoting the onset of septic shock.^{31,32}

Machine learning model performance

ML models offer significant advantages in identifying complex patterns within clinical data, improving prediction

accuracy and practical utility.³³ In this study, the performance of various models was compared using ROC and DCA curves, leading to the selection of the XGBoost model for further analysis.³⁴ The XGBoost model excels in several areas: (1) capturing high-order interactions between features and handling complex data relationships; (2) addressing discontinuous data and nonlinearity issues; and (3) reducing the impact of outliers through iterative weighting and optimization of the objective function. Previous studies have attempted to apply ML methods to predict complications associated with AP; however, most have not specifically targeted septic shock, and often lacked model interpretability. Liu et al.³⁵ developed an early prediction model for sepsis in ICU patients with AP using the MIMIC database, achieving an AUC of 0.985 in the test set. Nevertheless, their model did not distinguish patients at risk of septic shock and did not incorporate explainability techniques. Xia et al.³⁶ used a multi-center dataset of 604 patients from three hospitals in China to develop a model predicting 28-day septic shock, with the auto-encoder (AE) model reaching an AUC of 0.879—outperforming conventional scoring systems. However, the limited sample size and lack of explainable tools such as SHAP rendered the model a “black box,” restricting its clinical utility.

In contrast, our study is the first to construct an interpretable ML model for septic shock in patients with AP using



the emergency department data from the publicly available MIMIC-IV v3.0 database, ensuring high reproducibility and global applicability. We applied LASSO regression for feature selection and incorporated SHAP values to interpret the XGBoost model, thereby enhancing interpretability. SHAP summary plots revealed that eight key variables exhibited a “bipolar separation” pattern in SHAP values, indicating their strong influence on the predicted risk of septic shock.

In terms of predictive performance, the XGBoost model achieved an AUC of 0.780 in the validation cohort. To further visualize model predictions, we selected one patient who developed septic shock and one who did not. The XGBoost model estimated a risk probability close to 1 for the former and approximately 0 for the latter, providing an intuitive understanding of how the model distinguishes high-risk cases. These findings suggest that the SHAP-informed XGBoost model offers both strong predictive performance and enhanced interpretability in the early identification of septic shock risk among patients with AP.

Conclusion

This study demonstrates that ML models, particularly the SHAP-based interpretable XGBoost model, hold significant potential in predicting the risk of septic shock in patients

with AP. By incorporating clinical variables such as WBC, total bilirubin, and bicarbonate levels, the model offers a reliable and transparent method to assist emergency physicians in early risk stratification. The use of SHAP enhances the model's interpretability, providing insights into the contribution of each variable to the model's predictions. This approach offers a step forward in clinical decision support, aiding timely interventions and optimizing resource allocation in the management of AP-related septic shock. Future external validation across multiple centers will be essential to further assess the model's generalizability and clinical utility.

Limitations

Several limitations must be acknowledged in this study. First, the data used were derived from a single-center source, the MIMIC-IV database, which may limit the generalizability of the findings. Future research should aim to validate the model in multiple centers with diverse patient populations. Second, due to the high rate of missing data in the MIMIC-IV database, certain important clinical variables, such as pH levels and lactate, could not be included in the analysis. The exclusion of these variables may impact the model's predictive accuracy. Lastly, while our study

utilized a wide range of clinical parameters, additional factors such as imaging data, environmental factors, and lifestyle habits could further enhance the model's predictive power. These limitations should be addressed in future studies to improve the robustness of the model.

Ethics statement

The MIMIC-IV database is a publicly available dataset comprising de-identified patient information; therefore, no additional ethical approval is required for its use. Data extraction for this study was approved by the Institutional Review Board (IRB) of the Massachusetts Institute of Technology (MIT) (Certification No: 65828043, Named: Zhihao Li).

Acknowledgement

The authors would like to express their sincere gratitude to the editorial team and anonymous reviewers for their valuable feedback and insightful comments, which significantly improved the quality of this manuscript.

ORCID iD

Binglin Song  <https://orcid.org/0009-0008-8913-7172>

Ping Liu  <https://orcid.org/0009-0000-3563-9670>

Kangrui Fu  <https://orcid.org/0009-0001-7707-4767>

Author contributions

BS designed the study, performed the data analysis, and led the writing and revision of the manuscript. PL, KF, and CL contributed to data collection and processing, providing key technical support for the methodology. CL also provided overall guidance for the research, ensuring the scientific rigor and validity of the methods and conclusions.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

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

References

1. Matta B, Gougol A, Gao X, et al. Worldwide variations in demographics, management, and outcomes of acute pancreatitis. *Clin Gastroenterol Hepatol* 2020; 18: 1567–1575.e2.
2. Mederos MA, Reber HA and Girgis MD. Acute pancreatitis: a review. *JAMA* 2021; 325: 382–390.
3. Wang C, Zhang J, Liu L, et al. Early predictive value of pre-sepsis for secondary sepsis and mortality in intensive care unit patients with severe acute pancreatitis. *Shock* 2023; 59: 560–568.
4. Susak YM, Dirda OO, Fedorchuk OG, et al. Infectious complications of acute pancreatitis is associated with peripheral blood phagocyte functional exhaustion. *Dig Dis Sci* 2021; 66: 121–130.
5. Feng A, Ao X, Zhou N, et al. A novel risk-prediction scoring system for sepsis among patients with acute pancreatitis: a retrospective analysis of a large clinical database. *Int J Clin Pract* 2022; 2022: 5435656.
6. Mikó A, Vigh É, Mátrai P, et al. Computed tomography severity index vs. other indices in the prediction of severity and mortality in acute pancreatitis: a predictive accuracy meta-analysis. *Front Physiol* 2019; 10: 1002.
7. Kuo DC, Rider AC, Estrada P, et al. Acute pancreatitis: what's the score? *J Emerg Med* 2015; 48: 762–770.
8. Stirling AD, Moran NR, Kelly ME, et al. The predictive value of C-reactive protein (CRP) in acute pancreatitis—is interval change in CRP an additional indicator of severity? *HPB (Oxford)* 2017; 19: 874–880.
9. Khwannimit B, Bhurayanontachai R and Vattanavanit V. Comparison of the accuracy of three early warning scores with SOFA score for predicting mortality in adult sepsis and septic shock patients admitted to intensive care unit. *Heart Lung* 2019; 48: 240–244.
10. Van der Woude S, Van Doormaal F, Hutten B, et al. Classifying sepsis patients in the emergency department using SIRS, qSOFA or MEWS. *Neth J Med* 2018; 76: 158–166.
11. Hirota M, Shimosegawa T, Masamune A, et al. The sixth nationwide epidemiological survey of chronic pancreatitis in Japan. *Pancreatol* 2012; 12: 79–84.
12. Verma AA, Murray J, Greiner R, et al. Implementing machine learning in medicine. *CMAJ* 2021; 193: E1351–E1357.
13. Guan C, Ma F, Chang S, et al. Interpretable machine learning models for predicting venous thromboembolism in the intensive care unit: an analysis based on data from 207 centers. *Crit Care* 2023; 27: 406.
14. Wiggerthale J and Reich C. Explainable machine learning in critical decision systems: ensuring safe application and correctness. *AI* 2024; 5: 2864–2896.
15. Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Network Open* 2021; 4: e212240–e212240.
16. Johnson AE, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 2023; 10: 1.
17. 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: 102–111.
18. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315: 762–774.

19. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Stat Methodol* 1996; 58: 267–288.
20. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018; 2: 749–760.
21. Grondman I, Pirvu A, Riza A, et al. Biomarkers of inflammation and the etiology of sepsis. *Biochem Soc Trans* 2020; 48: 1–14.
22. Staubli SM, Oertli D and Nebiker CA. Laboratory markers predicting severity of acute pancreatitis. *Crit Rev Clin Lab Sci* 2015; 52: 273–283.
23. Patel JJ, Taneja A, Niccum D, et al. The association of serum bilirubin levels on the outcomes of severe sepsis. *J Intensive Care Med* 2015; 30: 23–29.
24. Guarino M, Perna B, Cesaro AE, et al. 2023 Update on sepsis and septic shock in adult patients: management in the emergency department. *J Clin Med* 2023; 12: 3188.
25. James JH, Luchette FA, McCarter FD, et al. Lactate is an unreliable indicator of tissue hypoxia in injury or sepsis. *Lancet* 1999; 354: 505–508.
26. Matyukhin I, Patschan S, Ritter O, et al. Etiology and management of acute metabolic acidosis: an update. *Kidney Blood Pressure Res* 2020; 45: 523–531.
27. Wu BU, Bakker OJ, Papachristou GI, et al. Blood urea nitrogen in the early assessment of acute pancreatitis: an international validation study. *Arch Intern Med* 2011; 171: 669–676.
28. Beier K, Eppanapally S, Bazick HS, et al. Elevation of blood urea nitrogen is predictive of long-term mortality in critically ill patients independent of “normal” creatinine. *Crit Care Med* 2011; 39: 305–313.
29. Schupp T, Weidner K, Rusnak J, et al. Diagnostic and prognostic significance of the prothrombin time/international normalized ratio in sepsis and septic shock. *Clin Appl Thromb Hemost* 2022; 28: 10760296221137893.
30. Delano MJ and Ward PA. The immune system’s role in sepsis progression, resolution, and long-term outcome. *Immunol Rev* 2016; 274: 330–353.
31. Turcato G, Zaboli A, Sibilio S, et al. Albumin as a prognostic marker of 30-day mortality in septic patients admitted to the emergency department. *Intern Emerg Med* 2023; 18: 2407–2417.
32. Barbosa-Silva MCG. Subjective and objective nutritional assessment methods: what do they really assess? *Curr Opin Clin Nutr Metab Care* 2008; 11: 248–254.
33. Chen T and Guestrin C. *XGBoost: A scalable tree boosting system*. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp.785–794.
34. Sagi O and Rokach L. Approximating XGBoost with an interpretable decision tree. *Inf Sci (Ny)* 2021; 572: 522–542.
35. Liu F, Yao J, Liu C, et al. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis. *BMC Surg* 2023; 23: 267.
36. Xia Y, Long H, Lai Q, et al. Machine learning predictive model for septic shock in acute pancreatitis with sepsis. *J Inflamm Res* 2024; 1443–1452. doi:10.2147/JIR.S441591