

Multimodal prediction models integrating radiomics and three-dimensional deep learning for acute respiratory distress syndrome in acute pancreatitis patients

Journal of International Medical Research

2026, Vol. 54(1) 1–14

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03000605251410432

journals.sagepub.com/home/imr



Jielu Zhou^{1*}, Yuying Wu^{1*}, Wen Liang^{1*},
Lin Liu¹, Chenyang Zhang¹, Yiping Shen²,
Meiyu Chen², Yu Wang³ , Chen Chao³,
Minyue Yin⁴, Jinzhou Zhu²  and Hailong Ge³

Abstract

Objectives: This study aimed to develop a multimodal predictive model that integrates clinical data, radiomics, and three-dimensional deep learning to forecast acute respiratory distress syndrome in patients with acute pancreatitis.

Methods: This retrospective study analyzed data from 759 patients with acute pancreatitis treated at three hospitals. Radiomics features were extracted from three-dimensional computed tomography images, and a three-dimensional deep learning model was developed using convolutional networks. These components were combined with clinical data using the XGBoost algorithm to construct a multimodal model. The performance of the model was compared with that of single-modal models and traditional scoring systems (Modified Computed Tomography Severity Index, Ranson score, and Bedside Index for Severity in Acute Pancreatitis), using area under the curve as the primary metric. Model interpretability was enhanced using variable

⁴Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, National Clinical Research Center for Digestive Disease, Beijing Digestive Disease Center, State Key Laboratory of Digestive Health, China

*These authors contributed equally.

Corresponding author:

Hailong Ge, Department of Hepatobiliary Surgery, Jintan Affiliated Hospital of Jiangsu University, Changzhou 213200, Jiangsu, China.
Email: ghl@jtrmy.com

¹Department of Geriatrics, Suzhou Kowloon Hospital, Shanghai Jiao Tong University School of Medicine, China

²Department of Gastroenterology, The First Affiliated Hospital of Soochow University, China

³Department of Hepatobiliary Surgery, Jintan Affiliated Hospital of Jiangsu University, China



importance analysis, SHapley Additive exPlanations, local interpretable model-agnostic explanations, calibration plots, and decision curve analysis.

Results: The multimodal model achieved area under the curve values of 0.872 (training set) and 0.876 (test set), outperforming traditional scores (Modified Computed Tomography Severity Index: 0.747 and 0.759; Ranson score: 0.575 and 0.568; and Bedside Index for Severity in Acute Pancreatitis: 0.748 and 0.757, respectively) and single-modal models (radiomics: 0.638 and 0.727 and deep learning: 0.756 and 0.727, respectively).

Conclusion: By integrating clinical tabular data, radiomics, and deep learning features, the multimodal model can predict the risk of acute respiratory distress syndrome in patients with acute pancreatitis at an early stage.

Keywords

Acute pancreatitis, acute respiratory distress syndrome, radiomics, deep learning, multimodal prediction models

Date received: 4 July 2025; accepted: 9 December 2025

Introduction

Acute pancreatitis (AP) is a common inflammatory disorder of the pancreas with a highly variable clinical course, ranging from mild, self-limiting disease to life-threatening severe forms.^{1,2} AP complicated by acute respiratory distress syndrome (ARDS) represents a particularly dangerous condition and is associated with significant morbidity and mortality.³ Previous studies have highlighted that ARDS is the most common and severe pulmonary complication of AP, with reported mortality rates as high as 44.5% due to the rapid progression of multiorgan failure. The pathophysiological mechanisms underlying ARDS in AP involve systemic inflammatory response syndrome (SIRS), cytokine storm, and damage to the alveolar-capillary membrane, which collectively lead to refractory hypoxemia.⁴ ARDS exacerbates pancreatic injury through bidirectional interactions, as pancreatic enzymes and inflammatory mediators circulate to the lungs, causing endothelial dysfunction, edema, and microthrombosis.⁵ Without early prediction and aggressive ventilatory

support (e.g. lung-protective strategies and extracorporeal membrane oxygenation (ECMO)), ARDS can contribute to prolonged intensive care unit (ICU) stays, secondary infections, and irreversible organ damage.^{6,7} Thus, ARDS associated with AP remains a critical determinant of poor prognosis and necessitates early risk stratification.⁸

In clinical practice, patient management relies heavily on the integration of diverse types of information, including tabular laboratory data, imaging, and audio data.⁹ Multimodal models that integrate data from multiple sources have demonstrated superior performance in medical tasks compared with single-modality models. These tasks include detection, classification, prediction, and prognosis.^{10–12} However, several challenges remain. For instance, different data modalities have distinct dimensional characteristics: clinical variables derived from electronic health records (EHRs) are typically one-dimensional, whereas imaging data are often two-dimensional or three-dimensional (3D).¹³ In terms of feature extraction, high-level features are often

manually identified or extracted using specialized software applications.¹⁰

This study aimed to develop and validate predictive models for early and noninvasive prediction of ARDS by integrating data from multiple modalities across three medical centers. The data modalities included EHRs, radiomics, and deep learning features extracted from 3D computed tomography (CT) scans.

Methods

Study design

This was a retrospective multicenter study. Data extracted from EHRs and imaging systems were obtained anonymously from three hospitals: Jintan Hospital affiliated to Jiangsu University, Suzhou Kowloon Hospital, and the First Affiliated Hospital of Soochow University. Patients from Jintan Hospital and Suzhou Kowloon Hospital were assigned to the training set, while those from the First Affiliated Hospital of Soochow University were assigned to the independent testing set.

Adult patients aged over 18 years who were admitted with a diagnosis of AP between January 2017 and December 2023 were included. The exclusion criteria were as follows: (a) chronic liver or renal disease; (b) hematological disease; (c) chronic, recurrent, idiopathic, or traumatic pancreatitis; pancreatic tumors; or a history of pancreatic resection; (d) pregnancy; and (e) a history of chemotherapy or radiotherapy. The study flowchart is presented in Figure 1.

Ethical considerations. This retrospective cohort study was approved by the Ethics Committee of the First Affiliated Hospital of Soochow University (approval number: 2022098). The need for informed consent was waived for this retrospective study, and all participants were managed in accordance with established guidelines.^{6,14}

ARDS was diagnosed according to the Berlin definition.¹⁵

Data collection

Data preparation. This study employed a retrospective approach to systematically collect clinical data, including demographic information (age and sex), clinical characteristics (e.g. etiological classification, including biliary, hyperlipidemia, alcoholic, and other causes), and comorbidities (e.g. hypertension and diabetes). Blood samples were collected within 24 h of admission for all patients to perform complete blood cell analysis, liver and kidney function tests, electrolyte measurements, lipid profiling, C-reactive protein assessment, and other laboratory evaluations. Abdominal non-contrast CT scans in DICOM format were obtained within 72 h of admission and scored using the Modified CT Severity Index (MCTSI).¹⁶ To comprehensively evaluate patient condition, the SIRS score, Bedside Index for Severity in Acute Pancreatitis (BISAP), and Ranson score were calculated within 24 h of AP onset to further quantify disease severity.¹⁷

Data cleaning. To ensure data quality and prevent potential data leakage, rigorous data cleaning procedures were conducted separately for each dataset. Variables with missing rates exceeding 20% were excluded to minimize the impact of incomplete data on analytical results. As shown in Figure 1, 79 patients were excluded from the training set because of missing data, while 51 patients were excluded in the test set. Table S1 lists the exact missing rates (%) for each clinical feature included in the multimodal model.

For the remaining variables with missing values, imputation was performed using the classification and regression tree (CART) algorithm. Classification trees were used for categorical variables, while regression trees were applied to continuous variables.

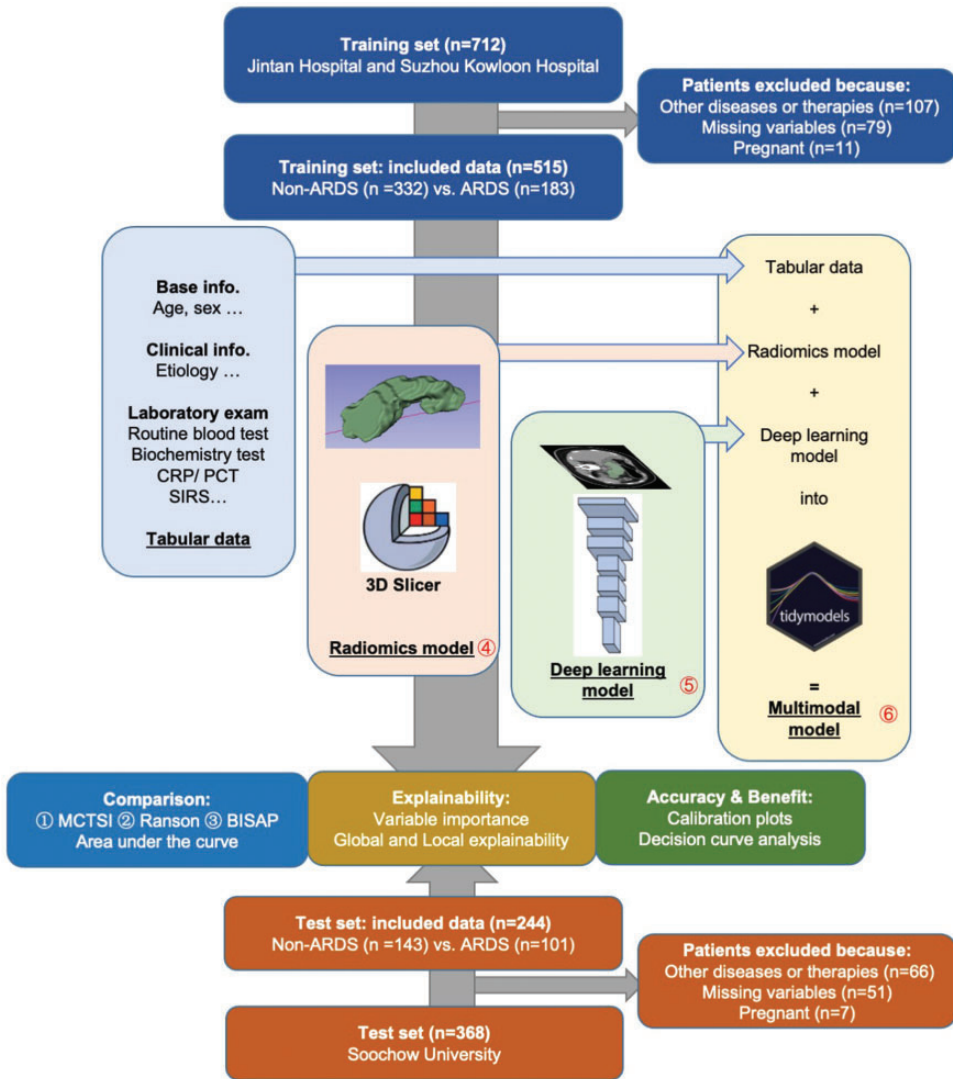


Figure 1. Flowchart of the study.

Development of models

With the occurrence of ARDS within 1 week of admission defined as the outcome variable, multiple predictive models were developed using different modalities: (a) radiomics; (b) 3D deep learning; and (c) a multimodal approach integrating clinical tabular data, radiomics, and deep learning.

Radiomics model. Feature extraction. The collected CT images were processed using 3D Slicer software (version 5.0.3) to manually delineate the pancreatic lesion regions and generate a 3D pancreatic model. Subsequently, 107 radiomics features were extracted using the built-in “Radiomics” module, which included seven categories: first-order statistics, gray-level co-occurrence

matrix, gray-level dependence matrix, gray-level run-length matrix, gray-level size-zone matrix, neighboring gray-tone difference matrix, and 3D-based shape features. Principal component (PC) analysis was applied to the 107 radiomics features for dimensionality reduction through linear transformation.^{12,18,19} PCs accounting for over 80% of the total explained variance were selected for subsequent modeling.

Consistency analysis. To ensure the robustness of the radiomics model, the intraclass correlation coefficient (ICC) was used to evaluate the consistency of the regions of interest (ROIs) during delineation and feature extraction. Two experienced radiologists independently delineated the ROIs. Both inter- and intraclass consistency tests were conducted to validate the reliability of the model. According to the ICC scoring criteria, consistency was categorized into three levels: low consistency (ICC < 0.4), moderate consistency (ICC between 0.4 and 0.75), and high consistency (ICC > 0.75), with ICC values ranging from 0 to 1.

Model development. After standardization, feature selection was performed using differential testing, Pearson correlation analysis, and Least Absolute Shrinkage and Selection Operator (LASSO) regression. This pipeline reduced the 12 PCs (retaining >80% variance from the original 107 radiomics features) to 5 final features, which were used to train the XGBoost model (R CRAN version: 1.7.8.1). The final model output served as a binary classifier to predict the occurrence of ARDS.

Deep learning model. To construct a binary deep learning predictive model for AP complicated by ARDS, two key steps were undertaken: (a) development of a pancreatic 3D segmentation model for noncontrast CT and (b) construction of a classification model based on semantically segmented 3D CT images.

Pancreatic 3D segmentation model. This study used the U-Net architecture as the baseline framework. The input noncontrast CT images were preprocessed by cropping to a size of 228 × 228 pixels, applying histogram equalization to enhance image contrast and maximize information entropy, and normalizing pixel values to eliminate scale-related effects. During model training, the Adam optimizer was used with a batch size of 20 and 200 epochs to ensure adequate data features and model coverage. Additionally, real-time data augmentation was employed to dynamically transform images during training, effectively enhancing the model's generalization ability.

Deep learning classification model. Based on the construction of a pancreatic 3D segmentation model, a deep learning classification model was further developed. This model takes 3D CT images of the pancreas as input and outputs the probability of occurrence of the primary outcome. Specifically, the 3D images are fed into a 3D ResNet50 encoder, where high-level features were extracted through a deep convolutional network and subsequently passed to a fully connected layer to further process and generate probability predictions for AP complicated with ARDS.

Multimodal fusion model. To fully leverage the complementary strengths of different data modalities, we constructed a multimodal fusion model using the XGBoost algorithm. Specifically, the predicted probabilities (i.e. continuous risk scores) generated by the radiomics model and the 3D deep learning model were used as two input features. These probabilities were combined with all available clinical tabular variables, following data cleaning and imputation, as candidate predictors. During training of the XGBoost fusion model, feature importance was automatically computed, and the top eight clinical features—ranked according to their contribution to prediction—were

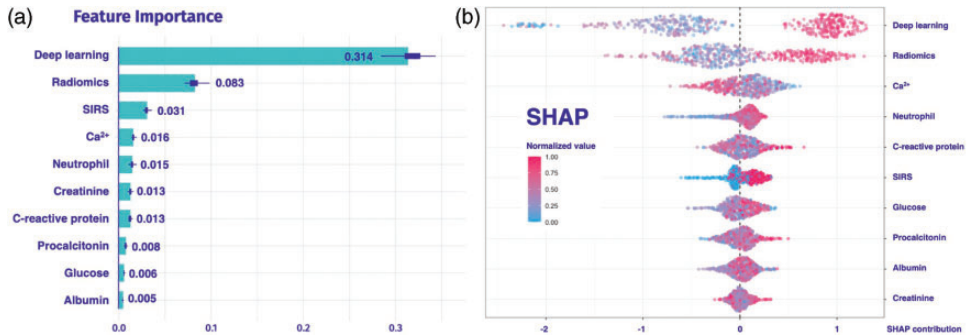


Figure 2. Global explainability of the multimodal model. (a) Variable importance of the multimodal model and (b) SHAP plots: Each point represents one case. The closer the variable value is to 1.0 at the x-axis, the higher the possibility of ARDS. ARDS: acute respiratory distress syndrome; SHAP: SHapley Additive exPlanations.

identified post hoc from the final model (see section “*Variable importance of the multimodal model*” and Figure 2(a)). Thus, the final multimodal model inputs consisted of the following: (a) the radiomics prediction score, (b) the deep learning prediction score, and (c) the top eight clinical variables selected based on XGBoost-derived feature importance within the fusion framework.

Statistical analysis

Receiver operating characteristic (ROC) curves were plotted, and the area under the curve (AUC) was calculated to evaluate the diagnostic performance of different models and existing scoring systems. The DeLong test was used to compare AUCs among the models and scoring systems.

Explainability and visualization

Variable importance was visualized using bar charts to illustrate the contribution of individual features to the prediction task. SHapley Additive exPlanations (SHAP) is a method used to explain predictions generated by machine learning models. This method is based on the concept of Shapley values from game theory, which provides an approach for fairly distributing the global contribution of each feature to the

model’s prediction. Meanwhile, local interpretable model-agnostic explanations (LIME) provide local explanations that elucidate the relationship between relevant features and the individual predictions by examining specific examples.^{20,21}

Calibration plots (also known as calibration curves) assess how well a model’s predicted probabilities align with the actual outcomes, reflecting the accuracy of the predicted probabilities. If the calibration curve closely follows the 45-degree reference line, it indicates that the model’s predicted probabilities are highly accurate.²²

Decision curve analysis (DCA) plots assess the clinical utility of a predictive model by comparing the net benefit across different decision thresholds. They reflect how effectively a model’s predictions translate into actionable decisions that benefit patients or users. DCA plots illustrate the net benefit of using a model across a range of threshold probabilities. The net benefit is calculated by considering true positives, false positives, and the relative harm associated with false positives. DCA plots help determine the range of threshold probabilities over which a model provides a positive net benefit, indicating its potential usefulness in clinical practice. Thus, DCA plots can be used to compare multiple models by

identifying which model provides the highest net benefit across varying threshold probabilities.

Software versions and platforms

Statistical analyses were conducted using R (version 4.1.0) and Python (version 3.9). For the development of machine learning algorithms, the tidymodels platform (version 0.2.0) was utilized. Deep learning models were constructed using the Keras Python platform with TensorFlow version 2.8.0 as the backend. The hardware used for this study was a Mac mini equipped with an 8-core Apple M1 processor (with an integrated graphic processing unit (GPU)) and 16 GB of memory to ensure efficient execution of the experiments.

Results

Characteristics of patients in the study

As shown in Table 1, a total of 515 patients with AP were included in the training set, of whom 183 developed ARDS. In the test set, 244 patients with AP were included, of whom 101 developed ARDS. In the training set, biliary disease was the most common etiological cause regardless of ARDS status; it was observed in 38.9% of patients without ARDS and 43.7% of patients with ARDS. Hyperlipidemia was the second most common cause (21.1% and 27.9%), followed by alcohol-related etiologies (8.1% and 4.4%, respectively). Similarly, in the test set, biliary factors remained the most common cause (51.7% and 48.5%, respectively).

Variable importance of the multimodal model

This study developed a multimodal prediction model for ARDS that integrated clinical tabular variables with prediction scores

derived from deep learning and radiomics models. During model development, the XGBoost algorithm was used to rank and incorporate variables according to their relative importance (Figure 2(a)). Specifically, the prediction scores derived from the deep learning and radiomics models were ranked as the most important features, followed by the presence of SIRS, calcium ion levels, neutrophil count, creatinine, C-reactive protein, procalcitonin, blood glucose, and albumin.

General explainability of the multimodal model: SHAP plots

By aggregating SHAP values across all instances, we obtained global insights into feature importance and interactions, enabling the identification of the most influential features in the model. As shown in Figure 2(b), features with values closer to 1.0 were associated with a higher likelihood of developing ARDS.

Local explainability of the multimodal model: LIME plots

Figure 3 illustrates the prediction process of the multimodal model for two randomly selected cases. For an AP patient without ARDS (Figure 3(a)), the model generated a predicted value of 0.026, consistent with the patient's observed outcome. The most influential contributors to this prediction included the deep learning model score (-1.272), radiomics model score (-1.035), neutrophil count ($11.7 \times 10^9/L$), procalcitonin (0.19 ng/mL), and blood glucose (7.48 mmol/L), with corresponding contributions of -0.239 , -0.075 , -0.012 , -0.012 , and -0.008 .

Conversely, Figure 3(b) presents the prediction for a patient with AP who developed ARDS. The model generated a predicted probability of 0.869, which was concordant with the patient's true outcome.

Table 1. Characteristics of patients at admission.

	Training set (n = 515)		p	Test set (n = 244)		p
	Non-ARDS (n = 332)	ARDS (n = 183)		Non-ARDS (n = 143)	ARDS (n = 101)	
Sex, n (%)						
Male	194 (58.4)	105 (57.4)	0.889	95 (66.4)	59 (58.4)	0.253
Female	138 (41.6)	78 (42.6)		48 (33.6)	42 (41.6)	
Age, years	51.52 (17.34)	51.21 (16.43)	0.840	52.88 (18.06)	52.36 (18.35)	0.826
Etiology, n (%)						
Biliary	129 (38.9)	80 (43.7)	0.047	74 (51.7)	49 (48.5)	0.595
Hyperlipidemia	70 (21.1)	51 (27.9)		24 (16.8)	22 (21.8)	
Alcoholic	27 (8.1)	8 (4.4)		10 (7.0)	4 (4.0)	
Others	106 (31.9)	44 (24.0)		35 (24.5)	26 (25.7)	
SIRS, n (%)						
No	198 (59.6)	91 (49.7)	0.038	95 (66.4)	48 (47.5)	0.005
Yes	134 (40.4)	92 (50.3)		48 (33.6)	53 (52.5)	
WBC count ($\times 10^9/L$)	14.20 (17.55)	15.85 (18.57)	0.318	13.22 (5.11)	14.76 (5.33)	0.024
N count ($\times 10^9/L$)	11.33 (5.22)	12.72 (4.48)	0.002	11.23 (5.10)	12.89 (5.18)	0.014
L count ($\times 10^9/L$)	1.30 (1.02)	1.14 (0.92)	0.081	1.20 (0.83)	1.24 (0.93)	0.691
PLT count ($\times 10^9/L$)	199.72 (66.95)	192.91 (60.30)	0.253	208.02 (76.75)	201.94 (69.05)	0.526
Ca ²⁺ (mmol/L)	2.11 (0.20)	2.03 (0.23)	<0.001	2.11 (0.18)	2.05 (0.21)	0.011
TG (mmol/L)	1.50 (0.91, 4.16)	2.12 (1.05, 5.45)	0.006	1.32 (0.84, 2.99)	1.51 (0.89, 4.08)	0.388
Glucose (mmol/L)	8.11 (3.51)	8.72 (3.33)	0.055	7.85 (3.12)	8.43 (3.52)	0.180
Albumin (g/L)	36.10 (5.45)	33.90 (5.24)	<0.001	35.65 (4.79)	33.79 (4.82)	0.003
ALT (U/L)	31.90 (15.70, 126.35)	37.04 (20.16, 80.71)	0.403	37.60 (19.75, 129.10)	47.37 (23.00, 97.00)	0.600
AST (U/L)	27.00 (18.00, 59.60)	30.10 (19.57, 51.75)	0.556	32.00 (18.20, 70.00)	31.77 (20.10, 68.10)	0.443
GGT (U/L)	77.40 (35.00, 215.65)	104.80 (45.97, 224.36)	0.213	93.30 (40.45, 225.95)	121.31 (58.00, 299.38)	0.165
ALP (U/L)	106.54 (77.61)	101.66 (74.99)	0.490	109.99 (93.75)	108.09 (67.11)	0.862
Creatinine ($\mu\text{mol/L}$)	71.43 (46.77)	78.31 (75.80)	0.204	70.80 (33.08)	78.89 (80.77)	0.283
Procalcitonin (ng/mL)	0.21 (0.09)	0.20 (0.06)	0.426	0.20 (0.07)	0.22 (0.14)	0.243
CRP (mg/L)	55.45 (7.95, 163.42)	109.00 (25.40, 196.71)	0.002	72.38 (16.85, 149.15)	110.10 (36.30, 200.19)	0.016

Continuous variables are presented as mean (standard deviation) or median (interquartile range); categorical variables are presented as count (percentage). p-values indicate comparisons between the two groups.

ALP: alkaline phosphatase; ALT: glutamic-pyruvic transaminase; AST: glutamic-oxaloacetic transaminase; CRP: C-reactive protein; GGT: glutamyl transpeptidase; L: lymphocyte; N: neutrophil; PLT: platelet; SIRS: systemic inflammatory response syndrome; TG: total triglycerides; WBC: white blood cell; Ca²⁺: calcium ion.

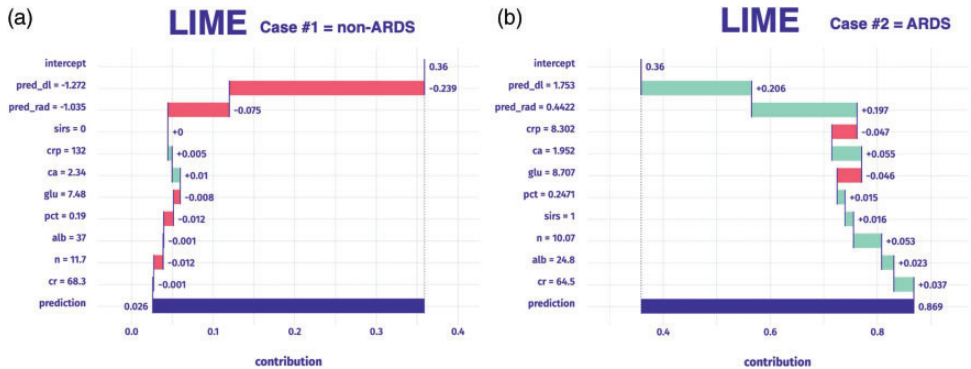


Figure 3. Local explainability of the multimodal model: LIME. (a) A randomly chosen patient with AP without ARDS and (b) a randomly chosen patient with AP and ARDS. alb: albumin; AP: acute pancreatitis; ARDS: acute respiratory distress syndrome; ca: Ca^{2+} (calcium ion); cr: creatinine; crp: C-reactive protein; glu: glucose; LIME: local interpretable model-agnostic explanations; n: neutrophil; pred_dl: deep learning model; pred_rad: radiomics model; pct: procalcitonin; SIRS: systemic inflammatory response syndrome.

The five most significant contributors to this prediction were the deep learning model score (1.753), radiomics model score (0.442), calcium ion concentration (1.952 mmol/L), neutrophil count (0.19 ng/mL), and CRP (8.302 mg/L), with corresponding contributions of +0.206, +0.197, +0.055, +0.053, and -0.047.

Performance of models and scoring systems

As shown in Figure 4, the multimodal model demonstrated superior performance in the training set ($\text{AUC} = 0.872$) compared with other models and established scoring systems (all p -values < 0.001). Similarly, in the test set, the multimodal model achieved the highest performance ($\text{AUC} = 0.876$).

Calibration plots

Figure S1 presents the calibration plots for the multimodal model. In the training set (Figure S1(A)), the multimodal model demonstrated good calibration, with minimal deviation between the observed and predicted probabilities (mean absolute error = 0.072). In the test set (Figure S1

(B)), the mean absolute error was 0.032, further indicating high accuracy of the predicted probabilities.

DCA plots

As shown in the DCA plot in Figure S2, the y-axis represents the net benefit. The blue line represents the multimodal model, the red line represents the radiomics model, and the green line represents the deep learning model. The “All” line represents the assumption that all patients with AP developed ARDS, while the “None” line represents the assumption that no patients with AP developed ARDS.

In the training set (Figure S2(A)), the multimodal model provided a positive net benefit across all patients, except at threshold probabilities of 72%–78%. The model also outperformed both the deep learning and radiomics models in terms of net benefit. In the test set (Figure S2(B)), the multimodal model provided a net benefit across all patients, except at threshold probabilities between 85% and 90%, again demonstrating superior performance compared with the deep learning and radiomics models.

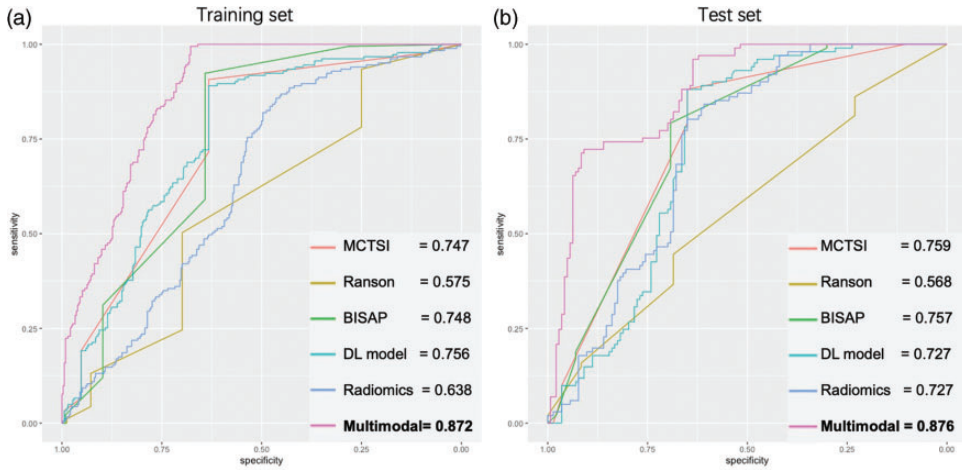


Figure 4. ROC analyses of models and scoring systems. (a) Training data and (b) test data.

Discussion

This study integrated clinical tabular data, radiomics features, and 3D deep learning features obtained at admission to develop a multimodal model for predicting ARDS in patients with AP. The model demonstrated robust performance in the training set, confirming its superiority over single-modality models and highlighting its favorable calibration and clinical utility. Moreover, the robustness and generalizability of the model were further validated using an independent test set. Additionally, the incorporation of explainable approaches, including global variable importance analysis and local prediction interpretation, enhanced the transparency and interpretability of the model.

In recent years, deep learning has emerged as a benchmark approach in machine learning,^{19,23} particularly following the advent of convolutional neural networks.²⁴ Deep learning has achieved significant progress in text and image processing, enabling simultaneous learning from diverse data types. The integration of large volumes of feature-rich data, which are increasingly available in modern medical

settings, is expected to enhance model performance.¹⁰ Artificial intelligence (AI) algorithms are particularly well suited to handle the nonlinear, high-dimensional, and complex characteristics of medical data.²⁵ Thus, integrating multiple data modalities can improve clinical acceptance and diagnostic accuracy, while AI-based tools can help reduce data redundancy and further enhance model performance.²¹

ARDS is a life-threatening form of respiratory failure characterized by acute bilateral pulmonary edema and refractory hypoxemia. Early identification of patients at high risk of ARDS, along with accurate prediction of disease severity, is crucial.²⁶ Previous studies have proposed several scoring systems to predict the severity of AP, including MCTSI, Ranson, and BISAP; however, their feasibility for predicting ARDS remains unclear. Li et al.²⁷ conducted a multicenter retrospective study involving 597 patients diagnosed with AP across 4 hospitals. Using multivariate logistic regression analysis, they identified four independent risk factors for both severe AP (SAP) and ARDS: heart rate, respiratory rate, serum calcium, and blood urea nitrogen. For ARDS prediction, the model

achieved AUC values of 0.892 and 0.833 in the training and test sets, respectively, outperforming SIRS and the quick Sepsis-Related Organ Failure Assessment. Zou et al.²⁸ developed predictive models for ARDS following SAP using artificial neural networks and logistic regression approaches. They collected clinical tabular data from 214 patients with SAP, randomly divided into a training set ($n=149$) and a test set ($n=65$). The artificial neural networks model demonstrated an accuracy of 80.0% and an AUC of 0.853. The most critical predictive variables in the artificial neural network model included the BISAP score, procalcitonin, prothrombin time, and serum calcium. Zhou et al.²⁹ developed several machine learning models to predict acute respiratory failure in patients with AP using data from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. The evaluated algorithms included logistic regression, decision tree, k-nearest neighbors, naive Bayes, and XGBoost. Among these models, the XGBoost model achieved the best predictive performance, with AUCs of 0.86 and 0.87 in the training and validation cohorts, respectively. Radiomics has also been explored for predicting ARDS in AP,³⁰ with studies suggesting that semiquantitative metrics derived from chest CT images, such as pleural effusion volume and the number of affected lung lobes with consolidation, may aid in the early identification of ARDS. However, this anthropometric-based approach is subject to subjective bias, and not all patients with AP undergo routine chest CT at admission. This limitation not only leads to unnecessary use of medical resources but also increases the economic burden on patients.

In this study, we developed a multimodal model that integrates clinical tabular data, radiomics features, and deep learning features derived from abdominal CT scans obtained at admission to predict ARDS in

patients with AP. The model achieved AUC values exceeding 0.85 in both the training and test sets, significantly outperforming single-modality models (radiomics and deep learning) and established clinical scoring systems, including MCTSI, Ranson, and BISAP. The model demonstrated greater efficiency in handling high-dimensional and multimodal data, thereby improving predictive accuracy and robustness. Moreover, we collected routine clinical test data at admission to extract all feature variables and subsequently optimized variable selection using the XGBoost algorithm, which identified SIRS, calcium ions, neutrophils, creatinine, C-reactive protein, procalcitonin, blood glucose, and albumin as the top eight tabular variables in the model's importance ranking, consistent with previous studies. Furthermore, the interpretability of the model was enhanced by visualizing the impact of key variables on prediction outcomes through global and local interpretability graphs, providing intuitive references for clinical decision-making. Finally, calibration and DCA plots further confirmed the model's predictive accuracy and clinical benefit.

However, this study has several limitations. First, the data samples were primarily derived from the Chinese population, which may restrict the applicability of the model to other racial and regional groups. Notably, the etiology of AP varies significantly across regions—biliary causes predominate in East Asia, whereas alcoholic or hypertriglyceridemic etiologies are more prevalent in Western populations. These etiological differences may influence systemic inflammatory profiles, imaging characteristics, and, consequently, the relevance of identified biomarkers (e.g. calcium and CRP) or radiomic patterns. Additionally, genetic and environmental factors may further influence disease progression and susceptibility to ARDS. Therefore, external validation in geographically and ethnically diverse cohorts is essential before clinical

implementation. Second, the retrospective collection of data from multiple centers resulted in the presence of missing data. Although strict inclusion and exclusion criteria, along with a relatively large sample size can, may mitigate these impacts, prospective studies are still required to further validate the performance of the multimodal model. Finally, the current model does not incorporate a real-time mechanism to assess the completeness and quality of input data and therefore cannot automatically identify or flag cases in which insufficient data may result in unreliable predictions. Future studies could incorporate uncertainty quantification methods to further enhance the clinical safety of the model.

Conclusion

This study developed and validated predictive models for the early and noninvasive prediction of ARDS in patients with AP by integrating data from multiple modalities across three medical centers. The data modalities included EHRs, radiomics features, and deep learning features extracted from 3D CT scans. This approach may assist medical practitioners in the early identification of patients at high risk of ARDS and potentially improve management strategies for individuals with AP.

Acknowledgments

Not applicable.

Author contributions

J. Zhou, Y. Wu, and L. Liu wrote the manuscript; W. Liang, C. Zhang, Y. Wang, and M. Yin collected the clinical data; Y. Shen, C. Chen, and M. Chen analyzed the clinical data; and H. Ge and J. Zhu contributed to the study design.

Data availability statement

The code used in this study is available upon reasonable request from the corresponding author.

Declaration of conflicting interests

The authors declare no competing interests and financial disclosure.


Funding

This study was supported by the National Natural Science Foundation of China (82000540), Youth Program of Suzhou Health Committee (KJXW2019001), the Open Fund of Key Laboratory of Hepatosplenic Surgery, Ministry of Education (GPKF202304), Science and Technology Projects of Jintan Municipal Health Commission (JTYXH-2025-1-02), Medical Education Collaborative Innovation Fund of Jiangsu University (No. JDYY2023042), and Suzhou key specialty (2241.07.01-0125).

Supplemental material

Supplemental material for this article is available online.

ORCID iDs

Yu Wang  <https://orcid.org/0009-0004-4063-8284>

Jinzhou Zhu  <https://orcid.org/0000-0003-0544-9248>

References

1. Banks PA, Bollen TL, Dervenis C; Acute Pancreatitis Classification Working Group, et al. Classification of acute pancreatitis – 2012: revision of the Atlanta classification and definitions by international consensus. *Gut* 2013; 62: 102–111.
2. Boxhoorn L, Voermans RP, Bouwense SA, et al. Acute pancreatitis. *Lancet* 2020; 396: 726–734.
3. Ge P, Luo Y, Okoye CS, et al. Intestinal barrier damage, systemic inflammatory response syndrome, and acute lung injury: a troublesome trio for acute pancreatitis. *Biomed Pharmacother* 2020; 132: 110770.
4. Schepers NJ, Bakker OJ, Besselink MG; Dutch Pancreatitis Study Group, et al. Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut* 2019; 68: 1044–1051.

5. Geng C, Zhang J, Wu S, et al. The role of ischaemia-modified albumin in the prognosis of acute pancreatitis and its correlation with the NF- κ B-mediated inflammatory response. *J Int Med Res* 2024; 52: 3000605 241287163.
6. Vivian E, Cler L, Conwell D, et al. Acute pancreatitis task force on quality: development of quality indicators for acute pancreatitis management. *Am J Gastroenterol* 2019; 114: 1322–1342.
7. Siriwardena AK, Jegatheeswaran S and Mason JM; PROCAP investigators. A procalcitonin-based algorithm to guide antibiotic use in patients with acute pancreatitis (PROCAP): a single-centre, patient-blinded, randomised controlled trial. *Lancet Gastroenterol Hepatol* 2022; 7: 913–921.
8. Yin M, Zhang R, Zhou Z, et al. Automated machine learning for the early prediction of the severity of acute pancreatitis in hospitals. *Front Cell Infect Microbiol* 2022; 12: 886935.
9. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021; 8: 53.
10. Huang SC, Pareek A, Seyyedi S, et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med* 2020; 3: 136.
11. Warner E, Lee J, Hsu W, et al. Multimodal machine learning in image-based and clinical biomedicine: survey and prospects. *Int J Comput Vis* 2024; 132: 3753–3769.
12. Yin M, Lin J, Wang Y, et al. Development and validation of a multimodal model in predicting severe acute pancreatitis based on radiomics and deep learning. *Int J Med Inform* 2024; 184: 105341.
13. Wang Y, Hong Y, Wang Y, et al. Automated multimodal machine learning for esophageal variceal bleeding prediction based on endoscopy and structured data. *J Digit Imaging* 2023; 36: 326–338.
14. Zheng Z, Ding YX, Qu YX, et al. A narrative review of acute pancreatitis and its diagnosis, pathogenetic mechanism, and management. *Ann Transl Med* 2021; 9: 69.
15. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 2012; 307: 2526–2533.
16. Mounzer R, Langmead CJ, Wu BU, et al. Comparison of existing clinical scoring systems to predict persistent organ failure in patients with acute pancreatitis. *Gastroenterology* 2012; 142: 1476–1482. quiz e15–16.
17. Bollen TL, Singh VK, Maurer R, et al. A comparative evaluation of radiologic and clinical scoring systems in the early prediction of severity in acute pancreatitis. *Am J Gastroenterol* 2012; 107: 612–619.
18. Zhong J, Hu Y, Xing Y, et al. A systematic review of radiomics in pancreatitis: applying the evidence level rating tool for promoting clinical transferability. *Insights Imaging* 2022; 13: 139.
19. Tong T, Gu J, Xu D, et al. Deep learning radiomics based on contrast-enhanced ultrasound images for assisted diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis. *BMC Med* 2022; 20: 74.
20. Lu C, Liu L, Yin M, et al. The development and validation of automated machine learning models for predicting lymph node metastasis in Siewert type II T1 adenocarcinoma of the esophagogastric junction. *Front Med (Lausanne)* 2024; 11: 1266278.
21. Lin J, Yang J, Yin M, et al. Development and validation of multimodal models to predict the 30-day mortality of ICU patients based on clinical parameters and chest X-rays. *J Imaging Inform Med* 2024; 37: 1312–1322.
22. Lin J, Yin M, Liu L, et al. The development of a prediction model based on random survival forest for the postoperative prognosis of pancreatic cancer: a SEER-based study. *Cancers (Basel)* 2022; 14: 4667.
23. He Y, Lin J, Zhu S, et al. Deep learning in the radiologic diagnosis of osteoporosis: a literature review. *J Int Med Res* 2024; 52: 3000605241244754.
24. Li Y, Zhu S, Wang Y, et al. Development and validation of deep learning models for bowel obstruction on plain abdominal radiograph. *J Int Med Res* 2024; 52: 3000605241271844.
25. Gao X, Lin J, Qu C, et al. Computer-aided diagnostic system with automated deep

- learning method based on the AutoGluon framework improved the diagnostic accuracy of early esophageal cancer. *J Gastrointest Oncol* 2024; 15: 535–543.
26. Tocoglu AG, Koksas AS, Toka B, et al. Validation of the Revised Atlanta Criteria in determining the severity of acute pancreatitis. *Eur J Gastroenterol Hepatol* 2023; 35: 1137–1142.
 27. Li YL, Zhang DD, Xiong YY, et al. Development and external validation of models to predict acute respiratory distress syndrome related to severe acute pancreatitis. *World J Gastroenterol* 2022; 28: 2123–2136.
 28. Zou K, Ren W, Huang S, et al. The role of artificial neural networks in prediction of severe acute pancreatitis associated acute respiratory distress syndrome: a retrospective study. *Medicine (Baltimore)* 2023; 102: e34399.
 29. Zhou LX, Zhou Q, Gao TM, et al. Machine learning predicts acute respiratory failure in pancreatitis patients: a retrospective study. *Int J Med Inform* 2024; 192: 105629.
 30. Song LJ and Xiao B. Medical imaging for pancreatic diseases: prediction of severe acute pancreatitis complicated with acute respiratory distress syndrome. *World J Gastroenterol* 2022; 28: 6206–6212.