Cite as: Ren X, Liu LJ, Zha JJ, Ye Y, Xu XD, Ye HW, Zhang Y. Development of a prediction model for acute respiratory distress syndrome in ICU patients with acute pancreatitis based on machine learning algorithms [J]. Chin J Clin Res, 2025, 38(8):1173-1181.

DOI: 10.13429/j.cnki.cjcr.2025.08.008

Development of a prediction model for acute respiratory distress syndrome in ICU patients with acute pancreatitis based on machine learning algorithms

REN Xia*, LIU Luojie, ZHA Junjie, YE Ye, XU Xiaodan, YE Hongwei, ZHANG Yan

**Department of Gastroenterology, Changshu No.1 People's Hospital, Changshu Hospital Affiliated to Soochow University, Changshu, Jiangsu 215500, China*

Corresponding author: ZHANG Yan, E-mail: 1429862753@qq.com

Abstract: Objective To develop and validate a predictive model based on machine learning algorithms to assess the risk of acute respiratory distress syndrome (ARDS) in patients with acute pancreatitis (AP) admitted to the intensive care unit (ICU) . **Methods** The relevant data of 857 AP patients from the Medical Information Mart for Intensive Care IV v2.2 (MIMIC-IV v2.2) database were retrospectively analyzed and were randomly divided into a training set (n=601) and an internal validation set (n=256) in a 7∶3 ratio. Additionally, The relevant data of 126 AP patients from the ICU of Changshu Hospital Affiliated to Soochow University from January 2019 to March 2024 were collected as an external test set. Patients were categorized into ARDS and non - ARDS groups based on the occurrence of ARDS. Demographic characteristics, initial vital signs, laboratory data, functional scores, and complications within the initial 24-hour of ICU admission were collected. Feature selection was performed using least absolute shrinkage and selection operator (LASSO) regression. Predictive models were constructed using seven machine learning algorithms: random forest (RF) , extreme gradient boosting (XGBoost) , light gradient boosting machine (LightGBM) , decision tree (DT) , logistic regression (LR) , support vector machine (SVM) , and K-nearest neighbors (KNN) . Model performance was evaluated using receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA) . Finally, model interpretability was enhanced through Shapley additive explanations (SHAP) analysis. **Results** In the MIMIC-IV database, 202 patients (23.57%) developed ARDS, while 26 patients (20.63%) developed ARDS in the external test set. Seven key variables were selected by LASSO regression from 43 variables in the training set to construct the models. Among various machine learning models, the RF model demonstrated the best performance with an area under the curve (AUC) of 0.780 (95%CI:0.721-0.846) in the internal validation set and 0.842 (95%CI:0.751- 0.917) in the external test set, outperforming the other six models. The calibration curve indicated that the predicted probabilities from the RF model had the smaller deviation from the actual probabilities compared to other models, showing the best overall predictive performance. SHAP analysis based on the RF model revealed that mechanical ventilation, Sequential Organ Failure Assessment (SOFA) score, body mass index (BMI) , peripheral oxygen saturation (SpO_2) and simplified acute physiology score (SAPS Ⅱ) were the main factors influencing ARDS risk. Mechanical ventilation increased the risk of ARDS from 16% to 37% . When the SOFA score exceeded 8, the ARDS risk rose significantly. The risk of ARDS elevated with increased BMI. While SpO_2 remained below 90%, ARDS risk stabilized at 30%; once SpO_2 surpassed 90%, the risk demonstrated a declining trend with further increases in SpO_2. For SAPS-II scores between 46 and 60, ARDS risk showed a pronounced upward trend. **Conclusion** The RF predictive model provides a reliable tool for assessing the risk of ARDS in AP patients and enhances model interpretability through the SHAP method, aiding in clinical decision-making.

Keywords: Acute pancreatitis; Acute respiratory distress syndrome; Intensive care medicine; Machine learning; Random forest; Shapley additive explanations

Fund program: Suzhou Science and Technology Development Plan Project (SLT2023006) ;Key Project of Changshu Science and Technology Development Plan (CSWS202209) ;Respiratory Disease Special Project of China International Medical Exchange Foundation (Z-2014-08-2309-1)

Acute pancreatitis (AP) is a common acute abdominal condition characterized by acute inflammation of the pancreas. Its incidence has been gradually increasing worldwide, with an estimated 13 to 49 cases per 100,000 people each year [1]. Over the past decade, the number of hospitalizations due to AP has increased by approximately 30%, and the annual healthcare costs related to this in the United States alone have reached $2.6 billion [2]. The overall mortality rate for AP is 1% to 5%, but approximately 20% of patients with mild AP may progress

to moderate severe acute pancreatitis (MSAP) or severe acute pancreatitis (SAP), accompanied by pancreatic necrosis and persistent organ dysfunction. In these cases, the mortality rate can rise to 10% to 50% [3-5]. Acute respiratory distress syndrome (ARDS) is often the first organ dysfunction to occur in SAP patients, with an incidence as high as 30%. It is also a major cause of early death in AP patients [6-8]. Timely identification of high-risk patients who may develop ARDS helps to reduce the severity of the disease, improve prognosis, and lower mortality.

With the rapid development of artificial intelligence (AI) technology, machine learning has become increasingly popular as a data analysis method due to its powerful pattern recognition capabilities and ability to handle complex data structures. In recent years, the application of machine learning in the field of AP has been growing, demonstrating good predictive performance in disease diagnosis [9], severity assessment [10], and complication prediction [11-12]. However, machine learning models are often referred to as "black-box" models due to their complex internal structures and decision-making processes. Current research on AP machine learning primarily focuses on model prediction accuracy, while overlooking the high demand for interpretability in the medical field. This limitation hinders the generalization ability of these models and restricts their application in disease diagnosis and prediction [10]. Lundberg *et al.* [13] proposed the Shapley additive explanations (SHAP) method based on Shapley values from game theory, which quantifies the contribution of each feature to the prediction result and provides an intuitive visual display.

Therefore, this study aims to use multicenter data to train and validate a series of models based on machine learning algorithms [including random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), decision tree (DT), logistic regression (LR), support vector machine (SVM), and K-nearest neighbors (KNN)], and perform least absolute shrinkage and selection operator (LASSO) regression analysis to select the best-performing model. The goal is to provide a visual explanation of the model for predicting the risk of ARDS in ICU patients with AP, assisting clinicians in making timely intervention decisions.

1 Objects and Methods

1.1 Data Source

This study retrospectively extracted data from the Medical Information Mart for Intensive Care v2.2 (MIMIC-IV v2.2) database (Dataset 1), which is a large, single-center, de-identified public database developed by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology (MIT). It contains detailed clinical data from 53,150 patients who were admitted to the ICU at the Beth Israel Deaconess Medical Center (BIDMC) in Boston between 2008 and 2019. This study has been approved through the NIH Collaborative Training Program (Certificate Number: 60227322), and the database access was granted. Additionally, 126 patients with MSAP and SAP who were admitted to the ICU at the Changshu Hospital Affiliated to Soochow University between January 2019 and March 2024 were selected as an external test set (Dataset 2). To ensure the protection of patient information, data cleaning was performed. This study was approved by the Ethics Committee of Changshu No.1 People's Hospital (Approved No. L202402).

1.2 Study Population

Patients admitted with a diagnosis of AP were selected from the MIMIC-IV database using the International Classification of Diseases codes ICD-9 (577.0) and ICD-10 (K85%). The inclusion criteria for the external test set population adhered to the *Guidelines for Diagnosis And Treatment of Acute Pancreatitis in China (2021)* [14]. The diagnostic criteria for ARDS followed the Berlin Definition [15]. Exclusion criteria included: age less than 18 years; ICU stay duration less than 24 hours; and pre-existing respiratory failure prior to ICU admission. The specific screening process is illustrated in **Figure 1.**

1.3 Data Extraction and Processing

Case data extraction from the public database was performed using SQL query language within pgAdmin 4 software. All clinical and laboratory variables were data collected within the first 24 hours after ICU admission. For variables with multiple measurements, only the first measured result was included in this study. A total of 50 variables were extracted from the database, including:

(1) Demographic characteristics: age, gender, race, height, body mass, insurance type;

(2) Vital signs: heart rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate (RR), temperature, oxygen saturation (SpO_2);

(3) Laboratory data: white blood cell count (WBC), hemoglobin (Hb), platelet count (PLT), albumin, hematocrit, creatinine (SCr), C-reactive protein, alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), glucose, Na, K, Cl, Ca, prothrombin time (PT), international normalized ratio (INR), pH, partial pressure of carbon dioxide (PCO_2), base excess (BE), lactate, bicarbonate (HCO_3^-);

(4) Functional scores: Glasgow Coma Scale (GCS) score, Sequential Organ Failure Assessment (SOFA) score, Simplified Acute Physiology Score II (SAPS II);

(5) Comorbidities: hypertension, diabetes, sepsis, myocardial infarction, chronic obstructive pulmonary disease (COPD), acute kidney injury;

(6) Treatment measures: continuous renal replacement therapy (CRRT), mechanical ventilation, cardiopulmonary resuscitation (CPR), heparin, aspirin, antibiotics, and vasoactive drugs.

Variables collected by Changshu Hospital Affiliated to Soochow University were identical to those listed above.

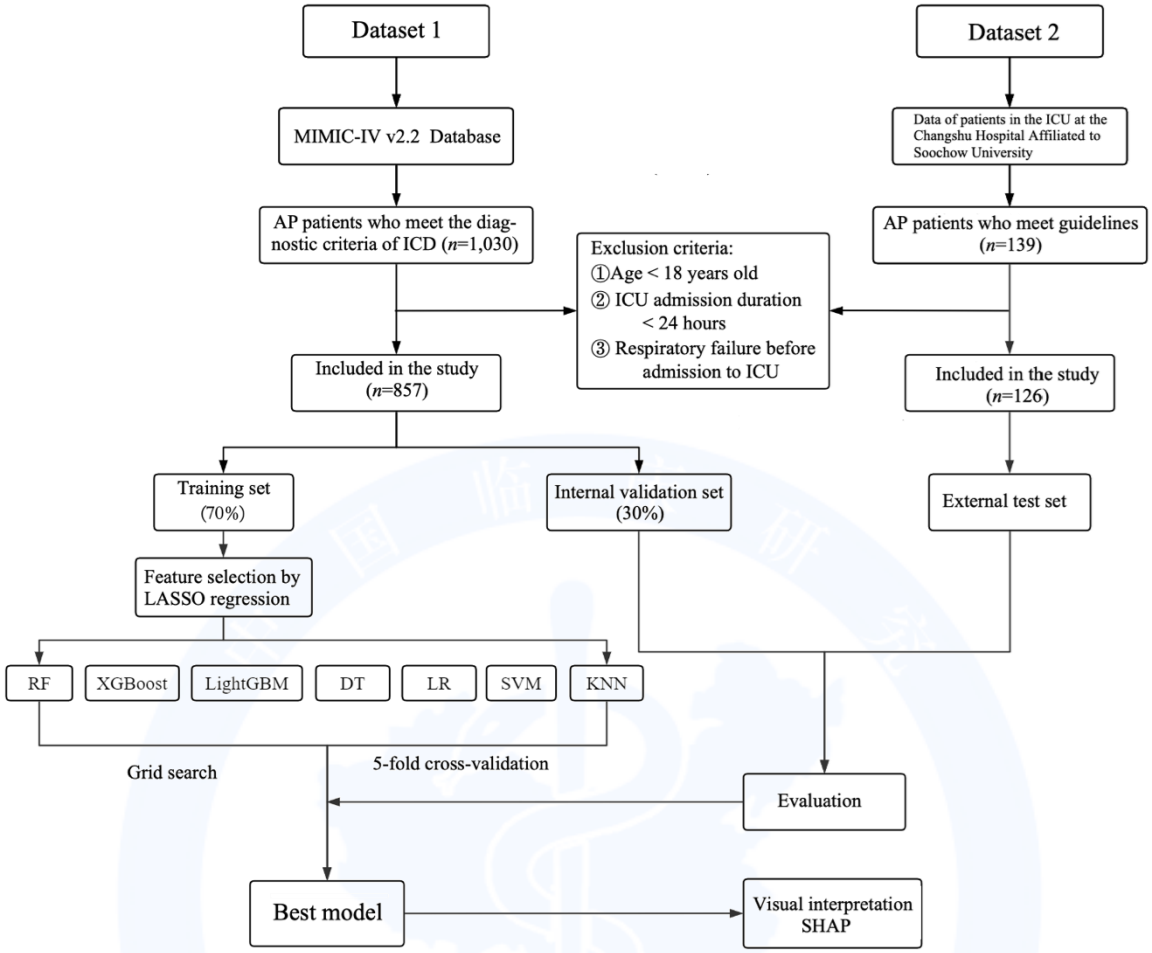


Fig.1 Research flowchart

For data missing less than 5% of values, mean imputation was used. For data missing 5% to 30% of values, multiple imputation was applied. Variables with missing values exceeding 30% were excluded from the analysis to minimize bias. To reduce multicollinearity among variables, Spearman rank correlation coefficients were calculated, and variables with coefficients $|r| > 0.8$ (BUN, hematocrit, AST, PT) were removed. A total of 43 variables were ultimately included in the analysis.

1.4 Statistical Methods

Data analysis was conducted using Stata 16.0, R 4.3.2, and Python 3.12.4 software. The processed MIMIC-IV dataset was split using simple random sampling, with 70% allocated to the training set and 30% to the internal validation set. Data from the Changshu Hospital Affiliated to Soochow University was used as the external test set. The $\bar{x} \pm s$ was used to describe the normally distributed continuous data, and independent sample *t*-tests were applied for group comparisons. For skewed continuous data, the $M(Q_L,Q_U)$ was used, and the Mann-Whitney *U* test was employed for inter-group comparisons. Categorical variables were expressed as *n* (%), with the chi-square test used for comparisons. Feature selection was performed using LASSO regression. $P<0.05$ was considered statistically significant.

1.5 Model Construction and Validation

LASSO regression was used to select variables after dimensionality reduction. Seven machine learning algorithms were employed to construct prediction models in the training cohort, including RF, XGBoost, LightGBM, DT, LR, SVM, and KNN. Five-fold cross-validation was applied during model construction to prevent overfitting, and grid search was used to adjust hyperparameters. The receiver operating characteristic (ROC) curve was used as the evaluation metric for the models, and the area under the curve (AUC), accuracy, sensitivity, specificity, and F1 score (the harmonic mean of precision and recall) were calculated. A calibration curve was plotted for model consistency analysis, and decision curve analysis (DCA) was applied to assess the clinical net benefit of the model.

1.6 Model Visual Interpretation

The SHAP method was used to reveal the importance of individual features and the interactions between different features by calculating the SHAP values for each feature variable. Based on the aforementioned evaluation metrics, the best diagnostic model was selected. Using

Python 3.12.4 software, SHAP feature importance plots, bee swarm plots, force plots, and partial dependence plots (PDP) for key variables were created to provide a visual interpretation of the model.

2 Results

2.1 Baseline Characteristics

A total of 983 AP patients were included in this study, with 857 patients from the MIMIC-IV database and 202 (23.57%) with concurrent ARDS. The external test set included 126 patients from the Changshu Hospital Affiliated to Soochow University, with 26 (20.63%) diagnosed with ARDS. The ARDS incidence rate was similar between the two groups. Baseline characteristics of both datasets are shown in **Table 1**. There was no statistically significant difference in demographic features, vital signs, laboratory data, functional scores, complications, or treatment measures between the training set and internal validation set (P>0.05), as shown in **Table 2.**

Tab.1 Comparison of baseline data between ARDS group and non-ARDS group of AP patients in different datasets

Variable	MIMIC Database				External test set			
	Non-ARDS group (n=655)	ARDS group (n=202)	$Z/\chi^2/t$ value	P value	Non-ARDS group (n=100)	ARDS group (n=26)	$Z/\chi^2/t$ value	P value
Demographic Characteristics								
Age [years ,$M(Q_L,Q_U)$]	61(47,76)	58(46,69)	2.230	0.026	52(40,66)	44(34,56)	1.595	0.111
Male [n(%)]	372(56.79)	125(61.88)	1.438	0.230	50(50.00)	15(57.69)	0.229	0.632
BMI [kg/m^2, $M(Q_L,Q_U)$]	27.69(23.96,32.34)	31.12(26.90,37.04)	5.735	<0.001	20.91(19.38,24.10)	23.50(21.06,27.53)	2.607	0.009
Medical Insurance [n(%)]	222(33.89)	62(30.69)	0.576	0.448	35(35.00)	12(46.15)	0.673	0.412
Vital Signs								
HR [beats/min, $M(Q_L,Q_U)$]	95(81,112)	103.5(88,119)	3.535	<0.001	100(80,116)	109(93,124)	1.707	0.088
DBP(mmHg, $\bar{x} \pm s$)	128.90 ± 24.30	121.61±27.82	3.599	<0.001	127.34 ± 25.06	120.04 ± 23.05	1.413	0.165
SBP(mmHg, $\bar{x} \pm s$)	73.57 ± 18.47	70.61±19.76	1.954	0.051	71.23 ± 17.96	67.77 ± 16.76	0.924	0.361
RR [beats/min, $M(Q_L,Q_U)$]	20(16,24)	22(18,28)	3.338	<0.001	19(15,24)	25(17,28)	2.356	0.019
Temperature [℃, $M(Q_L,Q_U)$]	36.9(36.6,37.2)	37.0(36.5,37.4)	2.252	0.074	36.8(36.6,37.2)	37.0(36.5,37.4)	0.946	0.346
SpO_2 [%, $M(Q_L,Q_U)$]	97(94,99)	96(94,99)	2.570	0.024	96(94,98)	95(92,98)	1.283	0.201
Laboratory Tests [$M(Q_L,Q_U)$]								
WBC(×10^9/L)	11.9(7.9,17.0)	13.7(9.9,20.0)	3.461	<0.001	10.6(7.3,16.6)	12.9(10.1,22.4)	2.586	0.010
Hb(g/dL, $\bar{x} \pm s$)	11.30 ± 2.30	11.49±2.59	0.423	0.338	11.45 ± 2.10	11.20 ± 2.32	0.497	0.622
PLT(×10^9/L)	186(132,254)	190(127,249)	0.208	0.836	174(132,234)	205(137,225)	0.805	0.423
Albumin (g/L)	3.1(2.7,3.5)	2.8(2.4,3.2)	6.047	<0.001	37.7 ± 5.2	32.0 ± 7.1	3.834	< 0.001
Creatinine (mg/dL)	1.0(0.7,1.5)	1.2(0.9,2.1)	4.482	<0.001	1.0(0.7,1.5)	1.7(1.0,2.6)	2.530	0.011
ALT(u/L)	59(26,127)	42(23,118)	1.763	0.078	84(34,147)	45(24,130)	1.182	0.238
Glucose (mg/dL)	125(102,161)	134(104,173)	1.515	0.130	133(105,165)	128(107,197)	0.021	0.986
Na^+(mmol/L)	138(135,141)	139(135,143)	2.570	0.010	138(135,141)	139(137,143)	1.673	0.095
K^+(mmol/L)	4.0(3.6,4.4)	4.1(3.7,4.7)	2.579	0.010	4.0(3.6,4.4)	4.0(3.5,4.6)	0.042	0.969
Cl^-(mmol/L)	104(100,108)	105(101,111)	3.220	0.001	104(100,108)	108(102,112)	2.202	0.028
INR	1.3(1.1,1.5)	1.3(1.2,1.6)	2.911	0.005	1.2(1.1,1.4)	1.4(1.2,1.6)	2.232	0.026
pH	7.38(7.32,7.43)	7.33(7.24,7.41)	5.251	<0.001	7.38(7.31,7.43)	7.33(7.21,7.38)	2.208	0.027
PCO_2(mmHg)	39(34,43)	42(35,48)	3.885	<0.001	38(34,42)	40(32,49)	0.673	0.503
BE(mmol/L)	-1(-5,1)	-3(-8,0)	3.133	0.002	-2(-6,0)	-3(-10,0)	1.294	0.197
Lactic Acid (mmol/L)	1.7(1.2,2.5)	1.9(1.3,3.0)	2.216	0.027	1.6(1.2,2.2)	2.3(1.3,3.3)	1.678	0.094
HCO_3^-(mmol/L)	22(18,25)	21(17,25)	1.781	0.075	22(18,25)	19(16,23)	1.353	0.177
Function Score [$M(Q_L,Q_U)$]								
GCS	15(14,15)	15(14,15)	0.164	0.870	15(14,15)	15(14,15)	0.345	0.075
SOFA	4(2,7)	8(5,12)	10.342	<0.001	4(2,7)	9(5,13)	3.356	< 0.001
SAPS II	31(22,42)	43(31,58)	8.751	<0.001	32(23,42)	48(38,62)	4.269	< 0.001
Complication [n(%)]								
Hypertension	313(47.79)	91(45.05)	0.361	0.548	41(41.00)	10(38.46)	<0.001	0.991
Diabetes	57(8.70)	17(8.42)	0.016	0.899	12(12.00)	1(3.85)	1.483	0.223
Sepsis	356(54.35)	169(83.66)	54.667	<0.001	35(35.00)	16(61.54)	4.981	0.026
Myocardial Infarction	10(1.53)	0	3.120	0.077	1(1.00)	0(0)	0.262	0.609
COPD	18(2.75)	8(3.96)	0.414	0.520	3(3.00)	2(7.69)	1.192	0.274
Acute Kidney Injury	390(59.54)	172(85.15)	43.718	<0.001	58(58.00)	17(65.38)	0.211	0.646
Treatment Measures [n(%)]								
CRRT	56(8.55)	56(27.72)	48.282	<0.001	32(32.00)	15(57.69)	4.777	0.029
Mechanical Ventilation	144(21.98)	126(62.38)	114.857	<0.001	32(32.00)	11(42.31)	0.571	0.450
Cardiopulmonary Resuscitation	9(1.37)	4(1.98)	0.380	0.538	1(1.00)	0	0.262	0.609
Heparin	595(90.84)	194(96.04)	5.025	0.025	93(93.00)	24(92.31)	0.015	0.903
Aspirin	212(32.37)	64(31.68)	0.009	0.924	28(28.00)	9(34.62)	0.175	0.676
Antibiotics	543(82.90)	197(97.52)	26.782	<0.001	79(79.00)	26(100)	6.552	0.010
Vasoactive Drugs	116(17.71)	93(46.04)	65.667	<0.001	20(20.00)	15(57.69)	12.795	< 0.001

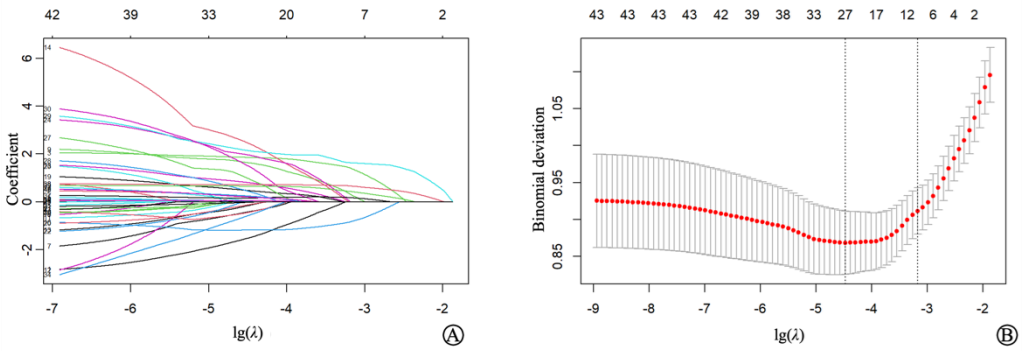
Tab.2 Comparison of baseline data of AP patients between training set and internal validation set

Variable	Internal validation set (n=256)	Training set (n=601)	$Z/\chi^2/t$ value	P value
Demographic Characteristics				
Age [years ,$M(Q_L,Q_U)$]	63(47,75)	59(47,73)	1.013	0.311
Male [n(%)]	144(56)	353(59)	0.359	0.549
BMI [kg/m^2, $M(Q_L,Q_U)$]	28.20(24.32,33.60)	28.40(24.34,33.33)	0.161	0.869
Medical Insurance [n(%)]	97(37.89)	187(31.15)	3.420	0.064
Vital Signs				
HR [beats/min, $M(Q_L,Q_U)$]	95(81,113)	97(83,114)	1.143	0.253
DBP(mmHg, $\bar{x} \pm s$)	126.09 ± 25.01	127.57 ± 25.50	0.678	0.498
SBP(mmHg, $\bar{x} \pm s$)	72.54 ± 18.68	73.01 ± 18.88	0.342	0.733
RR [beats/min, $M(Q_L,Q_U)$]	20.5(16,25)	20(17,25)	0.416	0.678
Temperature [℃, $M(Q_L,Q_U)$]	36.8(36.6,37.2)	36.9(36.6,37.3)	0.857	0.391
SpO_2 [%, $M(Q_L,Q_U)$]	96(94,99)	97(94,99)	0.778	0.436
Laboratory Tests [$M(Q_L,Q_U)$]				
WBC(×10^9/L)	12.0(8.6,17.6)	12.4(8.2,17.8)	0.152	0.892
Hb(g/dL, $\bar{x} \pm s$)	11.32 ± 2.39	11.35 ± 2.37	0.036	0.972
PLT(×10^9/L)	192(139,261)	185(127,250)	1.200	0.230
Albumin (g/L)	3.1(2.6,3.5)	3(2.6,3.5)	0.084	0.933
Creatinine (mg/dL)	1.1(0.7,1.7)	1(0.7,1.7)	0.699	0.485
ALT(u/L)	61(26,129)	54(25,118)	1.076	0.282
Glucose (mg/dL)	125(106,158)	127(101,167)	0.014	0.989
Na^+(mmol/L)	138(135,141)	138(135,141)	1.380	0.168
K^+(mmol/L)	4.0(3.6,4.4)	4.0(3.6,4.5)	0.243	0.808
Cl^-(mmol/L)	104(100,108)	104(100,109)	1.283	0.200
INR	1.3(1.1,1.4)	1.3(1.1,1.5)	1.283	0.200
pH	7.37(7.3,7.42)	7.37(7.30,7.43)	0.211	0.833
PCO_2(mmHg)	39(34,46)	39(34,44)	0.936	0.349
BE(mmol/L)	-2(-6,0)	-2(-6,1)	0.144	0.885
Lactic Acid (mmol/L)	1.7(1.1,2.8)	1.7(1.2,2.6)	0.201	0.841
HCO_3^-(mmol/L)	22(18,25)	22(18,25)	0.036	0.971
Function Score [$M(Q_L,Q_U)$]				
GCS	15(14,15)	15(14,15)	0.564	0.573
SOFA	5(3,7)	5(2,8)	1.014	0.311
SAPS II	33(24,43)	34(23,46)	0.277	0.782
Complication [n(%)]				
Hypertension	117(45.70)	287(47.75)	0.226	0.634
Diabetes	23(8.98)	51(8.49)	0.011	0.916
Sepsis	147(57.42)	378(62.9)	2.041	0.153
Myocardial Infarction	3(1.17)	7(1.16)	< 0.001	0.993
COPD	7(2.73)	19(3.16)	0.013	0.908
Acute Kidney Injury	171(66.8)	391(65.06)	0.170	0.681
Treatment Measures [n(%)]				
CRRT	33(12.89)	79(13.14)	0.010	0.920
Mechanical Ventilation	77(30.08)	193(32.11)	0.257	0.612
Cardiopulmonary Resuscitation	4(1.56)	9(1.50)	0.005	0.943
Heparin	239(93.36)	550(91.51)	0.603	0.437
Aspirin	87(33.98)	189(31.45)	0.419	0.517
Antibiotics	221(86.33)	519(86.36)	< 0.001	1.000
Vasoactive Drugs	60(23.44)	149(24.79)	0.113	0.737

2.2 Feature Selection and Model Evaluation

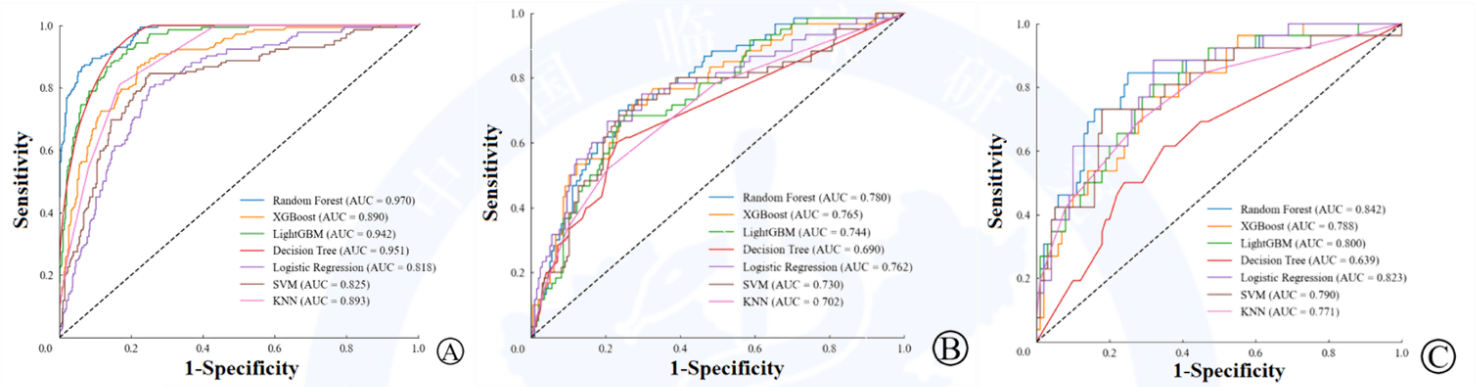
A total of 43 variables, after data preprocessing, were incorporated into LASSO regression for automatic feature selection. Using cross-validation, the optimal mean squared error λ was selected as 0.418 8, retaining 7 variables with non-zero coefficients: BMI, RR, serum albumin, SOFA score, sepsis, CRRT, and mechanical ventilation **(Figure 2)**. These features were used as predictor variables to construct the prediction model. **Figures 3A, 3B, and 3C** display the ROC curves for the training set, internal validation set, and external test set, respectively. ROC curve analysis of the internal validation set showed that among the 7 models, the RF model had the best predictive performance for ARDS occurrence in AP patients during ICU stay (AUC=0.780, 95%*CI*:0.721–0.846). The predictive performance ranking was RF, XGBoost, LR, LightGBM, SVM, KNN, and DT models, as shown in **Figure 3B.** The detailed performance metrics of the 7 machine learning models for predicting ARDS in AP patients in the internal validation set are shown in **Table 3**.

Subsequently, the RF model was evaluated for calibration and clinical applicability. The calibration curve showed that the prediction probability of the RF model had smaller deviation from the actual probability compared to other models, indicating higher predictive accuracy **(Figure 4A)**. The DCA curve showed that when the probability threshold was between 0.05 and 0.58, the DCA curve did not intersect with the two extreme curves, suggesting that using this model to predict ARDS risk in AP patients and making clinical interventions resulted in a good net benefit, demonstrating the model's clinical applicability (**Figure 4B**). ROC curve analysis of the external test set similarly showed that the RF model (AUC=0.842, 95%*CI*: 0.751–0.917) outperformed other machine learning models, consistent with the internal validation set results from MIMIC, proving the RF model's good generalizability (**Figure 3C**).



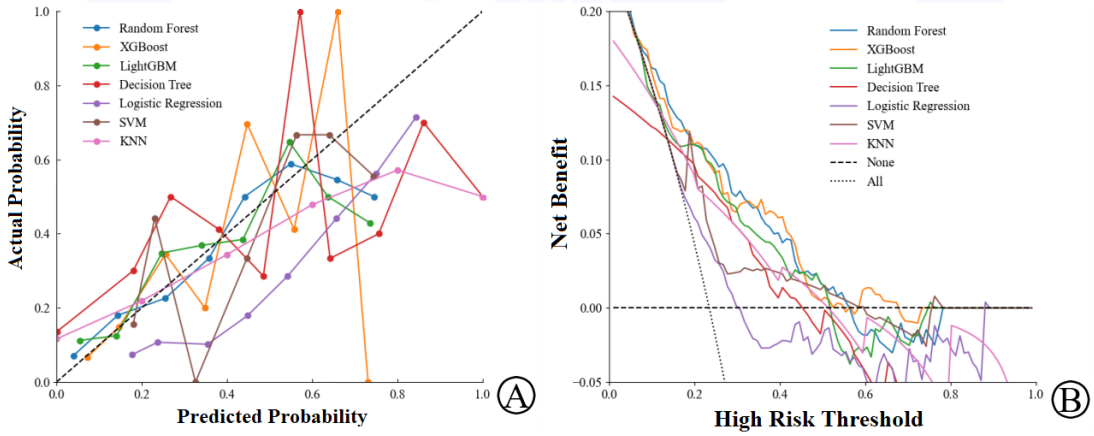
Note: A, clinical feature coefficient graph; B, cross-validation graph.

Fig.2 Variables screening by LASSO regression



Note: A, the training set; B, the internal validation set; C, the external test set.

Fig.3 ROC curves of 7 machine learning models predicting concurrent ARDS in AP patients



Note: A, calibration curve; B, DCA curve.

Fig.4 Calibration curves and DCA curves for predicting ARDS in AP patients by 7 machine learning models

Tab.3 Predictive value of 7 machine learning models for ARDS in AP patients

Model	AUC	95%*CI*	Accuracy	Precision	Sensitivity	Specificity	F1 Score
RF	0.780	0.721-0.846	0.781	0.556	0.333	0.918	0.417
XGBoost	0.765	0.692-0.835	0.770	0.522	0.200	0.944	0.289
LightGBM	0.744	0.674-0.817	0.781	0.553	0.350	0.913	0.429
DT	0.690	0.622-0.762	0.766	0.500	0.333	0.898	0.400
LR	0.762	0.689-0.837	0.746	0.471	0.667	0.770	0.552
SVM	0.730	0.657-0.806	0.777	0.579	0.183	0.959	0.278
KNN	0.702	0.632-0.776	0.770	0.513	0.333	0.903	0.404

2.3 Explanation of the RF Model Based on the SHAP Algorithm

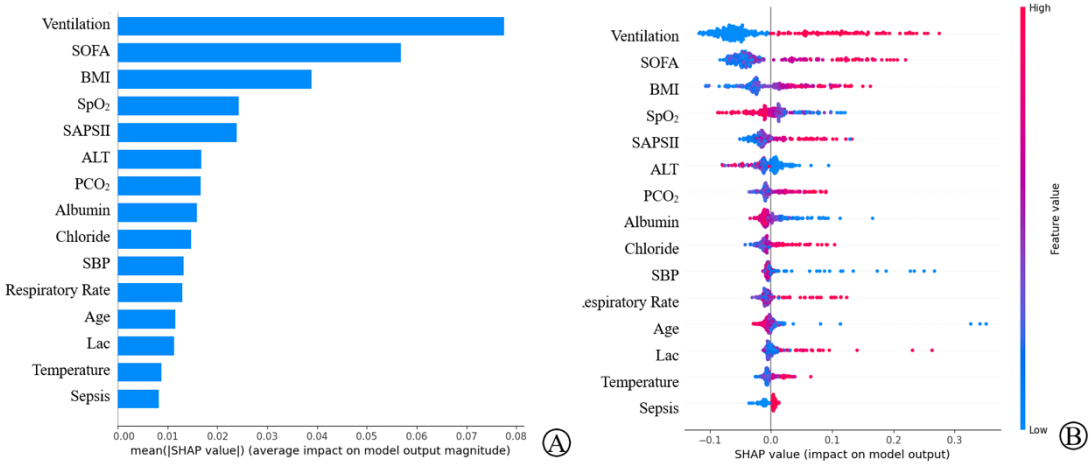
2.3.1 Global Sample Feature Explanation

The importance of each feature in the RF model's prediction results was shown in **Figure 5**. The variable importance plot ranked features based on the average absolute SHAP values for each feature. The top five variables, in order of importance, were mechanical ventilation, SOFA score, BMI, SpO_2, and SAPS II score (**Figure 5A**). The bee swarm plot revealed that among the top five variables, the use of mechanical ventilation, SOFA score, BMI, and SAPS II score increased the risk of ARDS (indicated by the red high-density region) (**Figure 5B**).

The top five important features were selected to plot the PDP. **Figure 6A** showed that mechanical ventilation increases the risk of ARDS from 16% to 37%. **Figure 6B** indicated that when the SOFA score was less than 8, the risk of ARDS remained stable at a relatively low level. However, when the score exceeded 8, the risk significantly increased. **Figure 6C** showed that the risk of ARDS increases with BMI. **Figure 6D** revealed a non-linear relationship between SpO_2 and ARDS risk when SpO_2 was below 90%. The risk remained around 30%, but as SpO_2 exceeded 90%, the risk decreased with increasing SpO_2. **Figure 6E** demonstrated that when the SAPS II score was between 46 and 60, the risk of ARDS significantly increased.

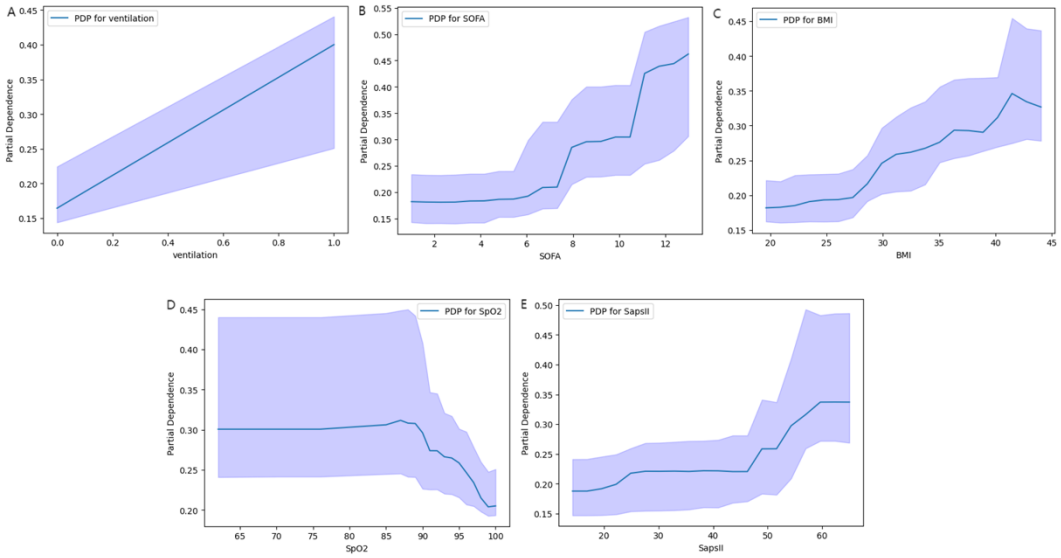
2.3.2 Single-Sample Prediction Feature Explanation

SHAP force plots were used to visualize the contribution of individual features to the prediction. **Figure 7** showed two randomly selected cases from the internal validation set. Case A (actual negative) had a predicted risk of 10%, which was significantly lower than the baseline. Case B (actual positive) had a predicted risk of 71%, which was clearly higher than the baseline. Both predictions aligned with clinical diagnoses, confirming the high reliability of the model's individualized predictions.



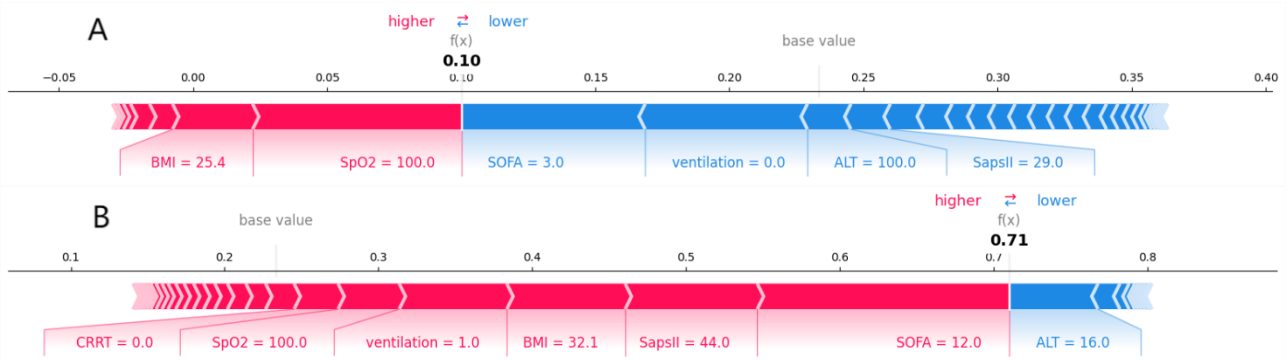
Note: A, variable importance diagram; B, bee colony diagram.

Fig.5 Variable importance diagram and bee colony diagram



Note: A, ventilation; B, SOFA; C, BMI; D, SpO2; E, SAPS II.

Fig.6 PDP of the top 5 variables in importance ranking



Note: A, patient instanced with predicted low ARDS risk; B, patient instanced with predicted high ARDS risk.

Fig.7 Attempt of random forest model

3 Conclusion

This study demonstrates the superior performance of the RF model, built using multicenter data, in predicting the risk of ARDS in AP patients (internal validation set AUC=0.780, external test set AUC=0.842). SHAP analysis reveals that mechanical ventilation, SOFA score, BMI, SpO_2, and SAPS II score are the core factors influencing the model's decision-making process, providing actionable early warning indicators for clinical use.

Various laboratory indicators and clinical scoring systems have been used to predict ARDS in AP patients. Laboratory indicators such as procalcitonin, white blood cell count, and albumin, as well as clinical scoring tools like APACHE II, Ranson score, and the bedside severity index of acute pancreatitis (BISAP), have all been shown to be associated with AP complicated by ARDS [4,6,16-17]. However, these methods still have limitations in predictive performance and assessment consistency, making it difficult to meet the clinical demand for accurate predictions. Furthermore, traditional regression models, such as LR in generalized linear regression and Cox regression in semiparametric survival analysis, although easy to operate and relatively interpretable, often require assumptions of linearity and homogeneity between input variables, making it difficult to capture nonlinear relationships between variables and leading to oversimplification of nonlinear interactions [18]. Additionally, these models tend to be less efficient when analyzing large sample data, exhibiting higher error rates compared to newer machine learning technologies. Their high sensitivity to multicollinearity among predictors may also degrade model performance [18-19].

In contrast, machine learning techniques can establish automated data analysis workflows, efficiently handle nonlinear and high-dimensional data, and uncover complex interactions between variables, thereby improving predictive accuracy [20]. With the widespread use of large databases, machine learning has shown excellent results in the clinical diagnosis and prognosis assessment of pancreatitis. Ren *et al.* [21] built a model using the MIMIC-IV and eICU-CRD databases to predict the in-hospital mortality of AP patients in the ICU, and the results showed that the Gaussian Naive Bayes (GNB) model achieved AUC values of 0.840 and 0.862 in the two databases, respectively. Liu *et al.* [12] conducted a study using both the MIMIC-III and MIMIC-IV databases to predict the risk of sepsis in AP patients, which also confirmed that machine learning methods outperformed traditional LR models and various scoring systems. This study developed the model and performed internal validation using the MIMIC-IV database. Feature selection was conducted using LASSO regression, and based on this, seven machine learning algorithms were developed, ultimately determining that the RF model demonstrated superior predictive performance in the early identification of high-risk ARDS patients. The RF algorithm, by integrating a large number of decision trees, outputs the class vote result of the majority of trees in classification tasks, demonstrating strong classification capability and generalization performance, especially in imbalanced data [22-24].The RF model was externally tested on data from AP patients treated in the ICU at Changshu Hospital Affiliated to Soochow University, and the results proved that it also exhibited good predictive performance in domestic patients.

Although machine learning algorithms have high predictive accuracy, their "black-box" nature limits their application in clinical settings [25]. Previous studies predicting ARDS in AP patients using machine learning techniques did not adequately quantify the influence of feature variables on model output and lacked effective methods to visualize these influences [26-27]. To overcome this limitation, this study used SHAP analysis to provide interpretability for the RF model, offering both global and personalized explanations to better understand the model's working mechanism. The results indicated that the five most important variables in prediction were mechanical ventilation, SOFA score, BMI, SpO_2 level, and SAPS II score. These variables can be used as key indicators for predicting ARDS in such patients. The PDP intuitively displayed the impact of individual features on the probability of ARDS, while the force plot achieved prediction contribution analysis for different feature values under randomly selected individual cases, helping clinicians better understand the decision-making process of the model.

Limitations of this study: (1) The data was collected from two databases, which may have regional and center biases. (2) The external test set sample size was small and requires multi-center, large-sample validation for generalization. (3) Although LASSO regression effectively reduced the number of features, it might overlook potentially important variables, and different feature selection methods may yield different results. (4) While SHAP analysis provided model interpretability, the explanation results depended on the model's prediction accuracy and data quality, and might not fully reveal the biological mechanisms behind complex interactions. Clinical knowledge and experimental validation were required for a more comprehensive understanding of the predictive results.

In conclusion, this study constructed an early prediction model for ARDS in AP patients using seven machine learning methods, with the RF model exhibiting the best predictive performance and generalization ability. The introduction of SHAP analysis enhanced the model's interpretability, providing clinicians with a reliable and intuitive decision-making tool that helps in the early identification of high-risk patients and optimization of treatment strategies.

Conflict of Interest None

References

[1] Sohail Z, Shaikh H, Iqbal N, et al. Acute pancreatitis: a narrative review[J]. J Pak Med Assoc, 2024, 74(5): 953-958.

[2] Trikudanathan G, Yazici C, Evans Phillips A, et al. Diagnosis and management of acute pancreatitis[J]. Gastroenterology, 2024, 167(4): 673-688.

[3] Petrov MS, Yadav D. Global epidemiology and holistic prevention of pancreatitis[J]. Nat Rev Gastroenterol Hepatol, 2019, 16(3): 175-184.

[4] Ding N, Guo CR, Song K, et al. Nomogram for the prediction of in-hospital incidence of acute respiratory distress syndrome in patients with acute pancreatitis[J]. Am J Med Sci, 2022, 363(4): 322-332.

[5] Zhou Y, Lu GT, Wang D, et al. Research progress on predictive models for acute pancreatitis complicated with acute respiratory distress syndrome [J]. Chin J Pancreatol, 2022, 22(3): 231-234. **[In Chinese]**

[6] Zhang WW, Zhang M, Kuang ZM, et al. The risk factors for acute respiratory distress syndrome in patients with severe acute pancreatitis: a retrospective analysis[J]. Medicine (Baltimore), 2021, 100(2): e23982.

[7] Fei Y, Gao K, Li WQ. Prediction and evaluation of the severity of acute respiratory distress syndrome following severe acute pancreatitis using an artificial neural network algorithm model[J]. HPB (Oxford), 2019, 21(7): 891-897.

[8] Guo SS, Lei JJ, Wang YJ, et al. Timing of blood purification therapy in acute pancreatitis with systemic inflammatory response syndrome [J]. Chin J Clin Res, 2024, 37(7): 1090-1096. **[In Chinese]**

[9] Gao X, Lin JX, Wu AR, et al. Application of machine learning model based on XGBoost algorithm in early prediction of patients with acute severe pancreatitis [J]. Chin Crit Care Med, 2023, 35(4): 421-426. **[In Chinese]**

[10] Kui B, Pintér J, Molontay R, et al. EASY-APP: an artificial intelligence model and application for early and easy prediction of severity in acute pancreatitis[J]. Clin Transl Med, 2022, 12(6): e842.

[11] Lin SW, Lu WB, 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[J]. Ren Fail, 2024, 46(1): 2303395.

[12] Liu F, Yao J, Liu CY, et al. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis[J]. BMC Surg, 2023, 23(1): 267.

[13] Lundberg SM, Lee SI. A unified approach to interpreting model predictions: 31st Conference on Neural Information Processing Systems[C/OL]. CA: Long Beach, 2017 (2017-11-26) [2025-04-01]. https://arxiv.org/abs/1705.07874v2.

[14] Chinese Pancreatic Surgery Association, Chinese Society of Surgery, Chinese Medical Association. Guidelines for diagnosis and treatment of acute pancreatitis in China (2021) [J]. Chin J Surg, 2021, 59(7): 578-587. **[In Chinese]**

[15] ARDS Definition Task Force. acute respiratory distress syndrome: The Berlin definition[J]. JAMA, 2012, 307(23): 2526-2533.

[16] Lin FY, Lu RL, Han DD, et al. A prediction model for acute respiratory distress syndrome among patients with severe acute pancreatitis: a retrospective analysis[J]. Ther Adv Respir Dis, 2022, 16: 17534666221122592.

[17] Li YL, Zhang DD, Xiong YY, et al. Development and external validation of models to predict acute respiratory distress syndrome related to severe acute pancreatitis[J]. World J Gastroenterol, 2022, 28(19): 2123-2136.

[18] Du M, Haag DG, Lynch JW, et al. Comparison of the tree-based machine learning algorithms to cox regression in predicting the survival of oral and pharyngeal cancers: analyses based on SEER database[J]. Cancers (Basel), 2020, 12(10): 2802.

[19] Yang DL, Kang J, Li YH, et al. Development of a predictive nomogram for acute respiratory distress syndrome in patients with acute pancreatitis complicated with acute kidney injury[J]. Ren Fail, 2023, 45(2): 2251591.

[20] Yin MY, Zhu JZ, Liu L, et al. Research advances in machine learning models for acute pancreatitis [J]. Journal of Clinical Hepatol, 2023, 39(12): 2978-2984. **[In Chinese]**

[21] Ren WS, Zou K, Huang S, et al. Prediction of in-hospital mortality of intensive care unit patients with acute pancreatitis based on an explainable machine learning algorithm[J]. J Clin Gastroenterol, 2024, 58(6): 619-626.

[22] Zhang WY, Chang YJ, Ding Y, et al. To establish an early prediction model for acute respiratory distress syndrome in severe acute pancreatitis using machine learning algorithm[J]. J Clin Med, 2023, 12(5): 1718.

[23] Masetic Z, Subasi A. Congestive heart failure detection using random forest classifier[J]. Comput Methods Programs Biomed, 2016, 130: 54-64.

[24] Roguet A, Eren AM, Newton RJ, et al. Fecal source identification using random forest[J]. Microbiome, 2018, 6(1): 185.

[25] The Lancet Respiratory Medicine. Opening the black box of machine learning[J]. Lancet Respir Med, 2018, 6(11): 801.

[26] Zou K, Ren WS, Pu XX, et al. The role of machine learning in predicting the occurrence of acute respiratory distress syndrome in patients with severe acute pancreatitis [J]. Chin J Pancreatol, 2023(6): 455-459. **[In Chinese]**

[27] Zou K, Ren WS, Huang S, et al. The role of artificial neural networks in prediction of severe acute pancreatitis associated acute respiratory distress syndrome: a retrospective study[J]. Medicine (Baltimore), 2023, 102(29): e34399.

Submission received: 2025-04-16 / **Revised:** 2025-05-15

·论 著·

基于机器学习算法构建重症监护病房急性胰腺炎并发急性呼吸窘迫综合征的风险预测模型

任夏[1]， 刘罗杰[1]， 查俊杰[2]， 叶叶[1]， 徐晓丹[1]， 叶宏伟[2]， 张岩[2]
1. 苏州大学附属常熟医院 常熟市第一人民医院消化内科， 江苏 常熟 215500；
2. 苏州大学附属常熟医院 常熟市第一人民医院重症医学科， 江苏 常熟 215500

摘要：目的 旨在开发和验证一种基于机器学习算法的预测模型，用于评估急性胰腺炎（AP）患者在重症监护病房（ICU）内发生急性呼吸窘迫综合征（ARDS）的风险。**方法** 回顾性分析美国重症监护医学信息数据库Ⅳ v2.2（MIMIC-Ⅳ v2.2）中的857例AP患者的相关资料，按7∶3的比例随机划分为训练集（*n*=601）和内部验证集（*n*=256），另收集2019年1月至2024年3月苏州大学附属常熟医院ICU 126例AP患者的相关资料作为外部测试集。根据是否并发ARDS将所有患者分为ARDS组与非ARDS组，收集其人口学特征、入ICU 24 h内初始的生命体征、实验室数据、功能评分及并发症情况，采用最小绝对收缩和选择算子（LASSO）回归进行特征选择，并使用随机森林（RF）、极端梯度提升（XGBoost）、轻量级梯度提升机（LightGBM）、决策树（DT）、逻辑回归（LR）、支持向量机（SVM）和K最近邻（KNN）7种机器学习算法构建预测模型。模型性能评估利用受试者工作特征（ROC）曲线、校准曲线及决策曲线分析（DCA），最后借助夏普利加性解释（SHAP）算法对模型进行可解释性分析。**结果** MIMIC-Ⅳ数据库中202例（23.57%）并发ARDS，外部测试集中26例（20.63%）并发ARDS。基于训练集数据，采用LASSO回归从43个变量中筛选出7个关键变量进行模型构建，多种机器学习模型比较结果显示，RF模型在内部验证集和外部测试集ROC曲线下面积（AUC）分别为0.780（95%*CI*为0.721~0.846）和0.842（95%*CI*为0.751~0.917），均高于其他6种模型；校准曲线显示RF模型的预测概率与实际概率的偏差较其他模型小，整体预测性能最佳。基于RF模型的SHAP算法分析表明，机械通气、序贯器官功能衰竭（SOFA）评分、身体质量指数（BMI）、脉博血氧饱和度（SpO_2）和简明急性生理功能Ⅱ（SAPSⅡ）评分是影响ARDS风险的主要因素。机械通气可使ARDS的发生风险从16%上升至37%；SOFA大于8分时ARDS风险会显著上升；ARDS发生风险会随着BMI的增加而升高；SpO_2低于90%时，ARDS发生风险维持在30%，当SpO_2超过90%后风险则随着SpO_2增加而呈下降趋势；SAPSⅡ评分在46~60分之间时，ARDS的风险呈明显上升趋势。**结论** 基于RF算法的预测模型为AP患者并发ARDS的风险评估提供了可靠工具，通过SHAP方法增强了模型的可解释性，有助于临床决策。

关键词： 急性胰腺炎；急性呼吸窘迫综合征；重症监护病房；机器学习；随机森林；夏普利加性解释
中图分类号： R563.8 TP181 **文献标识码：** A **文章编号：** 1674-8182(2025)08-1173-09

Development of a prediction model for acute respiratory distress syndrome in ICU patients with acute pancreatitis based on machine learning algorithms

REN Xia*, LIU Luojie, ZHA Junjie, YE Ye, XU Xiaodan, YE Hongwei, ZHANG Yan
*Department of Gastroenterology, Changshu No.1 People's Hospital, Changshu Hospital Affiliated to Soochow University, Changshu, Jiangsu 215500, China
Corresponding author: ZHANG Yan, E-mail: 1429862753@qq.com

Abstract: Objective To develop and validate a predictive model based on machine learning algorithms to assess the risk of acute respiratory distress syndrome (ARDS) in patients with acute pancreatitis (AP) admitted to the intensive

DOI: 10.13429/j.cnki.cjcr.2025.08.008
基金项目： 苏州市科技发展计划项目（SLT2023006）；常熟市科技发展计划重点项目（CSWS202209）；中华国际医学交流基金会呼吸疾病专项项目（Z-2014-08-2309-1）
通信作者： 张岩，E-mail：1429862753@qq.com
出版日期： 2025-08-20

QR code for English version

care unit(ICU). **Methods** The relevant data of 857 AP patients from the Medical Information Mart for Intensive Care Ⅳ v2.2 (MIMIC-Ⅳ v2.2) database were retrospectively analyzed and were randomly divided into a training set(n=601) and an internal validation set(n=256) in a 7∶3 ratio. Additionally, the relavent data of 126 AP patients from the ICU of Changshu Hospital Affiliated to Soochow University from January 2019 to March 2024 were collected as an external test set. Patients were categorized into ARDS and non-ARDS groups based on the occurrence of ARDS. Demographic characteristics, initial vital signs, laboratory data, functional scores, and complications within the initial 24-hour of ICU admission were collected. Feature selection was performed using least absolute shrinkage and selection operator (LASSO) regression. Predictive models were constructed using seven machine learning algorithms: random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), decision tree (DT), logistic regression (LR), support vector machine (SVM), and K-nearest neighbors (KNN). Model performance was evaluated using receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA). Finally, model interpretability was enhanced through Shapley additive explanations (SHAP) analysis. **Results** In the MIMIC-Ⅳ database, 202 patients (23.57%) developed ARDS, while 26 patients (20.63%) developed ARDS in the external test set. Seven key variables were selected by LASSO regression from 43 variables in the training set to construct the models. Among various machine learning models, the RF model demonstrated the best performance with an area under the curve (AUC) of 0.780 (95%*CI*: 0.721-0.846)in the internal validation set and 0.842 (95%*CI*: 0.751-0.917) in the external test set, outperforming the other six models. The calibration curve indicated that the predicted probabilities from the RF model had the smaller deviation from the actual probabilities compared to other models, showing the best overall predictive performance. SHAP analysis based on the RF model revealed that mechanical ventilation, sequential organ failure assessment (SOFA) score, body mass index (BMI), peripheral oxygen saturation (SpO_2) and simplified acute physiology score (SAPS Ⅱ) were the main factors influencing ARDS risk. Mechanical ventilation increased the risk of ARDS from 16% to 37%. When the SOFA score exceeded 8, the ARDS risk rose significantly. The risk of ARDS elevated with increased BMI. While SpO_2 remained below 90%, ARDS risk stabilized at 30%; once SpO_2 surpassed 90%, the risk demonstrated a declining trend with further increases in SpO_2. For SAPS-Ⅱ scores between 46 and 60, ARDS risk showed a pronounced upward trend. **Conclusion** The RF predictive model provides a reliable tool for assessing the risk of ARDS in AP patients and enhances model interpretability through the SHAP method, aiding in clinical decision-making.

Keywords: Acute pancreatitis; Acute respiratory distress syndrome; Intensive care unit; Machine learning; Random forest; Shapley additive explanations

Fund program: Suzhou Science and Technology Development Plan Project (SLT2023006); Key Project of Changshu Science and Technology Development Plan (CSWS202209); Respiratory Disease Special Project of China International Medical Exchange Foundation (Z-2014-08-2309-1)

急性胰腺炎(acute pancreatitis,AP)是一种以胰腺急性炎症为特征的常见急腹症,其发病率在全球范围内逐渐上升,估计每年每10万人中有13~49例发病[1]。在过去10年中,因AP住院的人数增加了约30%,美国每年在这方面的医疗花费就达到了26亿美元[2]。AP的总体死亡率约为1%~5%,但约20%的轻症AP患者可能进展为中度重症急性胰腺炎(moderate severe acute pancreatitis,MAP)或重症急性胰腺炎(severe acute pancreatitis,SAP),并伴随胰腺坏死及持续性器官功能障碍等,死亡率可随之上升为10%~50%[3-5]。急性呼吸窘迫综合征(acute respiratory distress syndrome,ARDS)往往是SAP患者最先出现的器官功能障碍,其发生率可高达30%,也是AP患者早期死亡的主要原因[6-8]。及时识别可能发生ARDS的高危患者有助于减轻疾病的严重程度、改善预后并降低死亡率。

随着人工智能(artificial intelligence, AI)技术的快速发展,机器学习作为一种数据分析方法,因其强大的模式识别能力和对复杂数据结构的处理能力而变得越来越流行。近年来,机器学习在AP领域的应用日益增多,已在AP的疾病诊断[9]、严重程度[10]和并发症预测[11-12]等多个方面展现出良好的预测效能。但机器学习模型因复杂的内部结构及决策过程而常被喻为“黑箱”模型,现有的关于AP机器学习的研究也多专注于模型预测的准确性判断,而忽视了医疗领域对结果可解释性的高要求,这在一定

程度上限制了这些模型的泛化能力，不利于在疾病诊断及预测中的应用[10]。Lundberg等[13]基于博弈论中的夏普利值提出夏普利加性解释（Shapley additive explanations，SHAP）方法，能量化模型中每个特征对预测结果的贡献，并提供直观的可视化展示。因此，本研究拟利用多中心数据，基于机器学习算法训练和验证一系列模型[包括随机森林（RF）、极端梯度提升（XGBoost）、轻量级梯度提升机（LightGBM）、决策树（DT）、逻辑回归（LR）、支持向量机（SVM）和K最近邻（KNN）]，通过最小绝对收缩和选择算子（LASSO）回归分析筛选出性能最佳的模型并进行可视化解释，用于预测重症监护病房（ICU）AP患者并发ARDS的风险，旨在帮助临床医师及时决策干预。

1　对象与方法

1.1　*数据来源*　本研究回顾性地从美国重症监护医学信息数据库Ⅳ v2.2（Medical Information Mart for Intensive Care v2.2，MIMIC-Ⅳ v2.2）提取数据（数据集1），该数据库为麻省理工学院计算生理学实验室开发的大型、单中心、去身份化的公开数据库，它收集了2008年至2019年期间在波士顿贝斯以色列女执事医疗中心（BIDMC）入住ICU的53 150名患者的详细就诊数据。本研究已通过美国国立卫生研究院合作机构培训计划（证书编号：60227322），并获得数据库访问许可。同时选取2019年1月至2024年3月在苏州大学附属常熟医院ICU入住的126例MAP及SAP患者作为外部测试集（数据集2），研究中为避免患者信息泄露，对患者数据进行了清洗。本研究经常熟市第一人民医院伦理委员会审核批准（批号：L202402）。

1.2　*研究人群*　根据MIMIC-Ⅳ数据库中国际诊断代码ICD-9（577.0）和ICD-10（K85%）选择入院时诊断为AP的患者，外部测试集人群的纳入标准符合《中国急性胰腺炎诊治指南（2021）》[14]，ARDS的诊断标准符合柏林定义[15]。排除标准包括：年龄小于18岁；入住ICU的时间少于24 h；在入住ICU之前已存在呼吸衰竭。具体筛选流程见图1。

1.3　*数据提取与处理*　公共数据库病例数据的提取利用pgAdmin 4软件中的SQL查询语言，所有临床和实验室变量均为转入ICU后24 h内的数据，对于有多个测量值的变量，在本研究中仅纳入首次测量结果。从数据库中共提取50个变量，包括以下内容，（1）人口统计学特征：年龄、性别、种族、身高、体质量、医保；（2）生命体征：心率、收缩压、舒张压、呼吸频率、体温、氧合指数；（3）实验室数据：白细胞计数、血红蛋白、血小板、白蛋白、红细胞压积、肌酐、C反应蛋白、丙氨酸氨基转移酶（ALT）、天门冬氨酸氨基转移酶（AST）、尿素氮、葡萄糖、钠、钾、氯、钙、凝血酶原时间（PT）、凝血酶原国际标准化比率（INR）、pH、二氧化碳分压（PCO_2）、剩余碱、乳酸、碳酸氢盐（HCO_3^-）；（4）功能评分：格拉斯哥昏迷（GCS）评分，序贯器官功能衰竭（SOFA）评分，简明急性生理功能评分Ⅱ（SAPSⅡ）；（5）合并症：高血压、糖尿病、脓毒症、心肌梗死、慢性阻塞性肺疾病（COPD）、急性肾损伤；（6）治疗措施：连续性肾替代治疗（CRRT）、机械通气、心肺复苏、肝素、阿司匹林、抗生素和血管活性药。苏州大学附属常

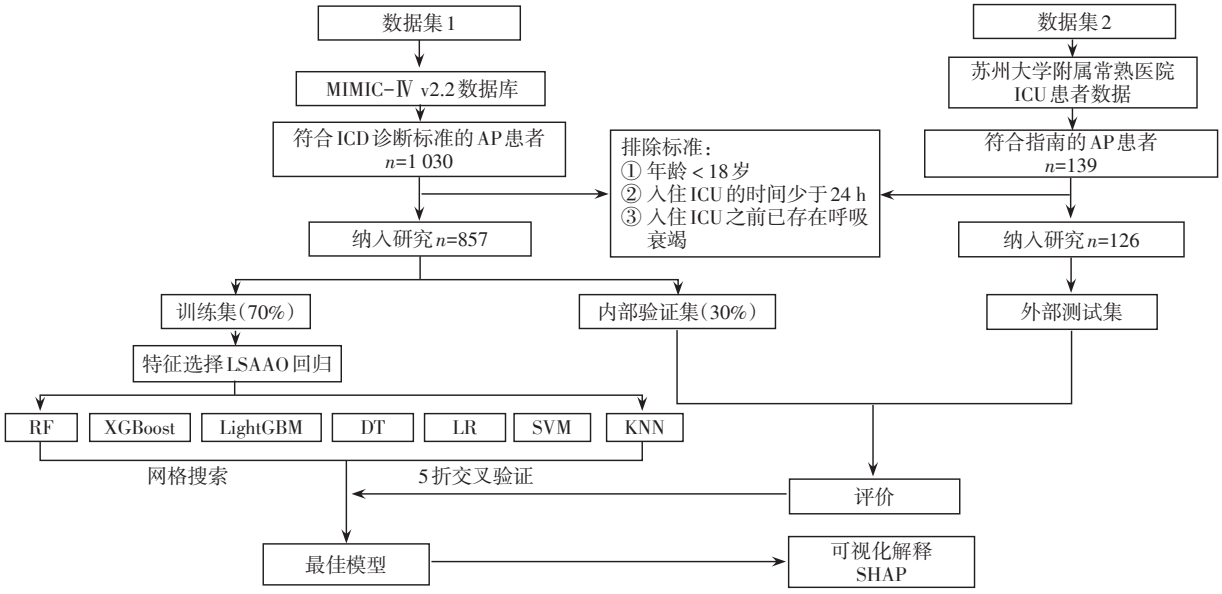


图1　研究流程图

Fig.1　Research flowchart

熟医院收集的变量同上。

对于缺失比例小于5%的数据，使用均值法插补；对于缺失5%~30%的数据，应用多重插补法进行填补；对于缺失值大于30%的变量，为减小偏差，选择不予纳入。为降低变量间的多重共线性，通过计算Spearman等级相关系数并移除系数大于0.8的变量（尿素氮、红细胞压积、AST、PT）。最终共43个变量纳入分析。

1.4　统计学方法　采用Stata 16.0软件、R 4.3.2软件和Python 3.12.4软件进行数据分析。通过简单随机抽样法对处理后的MIMIC-Ⅳ数据集进行拆分，其中70%用于训练集，30%用于内部验证集。利用苏州大学附属常熟医院的数据作为外部测试集。对符合正态分布的计量资料用 $\bar{x} \pm s$ 表示，两组间比较采用独立样本 t 检验；偏态分布的计量资料以 $M(Q_L, Q_U)$ 表示，组间比较采用Mann-Whitney U 检验。分类变量以例（%）表示，组间比较采用 χ^2 检验。通过LASSO回归进行特征选择。$P < 0.05$ 为差异有统计学意义。

1.5　构建模型及验证　使用LASSO回归分析降维后筛选出的变量，选择7种机器学习算法在训练队列中构建预测模型，包括RF、XGBoost、LightGBM、DT、LR、SVM和KNN。构建模型时使用五折交叉验证训练模型防止过拟合，并使用网格搜索调整超参数。采用受试者工作特征（ROC）曲线作为模型的评价指标，同时计算ROC曲线下面积（AUC）、精确度、敏感度、特异度和F1得分（精确度和召回率的调和平均值）。绘制校准曲线对模型进行一致性分析，应用决策曲线分析（DCA）评估模型的临床净获益。

1.6　模型可视化解释　SHAP方法通过计算每个特征变量的SHAP值揭示单个特征的重要性及不同特征间的相互作用。本研究根据上述评价指标筛选出最佳诊断模型，通过Python 3.12.4软件分别绘制SHAP特征重要性图、蜂群图、力图及重要变量的部分依赖图（PDP）对模型进行可视化解释。

2　结　果

2.1　基线特征　本次研究最终共纳入983例AP患者，其中MIMIC-Ⅳ数据库中纳入857例，并发ARDS者202例（23.57%）；苏州大学附属常熟医院作为外部测试集纳入126例，并发ARDS者26例（20.63%），两组患者ARDS发生率相近。两组数据集的基线资料见表1。训练集与内部验证集之间的人口学特征、基础生命体征、实验室数据、功能评分、并发症及治疗措施差异均无统计学意义（$P > 0.05$）。见表2。

2.2　特征选择及模型评价　将经过数据预处理的43个变量纳入LASSO回归进行自动特征选择，通过交叉验证，选择均方误差λ为0.418 8，最终保留7个具有非零系数的变量，包括：身体质量指数（BMI）、呼吸频率、血清白蛋白、SOFA、脓毒症、CRRT和机械通气（图2）。将这些特征作为预测变量构建预测模型，图3A、3B、3C分别为训练集、内部验证集和外部测试集的ROC曲线，内部验证集的ROC曲线分析显示，7种模型中，RF模型对AP患者住ICU期间发生ARDS的预测效果最好（AUC=0.780，95%*CI*为0.721~0.846），排序依次为RF、XGBoost、LR、LightGBM、SVM、KNN及DT模型，见图3B。以上7种预测AP患者发生ARDS的机器学习模型在内部验证集上的详细性能指标见表3。

随后对RF模型进行校准度及临床适用性分析，校准曲线显示，RF模型的预测概率与实际概率的偏差较其他模型小，RF模型有较高的预测准确性（图4A）。DCA曲线显示，当概率阈值在0.05~0.58之间时，DCA曲线与两条极端曲线不相交，用该模型预测AP患者发生ARDS风险并进行临床干预可获得良好的净获益，表明该模型临床适用性较好（图4B）。外部测试集的ROC曲线分析同样显示RF模型（AUC=0.842，95%*CI*为0.751~0.917）的预测效果优于其他机器学习模型，与MIMIC内部验证集结果相同，证明该RF模型拥有较好的泛化能力（图3C）。

2.3　基于SHAP算法解释RF模型

2.3.1　全局样本特征解释　RF算法构建的模型中每个特征对预测结果影响的重要性如图5所示，变量重要性图根据每个特征的平均SHAP绝对值进行排序，结果显示居前5位的变量依次为机械通气、SOFA评分、BMI、SpO_2和SAPSⅡ评分（图5A）；蜂群图揭示前5名变量中，使用机械通气及SOFA评分、BMI和SAPSⅡ评分的增加会驱动ARDS风险提升（红色高密度区）（图5B）。

选择重要性排名前5的特征绘制PDP图，图6A显示机械通气可使ARDS的发生风险从16%上升至37%；图6B显示SOFA评分在小于8分时，ARDS的风险基本平稳且维持在相对较低水平，大于8分后风险则会显著上升；图6C显示ARDS发生风险会随着BMI的增加而升高；图6D显示SpO_2低于90%时，ARDS发生风险与SpO_2之间并非线性关系，而是维持在30%左右，当SpO_2超过90%后风险则随着SpO_2增加而呈下降趋势；图6E显示SAPSⅡ评分在46~60分之间时，ARDS的风险呈明显上升趋势。

表1 不同数据集AP患者ARDS组与非ARDS组基线资料比较

Tab.1 Comparison of baseline data between ARDS group and non-ARDS group of AP patients in different datasets

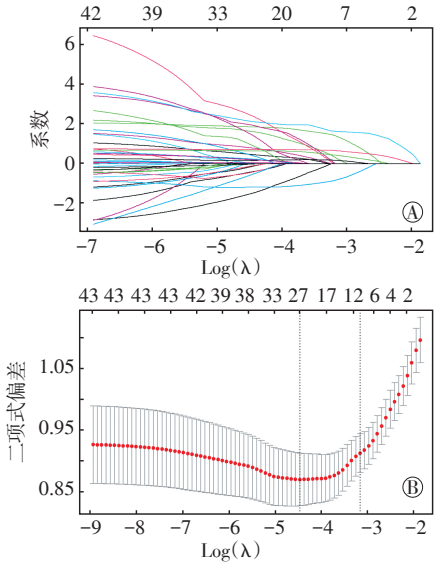
变量	MIMIC数据库				外部测试集			
	非ARDS组(n=655)	ARDS组(n=202)	$Z/\chi^2/t$值	P值	非ARDS组(n=100)	ARDS组(n=26)	$Z/\chi^2/t$值	P值
人口学特征								
年龄[岁, $M(Q_L,Q_U)$]	61(47,76)	58(46,69)	2.230	0.026	52(40,66)	44(34,56)	1.595	0.111
男性[例(%)]	372(56.79)	125(61.88)	1.438	0.230	50(50.00)	15(57.69)	0.229	0.632
BMI[kg/m^2, $M(Q_L,Q_U)$]	27.69(23.96,32.34)	31.12(26.90,37.04)	5.735	<0.001	20.91(19.38,24.10)	23.50(21.06,27.53)	2.607	0.009
医保[例(%)]	222(33.89)	62(30.69)	0.576	0.448	35(35.00)	12(46.15)	0.673	0.412
生命体征								
心率[次/min, $M(Q_L,Q_U)$]	95(81,112)	103.5(88,119)	3.535	<0.001	100(80,116)	109(93,124)	1.707	0.088
舒张压(mmHg, $\bar{x}\pm s$)	128.90±24.30	121.61±27.82	3.599	<0.001	127.34±25.06	120.04±23.05	1.413	0.165
收缩压(mmHg, $\bar{x}\pm s$)	73.57±18.47	70.61±19.76	1.954	0.051	71.23±17.96	67.77±16.76	0.924	0.361
呼吸[次/min, $M(Q_L,Q_U)$]	20(16,24)	22(18,28)	3.338	<0.001	19(15,24)	25(17,28)	2.356	0.019
体温[℃, $M(Q_L,Q_U)$]	36.9(36.6,37.2)	37.0(36.5,37.4)	2.252	0.074	36.8(36.6,37.2)	37.0(36.5,37.4)	0.946	0.346
SpO_2[%, $M(Q_L,Q_U)$]	97(94,99)	96(94,99)	2.570	0.024	96(94,98)	95(92,98)	1.283	0.201
实验室检查[$M(Q_L,Q_U)$]								
白细胞计数(×10^9/L)	11.9(7.9,17.0)	13.7(9.9,20.0)	3.461	<0.001	10.6(7.3,16.6)	12.9(10.1,22.4)	2.586	0.010
血红蛋白(g/dL, $\bar{x}\pm s$)	11.30±2.30	11.49±2.59	0.423	0.338	11.45±2.10	11.20±2.32	0.497	0.622
血小板计数(×10^9/L)	186(132,254)	190(127,249)	0.208	0.836	174(132,234)	205(137,225)	0.805	0.423
白蛋白(g/L)	3.1(2.7,3.5)	2.8(2.4,3.2)	6.047	<0.001	37.7±5.2	32.0±7.1	3.834	<0.001
肌酐(mg/dL)	1.0(0.7,1.5)	1.2(0.9,2.1)	4.482	<0.001	1.0(0.7,1.5)	1.7(1.0,2.6)	2.530	0.011
ALT(u/L)	59(26,127)	42(23,118)	1.763	0.078	84(34,147)	45(24,130)	1.182	0.238
葡萄糖(mg/dL)	125(102,161)	134(104,173)	1.515	0.130	133(105,165)	128(107,197)	0.021	0.986
Na^+(mmol/L)	138(135,141)	139(135,143)	2.570	0.010	138(135,141)	139(137,143)	1.673	0.095
K^+(mmol/L)	4.0(3.6,4.4)	4.1(3.7,4.7)	2.579	0.010	4.0(3.6,4.4)	4.0(3.5,4.6)	0.042	0.969
Cl^-(mmol/L)	104(100,108)	105(101,111)	3.220	0.001	104(100,108)	108(102,112)	2.202	0.028
INR	1.3(1.1,1.5)	1.3(1.2,1.6)	2.911	0.005	1.2(1.1,1.4)	1.4(1.2,1.6)	2.232	0.026
pH	7.38(7.32,7.43)	7.33(7.24,7.41)	5.251	<0.001	7.38(7.31,7.43)	7.33(7.21,7.38)	2.208	0.027
PCO_2(mmHg)	39(34,43)	42(35,48)	3.885	<0.001	38(34,42)	40(32,49)	0.673	0.503
剩余碱(mmol/L)	−1(−5,1)	−3(−8,0)	3.133	0.002	−2(−6,0)	−3(−10,0)	1.294	0.197
乳酸(mmol/L)	1.7(1.2,2.5)	1.9(1.3,3.0)	2.216	0.027	1.6(1.2,2.2)	2.3(1.3,3.3)	1.678	0.094
HCO_3^-(mmol/L)	22(18,25)	21(17,25)	1.781	0.075	22(18,25)	19(16,23)	1.353	0.177
功能评分[分, $M(Q_L,Q_U)$]								
GCS	15(14,15)	15(14,15)	0.164	0.870	15(14,15)	15(14,15)	0.345	0.075
SOFA	4(2,7)	8(5,12)	10.342	<0.001	4(2,7)	9(5,13)	3.356	<0.001
SAPSⅡ	31(22,42)	43(31,58)	8.751	<0.001	32(23,42)	48(38,62)	4.269	<0.001
并发症[例(%)]								
高血压	313(47.79)	91(45.05)	0.361	0.548	41(41.00)	10(38.46)	<0.001	0.991
糖尿病	57(8.70)	17(8.42)	0.016	0.899	12(12.00)	1(3.85)	1.483	0.223
脓毒症	356(54.35)	169(83.66)	54.667	<0.001	35(35.00)	16(61.54)	4.981	0.026
心肌梗死	10(1.53)	0	3.120	0.077	1(1.00)	0(0)	0.262	0.609
COPD	18(2.75)	8(3.96)	0.414	0.520	3(3.00)	2(7.69)	1.192	0.274
急性肾损伤	390(59.54)	172(85.15)	43.718	<0.001	58(58.00)	17(65.38)	0.211	0.646
治疗措施[例(%)]								
CRRT	56(8.55)	56(27.72)	48.282	<0.001	32(32.00)	15(57.69)	4.777	0.029
机械通气	144(21.98)	126(62.38)	114.857	<0.001	32(32.00)	11(42.31)	0.571	0.450
心肺复苏	9(1.37)	4(1.98)	0.380	0.538	1(1.00)	0	0.262	0.609
肝素	595(90.84)	194(96.04)	5.025	0.025	93(93.00)	24(92.31)	0.015	0.903
阿司匹林	212(32.37)	64(31.68)	0.009	0.924	28(28.00)	9(34.62)	0.175	0.676
抗生素	543(82.90)	197(97.52)	26.782	<0.001	79(79.00)	26(100)	6.552	0.010
血管活性药	116(17.71)	93(46.04)	65.667	<0.001	20(20.00)	15(57.69)	12.795	<0.001

注：SpO_2为脉搏血氧饱和度。

表 2　训练集和内部验证集 AP 患者基线资料比较

Tab.2　Comparison of baseline data of AP patients between training set and internal validation set

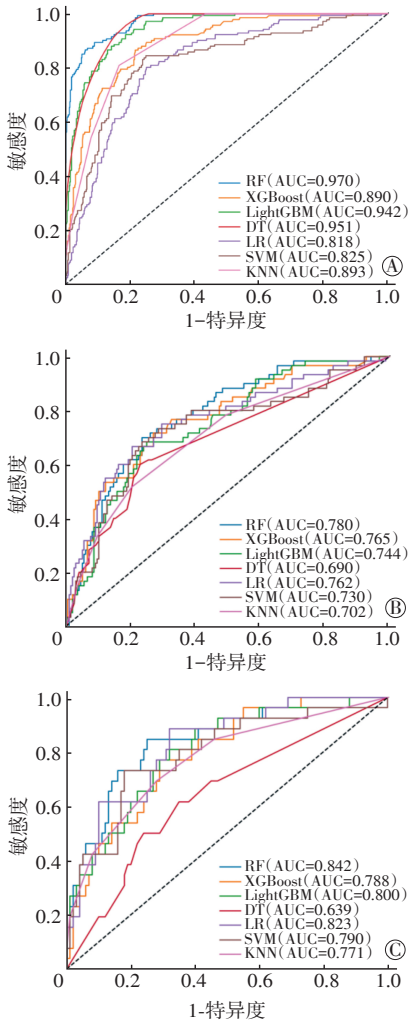
变量	内部验证集（n=256）	训练集（n=601）	$Z/\chi^2/t$ 值	P 值
人口学特征				
年龄[岁，$M(Q_L,Q_U)$]	63(47,75)	59(47,73)	1.013	0.311
男性[例(%)]	144(56)	353(59)	0.359	0.549
BMI[kg/m^2，$M(Q_L,Q_U)$]	28.20 (24.32,33.60)	28.40 (24.34,33.33)	0.161	0.869
医保[例(%)]	97(37.89)	187(31.15)	3.420	0.064
生命体征				
心率[次/min，$M(Q_L,Q_U)$]	95(81,113)	97(83,114)	1.143	0.253
舒张压(mmHg，$\bar{x}\pm s$)	126.09 ± 25.01	127.57 ± 25.50	0.678	0.498
收缩压(mmHg，$\bar{x}\pm s$)	72.54 ± 18.68	73.01 ± 18.88	0.342	0.733
呼吸[次/min，$M(Q_L,Q_U)$]	20.5(16,25)	20(17,25)	0.416	0.678
体温[℃，$M(Q_L,Q_U)$]	36.8(36.6,37.2)	36.9(36.6,37.3)	0.857	0.391
SpO_2[%，$M(Q_L,Q_U)$]	96(94,99)	97(94,99)	0.778	0.436
实验室检查[$M(Q_L,Q_U)$]				
白细胞计数(×10^9/L)	12.0(8.6,17.6)	12.4(8.2,17.8)	0.152	0.892
血红蛋白(g/dL，$\bar{x}\pm s$)	11.32 ± 2.39	11.35 ± 2.37	0.036	0.972
血小板计数(×10^9/L)	192(139,261)	185(127,250)	1.200	0.230
白蛋白(g/L)	3.1(2.6,3.5)	3(2.6,3.5)	0.084	0.933
肌酐(mg/dL)	1.1(0.7,1.7)	1(0.7,1.7)	0.699	0.485
ALT(U/L)	61(26,129)	54(25,118)	1.076	0.282
葡萄糖(mg/dL)	125(106,158)	127(101,167)	0.014	0.989
Na^+(mmol/L)	138(135,141)	138(135,141)	1.380	0.168
K^+(mmol/L)	4.0(3.6,4.4)	4.0(3.6,4.5)	0.243	0.808
Cl^-(mmol/L)	104(100,108)	104(100,109)	1.283	0.200
INR	1.3(1.1,1.4)	1.3(1.1,1.5)	1.283	0.200
pH	7.37(7.3,7.42)	7.37(7.30,7.43)	0.211	0.833
PCO_2(mmHg)	39(34,46)	39(34,44)	0.936	0.349
剩余碱(mmol/L)	-2(-6,0)	-2(-6,1)	0.144	0.885
乳酸(mmol/L)	1.7(1.1,2.8)	1.7(1.2,2.6)	0.201	0.841
HCO_3^-(mmol/L)	22(18,25)	22(18,25)	0.036	0.971
功能评分[分，$M(Q_L,Q_U)$]				
GCS	15(14,15)	15(14,15)	0.564	0.573
SOFA	5(3,7)	5(2,8)	1.014	0.311
SAPSⅡ	33(24,43)	34(23,46)	0.277	0.782
并发症[例(%)]				
高血压	117(45.70)	287(47.75)	0.226	0.634
糖尿病	23(8.98)	51(8.49)	0.011	0.916
脓毒症	147(57.42)	378(62.9)	2.041	0.153
心肌梗死	3(1.17)	7(1.16)	< 0.001	0.993
COPD	7(2.73)	19(3.16)	0.013	0.908
急性肾损伤	171(66.8)	391(65.06)	0.170	0.681
治疗措施[例(%)]				
CRRT	33(12.89)	79(13.14)	0.010	0.920
机械通气	77(30.08)	193(32.11)	0.257	0.612
心肺复苏	4(1.56)	9(1.50)	0.005	0.943
肝素	239(93.36)	550(91.51)	0.603	0.437
阿司匹林	87(33.98)	189(31.45)	0.419	0.517
抗生素	221(86.33)	519(86.36)	< 0.001	1.000
血管活性药	60(23.44)	149(24.79)	0.113	0.737



注：A 为临床特征系数图；B 为交叉验证图。

图 2　LASSO 回归筛选变量

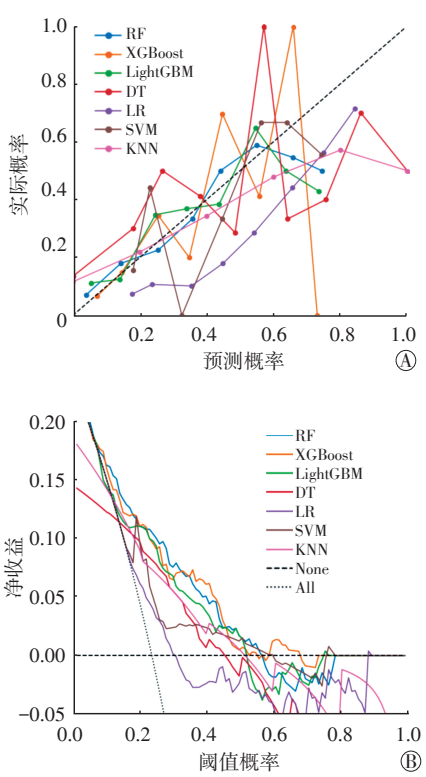
Fig.2　Variables screening by LASSO regression



注：A 为训练集；B 为内部验证集；C 为外部测试集。

图 3　7 种机器学习模型预测 AP 患者并发 ARDS 的 ROC 曲线

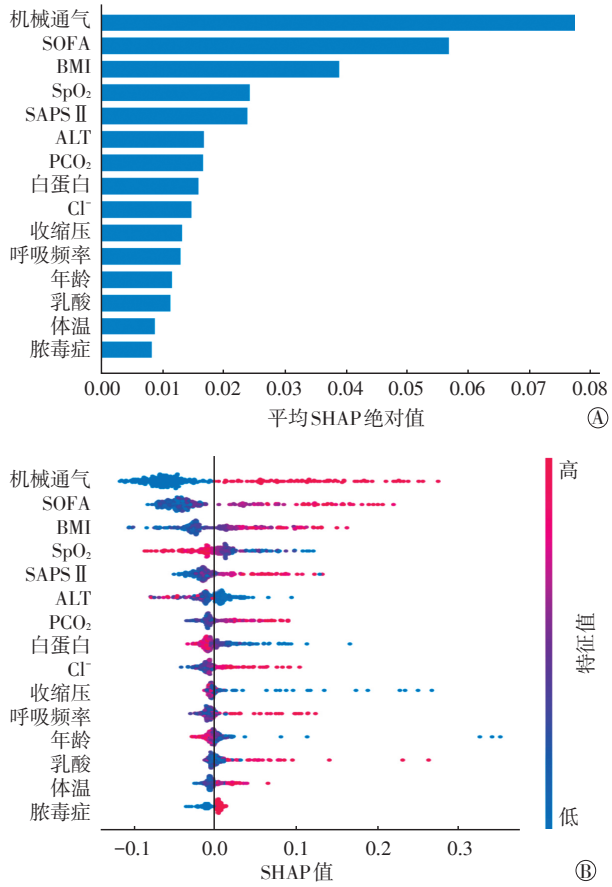
Fig.3　ROC curves of 7 machine learning models predicting ARDS in AP patients



注：A 为校正曲线；B 为 DCA 曲线。

图 4　7 种机器学习模型预测 AP 患者并发 ARDS 的校准曲线和 DCA 曲线

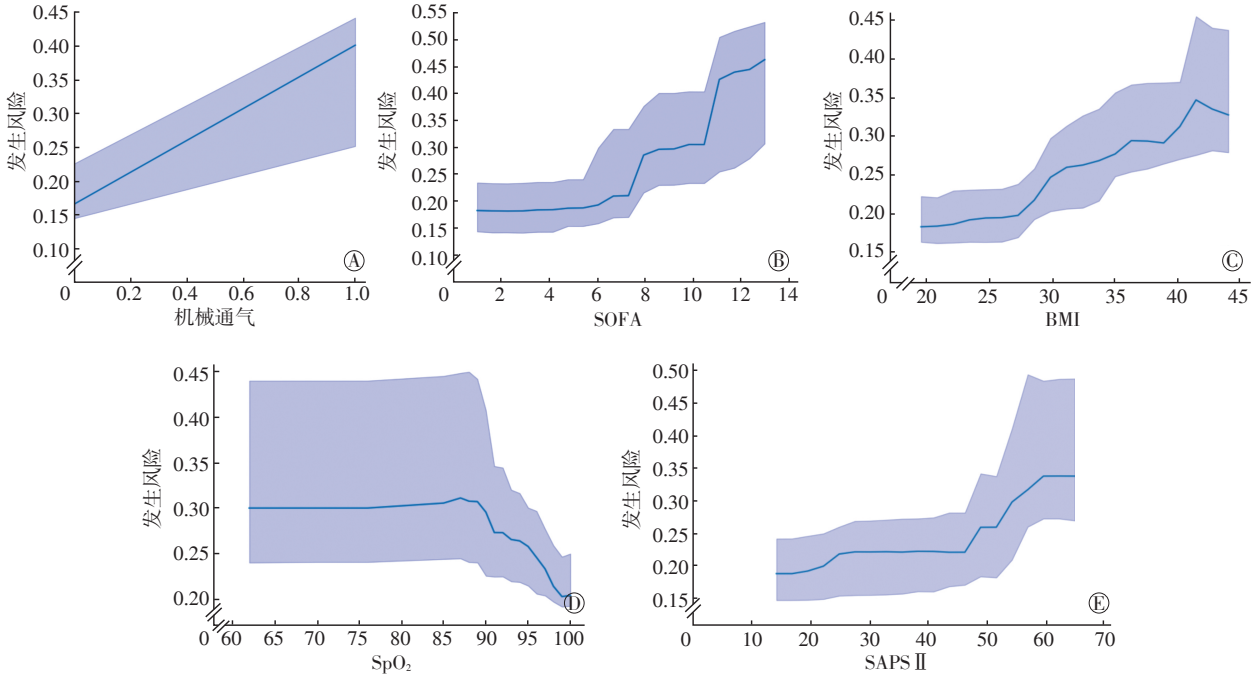
Fig.4　Calibration curves and DCA curves for predicting ARDS in AP patients by 7 machine learning models



注：A 为变量重要性图；B 为蜂群图。

图 5　变量重要性图及蜂群图

Fig.5　Variable importance diagram and bee colony diagram

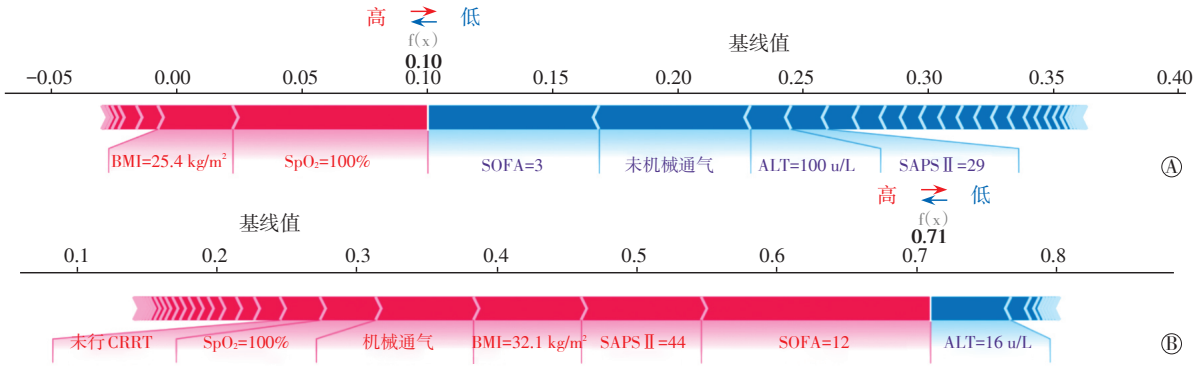


注：A 为机械通气；B 为 SOFA 评分；C 为 BMI；D 为 SpO_2；E 为 SAPS Ⅱ。

图 6　重要性排名前 5 位变量的 PDP 图

Fig.6　PDP of the top 5 variables in importance ranking

2.3.2　单样本预测特征解释　使用SHAP力图可视化个体预测特征贡献，图7为从内部验证集随机选择两个实例进行预测，病例A（实际阴性）预测风险10%显著低于基线，病例B（实际阳性）预测风险71%，明显高于基线，均与临床实际诊断一致，证实模型个体化预测的可靠性高。



注：A为预测低ARDS风险的患者实例；B为预测高ARDS风险的患者实例。

图7　随机森林模型的力图

Fig.7　Attempt of random forest model

表3　7种机器学习模型对AP患者并发ARDS的预测价值

Tab.3　Predictive value of 7 machine learning models for ARDS in AP patients

模型	AUC	95%*CI*	准确度	精度	敏感度	特异度	F1得分
RF	0.780	0.721~0.846	0.781	0.556	0.333	0.918	0.417
XGBoost	0.765	0.692~0.835	0.770	0.522	0.200	0.944	0.289
LightGBM	0.744	0.674~0.817	0.781	0.553	0.350	0.913	0.429
DT	0.690	0.622~0.762	0.766	0.500	0.333	0.898	0.400
LR	0.762	0.689~0.837	0.746	0.471	0.667	0.770	0.552
SVM	0.730	0.657~0.806	0.777	0.579	0.183	0.959	0.278
KNN	0.702	0.632~0.776	0.770	0.513	0.333	0.903	0.404

3　讨　论

本研究基于多中心数据构建的RF模型在预测AP患者并发ARDS风险中展现出优越性能（内部验证集AUC=0.780，外部测试集AUC=0.842）。SHAP揭示机械通气、SOFA评分、BMI、SpO_2和SAPS Ⅱ评分是影响模型决策的核心因素，为临床提供了可操作的预警指标。

目前已有多种实验室指标及临床评分系统被用于预测AP并发ARDS。降钙素原、白细胞计数、白蛋白等实验室指标，以及APACHE Ⅱ评分、Ranson评分、床旁胰腺炎严重度评分系统（BISAP）等临床评分工具，均被证实与AP并发ARDS相关[4, 6, 16-17]。然而，这些方法在预测效能和评估一致性上仍存在局限，难以满足临床对精准预测的需求。此外，传统回归模型，如广义线性回归中的LR和半参数生存分析模型中的Cox回归，尽管易于操作且具有一定可解释性，但这类模型通常需要对输入变量之间的线性和同质关系进行假设，难以充分捕捉变量之间的非线性关系，会过度简化非线性模型的交互效应[18]。此外，它们在分析大样本量数据时效率较低，误差与机器学习这些新技术相比偏大，其对预测因子间多重共线性的高敏感性也可能会降低模型性能[18-19]。

相比之下，机器学习技术能够建立自动化数据分析流程，高效地处理非线性及高维数据并揭示变量之间的复杂交互作用，从而提高预测的准确性[20]。随着大型数据库的广泛应用，机器学习在胰腺炎的临床诊断和预后评估中展现出良好的效果。Ren等[21]利用MIMIC-Ⅳ和eICU-CRD数据库构建预测ICU中AP患者住院死亡率的模型，结果显示高斯朴素贝叶斯（Gaussian Naive Bayes，GNB）模型在两个数据库上的AUC分别达到0.840和0.862。Liu等[12]使用双数据库MIMIC-Ⅲ和MIMIC-Ⅳ数据进行预测AP患者发生脓毒症风险的研究，也证明了机器学习方法优于传统LR模型及多种评分系统。本研究利用MIMIC-Ⅳ数据库进行模型构建及内部验证，通过LASSO回归算法进行特征选择，在此基础上开发7种机器学习算法，最终确定RF模型在高危ARDS患者的早期识别中表现出优越的预测性能。RF算法通过集成大量决策树，在分类任务中输出多数树的类别投票结果，在不平衡数据中展现出强大的分类能力和泛化性能[22-24]。笔者使用该RF模型在苏州大学附属常熟医院ICU收治的AP患者数据上进行外部测试，结果证明其在国内患者也能展现出良好的预测效能。

尽管机器学习算法在预测分析中具有较高的精度，但由于其“黑箱”特性，限制了其在临床环境中的应用[25]。此前有研究在利用机器学习技术对AP患者并发ARDS进行预测时，未能充分量化特征变量对

模型输出的影响，也缺乏有效的手段来可视化这些影响[26-27]。为克服这一不足，本研究基于SHAP算法对RF模型进行可解释性分析，提供了全局性和个性化解释，以便更好地理解模型的工作原理。结果显示最重要的5个变量依次为：机械通气、SOFA评分、BMI值、SpO_2水平和SAPSⅡ评分，这些变量可作为预测此类患者并发ARDS的关键指标。PDP图进一步直观地展示了单个特征对ARDS预测概率的影响，力图则实现了对随机选取特定个体情况下不同特征值的预测贡献分析，可帮助临床医师更深入地理解模型的决策过程。

本研究局限性：(1) 数据来自两个数据库，可能存在地域和中心偏差。(2) 外部测试集样本量较小，需多中心大样本验证泛化能力。(3) LASSO回归虽能有效减少特征数量，但也可能忽略潜在重要变量，不同的特征选择方法可能导致不同结果。(4) SHAP算法虽可提供模型解释性，但其解释结果依赖于模型预测准确性和数据质量，可能无法完全揭示复杂交互效应背后的生物学机制，需结合临床知识和实验验证以全面理解预测结果。

综上所述，本研究利用7种机器学习方法构建了早期预测AP患者并发ARDS的模型，其中RF模型表现出最优的预测性能和泛化能力。通过引入SHAP算法增强了模型的可解释性，为临床医师提供了直观可靠的辅助决策工具，有助于早期识别高危患者并优化治疗策略。

利益冲突　无

参考文献

[1] Sohail Z, Shaikh H, Iqbal N, et al. Acute pancreatitis: a narrative review[J]. J Pak Med Assoc, 2024, 74(5): 953-958.

[2] Trikudanathan G, Yazici C, Evans Phillips A, et al. Diagnosis and management of acute pancreatitis[J]. Gastroenterology, 2024, 167(4): 673-688.

[3] Petrov MS, Yadav D. Global epidemiology and holistic prevention of pancreatitis[J]. Nat Rev Gastroenterol Hepatol, 2019, 16(3): 175-184.

[4] Ding N, Guo CR, Song K, et al. Nomogram for the prediction of in-hospital incidence of acute respiratory distress syndrome in patients with acute pancreatitis[J]. Am J Med Sci, 2022, 363(4): 322-332.

[5] 周游，路国涛，王丹，等. 急性胰腺炎并发急性呼吸窘迫综合征预测模型的研究进展[J]. 中华胰腺病杂志，2022，22(3)：231-234.

[6] Zhang WW, Zhang M, Kuang ZM, et al. The risk factors for acute respiratory distress syndrome in patients with severe acute pancreatitis: a retrospective analysis[J]. Medicine (Baltimore), 2021, 100(2): e23982.

[7] Fei Y, Gao K, Li WQ. Prediction and evaluation of the severity of acute respiratory distress syndrome following severe acute pancreatitis using an artificial neural network algorithm model[J]. HPB (Oxford), 2019, 21(7): 891-897.

[8] 郭莎莎，雷静静，王玉洁，等. 发生全身炎症反应综合征的急性胰腺炎血液净化治疗时机选择[J]. 中国临床研究，2024，37(7)：1090-1096.

[9] 高欣，林嘉希，吴爱荣，等. 基于XGBoost算法的机器学习模型在早期预测重症急性胰腺炎中的应用[J]. 中华危重病急救医学，2023，35(4)：421-426.

[10] Kui B, Pintér J, Molontay R, et al. EASY-APP: an artificial intelligence model and application for early and easy prediction of severity in acute pancreatitis[J]. Clin Transl Med, 2022, 12(6): e842.

[11] Lin SW, Lu WB, 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[J]. Ren Fail, 2024, 46(1): 2303395.

[12] Liu F, Yao J, Liu CY, et al. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis[J]. BMC Surg, 2023, 23(1): 267.

[13] Lundberg SM, Lee SI. A unified approach to interpreting model predictions[EB/OL]. arXiv, (2017-05-22)[2025-04-01]. https://arxiv.org/abs/1705.07874v2.

[14] 中华医学会外科学分会胰腺外科学组. 中国急性胰腺炎诊治指南(2021)[J]. 中华外科杂志，2021，59(7)：578-587.

[15] ARDS Definition Task Force. Acute respiratory distress syndrome: the Berlin Definition[J]. JAMA, 2012, 307(23): 2526-2533.

[16] Lin FY, Lu RL, Han DD, et al. A prediction model for acute respiratory distress syndrome among patients with severe acute pancreatitis: a retrospective analysis[J]. Ther Adv Respir Dis, 2022, 16: 17534666221122592.

[17] Li YL, Zhang DD, Xiong YY, et al. Development and external validation of models to predict acute respiratory distress syndrome related to severe acute pancreatitis[J]. World J Gastroenterol, 2022, 28(19): 2123-2136.

[18] Du M, Haag DG, Lynch JW, et al. Comparison of the tree-based machine learning algorithms to Cox regression in predicting the survival of oral and pharyngeal cancers: analyses based on SEER database[J]. Cancers (Basel), 2020, 12(10): 2802.

[19] Yang DL, Kang J, Li YH, et al. Development of a predictive nomogram for acute respiratory distress syndrome in patients with acute pancreatitis complicated with acute kidney injury[J]. Ren Fail, 2023, 45(2): 2251591.

[20] 殷民月，朱锦舟，刘璐，等. 急性胰腺炎机器学习模型的研究进展[J]. 临床肝胆病杂志，2023，39(12)：2978-2984.

[21] Ren WS, Zou K, Huang S, et al. Prediction of in-hospital mortality of intensive care unit patients with acute pancreatitis based on an explainable machine learning algorithm[J]. J Clin Gastroenterol, 2024, 58(6): 619-626.

[22] Zhang WY, Chang YJ, Ding Y, et al. To establish an early prediction model for acute respiratory distress syndrome in severe acute pancreatitis using machine learning algorithm[J]. J Clin Med, 2023, 12(5): 1718.

[23] Masetic Z, Subasi A. Congestive heart failure detection using random forest classifier[J]. Comput Methods Programs Biomed, 2016, 130: 54-64.

[24] Roguet A, Eren AM, Newton RJ, et al. Fecal source identification using random forest[J]. Microbiome, 2018, 6(1): 185.

[25] The Lancet Respiratory Medicine. Opening the black box of machine learning[J]. Lancet Respir Med, 2018, 6(11): 801.

[26] 邹康，任文森，蒲鑫鑫，等. 机器学习在预测重症急性胰腺炎发生急性呼吸窘迫综合征中的作用[J]. 中华胰腺病杂志，2023，23(6)：455-459.

[27] Zou K, Ren WS, Huang S, et al. The role of artificial neural networks in prediction of severe acute pancreatitis associated acute respiratory distress syndrome: a retrospective study[J]. Medicine (Baltimore), 2023, 102(29): e34399.

收稿日期：2025-04-16　**修回日期**：2025-05-15　**编辑**：叶小舟