



OPEN ACCESS

EDITED BY

Francisco Epelde,
Parc Taulí Foundation, Spain

REVIEWED BY

Lixia Wang,
Cedars Sinai Medical Center, United States
Md. Enamul Hoq,
University of Arkansas for Medical Sciences,
United States
Wei Jun Dan Ong,
National University Health System, Singapore

*CORRESPONDENCE

Qiang Tang
✉ tangqiang678@zju.edu.cn
Shurui Wang
✉ shurui20200101@163.com
Chuanming Zheng
✉ zcm2611248@163.com

[†]These authors have contributed equally to this work

RECEIVED 04 June 2025

ACCEPTED 22 July 2025

PUBLISHED 12 August 2025

CITATION

Du Z, Ying Q, Yang Y, Ma H, Zhao H, Yang J, Wang Z, Zheng C, Wang S and Tang Q (2025) Machine learning-based predictive model for acute pancreatitis-associated lung injury: a retrospective analysis. *Front. Med.* 12:1638097. doi: 10.3389/fmed.2025.1638097

COPYRIGHT

© 2025 Du, Ying, Yang, Ma, Zhao, Yang, Wang, Zheng, Wang and Tang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning-based predictive model for acute pancreatitis-associated lung injury: a retrospective analysis

Zhaohui Du^{1†}, Qiaoling Ying^{2†}, Yisen Yang^{3†}, Huicong Ma¹, Hongchang Zhao¹, Jie Yang¹, Zhenjie Wang¹, Chuanming Zheng^{1*}, Shurui Wang^{4*} and Qiang Tang^{4*}

¹Department of Emergency Surgery, The First Affiliated Hospital of Bengbu Medical University, Bengbu, Anhui, China, ²Department of Radiation Oncology, The First Affiliated Hospital of Bengbu Medical University, Bengbu, Anhui, China, ³Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing, China, ⁴The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China

Background: Acute Pancreatitis-Associated Lung Injury (APALI) is one of the most severe and life-threatening systemic complications in acute pancreatitis patients, with high rates of morbidity and mortality. This study aims to develop a prediction model for the diagnosis of APALI based on machine learning algorithms.

Methods: This study included data from the First Affiliated Hospital of Bengbu Medical College (July 2012 to June 2022), which were randomly categorized into the training and testing set. And data from the Second Affiliated Hospital of Zhejiang University (January 2018 to April 2023) served as the external validation set. LASSO regression was applied to eliminate irrelevant or highly collinear independent variables. Six machine learning models were constructed, with evaluation metrics including Area Under Curve (AUC), accuracy, sensitivity, specificity, F1 score, and recall. The impact of model features was analyzed using SHapley Additive exPlanations (SHAP).

Results: A total of 1,975 patients with acute pancreatitis were randomly assigned to a training set (1,480 patients) and a testing set (495 patients). In the training set, 480 cases (32.43%) were diagnosed with APALI. The eXtreme Gradient Boosting (XGBoost) and Random Forest (RF) models demonstrated the best predictive performance, achieving the highest AUC (0.92 and 0.914, respectively), along with higher accuracy, F1 score, and recall in the testing set. Six particularly influential factors were identified and ranked as follows: CRP, BMI, neutrophil, calcium, lactate, and neutrophil-to-albumin ratio (NAR). The global interpretability of the XGBoost and RF models, along with these six features, is shown in the SHAP summary plot. These two models were selected as the optimal models for the development of an online calculator for clinical applications and risk stratification.

Conclusion: We developed and internally validated a machine learning model to predict APALI, showing strong performance in our study population. To support further research and clinical use, we created an open-access web-based risk calculator. Prospective multicenter validation is needed to confirm generalizability. If successful, the tool may support early risk identification and guide interventions to prevent APALI.

KEYWORDS

prediction model, machine learning, SHAP, acute pancreatitis (AP), lung injury

Introduction

Acute pancreatitis is characterized by abdominal pain, distension, nausea, vomiting, and systemic complications. The incidence of AP has been steadily increasing, with approximately 20% of patients progressing to severe acute pancreatitis (SAP) (1, 2). SAP can lead to a range of complications, among which acute lung injury is particularly severe (3), highlighting the need for early prediction of such outcomes. A multicenter retrospective study demonstrated that 92% of SAP may develop acute respiratory distress syndrome (ARDS), with a mortality rate of 37% (4), which underscores the clinical importance of identifying high-risk patients at an early stage. If not properly managed, ARDS can progress to multiple organ failure (MOF), which poses a significant threat to the patient's life (5). Despite the availability of various treatments for acute pancreatitis-associated acute lung injury (APALI), the mortality and morbidity rates remain high (6). Timely identification and intervention could mitigate or even alleviate acute lung injury, reducing both patient suffering and the economic burden (7, 8). Therefore, developing a clinical prediction model that can accurately identify lung injury at an early stage in AP is essential for improving patient outcomes. Given the multifaceted nature of acute lung injury, characterized by diverse clinical and biological features, no single clinical indicator can adequately represent the disease status (9). Recently, few studies have established a prediction model to identify APALI in patients with AP, for instance, Samanta et al. (10) proposed IL-6 and IL-8 as potential biomarkers for lung injury in AP, while Jia et al. (11) developed a nomogram-based tool using routine clinical data—however, both studies were limited by small sample sizes and lacked external validation. These limitations provided a strong rationale for the development of a robust, interpretable, and externally validated machine learning model as we present in this study.

In recent years, machine learning (ML) has been increasingly applied in critical care, offering notable advantages over conventional statistical methods. Evidence indicates that ML contributes significantly to the early diagnosis, severity assessment, and personalized treatment of AP (12–14). Algorithms such as logistic regression, random forests, and support vector machines each exhibit unique strengths in processing medical data (15, 16). Ong et al. demonstrated the utility of ML by developing an XGBoost-based model that outperformed traditional methods in predicting adverse outcomes—including readmission, mortality, and prolonged hospitalization—among patients requiring mechanical ventilation for more than 4 hours. The model achieved an AUC of 0.693 compared to 0.667 for conventional approaches ($p = 0.03$) and showed a 6.8% improvement in sensitivity at 95% specificity. Their study also highlighted important predictors, such as the Glasgow Coma Scale and duration of mechanical ventilation, while emphasizing the critical role of external validation and model interpretability in AI-driven applications within critical care (17). Although current ML models have limited accuracy in predicting APALI, ongoing technological advancements and the growing availability of electronic medical records are expected to enhance the precision

and timeliness of clinical decision support—particularly for identifying high-risk patients and supporting individualized risk assessments.

This study developed a clinical prediction model for APALI using machine learning algorithms and validated its performance on external datasets. By leveraging extensive clinical data, the model identified critical indicators associated with APALI risk, providing early warnings for physicians. This research may accelerate the early prediction of lung injury in AP and supports medical teams in implementing targeted interventions, ultimately improving recovery rates and quality of life.

Methods

Study population

This study included AP cases admitted to the Emergency Surgery Department of the First Affiliated Hospital of Bengbu Medical University from July 2012 to June 2022, which were used for both the training and validation cohorts. Additionally, cases from the Second Affiliated Hospital of Zhejiang University, admitted between January 2018 to April 2023, served as the external validation cohort. All patient data were de-identified prior to analysis. The study was approved by the ethics committees of both participating hospitals (2020KY073), and informed consent was obtained from all enrolled patients or their legally authorized representatives in accordance with the Declaration of Helsinki.

Inclusion Criteria: Diagnosis of AP was based on the consensus of the International Association of Pancreatology (IAP), requiring at least three of the following conditions: typical clinical manifestations (e.g., abdominal pain); elevated serum amylase and/or lipase levels (generally exceeding three times the normal value), and imaging findings consistent with pancreatitis. **Exclusion Criteria:** Traumatic pancreatitis, acute exacerbation of chronic obstructive pulmonary disease (AECOPD), ARDS due to causes other than pancreatitis, and age under 14 years.

Definition

According to the guidelines of the American Thoracic Society (ATS) and the European Respiratory Society (ERS), acute lung injury is defined as follows (18–21): (1) rapid onset, with acute respiratory distress typically occurring within hours; (2) hypoxemia is defined as a $\text{PaO}_2/\text{FiO}_2$ ratio of 200–300 mmHg; (3) Pulmonary infiltrates on chest X-rays or CT scans, excluding cardiogenic causes; and (4) The absence of cardiogenic pulmonary edema confirms that the injury is not due to heart failure or fluid overload. Diagnosis of acute lung injury was confirmed independently by two board-certified radiologists who were blinded to patient clinical outcomes. In cases of disagreement, a third senior radiologist adjudicated the final classification by consensus. Standardized diagnostic criteria were uniformly applied across all cases.

Data collection

Clinical data were collected from patients with AP, including vital signs, demographic information, CT grade, and laboratory makers such as calcium (Ca^{2+}), blood glucose (BG), lactate (Lac), serum lipase (LPS), serum amylase (AMY), urinary amylase (UAMY), triglycerides (TG), total cholesterol (TC), procalcitonin (PCT), heparin-binding protein (HBP), C-reactive protein (CRP), albumin, globulin, and the neutrophil-to-albumin ratio (NAR). Data on the presence of mechanical ventilation, oxygen partial pressure, concentration of inspired oxygen, and pleural effusion were also collected.

All laboratory variables were standardized to be collected within 24 h of hospital admission to ensure temporal consistency across patients. Elective admissions were excluded from the analysis to avoid confounding due to different baseline risk profiles and disease progression patterns. CT imaging was independently reviewed by two board-certified radiologists with 5–10 years of experience in abdominal imaging. Severe AP was defined as a CT Severity Index (CTSI) score ≥ 5 , in accordance with the Revised Atlanta Classification. Discrepancies in CT grading were resolved by consensus with a senior radiologist with over 15 years of clinical experience. All CT evaluations and clinical/laboratory data were obtained within the first 24 h of hospital admission. Despite the class imbalance, no oversampling or class weighting was applied during analysis.

Model construction and evaluation

To deploy the clinical predictive model, we first applied 10-fold cross-validated LASSO regression for preliminary feature selection. This step aimed to eliminate variables with low contributions and address multicollinearity among highly correlated predictors. The number of selected variables was determined using the λ that minimized the cross-validation error. These selected features were then used as inputs for subsequent machine learning models.

We constructed six machine learning models, including Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Network (NNET). During the data preprocessing stage, missing values were imputed with either the median or mode, depending on the variable type, and categorical variables were encoded using one-hot encoding. The dataset was split into training and test sets in a 3:1 ratio. Hyperparameter optimization was conducted using grid search with 25 bootstrapped resamples of the training set to ensure robust tuning.

Model performance was assessed using multiple metrics, including the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, F1 score, and recall. To further ensure reliability, external validation data were used, and AUC, accuracy, sensitivity, specificity, F1 score, and recall were recalculated.

Model interpretation

To clarify the contribution of each feature to the final model, we used Shapley Additive Explanations (SHAP) to interpret and visualize the impact of individual variables. We assessed feature importance by calculating the mean absolute SHAP values.

Additionally, we plotted SHAP values for each feature across all samples to better understand the overall patterns and the influence of features on the dataset. Three SHAP examples were provided to illustrate these concepts.

Statistical analysis

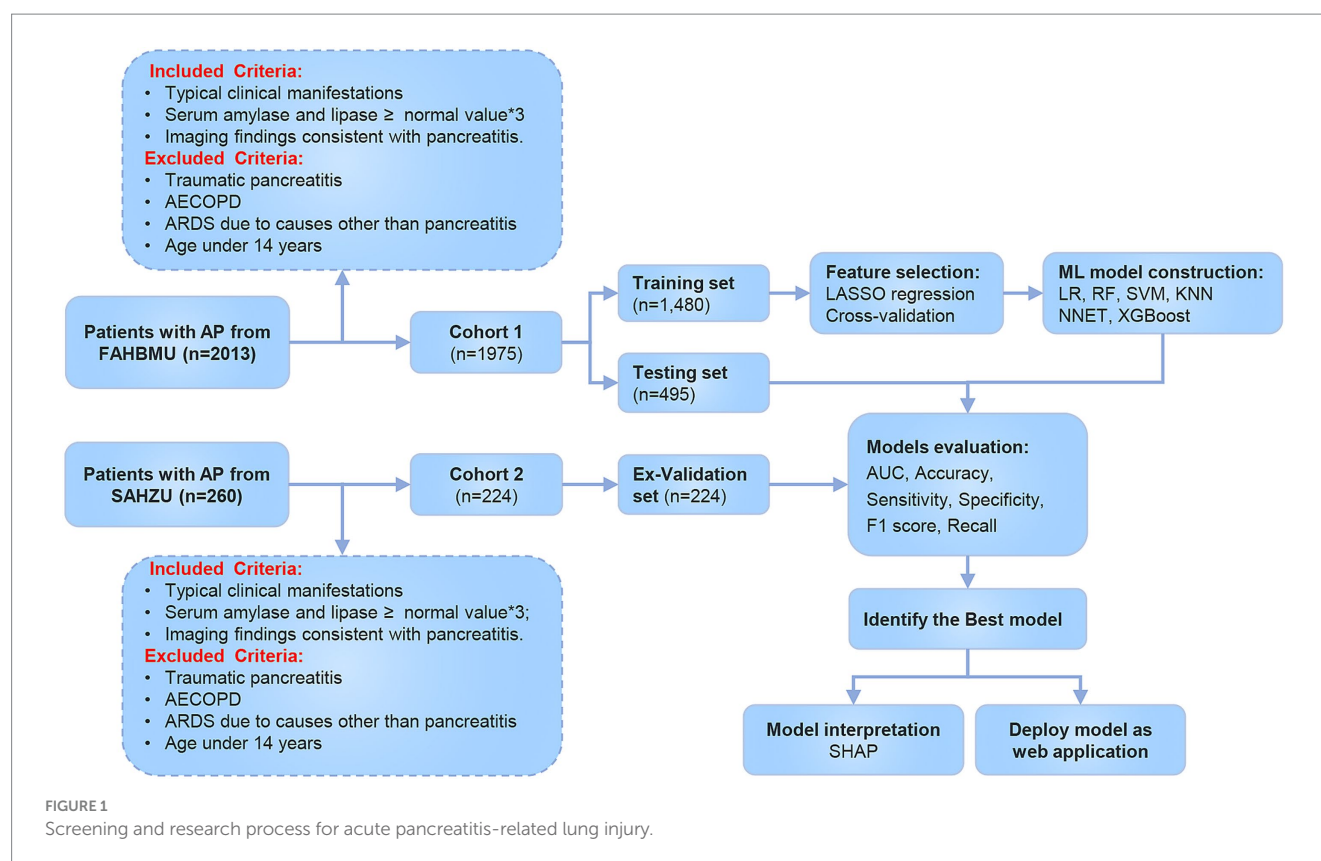
Binary variables were summarized as counts and proportions, and comparisons were performed using chi-square tests or Fisher's exact tests, as appropriate. Continuous variables with a normal distribution were compared using independent t-tests, with results presented as mean \pm standard deviation. For non-normally distributed variables, the Mann–Whitney U test was employed. A p -value < 0.05 was considered indicative of statistical significance. All statistical analyses were conducted using R (v4.2.3), leveraging the packages “tidymodels,” “glmnet,” “kernelshap,” and “shapviz.”

Results

Patient characteristics

The flowchart of the study is shown in [Figure 1](#). A total of 1,975 patients with AP from the First Affiliated Hospital of Bengbu Medical University were included in the study. The patients were randomly assigned to a training set (1,480 cases) and a validation set (495 cases) in a 3:1 ratio ([Table 1](#)). In the training set, 480 cases (32.43%) of APALI were identified, consisting of 268 males (55.83%) and 212 females (44.17%). Among the 1,000 cases (67.57%) without APALI, 566 were male (56.60%) and 434 were female (43.40%) ([Table 2](#)). The gender distribution did not differ significantly between the two groups. [Table 2](#) and [Supplementary Table S1](#) summarize the baseline characteristics of the training and testing set. Compared to the non-APALI group, patients in the APALI group were significantly younger (50 [38, 66] vs. 47 [35, 65]) ($p < 0.05$). In addition, the APALI group showed significantly higher levels of calcium ions, neutrophil count, lymphocyte count, lactate, BMI, pulse, blood glucose, CRP, and other markers. In comparison, albumin levels were notably lower ($p < 0.05$). More importantly, levels of NLR, CRP, Triglyceride-Glucose Index, PLR, NPR, NAR, blood amylase, and urinary amylase were significantly higher in the APALI group compared to the non-APALI group ($p < 0.05$). The APALI group also had a higher prevalence of pleural effusion. In the testing set of 495 patients, 161 (32.53%) had APALI, with 93 males (57.76%) and 68 females (42.24%). In the non-APALI group ($n = 334$), 182 were males (54.49%) and 152 were females (45.51%). The gender distribution was similar to the training set, with consistent findings in the validation set ([Supplementary Table S1](#)).

For the external validation, 224 additional patients with AP were recruited from the Second Affiliated Hospital of Zhejiang University ([Supplementary Table S2](#)). The cohort included 151 cases in the non-APALI group (70.20% males and 29.80% females) and 73 cases in the APALI group (66.75% males and 34.25% females). The characteristics of this cohort were consistent with those of the training set ($p < 0.05$). Compared to the training and testing set, patients in the external validation set were significantly older and had markedly higher levels of Ca^{2+} , neutrophil count, lymphocyte count, lactate, and



CRP further supporting the external validation cohort better represents a wide range of clinical scenarios, enhancing the generalizability and reliability of the study findings.

Feature selection

We performed LASSO regression analysis combined with cross-validation to identify potential risk factors associated with APALI. LASSO regression incorporates a penalty term to mitigate multicollinearity, optimize variable selection, and enhance model stability and interpretability. After the LASSO regression selection, 25 candidate predictors were determined, including Age, Respiratory Rate, Calcium ions, Platelet, Neutrophil count, Lymphocyte, Globulin, Cholesterol, Pleural effusion, CT grade, Lactate, BMI, Temperature, Pulse, SBP, DBP, Red blood cell width, Blood Glucose, NLR, CRP, and NAR (Figures 2a,b). These variables were used as potential predictors in development of the machine learning model. Six machine learning methods—logistic regression, random forest, XGBoost, SVM, KNN, NNET—were applied to construct the risk models.

Model evaluation in the training/testing set and the external validation set

The comparative performance of all models is summarized in Tables 3, 4 and Figures 3a–f. After evaluating six machine learning algorithms on both training and testing datasets, XGBoost was selected as the primary model due to its superior performance on the test set, achieving an AUC of 0.91 (95% CI: 0.89–0.94), an F1

score of 0.90 (95% CI: 0.88–0.92), a sensitivity of 0.95 (95% CI: 0.92–0.97), and a specificity of 0.68 (95% CI: 0.61–0.75). The random forest (RF) model attained the highest AUC of 0.92 (95% CI: 0.89–0.94) and was retained as a secondary tool for reassessing uncertain cases. Both tree-based models outperformed linear models (e.g., LR: AUC = 0.90 [95% CI: 0.87–0.93], F1 = 0.88 [95% CI: 0.86–0.91]) and (NNET: AUC = 0.81 [95% CI: 0.77–0.85], F1 = 0.82 [95% CI: 0.78–0.85]) across all metrics. The comparable recall (XGBoost: 0.95 [95% CI: 0.92–0.97]; RF: 0.93–0.97) and F1 scores (XGBoost: 0.90; RF: 0.89) further demonstrate their robust feature-learning capabilities. Overall, XGBoost appears more suitable for clinical applications, whereas RF may serve as a complementary tool for exploratory or confirmatory analyses. The ROC curves for all models in the test set are presented in Figure 3g.

To evaluate the generalizability of our model, we tested its performance on an external validation set to classify patients with APALI and non-APALI. The performance of the models is summarized in Supplementary Table S3. All models in the independent external validation set demonstrated strong discriminative performance, with ROC-AUC values consistently exceeding 0.750. Among them, the RF model exhibited statistically superior performance (DeLong's test, $p^* < 0.001$). Notably, the XGBoost model also displayed competitive performance, with marginally higher metrics compared to the remaining models, including an AUROC of 0.990, accuracy (0.953), F1 score (0.900), recall (0.987), sensitivity (0.987), and specificity (0.975). The result indicated that the XGBoost and RF models exhibited good performance and clinical utility, whereas the SVM, KNN, and NNET models showed relatively lower performance Supplementary Figure S1.

TABLE 1 The characteristics of patients with APALI.

Characteristic	Training set (<i>N</i> = 1,480) ^a	Testing set (<i>N</i> = 495) ^a	External validation set (<i>N</i> = 224) ^a
Outcome			
Non-APALI	1,000 (68.57%)	334 (67.47%)	151 (67.41%)
APALI	480 (32.43%)	161 (32.53%)	73 (32.59%)
Gender			
Male	834 (56.35%)	275 (55.56%)	154 (68.75%)
Female	646 (43.65%)	220 (44.44%)	70 (31.25%)
Age (years)	49 (37, 66)	49 (38, 65)	53 (41, 66)
Respiratory rate (cpm)	21.20 (2.81)	21.07 (2.70)	20.71 (13.03)
Ca ²⁺ (mmol/l)	1.07 (1.00, 1.15)	1.08 (1.01, 1.15)	1.76 (1.26, 2.07)
Platelet (×10 ⁹ /l)	206.0 (159.25, 256.0)	194.0 (150.25, 255.0)	199.0 (159.0, 269.5)
Neutrophil (×10 ⁹ /l)	9.84 (6.95, 13.79)	9.78 (6.40, 13.42)	11.20 (7.90, 16.01)
Lymphocyte (×10 ⁹ /l)	1.12 (0.79, 1.57)	1.15 (0.78, 1.62)	1.05 (0.75, 1.51)
Globulin (g/l)	33.50 (29.7, 37.5)	33.15 (28.98, 37.43)	31.70 (28.83, 35.23)
Albumin (g/l)	39.30 (34.6, 43.0)	39.15 (35.0, 43.2)	35.90 (31.0, 40.5)
Total cholesterol (mmol/l)	4.01 (3.13, 5.46)	3.95 (3.11, 5.35)	3.84 (2.87, 5.22)
Pleural effusion	488 (32.97%)	177 (35.76%)	67 (29.91%)
CTSI			
A	42 (2.84%)	12 (2.42%)	44 (19.64%)
B	758 (51.22%)	256 (51.72%)	79 (35.27%)
C	618 (41.75%)	213 (43.03%)	51 (22.77%)
D	62 (4.19%)	14 (2.83%)	14 (6.25%)
E	0 (0%)	0 (0%)	36 (16.07%)
Lactate (mmol/l)	1.28 (0.84, 2.23)	1.30 (0.84, 2.21)	1.90 (1.08, 2.75)
BMI	25.15 (3.09)	25.08 (3.15)	24.55 (3.52)
Temperature (°C)	36.88 (0.54)	36.84 (0.49)	36.89 (1.09)
Pulse (bpm)	92.95 (82.0, 100.0)	92.95 (81.0, 99.0)	92.95 (84.0, 101.0)
SBP (mmHg)	133.51 (126.0, 140.0)	133.51 (125.0, 139.0)	133.51 (127.0, 141.0)
DBP (mmHg)	79.19 (76.0, 87.0)	79.19 (76.0, 88.0)	79.19 (76.0, 86.0)
RDW	13.69 (1.40)	13.65 (1.29)	13.63 (1.25)
Blood glucose (mmol/l)	7.82 (6.15, 11.23)	7.83 (6.01, 10.96)	7.33 (5.85, 9.82)
NLR	8.91 (5.15, 14.23)	8.65 (5.11, 13.67)	11.07 (6.86, 15.75)
C-reactive protein (mg/l)	46.92 (12.29, 90.0)	45.00 (11.18, 90.0)	137.00 (35.0, 229.90)
TyG	3.44 (3.07, 3.98)	3.42 (3.01, 3.96)	2.71 (2.27, 3.18)
PLR	179.08 (126.49, 260.76)	180.75 (123.17, 245.23)	180.61 (110.37, 283.96)
NPR	0.05 (0.03, 0.07)	0.05 (0.03, 0.07)	0.05 (0.04, 0.08)
NAR	0.26 (0.18, 0.37)	0.25 (0.17, 0.36)	0.32 (0.22, 0.46)
Amylase (ln) ^b	5.37 (4.39, 6.42)	5.51 (4.44, 6.50)	5.63 (4.63, 6.48)
Urinary Amylase (ln) ^b	6.61 (5.51, 8.30)	6.86 (5.82, 8.47)	5.96 (5.25, 6.96)
Triglyceride (ln) ^b	0.52 (0.08, 1.33)	0.52 (−0.03, 1.28)	0.46 (0.12, 1.05)
Procalcitonin (ng/ml)	−0.48 (−2.12, 0.66)	−0.63 (−2.21, 0.66)	0.75 (0.23, 1.72)
SII (ln) ^b	10.53 (9.98, 11.03)	10.47 (9.95, 10.92)	7.75 (7.13, 8.26)

^aMean (SD), Median (Q1, Q3); *n* (%).^bln, Natural logarithm.

CTSI, CT Severity Index; BMI, Body Mass Index; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; RDW, Red Cell Distribution Width; NLR, Neutrophil-to-Lymphocyte Ratio; TyG, Triglyceride-Glucose Index; PLR, Platelet-to-Lymphocyte Ratio; NPR, Neutrophil-to-Platelet Ratio; NAR, Neutrophil-to-Albumin Ratio; SII, Systemic Inflammation Index.

TABLE 2 The characteristics of patients with acute pancreatitis in the training set.

Characteristic	Non-APALI (N = 1,000) ^a	APALI (N = 480) ^a	P-value ^b
Gender			0.80
Male	566 (56.60%)	268 (55.83%)	
Age (years)	50 (38, 66)	47 (35, 65)	0.02
Respiratory rate (cpm)	21.08 (2.76)	21.48 (2.89)	0.002
Ca ²⁺ (mmol/l)	1.07 (1.00, 1.13)	1.10 (1.00, 1.21)	<0.001
Platelet (×10 ⁹ /l)	204.0 (157.75, 253.0)	211.0 (162.0, 263.75)	0.30
Neutrophil (×10 ⁹ /l)	8.84 (5.88, 12.07)	12.80 (9.76, 16.90)	<0.001
Lymphocyte (×10 ⁹ /l)	1.10 (0.78, 1.52)	1.16 (0.82, 1.67)	0.05
Globulin (g/l)	33.60 (29.80, 37.70)	33.40 (29.50, 37.20)	0.80
Albumin (g/l)	40.30 (35.70, 43.45)	37.10 (32.60, 41.50)	<0.001
Total cholesterol (mmol/l)	4.05 (3.24, 5.46)	3.94 (2.94, 5.46)	0.06
Pleural Effusion	262 (26.20%)	226 (47.08%)	<0.001
CTSI			<0.001
A	37 (3.70%)	5 (1.04%)	
B	606 (60.60%)	152 (31.67%)	
C	338 (33.80%)	280 (58.33%)	
D	19 (1.90%)	43 (8.96%)	
Lactate (mmol/l)	1.15 (0.80, 2.01)	1.56 (0.94, 2.90)	<0.001
BMI	24.52 (3.08)	26.44 (2.70)	<0.001
Temperature (°C)	36.85 (0.51)	36.96 (0.58)	<0.001
Pulse (bpm)	92.95 (80.0, 98.0)	92.95 (87.0, 102.0)	<0.001
SBP (mmHg)	133.51 (125.0, 141.0)	133.51 (126.75, 138.0)	0.80
DBP (mmHg)	79.19 (76.0, 87.0)	79.19 (76.0, 85.0)	0.20
RDW	13.63 (1.41)	13.81 (1.37)	<0.001
Blood glucose (mmol/l)	7.68 (6.11, 10.76)	8.22 (6.31, 11.96)	0.019
NLR	8.03 (4.42, 13.06)	10.63 (7.31, 16.65)	<0.001
C-reactive protein (mg/l)	27.65 (7.83, 76.18)	90.0 (48.72, 90.0)	<0.001
TyG	3.44 (3.08, 3.93)	3.44 (3.03, 4.09)	0.80
PLR	181.0 (128.67, 264.01)	174.59 (124.19, 251.53)	0.30
NPR	0.04 (0.03, 0.06)	0.06 (0.04, 0.09)	<0.001
NAR	0.23 (0.15, 0.31)	0.35 (0.25, 0.46)	<0.001
Amylase (ln) ^c	5.51 (4.45, 6.55)	5.05 (4.28, 6.12)	<0.001
Urinary amylase (ln) ^c	6.70 (5.59, 8.36)	6.42 (5.34, 8.04)	0.004
Triglyceride (ln) ^c	0.48 (0.07, 1.28)	0.62 (0.08, 1.38)	0.13
Procalcitonin (ng/ml)	−0.40 (−2.21, 0.66)	−0.56 (−1.97, 0.66)	0.40
SII (ln) ^c	10.53 (10.00, 11.02)	10.52 (9.97, 11.04)	>0.90

^aMean (SD), Median (Q1, Q3); *n* (%).
^bPearson's Chi-squared test; Wilcoxon rank sum test.
^cln, Natural logarithm.
CTSI, CT Severity Index; BMI, Body Mass Index; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; RDW, Red Cell Distribution Width; NLR, Neutrophil-to-Lymphocyte Ratio; TyG, Triglyceride-Glucose Index; PLR, Platelet-to-Lymphocyte Ratio; NPR, Neutrophil-to-Platelet Ratio; NAR, Neutrophil-to-Albumin Ratio; SII, Systemic Inflammation Index.

Model interpretation and variables of importance

Among the models evaluated, Random Forest, XGBoost, and Logistic Regression demonstrated favorable predictive performance. To enhance our understanding of model decisions, we utilized Shapley

Additive Explanations (SHAP), which provides insights into the significance of individual features and their interactions within the model. SHAP values quantify the contribution of each clinical variable, with positive values reflecting an increased probability of developing APALI, and negative values suggesting a decreased likelihood. The global interpretability of these models was visualized

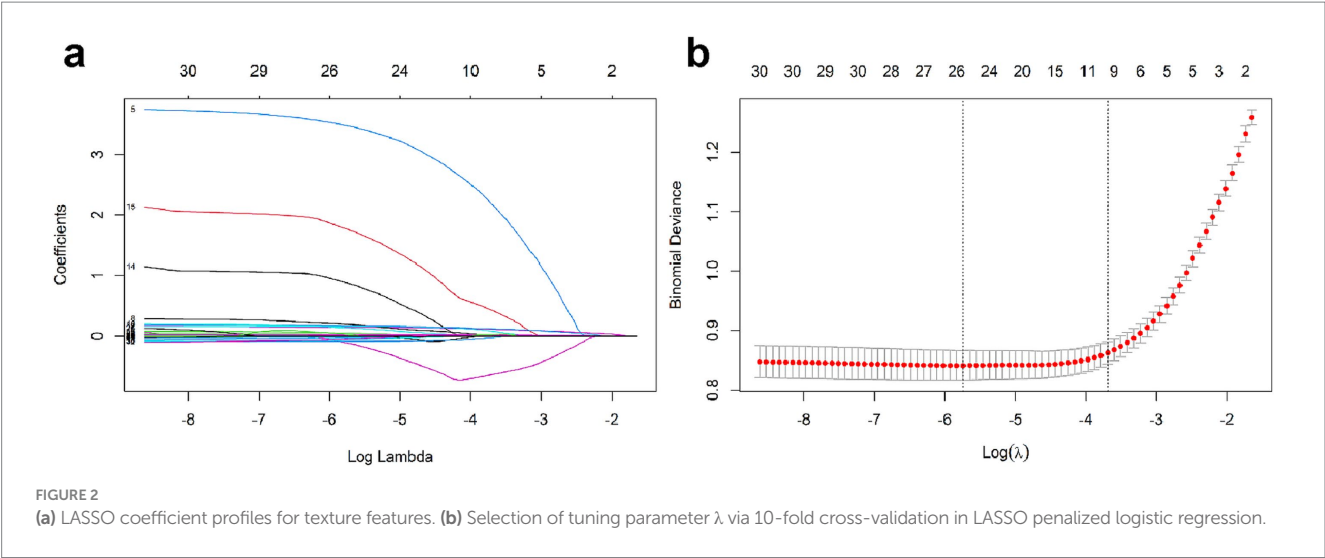


TABLE 3 Comparison of the performance of the six models in training set.

Model	Training set					
	Accuracy	AUC	F1	Recall	Sensitivity	Specificity
KNN	0.79 (0.77, 0.81)	0.84 (0.82, 0.86)	0.85 (0.84, 0.87)	0.91 (0.89, 0.93)	0.91 (0.89, 0.93)	0.54 (0.50, 0.58)
LR	0.81 (0.79, 0.83)	0.87 (0.86, 0.89)	0.87 (0.85, 0.88)	0.90 (0.88, 0.92)	0.90 (0.88, 0.92)	0.63 (0.59, 0.67)
NNET	0.79 (0.77, 0.81)	0.86 (0.84, 0.88)	0.85 (0.83, 0.86)	0.87 (0.85, 0.89)	0.87 (0.85, 0.89)	0.63 (0.59, 0.67)
RF	0.83 (0.81, 0.84)	0.90 (0.88, 0.91)	0.88 (0.86, 0.89)	0.94 (0.92, 0.95)	0.94 (0.92, 0.95)	0.60 (0.55, 0.64)
SVM	0.79 (0.77, 0.81)	0.87 (0.85, 0.88)	0.84 (0.82, 0.85)	0.80 (0.78, 0.83)	0.80 (0.78, 0.83)	0.75 (0.71, 0.79)
XGboost	0.82 (0.80, 0.84)	0.90 (0.88, 0.91)	0.87 (0.85, 0.89)	0.90 (0.88, 0.91)	0.90 (0.88, 0.91)	0.66 (0.62, 0.71)

LR, Logistic regression; RF, Random Forest; XGBoost, Extreme Gradient Boosting; SVC, Support vector Classifier; KNN, k-nearest neighbor; NNET, Neural Network.

TABLE 4 Comparison of the performance of the six models in testing set.

Model	Testing set					
	Accuracy	AUC	F1	Recall	Sensitivity	Specificity
KNN	0.78 (0.74, 0.81)	0.82 (0.78, 0.86)	0.85 (0.82, 0.87)	0.91 (0.88, 0.94)	0.91 (0.88, 0.94)	0.50 (0.42, 0.57)
LR	0.83 (0.80, 0.86)	0.90 (0.87, 0.93)	0.88 (0.86, 0.91)	0.94 (0.92, 0.97)	0.94 (0.92, 0.97)	0.61 (0.53, 0.68)
NNET	0.74 (0.70, 0.78)	0.81 (0.77, 0.85)	0.82 (0.78, 0.85)	0.84 (0.80, 0.88)	0.84 (0.80, 0.88)	0.53 (0.45, 0.61)
RF	0.84 (0.81, 0.87)	0.92 (0.89, 0.94)	0.89 (0.87, 0.91)	0.95 (0.93, 0.97)	0.95 (0.93, 0.97)	0.62 (0.55, 0.69)
SVM	0.82 (0.79, 0.85)	0.89 (0.86, 0.92)	0.87 (0.84, 0.89)	0.86 (0.83, 0.90)	0.86 (0.83, 0.90)	0.74 (0.67, 0.81)
XGboost	0.86 (0.83, 0.89)	0.91 (0.89, 0.94)	0.90 (0.88, 0.92)	0.95 (0.92, 0.97)	0.95 (0.92, 0.97)	0.68 (0.61, 0.75)

LR, Logistic regression; RF, Random Forest; XGBoost, Extreme Gradient Boosting; SVC, Support vector Classifier; KNN, k-nearest neighbor; NNET, Neural Network.

using the SHAP summary plot, which ranked the importance of variables across clinical outcomes. As shown in Figure 4, the top 10 most influential features were identified. The variables significantly impacting the model's predictions included CRP, neutrophil count, NAR, BMI, calcium ion levels, lactate, age, CT grade, Lym, blood amylase, and pleural effusion.

To enhance model simplicity while maintaining predictive performance, six variables—CRP, BMI, Neu, Calcium, Lactate, and NAR—were selected based on variable importance rankings and clinical relevance for model retraining. As illustrated in Figures 5a–d, the receiver operating characteristic (ROC) curves and calibration plots demonstrated that both the random forest (RF) and XGBoost

models exhibited strong predictive performance. Furthermore, decision curve analysis (DCA) indicated that both models provided greater net clinical benefit across a wide range of clinically relevant threshold probabilities compared to the “treat-all” and “treat-none” strategies (Supplementary Figure S2). Additionally, the global interpretability of the XGBoost and RF model, along with the six most influential features, is depicted in the SHAP summary plot (Figures 6a,b). To further illustrate the model's interpretability, we present two representative cases. The first cases describe patients who did not develop APALI and had low SHAP prediction scores (Figures 7a,b). The abdominal CT (Figure 7c) and chest CT (Figures 7d,e) demonstrated that this AP patient did not have signs of

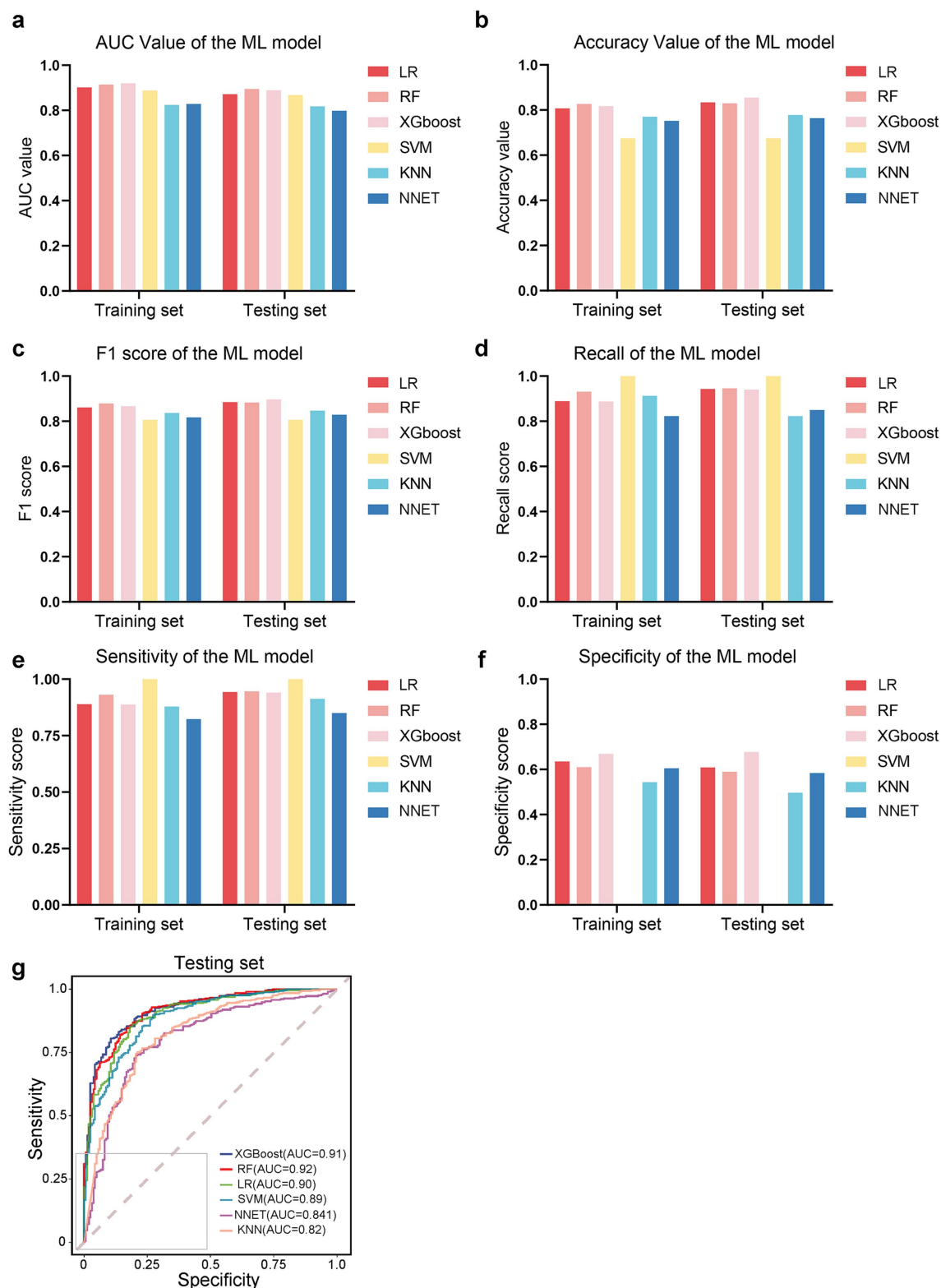


FIGURE 3 Performance comparison of machine learning models. (a–f) Bar plots or metrics distributions (AUC, accuracy, F1 score, recall, sensitivity, specificity) across models. (g) Receiver operating characteristic (ROC) curves for each model in the testing set.

lung injury. In contrast, the second case involved a patient diagnosed with APALI, showing a high SHAP prediction score (Figures 7f,g). And the abdominal CT (Figure 7h) and chest CT (Figures 7i,j) demonstrated that this AP patient had obvious lung consolidation with associated pleural effusion. All of these results proved the accuracy of the prognostic prediction system.

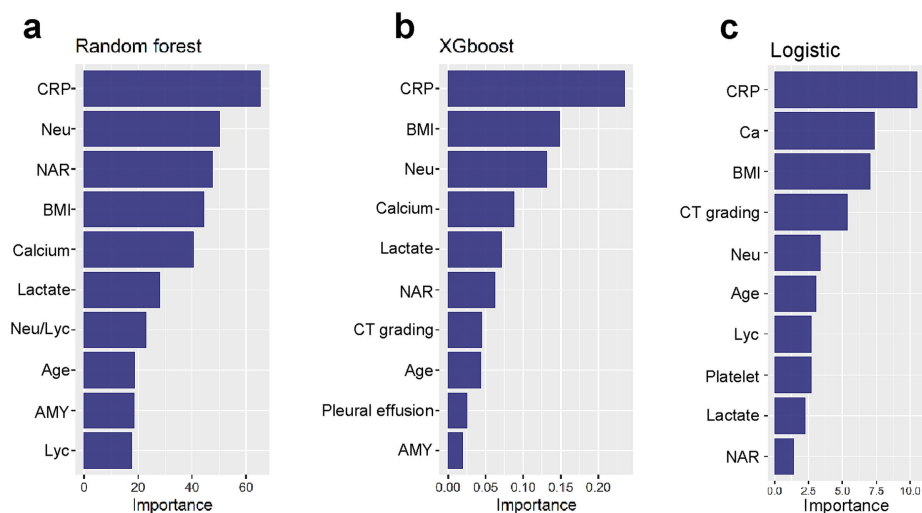


FIGURE 4

(a–c) Global interpretability of top-performing models (XGBoost, RF, LR) via SHAP. Summary plots rank the top 10 clinical features by mean absolute SHAP values, indicating their predictive contribution to APALI. Key features are CRP, neutrophil count, NAR, BMI, calcium ions, lactate, age, CT grade, lymphocytes (Lym), blood amylase, and pleural effusion.

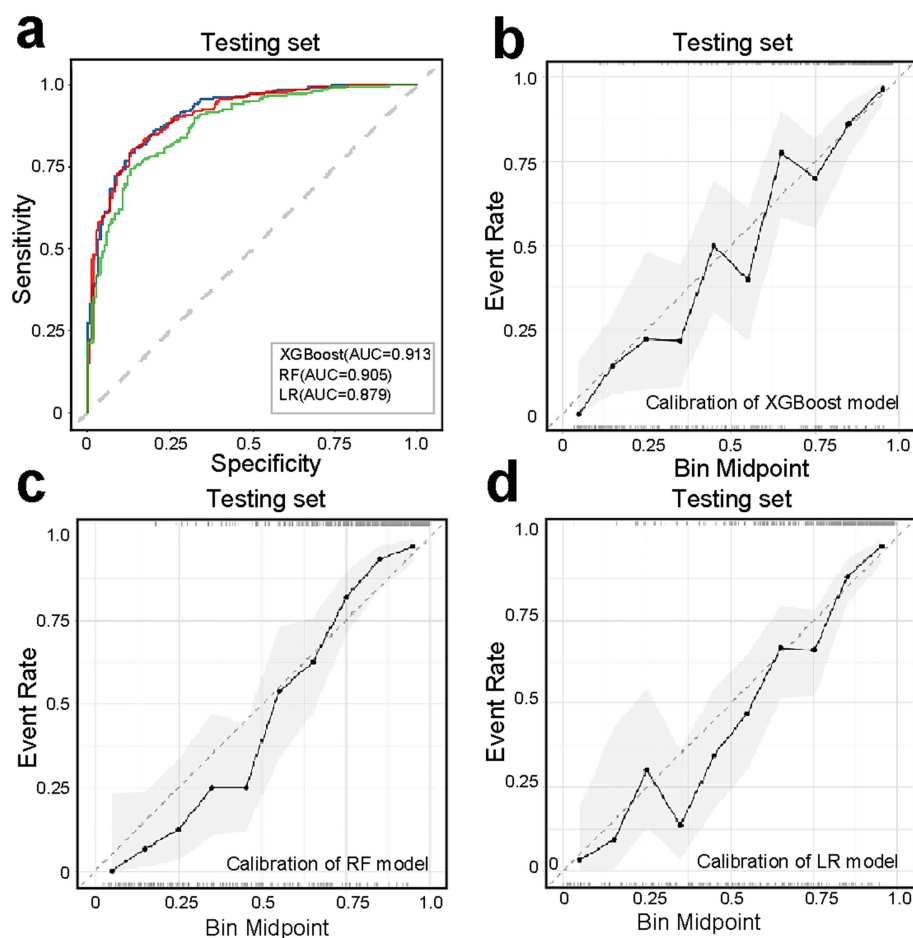


FIGURE 5

Performance evaluation of simplified models (XGBoost, RF, LR) using six key predictors (CRP, BMI, neutrophil count, calcium ions, lactate, NAR). (a) ROC curves demonstrating maintained predictive accuracy for XGBoost, RF, LR model despite feature reduction. (b–d) Calibration curves assessing agreement between predicted probabilities and observed outcomes, with closer-to-diagonal curves indicating better reliability.

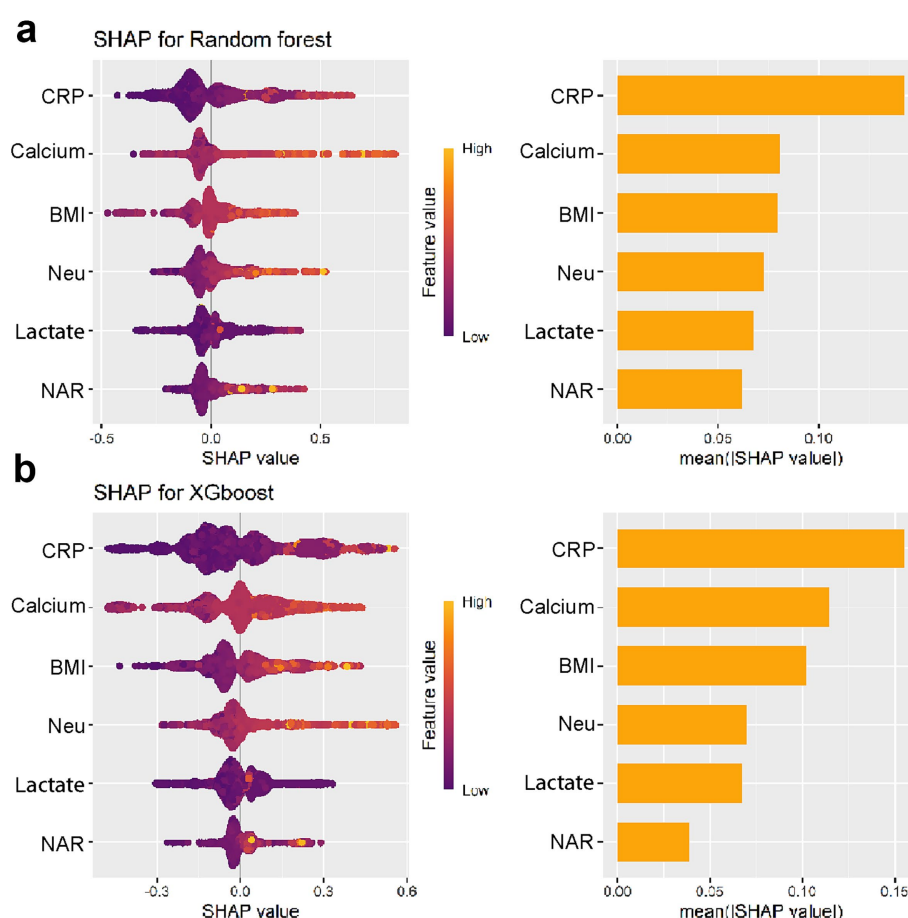


FIGURE 6

Global interpretability of the simplified XGBoost (a) and RF (b) models using SHAP summary plots.

Application of the model

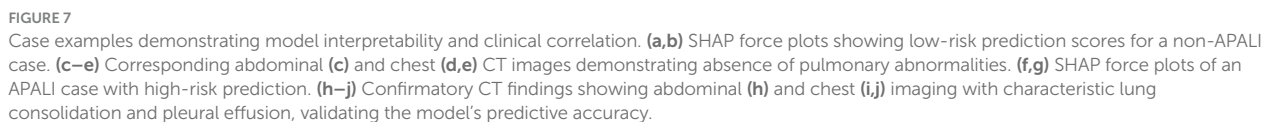
To enhance the convenience and practice utility of the developed model, we created a web-based tool to facilitate clinicians in clinical decision-making (accessible at: <https://yyiyis.shinyapps.io/APALI/>). Using this tool, clinicians can input key clinical data, such as BMI, Neu (neutrophil), ALB (albumin), CRP, and calcium ion levels, to predict likelihood of APALI in patients with AP (Figure 8).

Discussion

This study is the first and most comprehensive to apply machine learning methods to develop a model for predicting the occurrence of acute pancreatitis-associated lung injury (APALI). The RF and XGBoost models outperformed four other machine learning algorithms in the validation set, demonstrating superior AUC, accuracy, and F1 score, and exhibiting strong discriminatory power and calibration performance. We identified CRP, BMI, neutrophil count, calcium ions, lactate, and NAR as the most important independent predictors for APALI. Additionally, we developed a web-based tool to enhance the model's convenience and practical utility. This study may facilitate the early detection of lung injury in AP and support healthcare teams

in delivering targeted interventions, accelerating recovery, and improving quality of life.

Currently, there are limited studies on the prediction for ALI in AP. Jayanta Samanta identified that IL-6 and IL-8 could predict the development of ALI in AP and may serve as a composite biomarker (10). Lawrence Owusu proposed that γ -enolase could serve as an early indicator of lung tissue damage, even before significant histopathological injury is evident (22). Although the efficacy of these biomarkers has been partially validated, their complexity and limited availability may restrict their clinical applicability. Mengyu Jia developed a predictive model incorporating routine clinical data, such as diabetes, oxygen supplementation, neutrophil count, and D-dimer levels, and visualized the model using a nomogram (11). While this study based on relatively small cohorts (91 cases) which may undermine the stability of the model and restrict its practical applicability and suggest that machine learning models may outperform traditional manual scoring systems. Fei et al. compared artificial neural networks with logistic regression in predicting acute ALI among 217 patients with SAP. The ANN model achieved a sensitivity of 87.5%, specificity of 83.3%, and overall accuracy of 84.43%, significantly outperforming logistic regression (23). Ong et al. demonstrated that ML models (XGBoost) modestly outperformed conventional statistical methods in predicting poor post-ICU outcomes for mechanically ventilated patients, albeit with limited



AUC, along with higher accuracy, F1 score, and recall in the testing set. And we identified six key risk factors (CRP, BMI, neutrophil count, calcium ions, lactate, and NAR), which may predict the occurrence of APALI in the early phase of the disease. In our study, although both KNN and logistic regression models performed reasonably well, their overall predictive performance was inferior to

Predicting model for Acute Pancreatitis-Associated Lung Injury

Predicting probability

Input parameters

CRP(mg/l):

BMI:

Neu ($\times 10^9/l$):

Albumin (g/l):

Calcium

Lactate

Calculated N_ALB: 0.246

Submit

Status/Output

[1] "Calculation complete. The prediction results are for reference only."

Model	Prob of APALI
Random forest	0.49
XG Boost	0.55

FIGURE 8

Web-based clinical decision support tool for APALI risk prediction. Screenshot of the interactive interface showing input parameters (CRP, BMI, neutrophil count/albumin, calcium, lactate) and real-time risk calculation. Example output displaying predicted APALI probability with interpretative guidance. The tool is publicly available at: <https://yyiyis.shinyapps.io/APALI/>.

that of the random forest and XGBoost models. Compared to previous studies, our research is based on a larger-scale population dataset and demonstrates significant advantages across multiple key predictive indicators. The model's predictive performance has notably improved, particularly in terms of accuracy, sensitivity, and specificity, achieving high levels in these areas. These results indicate that our model not only holds strong clinical application value but also has considerable potential for widespread implementation, offering robust support for early disease prediction and intervention in broader clinical practices.

This study demonstrates that the neutrophil-to-albumin ratio (NAR) is a significant predictor of APALI, a relationship seldom explored in previous research. Neutrophils play a central role in APALI pathogenesis by releasing inflammatory mediators—such as myeloperoxidase (MPO) (24, 25), matrix metalloproteinases (MMP-8 and MMP-9), neutrophil elastase, and other proteolytic enzymes—that collectively drive systemic inflammation and tissue damage. Albumin reflects nutritional and hepatic status and declines during inflammation, infection, and organ dysfunction (26, 27). By integrating inflammatory burden and nutritional status, NAR has recently emerged as a robust prognostic biomarker across multiple clinical contexts. Elevated NAR predicts poor outcomes in sepsis, cardiovascular disease, and various malignancies, reflecting its strong association with systemic inflammation and organ failure (28). For example, in sepsis, high NAR correlates with increased mortality and organ dysfunction (29). Similar patterns appear in colorectal, gastric, and lung cancers, likely reflecting a proinflammatory tumor

microenvironment (30). Furthermore, elevated NAR has been linked with adverse long-term outcomes in cardiovascular disease, including higher risks of heart failure and cardiac events (31–34).

Large-scale population studies further support NAR's prognostic value, as well as that of its variant, the neutrophil-to-prealbumin ratio. Feng et al. reported that each 10-point increase in the Life's Crucial 9 score corresponded to a 28% reduction in COPD odds, with NAR mediating 4.8% of the effect (35). Han et al. identified a J-shaped relationship between NAR and all-cause or cardiovascular mortality in CKD, partially mediated by eGFR (36). Li et al. found that CKD patients with cardiovascular disease and elevated NPAR had significantly increased mortality risks (37). Ma et al. demonstrated a linear association between log-transformed NPAR and albuminuria, and Gao et al. identified NAR as a key predictor of poor outcomes after endovascular stroke therapy (38, 39). In our study, NAR ranked third in importance in the random forest model—after CRP and neutrophil count—and was strongly and positively correlated with APALI. Given its simplicity and predictive strength, NAR may serve as an efficient and early biomarker for identifying patients at high risk of APALI.

Lactate levels reflect tissue hypoxia, metabolic disturbances, and inflammatory responses, all of which are prevalent in AP (40). Lactate accumulation arises from microcirculatory dysfunction and tissue hypoxia, central to the pathogenesis of acute lung injury and ARDS (40). Studies have shown that persistently elevated lactate levels are associated with poor outcomes, complications such as ARDS, and increased

mortality in AP (41). Therefore, lactate monitoring is essential for prognostication and guiding therapeutic interventions, including fluid resuscitation and oxygen therapy (42, 43). The relationship between BMI and AP, as well as its associated lung injuries, such as ARDS, represents an important area of research (44, 45). Our study showed that BMI was ranked highly in both the Random Forest and XGBoost model. Obesity negatively impacts pancreatic health by disrupting fat metabolism, inducing insulin resistance, and promoting the release of pro-inflammatory cytokines. These factors contribute to systemic inflammation in patients with AP, increasing the risk of lung injury (44–46). An elevated BMI may further amplify inflammatory responses, as excess adipose tissue releases inflammatory mediators that worsen pancreatic damage and contribute to pulmonary complications (47). Therefore, a high BMI can serve as a critical early warning indicator for pulmonary injury and its progression in severe AP, playing a crucial role in the early identification and treatment of pulmonary complications.

We employed Shapley Additive Explanations (SHAP) to interpret the model and identify key clinical features influencing predictions. The SHAP values indicated that CRP, BMI, neutrophil count, calcium levels, lactate, and NAR were the most significant predictors of acute lung injury. Notably, SHAP visualizations provided clarity on the contribution of each variable and highlighted feature interactions, offering valuable insights into the underlying mechanisms of acute lung injury in AP. These findings support the utility of machine learning models in predicting the risk of acute lung injury in AP patients. By identifying key clinical features such as NAR, lactate, and BMI, our model could aid clinicians in early detection and personalized intervention. This predictive capability may facilitate timely therapies to reduce systemic inflammation and prevent progression to acute respiratory distress syndrome (ARDS). Despite strong predictive performance, real-world use of the APALI model faces three main challenges. First, technical integration: hospitals with fragmented IT systems may need middleware to support real-time use, even though the model relies on structured EHR data. Second, workflow compatibility: clinician trust requires explainability (e.g., SHAP plots) and smooth integration into systems like order entry. Third, scalability: retraining may be needed to adapt to local case-mix differences.

This study has several limitations. As a retrospective case-control study, it may be affected by selection bias, and some patients may have received treatment before laboratory data were collected, potentially influencing the results. Although we employed rigorous methodology, the external validation yielded exceptionally high performance. While we applied strict data separation and anti-leakage procedures, this result may reflect specific characteristics of the external dataset, such as small sample size, case homogeneity, and a higher proportion of critically ill patients from a tertiary referral center. Therefore, this finding should be interpreted with caution. To enhance the accuracy, generalizability, and clinical applicability of the model, future studies should focus on large-sample, multi-center prospective research, incorporate real-time clinical data, and minimize treatment-related biases.

Conclusion

In this study, we developed a clinical prediction model for the early identification of lung injury in patients with AP by utilizing machine learning algorithms to analyze clinical data. The model

exhibited robust predictive performance, validated through external testing and individual assessments, underscoring its clinical utility for timely interventions in AP-associated lung injury. Additionally, we created a web-based calculator to facilitate the model's application in clinical practice, enabling healthcare professionals to make faster decisions and potentially improve patient outcomes. This tool represents a valuable resource for clinicians managing AP, ensuring timely and appropriate care for patients at high risk of lung injury. Future work will focus on enhancing the model through multi-center validation, incorporating diverse clinical variables, and optimizing the tool for broader clinical application.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.

Ethics statement

This retrospective multicenter study examined acute pancreatitis cases from two academic medical centers. The derivation cohort (July 2012–June 2022) consisted of patients from Bengbu Medical University's First Affiliated Hospital, used for model training and internal validation. The external validation cohort (January 2018–April 2023) comprised patients from Zhejiang University's Second Affiliated Hospital. The study was approved by both institutions' ethics committees, with all patient data anonymized and written informed consent obtained in accordance with the Declaration of Helsinki.

Author contributions

ZD: Conceptualization, Writing – original draft. QY: Investigation, Writing – review & editing, Data curation, Formal analysis. YY: Formal analysis, Writing – review & editing, Software. HM: Validation, Writing – review & editing, Supervision. HZ: Supervision, Validation, Methodology, Writing – review & editing. JY: Software, Validation, Writing – review & editing, Supervision. ZW: Writing – review & editing, Supervision. CZ: Supervision, Writing – review & editing. SW: Validation, Writing – review & editing, Project administration, Methodology, Software, Supervision, Investigation. QT: Software, Supervision, Methodology, Data curation, Resources, Conceptualization, Validation, Investigation, Formal analysis, Writing – original draft, Funding acquisition, Writing – review & editing, Visualization, Project administration.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded by the Key Project of Natural Science Research Project of Anhui Provincial Education Department in 2022 (No. KJ2022AH051437), the Natural science Project of Bengbu Medical University in 2021 (No.

2021byzd175), and National Natural Science Foundation of China (NSFC, No. 82303843).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

References

- Avery B, Nathens J, Randall C, Richard J. Management of the critically ill patient with severe acute pancreatitis – ScienceDirect. *Reanimation*. (2005) 14:148–50. doi: 10.1097/01.ccm.0000148222.09869.92
- Li XY, He C, Zhu Y, Lu NH. Role of gut microbiota on intestinal barrier function in acute pancreatitis. *World J Gastroenterol*. (2020) 26:2187–93. doi: 10.3748/wjg.v26.i18.2187
- Petrov MS, Yadav D. Global epidemiology and holistic prevention of pancreatitis. *Nat Rev Gastroenterol Hepatol*. (2019) 16:175–84. doi: 10.1038/s41575-018-0087-5
- Schepers NJ, Bakker OJ, Besselink MG, Ahmed Ali U, Bollen TL, Gooszen HG, et al. Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut*. (2019) 68:1044–51. doi: 10.1136/gutjnl-2017-314657
- Sousa MARD, Teixeira GDS, Marquesa RB, Sousa LMRD, Ramos RM, Bento RRDF, et al. Therapeutic actions of methyl eugenol in acute lung inflammation induced in rats. *S Afr J Bot*. (2024) 169:341–9. doi: 10.1016/j.sajb.2024.04.023
- Wu X, Yao J, Hu Q, Kang H, Miao Y, Zhu L, et al. Emodin ameliorates acute pancreatitis-associated lung injury through inhibiting the alveolar macrophages Pyroptosis. *Front Pharmacol*. (2022) 13:873053. doi: 10.3389/fphar.2022.873053
- Leclair V, Zheng B, Diederichsen LP. Extracorporeal membrane oxygenation for acute lung injury in idiopathic inflammatory myopathies-a potential lifesaving intervention: reply. *Rheumatology (Oxford)*. (2024) 64:2333–4. doi: 10.1093/rheumatology/keae419
- Zheng B, Eline E, Xu L, Huang K, Hermans G, Perch M, et al. Extracorporeal membrane oxygenation for acute lung injury in idiopathic inflammatory myopathies-a potential lifesaving intervention. *Rheumatology (Oxford)*. (2024) 64:2204–8. doi: 10.1093/rheumatology/keae311
- Kutumova EO, Akberdin IR, Egorova VS, Kolesova EP, Parodi A, Pokrovsky VS, et al. Physiologically based pharmacokinetic model for predicting the biodistribution of albumin nanoparticles after induction and recovery from acute lung injury. *Heliyon*. (2024) 10:e30962. doi: 10.1016/j.heliyon.2024.e30962
- Samanta J, Singh S, Arora S, Muktesh G, Aggarwal A, Dhaka N, et al. Cytokine profile in prediction of acute lung injury in patients with acute pancreatitis. *Pancreatol*. (2018) 18:878–84. doi: 10.1016/j.pan.2018.10.006
- Jia M, Xu X, Zhou S, Liu H, Zhao Y, Xu Y, et al. Prediction of acute lung injury in severe acute pancreatitis by routine clinical data. *Eur J Gastroenterol Hepatol*. (2023) 35:36–44. doi: 10.1097/MEG.0000000000002458
- Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
- Deo RC. Machine learning in medicine. *Circulation*. (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
- Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther*. (2020) 51:675–87. doi: 10.1016/j.beth.2020.05.002
- Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J Transl Med*. (2022) 20:215. doi: 10.1186/s12967-022-03364-0
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. (2019) 110:12–22. doi: 10.1016/j.jclinepi.2019.02.004
- Ong WJD, How CH, Chong WHK, Khan FA, Ngiam KY, Kansal A. Outcome prediction for adult mechanically ventilated patients using machine learning models and

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1638097/full#supplementary-material>

- comparison with conventional statistical methods: a single-Centre retrospective study. *Intell -Based Med*. (2024) 10:100165. doi: 10.1016/j.ibmed.2024.100165
- Bos LD, Ware LB. Acute respiratory distress syndrome: causes, pathophysiology, and phenotypes. *Lancet*. (2022) 400:1145–56. doi: 10.1016/S0140-6736(22)01485-4
- Gorman EA, O'Kane CM, McAuley DF. Acute respiratory distress syndrome in adults: diagnosis, outcomes, long-term sequelae, and management. *Lancet*. (2022) 400:1157–70. doi: 10.1016/S0140-6736(22)01439-8
- Matthay MA, Zemans RL, Zimmerman GA, Arabi YM, Beitler JR, Mercat A, et al. Acute respiratory distress syndrome. *Nat Rev Dis Prim*. (2019) 5:18. doi: 10.1038/s41572-019-0069-0
- Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, et al. Acute respiratory distress syndrome: the Berlin definition. *JAMA*. (2012) 307:2526–33. doi: 10.1001/jama.2012.5669
- Owusu L, Xu C, Chen H, Liu G, Zhang G, Zhang J, et al. Gamma-enolase predicts lung damage in severe acute pancreatitis-induced acute lung injury. *J Mol Histol*. (2018) 49:347–56. doi: 10.1007/s10735-018-9774-3
- Fei Y, Gao K, Li WQ. Artificial neural network algorithm model as powerful tool to predict acute lung injury following to severe acute pancreatitis. *Pancreatol*. (2018) 18:892–9. doi: 10.1016/j.pan.2018.09.007
- Davies MJ. Myeloperoxidase: mechanisms, reactions and inhibition as a therapeutic strategy in inflammatory diseases. *Pharmacol Ther*. (2021) 218:107685. doi: 10.1016/j.pharmthera.2020.107685
- Valadez-Cosmes P, Raftopoulou S, Mihalic ZN, Marsche G, Kargl J. Myeloperoxidase: growing importance in cancer pathogenesis and potential drug target. *Pharmacol Ther*. (2022) 236:108052. doi: 10.1016/j.pharmthera.2021.108052
- Eckart A, Struja T, Kutz A, Baumgartner A, Baumgartner T, Zurluh S, et al. Relationship of nutritional status, inflammation, and serum albumin levels during acute illness: a prospective study. *Am J Med*. (2020) 133:713–722.e7. doi: 10.1016/j.amjmed.2019.10.031
- Sheinenzon A, Shehadeh M, Michelis R, Shaoul E, Ronen O. Serum albumin levels and inflammation. *Int J Biol Macromol*. (2021) 184:857–62. doi: 10.1016/j.ijbiomac.2021.06.140
- Shen H, Dai S, Wang M, Gu S, Xu W, Xu G, et al. Preprocedural neutrophil to albumin ratio predicts in-stent restenosis following carotid angioplasty and stenting. *J Stroke Cerebrovasc Dis*. (2019) 28:2442–7. doi: 10.1016/j.jstrokecerebrovasdis.2019.06.027
- Sutherland A, Thomas M, Brandon RA, Brandon RB, Lipman J, Tang B, et al. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis. *Crit Care*. (2011) 15:1–11. doi: 10.1186/cc10274
- Xie H, Wei L, Liu M, Liang Y, Yuan G, Gao S, et al. Neutrophil-albumin ratio as a biomarker for postoperative complications and long-term prognosis in patients with colorectal cancer undergoing surgical treatment. *Front Nutr*. (2022) 9:976216. doi: 10.3389/fnut.2022.976216
- Sahin DY, Elbasan Z, Gur M, Yldz A, Akpınar O, İcen YK, et al. Neutrophil to lymphocyte ratio is associated with the severity of coronary artery disease in patients with ST-segment elevation myocardial infarction. *Angiology*. (2013) 64:423–9. doi: 10.1177/0003319712453305
- Yu YY, Lin YT, Chuang H-C, Huang C-Y, Fang T-L, Tsai F-M, et al. Prognostic utility of neutrophil-to-albumin ratio in surgically treated oral squamous cell carcinoma. *Head Neck*. (2023) 45:2839–50. doi: 10.1002/hed.27511
- Stefanescu S, Turcustiolica A, Shelby ES, Matei M, Subtirelu MS, Meca AD, et al. Prediction of treatment outcome with inflammatory biomarkers after 2 months of

therapy in pulmonary tuberculosis patients: preliminary results. *Pathogens*. (2021) 10:789. doi: 10.3390/pathogens10070789

34. Xiao L, Li F, Sheng Y, Hou X, Liao X, Zhou P, et al. Predictive value analysis of albumin-related inflammatory markers for short-term outcomes in patients with in-hospital cardiac arrest. *Expert Rev Clin Immunol*. (2024) 21:249–57. doi: 10.1080/1744666X.2024.2399700

35. Feng J, Gong H. Neutrophil-to-albumin ratio mediates the association between life's crucial 9 and chronic obstructive pulmonary disease. *Front Med (Lausanne)*. (2025) 12:1610945. doi: 10.3389/fmed.2025.1610945

36. Han Y, Jiao X, Yang M, Liu C, Zhao D. Association of neutrophil-to-albumin ratio with all-cause and cardiovascular mortality in community-dwelling individuals with chronic kidney disease: evidence from the NHANES 1999–2018. *BMC Cardiovasc Disord*. (2025) 25:457. doi: 10.1186/s12872-025-04935-x

37. Li J, Yang M, Zhang X, Huang R, Zhang Y, Fan K. Neutrophil to albumin ratio predicts cardiovascular and all cause mortality in CVD patients with abnormal glucose metabolism. *Sci Rep*. (2025) 15:21976. doi: 10.1038/s41598-025-08130-y

38. Gao W, Sun J, Yu L, She J, Zhao Y, Cai L, et al. Distinct trajectory patterns of neutrophil-to-albumin ratio predict clinical outcomes after endovascular therapy in large vessel occlusion stroke. *Front Aging Neurosci*. (2025) 17:1570662. doi: 10.3389/fnagi.2025.1570662

39. Ma X, Qian Y, Qian C, Lin H, Sun Y. Association between inflammation indicators and albuminuria in US adults: a cross-sectional study. *Sci Rep*. (2025) 15:21496. doi: 10.1038/s41598-025-06540-6

40. Vincent JL, Quintairos ESA, Couto L Jr, Taccone FS. The value of blood lactate kinetics in critically ill patients: a systematic review. *Crit Care*. (2016) 20:257. doi: 10.1186/s13054-016-1403-5

41. Lyu S, Liu S, Guo X, Zhang Y, Liu Z, Shi S, et al. hP-MSCs attenuate severe acute pancreatitis in mice via inhibiting NLRP3 inflammasome-mediated acinar cell pyroptosis. *Apoptosis*. (2024) 29:920–33. doi: 10.1007/s10495-024-01946-5

42. Cha M, Park J. Utilizing point-of-care lactate testing for rapid prediction of clinical outcomes in patients with acute gastrointestinal bleeding in the emergency department. *Heliyon*. (2024) 10:e38184. doi: 10.1016/j.heliyon.2024.e38184

43. Kim S, Lee S, Ahn S, Park J, Moon S, Cho H, et al. The prognostic utility of lactate/albumin*age score in septic patient with normal lactate level. *Heliyon*. (2024) 10:e37056. doi: 10.1016/j.heliyon.2024.e37056

44. Yousefshahi F, Samadi E, Paknejad O, Movafegh A, Barkhordari K, Bastan Hagh E, et al. Prevalence and risk factors of hypoxemia after coronary artery bypass grafting: the time to change our conceptions. *J Tehran Heart Cent*. (2019) 14:74–80. doi: 10.18502/jthc.v14i2.1375

45. Ni YN, Luo J, Yu H, Wang YW, Hu YH, Liu D, et al. Can body mass index predict clinical outcomes for patients with acute lung injury/acute respiratory distress syndrome? A meta-analysis. *Crit Care*. (2017) 21:36. doi: 10.1186/s13054-017-1615-3

46. Sandby K, Krarup T, Chabanova E, Geiker NRW, Magkos F. Liver fat accumulation is associated with increased insulin secretion independent of total, visceral, and pancreatic fat. *J Clin Endocrinol Metab*. (2024) 110:dgae572. doi: 10.1210/clinem/dgae572

47. Karki R, Kanneganti TD. The 'cytokine storm': molecular mechanisms and therapeutic prospects. *Trends Immunol*. (2021) 42:681–705. doi: 10.1016/j.it.2021.06.001