

RESEARCH

Open Access



Predictive model for early fluid requirement in severe acute pancreatitis

Ayijiang Jiamaliding^{1†}, Gaolei Sun^{1†}, Junhong Wang^{1†}, Minjie Duan² and Qingbian Ma^{1*}

Abstract

Background Severe acute pancreatitis (SAP) involves significant fluid loss, necessitating precise fluid therapy. However, standardized protocols are lacking. This study aimed to develop a machine learning-based prediction model for fluid requirement in SAP patients.

Methods We conducted a retrospective observational study of SAP patients admitted to Peking University Third Hospital (2016–2020) within 48 h of onset. Fluid requirement was quantified by 48-hour rehydration volume. Using the Lasso algorithm, we screened predictive variables and trained five models: Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGBoost), LightGBM, CatBoost, and multiple linear regression. Model performance was evaluated via mean absolute error (MAE), root mean square error (RMSE), R^2 , and fitting curves. The soft voting method was used to fuse the above five prediction models to improve the performance of the model. The optimal model was interpreted using SHapley Additive exPlanations (SHAP). To illustrate potential clinical use, ten randomly selected cases from the test set are presented.

Results Among 308 included patients, 90% were allocated to the training set. Sixteen key variables were selected for prediction. Among the five machine learning algorithms used to build prediction models, the MAE and RMSE values of the XGBoost algorithm were the smallest and the R^2 value was the closest to 1, which indicated that XGBoost was the best-performing model in our study. After model fusion, the model performance was further improved. SHAP analysis of the optimal model, XGBoost, revealed the relative importance of each predictor. To illustrate the model's practical application, we present the predictions for 10 randomly selected test-set cases. The differences between the predicted and actual fluid volumes in these individual cases ranged from 31.07 to 329.80 mL, serving as concrete examples of how the model could be used at the bedside.

Conclusion In this study, we developed the Fluid Requirement Predicting Model for SAP (FRPM-SAP), which can predict the specific amount of fluid loss in SAP patients. The predictive performance was good, demonstrating that the model has practical application for guiding clinicians in their assessment of 48-hour rehydration volume.

Keywords Severe acute pancreatitis, Fluid requirement, Machine learning

[†]Ayijiang Jiamaliding, Gaolei Sun and Junhong Wang contributed equally to this work.

*Correspondence:
Qingbian Ma
maqingbian@126.com

¹Department of Emergency, Peking University Third Hospital, Beijing 100191, China

²Chevidence Lab of Child and Adolescent Health, National Clinical Research Center for Child Health and Disorders, Children's Hospital of Chongqing Medical University, Chongqing 400014, China



Introduction

Acute pancreatitis (AP) is a common gastrointestinal disorder requiring hospitalization, with 15–20% of cases progressing to severe acute pancreatitis (SAP). SAP is life-threatening, associated with multiple organ failure and a mortality rate of 20–40% [1]. Early fluid resuscitation is critical in SAP management, as inadequate or excessive fluid therapy can lead to complications such as pulmonary edema or abdominal compartment syndrome [2–8]. However, current clinical practice lacks precise protocols [9, 10], which rely heavily on empirical judgment.

Existing predictive models for AP primarily focus on complications or mortality, neglecting early fluid resuscitation—a key determinant of patient outcomes [11]. While guidelines such as the Surviving Sepsis Campaign (2021) [9] and WSES (2019) [10] emphasize early goal-directed fluid therapy, they lack tools for individualized volume assessment. This gap limits clinicians' ability to tailor fluid resuscitation. To address this, we developed the Fluid Requirement Predicting Model for SAP (FRPM-SAP) via machine learning (ML). Unlike traditional statistical methods, ML captures nonlinear relationships among variables, enabling accurate prediction of fluid loss. Our model leverages 16 easily accessible clinical parameters obtained during initial emergency visits, offering a rapid, actionable tool for fluid management.

This study makes significant advancements by focusing on early fluid requirement in SAP, specifically by targeting the critical first 48 h where timely intervention has the greatest impact. Our approach combines machine learning precision through five advanced algorithms with SHAP interpretability to enhance clinical interpretability, effectively bridging the gap between complex ML outputs and practical decision-making. By integrating capabilities, the FRPM-SAP model aims to offer a practical tool for assisting personalized fluid resuscitation, ultimately reducing complications and improving patient outcomes.

Methods

Study design and participants

This single-center retrospective study enrolled consecutive SAP patients admitted to the Emergency Department of Peking University Third Hospital between January 2016 and December 2020. This study complied with the Declaration of Helsinki [12] and China's Personal Information Protection Law (PIPL) [13]. All the data were anonymized before analysis and stored on password-protected hospital servers. Access was restricted to study team members under confidentiality agreements.

Patients were included if they (1) met the 2012 Atlanta criteria for SAP diagnosis [14]; (2) received protocolized fluid therapy within 48 h of admission [7, 9, 15–17], defined as achieving lactate clearance $\geq 10\%$, heart

rate (HR) < 120 bpm, mean arterial pressure (MAP) 65–85 mmHg, and urine output >0.5 mL/kg/h; and (3) had complete 48-hour fluid records with key parameters (lactate, HR, MAP, and urine output) measured at admission and 48 h post-admission. Exclusion criteria were as follows: (1) age < 16 years; (2) pregnancy; (3) cardiac arrest on arrival; or (4) $>20\%$ missing data for key variables.

Patients were randomly split into training (90%, $n=277$) and test (10%, $n=31$) sets. The sample size met the requirement of 10–20 times the number of predictors (16 variables) for stable modeling. The model's performance was primarily evaluated through internal validation on the held-out test set ($n=31$). To demonstrate the model's potential for clinical application with concrete examples, we randomly selected 10 cases from the test set ($n=31$) for a detailed case presentation. This selection was intended solely for illustrative purposes and was not used to assess the model's performance.

Crucially, to ensure the prospective utility of the predictive model, a strict temporal distinction was maintained between predictor and outcome variables. All predictors were collected exclusively at the initial emergency department presentation, prior to the initiation of the 48-hour rehydration protocol and the assessment of its total volume. The outcome variable—the total 48-hour rehydration volume—was calculated retrospectively after this period. The parameters used solely for assessing adequate fluid therapy (lactate clearance, HR, MAP, urine output) at 48 h were not used as predictors, thus preventing any information leakage.

Data collection and preprocessing

A standard strategy was developed to avoid inconsistency. Data were anonymized and collected via Research Electronic Data Capture (REDCap), a platform provided by the Department of Emergency Medicine and the Clinical Epidemiology Research Centre of the Third Hospital of Peking University. Demographic data, clinical manifestations, laboratory examination data, and abdominal CT examination data were all manually uploaded into REDCap hospital information.

For cases with missing variables, a common method of filling missing values is to use the median or mean. These methods are relatively easy, but they can affect the distribution of the original data and increase the bias of the dataset itself. In our study, the missForest algorithm from the Python tool library `missingpy` was used to input continuous variables, and plurality was used to input discrete variables.

MissForest is a robust, non-parametric method that assumes the data are Missing at Random (MAR). We believe the MAR assumption is reasonable in our clinical context for the following reasons: First, the majority of

missingness occurred in specialized laboratory tests (e.g., cardiac markers, detailed lipid profiles), which were likely omitted based on the clinician's initial assessment rather than the patient's unobserved condition (e.g., a test was not ordered due to hospital protocol or lack of specific indication, not because the clinician intuitively knew the value was pathological). Second, our dataset is rich with strong predictors from demographics, vital signs, routine labs, and imaging. These observed variables are likely to capture the underlying reasons for the missingness of other variables, thus making the MAR assumption more plausible.

Descriptive statistics were performed using SPSS 20.0 (IBM Inc., Armonk, NY, USA). Normally distributed variables were expressed as the mean \pm standard deviation ($\bar{x} \pm s$), and abnormally distributed variables were expressed as the median and interquartile range [median (25th percentile, 75th percentile)]. Count data was expressed as frequencies and percentages [n(%)].

Prior to feature selection and model training, all continuous variables were standardized to a common scale to ensure that the model coefficients would be comparable and to improve the convergence of the machine learning algorithms. We applied the MinMaxScaler from the scikit-learn library to rescale the values of each continuous variable to a range between 0 and 1. This normalization process was applied to the dataset after imputation of missing values and before performing Lasso regression and training the machine learning models.

Feature selection and modeling

All subsequent analyses, including feature selection and model training, were performed using the preprocessed dataset (after imputation and standardization). In the fields of mathematics and statistics, screening parameters refers to the process of extracting a subset of useful features from the original data. This process can reduce the number of features, reduce noise, increase the speed of model training, strengthen the model's generalizability, reduce overfitting, and make the model more convenient to be applied in the clinic. In this study, we used the "glmnet" package in R (version 4.1–7.1) to perform The Least Absolute Shrinkage and Selection Operator (Lasso) analysis. Lasso regression is particularly suitable for high-dimensional data as it performs both variable selection and regularization by imposing an L1 penalty on the regression coefficients. This approach not only reduces the number of features but also minimizes the risk of overfitting by shrinking the coefficients of less important variables to zero. The optimal penalty parameter (λ) was determined through 10-fold cross-validation, a process which partitions the training data into 10 folds of roughly equal size. The model is fit 10 times, each time using 9 folds for training and the remaining 1 fold for

validation, to compute the mean cross-validated error for each λ value. The value of λ that resulted in the minimum mean cross-validated error was selected as optimal. This approach ensures that the feature selection is robust and generalizable.

Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), CatBoost and multiple linear regression machine learning methods were established. The methodology of these methods is described in the supplementary materials. The SHapley Additive explanation (SHAP) (version 0.39) method was used to explain the optimal model among the five models, providing a unified framework to explain individual predictions by quantifying the contribution of each feature. This approach aligns with recent advances in explainable AI for tree-based models [18], ensuring transparency in clinical decision support.

To further enhance predictive performance, we implemented a soft voting ensemble approach to integrate predictions from all five machine learning models. Model fusion is a technique that combines multiple base estimators to solve a problem of interest, often outperforming any single constituent model. Among various fusion methods, the voting method is most commonly used. Specifically, for regression tasks, it computes the final prediction as a weighted average of all models' outputs, rather than a simple average. This approach leverages the principle that a properly weighted combination of diverse models can reduce prediction variance and improve model robustness, which is particularly valuable in clinical prediction tasks where stability is crucial. In our implementation, each model's contribution to the final prediction was determined by its validation set performance, with higher-performing models receiving greater influence. The specific weight values were objectively derived to ensure the ensemble optimally balanced accuracy and diversity. Accordingly, XGBoost was assigned the highest weight due to its superior individual performance, followed by CatBoost and GBDT, whereas LightGBM and linear regression received proportionally lower weights on the basis of their relative MAE values. The ensemble was implemented via scikit-learn's VotingRegressor. This strategically weighted integration allowed the fusion model to effectively mitigate individual model weaknesses while preserving their collective predictive power.

Results

As shown in Fig. 1, 308 cases were included in the final analysis. Among the included patients, 65% were male, and the mean age was 53 years. Supplementary Table 1 in the supplementary materials details the demographics,

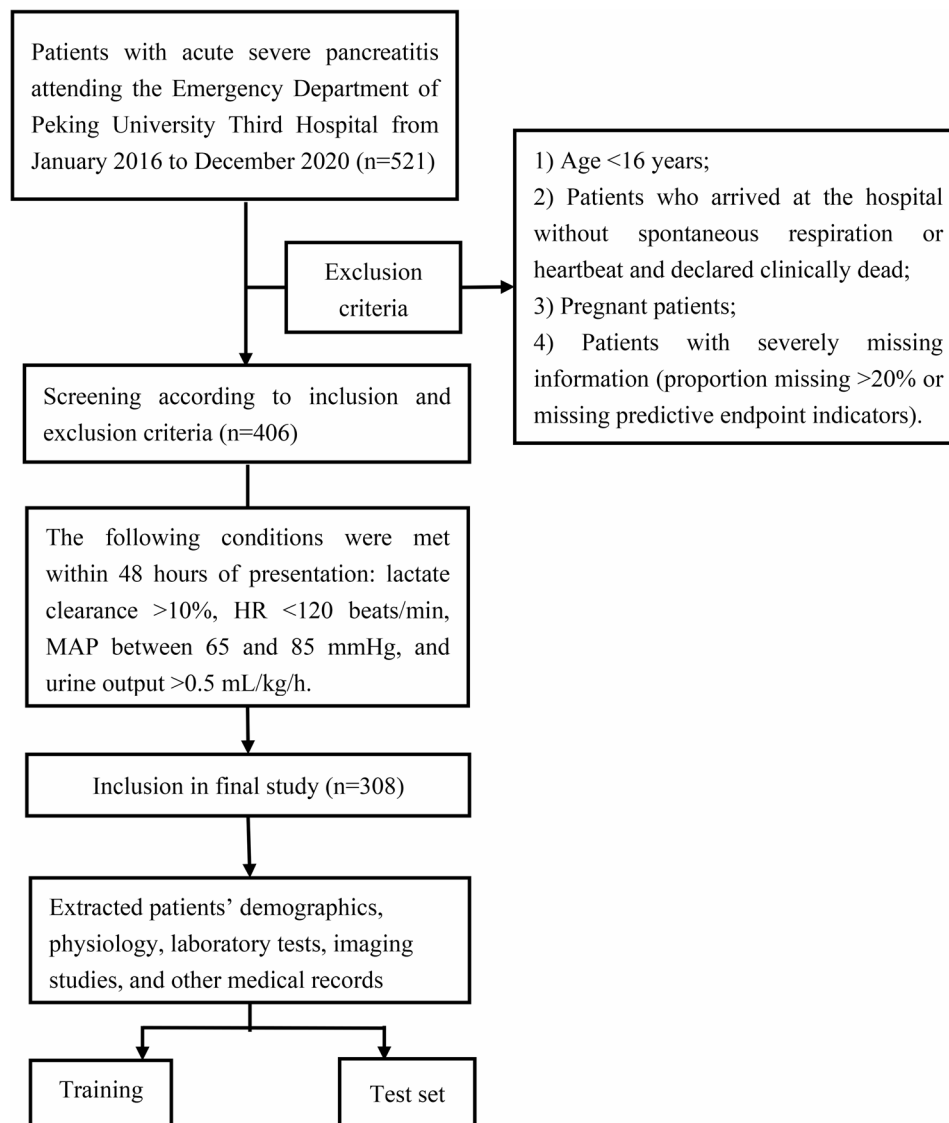


Fig. 1 Schematic diagram of the present study

physiological characteristics, and laboratory results of the studied population.

Results of parameter screening by applying Lasso

A total of 101 variables were collected and collated, 22 of which were excluded due to a large proportion of missing data. We applied Lasso regression to select predictive features from the remaining 79 parameters (Fig. 2). Cross-validation identified the optimal penalty parameter (λ), ensuring that the selected features contribute significantly to the prediction of fluid loss in SAP patients. At this value of λ , 16 variables were retained, each contributing significantly to the prediction of fluid loss. Figure 2 shows the cross-validation curve for Lasso regression, where the optimal penalty parameter (λ) was selected to minimize the mean squared error. The vertical axis

represents the mean squared error of the model, while the horizontal axis represents the value of $\log(\lambda)$. The upper horizontal axis indicates the number of non-zero coefficient features. When the number of features is 16, the mean squared error is the smallest, indicating the optimal balance between model complexity and predictive accuracy.

The importance ranking of the 16 variables in Table 1 shows that of the four physical signs, three were highly important and consistent with the results of clinical assessment, namely, the presence of cyanosis of the skin at the time of consultation, abdominal muscle tension, and upper abdominal tenderness. In terms of laboratory tests, the importance of serum calcium, CTSI, and arterial blood lactate values was greater and more consistent with the clinical findings. These features are clinically

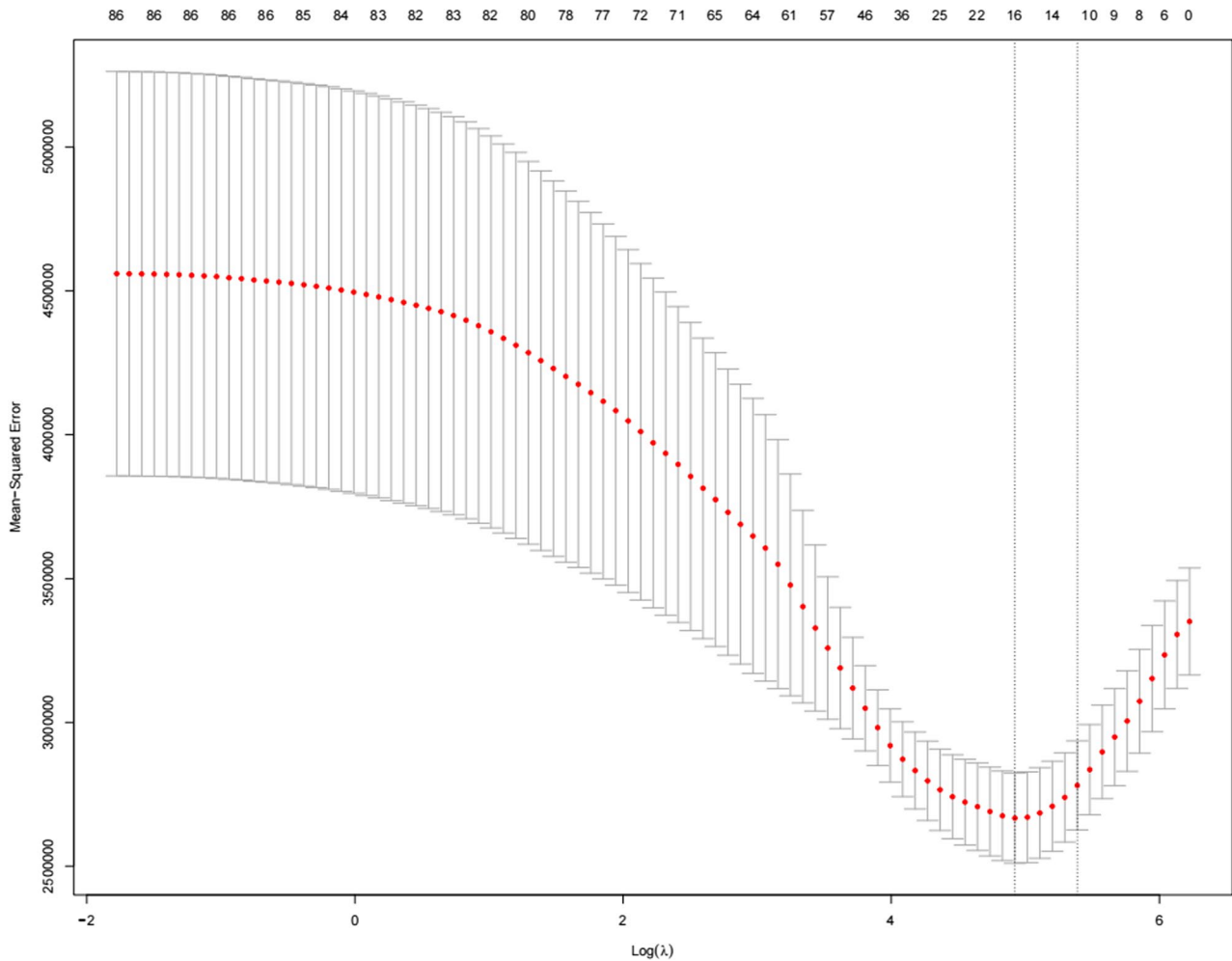


Fig. 2 Penalty strength cross-validation plot. The vertical axis represents the mean square error of the model, the lower horizontal axis represents the value of $\log(\lambda)$, and the upper horizontal axis represents the number of non-zero coefficient features. When the number of features is 16, the mean square error is the smallest

Table 1 Importance rankings of all features

Number	Variant	Coefficient
1	skin color	1207.88
2	abdominal muscle tension	408.59
3	upper abdominal tenderness	298.28
4	serum calcium	-271.43
5	respiratory rate	252.73
6	CTSI	221.21
7	diarrhea	220.81
8	abdominal distension	195.02
9	RBC	185.68
10	arterial blood lactate	130.57
11	SI	19.04
12	BE	-13.28
13	age	-2.43
14	HB	1.52
15	ALT(U/L)	0.90
16	AST(U/L)	0.07

relevant, as they reflect the severity of tissue hypoperfusion and systemic inflammation, which are key drivers of fluid imbalance in SAP patients. The feature importance performances of the four machine learning algorithms based on decision trees are shown in Fig.S1 (a-d) in the Supplementary materials.

Results of internal validation of the model

The prediction performance of all five machine learning algorithms is shown in Table 2. Among them, the XGBoost model had the smallest MAE(537.131) and RMSE(730.566) values, and the $R^2(0.743)$ was the closest to 1, indicating the best predictive performance. The CatBoost model closely followed, with an MAE of 556.977 and an RMSE of 760.650, whereas the GBDT and LightGBM models presented slightly greater errors. The multiple linear regression model performed the worst, with the highest MAE (734.843) and RMSE (1038.489) and the lowest R^2 (0.480).

Table 2 Predictive performance of the five machine learning models and the fusion model

Model	MAE	RMSE	R ²
XGBoost	537.131	730.566	0.743
CatBoost	556.977	760.650	0.721
GBDT	582.221	781.970	0.705
LightGBM	640.054	808.831	0.685
Linear Regression	734.843	1038.489	0.480
Ensemble model	496.917	701.654	0.763

To provide a comprehensive rationale for selecting XGBoost as the optimal model, we discuss the strengths and limitations of each algorithm below:

GBDT is an ensemble learning method that builds decision trees sequentially, with each tree correcting the errors of the previous tree. It is known for its robustness and ability to handle complex, nonlinear relationships in the data. However, GBDT can be computationally expensive and prone to overfitting if not properly regularized. In our study, GBDT performed well but was outperformed by XGBoost in terms of both the MAE and the RMSE.

XGBoost is an advanced implementation of gradient boosting that incorporates several optimizations, including regularization to prevent overfitting, parallel processing for faster training, and handling of missing values [19]. These features make XGBoost highly efficient and accurate, particularly for structured data. In our analysis, XGBoost achieved the smallest MAE and RMSE values, with R² values closest to 1, indicating the best predictive performance among the five models. Its ability to handle high-dimensional data and its robustness against overfitting make it the optimal choice for our predictive model.

The LightGBM is designed for efficiency and speed, particularly with large datasets. It uses a leaf-wise tree growth strategy, which can lead to faster training times than depthwise growth. However, this approach can also make LightGBM more prone to overfitting on smaller datasets. In our study, LightGBM performed reasonably well but was slightly less accurate than XGBoost, with higher MAE (640.054) and RMSE (808.831) values.

CatBoost is another gradient boosting algorithm that excels in handling categorical data and is robust to overfitting [20]. It uses ordered boosting and a novel algorithm for processing categorical features, which can improve model performance. However, CatBoost can be slower to train than XGBoost and LightGBM. In our analysis, CatBoost performed similarly to GBDT but was outperformed by XGBoost, with an MAE of 556.977 and RMSE of 760.650.

Multiple linear regression is a traditional statistical method that assumes a linear relationship between the independent and dependent variables. While it is simple and interpretable, it often fails to capture complex,

nonlinear relationships in the data. In our study, linear regression had the highest MAE (734.843) and RMSE (1038.489), and the lowest R² (0.480), indicating poor predictive performance compared with the other machine learning models.

In conclusion, while all five models demonstrated varying degrees of predictive ability, XGBoost emerged as the most accurate and reliable model for predicting fluid loss in SAP patients. Its combination of high predictive accuracy, robustness, and interpretability makes it a valuable tool for guiding fluid resuscitation strategies in clinical practice. To visualize the validation results, fitting curves were constructed for each model via the test set, as shown in Fig.S2 (a-e) in the Supplementary materials.

The five prediction models were fused via the soft voting method, with each model's contribution weighted according to its validation set performance. The resulting ensemble model achieved superior predictive accuracy, with MAE (496.917) and RMSE (701.654) values that were 7.5% and 4.0% lower, respectively, than those of best-performing individual model (XGBoost). The ensemble's R² (0.763) also showed a 2.7% improvement over XGBoost, demonstrating that the weighted combination effectively leveraged the strengths of each constituent model. This performance enhancement validates the soft voting approach's ability to mitigate individual model weaknesses while preserving their collective predictive power. Notably, the weight distribution, with XGBoost contributing most significantly (28%) and linear regression the least (5%), reflects the models' relative accuracies while maintaining beneficial diversity in the ensemble.

Results of model interpretation and application

In this study, the SHAP method was used to interpret the XGBoost model, which is the optimal model among the five single models. The SHAP values for each variable in all the training set samples were averaged to derive the importance value of that feature, as shown in Fig. 3(a). Base excess (BE) contributes the most to the model, followed by the shock index and lactate value. Figure 3(b) further visualizes the positive or negative relationship of each feature to the target value. The high feature value of BE at the time of patient visit has a negative Shapley value, showing that BE was negatively correlated with fluid loss. SI is another important feature that is negatively correlated with fluid loss. The lactate concentration was positively correlated with fluid loss. Similar interpretations can be obtained for the remaining variables. Fig. S3 and Fig.S4 in the Supplementary materials provide examples of the features involved in an individual prediction of the amount of fluid loss.

Supplementary Table 2 shows the 10 cases from the test set that were investigated. Taking Patient 10 as an example, it can be assumed that the amount of fluid loss of this

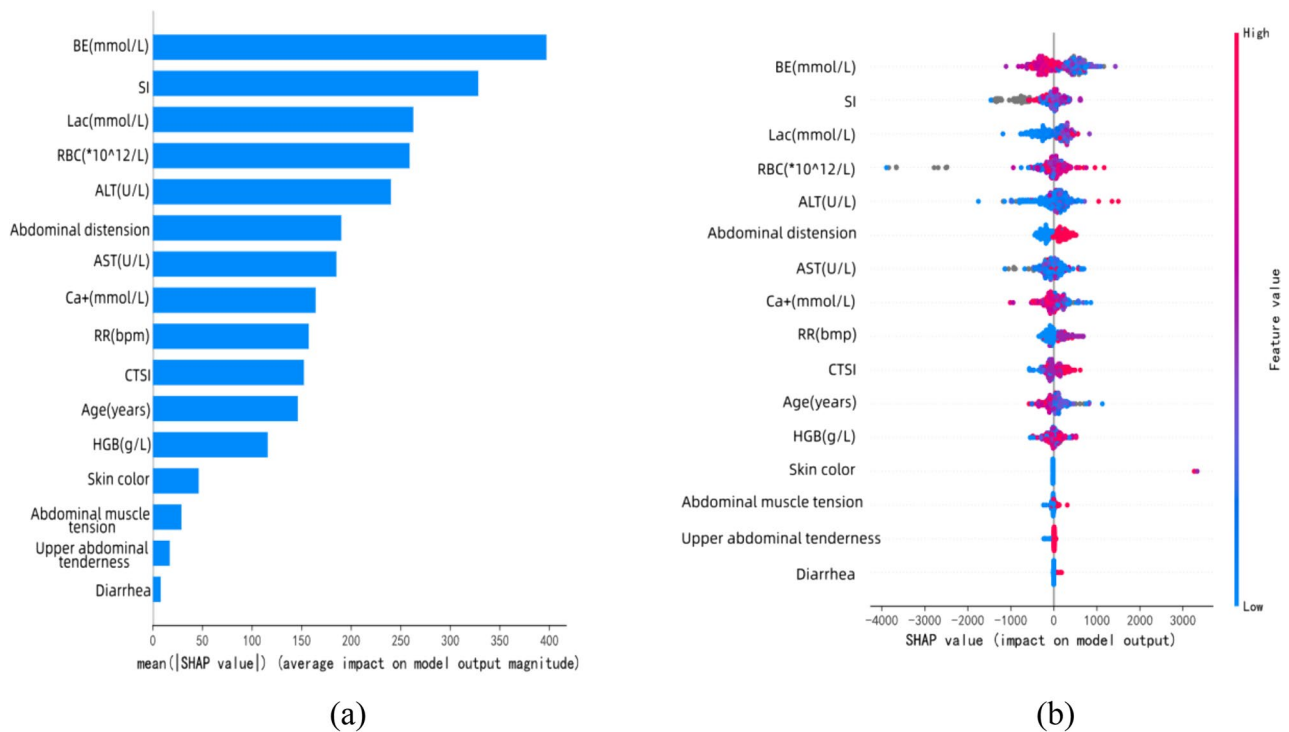


Fig. 3 **a** Mean SHAP values of variables in the prediction model. **b** Honeycomb graph presentation of SHAP value ranking. In **b**, each point on the graph is the Shapley value for a feature, the y-axis position of which is determined by the feature, in order of importance, and the x-axis position is determined by the Shapley value. The color indicates the feature value: the closer to red, the higher the eigenvalue is, and the closer to blue, the lower the eigenvalue is. The overlapping points float in the y-axis, making it is possible to understand the positive-negative distribution of Shapley values for each feature

patient was 6,268.93 mL, and the amount of fluid replenishment for the next 48 h after this patient's visit should have been this value. This means that the exact amount of fluid replenishment for 48 h can be determined by inputting the values of the above 16 variables into the computational model after the SAP patient's visit.

Discussion

Performance of the predictive model for measuring early fluid requirement in the SAP

SAP induces a characteristic systemic inflammatory response [1] that promotes rapid fluid shift into third spaces through increased vascular permeability. This pathophysiological basis directly supports our model's design focusing on early volume prediction, as accurate assessment of these losses within the initial 48 h is critical for guiding resuscitation. In this study, the FRPM-SAP was developed to assess fluid supplement volume in SAP patients. Based on the comparison of the five models' predictive performances, the XGBoost model was identified as the optimal model. The soft voting ensemble approach further enhances this performance by strategically combining all five models through accuracy-weighted integration. The optimal model, XGBoost, was subsequently interpreted via the SHAP tool, and its results are consistent with those of clinical observations

and related studies [8, 21–28], confirming the relevance of the selected parameters in predicting fluid requirement in pancreatitis patients. Finally, the model's application was demonstrated using 10 randomly selected clinical cases from the test set, providing a tangible illustration of its potential for real-world use.

Recent evidence has emphasized individualized fluid therapy for SAP [7, 8, 29]. Our model aligns with this need by providing an accurate estimate of fluid loss within the first 48 h of admission and assisting clinicians in tailoring fluid resuscitation strategies to individual patient needs, thereby reducing the risk of complications such as pulmonary edema or abdominal compartment syndrome. The model's reliance on easily accessible clinical parameters, such as arterial blood lactate, BE, and the shock index, makes it particularly suitable for use in the emergency department, where rapid decision-making is critical. Moreover, the interpretability of the model, enhanced by the SHAP method, allows clinicians to understand the contribution of each variable to the predicted fluid loss, increasing their confidence in the model's recommendations.

Although our model was developed using data from a single center, which may introduce selection bias and limit the generalizability of the findings, its reliance on widely recognized and routinely collected clinical

parameters suggests strong potential for broader applicability. Key predictors such as arterial blood lactate, BE, and the shock index reflect fundamental pathophysiological processes (e.g., tissue hypoperfusion and systemic inflammation) that are common across diverse SAP patient populations. Therefore, while external validation is necessary, the model's design based on universally relevant physiology supports its potential for use in other tertiary care settings. Furthermore, the practicality of the FRPM-SAP model extends to resource-limited environments. Its simplicity and dependence on clinical data that are readily available even in settings without advanced diagnostic tools make it particularly valuable. By enabling more precise and personalized fluid management at the critical early stages of SAP, the model has the potential to improve clinical outcomes across a wide spectrum of healthcare systems. Future studies should prioritize multicenter validation to confirm its generalizability and refine its adaptability to different clinical environments.

Compared with previous approaches for SAP fluid management, our FRPM-SAP demonstrates measurable advancements through machine learning integration. While traditional protocols such as those by de-Madaria [4] and Kuwabara [5] established volume-based resuscitation frameworks, they lack quantitative prediction models for individual patients. To our knowledge, our approach represents one of the first studies to provide a validated 48-hour fluid loss prediction model for SAP, incorporating several features such as (a) dynamic weighting of metabolic markers through SHAP-guided interpretation, (b) ensemble modeling that reduces variance compared with single-algorithm approaches, and (c) early prediction using admission parameters without requiring serial measurements.

Factors influencing early fluid requirement in patients with severe acute pancreatitis

The particular clinical working environment was taken into account in the selection of parameters. First, the parameters selected in this study are those that can be rapidly obtained in the emergency department, which allows the model to be applied in the early stages of patient consultation. Second, the universality of the model was considered: most of the indicators that are routinely collected in the emergency department were selected in this study, which makes the model applicable in hospitals of different levels.

The use of Lasso regression for feature selection not only reduced the dimensionality of our dataset but also enhanced the model's generalizability by eliminating redundant or irrelevant variables. This approach is particularly important in clinical settings, where overfitting can lead to models that perform well on training data but poorly in real-world applications. By retaining only the

most significant features, our model achieves a balance between complexity and predictive accuracy. The final set of 16 variables collectively captures the pathophysiological mechanisms underlying fluid loss in SAP. While our model focuses on SAP, its parameters are consistent with the general principles of sepsis resuscitation [9], suggesting broader applicability.

- 1) Arterial blood lactate: The early stage of AP leads to hypovolemia and microcirculatory disorders due to the systemic inflammatory response resulting in fluid loss in the third interstitial space, which in turn leads to inadequate tissue perfusion and hypoxia and ultimately to elevated lactate [24]. The lactate indicator contributes to the prediction of fluid loss in the models in this study, with a coefficient of 130.57. The higher the lactate concentration is, the greater the volume of fluid added. The strong predictive power of lactate in our model corroborates recent work by Zeng et al. [24], who demonstrated that dynamic lactate clearance rates outperform static values in the prediction of SAP outcomes.
- 2) BE: BE reflects the status of tissue perfusion, the level of cellular energy metabolism, and the functional status of each organ. When the effective circulating blood volume is reduced, tissue perfusion is insufficient, leading to a significant increase in the negative value of BE. This increase causes changes in the osmotic pressure inside and outside the cell membrane of the organs in vivo, leading to edema and apoptosis of the epithelial cells and triggering platelet-activating factor, intestinal bacterial endotoxin, and related metabolites to act on the tissues and organs together to cause an excessive inflammatory response, resulting in multiorgan failure [25]. The relationship between BE and the fluid loss measure revealed by the model in this study is compatible with clinical logic.
- 3) SI: Traditionally, HR and systolic blood pressure have been widely used to assess the hemodynamic status of patients upon arrival at the emergency department. In our study, a proportion of patients presented with volume deficits, such as rapid heart rate (123, 39.9%) and low blood pressure (80, 30%), at disease onset. However, even in critically ill patients, these parameters may be normal at the onset of illness and may lead to delayed intervention. Patients with an SI >1 are more likely to respond to fluid resuscitation [26]. Therefore, the SI can be used as an alternative to traditional vital signs of HR and systolic blood pressure to assess the hemodynamic stability of patients, especially when assessing their fluid volume responsiveness is assessed. The results

of this study, analyzed by the Lasso method, suggest an importance coefficient of 19.04 for SI, suggesting that SI is associated with fluid loss in SAP patients.

- 4) Respiratory rate: Patients with SAP often suffer from respiratory distress due to pain, interstitial lung edema, pleural effusion, or other reasons, and their respiratory rate is accelerated. As it increases, the amount of fluid loss through respiration increases accordingly; that is, more fluid replacement is needed.
- 5) Serum calcium: A large amount of calcium ions flow into alveolar cells in SAP, leading to pancreatic necrosis and severe inflammatory response syndrome, while pancreatitis produces a large amount of lipase to decompose fats, generating a large amount of fatty acids; these acids react with calcium ions in the body, thus leading to the loss of calcium ions in the bloodstream, resulting in a decrease in the level of serum calcium [27]. Calcium ions can be used as an indicator to reflect the severity of disease in AP patients, thus indirectly predicting fluid requirement. The prediction model in this study also suggested that the lower the serum calcium is, the greater the fluid requirement.
- 6) Age: The results of this study suggest a negative correlation between age and 48-hour fluid supplement volume. Elderly patients tend to have underlying cardiopulmonary diseases that increase their susceptibility to complications such as pulmonary edema and heart failure due to fluid overload.
- 7) Hct and HB: Previous studies have shown that Hct is correlated with volume, but in this study, Lasso analysis suggested that the erythrocyte count and hemoglobin level, especially the erythrocyte count, which has an importance coefficient of 185.68, are correlated with fluid loss in SAP patients. The possible reasons for these results were considered. First, the rapid activation of the inflammatory response, which occurs in the early stages of disease in SAP patients, and high oxidative stress can reduce erythrocyte survival, affect erythrocyte lifespan and deformability and reduce RBC and Hb in peripheral blood [28]. Second, the range of variability of Hct is smaller than that of the erythrocyte count and hemoglobin levels. Therefore, the erythrocyte count and hemoglobin concentration were included in this study.
- 8) Physical signs: Abdominal distension, abdominal muscle tension, epigastric tenderness, and skin color were included in this study. The first three can intuitively reflect the severity of AP, and the last can reflect tissue perfusion and the degree of volume insufficiency. These parameters can indirectly

reflect the amount of fluid inflammatory exudation and fluid imbalance in the third interstitial space. However, it is also important to recognize that these parameters are highly subjective and depend to a large extent on the clinician's judgment, which may result in some bias.

- 9) CTSI: Imaging examinations, such as abdominal ultrasound, abdominal CT, ERCP, magnetic resonance imaging, and magnetic resonance cholangiopancreatography, play important roles in the diagnosis and treatment of pancreatic diseases. Among the imaging techniques, CT has high sensitivity and accuracy in observing the morphology of the pancreas and the surrounding diseased tissues; it can more intuitively reflect the presence and degree of hemorrhage or necrosis of pancreatic tissues and is the preferred imaging method for assessing the severity and complications of AP. In 1990, Balthazar proposed the CT Severity Index (CTSI), which correlates the original grading system with pancreatic necrosis and the extent of necrosis; the CTSI was the first clinically recognized scoring system for assessing AP imaging [30]. The more severe the degree of pancreatic exudation and necrosis is, the greater the degree of fluid loss, and a greater rehydration volume is needed. In this study, the CTSI was used to predict the amount of fluid loss in SAP patients, which is consistent with the results of previous studies.

Necessity of the predictive model

Clinicians measure the amount of fluid replacement for SAP patients on the basis of experience. In this study, 98 (24.1%) of the 308 patients did not meet the criteria for appropriate fluid therapy. Among these 98 patients, 56 (13.8%) experienced different degrees of acute pulmonary edema, acute heart failure, and other complications due to excessive fluid replacement, while 42 (10.3%) patients presented with narrowing of the inferior vena cava and elevated lactic acid, indicating that they did not have enough fluid replacement. Therefore, there is an urgent need for a general and accurate predictive model to guide the rehydration regimens of SAP patients in clinical practice, which is particularly important for junior physicians.

Compared with traditional models, machine learning models are more capable of handling multiple redundant and interrelated variables, and are increasingly being used in clinical research. Machine learning models were applied in this study to construct a predictive model for the early assessment of fluid loss in SAP patients at the time of emergency primary diagnosis. In this scenario, the physician has little information about the patient,

including a preliminary history, physical examination, and simple laboratory results; however, the physician needs to make a rapid judgment of the overall fluid loss and then decide the fluid supplement volume. Establishing such a prediction model requires the consideration of many factors, not only indicators that can assess fluid capacity but also some reference indicators specific to pancreatitis. SAP patients in the emergency department often have characteristics such as advanced age, comorbid chronic diseases, and a history of repeated visits. With reference to previous literature and actual clinical experience, this study collected a large amount of clinical data and tried to employ all the data to comprehensively assess fluid loss in SAP patients. However, increasing the number of variables only led to an increase in the “noise” of the model, resulting in the model being overfit and unable to be extrapolated, which made the model more difficult to use. Therefore, the accessibility of the variables was taken into account when selecting variables for this study; specifically, the model variables were restricted to information that could be obtained within a few hours of the patient’s visit to the clinic.

Despite the great success of machine learning methods in the medical field, few models can be practically applied in the clinical setting, largely due to the lack of explanatory properties of machine learning algorithms, which are considered “black boxes.” A medical aid diagnostic system must be understandable and interpretable. Ideally, it should be able to explain the complete application methodology and logic of the model to all relevant parties to gain their trust. In this study, the SHAP tool was used to explain the contribution of each feature to the predicted value in each patient sample. This interpretability is critical for clinician adoption, as it bridges the gap between complex machine learning outputs and actionable clinical insights [18, 31]. Finally, the specific application methodology is demonstrated using data from 10 patients, which increased transparency and facilitated clinicians’ understanding and application.

Limitations of this study

This study has several limitations. First, this was a single-center retrospective study, and the predictive results of the model relied on the selection of cases in the training set. Therefore, the variables are closely related to the rehydration habits of the hospital from which the cases originated. The single-center nature of the dataset may introduce selection bias and limit the generalizability of the findings. The patient population and clinical practices at our institution may not fully represent those in other hospitals or regions, potentially affecting the model’s applicability in diverse settings. Future studies should incorporate multicenter data and conduct external validations in diverse patient populations. This would

enhance the reliability and generalizability of the model, ensuring its effectiveness across different healthcare environments. Second, the sample size of the independent test set ($n=31$) is relatively small, which may affect the precision of the performance estimates. We employed strategies to enhance model robustness, including regularized feature selection via Lasso regression and the use of a soft-voting ensemble, which are known to mitigate overfitting. Nevertheless, the performance metrics obtained from this test set should be interpreted with caution, and external validation in a larger cohort is essential to confirm the model’s generalizability. Third, and most importantly, our outcome variable was the actual fluid volume administered, not an independently verified optimal volume. While the model accurately predicts what was done in our clinical practice, and this likely correlates with true physiological need, it does not necessarily predict the ideal fluid volume. A subset of patients (31.8%) did not meet the optimal fluid therapy endpoints, indicating that clinical practice itself is imperfect. Therefore, the model encapsulates ‘collective clinical wisdom’ which includes both correct and potentially suboptimal decisions. It serves as a data-driven starting point for resuscitation, which should then be fine-tuned based on ongoing clinical assessment. Future research should aim to define and predict a truly optimal fluid volume based on hard outcomes. Furthermore, although the missForest algorithm was effective in imputing missing values, it assumes that the missing data mechanism is MAR, which may not always hold true in clinical settings. If the data are not MAR, the imputed values may introduce bias. Future studies could explore alternative imputation methods, such as multiple imputation or deep learning-based approaches, to further improve the robustness of the predictive models. Finally, lactate clearance, heart rate, mean arterial pressure, and urine output were selected as criteria for appropriate fluid therapy in this study; however, SAP is a disease that requires comprehensive treatment, and many factors influencing these metrics do not depend entirely on fluid resuscitation therapy. Thus, there may be some systematic bias. Future studies should consider longitudinal data collection to capture the dynamic changes in patient conditions and further refine the predictive model.

Conclusions

For SAP patients, we developed FRPM-SAP using easily accessible parameters obtained at emergency department presentation. The use of Lasso regression for feature selection ensured that only the most significant variables were retained, minimizing the risk of overfitting and enhancing the model’s generalizability. Among five machine learning methods applied, the XGBoost model emerged as the optimal model, and its performance

was further improved through a soft voting ensemble. Enhanced interpretability, provided by SHAP analysis, offers insights into the contribution of key predictors. This study addresses a critical gap in SAP management by providing a practical, data-driven tool for guiding fluid resuscitation. Future plans include multicenter validation and the development of a clinician-friendly website for real-time application, aiming to confirm the model's reliability and facilitate its integration into clinical practice.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12876-025-04465-y>.

Supplementary Material 1.

Acknowledgements

Corresponding Author: Qingbian Ma (maqingbian@126.com), Department of Peking University Third Hospital, No. 49 Huayuan North Road, Haidian District, Beijing, China.

Authors' contributions

Ayijiang Jiamaliding contributed to the conception and design of the study, collection and analysis of the data and drafting of the manuscript. Sun GL participated in data collection. Duan MJ contributed to the statistical analysis and interpretation of the data. Wang JH and Ma QB involved in the study design and revision of the manuscript.

Funding

No funding was obtained for this manuscript.

Data availability

The datasets generated and analyzed during this study are not publicly available due to patient privacy concerns but may be available from the corresponding author upon reasonable request and with permission from the institutional ethics committee.

Declarations

Ethics approval and consent to participate

Ethics approval was given by the Biomedical Ethics Committee of Peking University prior to the study's commencement (IRB00006761-M2021163), and informed consent was waived because of retrospective anonymized data analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 May 2025 / Accepted: 7 November 2025

Published online: 22 November 2025

References

- Boxhoorn L, et al. Acute pancreatitis. *Lancet*. 2020;396(10252):726–34.
- De Waele JJ, Leppaniemi AK. Intra-abdominal hypertension in acute pancreatitis. *World J Surg*. 2009;33(6):1128–33.
- Duchesne JC, Balogh KL, Malbrain ZJ. Role of permissive hypotension, hypertonic resuscitation and the global increased permeability syndrome in patients with severe hemorrhage: adjuncts to damage control resuscitation to prevent intra-abdominal hypertension. *Anaesthesiol Intensive Ther*. 2015;47(2):143–55.
- de-Madaria E, Sánchez-Payá S-SG, Lopez-Font J, Martínez I, Gómez-Escolar J, Sempere L, Sánchez-Fortún L, Pérez-Mateo C. Influence of fluid therapy on the prognosis of acute pancreatitis: a prospective cohort study. *Am J Gastroenterol*. 2011. Oct;106(10):1843–50.
- Kuwabara K, et al. Early crystalloid fluid volume management in acute pancreatitis: association with mortality and organ failure. *Pancreatol*. 2011;11(3):351–61.
- Schepers NJ, et al. Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut*. 2019;68(6):1044–51.
- Crosignani A, et al. Intravenous fluid therapy in patients with severe acute pancreatitis admitted to the intensive care unit: a narrative review. *Ann Intensive Care*. 2022;12(1):98.
- Machicado JD, Papachristou GI. Intravenous fluid resuscitation in the management of acute pancreatitis. *Curr Opin Gastroenterol*. 2020;36(5):409–16.
- Evans L, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. 2021;47(11):1181–247.
- Leppaniemi A, et al. 2019 WSES guidelines for the management of severe acute pancreatitis. *World J Emerg Surg*. 2019;14:27.
- Sabrie N et al. *Performance of Clinical Risk Prediction Models for Post-ERCP Pancreatitis: A Systematic Review*. *Pancreas*, 2025.
- Association WM. *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects, in 9th revision*. 2023.
- Congress SC. o.t.N.P.s., *Personal Information Protection Law of the People's Republic of China (PIPL)*. 2021.
- Banks PA, et al. Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013;62(1):102–11.
- Hirota M, et al. Fundamental and intensive care of acute pancreatitis. *J Hepatobiliary Pancreat Sci*. 2010;17(1):45–52.
- Nassar TI. AKI associated with acute pancreatitis. *Clin J Am Soc Nephrol*. 2019. Jul;5(7):1106–15.
- Nguyen HB, et al. Early lactate clearance is associated with improved outcome in severe sepsis and septic shock*. *Crit Care Med*. 2004;32(8):1637–42.
- Lundberg SM, et al. From local explanations to global Understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56–67.
- Chen T, Guestrin C. *XGBoost: A Scalable Tree Boosting System, in 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. San Francisco, California, USA.
- L P et al. *CatBoost: unbiased boosting with categorical features, in 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*. 2018. Montreal, Canada.
- Vincent JL, et al. The value of blood lactate kinetics in critically ill patients: a systematic review. *Crit Care*. 2016;20(1):257.
- Nichol A, et al. Dynamic lactate indices as predictors of outcome in critically ill patients. *Crit Care*. 2011;15(5):R242.
- Maraví Poma E, Petrov ZOF, Navarro Soto MS, Laplaza Santos S, Morales Alava C, Darnell Martín F, Gorraiz López A, Bolado Concejo B, Casi Villarroya F, Aizcorbe Garralda M, Albeniz Arbizu M. Sánchez-Izquierdo Riera JA, Tirapu León JP, Bordejé Laguna L, López camps V, Marcos Neira P, regidor Sanz E, Jiménez Mendioroz F., *Recommendations for intensive care management of acute pancreatitis*. *Med Intensiva*. 2013. Apr;37(3):163–79. English, Spanish.
- Zeng J, et al. Prognostic value of arterial lactate metabolic clearance rate in moderate and severe acute pancreatitis. *Dis Markers*. 2022;2022:p9233199.
- Hamada S, Masamune A, Shimosegawa T. Management of acute pancreatitis in japan: analysis of nationwide epidemiological survey. *World J Gastroenterol*. 2016;22(28):6335–44.
- Sam A, et al. The shock index and the simplified PESI for identification of low-risk patients with acute pulmonary embolism. *Eur Respir J*. 2011;37(4):762–6.
- Gutierrez-Jimenez AA, Castro-Jimenez E, Lagunes-Cordoba R. Total serum calcium and corrected calcium as severity predictors in acute pancreatitis. *Rev Gastroenterol Mex*. 2014;79(1):13–21.
- Senol K, et al. Red cell distribution width as a predictor of mortality in acute pancreatitis. *Am J Emerg Med*. 2013;31(4):687–9.
- Crosignani A et al. *Intravenous fluid therapy in patients with severe acute pancreatitis admitted to the intensive care unit: a narrative review (vol 12, 98, 2022)*. *Annals of Intensive Care*, 2023. 13(1).

30. Yoshihisa T, Hiroyoshi NT. Early diagnosis of pancreatic necrosis based on perfusion CT To predict the severity of acute pancreatitis. *J Gastroenterol.* 2017;52(1):1–10.
31. Kelly CJ et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*, 2019. 17(1).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.