![frontiers | Frontiers in Cellular and Infection Microbiology]

Usefulness of Random Forest Algorithm in Predicting Severe Acute Pancreatitis

Wandong Hong[1*], Yajing Lu[1], Xiaoying Zhou[2], Shengchun Jin[2], Jingyi Pan[2], Qingyi Lin[2], Shaopeng Yang[2], Zarrin Basharat[3], Maddalena Zippi[4] and Hemant Goyal[5]

[1] Department of Gastroenterology and Hepatology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [2] School of the First Clinical Medical Sciences, Wenzhou Medical University, Wenzhou, China, [3] Jamil-ur-Rahman Center for Genome Research, Dr. Panjwani Centre for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan, [4] Unit of Gastroenterology and Digestive Endoscopy, Sandro Pertini Hospital, Rome, Italy, [5] Department of Medicine, The Wright Center for Graduate Medical Education, Scranton, PA, United States

Background and Aims: This study aimed to develop an interpretable random forest model for predicting severe acute pancreatitis (SAP).

Methods: Clinical and laboratory data of 648 patients with acute pancreatitis were retrospectively reviewed and randomly assigned to the training set and test set in a 3:1 ratio. Univariate analysis was used to select candidate predictors for the SAP. Random forest (RF) and logistic regression (LR) models were developed on the training sample. The prediction models were then applied to the test sample. The performance of the risk models was measured by calculating the area under the receiver operating characteristic (ROC) curves (AUC) and area under precision recall curve. We provide visualized interpretation by using local interpretable model-agnostic explanations (LIME).

Results: The LR model was developed to predict SAP as the following function: -1.10-0.13×albumin (g/L) + 0.016 × serum creatinine (μmol/L) + 0.14 × glucose (mmol/L) + 1.63 × pleural effusion (0/1)(No/Yes). The coefficients of this formula were utilized to build a nomogram. The RF model consists of 16 variables identified by univariate analysis. It was developed and validated by a tenfold cross-validation on the training sample. Variables importance analysis suggested that blood urea nitrogen, serum creatinine, albumin, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, calcium, and glucose were the most important seven predictors of SAP. The AUCs of RF model in tenfold cross-validation of the training set and the test set was 0.89 and 0.96, respectively. Both the area under precision recall curve and the diagnostic accuracy of the RF model were higher than

that of both the LR model and the BISAP score. LIME plots were used to explain individualized prediction of the RF model.

Conclusions: An interpretable RF model exhibited the highest discriminatory performance in predicting SAP. Interpretation with LIME plots could be useful for individualized prediction in a clinical setting. A nomogram consisting of albumin, serum creatinine, glucose, and pleural effusion was useful for prediction of SAP.

Keywords: random forest, nomogram, acute pancreatitis, predictor, artificial intelligence, LIME plot

HIGHLIGHTS

1. An interpretable random forest model exhibited the highest discriminatory performance in SAP prediction.
2. Interpretation with LIME plots could be useful for individualized prediction in a clinical setting.
3. A nomogram comprising albumin, serum creatinine, glucose, and pleural effusion is a useful predictor of SAP.

INTRODUCTION

Acute pancreatitis (AP) is one of the most common gastrointestinal problems for hospital admission globally (Hong et al., 2020). While most patients with AP will recover within a week of a mild course and are often self-limiting, 20% of patients progress to severe disease with a historical mortality risk as high as 30% (Trikudanathan et al., 2019). In the absence of specific treatment in the early phase, initial management of severe acute pancreatitis (SAP) focuses on supportive care such as fluid resuscitation, pain control, and nutritional support, aimed to minimize the impact of systemic inflammatory response syndrome (Lee and Papachristou, 2019). Patients with SAP often need to be transferred to the intensive care unit once organ failure occurs. Therefore, it is important to recognize predictors for severe disease in the early phase of AP, to select those patients who would benefit most from enhanced surveillance or early interventions. Early case identification and classification of disease severity could improve the clinical outcomes (Hong et al., 2021).

Many clinical scoring systems have been developed for the prediction of disease severity, such as the Ranson, chronic health evaluation (APACHE-II) score, Pancreatitis Outcome Prediction (POP) Score (Harrison et al., 2007), and Bedside Index Of Severity In Acute Pancreatitis (BISAP) (Wu et al., 2008). However, the existing scoring systems have moderate accuracy in predicting the severity of AP (Mounzer et al., 2012). Recently, Langmead et al. reported that a 5-cytokine panel consisting of angiopoietin-2, hepatocyte growth factor, interleukin-8, resistin, and soluble tumor necrosis factor receptor 1A accurately predicts persistent organ failure early in the disease process and significantly outperforms the prognostic accuracy of existing laboratory tests and clinical scores (Langmead et al., 2021). However, the test of cytokine is not routinely available,

resulting in limited use in clinical practice. Several laboratory indexes such as total cholesterol (Hong et al., 2020), low-density lipoprotein cholesterol (Hong et al., 2018), albumin (Hong et al., 2017a), and blood urea nitrogen (BUN) (Lin et al., 2017) have been proposed as single predictors of severity of AP. Recently, Takeda et al. reported that fluid sequestration is a useful parameter in the early identification of SAP (Takeda et al., 2019). Yan et al. described that pleural effusion volume could be a reliable radiologic biomarker in the prediction of severity and clinical outcomes of AP (Yan et al., 2021). Gibor et al. reported that circulating cell-free DNA in patients with acute biliary pancreatitis is associated with disease markers and prolonged hospitalization time (Gibor et al., 2020). However, these single prediction markers are easy to use in practice but lack high accuracy.

Recently, artificial intelligence methods are also being used in the clinical setting for disease prediction or aiding in making decisions. Among several methods, a random forest (RF) is a group of many decision trees, each of which is characterized by a tree-like structure (Genuer and Poggi, 2020). It will randomly choose features and make observations, build a forest of decision trees, and then average out the results (James et al., 2013). RF allows considering qualitative and quantitative explanatory variables together, without preprocessing (Genuer and Poggi, 2020). Random forests are adapted to both supervised classification problems and regression problems (Genuer and Poggi, 2020). In addition, RF can handle datasets with many predictor variables, while also performing very well. Additionally, it can obtain variable importance ranking when used for prediction modeling (Speiser et al., 2019). RF, as a traditional machine learning method, has been shown to outperform other techniques for sets of features in a variety of different settings. RF has recently demonstrated high performance in risk classification and disease prediction (Yu et al., 2021). Lo et al. developed RF model for forecasting allergenic pollen in North America (Lo et al., 2021). Lin et al. reported that using the RF model could predict environmental risk factors in relation to health outcomes among school children from Romania (Lin et al., 2021b). Roguet et al. reported that RF classification with 16S rRNA gene amplicons offers an accurate solution for identifying host microbial signatures (useful in detecting human and animal fecal contamination in environmental samples) (Roguet et al., 2018). Yang et al. provided an RF prediction model for 3-year risk assessment of cardiovascular disease (Yang et al., 2020).

However, to the best of our knowledge, the use of RF model in predicting disease severity in patients with AP has not been performed yet. The aim of this study was to develop an RF model and compare it with a traditional logistic regression (LR) model for prediction of SAP.

PATIENTS AND METHODS

Inclusion and Exclusion Criteria

We conducted a *post-hoc* analysis of a previously reported retrospective cohort study in the First Affiliated Hospital of Wenzhou Medical University in mainland China (Hong et al., 2018). Patients with AP admitted to the First Affiliated Hospital of Wenzhou Medical University within 72 h of symptom onset (from April 1, 2012 to December 31, 2015) were consecutively enrolled in the study (Hong et al., 2020). The diagnosis of AP was based on the presence of two of the following three features: (1) abdominal pain consistent with AP; (2) serum amylase and/ or lipase more than three times that of the normal; (3) abdominal imaging findings (Hong et al., 2020). As previously described (Hong et al., 2020), exclusion criteria included endoscopic or trauma related pancreatitis, chronic pancreatitis, pancreatic tumor, history of surgery operation/taking hypolipidemia drugs, malnutrition, and chronic liver or renal disease.

Data Collection

The clinical and laboratory data on admission were obtained with data collection forms from electronic medical records. These data included blood chemica+l analysis, liver, and renal function testing, glucose, lipids, coagulation testing, serum calcium, C-reaction protein, and pleural effusion (Hong et al., 2020).

Definition of Severity and Study Endpoint

SAP is defined as a persistent organ failure (>48 h) in patients. Organ failure for this study was defined according to a Marshall score ≥2, meaning that at least one organ system (respiratory, cardiovascular, renal) must be affected (Hong et al., 2018). The primary study endpoint was the occurrence of SAP during hospitalization.

Sample Size and Missing Values

The calculation of the sample size of this study was according to our previous study (Hong et al., 2020). There were missing values in serum calcium and C-reactive protein data. To handle this issue, missing values were imputed using Multiple Imputations by Chained Equations (MICE) when performing LR and RF analysis (Royston, 2005). The MICE has emerged as one of the principal statistical approaches for dealing with missing data. The missing values were replaced by the estimated plausible values to create a "complete" dataset (Royston, 2005).

Statistical Analysis

Categorical values were described by count and proportions and compared by the χ2 test or Fisher's exact test. According to the results of the Shapiro-Wilk test, continuous values were expressed by mean ± SD or median and Inter Quartile Range (IQR) and compared using Student's t-test or the Wilcoxon's non-parametric test. The discriminative power of the predictor was assessed by calculating the area under the receiver operating characteristic (ROC) curves (AUC) (Hong et al., 2019). A variable with an AUC above 0.7 was considered useful (Hong et al., 2017b).

The data samples (of 648 patients) were randomly split into training and test sets according to a division of 3:1 (487 *vs.* 161 patients). The RF model was developed on the training set and independently validated on the test set by using "randomForest" (Liaw and Wiener, 2002) and "caret" package (Kuhn, 2008). When we built and tuned the RF model on a training set, we used tenfold cross-validation as the resampling method to avoid overfitting of the model in new data (Kuhn, 2008). The training set was divided into 10 equal-sized sub-samples in which 9 sub-samples were for the training and the remaining ones for testing over all possible permutations. Analysis was repeated 10 times (folds) (Hong et al., 2019). The AUC was calculated for each of the 10 analyses, using only the respective test data (Hong et al., 2019). Then this iteration process was repeated 10 times. At last, the mean AUC with 95%CI, as well as area under precision recall curve was calculated and compared (Saito and Rehmsmeier, 2015; Hong et al., 2019).

After training the RF model, a general approach of interpretability is to identify important variables (features) in the model (Staniak and Biecek, 2018). The RF algorithm estimates the importance of a variable by looking at how much prediction error increases when Out-Of-Bag (OOB) data for that variable are permuted, while all others are left unchanged (Liaw and Wiener, 2002; Genuer and Poggi, 2020). The variable importance is a global explanation of relative importance of each feature in the RF model (Kuhn, 2008). Variables having high importance are drivers of the outcome and their values have a significant impact on the result values.

To overcome the black box problem of the RF model output and improve its interpretability, the local interpretable model-agnostic explanations (LIME) plot was used to explain the individualized prediction (Deshmukh and Merchant, 2020). LIME is a novel explanation technique that explains the predictions of any classifier in an interpretable and faithful manner, by learning an interpretable model locally around the prediction (Ribeiro et al., 2016). The training of the local interpretable model involves giving weight to the disturbance input, followed by the observation of the general (black box) model output, which gives a basis for interpretation of the prediction results (Pan et al., 2020). This feature is deemed important if perturbations at the local level produce a change in the general model while the value of the target feature is determined by the level of change it determines (Bramhall et al., 2020). Local explanation detects variables' contribution at the local level. In other words, LIME could provide easily understood explanations of clinical factors in the RF models, which contribute to each prediction for the individual patient (Petch et al., 2022). LIME was performed by using the "lime" package ("Lime: Local Interpretable Model-Agnostic Explanations, 2021", http://cran.itam.mx/web/packages/lime/index.html), in which two types of inputs (tabular and text) are supported (Pedersen and Benesty, 2021).

A forward-conditional stepwise LR analysis was also applied on the training set. The conditional probabilities for stepwise entry and removal of a factor were 0.05 and 0.06, respectively (Hong et al., 2020). Based on the results of LR, a nomogram was developed to predict SAP. Model calibration, reflecting the link between predicted and observed risk, was evaluated by the Hosmer-Lomeshow goodness of fit test, as well as plotting the predicted *vs.* observed deciles of predicted risk (Hong et al., 2017b). Odds ratios (OR) were calculated, with 95%CI.

We selected the best cut-off point, where the number of true positives was the highest possible (sensitivity >90%). This was done by selecting a threshold value at a point where the longest increase in the specificity of the slope declines for all models and scores. The sensitivity, specificity, and accuracy were calculated and compared (Saito and Rehmsmeier, 2015).

A two-tailed P-value of less than 0.05 was considered statistically significant. All statistical analysis were performed in the R and STATA software. A data flow diagram of our study is shown in **Supplementary Figure S1**.

RESULTS

Baseline Characteristics

Of all the patients, the hospital mortality was 1.54%. There were 247 (58.8%) men and their median age was 53 (42.0–64.5) years. The most common etiology of AP was biliary (42.4%). The median time interval between onset and admission was 2 (IQR 1-2) days. Of these patients 10% developed SAP during hospitalization. The median length of the hospital stay was 10 (IQR 7-14) days. The baseline characteristics of the patients in the training and test sets are shown in **Table 1**.

Univariate Analysis on the Training Sample

As shown in **Table 2**, 16 variables, namely, systemic inflammatory response syndrome (SIRS), hematocrit, platelets, prothrombin time, albumin, aspartate aminotransferase (AST), glucose, serum creatinine, blood urea nitrogen (BUN), total cholesterol, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglyceride, serum calcium, C-reactive protein (CRP), and pleural effusion were significantly associated with the development of SAP, as inferred by univariate analysis.

Models Development, Calibration, Tenfold Cross-Validation on the Training Sample

Variables significantly linked to the development of SAP in the univariate analysis were assessed by stepwise LR analysis. LR identified the following four independent variables as predictive of SAP: albumin (OR 0.88, 95%CI 0.81-0.95, P=0.002), serum creatinine (OR 1.02, 95%CI 1.01-1.03, P=0.002), glucose (OR 1.15, 95%CI 1.07-1.24, P<0.001), and pleural effusion (OR 5.11, 95%CI

TABLE 1 | Comparison of clinical and laboratory findings among patients, with and without SAP (training sample set).

Variable	Training set (n = 487)	Test set (n = 161)	P-value
Age, years (IQR)	47 (37,61)	49 (36,64)	0.501
Male sex, N (%)	301 (61.81)	103 (63.98)	0.623
Duration of symptoms (days, IQR)	1.83 ± 0.80	1.78 ± 0.78	0.515
BMI, kg/m^2 (IQR)	23.5 (21.1-26.3)	23.9 (21.5-21.5)	0.573
SIRS, N (%)	191 (39.22)	65 (40.37)	0.795
Biliary etiology, N (%)	207 (42.51)	68 (42.24)	0.584
Laboratory findings			
Hematocrit (l/l)	0.42 (0.38-0.46)	0.42 (0.38-0.46)	0.693
Platelets (109/L)	199 (161-233)	195 (157-233)	0.472
Prothrombin time, s (IQR)	13.8 (13.1-14.6)	13.7 (13.0-14.5)	0.278
Albumin, g/L (IQR)	36.3 (32.6-39.9)	36.4 (34.0-39.8)	0.191
Total bilirubin, mmol/L (IQR)	20 (14-31)	20 (13-32)	0.916
ALT, U/L (IQR)	43 (19-119)	31 (19-82)	0.055
AST, U/L (IQR)	39 (22-88)	28 (19-71)	0.012
Glucose, mmol/L (IQR)	7.9 (6.3-10.5)	8.4 (6.7-11.3)	0.128
Serum creatinine, μmol/L (IQR)	64 (54-77)	64 (55-76)	0.882
BUN, mmol/L (IQR)	4.8 (3.7-6.1)	4.9 (4.0-6.2)	0.346
Total cholesterol, mmol/L (IQR)	4.79 (3.8-6.2)	4.8 (3.8-6.1)	0.970
HDL, mmol/L (IQR)	1.0 (0.7-1.3)	1.0 (0.8-1.3)	0.461
LDL, mmol/L (IQR)	2.5 (1.9-3.2)	2.2 (1.8-3.0)	0.100
Triglyceride (mg/dL), mmol/L (IQR)	1.3 (0.8-3.4)	1.3 (0.8-3.6)	0.995
Serum calcium, mmol/L (IQR)	2.7 (2.1-2.3)	2.2 (2.1-2.3)	0.051
C-reactive protein, mg/L (IQR)	35.0 (11.7-90.0)	29.4 (8.7-85.3)	0.415
Pleural effusion, N (%)	89 (18.28)	35 (21.74)	0.333
Patients with SAP, N (%)	49 (10.1)	16 (9.9)	0.0092
Length of hospital stay, days (IQR)	10 (7-13)	11 (7-15)	0.964
Hospital mortality, N (%)	9 (1.85)	1 (0.62)	0.274

Data were mean ± standard deviation, or numbers and percentages, or median (25th–75th percentile), as appropriate. N, number; IQR, interquartile range; BMI, body mass index; SIRS, systemic inflammatory response syndrome; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol.

TABLE 2 | Comparison of clinical and laboratory findings between patients, with and without SAP in the training sample (487 patients).

Variable	No-SAP (n = 438)	SAP (n = 49)	P-value
Age, years (IQR)	46 (37-60)	51 (38-66)	0.115
Male sex, N (%)	270 (61.6)	31 (63.3)	0.825
Duration of symptoms (days, IQR)	1.8 ± 0.8	1.9 ± 0.8	0.799
BMI, kg/m^2 (IQR)	23.4 (20.9-26.3)	24.4 (22.1-26.6)	0.083
SIRS, N (%)	157 (35.8)	34 (69.4)	<0.001
Biliary etiology, N (%)	190 (43.4)	17 (34.7)	0.243
Laboratory findings			
Hematocrit (l/l)	0.42 (0.38-0.45)	0.44 (0.41-0.49)	0.007
Platelets (10^9/L)	202 (167-234)	184 (135-208)	0.005
Prothrombin time, s (IQR)	13.8 (13.1-14.6)	14.6 (13.6-15.3)	0.004
Albumin, g/L (IQR)	37.1 (33.3-39.3)	30.4 (27.5-33.9)	<0.001
Total bilirubin, mmol/L (IQR)	20 (14-31)	20 (15-28)	0.631
ALT, U/L (IQR)	43 (18-121)	48 (24-77)	0.868
AST, U/L (IQR)	36 (21-88)	60 (41-85)	0.005
Glucose, mmol/L (IQR)	7.7 (6.2-10.0)	10.2 (8.2-14.4)	<0.001
Serum creatinine, μmol/L (IQR)	63 (54-76)	81 (59-154)	<0.001
BUN, mmol/L (IQR)	4.6 (3.6-5.9)	7.3 (5.1-11.4)	<0.001
Total cholesterol, N			0.001
<160 mmol/L	203 (95.75)	9 (9.45)	
160-240 mmol/L	131 (83.97)	25 (16.03)	
>240 mmol/L	104 (87.39)	15 (12.61)	
HDL, mmol/L (IQR)	1.0 (0.8-1.3)	0.6 (0.4-1.0)	<0.001
LDL, mmol/L (IQR)	2.6 (2.0-3.3)	1.6 (1.3-2.4)	<0.001
Triglyceride (mg/dL), mmol/L (IQR)	1.3 (0.8-3.3)	2.4 (1.3-7.2)	<0.001
Serum calcium, mmol/L (IQR)	2.2 (2.1-2.3)	1.9 (1.6-2.1)	<0.001
C-reactive protein, mg/L (IQR)	30.5 (10.4-87.8)	80.0 (28.4-90.0)	0.003
Pleural effusion, N (%)	59 (13.47)	30 (61.22)	<0.001

Data were mean ± standard deviation, or numbers and percentages, or median (25th–75th percentile), as appropriate. N, number; IQR, interquartile range; BMI, body mass index; SIRS, systemic inflammatory response syndrome; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol.

2.38-10.94, P<0.001). The LR model was developed to predict SAP as the following function: -1.10-0.13×albumin (g/L) + 0.016× serum creatinine (μmol/L) +0.14 × glucose(mmol/L) + 1.63 × pleural effusion (0/1)(No/Yes). The coefficients of this formula were utilized to build a nomogram for the prediction of SAP

(Figure 1). The Hosmer-Lemeshow goodness-of-fit test was significant (P=0.87), suggesting that our prediction model fit the actual data well.

The same 16 variables (SIRS, hematocrit, platelets, prothrombin time, albumin, AST, glucose, serum creatinine, BUN, cholesterol,
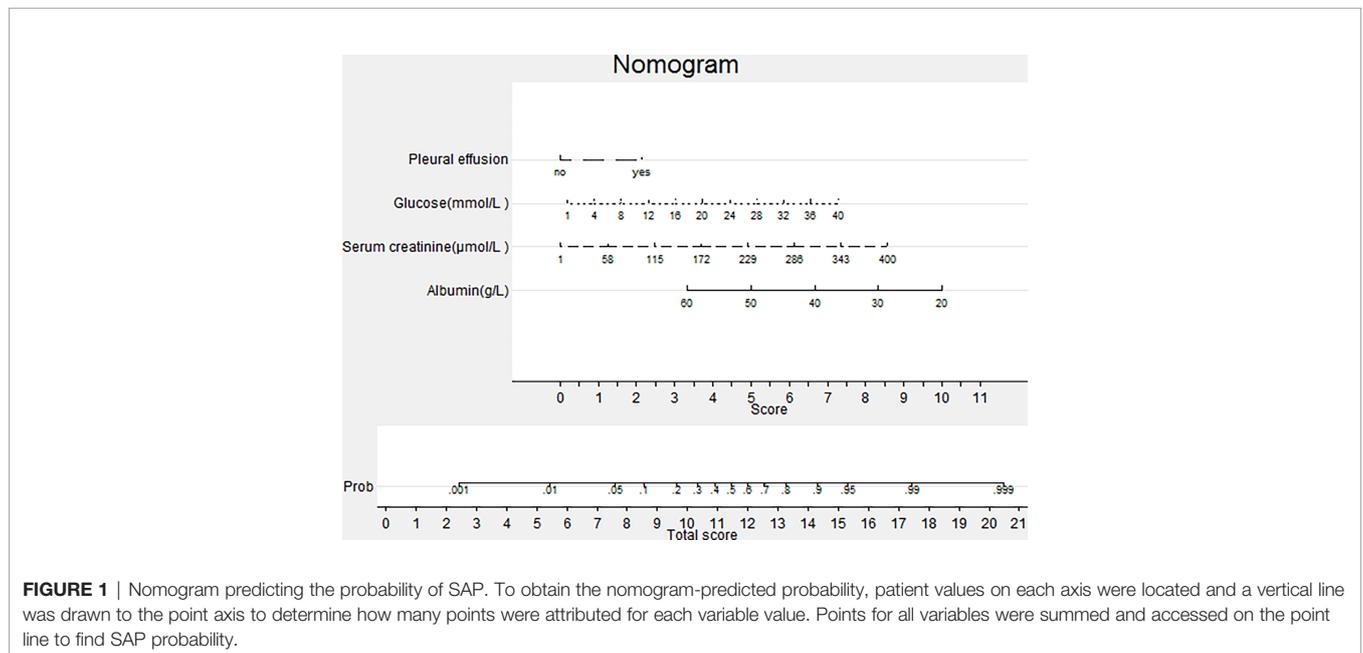

FIGURE 1 | Nomogram predicting the probability of SAP. To obtain the nomogram-predicted probability, patient values on each axis were located and a vertical line was drawn to the point axis to determine how many points were attributed for each variable value. Points for all variables were summed and accessed on the point line to find SAP probability.

HDL, LDL, triglyceride, serum calcium, C-reactive protein, and pleural effusion) were used for the RF model. As shown in **Figure 2**, based on variable important analysis of the RF model, serum creatinine, albumin, blood urea nitrogen, HDL, LDL, calcium, and glucose were the most important 7 predictors of SAP. **Figure 3** depicts the results of tenfold cross-validation. It indicated that the RF model achieved a higher mean AUC (AUC=0.89[95% CI, 0.83-0.95]) than that of the LR model (mean AUC =0.85[95% CI, 0.78-0.92]) (p=0.026). The area under the precision recall curve of the RF model (0.58) was also higher than that of the LR model (0.55) (**Figure 4**). The calibration plots indicate adequate predicted probabilities against observed proportions of SAP for both RF and LR models (**Figure 5**).

Validation and Comparison of Prediction Models on the Test Samples

The ROC curves for the RF model, the LR model, and the BISAP score for the prediction of SAP are shown in **Figure 6**. The RF model achieved the highest AUC (AUC=0.96[95% CI, 0.93-0.99]), followed by the LR model (AUC =0.92[95% CI, 0.87-0.97]) and the BISAP score (AUC=0.84[95% CI, 0.73-0.93]) (P=0.03). The area under precision recall curve of the RF model (0.67) was higher than that of both the LR model (0.57) and the BISAP score (0.576) (**Figure 7**).

The RF model achieved a sensitivity of 93.8%, specificity of 82.8%, and a diagnostic accuracy of 83.9. As a comparison, the LR model achieved a similar sensitivity of 93.8%, a lower
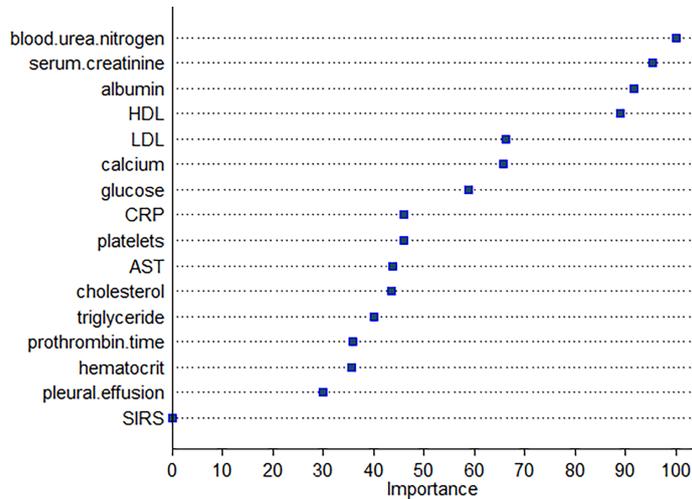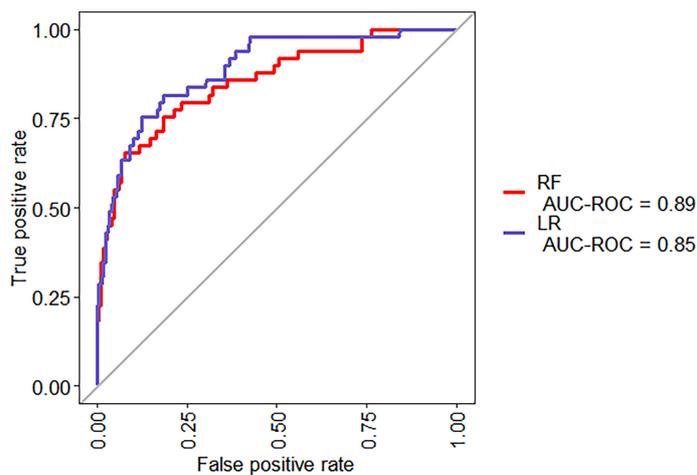

FIGURE 2 | Variable importance plot of the RF for SAP.

FIGURE 3 | ROC curves for the RF and LR models, for a tenfold cross-validation on the training set.
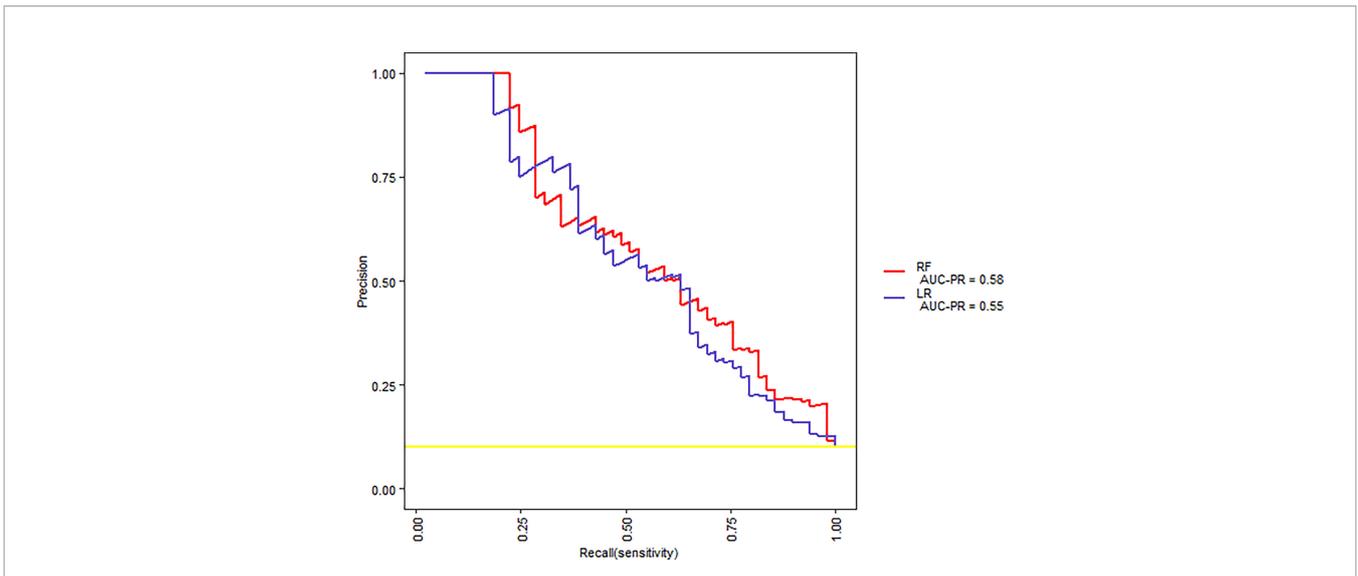
FIGURE 4 | The precision-recall curves for RF and LR models for tenfold cross-validation on the training set.
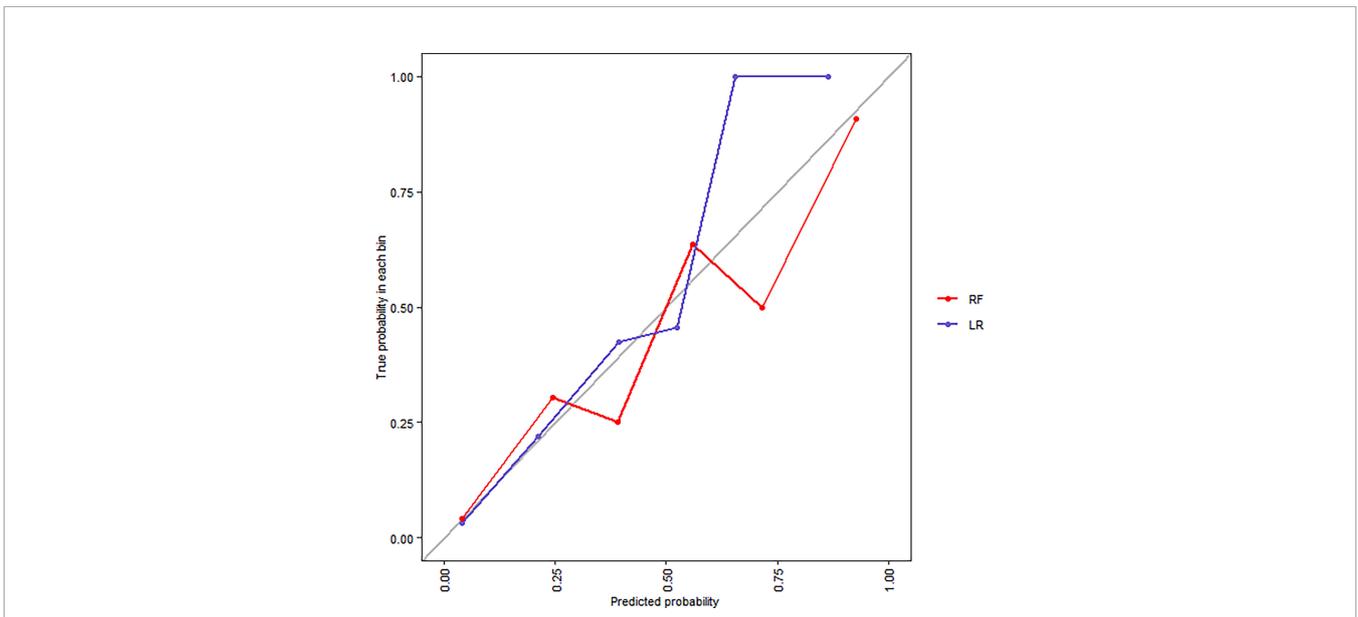

FIGURE 5 | Calibration plots for RF and LR models for tenfold cross-validation on the training set.

specificity of 79.3%, and 80.8% diagnostic accuracy. Both diagnostic performance of the RF and LR models was better than that of the BISAP score (**Table 3**).

Explanation: Individualized Prediction on The Test Sample

To clarify the model prediction for individual patients, the LIME plot was generated. It shows two typical predictions made by the RF model, in which one was for non-SAP and the other was for SAP patients (**Figure 8**). The bar charts represent the influence that individual covariates have on the overall prediction (Chan et al.,

2022). The length of the bar indicates the magnitude (absolute value), while the color indicates the sign (red for negative, blue for positive) of the estimated coefficient (Biecek and Burzykowski, 2021). In other words, the length of the bar for each feature indicates the importance (weight) of that feature in making the prediction. A longer bar, therefore, indicates a feature that contributes more toward or against the prediction (Lin et al., 2021a).

As shown in **Figure 8**, the first case (case 49) is a non-SAP patient, who was correctly classified based on the RF model, with a predicted probability of 0.97 as non-SAP. The second case (case 51) is an SAP patient, who was correctly classified based on the
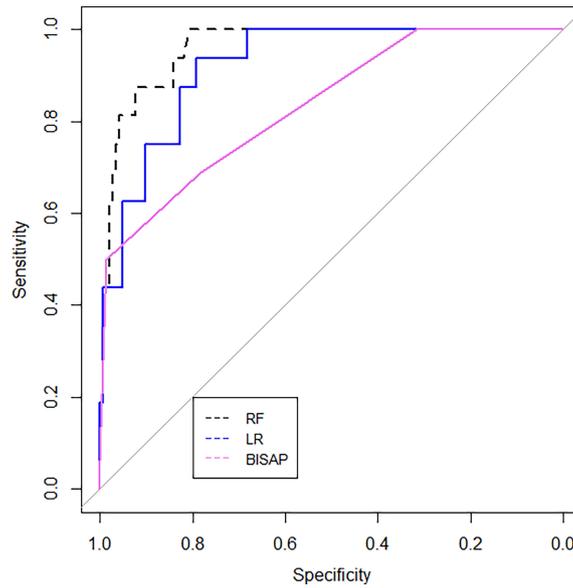
FIGURE 6 | ROC curves for the RF and LR models and BISAP scores, applied on the test set.

TABLE 3 | Diagnostic values of various models of SAP.

Variable	Cut-off value	Sensitivity	Specificity	LR+	LR-	Accuracy
RF model	0.13	93.8%	82.8%	5.44	0.08	83.9%
LR model	0.08	93.8%	79.3%	4.53	0.08	80.8%
BISAP score	2	68.8%	78.6%	3.22	0.40	77.64%

LR+, Positive likelihood ratio; LR-, negative likelihood ratio.

RF model, with a predicted probability of 0.82 as SAP. The levels of creatinine, BUN, glucose, triglyceride, and total cholesterol were positively correlated with the development of SAP. Patients with SAP had lower levels of HDL, albumin, and calcium than that of non-SAP people.

DISCUSSION

Albumin is one of the most important proteins in plasma and plays a role in maintaining osmotic pressure, antioxidants, and scavenging free radicals (Viasus et al., 2013). Albumin has also long been considered a negative acute-phase protein, with reduced production in inflammation, paving the way for inflammatory cytokines (Charlie-Silva et al., 2019). Serum albumin levels undoubtedly decrease in inflammatory states, which may result in shorter half-life and a larger interstitial pool (Barle et al., 2006) as well as capillary leak (Soeters et al., 2019), during the inflammatory process. Excessive oxidative stress is associated with damage to acinar cells which has been observed in cerulein-induced mouse models of AP (Shen et al., 2018). In addition, clinical evidence suggests that oxidative stress is common in the early phase of AP (Hackert and Werner, 2011). Therefore, it was suggested that

decreased albumin may reduce the ability to counterwork oxidative stress-induced acinar damage by binding reactive oxygen species in AP (Xu et al., 2020; Belinskaia et al., 2021). Xu et al. has reported that albumin is an independent predictor for SAP and in-hospital mortality in AP patients (Xu et al., 2020). Our previous study also indicated that hypoalbuminemia within 24 h of admission is independently associated with the development of persistent organ failure and mortality in AP (Hong et al., 2017a). Ocskay et al. (Ocskay et al., 2021) reported that the incidence of hypoalbuminemia was 35.7% during hospitalization and it was dose-dependent, associated with severity and mortality in AP. In our current study, the LR analysis indicated that albumin (p<0.001) is an independent predictor of SAP (**Figure 1**). Based on the RF model, albumin is also an important predictor of SAP, based on variable importance analysis (**Figure 2**). These results are consistent with previous reports.

Creatinine is primarily generated by muscle mass and dietary intake. It is eliminated from the glomerular filtration membrane (Stevens et al., 2006; Earley et al., 2012) and serves as the most widely used functional biomarker of the kidney, which can reflect renal injury in AP (Earley et al., 2012). Apart from renal injury, it also has been reported that the level of serum creatinine is associated with pancreatic necrosis (Muddana et al., 2009; Papachristou et al., 2010; Lipinski et al., 2013). The possible
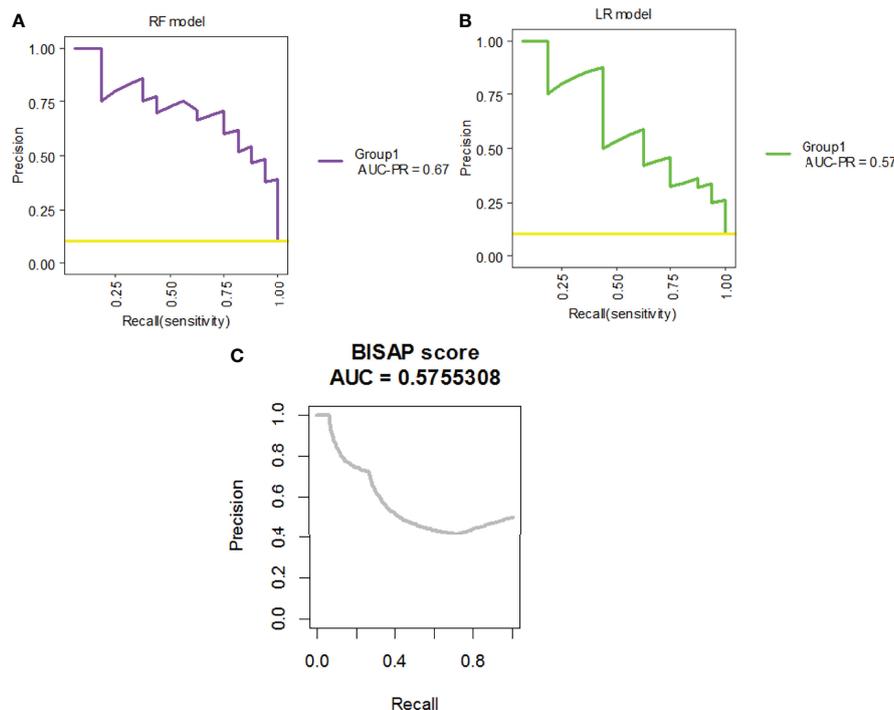
FIGURE 7 | The precision-recall curves for the **(A)** RF model, **(B)** LR model, and **(C)** BISAP score applied on the test set.

explanation is that necrotic cells release a large number of toxic substances and pro-inflammatory factors to cause renal injury, manifesting the elevation of serum creatinine. Therefore, the rise of creatinine may be attributed to the renal injury and pancreatic necrosis along with SAP. Wilkman et al. reported that increased creatinine levels are independently associated with 90-day mortality in AP patients (Wilkman et al., 2013). Wan et al. suggested that serum creatinine levels within 24 h of admission are effective for predicting persistent organ failure in AP (Wan et al., 2019). In addition, several scoring systems, which take creatinine as an index, are widely used in the clinical settings, including the Acute Physiology and Chronic Health Evaluation (APACHE) II, Sequential Organ Failure Assessment (SOFA) score for predicting the severity of pancreatitis, and modified Marshall scoring system for assessing organ dysfunction occurrence in SAP (Mederos et al., 2021). Our study indicated serum creatinine could be a useful predictor in both the RF and the LR model for predicting SAP (**Figures 1**, **2**).

The lipoprotein profile, especially HDL, is markedly decreased in inflammation and the accompanying acute-phase (Jahangiri, 2010). The mechanisms causing low serum HDL and LDL levels in the acute phase of AP remain largely unknown (Hong et al., 2017b; Hong et al., 2018). Jahangiri et al. (Jahangiri, 2010) suggested that it was related to a decreased rate of lipoprotein synthesis in the liver, general catabolism, and activation of the inflammatory system in the acute phase of the disease. Another explanation for the low serum HDL levels is that it may be due to increased expression of the Toll-like receptors (TLRs), especially TLR-4 expression (Zhang et al., 2010). It was reported that

stimulated TLR-4 expression suppresses HDL levels (Liao et al., 1999). Khan et al. found that serum lipid concentrations such as HDL cholesterol and LDL cholesterol were associated with patients of SAP in all etiologies (Khan et al., 2013). However, Bugdaci et al. (2011) found a significant association between decreased HDL level and severity of the disease only in alcoholic and hypotriglyceridemic pancreatitis. In hypertriglyceridemic status, it is demonstrated that free fatty acids (FFAs) damage acinar cells and cause pancreatitis attack due to premature activation of trypsinogen, by creating an acidic environment (Okura et al., 2004; Guo et al., 2019). It has been reported that HDL takes part in FFA clearance (Asztalos et al., 2007) so that decreased HDL in hypertriglyceridemic AP cases may lead to an increase in FFA, and further damage acinar cells. Therefore, it has been suggested that an increase in HDL may be helpful for recovery from the disease by contributing to antioxidants (Bugdaci et al., 2011) and anti-inflammatory effect (Murphy and Woollard, 2010). On the other hand, in comparison, few studies are available about the pathophysiological mechanism of decreased LDL in SAP. Our study indicated both HDL and LDL were useful predictors for SAP (**Figure 2**).

Pleural effusion occurs in 3–50% patients with AP, based on a previous study (Basran et al., 1987; Kumar et al., 2019; Peng et al., 2020). The effusion can be asymptomatic and often hemorrhagic, usually resolving as pancreatitis subsides (Basran et al., 1987). Several mechanisms of pleural effusion in pancreatitis have been proposed, such as the trans-diaphragmatic lymphatic blockage, the pancreatic pleural fistula caused by the rupture of the pancreatic duct, and the fluid exudation from the sub-pleural
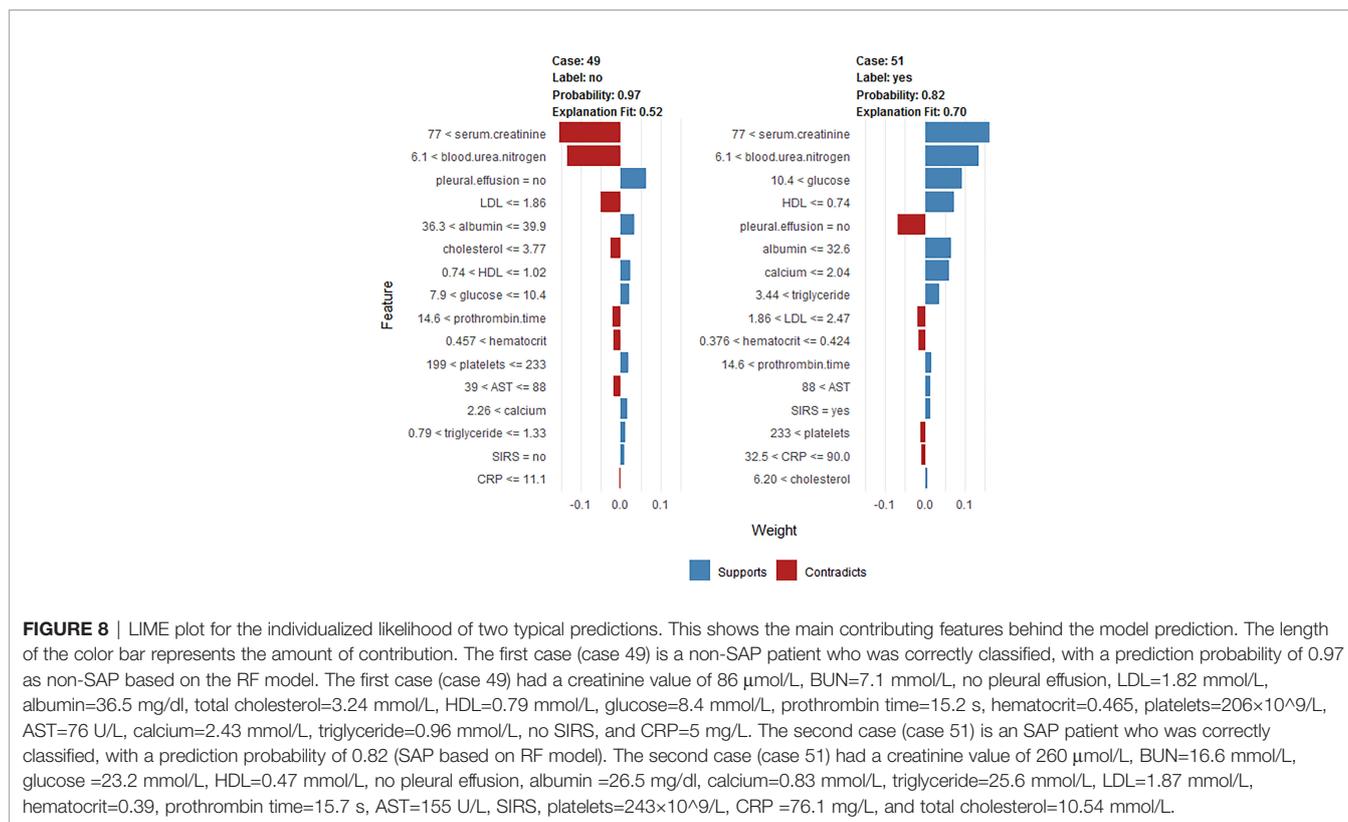
FIGURE 8 | LIME plot for the individualized likelihood of two typical predictions. This shows the main contributing features behind the model prediction. The length of the color bar represents the amount of contribution. The first case (case 49) is a non-SAP patient who was correctly classified, with a prediction probability of 0.97 as non-SAP based on the RF model. The first case (case 49) had a creatinine value of 86 μmol/L, BUN=7.1 mmol/L, no pleural effusion, LDL=1.82 mmol/L, albumin=36.5 mg/dl, total cholesterol=3.24 mmol/L, HDL=0.79 mmol/L, glucose=8.4 mmol/L, prothrombin time=15.2 s, hematocrit=0.465, platelets=206×10^9/L, AST=76 U/L, calcium=2.43 mmol/L, triglyceride=0.96 mmol/L, no SIRS, and CRP=5 mg/L. The second case (case 51) is an SAP patient who was correctly classified, with a prediction probability of 0.82 (SAP based on RF model). The second case (case 51) had a creatinine value of 260 μmol/L, BUN=16.6 mmol/L, glucose =23.2 mmol/L, HDL=0.47 mmol/L, no pleural effusion, albumin =26.5 mg/dl, calcium=0.83 mmol/L, triglyceride=25.6 mmol/L, LDL=1.87 mmol/L, hematocrit=0.39, prothrombin time=15.7 s, AST=155 U/L, SIRS, platelets=243×10^9/L, CRP =76.1 mg/L, and total cholesterol=10.54 mmol/L.

diaphragmatic vessels into the pleural cavity (Kumar et al., 2019). Pleural effusion is reported to be associated with a severe course for initial risk assessment severity in AP and a sign of SAP (Heller et al., 1997; Tenner et al., 2013; Lankisch et al., 2015). Yan et al. reported that pleural effusion volume quantified on chest CT was positively associated with the duration of hospitalization (Yan et al., 2021). As a prognostic factor, pleural effusion has been incorporated in SAP severity predictive systems such as the Bedside Index for Severity in Acute Pancreatitis (BISAP) score (Gao et al., 2015), the Panc 3 score (Brown et al., 2007), and the Extra Pancreatic Inflammation on CT (EPIC) score (De Waele et al., 2007). Following the above outcomes, the present study suggested that pleural effusion (OR 5.11, 95%CI 2.38-10.94) was an independent risk factor for SAP (**Figure 1**).

The mechanism of BUN elevation in AP is thought to be based on the loss of intravascular volume, caused by interstitial extravasations owing to the systemic inflammatory response syndrome and an AP promoted direct renal injury mechanism. It has been reported that BUN, as a single predictor, had moderate accuracy in predicting persistent organ failure in AP (Mounzer et al., 2012). Koutroumpakis et al. reported that the rise in BUN at 24 h was the most accurate in predicting persistent organ failure and pancreatic necrosis (Koutroumpakis et al., 2015). Li et al. suggested that BUN was an independent risk factor to predict in-hospital mortality (Li et al., 2020). Valverde-Lopez et al. indicated that BUN was the best predictor of SAP after 48 h (Valverde-Lopez et al., 2017). BUN is also included in many scoring systems for AP, such as BISAP, JSS, and Glasgow

score (Mounzer et al., 2012). Consistent with these reports, our study shows that BUN is the most important predictor of the RF model based on variable important analysis (**Figure 2**).

Decreased levels of serum calcium are commonly seen in critical illness, and hypocalcemia is significantly more frequent in patients with SAP (Peng et al., 2017). The mechanisms of hypocalcemia in SAP may be multi-factorial, such as abnormalities of parathyroid hormone secretion and action as well as vitamin D deficiency, binding of calcium in areas of fat necrosis, likely to contribute to the medication side effects (Weir et al., 1975; Steele et al., 2013). Serum calcium levels are closely related to the severity of the disease and its complications in AP. It has been incorporated in several clinical scoring systems as such as Pancreatitis Outcome Prediction (POP) Score, and Simple Prognostic Score (Harrison et al., 2007; Gonzálvez-Gasch et al., 2009). Mentula et al. suggested that serum calcium was the best single marker in predicting organ failure in AP after 24 h of symptom onset (Mentula et al., 2005). He et al. indicated that serum calcium was one of the independent predictors of the severity of AP in elderly patients (He et al., 2021). Serum calcium was also considered a significant factor in predicting early death in SAP (Shinzeki et al., 2008). As expected, our study indicated that calcium could be a useful predictor of SAP in the RF model (**Figure 2**).

Clinical evidence shows hyperglycemia is the common early feature of AP and abnormal glucose metabolism is present in almost 40% of AP patients (Banks et al., 2013; Chen et al., 2021). According to the traditional view, the mechanism is that the damage of organisms caused by AP activate the neuroendocrine system and

lead to the secretion of many stress hormones (Binker and Cosen-Binker, 2014; Sun et al., 2019; Lu et al., 2021). Meanwhile, it is also related to the damage of the endocrine pancreas caused by SAP attacks. The association between hyperglycemia and adverse clinical outcomes in critically ill patients has been demonstrated in several observational studies, which suggest that high levels of glucose during the progression of AP can promote the release of inflammatory cytokines. These, in turn, influence disease progression (Sun et al., 2019; Chen et al., 2021). Sun et al. has suggested that the level of glucose in serum is positively correlated with the APACHE II scores, TNF-α, and CRP in AP (Sun et al., 2019). However, transient stress hyperglycemia in critically ill patients is considered harmless in some studies, indicating that the body has normal immune regulation ability (Lu et al., 2021), the subsequent derangement of glucose homeostasis could cause damage to the body (Pendharkar et al., 2016). Blood glucose-related indicators are associated with in-hospital mortality in critically ill patients with AP (Lu et al., 2021). Our LR model also shows that glucose is a useful predictor of SAP (**Figure 1**).

Machine learning has been extensively used for the prediction of severity or complication of AP (Zhou et al., 2022). Thapa et al. has reported that an XGBoost model could predict which patients would require treatment for SAP (Thapa et al., 2022). Early prediction of SAP using machine learning has also been attempted (Thapa et al., 2022). Jin et al. reported that the multilayer perception-artificial neural network (MPL-ANN) model based on routine blood and serum biochemical indexes could reliably predict disease severity in patients with AP (Jin et al., 2021). Choi et al. combined clinical (i.e., APACHE-II and BISAP scores) and radiologic (i.e., Balthazar grade and EPIC score) scoring systems by classification tree analysis for predicting SAP (Choi et al., 2018). Xu et al. reported that adaptive boosting algorithm (AdaBoost) could predict development of multiple organ failure, complicated by moderately severe or severe AP (Xu et al., 2021). However, the above models were limited due to lack of individualized prediction on the test sample. Implementation on such data remains challenging because of the low interpretability of results of machine learning (Yu et al., 2021). Our study indicated that, compared to the LR model and BISAP score, RF exhibited the highest discriminatory performance for the prediction of SAP on both training and test samples (**Figures 3**, **4**, **6**, **7**). Using the RF model, we could illustrate key features and establish a prediction model, with high accuracy in patients with SAP. The LIME plot could provide a visual illustration of the individualized interpretation of the importance of different features, which might help clinical doctors to understand results of the RF model (**Figure 8**). The LR model (nomogram) achieved a sensitivity of 93.8%, acceptable specificity of 79.3%, and diagnostic accuracy of 80.8% (**Table 3**). Though the diagnostic performance of the LR model (nomogram) is inferior to the RF model, it is simple and intuitive to calculate the prediction probability of a result, which makes it valuable in predicting SAP (**Figure 1**).

To the best of our knowledge, this is the first study to develop an interpretable RF model for SAP prediction. The strength of this study is a large sample size, which enables a strong statistical power. Both patients in ICU and the general ward were enrolled in this study, thus reducing selection bias. However, our study has some limitations, even if it has been internally validated by tenfold cross-validation technique and test set, testing the performance of our RF model in an external/other independent data set is necessary. In addition, even if effective, RF models are sophisticated and difficult to understand, and thus, comparable to a 'black box'. We have, therefore, demonstrated that by utilizing Lime plots, the results could be more easily interpreted (Al'Aref et al., 2020). At last, we did not evaluate the RF model and single predictors for other clinical outcomes such as patient survival and organ failure occurrence, intensive care unit (ICU) admission, and SAP recurrent rate. It would be interesting to carry out a large-sample prospective study to determine whether our model and other variables such as serum creatinine, albumin, BUN, HDL, LDL, calcium, and glucose play a significant role in predicting these clinical outcomes.

In conclusion, an interpretable RF model exhibited the highest discriminatory performance to predict SAP. Interpretation with LIME plots could be useful for individualized prediction in the clinical setting. A nomogram consisting of albumin, serum creatinine, glucose, and pleural effusion is also useful for the prediction of SAP.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

WH conceived the study and carried out the majority of the work. WH participated in data collection and conducted data analysis. WH, YL, XZ, SJ, JP, QL, and SY drafted the manuscript. ZB, MZ, and HG helped to finalize the manuscript. All the authors read and approved the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2022.893294/full#supplementary-material

Supplementary Figure 1 | Data flow diagram of this study.

REFERENCES

Al'Aref, S. J., Maliakal, G., Singh, G., Van Rosendael, A. R., Ma, X., Xu, Z., et al. (2020). Machine Learning of Clinical Variables and Coronary Artery Calcium Scoring for the Prediction of Obstructive Coronary Artery Disease on Coronary Computed Tomography Angiography: Analysis From the CONFIRM Registry. *Eur. Heart J.* 41, 359–367. doi: 10.1093/eurheartj/ehz565

Asztalos, B. F., Schaefer, E. J., Horvath, K. V., Yamashita, S., Miller, M., Franceschini, G., et al. (2007). Role of LCAT in HDL Remodeling: Investigation of LCAT Deficiency States. *J. Lipid Res.* 48, 592–599. doi: 10.1194/jlr.M600403-JLR200

Banks, P. A., Bollen, T. L., Dervenis, C., Gooszen, H. G., Johnson, C. D., Sarr, M. G., et al. (2013). Classification of Acute Pancreatitis–2012: Revision of the Atlanta Classification and Definitions by International Consensus. *Gut* 62, 102–111. doi: 10.1136/gutjnl-2012-302779

Barle, H., Hammarqvist, F., Westman, B., Klaude, M., Rooyackers, O., Garlick, P. J., et al. (2006). Synthesis Rates of Total Liver Protein and Albumin Are Both Increased in Patients With an Acute Inflammatory Response. *Clin. Sci. (Lond.)* 110, 93–99. doi: 10.1042/CS20050222

Basran, G. S., Ramasubramanian, R., and Verma, R. (1987). Intrathoracic Complications of Acute Pancreatitis. *Br. J. Dis. Chest* 81, 326–331. doi: 10.1016/0007-0971(87)90180-X

Belinskaia, D. A., Voronina, P. A., Shmurak, V. I., Jenkins, R. O., and Goncharov, N. V. (2021). Serum Albumin in Health and Disease: Esterase, Antioxidant, Transporting and Signaling Properties. *Int. J. Mol. Sci.* 22:1–37. doi: 10.3390/ijms221910318

Biecek, P., and Burzykowski, T. (2021). *Explanatory Model Analysis : Explore, Explain, and Examine Predictive Models* (New York, US: Chapman and Hall/CRC).

Binker, M. G., and Cosen-Binker, L. I. (2014). Acute Pancreatitis: The Stress Factor. *World J. Gastroenterol.* 20, 5801–5807. doi: 10.3748/wjg.v20.i19.5801

Bramhall, S., Horn, H., Tieu, M., and Lohia, N. (2020). Qlime-A Quadratic Local Interpretable Model-Agnostic Explanation Approach. *SMU Data Sci. Rev.* 3, 4.

Brown, A., James-Stevenson, T., Dyson, T., and Grunkenmeier, D. (2007). The Panc 3 Score: A Rapid and Accurate Test for Predicting Severity on Presentation in Acute Pancreatitis. *J. Clin. Gastroenterol.* 41, 855–858. doi: 10.1097/01.mcg.0000248005.73075.e4

Bugdaci, M. S., Sokmen, M., Zuhur, S. S., and Altuntas, Y. (2011). Lipid Profile Changes and Importance of Low Serum α-Lipoprotein Fraction (High-Density Lipoprotein) in Cases With Acute Pancreatitis. *Pancreas* 40, 1241–1244. doi: 10.1097/MPA.0b013e3182211bbf

Chan, M. C., Pai, K. C., Su, S. A., Wang, M. S., Wu, C. L., and Chao, W. C. (2022). Explainable Machine Learning to Predict Long-Term Mortality in Critically Ill Ventilated Patients: A Retrospective Study in Central Taiwan. *BMC Med. Inform. Decis. Mak.* 22, 75. doi: 10.1186/s12911-022-01817-6

Charlie-Silva, I., Klein, A., Gomes, J. M. M., Prado, E. J. R., Moraes, A. C., Eto, S. F., et al. (2019). Acute-Phase Proteins During Inflammatory Reaction by Bacterial Infection: Fish-Model. *Sci. Rep.* 9, 4776. doi: 10.1038/s41598-019-41312-z

Chen, Y., Tang, S., and Wang, Y. (2021). Prognostic Value of Glucose-To-Lymphocyte Ratio in Critically Ill Patients With Acute Pancreatitis. *Int. J. Gen. Med.* 14, 5449–5460. doi: 10.2147/IJGM.S327123

Choi, H. W., Park, H. J., Choi, S. Y., Do, J. H., Yoon, N. Y., Ko, A., et al. (2018). Early Prediction of the Severity of Acute Pancreatitis Using Radiologic and Clinical Scoring Systems With Classification Tree Analysis. *AJR Am. J. Roentgenol.* 211, 1035–1043. doi: 10.2214/AJR.18.19545

Deshmukh, F., and Merchant, S. S. (2020). Explainable Machine Learning Model for Predicting GI Bleed Mortality in the Intensive Care Unit. *Am. J. Gastroenterol.* 115, 1657–1668. doi: 10.14309/ajg.0000000000000632

De Waele, J. J., Delrue, L., Hoste, E. A., De Vos, M., Duyck, P., and Colardyn, F. A. (2007). Extrapancreatic Inflammation on Abdominal Computed Tomography as an Early Predictor of Disease Severity in Acute Pancreatitis: Evaluation of a New Scoring System. *Pancreas* 34, 185–190. doi: 10.1097/mpa.0b013e31802d4136

Earley, A., Miskulin, D., Lamb, E., Levey, A., and Uhlig, K. (2012). Estimating Equations for Glomerular Filtration Rate in the Era of Creatinine Standardization: A Systematic Review. *Ann. Intern. Med.* 156, 785–795, W-270, W-271, W-272, W-273, W-274, W-275, W-276, W-277, W-278. doi: 10.7326/0003-4819-156-11-201203200-00391

Gao, W., Yang, H. X., and Ma, C. E. (2015). The Value of BISAP Score for Predicting Mortality and Severity in Acute Pancreatitis: A Systematic Review and Meta-Analysis. *PloS One* 10, e0130412. doi: 10.1371/journal.pone.0142025

Genuer, R., and Poggi, J.-M. (2020). *Random Forests With R* (Cham, Switzerland: Springer).

Gibor, U., Perry, Z., Netz, U., Kirshtein, B., Mizrahi, S., Czeiger, D., et al. (2020). Circulating Cell-Free DNA in Patients With Acute Biliary Pancreatitis: Association With Disease Markers and Prolonged Hospitalization Time-A Prospective Cohort Study. *Ann. Surg.* doi: 10.1097/SLA.0000000000004679

Gonzálvez-Gasch, A., De Casasola, G., Martín, R., Herreros, B., and Guijarro, C. (2009). A Simple Prognostic Score for Risk Assessment in Patients With Acute Pancreatitis. *Eur. J. Intern. Med.* 20, e43–e48. doi: 10.1016/j.ejim.2008.09.014.

Guo, Y. Y., Li, H. X., Zhang, Y., and He, W. H. (2019). Hypertriglyceridemia-Induced Acute Pancreatitis: Progress on Disease Mechanisms and Treatment Modalities. *Discov. Med.* 27, 101–109.

Hackert, T., and Werner, J. (2011). Antioxidant Therapy in Acute Pancreatitis: Experimental and Clinical Evidence. *Antioxid. Redox Signal.* 15, 2767–2777. doi: 10.1089/ars.2011.4076

Harrison, D., D'amico, G., and Singer, M. (2007). The Pancreatitis Outcome Prediction (POP) Score: A New Prognostic Index for Patients With Severe Acute Pancreatitis. *Crit. Care Med.* 35, 1703–1708. doi: 10.1097/01.CCM.0000269031.13283.C8

Heller, S. J., Noordhoek, E., Tenner, S. M., Ramagopal, V., Abramowitz, M., Hughes, M., et al. (1997). Pleural Effusion as a Predictor of Severity in Acute Pancreatitis. *Pancreas* 15, 222–225. doi: 10.1097/00006676-199710000-00002

He, F., Zhu, H., Li, B., Li, X., Yang, S., Wang, Z., et al. (2021). Factors Predicting the Severity of Acute Pancreatitis in Elderly Patients. *Aging Clin. Exp. Res.* 33, 183–192. doi: 10.1007/s40520-020-01523-1

Hong, W., Chen, Q., Qian, S., Basharat, Z., Zimmer, V., Wang, Y., et al. (2021). Critically Ill vs. Non-Critically Ill Patients With COVID-19 Pneumonia: Clinical Features, Laboratory Findings, and Prediction. *Front. Cell Infect. Microbiol.* 11, 550456. doi: 10.3389/fcimb.2021.550456.

Hong, W., Lillemoe, K. D., Pan, S., Zimmer, V., Kontopantelis, E., Stock, S., et al. (2019). Development and Validation of a Risk Prediction Score for Severe Acute Pancreatitis. *J. Transl. Med.* 17, 146. doi: 10.1186/s12967-019-1903-6

Hong, W., Lin, S., Zippi, M., Geng, W., Stock, S., Basharat, Z., et al. (2017a). Serum Albumin Is Independently Associated With Persistent Organ Failure in Acute Pancreatitis. *Can. J. Gastroenterol. Hepatol.* 2017, 5297143. doi: 10.1155/2017/5297143

Hong, W., Lin, S., Zippi, M., Geng, W., Stock, S., Zimmer, V., et al. (2017b). High-Density Lipoprotein Cholesterol, Blood Urea Nitrogen, and Serum Creatinine Can Predict Severe Acute Pancreatitis. *BioMed. Res. Int.* 2017, 1648385. doi: 10.1155/2017/1648385

Hong, W., Zimmer, V., Basharat, Z., Zippi, M., Stock, S., Geng, W., et al. (2020). Association of Total Cholesterol With Severe Acute Pancreatitis: A U-Shaped Relationship. *Clin. Nutr.* 39, 250–257. doi: 10.1016/j.clnu.2019.01.022

Hong, W., Zimmer, V., Stock, S., Zippi, M., Omoshoro-Jones, J. A., and Zhou, M. (2018). Relationship Between Low-Density Lipoprotein Cholesterol and Severe Acute Pancreatitis ("the Lipid Paradox"). *Ther. Clin. Risk Manag.* 14, 981–989. doi: 10.2147/TCRM.S159387

Jahangiri, A. (2010). High-Density Lipoprotein and the Acute Phase Response. *Curr. Opin. Endocrinol. Diabetes Obes.* 17, 156–160. doi: 10.1097/MED.0b013e328337278b

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning With Applications in R* (Cham, Switzerland:Springer).

Jin, X., Ding, Z., Li, T., Xiong, J., Tian, G., and Liu, J. (2021). Comparison of MPL-ANN and PLS-DA Models for Predicting the Severity of Patients With Acute Pancreatitis: An Exploratory Study. *Am. J. Emerg. Med.* 44, 85–91. doi: 10.1016/j.ajem.2021.01.044

Khan, J., Nordback, I., and Sand, J. (2013). Serum Lipid Levels Are Associated With the Severity of Acute Pancreatitis. *Digestion* 87, 223–228. doi: 10.1159/000348438

Koutroumpakis, E., Wu, B. U., Bakker, O. J., Dudekula, A., Singh, V. K., Besselink, M. G., et al. (2015). Admission Hematocrit and Rise in Blood Urea Nitrogen at 24 H Outperform Other Laboratory Markers in Predicting Persistent Organ Failure and Pancreatic Necrosis in Acute Pancreatitis: A *Post Hoc* Analysis of Three Large Prospective Databases. *Am. J. Gastroenterol.* 110, 1707–1716. doi: 10.1038/ajg.2015.370

Kuhn, M. (2008). Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05

Kumar, P., Gupta, P., and Rana, S. (2019). Thoracic Complications of Pancreatitis. *JGH Open* 3, 71–79. doi: 10.1002/jgh3.12099

Langmead, C., Lee, P. J., Paragomi, P., Greer, P., Stello, K., Hart, P. A., et al. (2021). A Novel 5-Cytokine Panel Outperforms Conventional Predictive Markers of Persistent Organ Failure in Acute Pancreatitis. *Clin. Transl. Gastroenterol.* 12, e00351. doi: 10.14309/ctg.0000000000000351

Lankisch, P. G., Apte, M., and Banks, P. A. (2015). Acute Pancreatitis. *Lancet* 386, 85–96. doi: 10.1016/S0140-6736(14)60649-8

Lee, P. J., and Papachristou, G. L. (2019). New Insights Into Acute Pancreatitis. *Nat. Rev. Gastroenterol. Hepatol.* 16, 479–496. doi: 10.1038/s41575-019-0158-2

Liao, W., Rudling, M., and Angelin, B. (1999). Endotoxin Suppresses Mouse Hepatic Low-Density Lipoprotein-Receptor Expression *via* a Pathway Independent of the Toll-Like Receptor 4. *Hepatology* 30, 1252–1256. doi: 10.1002/hep.510300524

Liaw, A., and Wiener, M. (2002). Classification and Regression by Randomforest. *R News* 2, 18–22.

Lime: Local Interpretable Model-Agnostic Explanations. (2021) Available at: https://cran.r-project.org/web/packages/lime/index.html (Accessed 2022-4-20).

Lin, S., Hong, W., Basharat, Z., Wang, Q., Pan, J., and Zhou, M. (2017). Blood Urea Nitrogen as a Predictor of Severe Acute Pancreatitis Based on the Revised Atlanta Criteria: Timing of Measurement and Cutoff Points. *Can. J. Gastroenterol. Hepatol.* 2017, 9592831. doi: 10.1155/2017/9592831

Lin, M. Y., Li, C. C., Lin, P. H., Wang, J. L., Chan, M. C., Wu, C. L., et al. (2021a). Explainable Machine Learning to Predict Successful Weaning Among Patients Requiring Prolonged Mechanical Ventilation: A Retrospective Cohort Study in Central Taiwan. *Front. Med. (Lausanne)* 8, 663739. doi: 10.3389/fmed.2021.663739

Lin, Z., Lin, S., Neamtiu, I. A., Ye, B., Csobod, E., Fazakas, E., et al. (2021b). Predicting Environmental Risk Factors in Relation to Health Outcomes Among School Children From Romania Using Random Forest Model - An Analysis of Data From the SINPHONIE Project. *Sci. Total Environ.* 784, 147145. doi: 10.1016/j.scitotenv.2021.147145

Lipinski, M., Rydzewski, A., and Rydzewska, G. (2013). Early Changes in Serum Creatinine Level and Estimated Glomerular Filtration Rate Predict Pancreatic Necrosis and Mortality in Acute Pancreatitis: Creatinine and eGFR in Acute Pancreatitis. *Pancreatology* 13, 207–211. doi: 10.1016/j.pan.2013.02.002

Li, C., Ren, Q., Wang, Z., and Wang, G. (2020). Early Prediction of in-Hospital Mortality in Acute Pancreatitis: A Retrospective Observational Cohort Study Based on a Large Multicentre Critical Care Database. *BMJ Open* 10, e041893. doi: 10.1136/bmjopen-2020-041893

Lo, F., Bitz, C. M., and Hess, J. J. (2021). Development of a Random Forest Model for Forecasting Allergenic Pollen in North America. *Sci. Total Environ.* 773, 145590. doi: 10.1016/j.scitotenv.2021.145590

Lu, Y., Zhang, Q., and Lou, J. (2021). Blood Glucose-Related Indicators Are Associated With in-Hospital Mortality in Critically Ill Patients With Acute Pancreatitis. *Sci. Rep.* 11, 15351. doi: 10.1038/s41598-021-94697-1

Mederos, M., Reber, H., and Girgis, M. (2021). Acute Pancreatitis: A Review. *JAMA* 325, 382–390. doi: 10.1001/jama.2020.20317

Mentula, P., Kylänpää, M., Kemppainen, E., Jansson, S., Sarna, S., Puolakkainen, P., et al. (2005). Early Prediction of Organ Failure by Combined Markers in Patients With Acute Pancreatitis. *Br. J. Surg.* 92, 68–75. doi: 10.1002/bjs.4786.

Mounzer, R., Langmead, C. J., Wu, B. U., Evans, A. C., Bishehsari, F., Muddana, V., et al. (2012). Comparison of Existing Clinical Scoring Systems to Predict Persistent Organ Failure in Patients With Acute Pancreatitis. *Gastroenterology* 142, 1476–1482; quiz e1415-1476. doi: 10.1053/j.gastro.2012.03.005

Muddana, V., Whitcomb, D. C., Khalid, A., Slivka, A., and Papachristou, G. I. (2009). Elevated Serum Creatinine as a Marker of Pancreatic Necrosis in Acute Pancreatitis. *Am. J. Gastroenterol.* 104, 164–170. doi: 10.1038/ajg.2008.66

Murphy, A. J., and Woollard, K. J. (2010). High-Density Lipoprotein: A Potent Inhibitor of Inflammation. *Clin. Exp. Pharmacol. Physiol.* 37, 710–718. doi: 10.1111/j.1440-1681.2009.05338.x

Ocsay, K., Vinko, Z., Nemeth, D., Szabo, L., Bajor, J., Godi, S., et al. (2021). Hypoalbuminemia Affects One Third of Acute Pancreatitis Patients and Is Independently Associated With Severity and Mortality. *Sci. Rep.* 11, 24158. doi: 10.1038/s41598-021-03449-8

Okura, Y., Hayashi, K., Shingu, T., Kajiyama, G., Nakashima, Y., and Saku, K. (2004). Diagnostic Evaluation of Acute Pancreatitis in Two Patients With Hypertriglyceridemia. *World J. Gastroenterol.* 10, 3691–3695. doi: 10.3748/wjg.v10.i24.3691

Pan, P., Li, Y., Xiao, Y., Han, B., Su, L., Su, M., et al. (2020). Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. *J. Med. Internet Res.* 22, e23128. doi: 10.2196/23128

Papachristou, G. I., Muddana, V., Yadav, D., and Whitcomb, D. C. (2010). Increased Serum Creatinine Is Associated With Pancreatic Necrosis in Acute Pancreatitis. *Am. J. Gastroenterol.* 105, 1451–1452. doi: 10.1038/ajg.2010.92

Pedersen, T. L., and Benesty, M. (2021) *Understanding Lime*. Available at: https://cran.r-project.org/web/packages/lime/vignettes/Understanding_lime.html (Accessed 2022-04-20).

Pendharkar, S. A., Asrani, V. M., Xiao, A. Y., Yoon, H. D., Murphy, R., Windsor, J. A., et al. (2016). Relationship Between Pancreatic Hormones and Glucose Metabolism: A Cross-Sectional Study in Patients After Acute Pancreatitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* 311, G50–G58. doi: 10.1152/ajpgi.00074.2016

Peng, T., Peng, X., Huang, M., Cui, J., Zhang, Y., Wu, H., et al. (2017). Serum Calcium as an Indicator of Persistent Organ Failure in Acute Pancreatitis. *Am. J. Emerg. Med.* 35, 978–982. doi: 10.1016/j.ajem.2017.02.006

Peng, R., Zhang, L., Zhang, Z. M., Wang, Z. Q., Liu, G. Y., and Zhang, X. M. (2020). Chest Computed Tomography Semi-Quantitative Pleural Effusion and Pulmonary Consolidation Are Early Predictors of Acute Pancreatitis Severity. *Quant. Imaging Med. Surg.* 10, 451–463. doi: 10.21037/qims.2019.12.14

Petch, J., Di, S., and Nelson, W. (2022). Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Can. J. Cardiol.* 38, 204–213. doi: 10.1016/j.cjca.2021.09.004

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016) *Why Should I Trust You?": Explaining the Predictions of Any Classifier*. Available at: https://arxiv.org/abs/1602.04938.

Roguet, A., Eren, A. M., Newton, R. J., and Mclellan, S. L. (2018). Fecal Source Identification Using Random Forest. *Microbiome* 6, 185. doi: 10.1186/s40168-018-0568-3

Royston, P. (2005). Multiple Imputation of Missing Values: Update of Ice. *Stata J.* 5, 527–536. doi: 10.1177/1536867X0500500404

Saito, T., and Rehmsmeier, M. (2015). The Precision-Recall Plot is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PloS One* 10, e0118432. doi: 10.1371/journal.pone.0118432

Shen, A., Kim, H. J., Oh, G. S., Lee, S. B., Lee, S., Pandit, A., et al. (2018). Pharmacological Stimulation of NQO1 Decreases NADPH Levels and Ameliorates Acute Pancreatitis in Mice. *Cell Death Dis.* 10, 5. doi: 10.1038/s41419-018-1252-z

Shinzeki, M., Ueda, T., Takeyama, Y., Yasuda, T., Matsumura, N., Sawa, H., et al. (2008). Prediction of Early Death in Severe Acute Pancreatitis. *J. Gastroenterol.* 43, 152–158. doi: 10.1007/s00535-007-2131-z

Soeters, P. B., Wolfe, R. R., and Shenkin, A. (2019). Hypoalbuminemia: Pathogenesis and Clinical Significance. *JPEN J. Parenter. Enteral Nutr.* 43, 181–193. doi: 10.1002/jpen.1451

Speiser, J. L., Miller, M. E., Tooze, J., and Ip, E. (2019). A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling. *Expert Syst. Appl.* 134, 93–101. doi: 10.1016/j.eswa.2019.05.028

Staniak, M., and Biecek, P. (2018). Explanations of Model Predictions With Live and Breakdown Packages. *R J.* 10, 395–409. doi: 10.32614/RJ-2018-072

Steele, T., Kolamunnage-Dona, R., Downey, C., Toh, C., and Welters, I. (2013). Assessment and Clinical Course of Hypocalcemia in Critical Illness. *Crit. Care (Lond. England)* 17, R106. doi: 10.1186/cc12756

Stevens, L., Coresh, J., Greene, T., and Levey, A. (2006). Assessing Kidney Function–Measured and Estimated Glomerular Filtration Rate. *N. Engl. J. Med.* 354, 2473–2483. doi: 10.1056/NEJMra054415

Sun, Y. F., Song, Y., Liu, C. S., and Geng, J. L. (2019). Correlation Between the Glucose Level and the Development of Acute Pancreatitis. *Saudi J. Biol. Sci.* 26, 427–430. doi: 10.1016/j.sjbs.2018.11.012

Takeda, T., Nakai, Y., Mizuno, S., Suzuki, T., Sato, T., Hakuta, R., et al. (2019). Fluid Sequestration Is a Useful Parameter in the Early Identification of Severe Disease of Acute Pancreatitis. *J. Gastroenterol.* 54, 359–366. doi: 10.1007/s00535-018-1531-6

Tenner, S., Baillie, J., Dewitt, J., and Vege, S. S. (2013). American College of Gastroenterology Guideline: Management of Acute Pancreatitis. *Am. J. Gastroenterol.* 108, 1400–1415; 1416. doi: 10.1038/ajg.2013.218

Thapa, R., Iqbal, Z., Garikipati, A., Siefkas, A., Hoffman, J., Mao, Q., et al. (2022). Early Prediction of Severe Acute Pancreatitis Using Machine Learning. *Pancreatology* 22, 43–50. doi: 10.1016/j.pan.2021.10.003

Trikudanathan, G., Wolbrink, D. R. J., Van Santvoort, H. C., Mallery, S., Freeman, M., and Besselink, M. G. (2019). Current Concepts in Severe Acute and Necrotizing Pancreatitis: An Evidence-Based Approach. *Gastroenterology* 156, 1994–2007.e1993. doi: 10.1053/j.gastro.2019.01.269

Valverde-Lopez, F., Matas-Cobos, A. M., Alegria-Motte, C., Jimenez-Rosales, R., Ubeda-Munoz, M., and Redondo-Cerezo, E. (2017). BISAP, RANSON, Lactate and Others Biomarkers in Prediction of Severe Acute Pancreatitis in a European Cohort. *J. Gastroenterol. Hepatol.* 32, 1649–1656. doi: 10.1111/jgh.13763

Viasus, D., Garcia-Vidal, C., Simonetti, A., Manresa, F., Dorca, J., Gudiol, F., et al. (2013). Prognostic Value of Serum Albumin Levels in Hospitalized Adults With Community-Acquired Pneumonia. *J. Infect.* 66, 415–423. doi: 10.1016/j.jinf.2012.12.007

Wan, J., Shu, W., He, W., Zhu, Y., Zeng, H., Liu, P., et al. (2019). Serum Creatinine Level and APACHE-II Score Within 24 H of Admission Are Effective for Predicting Persistent Organ Failure in Acute Pancreatitis. *Gastroenterol. Res. Pract.* 2019, 8201096. doi: 10.1155/2019/8201096

Weir, G., Lesser, P., Drop, L., Fischer, J., and Warshaw, A. (1975). The Hypocalcemia of Acute Pancreatitis. *Ann. Intern. Med.* 83, 185–189. doi: 10.7326/0003-4819-83-2-185

Wilkman, E., Kaukonen, K. M., Pettilä, V., Kuitunen, A., and Varpula, M. (2013). Early Hemodynamic Variables and Outcome in Severe Acute Pancreatitis: A Retrospective Single-Center Cohort Study. *Pancreas* 42, 272–278. doi: 10.1097/MPA.0b013e318264c9f7

Wu, B. U., Johannes, R. S., Sun, X., Tabak, Y., Conwell, D. L., and Banks, P. A. (2008). The Early Prediction of Mortality in Acute Pancreatitis: A Large Population-Based Study. *Gut* 57, 1698–1703. doi: 10.1136/gut.2008.152702

Xu, X., Ai, F., and Huang, M. (2020). Deceased Serum Bilirubin and Albumin Levels in the Assessment of Severity and Mortality in Patients With Acute Pancreatitis. *Int. J. Med. Sci.* 17, 2685–2695. doi: 10.7150/ijms.49606

Xu, F., Chen, X., Li, C., Liu, J., Qiu, Q., He, M., et al. (2021). Prediction of Multiple Organ Failure Complicated by Moderately Severe or Severe Acute Pancreatitis Based on Machine Learning: A Multicenter Cohort Study. *Mediators Inflamm.* 2021, 5525118. doi: 10.1155/2021/5525118

Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., et al. (2020). Study of Cardiovascular Disease Prediction Model Based on Random Forest in Eastern China. *Sci. Rep.* 10, 5245. doi: 10.1038/s41598-020-62133-5

Yan, G., Li, H., Bhetuwal, A., Mcclure, M. A., Li, Y., Yang, G., et al. (2021). Pleural Effusion Volume in Patients With Acute Pancreatitis: A Retrospective Study From Three Acute Pancreatitis Centers. *Ann. Med.* 53, 2003–2018. doi: 10.1080/07853890.2021.1998594

Yu, F., Wei, C., Deng, P., Peng, T., and Hu, X. (2021). Deep Exploration of Random Forest Model Boosts the Interpretability of Machine Learning Studies of Complicated Immune Responses and Lung Burden of Nanoparticles. *Sci. Adv.* 7:1–14. doi: 10.1126/sciadv.abf4130

Zhang, X., Zhu, C., Wu, D., and Jiang, X. (2010). Possible Role of Toll-Like Receptor 4 in Acute Pancreatitis. *Pancreas* 39, 819–824. doi: 10.1097/MPA.0b013e3181ca065c

Zhou, Y., Ge, Y. T., Shi, X. L., Wu, K. Y., Chen, W. W., Ding, Y. B., et al. (2022). Machine Learning Predictive Models for Acute Pancreatitis: A Systematic Review. *Int. J. Med. Inform.* 157, 104641. doi: 10.1016/j.ijmedinf.2021.104641

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.

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.