RESEARCH

Interpretable prediction of 30-day mortality in patients with acute pancreatitis based on machine learning and SHAP

Xiaojing Li[1†], Yueqin Tian[2†], Shuangmei Li[1], Haidong Wu[1*] and Tong Wang[1*]

Abstract

Background Severe acute pancreatitis (SAP) can be fatal if left unrecognized and untreated. The purpose was to develop a machine learning (ML) model for predicting the 30-day all-cause mortality risk in SAP patients and to explain the most important predictors.

Methods This research utilized six ML methods, including logistic regression (LR), k-nearest neighbors(KNN), support vector machines (SVM), naive Bayes (NB), random forests(RF), and extreme gradient boosting(XGBoost), to construct six predictive models for SAP. An extensive evaluation was conducted to determine the most effective model and then the Shapley Additive exPlanations (SHAP) method was applied to visualize key variables. Utilizing the optimized model, stratified predictions were made for patients with SAP. Further, the study employed multivariable Cox regression analysis and Kaplan-Meier survival curves, along with subgroup analysis, to explore the relationship between the machine learning-based score and 30-day mortality.

Results Through LASSO regression and recursive feature elimination (RFE), 25 optimal feature variables are selected. The XGBoost model performed best, with an area under the curve (AUC) of 0.881, a sensitivity of 0.5714, a specificity of 0.9651 and an F1 score of 0.64. The first six most important feature variables were the use of vasopressor, high Charlson comorbidity index, low blood oxygen saturation, history of malignant tumor, hyperglycemia and high APSIII score. Based on the optimal threshold of 0.62, patients were divided into high and low-risk groups, and the 30-day survival rate in the high-risk group decreased significantly. COX regression analysis further confirmed the positive correlation between high-risk scores and 30-day mortality. In the subgroup analysis, the model showed good risk stratification ability in patients with different gender, renal replacement therapy and with or without a history of malignant tumor, but it was not effective in predicting peripheral vascular disease.

Conclusions the XGBoost model effectively predicts the severity of SAP, serving as a valuable tool for clinicians to identify SAP early.

[†]Xiaojing Li and Yueqin Tian these authors contributed equally.

*Correspondence:
Haidong Wu
wywhd@163.com
Tong Wang
tongwang316@163.com

Full list of author information is available at the end of the article

Introduction

Acute pancreatitis (AP) is an inflammatory disease of the pancreas whose incidence and hospitalization rates have increased in recent years, affecting more than 3 million U.S. patients annually [1]. Variables that influence disease severity in AP people include comorbidities and demographic variables such as older age, type 2 diabetes [2], cardiovascular disease, kidney disease [3] and obesity [2, 4]. According to statistics, nearly 25% of AP patients develop SAP due to serious complications and have to be transferred to the intensive care unit (ICU) for treatment and the mortality rate of SAP patients is as high as 30% [5–7]. This imposes a huge health and economic burden on patients and society. Therefore, early risk assessment and timely treatment of patients with AP are very important to improve their clinical outcomes.

Many researchers have tried to evaluate and predict the severity and clinical prognosis of AP with some laboratory tests, scoring models, predictive models, etc., but they have different limitations. For example, C-reactive protein (CRP) levels were significantly higher in SAP patients 48 h later, but less accurate at admission [8]. The Acute Physiological and Chronic Health Assessment (APACHE II) score was primarily designed to assess critically ill patients rather than AP patients [9]. The Acute Pancreatitis Severity Bedside Index (BISAP) was used primarily to identify the severity of AP and its mortality, while the Harmless Acute Pancreatitis Score (HAPS) was more sensitive and accurate in identifying mild AP [10, 11]. In addition, the CT score is often used to assess the severity of AP, but it mainly focuses on the qualitative assessment of local injuries in and around the pancreas, ignoring important clinical symptoms, signs and biochemical indicators [12]. A study of the accuracy of existing clinical scoring models by Mounzer et al., suggests that existing methods may not meet the need to accurately predict SAP risk [13]. Therefore, it is necessary to develop a new model with high accuracy to predict SAP risk.

Currently, machine learning (ML) algorithms are increasingly being used to solve medical problems, building models based on training data sets that can improve risk prediction [14, 15] and drug interactions [16, 17] for a variety of diseases, including AP. Compared with traditional logistic regression and linear regression, ML models produce more stable predictions due to their advantages of higher-order nonlinear interactions. Most previous studies have used ML models to predict the severity of AP and fewer focus on predicting clinical outcomes of ICU patients. SHAP can explain the contribution of variables in an ML model to predict risk, making up for the unexplainable shortcomings of ML [18]. This study aims to develop a prognostic model for SAP patients by combining multiple ML algorithms and SHAP values to provide a powerful tool for clinical decision-making. To the best of our knowledge, this is the first study to apply explicable ML to predict clinical outcomes for SAP patients in the ICU.

Methods

Study population

This study utilized the Medical Information Mart for Intensive Care (MIMIC)-IV database for analysis. The database is a public resource that compiles information on patients hospitalized in the ICU of Beth Israel Deaconess Medical Center during 2001–2012. Access to the database was granted by the Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center, and the collection of the original data was obtained with consent. Patient information included in the MIMIC-IV database is anonymous; accordingly, informed consent was not required.

This study included AP patients who met the International Classification of Diseases, Ninth Revision (ICD-9) code of 577.0, were over 18 years of age, and had an ICU stay of more than 24 h. If a single patient had multiple ICU admission records, only data related to the first ICU admission were analyzed.

The study data was extracted from the raw data using Structured Query Language (SQL) with DataGlip (v 2021.2.1) and further processed in R (v 4.4.0, the R Foundation for Statistical Computing) for retrieval of subject information from the database. Baseline characteristics within 24 h of hospital entry were captured.

Data collection

The data analyzed in this study encompassed demographic characteristics, including age, sex, and race, along with vital signs recorded within the first 24 h of admission. These vital signs comprised temperature, heart rate, respiratory rate, blood pressure, and oxygen saturation (SpO2), with the average, minimum, and maximum values of each parameter being documented from multiple measurements within this time frame. Laboratory results were also examined, specifically white blood cell (WBC) count, hemoglobin, platelet count, serum creatinine, albumin, bilirubin, calcium, potassium, and lactic acid levels, recording their average, minimum, and maximum values over 24 h. Additionally, various clinical scores were noted, including SOFA (Sequential Organ Failure Assessment), SIRS (Systemic Inflammatory Response Syndrome), SAPS III (Simplified Acute Physiology Score

III), OASIS (Oxford Acute Severity of Illness Score), and the Charlson co-morbidity index. Further data collection encompassed the presence of septicemia, myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic lung disease, rheumatic disease, peptic ulcer disease, diabetes, liver disease, paraplegia, malignant tumors, metastatic solid tumors, acute respiratory distress syndrome (ARDS), and acute kidney injury (AKI) stage. The primary endpoint was all-cause mortality within 30 days.

In this modeling analysis, categorical variables such as gender, past medical history, and treatment modalities were transformed into binary numeric formats. Gender was coded as 1 for males and 0 for females. For past medical history, conditions including Myocardial infarction, Congestive Heart Failure, and Peripheral Vascular Disease were coded as 1 if present, and 0 otherwise. Treatment variables involving the use of pressor drugs, diuretics, sedatives, and Continuous Renal Replacement Therapy (CRRT) were similarly coded as 1 when utilized and 0 when not. This binary coding method facilitates the effective integration of categorical variables into statistical analyses.

Data processing
Missing values are frequently encountered in MIMIC-IV databases. When the proportion of missing data for a variable exceeds 30%, that variable is excluded from further analysis. Conversely, if the proportion of missing data is less than 30%, the "mice" package (version 4.1.2) [19] implemented in R is employed to perform multiple imputations, thereby minimizing bias. Specifically, variables with a normal distribution are imputed using mean interpolation, while those with non-normal distributions are addressed through median interpolation.

Feature selection
First, we used RFE based on five-fold cross-validation to select features from the train set [20]. In the study, the "mlbench" (v 2.1-5) and "caret" (v 6.0–94) packages within the R programming environment were utilized to perform feature selection on training datasets via RFE with five-fold cross-validation. This iterative RFE method builds a model to identify and remove the most significant features, then reassesses the remaining features until all have been evaluated, aiming to identify an optimal feature subset. The five-fold cross-validation ensures that each subset of the original dataset is used once as validation data, facilitating robust model training. Performance metrics such as accuracy are calculated in each iteration, following the removal of the least important features, to evaluate the efficacy of the reduced feature set. Through this methodical process, the most effective subset of

characteristics is ascertained, enhancing the predictive accuracy of the modeling approach.

Besides, in this study, we also employed the LASSO method for variable selection with the "glmnet" package (v 4.1-8). LASSO regression is a shrinkage estimation method used to address multicollinearity between covariates. When multiple correlated predictors are present, LASSO selects one and ignores others or sets some regression coefficients to zero. It is worth pointing out that the λ value is determined when the cross-validation error is within one standard error (SE) of its minimum because LASSO regression uses cross-validation to select the λ value based on the 1-SE criterion. We obtained a subset of features selected by LASSO.

Finally, we take the intersection of the two selection results to obtain the final subset of features.

Model establishment and evaluation
499 patients were grouped into a train and test set in a 4:1 ratio by stratified random sampling. The training set was preprocessed using a synthetic minority oversampling technique combined with an edited nearest neighbor (SMOTE+ENN) technique to balance the positive and negative classes [21]. This preprocessing was executed using the "smotefamily" package (v 1.4.0), which includes SMOTE in R. And then, based on the training dataset, we established six ML models, including LR, KNN using the "kknn" package(v 1.3.1), SVM with the "e1071" package (v 1.7–14), NB also through "e1071" package (v 1.7–14), RF using the" randomForest" package (v 4.7–1.1) and XGBoost utilizing the "xgboost" package (v 1.7.7.1) for predicting 30-day all-cause mortality in SAP patients [22]. The hyperparameters of these ML models were optimized using the quintuple cross-validation method provided by the Grid Search algorithm implemented via the "caret" package (v 6.0–94) in R.

Subsequently, we evaluated and compared the performance of each model in the test set. To avoid bias and overfitting and obtain more stable predictive performance, we repeated these ML methods 100 times with different random seeds and computed the average performance over these 100 repeats [23]. Finally, multiple indicators including AUC, sensitivity, specificity and F1 score were comprehensively evaluated, and the "pROC" package (version 1.18.5) was utilized to compute the AUC.

Model explanation
In this study, we employed SHAP as a method of interpretability to enhance the transparency of our predictive model. SHAP is recognized for its post-hoc analytical capacity to quantify the impact of individual features on the output of the model both individually and collectively, thereby clarifying the model's operational mechanisms [18, 24]. Specifically, it calculates the Shapley value

for each attribute of a data point using specific algorithms, indicating that feature contributions are additive. This approach facilitates comprehensive explanations of how each feature influences the predictive accuracy and likelihood in each data set.

Unlike traditional Feature Importance metrics commonly associated with many machine learning models, SHAP analysis offers greater statistical depth and interpretability, as evidenced by several prior studies [25, 26]. Therefore, to elucidate the decision-making processes underlying our model, we implemented the SHAP methodology. In this investigation, the "xgboost" package (v 1.7.7.1) alongside SHAP analysis was utilized to ascertain the critical predictors of 30-day all-cause mortality in patients with acute pancreatitis. This approach effectively highlighted the most impactful variables within the model.

The calculation formula of SHAP is as follows:

$$\varphi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (f(S \cup \{i\}) - f(S))$$

where:

- $\varphi_i(f)$ is the SHAP value for feature i, quantifying its contribution to the prediction.
- N is the set of all features included in the model.
- S is any subset of the features that does not include feature i.
- $f(S)$ is the model output utilizing the features in subset S.
- $f(S \cup \{i\})$ is the model output when feature i is added to the subset S.
- $|S|!$ is the factorial of the number of elements in subset S, accounting for permutations of features within S.
- $(|N| - |S| - 1)!$ is the factorial representing the number of features not in subset S excluding the feature i.
- The factor $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ normalizes the influence of subsets of various sizes by accounting for the number of possible permutations of the features in and out of S, ensuring a balanced contribution from all subsets.

Model predictions

In the predictive segment of the study, patients in train set were stratified into low-risk and high-risk groups based on the optimized Jordan index to assess the predictive capability of the model. The evaluation was conducted using the following strategies:

(1) Log-rank Test: The variance of Kaplan-Meier survival curves was analyzed using the log-rank test to determine if the differences in survival time distributions

between the different risk groups were statistically significant. This statistical testing was conducted using the "survival" package (v 3.6-4) in R.

(2) Multivariate Cox Regression Analysis: This analysis assessed the correlation between the risk classification predicted by the machine learning model and 30-day all-cause mortality. It also considered potential confounding factors to ascertain the independent impact of risk prediction on forecasting the short-term risk of mortality. The analysis was performed using the "survival" package (v 3.6-4) and the "forestplot" package (version 3.1.3).

(3) Subgroup Analysis: The predictive effectiveness of the machine learning model was further evaluated in various subgroups, including sex, age, artificial renal replacement therapy, peripheral vascular disease, cerebrovascular disease, and malignant tumor. This analysis aimed to identify specific patient characteristics or clinical conditions that may influence the accuracy and reliability of the model's predictions. The "forestplot" package (version 3.1.3) was utilized to visualize the results and interactions within the subgroups.

This analysis aimed to identify specific patient characteristics or clinical conditions that may influence the accuracy and reliability of the model's predictions.

Statistical analysis

Continuous data were depicted as mean±standard deviation (SD) or median (interquartile range [IQR]) and compared with Student's t test or Mann-Whitney U test. Categorical variables were expressed as numbers (percentages) and compared using the chi-square test. The normality of data distribution was evaluated with the Shapiro-Wilk test. Non-normally distributed data or data exhibiting heterogeneity of variance were compared using the Kruskal-Wallis or Mann-Whitney U test. The P-value<0.05 was deemed as statistically meaningful and the statistical analysis was conducted with R (v 4.4.0).

Results

Patient characteristics

The flowchart was shown in Fig. 1. Initially, the dataset included data from 1,069 patients.we applied specific inclusion criteria: patients were admitted to the ICU for the first time, aged between 17 and 90 years, and stayed in the ICU for more than 24 h. This refinement process narrowed the number of suitable patients down to 499, including 300 men (60.1%) and 199 women (39.9%). At 30 days follow-up, 74 patients (14.8%) died. 499 patients were separated into a train set (399) and a test set (100). The general characteristics of the patients were presented in Table 1. There is no significant difference between the training dataset and the test dataset except for the variables "Potassium" and "Aniongap". Screened
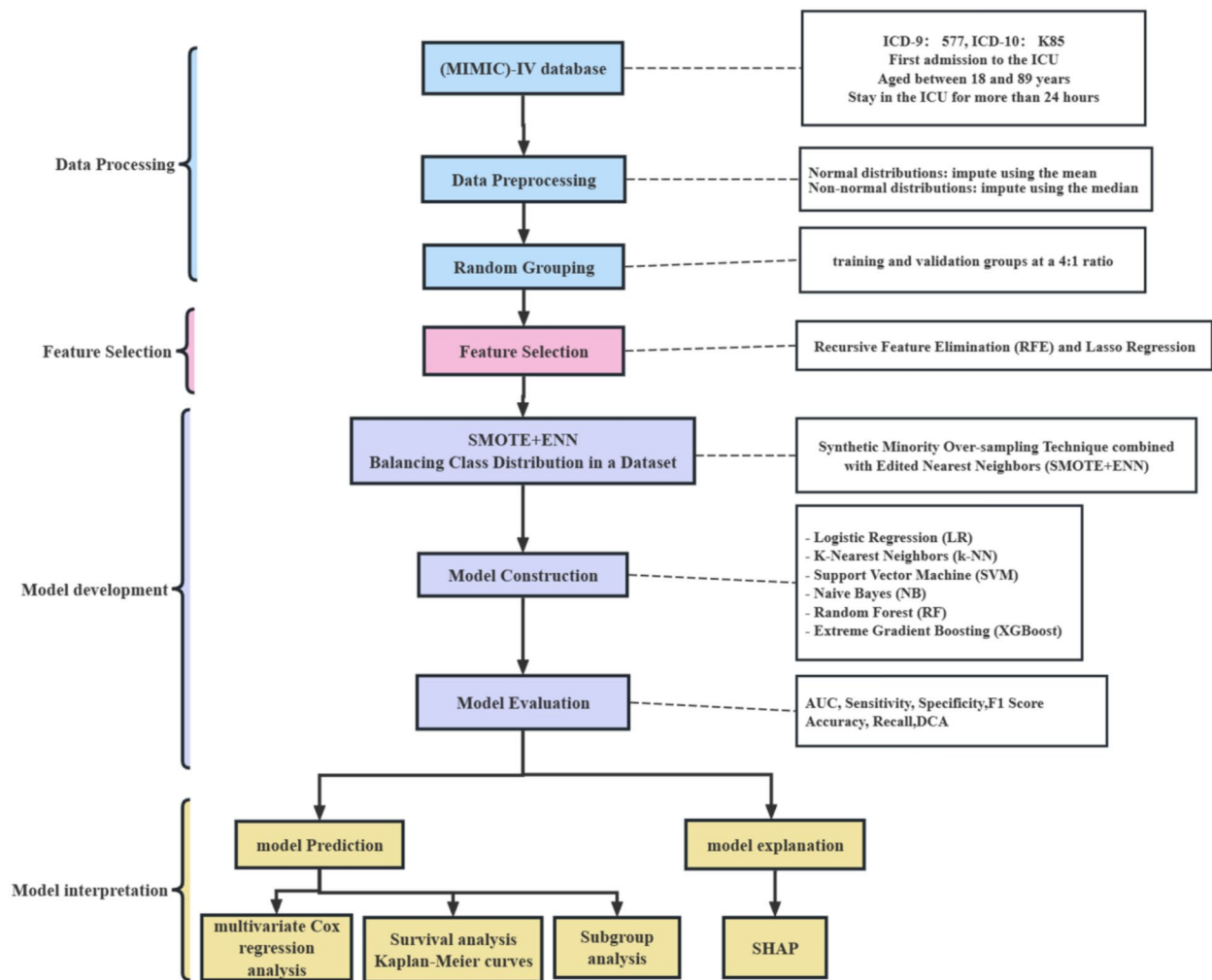
Fig. 1 The flowchart of this study

characteristics differed between fatal and nonfatal subjects in the training cohort (Table 2).

Feature selection
The initial dataset comprised 230 variables. Due to the presence of missing values, 87 of these variables were discarded, leaving 143 variables available for further analysis. By employing the LASSO method with an optimal lambda value of approximately 0.0127, the number of significant variables was reduced to 47. These attributes demonstrated minimal errors in predictive modeling. Additionally, recursive feature elimination (RFE) is used to identify another set of 53 important features based on accuracy:0.8547, Kappa:0.04696. A subsequent comparison and intersection of the attributes selected by both LASSO and RFE methods further refined this to the 25 most pertinent attributes, which were used to enhance the performance of the machine learning models (Fig. 2).

The 25 features included age, temperature_mean, mbp_max, sbp_max, SPO_2_max, SPO_2_min, SPO_2_mean, Charlson Comorbidity Index, APSIII, bun_min, anion gap_min, wbc_min, mbp_min, bilirubin_total, alp_min, glucose_min, rdw_max, PTT_min, rheumatic disease, metastatic solid tumor, peripheral vascular disease, myocardial infarct, malignant cancer, continuous renal replacement therapy (CRRT) and vasopressor.

Model hyper-parameters
Utilizing a dataset composed of 25 variables, advanced modeling techniques such as LR, SVM, RF, Random Forest (RF), and XGBoost were implemented. The training set was preconditioned using the SMOTE+ENN to ensure a balanced representation of positive and negative classifications. A grid search method was employed to identify the most effective hyperparameters for each model. Details regarding the adjustment of these

Table 1 The basic demographic characteristics of all patients

	Total (*N* = 499)	Test (*N* = 100)	Train (*N* = 399)	*P*
Demographic Characteristics				
Age (years), Median[IQR]	58.0 [45.0, 71.0]	59.0 [45.0, 75.0]	57.0 [45.0, 70.0]	0.495
Male, N (%)	300 (60.1)	66 (66)	234 (58.6)	0.219
Race (White), N (%)	303 (60.7)	54 (54)	249 (62.4)	0.154
Vital Signs, **Median**[IQR]				
Heart_Rate_Mean(bpm)	98.2 [85.1, 110.3]	98.8 [82.5, 110.6]	97.6 [85.2, 110.2]	0.602
Sbp_Mean(mmHg)	118.0 [107.6, 133.5]	117.1 [108.2, 130.4]	118.1 [107.5, 134.4]	0.748
Dbp_Mean(mmHg)	65.2 [57.6, 74.5]	63.8 [57.8, 72.6]	65.6 [57.5, 75.0]	0.510
Mbp_Mean(mmHg)	79.5 [71.7, 88.9]	77.8 [72.5, 86.4]	80.0 [71.3, 90.4]	0.631
Resp_Rate_Mean (bpm)	21.3 [18.4, 24.4]	21.4 [18.6, 24.8]	21.2 [18.4, 24.3]	0.392
Temperature_Mean(℃)	37.0 [36.6, 37.5]	37.0 [36.7, 37.5]	37.0 [36.6, 37.5]	0.405
Spo2_Mean(%)	96.6 [95.0, 98.0]	97.0 [95.2, 98.4]	96.5 [95.0, 98.0]	0.339
Glucose_Mean (mg/dL)	138.8 [110.9, 177.1]	138.8 [113.9, 168.2]	138.6 [110.3, 180.8]	0.674
Scores, Median[IQR]				
SOFA	6.0 [3.0, 9.0]	7.0 [3.5, 9.0]	6.0 [3.0, 10.0]	0.757
SIRS	3.0 [3.0, 4.0]	3.0 [2.5, 4.0]	3.0 [3.0, 4.0]	0.392
SAPSIII	38.0 [27.0, 50.0]	39.0 [28.5, 46.0]	38.0 [27.0, 51.0]	0.919
OASIS	35.0 [29.0, 42.0]	34.5 [30.0, 41.5]	35.0 [29.0, 42.0]	0.838
APSIII	54.0 [40.0, 74.0]	53.5 [40.0, 65.0]	54.0 [39.5, 75.0]	0.527
Charlson Comorbidity Index	4.0 [1.0, 6.0]	4.0 [1.5, 6.0]	4.0 [1.0, 6.0]	0.649
Laboratory Blood Tests				
Complete Blood Count, **Median**[IQR]				
WBC(10*9/L)	13.1 [9.1, 19.0]	11.9 [8.9, 16.8]	13.3 [9.1, 19.5]	0.188
Hemoglobin(g/L)	11.2 [9.2, 13.2]	11.2 [9.6, 13.2]	11.2 [9.2, 13.2]	0.623
Platelet(10*9/L)	205.0 [140.5, 312.0]	200.5 [156.5, 321.5]	209.0 [139.5, 312.0]	0.755
Hematocrit (%)	34.3 [28.6, 39.4]	34.0 [29.4, 39.5]	34.3 [28.4, 39.4]	0.779
Biochemistry, **Median[IQR]**				
Bicarbonate(mmol/L)	21.0 [17.0, 25.0]	22.0 [18.0, 25.0]	21.0 [17.0, 25.0]	0.084
Calcium (mmol/L)	8.1 [7.1, 9.0]	8.2 [7.2, 9.1]	8.1 [7.1, 8.9]	0.282
Chloride(mmol/L)	103.0 [98.0, 108.0]	104.0 [98.5, 108.0]	103.0 [98.0, 108.0]	0.612
Creatinine (mg/dL)	1.1 [0.8, 2.0]	1.1 [0.7, 1.7]	1.1 [0.8, 2.1]	0.362
Glucose(mmol/L)	125.0 [99.0, 186.5]	126.5 [101.0, 177.5]	125.0 [98.0, 187.0]	0.947
Sodium(mmol/L)	138.0 [135.0, 142.0]	139.0 [135.0, 142.5]	138.0 [135.0, 142.0]	0.491
Potassium(mmol/L)	4.1 [3.6, 4.8]	4.0 [3.5, 4.5]	4.2 [3.7, 4.8]	0.013
Total Bilirubin (mmol/L)	0.7 [0.4, 2.2]	0.7 [0.4, 1.9]	0.7 [0.4, 2.2]	0.760
ALP (U/L)	96.0 [61.0, 184.5]	77.5 [57.0, 170.0]	102.0 [62.5, 188.5]	0.150
AST (U/L)	87.0 [37.5, 354.5]	105.0 [37.0, 527.0]	80.0 [37.5, 341.0]	0.321
Coagulation, **Median[IQR]**				
INR	1.3 [1.1, 1.5]	1.3 [1.1, 1.5]	1.3 [1.1, 1.5]	0.808
PT (s)	14.5 [13.0, 18.3]	14.5 [13.1, 18.0]	14.4 [13.0, 18.4]	0.735
Blood Gas, **Median[IQR]**				
Anion gap (mmol/L)	16.0 [13.0, 21.0]	15.0 [13.0, 18.0]	17.0 [13.0, 22.0]	0.030
Lactate_Min(mmol/L)	1.3 [1.0, 1.8]	1.3 [0.9, 1.9]	1.3 [1.0, 1.8]	0.621
Lactate_Max(mmol/L)	1.9 [1.1, 4.0]	1.8 [1.0, 4.6]	1.9 [1.1, 3.9]	0.862
pH_Min	7.3 [7.1, 7.4]	7.3 [7.2, 7.4]	7.3 [7.1, 7.4]	0.939
pH_Max	7.4 [7.3, 7.4]	7.4 [7.3, 7.4]	7.4 [7.3, 7.4]	0.393
Po2_Min(mmHg)	74.0 [47.0, 120.0]	72.5 [45.5, 109.0]	76.0 [47.0, 124.5]	0.369
Po2_Max(mmHg)	141.0 [91.0, 280.0]	142.5 [91.5, 266.0]	139.0 [90.5, 281.5]	0.930
Pco2_Min(mmHg)	34.0 [27.0, 40.5]	34.0 [27.5, 40.0]	34.0 [27.0, 41.0]	0.860
Pco2_Max(mmHg)	42.0 [33.0, 51.0]	44.0 [34.0, 56.0]	42.0 [33.0, 50.0]	0.139
Baseexcess_Min(mmol/L)	-4.0 [-10.0, 1.0]	-3.0 [-9.0, 1.0]	-4.0 [-10.0, 1.0]	0.292
Baseexcess_Max(mmol/L)	-1.0 [-6.0, 1.0]	0.0 [-5.0, 1.5]	-2.0 [-7.0, 1.0]	0.465
Totalco2_Min(mEq/L)	22.0 [17.0, 29.0]	24.0 [17.0, 28.0]	22.0 [17.0, 29.0]	0.692

Table 1 (continued)

	Total (*N* = 499)	Test (*N* = 100)	Train (*N* = 399)	*P*
Totalco2_Max(mEq/L)	26.0 [21.0, 30.0]	27.0 [22.0, 30.0]	25.0 [21.0, 30.0]	0.123
Complications				
sepsis3, N (%)	386 (77.4)	80 (80)	306 (76.7)	0.566
Myocardial Infarct, N (%)	47 (9.4)	7 (7)	40 (10)	0.463
Congestive Heart Failure, N (%)	100 (20)	24 (24)	76 (19)	0.334
Peripheral Vascular Disease, N (%)	29 (5.8)	6 (6)	23 (5.8)	1.000
Cerebrovascular Disease, N (%)	36 (7.2)	7 (7)	29 (7.3)	1.000
Dementia, N (%)	10 (2)	2 (2)	8 (2)	1.000
Chronic Pulmonary Disease, N (%)	113 (22.6)	22 (22)	91 (22.8)	0.969
Rheumatic Disease, N (%)	14 (2.8)	1 (1)	13 (3.3)	0.320
Peptic Ulcer Disease, N (%)	23 (4.6)	6 (6)	17 (4.3)	0.431
Diabetes, N (%)	160 (32.1)	29 (29)	131 (32.8)	0.539
Liver Disease, N (%)	147 (29.5)	33 (33)	114 (28.6)	0.456
Paraplegia, N (%)	8 (1.6)	2 (2)	6 (1.5)	0.664
Malignant Cancer, N (%)	41 (8.2)	13 (13)	28 (7)	0.081
Metastatic Solid Tumor, N (%)	15 (3)	3 (3)	12 (3)	1.000
ARDS, N (%)	5 (1)	1 (1)	4 (1)	1.000
AKI stage, N (%)	210 (42.1)	46 (46)	164 (41.1)	0.816
Stage 1	79 (15.8)	15 (15)	64 (16)	
Stage 2	156 (31.3)	28 (28)	128 (32.1)	
Stage 3	54 (10.8)	11 (11)	43 (10.8)	
Treatment				
Vasopressor, N (%)	234 (46.9)	49 (49)	185 (46.4)	0.719
Sedatives, N (%)	253 (50.7)	52 (52)	201 (50.4)	0.858
Diuretics, N (%)	255 (51.1)	54 (54)	201 (50.4)	0.592
CRRT, N (%)	75 (15)	16 (16)	59 (14.8)	0.883
Prognosis				
30-day mortality, N (%)	74 (14.8)	14 (14)	60 (15)	0.917

Heart_Rate_Mean, Sbp_Mean, Dbp_Mean, Mbp_Mean, Resp_Rate_Mean, Temperature_Mean, Spo2_Mean, Glucose_Mean: Mean values in the first 24 h of ICU admission; Lactate_Min, pH_Min, Po2_Min, Pco2_Min, Baseexcess_Min, Totalco2_Min: Min values in the first 24 h of ICU admission; Lactate_Max, pH_Max, Po2_Max, Pco2_Max, Baseexcess_Max, Totalco2_Max: Max values in the first 24 h of ICU admission

parameters are thoroughly outlined in Supplementary Table S1. Parameters not specifically adjusted adhered to their default settings.

Performance of ML models

The analysis involved training and evaluating six prevalent machine learning models: LR, KNN, NB, RF, SVM and XGBoost, as detailed in Fig. 3; Table 3.

The evaluation metrics included sensitivity, specificity, recall, accuracy, F1 score and AUC. As per the results in Fig. 3; Table 3, the XGBoost model outperformed the other models with an AUROC of 0.881, an accuracy of 0.91, and an F1 score of 0.64. In terms of sensitivity, which is identical to the recall rate, Naive Bayesian and SVM demonstrated commendable performance, both registering a value of 0.6429, whereas KNN lagged with a rate of 0.4286. RF showcased the highest specificity at 0.9651, outperforming other models, which ranged between 0.8721 and 0.9302 in this metric. Concurrently, XGBoost attained the highest accuracy, whereas KNN scored the lowest at 0.81.

Overall, the data illustrates that the XGBoost model offers superior predictive capabilities. Conversely, KNN's overall effectiveness was comparatively lower as indicated in Fig. 3; Table 3.

In conclusion, given its robust performance across various metrics, the XGBoost model is selected for further analytical pursuits.

Visualization of feature importance

To intuitively interpret the chosen variables, the SHAP values were utilized to show what effect these elements had on the 30-day mortality in the model. In general, the more important the SHAP value of a feature is, the more influence it has on the model. We then ranked the importance of features in the best-performing XGBoost model.

In the modeling section, 25 variables were initially analyzed, out of which the XGBoost model ultimately selected 18 for inclusion in the model. Figure 4 displays these chosen variables, sorted by their importance. This ranking illustrates the extent to which each variable influences the model's predictive performance, aiding in

Table 2 Differences in screening characteristics between dead and non-dead subjects in the training group

	ALL (*N*=399)	Alive (*N*=339)	Dead (*N*=60)	*P*
Age(years), Median(IQR)	57.0 [45.0, 70.0]	56.0 [43.0, 67.0]	70.0 [57.0, 79.5]	< 0.001
Mbp_Min(mmHg), Median(IQR)	60.0 [51.0, 71.0]	62.0 [52.2, 72.0]	53.5 [47.8, 62.5]	< 0.001
Mbp_Max(mmHg), Median(IQR)	107.0 [93.5, 119.0]	108.0 [94.0, 119.8]	97.5 [88.0, 114.0]	0.002
Sbp_Max(mmHg), Median(IQR)	150.0 [136.0, 168.0]	152.0 [137.5, 169.5]	139.5 [128.0, 158.0]	< 0.001
Spo2_Max(%),Median(IQR)	100.0 [99.0, 100.0]	100.0 [99.0, 100.0]	100.0 [99.0, 100.0]	0.511
Spo2_Min(%),Median(IQR)	92.0 [90.0, 94.0]	92.0 [90.0, 94.0]	90.5 [87.5, 93.0]	< 0.001
Spo2_Mean(%),Median(IQR)	96.5 [95.0, 98.0]	96.6 [95.2, 98.1]	95.9 [94.1, 97.5]	0.013
Temperature_Mean(℃),Median(IQR)	37.0 [36.6, 37.5]	37.0 [36.7, 37.5]	36.8 [36.5, 37.2]	0.003
APSIII, Median(IQR)	54.0 [39.5, 75.0]	51.0 [37.0, 69.5]	77.5 [62.0, 92.0]	< 0.001
Charlson Comorbidity Index, Median(IQR)	4.0 [1.0, 6.0]	3.0 [1.0, 5.0]	6.0 [5.0, 8.0]	< 0.001
WBC_Min(10*9/L), Median(IQR)	10.8 [7.4, 15.9]	10.6 [7.2, 15.5]	12.8 [7.9, 18.5]	0.103
Glucose_Min(mg/dL), Median(IQR)	103.0 [84.0, 125.0]	101.0 [81.0, 123.0]	113.0 [88.0, 133.0]	0.066
PTT_Min(s), Median(IQR)	28.7 [25.7, 33.8]	28.4 [25.6, 32.8]	31.1 [26.5, 40.2]	0.003
Rdw_Max(%), Median(IQR)	15.0 [14.0, 16.5]	14.8 [13.9, 16.2]	15.9 [14.8, 18.4]	< 0.001
Bun_Min(mg/dL), Median(IQR)	19.0 [12.0, 36.0]	18.0 [11.0, 31.5]	35.5 [18.0, 63.0]	< 0.001
Aniongap_Min(mmol/L), Median(IQR)	13.0 [11.0, 16.0]	13.0 [11.0, 15.0]	16.0 [12.0, 18.0]	< 0.001
Total Bilirubin (mmol/L), Median(IQR)	0.7 [0.4, 2.2]	0.7 [0.4, 2.0]	0.9 [0.5, 4.3]	0.038
Alp_Min(U/L), Median(IQR)	85.0 [58.5, 143.5]	85.0 [60.0, 141.0]	84.0 [46.0, 153.0]	0.523
Metastatic Solid Tumor, N (%)	12 (3)	5 (1.5)	7 (11.7)	< 0.001
Peripheral Vascular Disease, N (%)	23 (5.8)	14 (4.1)	9 (15)	0.003
Myocardial Infarct, N (%)	40 (10)	26 (7.7)	14 (23.3)	< 0.001
Malignant Cancer, N (%)	28 (7)	16 (4.7)	12 (20)	< 0.001
Rheumatic Disease, N (%)	13 (3.3)	9 (2.7)	4 (6.7)	0.115
CRRT, N (%)	59 (14.8)	35 (10.3)	24 (40)	< 0.001
Vasopressor, N (%)	185 (46.4)	138 (40.7)	47 (78.3)	< 0.001

Mbp_Min, Spo2_Min, Glucose_Min, WBC_Min, PTT_Min, Bun_Min, Aniongap_Min, Alp_Min: Min values in the first 24 h of ICU admission; Mbp_Max, Rdw_Max, Sbp_Max, Spo2_Max: Max values in the first 24 h of ICU admission; Spo2_Mean, Temperature_Mean: Mean values in the first 24 h of ICU admission;

a deeper understanding of the decision-making process within the model.

Figure 5 displays the 18 predictors evaluated by the mean SHAP value. The feature ranks (Y-axis) refer to the significance of each feature for the prediction model, and the SHAP values (X-axis) correspond to the impact of each feature on each sample model. The relationship between the size of the characteristic value and the predicted impact can be seen through the color, and the distribution of the characteristic value is displayed (blue indicates the high-risk value, while yellow indicates the low-risk value).

As depicted in Fig. 5, the risk factors for 30-day all-cause mortality are as follows: higher APSIII scores, older age, and an increased Charlson comorbidity index. Past medical history factors include peripheral vascular disease, rheumatic diseases, and a history of malignancy. Within the first 24 h in the ICU, crucial indicators include lower average oxygen saturation, higher blood glucose levels, lower body temperature, and lower mean arterial pressure. Blood tests within this time frame show longer PTT durations, increased maximum red cell distribution width, and larger anion gaps which are significant markers. Additionally, the use of vasoactive medications

and Continuous Renal Replacement Therapy (CRRT) are highlighted as treatment factors.

Classification and risk stratification of predictive scores
XGBoost model was applied for predicting and stratifying the likelihood of 30-day all-cause mortality in SAP patients of the training set. All subjects in the training set were categorized into high-risk and low-risk groups, taking the maximum proximity entry index to be the best cutoff (0.62).

Survival Analysis
As shown in the Kaplan-Meier curves, the 30-day survival rate for patients identified as high-risk by the XGBoost model decreases over time, suggesting that these individuals are more likely to succumb. This observation is statistically significant (logarithmic rank test: $p < 0.0001$, Fig. 6).

COX regression analysis
The correlation between high ML risk and 30-day mortality in SAP patients remained after adjustment for the first 9 most impactful variables. (adjusted HR:10.61; 95% CI:5.47–20.60; $p < 0.001$). Multivariable COX regression analysis is shown in Fig. 7.
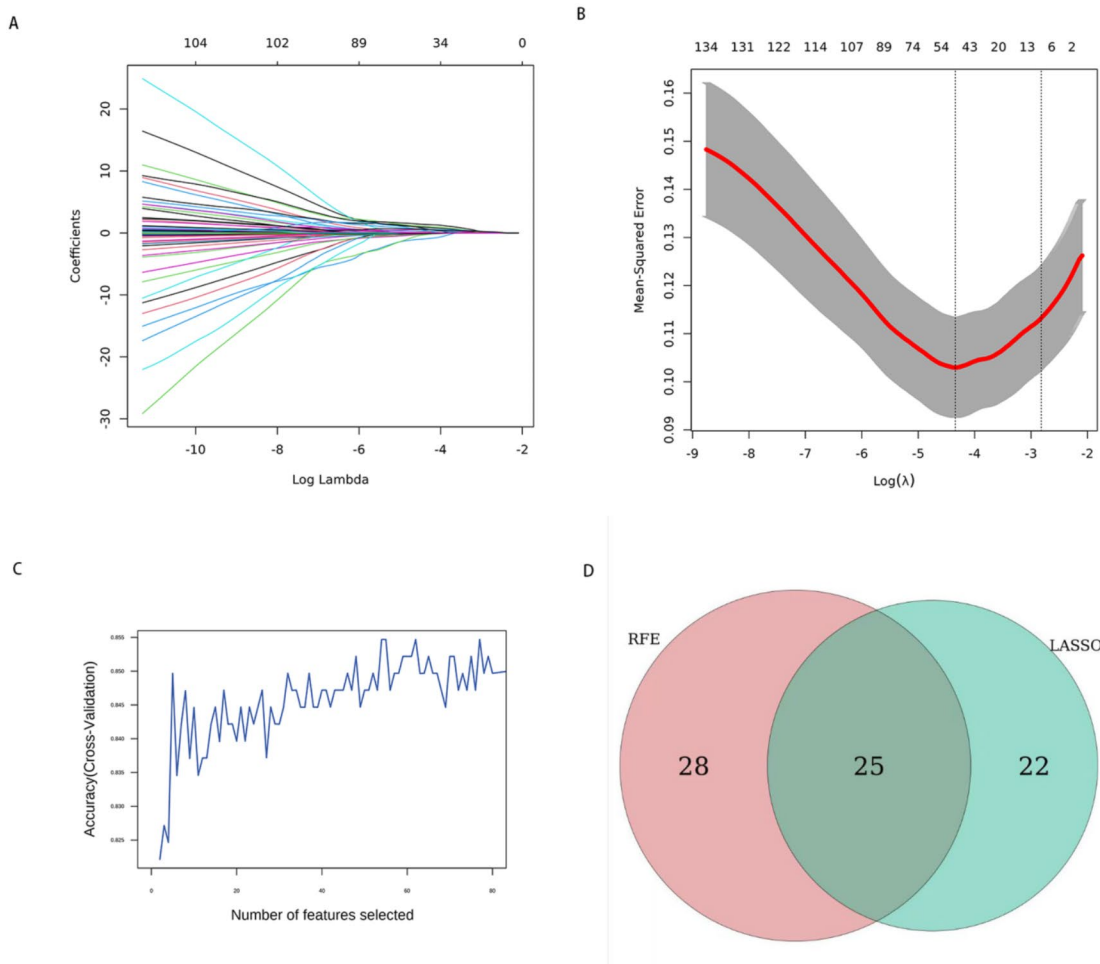
Fig. 2 Feature selection. **A**: Changes in the Coefficients of 104 Predictive Variables with Variations in the Regularization Parameter λ; **B**: Feature selection in Lasso regression analysis: The horizontal axis represents the change of λ, the vertical axis shows the coefficient of each characteristic variable; **C**: Features selected by RFE with the 5-fold Cross-Validation: This graph demonstrates the prediction accuracy of different feature sets selected through Recursive Feature Elimination (RFE) using 5-fold cross-validation. The horizontal axis displays the number of features included in the model, while the vertical axis represents the accuracy achieved by each subset; **D**: Venn Diagram Showing the Intersection Between Lasso Regression and Recursive Feature Elimination (RFE) with Five-Fold Cross-Validation

Subgroup analysis

Then, subgroup analysis is performed to further verify the predictive value of the model (Table 4). The subgroup analysis revealed that the model effectively distinguished between high-risk and low-risk patients with severe acute pancreatitis, irrespective of gender, age, history of continuous renal replacement therapy, or cancer. However, it was unable to predict the 30-day mortality risk for patients with severe acute pancreatitis who also suffered from peripheral vascular disease. To further evaluate the robustness of the results, we tested cross-interactions between high- and low-risk groups and age, gender, CRRT, peripheral vascular disease, or malignant cancer. In the high- and low-risk groups, there were no interactions found between age, gender, CRRT, peripheral vascular disease, or malignant cancer.

Discussion

In this research, we developed and tested an interpretable ML-based risk stratification tool for predicting the risk of all-cause mortality in SAP patients during a 30-day follow-up period. In this study, we applied 6 ML methods to construct the scoring system, among which XGBoost showed the best performance. The average AUC of this risk score is 0.881, sensitivity of 0.5714, specificity of 0.9651 and F1 score of 0.64which is significantly better than other currently available risk scores. Although ML models are often unable to output intrinsic explanations, we solve the problem of ML model interpretation by applying a state-of-the-art technique called an interpretable ML tool called SHAP. SHAP can help us to identify the first six most important feature variables of SAP mortality use of vasopressor, high Charlson comorbidity
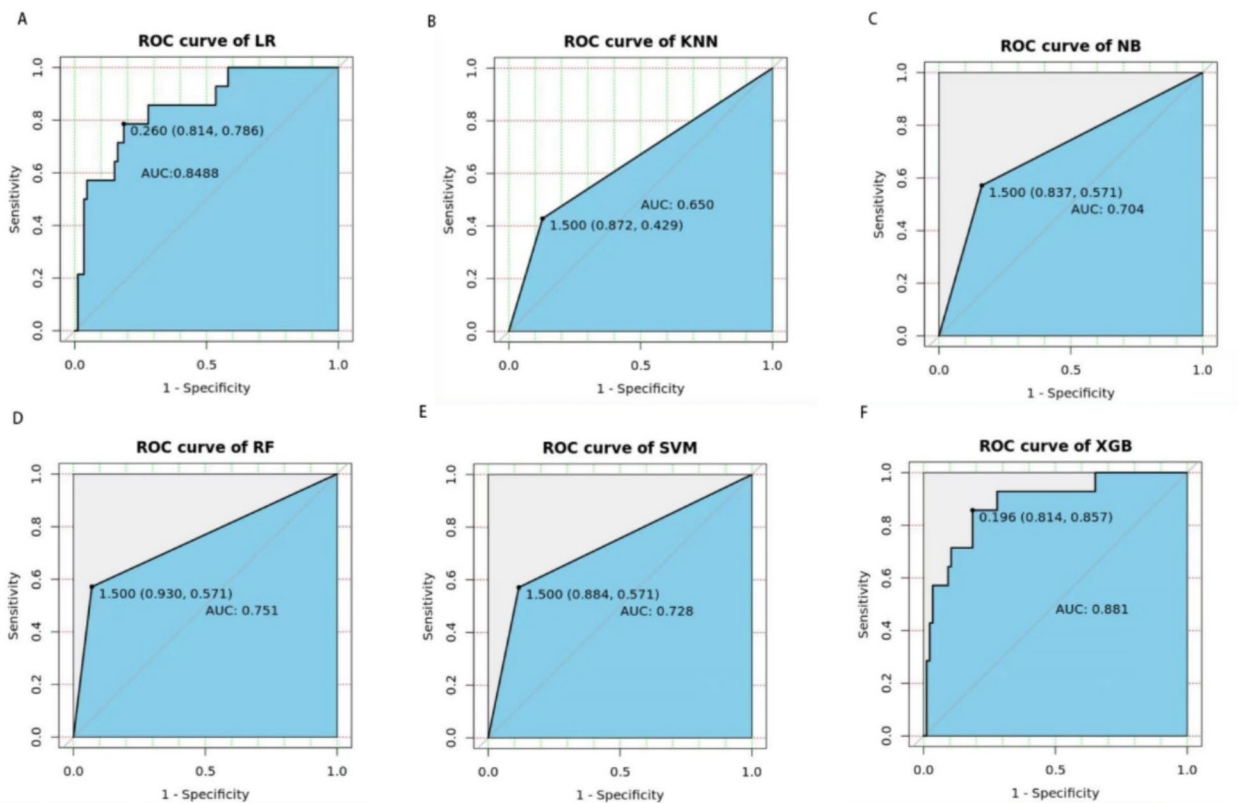
Fig. 3 AUC Comparison for Different Models: **A**: Logistic Regression (LR); **B**: K-Nearest Neighbors (KNN); **C**:Naive Bayes (NB); **D**:Random Forests (RF); **E**:Support Vector Machines (SVM); **F**: Extreme Gradient Boosting (XGBoost)

Table 3 Results of ML modeling of 30-day follow-up mortality in SAP patients

	LR	KNN	NB	RF	SVM	XGBoost
Sensitivity	0.5714	0.4286	0.6429	0.5714	0.6429	0.5714
Specificity	0.8721	0.8721	0.8721	0.9302	0.8837	0.9651
Recall	0.5714	0.4286	0.6429	0.5714	0.6429	0.5714
Accuracy	0.83	0.81	0.84	0.88	0.85	0.91
F1 score	0.4848	0.3871	0.5294	0.5714	0.5455	0.64
AUROC	0.8488	0.65	0.704	0.751	0.728	0.881

LR: Logistic Regression; KNN: K-Nearest Neighbors; NB: Naive Bayes; RF: Random Forests ; SVM: Support Vector Machines; XGBoost: Extreme Gradient Boosting

index, low blood oxygen saturation, history of malignant tumor, hyperglycemia and high APSIII score.

Compared with laboratory tests, predictive scores or models in previous studies, the risk stratification tool built with XGBoost as the main algorithm in this study can more accurately predict the 30-day mortality of acute severe pancreatitis. In previous studies, as shown in Supplementary Table 2, the AUC values for predicting 28-day mortality of SAP patients, including WBC, PLR, NLR, RDW, CRP, Bedside index of acute pancreatitis severity (BISAP), CTSI and APACHE II, were 0.796, 0.697, 0.749, 0.722, 0.595, 0.812, 0.84 and 0.78, both significantly lower than the performance of the scoring model established in this study [27–29]. In addition, these scoring systems have different limitations. For example, the APACHE II score is mainly used for critically ill patients rather

than for AP patients and requires invasive tests such as blood gas tests [9]. Ranson scores measure 48 h of data to predict prognosis, leading to delays in patient risk management [30]; Although the CTSI score may provide essential information on the diagnosis of AP, the availability of instruments may limit the application of the score and neglect the evaluation of clinical signs and symptoms [12]. The Harmless acute pancreatitis score (HAPS) was designed to identify mild acute pancreatitis [11].

Recently, several researchers have built ML models to predict the severity of AP patients and to identify SAP patients early [31, 32]. BalazsKui et al. used decision trees, random forest, logistic regression, SVM, CatBoost and XGBoost to construct ML models to identify the severity of AP patients at an early stage. The XGBoost
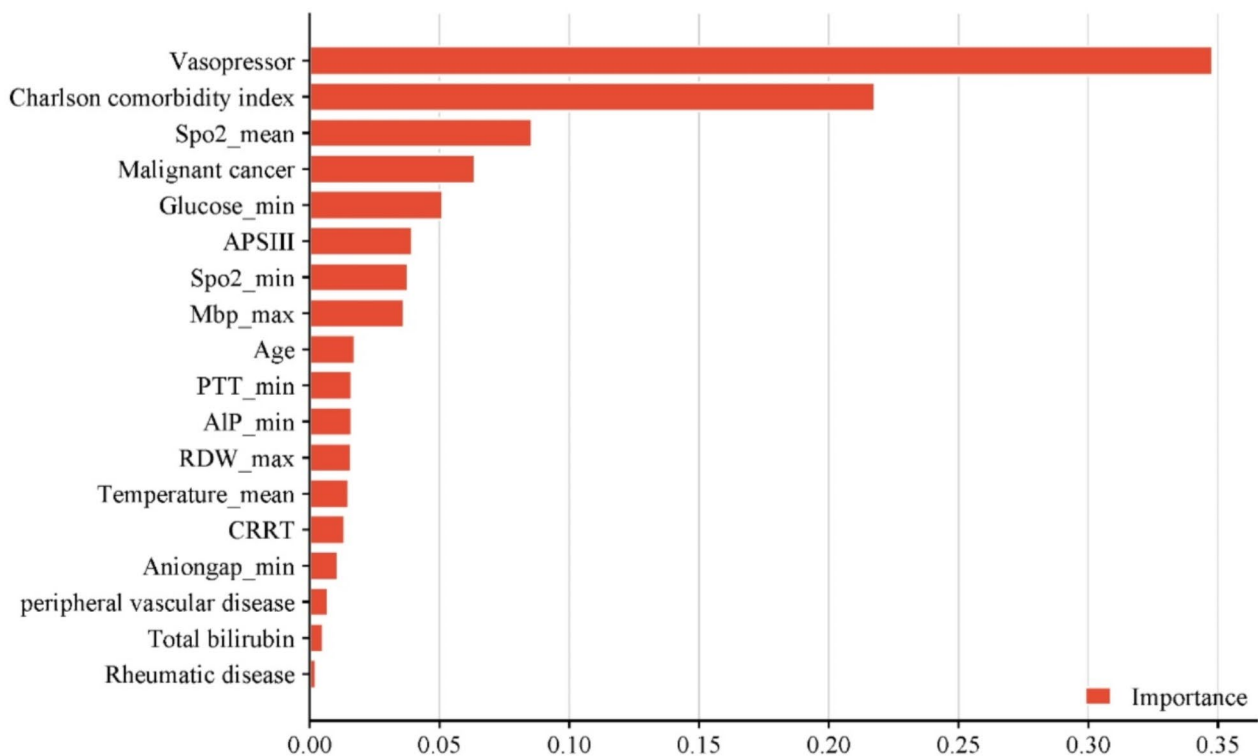
Fig. 4 The importance ranking of the top 18 variables based on the mean (|SHAP value|). APSIII: Acute Physiology Score III; Mbp_max: Mean Blood Pressure maximum; PTT_min: Partial Thromboplastin Time minimum; AlP_min: Alkaline Phosphatase minimum; RDW_max: Red Cell Distribution Width maximum; CRRT: Continuous Renal Replacement Therapy

classifier had the strongest predictive power, with an average AUC of about 0.81. In addition, Anjuli K Luthra et al. compared the ability of GBM ML and multivariable logistic regression to predict mortality in patients with acute biliary pancreatitis. the GBM ML model had higher PPV (47.3% vs. 35.9%) and lower sensitivity (40.1% vs. 46.7%) compared with the GBM ML model multivariable logistic regression, respectively [33]. This study was aimed at SAP patients in the ICU. Six ML methods, including XGBoost and LR, were used to establish models for predicting patient mortality within 30 days. The model constructed by XGBoost had excellent prediction performance, with an AUC value as high as 0.881, sensitivity of 0.5714, and specificity of 0.9651. The PPV is 0.7273. Therefore, this model is of great value for the death prognosis of patients with severe pancreatitis and is of great significance for the patient risk management of clinicians.

Although many studies have proven the predictive power of clinical elements on the negative outcome of pancreatitis, the present study further identifies significant predictors of all-cause mortality in AP patients. Previous studies have shown that clinical characteristics, demographic characteristics, and treatment status are important bases for patient risk assessment. Consistent with previous literature and clinical experience, the first six key variables involved in this model, including the use of vasopressor, high Charlson comorbidity index, low blood oxygen saturation, history of malignant tumor, hyperglycemia and high APSIII score, are important in predicting the poor prognosis of SAP patients. For example, AP Patients treated with vasopressors had a higher risk of death compared to non-users during the followup period [34]. Patients with a high Charlson Comorbidity Index had higher mortality [35]. Lower SPO2 is correlated with higher fatality in AP patients [36]; The mortality rate of AP patients with diabetes was markedly higher than that of AP patients without diabetes (1.7%). AP may be the first symptom of pancreatic cancer, and patients with malignant tumors such as pancreatic cancer are more likely to have a poor prognosis [37]. The CRRT plays an important role in the treatment of SAP patients [38]. Lower mean blood pressure and higher BUN were independent risk variables for mortality in SAP patients [39]. Older SAP patients are three times more likely to die than younger patients [40]; Hypothermia can lead to worse outcomes in SAP patients [41]; In addition, longer partial thromboplastin time (PTT) [42], higher red cell distribution width (RDW) [43], alkaline phosphatase (ALP) [44], rheumatic disease [45], peripheral vascular
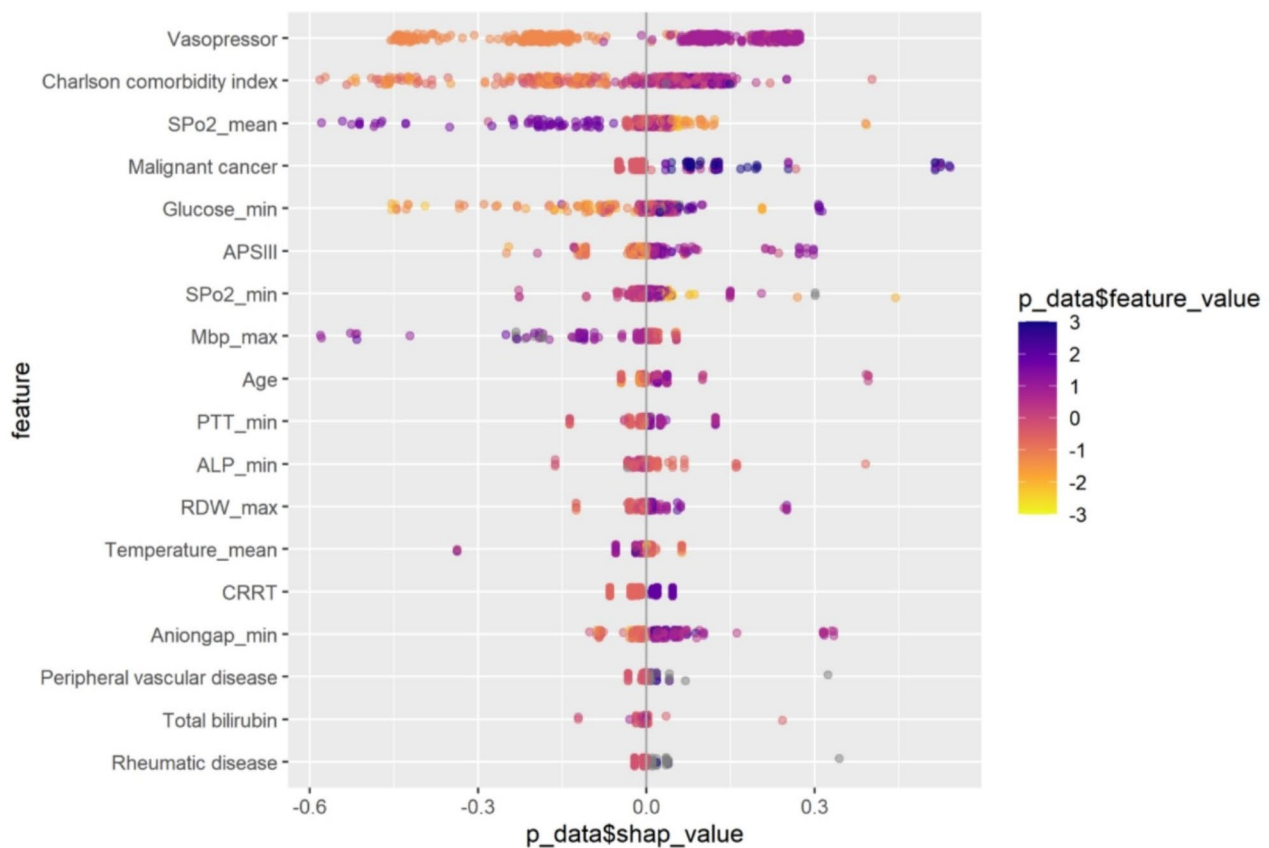
Fig. 5 The importance ranking of the 18 risk variables with stability and explanatory properties performed by the optimal model. The higher the SHAP value of the feature, the higher the risk of death for the patient. The blue portion of the feature's value represents higher values

disease [46], higher total bilirubin [47], metastatic solid tumor [48], myocardial infarct [49], SBP [39], and less WBC [50] were closely related to the high mortality of SAP patients. These results suggest that these variables are effective predictors of mortality in SAP and can prospectively provide a basis for clinicians' risk management of SAP patients.

The research model has the following advantages: First, compared with traditional logistic regression and linear regression, the ML model uses high-order nonlinear interaction, and its prediction performance is better and more stable [17, 51]. In this study, six kinds of ML models such as XGBoost were used to build prediction models and the best ones were selected for research. Second, the black-box nature of ML algorithms limits the interpretability of predictive models, while the AI tool SHAP identifies key variables and quantifies the impact of individual features on the ultimate prediction [24, 52]. Therefore, this study uses SHAP values to explain the critical variables contained in the predictive model to help clinical practitioners understand and apply the model and contribute to patient risk management. Finally, the key features involved in this model are objective data, which avoids the subjective bias of physicians. At the same time,

some limitations existed in this research. First of all, the present study only included patients from one hospital, which may cause some bias. We will further expand the scope of the study to cover patients from various areas and hospitals to optimize the performance of this model. Second, we focused only on common ML approaches to modeling and did not evaluate the performance of these models against currently used risk models. Third, based on the advantages of deep learning for building medical models, we will try to build prognostic models of AP through deep learning and conduct in-depth studies combining more comprehensive data and patient information to improve prediction.

Conclusion

In the explanatory machine learning model, the best prognostic model of severe acute pancreatitis was the XGBoost model, with an average AUC value of 0.881±0.033. The first six most important characteristic variables were the use of vasopressor, high Charlson comorbidity index, low blood oxygen saturation, history of malignant tumor, hyperglycemia and high APSIII score. Based on a machine learning algorithm, the model excavates the key clinical indicators that affect
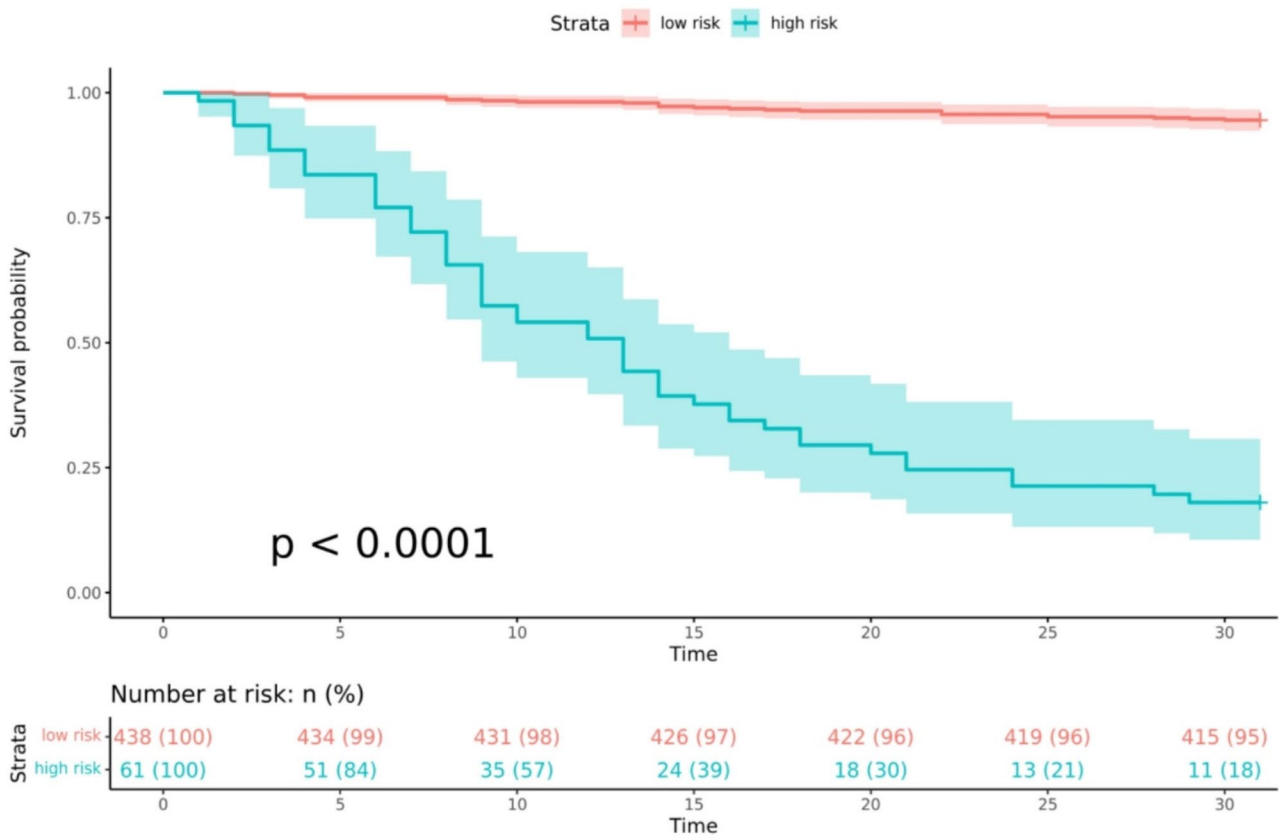
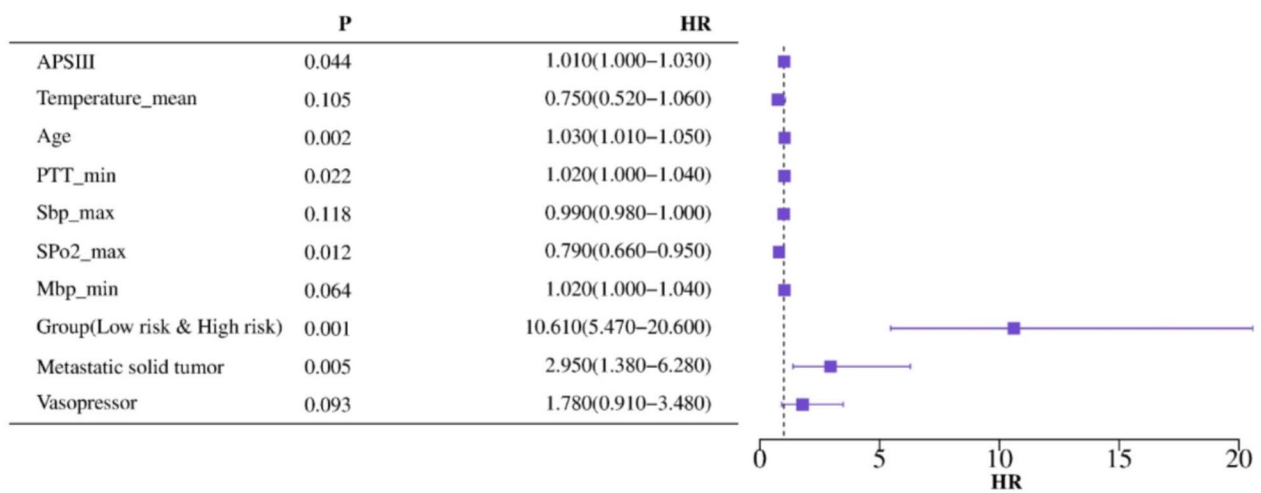**Fig. 6** Kaplan-Meier curves for low and high ML risk groups

	P	HR
APSIII	0.044	1.010(1.000–1.030)
Temperature_mean	0.105	0.750(0.520–1.060)
Age	0.002	1.030(1.010–1.050)
PTT_min	0.022	1.020(1.000–1.040)
Sbp_max	0.118	0.990(0.980–1.000)
SPo2_max	0.012	0.790(0.660–0.950)
Mbp_min	0.064	1.020(1.000–1.040)
Group(Low risk & High risk)	0.001	10.610(5.470–20.600)
Metastatic solid tumor	0.005	2.950(1.380–6.280)
Vasopressor	0.093	1.780(0.910–3.480)

Fig. 7 Multivariable Cox regression for 30-day all-cause death

the prognosis of critically ill patients and shows excellent predictive ability and risk stratification potential through comprehensive evaluation, which provides data support for treatment strategies.

Table 4 Subgroup analysis of diverse populations (Low-risk group as reference)

	Low risk group	High risk group	HR	95%CI	*P*	*P* for interaction
Sex						0.123
Female	reference		10.55	2.98–37.39	*p* < 0.001	
Male	reference		13.53	6.20-29.52	*p* < 0.001	
Age						0.739
≤ 57 years	reference		8.66	2.01–37.34	*p* = 0.004	
>57 years	reference		10.51	4.88–22.65	*p* < 0.001	
CRRT						0.797
No	reference		10.78	4.75–24.46	*p* < 0.001	
Yes	reference		29.53	5.34-163.22	*p* < 0.001	
Peripheral vascular disease						0.819
No	reference		9.61	4.70-19.65	*p* < 0.001	
Yes	reference		39.31	0.59-2615.70	*p* = 0.086	
Malignant cancer						0.142
No	reference		12.17	5.75–25.75	*p* < 0.001	
Yes	reference		21.82	1.69-282.16	*p* = 0.018	

Abbreviations

AP	Acute pancreatitis
SAP	Severe acute pancreatitis
ML	Machine learning
AUC	The area under the ROC curve
SHAP	SHAPley additive explanation
ICU	Intensive care unit
CRP	C-reactive protein
APACHE II	Acute physiological and chronic health assessment
BISAP	Acute pancreatitis severity bedside index
HAPS	Harmless acute pancreatitis score
MIMIC	Medical information mart for intensive care
MIT	Massachusetts institute of technology
ICD-9	International classification of diseases, ninth revision
SQL	Structured query language
WBC	White blood cell
RFE	Recursive feature elimination
SE	Standard error
SMOTE + ENN	Synthetic minority oversampling technique combined with an edited nearest neighbor
LR	Logistic regression
KNN	K-nearest neighbors
SVM	Support vector machines
NB	Naive bayesian
RF	Random forest
XGBoost	Extreme gradient boosting
IQR	Interquartile range
SD	Standard deviation
MBP	Mean blood pressure
RDW	Red blood cell distribution width
SBP	Systolic blood pressure
BUN	Blood urea nitrogen
CRRT	Continuous renal replacement therapy
PTT	Partial thromboplastin time
ALP	Alkaline phosphatase

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12911-024-02741-7.

Supplementary Material 1

Supplementary Material 2

Author contributions
Xiaojing Li: conceived and designed the study, collected the data, analyzed the data, and drafted the manuscript; Yueqin Tian: conceived and designed the study, collected the data, analyzed the data, and drafted the manuscript; Shuangmei Li: collected the data, analyzed the data and draw the figures. Haidong Wu: conceived and designed the study, and revised the manuscript; Tong Wang: conceived and designed the study, and revised the manuscript. All the listed authors have read and approved the submitted manuscript.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Emergency, the Eighth Affiliated Hospital of Sun Yat-sen University, Shenzhen 518033, China

²Department of Rehabilitation Medicine, The Third Affiliated Hospital, Sun Yat-sen University, No. 600, Tianhe Road, Guangzhou 510630, Guangdong, China

References

1. Mederos MA, Reber HA, Girgis MD. <ArticleTitle Language="En">Acute Pancreatitis: a review. JAMA. 2021;325:382–90.
2. Moran RA, García-Rayado G, de la Iglesia-García D, Martínez-Moneo E, Fort-Martorell E, Lauret-Braña E, et al. Influence of age, body mass index and comorbidity on major outcomes in acute pancreatitis, a prospective nationwide multicentre study. United Eur Gastroenterol J. 2018;6:1508–18.
3. Murata A, Ohtani M, Muramatsu K, Matsuda S. Influence of comorbidity on outcomes of older patients with acute pancreatitis based on a national administrative database. Hepatobiliary Pancreat Dis Int. 2015;14:422–8.
4. Dobszai D, Mátrai P, Gyöngyi Z, Csupor D, Bajor J, Erőss B, et al. Body-mass index correlates with severity and mortality in acute pancreatitis: A meta-analysis. World J Gastroenterol. 2019;25:729–43.
5. Pavlidis P, Crichton S, Lemmich Smith J, Morrison D, Atkinson S, Wyncoll D, et al. Improved outcome of severe acute pancreatitis in the intensive care unit. Crit Care Res Pract. 2013;2013:897107.
6. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Classification of acute pancreatitis–2012: revision of the Atlanta classification and definitions by international consensus. Gut. 2013;62:102–11.
7. Sternby H, Bolado F, Canaval-Zuleta HJ, Marra-López C, Hernando-Alonso AI, Del-Val-Antoñana A, et al. Determinants of Severity in Acute Pancreatitis: A Nation-wide Multicenter Prospective Cohort Study. Ann Surg. 2019;270:348–55.
8. Farkas N, Hanák L, Mikó A, Bajor J, Sarlós P, Czimmer J, et al. A Multicenter, International Cohort Analysis of 1435 Cases to Support Clinical Trial Design in Acute Pancreatitis. Front Physiol. 2019;10:1092.
9. Larvin M, McMahon MJ. APACHE-II score for assessment and monitoring of acute pancreatitis. Lancet. 1989;2:201–5.
10. Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. Gut. 2008;57:1698–703.
11. Lankisch PG, Weber-Dany B, Hebel K, Maisonneuve P, Lowenfels AB. The harmless acute pancreatitis score: a clinical algorithm for rapid initial stratification of nonsevere disease. Clin Gastroenterol Hepatol. 2009;7:702–5. quiz 607.
12. Choi HW, Park HJ, Choi S-Y, Do JH, Yoon NY, Ko A, et al. Early Prediction of the Severity of Acute Pancreatitis Using Radiologic and Clinical Scoring Systems With Classification Tree Analysis. AJR Am J Roentgenol. 2018;211:1035–43.
13. Mounzer R, Langmead CJ, Wu BU, Evans AC, Bishehsari F, Muddana V, et al. Comparison of existing clinical scoring systems to predict persistent organ failure in patients with acute pancreatitis. Gastroenterology. 2012;142:1476–82. quiz e15-16.
14. Ji M-Y, Yuan L, Lu S-M, Gao M-T, Zeng Z, Zhan N, et al. Glandular orientation and shape determined by computational pathology could identify aggressive tumor for early colon carcinoma: a triple-center study. J Transl Med. 2020;18:129.
15. Qiu Q, Nian Y-J, Guo Y, Tang L, Lu N, Wen L-Z, et al. Development and validation of three machine-learning models for predicting multiple organ failure in moderately severe and severe acute pancreatitis. BMC Gastroenterol. 2019;19:118.
16. Vo TH, Nguyen NTK, Kha QH, Le NQK. On the road to explainable AI in drug-drug interactions prediction: A systematic review. Comput Struct Biotechnol J. 2022;20:2112–23.
17. Hung TNK, Le NQK, Le NH, Van Tuan L, Nguyen TP, Thi C, et al. An AI-based Prediction Model for Drug-drug Interactions in Osteoporosis and Paget's Diseases from SMILES. Mol Inf. 2022;41:e2100264.
18. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. pp. 4768–77.
19. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011;45:1–67.
20. Chen Q, Meng Z, Su R. WERFE: A Gene Selection Algorithm Based on Recursive Feature Elimination and Ensemble Strategy. Front Bioeng Biotechnol. 2020;8:496.
21. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl. 2004;6:20–9.
22. A review of supervised. machine learning algorithms | IEEE Conference Publication | IEEE Xplore [Internet]. [cited 2023 Jul 1]. https://ieeexplore.ieee.org/document/7724478
23. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Comput Biol Med. 2021;137:104813.
24. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018;2:749–60.
25. Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M, Explainable. AI for Retinoblastoma Diagnosis: Interpreting Deep Learning Models with LIME and SHAP. Diagnostics (Basel). 2023;13:1932.
26. Aldughayfiq B, Sampalli S. Digital Health in Physicians' and Pharmacists' Office: A Comparative Study of e-Prescription Systems' Architecture and Digital Security in Eight Countries. OMICS. 2021;25:102–22.
27. Zhou Y, Han F, Shi X-L, Zhang J-X, Li G-Y, Yuan C-C, et al. Prediction of the severity of acute pancreatitis using machine learning models. Postgrad Med. 2022;134:703–10.
28. Papachristou GI, Muddana V, Yadav D, O'Connell M, Sanders MK, Slivka A, et al. Comparison of BISAP, Ranson's, APACHE-II, and CTSI scores in predicting organ failure, complications, and mortality in acute pancreatitis. Am J Gastroenterol. 2010;105:435–41. quiz 442.
29. Akdur G, Bardakcı O, Das M, Akdur O, Beyazit Y. Diagnostic utility of hematological indices in predicting adverse outcomes and severity of acute pancreatitis based on BISAP and modified Glasgow score. Ulus Travma Acil Cerrahi Derg. 2022;28:268–75.
30. Ranson JH, Rifkind KM, Roses DF, Fink SD, Eng K, Localio SA. Objective early identification of severe acute pancreatitis. Am J Gastroenterol. 1974;61:443–51.
31. Thapa R, Iqbal Z, Garikipati A, Siefkas A, Hoffman J, Mao Q, et al. Early prediction of severe acute pancreatitis using machine learning. Pancreatology. 2022;22:43–50.
32. Yuan L, Ji M, Wang S, Wen X, Huang P, Shen L, et al. Machine learning model identifies aggressive acute pancreatitis within 48 h of admission: a large retrospective study. BMC Med Inf Decis Mak. 2022;22:312.
33. Luthra AK, Porter K, Hinton A, Chao W-L, Papachristou GI, Conwell DL, et al. A Comparison of Machine Learning Methods and Conventional Logistic Regression for the Prediction of In-Hospital Mortality in Acute Biliary Pancreatitis. Pancreas. 2022;51:1292–9.
34. Shi H, Sun S-Y, He Y-S, Peng Q. Association between early vasopressor administration and in-hospital mortality in critically ill patients with acute pancreatitis: A cohort study from the MIMIC-IV database. Eur Rev Med Pharmacol Sci. 2023;27:787–98.
35. Knudsen JS, Heide-Jørgensen U, Mortensen FV, Sørensen HT, Ehrenstein V. Acute pancreatitis: 31-Year trends in incidence and mortality - A Danish population-based cohort study. Pancreatology. 2020;20:1332–9.
36. Miller J, Wu Y, Safa R, Marusca G, Bhatti S, Ahluwalia G, et al. Derivation and validation of the ED-SAS score for very early prediction of mortality and morbidity with acute pancreatitis: a retrospective observational study. BMC Emerg Med. 2021;21:16.
37. Umans DS, Hoogenboom SA, Sissingh NJ, Lekkerkerker SJ, Verdonk RC, van Hooft JE. Pancreatitis and pancreatic cancer: A case of the chicken or the egg. World J Gastroenterol. 2021;27:3148–57.
38. Sun S, He L, Bai M, Liu H, Li Y, Li L, et al. High-volume hemofiltration plus hemoperfusion for hyperlipidemic severe acute pancreatitis: a controlled pilot study. Ann Saudi Med. 2015;35:352–8.
39. Wilkman E, Kaukonen K-M, Pettilä V, Kuitunen A, Varpula M. Early hemodynamic variables and outcome in severe acute pancreatitis: a retrospective single-center cohort study. Pancreas. 2013;42:272–8.
40. Gardner TB, Vege SS, Chari ST, Pearson RK, Clain JE, Topazian MD, et al. The effect of age on hospital outcomes in severe acute pancreatitis. Pancreatology. 2008;8:265–70.
41. Andraus W, Jukemura J, Dutra F, Bechara E, Cunha JEM, Leite KRM, et al. Oxidative stress is enhanced by hypothermia imposed on cerulein-induced pancreatitis in rats. Clin (Sao Paulo). 2007;62:483–90.

42. Badhal SS, Sharma S, Saraya A, Mukhopadhyay AK. Prognostic significance of D-dimer, natural anticoagulants and routine coagulation parameters in acute pancreatitis. Trop Gastroenterol. 2012;33:193–9.

43. Zhou H, Mei X, He X, Lan T, Guo S. Severity stratification and prognostic prediction of patients with acute pancreatitis at early phase: A retrospective study. Med (Baltim). 2019;98:e15275.

44. Simsek O, Kocael A, Kocael P, Orhan A, Cengiz M, Balcı H, et al. Inflammatory mediators in the diagnosis and treatment of acute pancreatitis: pentraxin-3, procalcitonin and myeloperoxidase. Arch Med Sci. 2018;14:288–96.

45. Wang Q, Li M, Qian J, Lu C, Lü H. [Analysis of clinical features of autoimmune disease-related pancreatitis]. Zhonghua nei ke za zhi. 2008;47:999–1002.

46. Li M, Bai X, Xu K, Wu X, Guo T, Jiang Q, et al. Peripancreatic vascular involvement in patients with type 1 autoimmune pancreatitis. Hepatobiliary Surg Nutr. 2022;11:355–62.

47. Xu X, Ai F, Huang M. Deceased serum bilirubin and albumin levels in the assessment of severity and mortality in patients with acute pancreatitis. Int J Med Sci. 2020;17:2685–95.

48. Huang Y-W, Yang J-C, Chang Y-L, Tsang Y-M, Wang T-H. Acute pancreatitis combined with acute Budd-Chiari syndrome as the initial manifestation of small cell lung cancer. J Formos Med Assoc. 2005;104:431–5.

49. Luo Y, Li Z, Ge P, Guo H, Li L, Zhang G, et al. Comprehensive Mechanism, Novel Markers and Multidisciplinary Treatment of Severe Acute Pancreatitis-Associated Cardiac Injury – A Narrative Review. J Inflamm Res. 2021;14:3145–69.

50. Matsuda Y, Masuda Y, Shimoji K, Matsukawa M, Kinowaki Y, Fukumura Y, et al. Severe Acute Pancreatitis in Autopsies Associated With Surgeries and Severe Inflammatory Diseases. Pancreas. 2019;48:1321–8.

51. Pearce CB, Gunn SR, Ahmed A, Johnson CD. Machine learning can improve prediction of severity in acute pancreatitis using admission values of APACHE II score and C-reactive protein. Pancreatology. 2006;6:123–31.

52. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017;318:517–8.

Publisher's note