Article

*Discovery Medicine* 2025; 37(192): 73–92
https://doi.org/10.24976/Discov.Med.202537192.7

# Construction of a Multi-View Deep Learning Model for the Severity Classification of Acute Pancreatitis

Kailai Xiang[1,2], Dong Shang[1,2,*]

[1]Department of General Surgery, First Affiliated Hospital of Dalian Medical University, 116011 Dalian, Liaoning, China
[2]Clinical Laboratory of Integrative Medicine, First Affiliated Hospital of Dalian Medical University, 116011 Dalian, Liaoning, China
*Correspondence: shangdong@dmu.edu.cn (Dong Shang)
Published: 20 January 2025

**Background: Acute pancreatitis (AP) is a prevalent pathological condition of abdomen characterized by sudden onset, high incidence and complex progression. Timely assessment of AP severity is crucial for informing intervention decisions so as to delay deterioration and reduce mortality rates. Existing AP-related scoring systems can only assess current condition of patients and utilize only a single type of clinical data, which is of great limitation. Therefore, it is imperative to establish more accurate and data-compatible methods for predicting the severity of AP. The artificial intelligence (AI) algorithm based on artificial neural network (ANN) allow for the adaptive feature extraction for objective task through its internal complex network, instead of the hand-crafted methods commonly used in traditional machine learning (ML) algorithms. In this study, we delve into the final severity classification prediction of newly admitted AP patients, using deep learning (DL) algorithm to develop multi-view models, incorporated with patients' demographic information, vital signs, AP-related laboratory indexes and admission computed tomography (CT) images.**

**Methods: The pancreatitis database in the platform of Clinical Data Research Center of Acute Abdominal Surgery at the First Affiliated Hospital of Dalian Medical University was used to gather AP cases. Deep neural network (DNN) and convolutional neural network (CNN) were utilized to construct models. The DNN prediction models with clinical data as input, the CNN prediction models with admission CT as input, and the multi-view models combining the two inputs were respectively established to predict the severity of AP.**

**Results: DL models for AP severity classification based on clinical indexes, imaging data and merged data were constructed. The multi-view model based on merged data offered more accurate prediction of the final severity classification of AP, with an overall accuracy rate of 80.26% (95% confidence interval (CI): 79.58%–80.94%). The constituent accuracy rates for mild acute pancreatitis, moderately severe acute pancreatitis, and severe acute pancreatitis were 91.69% (95% CI: 87.80%–95.57%), 64.90% (95% CI: 58.85%–70.95%), and 75.56% (95% CI: 68.58%–82.53%), respectively.**

**Conclusion: The multi-view models using clinical indexes and imaging data as input outperform single-view models in AP severity prediction.**

**Keywords:** acute pancreatitis severity classification; computer vision; machine learning; deep learning; multi-view learning

## Introduction

Acute pancreatitis (AP) is a common digestive disorder characterized by rapid onset, high incidence, intricate progression and multiple complications [1]. Common causes of AP include cholelithiasis, alcoholism, hyperlipidemia, trauma, and infection [2]. The trajectories and outcomes vary across the spectrum of AP with different severity. The 1992 Atlanta classification divided AP into mild acute pancreatitis (MAP) and severe acute pancreatitis (SAP) based on disease severity. Most of the MAPs are self-limiting, having short courses of disease and good prognoses. Patients with SAP may develop local or systemic complications, organ failure and other serious events during the course of the disease. It has been observed in clinical practice that SAPs are often accompanied by pancreatic tissue necrosis, dysfunction of the heart, lungs, kidneys, liver and other organs, and even multiple organ failure. Compared to MAP, SAP results in extended hospital stay, poor prognosis and higher mortality [3]. However, many studies on AP have revealed that a subset of patients with SAP experience a short duration of organ dysfunction, favorable clinical outcomes and a low mortality rate, owing to the medical advancements, while the others grapple with critical, persistent organ dysfunction, challenging treatment, and a relatively high mortality rate.

The 2012 Atlanta conference improved the AP severity classification [4] by further dividing SAP into moderately severe acute pancreatitis (MSAP) and SAP, while retaining the original classification for MAP. In this edition of severity classification, MAP is defined as absence of organ dysfunction and local or systemic complications;

MSAP involves transient organ dysfunction (less than 48 hours) and/or local or systemic complications; and SAP is characterized by persistent (more than 48 hours) organ dysfunction, with or without local or systemic complications. Given the stable clinical course of MAP, the corresponding treatment does not need to be overly aggressive. Diet management and moderate rehydration should be appropriately integrated into MAP treatment regimen. Patients with MAP usually recover within a relatively short period of time. As for MSAP, close monitoring and nutritional support are required for the more severe symptoms. Patients with MSAP require more potent drugs to alleviate pain, decrease inflammation and address complications. Despite the complications, MSAP generally has a good prognosis with a low fatality rate. As the most severe variant of the disorder, SAP is highly lethal and requires aggressive medical intervention such as mechanical ventilation, renal replacement therapy, or other organ support and long-term monitoring [4]. In some instances, surgery may be warranted to tackle infectious pancreatic necrosis, bleeding, or other serious complications. Although MAP is a mild, self-limiting disease, it may lead to organ dysfunction and deteriorate into either MSAP or SAP if pancreatic necrosis and secondary infection come into the picture, further complicating the disorder progression [5–7]. Moreover, distinguishing between patients with MSAP or SAP and those with MAP can be a challenging endeavor particularly in the early stage of the disease, for their early symptoms may be very similar, which necessitates close monitoring and additional clinical and laboratory evaluation for accurate classification of AP severity [8,9]. The overall condition of AP patients is often marked by rapid, ever-changing manifestations. At present, close monitoring is often used in clinical practice to assess the severity of AP in real time. With accurate prediction of the final AP severity at early stage, physicians can adjust the treatment regimen in time to forestall the deterioration of the disease, prevent the occurrence of complications, improve the prognosis, and reduce the unnecessary wastage of medical resources.

The clinical scoring systems used for AP are relatively complex. Most of the existing scoring systems are based on vital signs, laboratory indexes and imaging data. The scoring systems currently used in clinical practice for early assessment of AP severity include the Ranson score, Acute Physiology and Chronic Health Evaluation II (APACHE-II) score, bedside index for severity in acute pancreatitis (BISAP) score, and computed tomography severity index (CTSI) score. Although these scoring methods have been applied in clinical practice for a long time, it should be noted that these methods are obviously inadequate for feature extraction and it is difficult to integrate patients' baseline data, laboratory indexes, imaging and other multidimensional data with most of these methods. At present, the technical means used to process clinical data are subjected to continuous upgrades and enhancements. For instance, the majority of clinical data such as laboratory indexes and baseline information can be quantified or categorized, and subsequently analyzed using various classical machine learning (ML) algorithms such as support vector machine (SVM) and random forest (RF) [10,11]. The combination of computer vision and ML confers an obvious benefit in the processing of medical image data. For example, computer vision and ML can be applied to medical imaging to detect and diagnose diseases [12]. The algorithms can automatically identify and highlight abnormal areas in the image, assisting doctors in making more accurate judgments. ML algorithm can segment medical images and separate different structures or tissues in the image, a feature beneficial for further image processing analysis and disease diagnosis [13]. Computer vision and ML can also be leveraged in the automatic generation of medical image reports, thereby reducing the work burden of physicians [14]. It is worth noting that traditional computer vision generally relies on imagomics for feature extraction of imaging. Its essence lies in the application of multiple artificially designed feature filters to the original image, which are then quantified as features. The semantic extraction of these predefined feature filters is generally relatively low and not tailored to specific tasks [15]. Therefore, for diseases with a complex course like AP, an evaluation model based on scientifically grounded and accurate technical means that can integrate patients' multi-dimensional clinical information is urgently needed for clinical application.

Artificial intelligence (AI), a technology that simulate human intelligence, has achieved gradual maturation in recent years. Currently, AI is widely applied in speech recognition, image recognition, natural language processing, machine translation, medical diagnosis, financial prediction, etc. [16–21]. Deep learning, an ML technology based on artificial neural network (ANN), represents the key technology in the field of AI at the present stage [22]. ANN consists of multiple layers of neurons, each containing multiple neurons and forming connections with neurons in adjacent layers. The neural network receives data organized in the form of tensors through the input layer, and the final output is calculated and transformed by the internal network structure. The neural network backpropagates the error between the result of the forward propagation of data and the actual label, gradually adjusting the weights and biases of neurons to minimize the error, thereby optimizing the performance of the model [23]. Compared to traditional ML methods, ANN can handle more complex nonlinear relationships and possess stronger adaptability and generalization abilities. Its performance in processing complex tasks far exceeds that of other traditional ML models.

There are various deep learning (DL) architectures, such as convolutional neural network (CNN) autoencoder, deep belief network, and recurrent neural network (RNN). In medical diagnosis and prediction, AI utilizes different types of DL models to process various clinical variables

from patients and conduct subsequent analyses. For instance, in the case of numerical data such as laboratory indexes and vital signs, the feed-forward neural network based on multi-layer perceptron can leverage its powerful fitting ability to create a model equipped with advanced abstract feature representation ability through layer-by-layer calculation and training [24]. For image data such as medical images, after converting the images into a numerical matrix, CNN can extract and characterize the visual semantic features of the image layer by layer, in which the high-level semantics of the deep layer are derived from the low-level features of the former layer, enabling automatic adaptation for specific task. This method used for data feature learning by the model itself no longer relies on manually designed feature engineering used in traditional computer vision [25]. For time series data and text data, various types of RNNs also play important roles [26,27].

This study aimed to establish DL models that combine computed tomography (CT) images and various clinical indexes of AP patients to conduct multi-view analysis for predicting the severity classification of newly admitted AP patients at an early stage. In this study, deep neural network (DNN) was employed to process non-imaging clinical data of AP patients, while CNN was used to process AP images. It is worth noting that the image processing of AP is different from that of disorders in other organs. The peripancreatic manifestations are crucial and strongly correlated with the severity of AP. Therefore, the method of organ segmentation and further feature extraction is not suitable in the context of AP [28,29]. Thus, we applied CNN with attention mechanism to process AP imaging data, allowing the algorithm to automatically focus on the parts of the CT image related to AP severity [30]. At the same time, we combined the aforementioned two types of data to develop a classification prediction model for assessing AP severity.

## Materials and Methods

### *Data Collection*

The pancreatitis database in the platform of Clinical Data Research Center of Acute Abdominal Surgery in the First Affiliated Hospital of Dalian Medical University was used in this study. The relevant clinical data of AP patients collected in the database from January 2011 to November 2022 were retrospectively analyzed. The inclusion criteria are as follows: (i) the patient is 18 years of age or older, diagnosed with AP; (ii) the detention time outside the hospital is less than 3 days; and (iii) relevant clinical data are complete. Individuals with the following characteristics were excluded: (i) patients under immunosuppressed state (taking immunosuppressant drugs or caused by certain diseases); (ii) patients with malignant tumor or severe organ dysfunction; (iii) pregnant or lactating women; (iv) organ transplant recipients; (v) patients with traumatic pancreatitis; (vi) patients with chronic pancreatitis; (vii) patients with pancreatic space-occupying lesions; (viii) patients with a history of pancreatic surgery; and (ix) patients who discontinue treatment or transfer to another hospital during treatment.

The treatment methods for AP include the following:

(1) Medical treatment: fasting, inhibition of gastric acid secretion, fluid infusion, catharsis, oxygen inhalation, anti-infection, lipid-lowering, gastrointestinal decompression, inhibition of pancreatic enzyme activity and pancreatic secretion, correction of electrolyte disturbance and acid-base imbalance, nutritional support, blood purification, etc.

(2) Surgical treatment: patients with obstructive biliary pancreatitis are treated with endoscopic retrograde cholangiopancreatography, common bile duct exploration, and T-tube drainage to remove the obstruction; patients with non-obstructive biliary pancreatitis require surgery to remove the biliary cause after the disease is relieved; patients with infectious pancreatic necrosis should undergo treatment with catheter drainage and removal of necrotic tissue [5].

A total of 934 eligible cases (395 MAPs, 340 MSAPs, 199 SAPs) with complete clinical indexes and 1311 cases (665 MAPs, 438 MSAPs, 208 SAPs) with valid admission CT images were collected based on the inclusion and exclusion criteria. There were 772 cases (388 MAPs, 246 MSAPs, 138 SAPs) having both clinical data and admission CT images. Demographic information and vital signs of the patients were collected, including age, gender, systolic pressure, diastolic pressure, heart rate, body temperature, history of hypertension, history of diabetes, history of smoking, history of alcoholism, and history of AP. Laboratory indexes of patients within 48 hours after admission were collected: blood glucose (GLU), aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TBIL), white blood cell count (WBC), blood urea nitrogen (BUN), hematocrit (HCT), creatinine (CRE), amylase (AMY), lipase (LPS), hemoglobin (HB), platelet (PLT), albumin (ALB), blood calcium ($Ca^{2+}$), blood sodium ($Na^{+}$), blood potassium ($Ka^{+}$), blood chloride ($Cl^{-}$), blood pH, blood bicarbonate ($HCO_3{}^{-}$), base excess (BE) and oxygenation index (OI). Plain CT scans of AP patients at admission were also collected. The cases were divided into a training set and a validation set at a ratio of 4:1 for 5-fold cross-validation (CV). The flowchart of this study is depicted in Fig. 1.

### *AP-Related Clinical Criteria*

#### Diagnostic Criteria for AP

The diagnosis of AP must meet two of the following criteria: (i) persistent pain in the upper abdomen (acute, severe, and often radiates to the back); (ii) serum lipase or amylase level of at least 3 times higher than the upper limit of normal; (iii) imaging evidence showing the characteristic manifestations of AP [31].

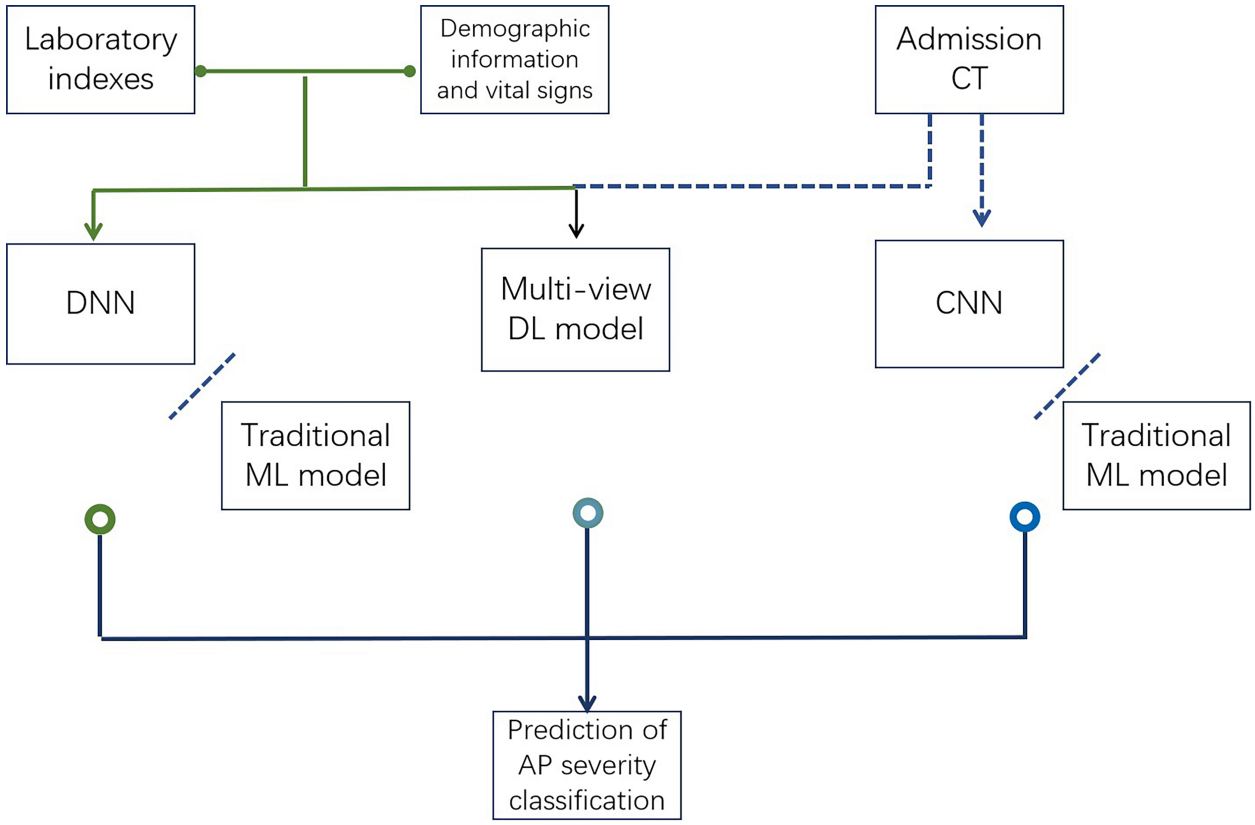


**Fig. 1. The flowchart depicting the process of establishing prediction models for different classifications of acute pancreatitis (AP) severity.** The deep neural network (DNN) model was developed using laboratory indexes, demographic information and vital signs as input. The convolutional neural network (CNN) model was developed using admission plain computed tomography (CT) scans as input. The multi-view deep learning (DL) model was established combining two types of data. DL models were compared to the corresponding traditional machine learning (ML) algorithms. At the same time, the three DL models were compared to each other.

AP-Related Organ Dysfunction

The function of respiratory system, cardiovascular system and kidney among the AP patients were evaluated. Organ dysfunction is defined according to the modified Marshall score [4] (Table 1).

Complications of AP

Local complications of AP include acute peripancreatic fluid collection (APFC), acute necrotic collection (ANC), pancreatic pseudocyst (PPC), and walled-off necrosis (WON). Systemic complications of AP, including preexisting comorbidity (such as coronary heart disease and chronic obstructive pulmonary disease), aggravated the condition of AP [4].

### *Statistical Analysis*

The measurement data complying with a normal distribution are expressed as mean ± standard error, and those not compliant with a normal distribution are expressed presented as median (25th percentile, 75th percentile). Analysis of variance (ANOVA) was used to analyze normally distributed variables that satisfied the homogeneity of variance among groups. Chi-square test was used to analyze data of nominal variables. Kruskal–Wallis rank-sum test was used for data analysis of non-normally distributed variables, which did not satisfy the homogeneity of variance among groups or ordinal variables. In addition, post-hoc Tukey test was utilized after ANOVA, whereas Nemenyi test was used for multiple comparisons following the rank-sum test. A significance level of $p < 0.05$ was considered statistically significant.

### *Data Processing*

Processing of Imaging Data

The representative slice from the admission CT scan and its adjacent upper and lower levels were fused into a 2.5-dimensional image. A wide window (window width [WW]: 300 Hu, window level [WL]: 40 Hu), a narrow window (WW: 100 Hu, WL: 40 Hu) and an intermediate window (WW: 200 Hu, WL: 40 Hu) were selected as the three channels of the input image according to the habits of imaging physicians for observing AP imaging, and the CT values were converted into pixel values ranging from 0 to 255. Subsequently, the horizontal contour of the abdomen was extracted using the level-set method, and the surrounding noise interference was removed (Fig. 2) [32].

**Table 1. Modified Marshall scoring system for organ dysfunction.**

| Organ system | Score (≥2 in any system defining the presence of organ failure) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| OI (mmHg) | >400 | 301–400 | 201–300 | 101–200 | ≤101 |
| CRE (μmol/L) | ≤134 | 134–169 | 170–310 | 311–439 | >439 |
| Systolic pressure (mmHg) | >90 | <90, fluid-responsive | <90, not fluid-responsive | <90, pH <7.3 | <90, pH <7.2 |

Abbreviations: CRE, creatinine; OI, oxygenation index.

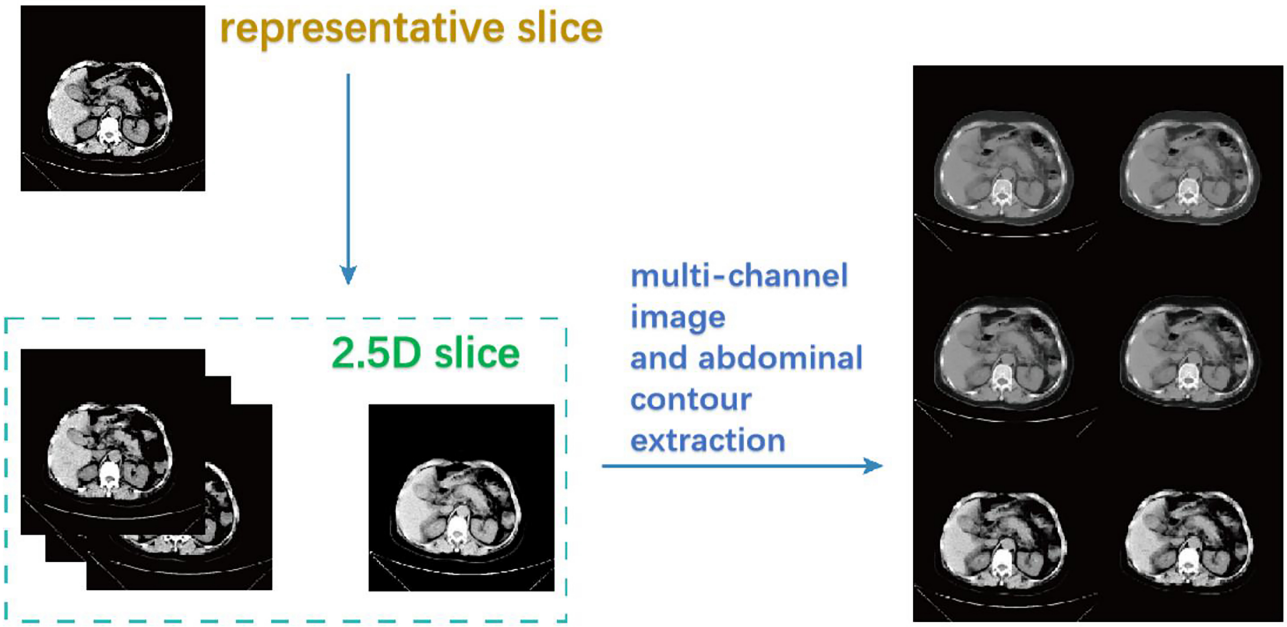


**Fig. 2. Processing of admission CT scans.** The representative CT slices and the adjacent slices were fused into a 2.5-dimensional slice. A wide window (window width (WW): 300 Hu, window level (WL): 40 Hu), a narrow window (WW: 100 Hu, WL: 40 Hu) and an intermediate window (WW: 200 Hu, WL: 40 Hu) derived from Digital Imaging and Communications in Medicine (DICOM) data were merged as a three-channel picture. Eventually, the level-set method was employed to extract the horizontal contour of the patient's abdomen and remove the surrounding noise interference.

Processing of Non-Imaging Clinical Data

To mitigate the impact of variable orders of magnitude and units, the training set data was standardized using z-scores, and the same standardization was applied to the validation set.

## *Model Construction*

Training Conditions

DL models were developed using the PyTorch framework (version 1.12.0) and implemented on 10 NVIDIA GeForce RTX3090 graphics cards. Samples of various severity levels were selected and divided into training and validation groups in proportion to ensure a balanced representation of categories.

The AP Severity Prediction Model with Clinical Indexes as Input

This section includes the DL model with clinical indexes as input and the SVM model as the baseline model. A total of 934 cases with complete clinical indexes were used for DNN modeling. The network consists of an input layer, two hidden layers, and an output layer. Grid search was applied for epochs in a range of 100–300 with a step of 100 and for initial learning rate in a range of 0.005–0.02 with a step of 0.005. According to the average accuracy, the training parameters were set as follows: batch size = 64, epochs = 300, initial learning rate = 0.01, and optimizer = stochastic gradient descent (SGD) with momentum. The SVM model was developed using scikit-learn, with a penalty coefficient C = 1 and a radial basis function kernel.

The AP Severity Prediction Model with Imaging Data as Input

The imaging data of 1311 AP patients with valid admission CT data were used to establish imaging models for predicting the severity of AP, including the Visual Geometry Group 16 (VGG16) framework with or without attention mechanism. The main structure of the VGG16 model is as follows:

(1) Input layer: The input layer is used to accept imaging data which are usually 224 × 224 pixel images.

(2) Convolutional layers: VGG16 contains 13 convolutional layers, which use small-sized 3 × 3 convolution kernels to conduct convolution operations on the input images for feature extraction. Convolutional layers at different depths can extract different levels of abstract features from the input data. Low-level features can be edges and textures, while higher-level features can be patterns that are more abstract and complex.

(3) Pooling layer: VGG16 uses a maximum pooling operation, using a 2 × 2 window to select the maximum value of each region to reduce the spatial resolution of the feature map, and retain the most prominent features in each region. This helps to reduce the complexity of the model and improve computational efficiency while preventing overfitting. Through the pooling operation, the receptive field of the network can be gradually increased. This allows the network to capture a larger range of features, helping to learn more global and abstract feature patterns.

(4) Fully connected layer: After the convolutional layer, VGG16 contains three fully connected layers, of which the first two fully connected layers have 4096 neurons, and the last fully connected layer outputs the corresponding number of neurons according to the number of categories of the classification task, which is used to map the high-level features extracted by the framework to specific categories.

(5) Activation function: Rectified linear unit (ReLU) is usually used as the activation function between each layer to bring in nonlinear factors, so that the neural network can learn and represent more complex relationships.

(6) Dropout layer: The fully connected layers of VGG16 typically include a dropout layer between them, forcing the network to be independent of specific neurons during training by randomly turning off neurons with a certain probability at training time. This helps to reduce the overfitting of the model to the training data and improve the generalizability of the model to the unseen data.

(7) Output layer: The output of the last fully connected layer is usually converted to class probabilities by the softmax function for image classification (Fig. 3) [33,34].

The shape and size of the pancreas vary from individual to individual. At the same time, there are many local complications of AP, which may only affect the pancreas or the surrounding tissues. The extent of peripheral pancreatic involvement varies significantly, leading to complex imaging manifestations. As a result, the traditional method of organ segmentation may no longer be applicable. Considering the characteristics of AP images, it is challenging to segment and extract regions of interest (ROIs). On the contrary, using the attention mechanism is more effective for automatically capturing critical regions and reducing the influence of irrelevant regions.

In this part, we integrated attention mechanism into DL network. We utilized global features and local features of different scales to generate local attention features, which were then employed for classification. This approach prompts the model to simultaneously focus on lesions of different scales (Fig. 4). At the same time, attention was visualized by displaying the heat map of the model's focus area [35]. The training parameters were set as the grid search methods described above: batch size = 64, epochs = 300, initial learning rate = 0.01, and optimizer = SGD with momentum.

### The AP Severity Prediction Model with Clinic Indexes and Imaging Data as Input

In this section, we constructed multi-view models composed of CNN with attention mechanism and DNN. The clinical data of 772 cases with complete set of data (both clinical data and admission CT images) were used for modeling. The framework of image-related attention mechanism model was VGG, and the training parameters were set as the grid search methods described above: batch size = 64, epochs = 300, initial learning rate = 0.01, and optimizer = SGD with momentum. The effect of the model was compared to that of the single-view model. Non-imaging clinical data and imaging data were input into the model through two separate paths. Then, the feature fusion was completed in the high semantic layer of CNN, and the combined features were classified by the subsequent fully connected layers (Fig. 5) [36].

## Result

### *Overview of Patients' Clinical Data*

Statistical test of clinical indexes was performed on 934 AP patients with complete clinical data. Table 2 displays the age, gender, medical history, including hypertension, diabetes, AP, smoking, and alcoholism, categorized by the severity of AP. The clinical indexes that demonstrated significant differences are presented in Table 3.

It can be seen that with the increase in the severity of AP, the clinical index of patients has different degrees of change, and the corresponding relationship between the changes of the patient's overall internal environment and the severity of AP needs to be further amplified with the help of ML models.

Subsequently, statistical test of Balthazar scores of 1311 AP patients' admission CT scans was conducted. The distribution of scores revealed significant differences ($\chi^2$ = 910.06, $p$ = 0.00) in CT imaging among MAP, MSAP and SAP, as shown in Table 4.

### *Performance of DL Models with Clinical Indexes as Input*

In this study, a MAP/MSAP/SAP three-classification DNN model was established. The overall accuracy, macro-

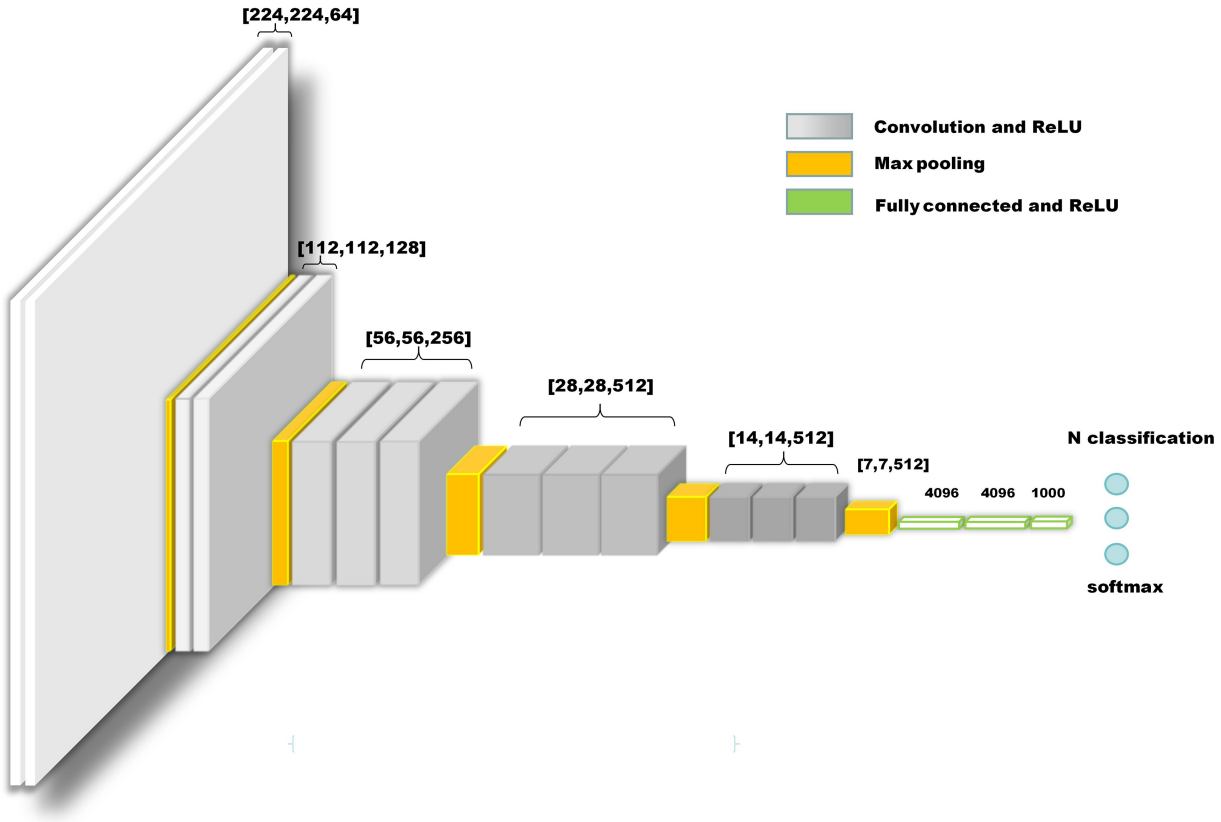


**Fig. 3. Model of convolutional neural network.** The network architecture consists of convolution layers, pooling layers, nonlinear activation functions, and other basic units stacked, and finally outputs the classification or regression results. The feature map size of the input in the network gradually decreases, while the number of channels gradually increases. ReLU, rectified linear unit.

**Table 2. Age, gender, history of hypertension, history of diabetes, history of AP, history of smoking and history of alcoholism of AP patients of different severity classifications.**

| Indexes | | MAP (n = 395) | SAP (n = 340) | AP (n = 199) | F/$\chi^2$ | *p* |
|---|---|---|---|---|---|---|
| Age | | 56 (39, 70) | 53 (38, 68) | 57 (36, 72) | 5.84 | 0.05 |
| Gender | Male (n = 569) | 242 | 201 | 126 | 0.964 | 0.62 |
| | Female (n = 365) | 153 | 139 | 73 | | |
| Hypertension history | Yes (n = 256) | 95 | 101 | 60 | 3.89 | 0.14 |
| | No (n = 678) | 300 | 239 | 139 | | |
| Diabetes history | Yes (n = 236) | 87 | 97 | 52 | 4.20 | 0.12 |
| | No (n = 698) | 308 | 243 | 147 | | |
| Smoking history | Yes (n = 137) | 48 | 53 | 36 | 4.09 | 0.13 |
| | No (n = 797) | 347 | 287 | 163 | | |
| Drinking history | Yes (n = 146) | 50 | 58 | 38 | 4.98 | 0.08 |
| | No (n = 788) | 345 | 282 | 161 | | |
| AP history | Yes (n = 173) | 79 | 59 | 35 | 1.00 | 0.61 |
| | No (n = 761) | 316 | 281 | 164 | | |

Abbreviations: MAP, mild acute pancreatitis; SAP, severe acute pancreatitis.

sensitivity, micro-sensitivity, macro-specificity and micro-specificity of the DNN model on the validation set were 71.83% (95% confidence interval (CI): 68.47%–75.19%), 70.54% (95% CI: 66.85%–74.24%), 71.83% (95% CI: 68.47%–75.19%), 85.09% (95% CI: 83.31%–86.88%) and 85.91% (95% CI: 84.24%–87.59%). The prediction accuracy for MAP was 86.08% (95% CI: 80.75%–91.41%), with a sensitivity of 86.08% (95% CI: 80.75%–91.41%), a specificity of 79.81% (95% CI: 73.59%–86.04%) and an area under the curve (AUC) of receiver operating character-

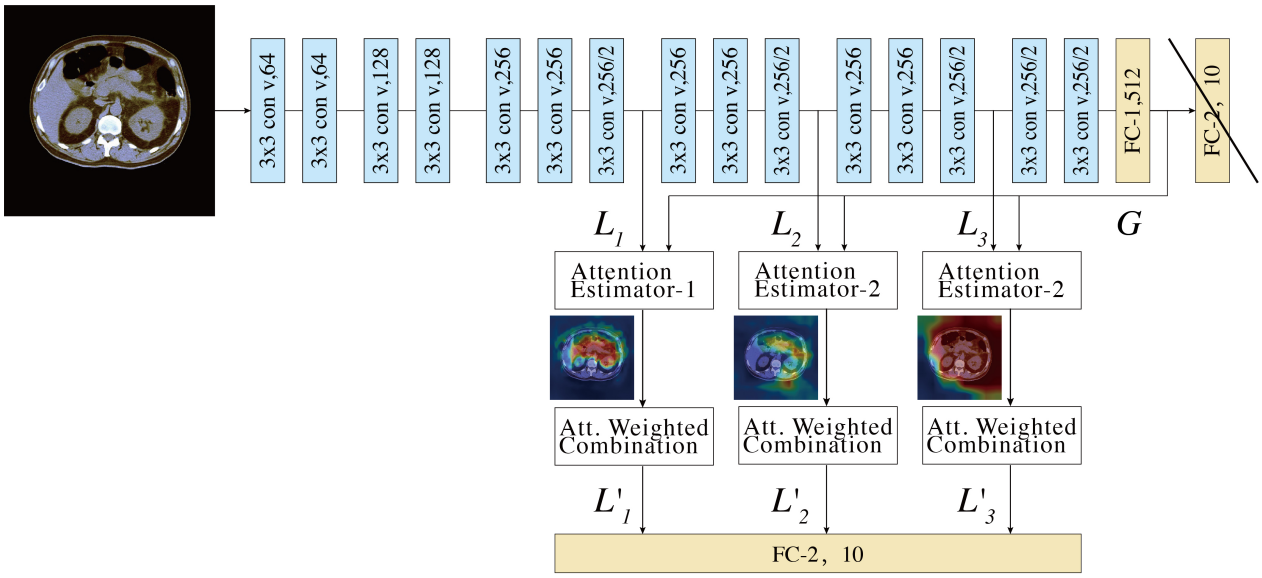


**Fig. 4. Convolutional neural networks with attention mechanisms.** The final classification layer fully connected (FC)-2 of the Visual Geometry Group (VGG) network is removed, and the output *G* of the fully connected layer FC-1 is regarded as the global feature. *L [*$L_1$, $L_2$, $L_3$*]* are the output of intermediate features at different scales in the VGG network (local features). The attention estimator integrates *L* and *G* as inputs and computes the attention map, which acts on each channel of $L_1$, $L_2$, $L_3$to obtain the weighted local feature *L' [*$L'_1$, $L'_2$, $L'_3$*]*. Finally, *L'* is fed into the fully connected layer FC-2 for final classification.

**Table 3. Overview of clinical indexes demonstrating significant differences between different AP severity classifications.**

| Indexes | MAP (n = 395) | MSAP (n = 340) | SAP (n = 199) | F/$\chi^2$ | *p* |
|---|---|---|---|---|---|
| WBC ($10^9$/L) | 11.73 (8.89, 15.6)$^{a,b}$ | 14.52 (11.48, 18.35) | 15.9 (10.69, 19.00) | 64.74 | <0.01 |
| ALB (g/L) | 41.17 ± 4.45$^{a,b}$ | 38.68 ± 5.37$^{b}$ | 35.58 ± 6.75 | 182.80 | <0.01 |
| $Ca^{2+}$ (mmol/L) | 2.24 (2.14, 2.32)$^{a,b}$ | 2.13 (1.94, 2.24)$^{b}$ | 1.77 (1.59, 1.95) | 320.80 | <0.01 |
| $K^{+}$ (mmol/L) | 3.98 (3.65, 4.21)$^{b}$ | 3.91 (3.60, 4.29)$^{b}$ | 4.80 (4.11, 5.80) | 20.71 | <0.01 |
| $Na^{+}$ (mmol/L) | 139.00 (136.00, 141.00)$^{a,b}$ | 138 (135.00, 141.00) | 137.00 (134.00, 140.00) | 32.53 | <0.01 |
| LPS (mmol/L) | 1884.00 (1091.00, 6956.00)$^{b}$ | 2460.00 (868.00, 6411.00) | 1819.00 (669.00, 4886.00) | 10.46 | <0.01 |
| CRE (μmoI/L) | 64.00 (53.00, 79.00)$^{b}$ | 64.00 (52.00, 83.75)$^{b}$ | 92.00 (61.00, 182.00) | 81.58 | <0.01 |
| BUN (mmol/L) | 4.98 (3.94, 6.37)$^{b}$ | 5.13 (3.95, 6.71)$^{b}$ | 7.0 (4.86, 12.48) | 73.26 | <0.01 |
| ALT (mmol/L) | 68.00 (28.00, 239.00)$^{b}$ | 47.50 (25.00, 192.50) | 40.00 (24.00, 124.00) | 11.88 | <0.01 |
| GLU (mmol/L) | 7.54 (6.20, 10.02)$^{a,b}$ | 9.61 (7.22, 14.04) | 10.47 (7.40, 14.91) | 78.72 | <0.01 |
| $PaO_2/FiO_2$ (mmHg) | 376.19 (342.86, 414.29)$^{ab}$ | 300.00 (260.00, 364.12)$^{b}$ | 245.71 (209.09, 283.33) | 346.23 | <0.01 |
| pH | 7.40 (7.38, 7.43)$^{b}$ | 7.40 (7.36, 7.43)$^{b}$ | 7.37 (7.31, 7.42) | 34.15 | <0.01 |
| $HCO_3^{-}$ (mmol/L) | 22.90 (21.70, 24.00)$^{a,b}$ | 22.30 (20.50, 23.80)$^{b}$ | 19.50 (15.90, 22.50) | 101.13 | <0.01 |
| BE (mmol/L) | –1.80 (–3.30, –0.50)$^{a,b}$ | –2.60 (–4.90, –0.70)$^{b}$ | –5.90 (–11.10, –2.10) | 97.97 | <0.01 |
| Systolic pressure (mmHg) | 134.00 (120.00, 145.00)$^{b}$ | 140.00 (120.00, 150.00)$^{b}$ | 129.00 (110.00, 140.00) | 24.94 | <0.01 |
| Diastolic pressure (mmHg) | 80.00 (73.00, 90.00) | 80.00 (77.00, 90.00)$^{b}$ | 80.00 (69.00, 90.00) | 8.23 | <0.01 |
| Heart rate (per min) | 80.00 (72.00, 88.00)$^{a,b}$ | 88.00 (78.00, 104.00)$^{b}$ | 110.00 (92.00, 126.00) | 213.55 | <0.01 |

Notes: $^{a}$ $p < 0.05$ compared with MSAP, $^{b}$ $p < 0.05$ compared with SAP. Abbreviations: MSAP, moderately severe acute pancreatitis; WBC, white blood cell count; ALB, albumin; LPS, lipase; CRE, creatinine; BUN, blood urea nitrogen; ALT, alanine aminotransferase; GLU, blood glucose; BE, base excess; MAP, mild acute pancreatitis; SAP, severe acute pancreatitis.

istic (ROC) of 0.89 (95% CI: 0.87–0.92). The MSAP prediction accuracy was 57.35% (95% CI: 49.50%–65.21%), with a sensitivity of 57.35% (95% CI: 49.50%–65.21%), a specificity of 81.86% (95% CI: 78.33%–85.40%) and an AUC of 0.75 (95% CI: 0.72–0.78). For SAP, the accuracy was 68.21% (95% CI: 58.50%–77.91%), with the sensitivity, specificity and AUC being 68.21% (95% CI: 58.50%–77.91%), 93.61% (95% CI: 91.10%–96.11%) and 0.91 (95% CI: 0.88–0.95), respectively. As a comparison model, the three-classification SVM model was con-

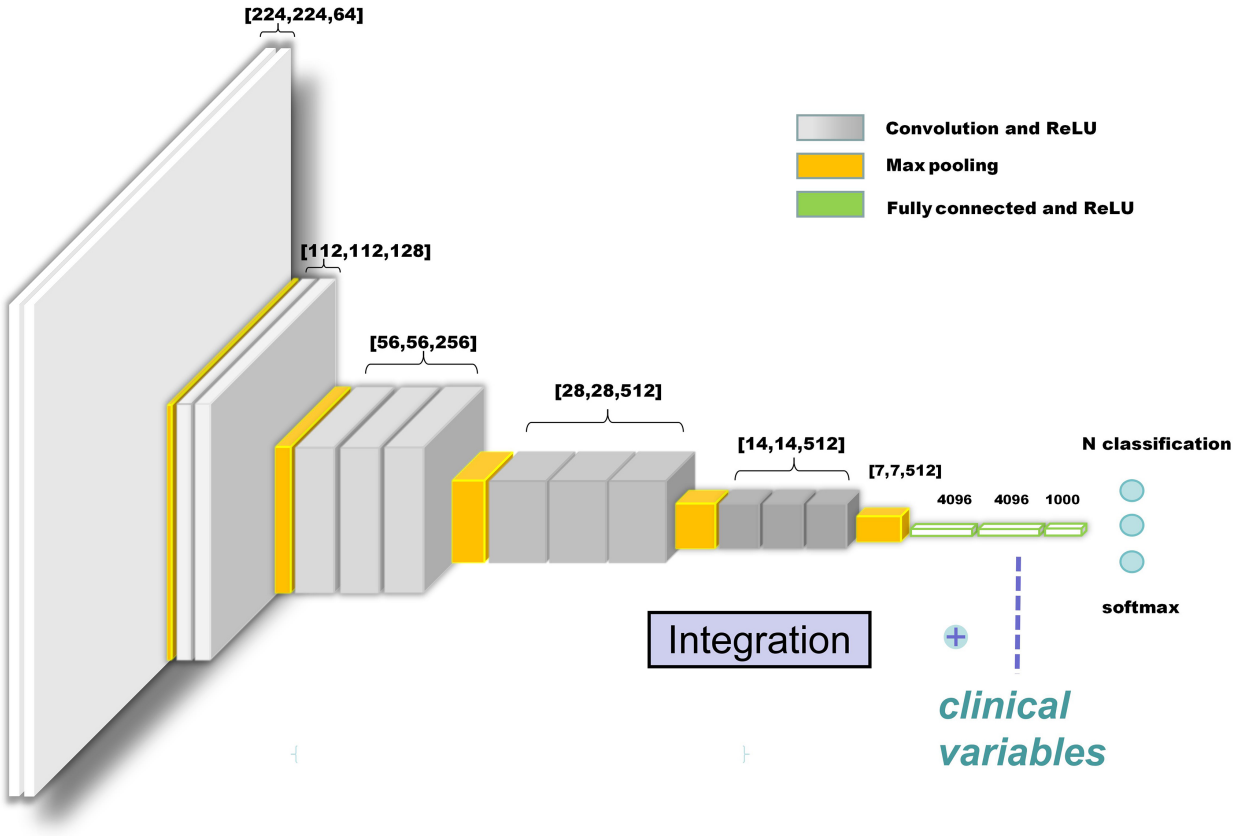


**Fig. 5. Multi-view DL model constructed using clinical indexes and imaging data for predicting severity classification of AP.** Image data were propagated through the network using the conventional CNN methods, and then integrated with clinical data at a higher semantic layer to collectively predict the output.

**Table 4. Balthazar scores of admission CT of MAP/MSAP/SAP.**

| Severity | Score | | | | | Total |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | |
| MAP[a,b] | 207 | 158 | 300 | 0 | 0 | 665 |
| MSAP[b] | 0 | 6 | 72 | 206 | 154 | 438 |
| SAP | 0 | 0 | 10 | 31 | 167 | 208 |
| Total | 207 | 164 | 382 | 237 | 321 | 1311 |

Notes: [a] $p < 0.05$ compared with MSAP; [b] $p < 0.05$ compared with SAP. Score 0: normal pancreas and adjacent tissues; Score 1: localized or diffuse enlargement of the pancreas; Score 2: inflammatory infiltration involving peripancreatic fat; Score 3: single pancreatic or peripancreatic fluid collection in addition to pancreatic disease; Score 4: presence of two or more areas of fluid and gas accumulation in or around the pancreas.

structed with an overall accuracy of 69.03% (95% CI: 65.81%–72.25%), a macro-sensitivity of 68.05% (95% CI: 65.33%–70.76%), a micro-sensitivity of 69.03% (95% CI: 65.81%–72.25%), a macro-specificity of 83.68% (95% CI: 82.16%–85.20%), and a micro-specificity of 84.52% (95% CI: 82.90%–86.13%). For MAP, the prediction accuracy was 87.85% (95% CI: 84.78%–90.91%), with a sensitivity of 87.85% (95% CI: 84.78%–90.91%), a specificity of 76.64% (95% CI: 73.25%–80.02%), and an AUC of 0.90 (95% CI: 0.88–0.93). For MSAP, the prediction accuracy was 47.06% (95% CI: 38.59%–55.53%), with a sensitivity of 47.06% (95% CI: 38.59%–55.53%), a specificity of 82.71% (95% CI: 81.51%–83.91%), and an AUC of 0.74 (95% CI: 0.70–0.77). In the case of SAP, the prediction accuracy was 69.23% (95% CI: 63.27%–75.19%), with the sensitivity, specificity and AUC being 69.23% (95% CI: 63.27%–75.19%), 91.70% (95% CI: 88.47%–94.94%), and 0.93 (95% CI: 0.91–0.96), respectively (Fig. 6).

Meanwhile, MAP/MSAP binary classification models were also established in this study. The overall accuracy of the DNN binary classification model in the validation set was 80.82% (95% CI: 79.72%–81.92%), with a macro-sensitivity of 80.45% (95% CI: 79.53%–81.37%), a micro-sensitivity of 80.82% (95% CI: 79.72%–81.92%), a macro-specificity of 80.45% (95% CI: 79.53%–81.37%), a micro-specificity of 80.82% (95% CI: 79.72%–81.92%), and an AUC of 0.86 (95% CI: 0.84–0.87). The prediction accuracies for MAP and MSAP were 85.32% (95% CI: 80.65%–

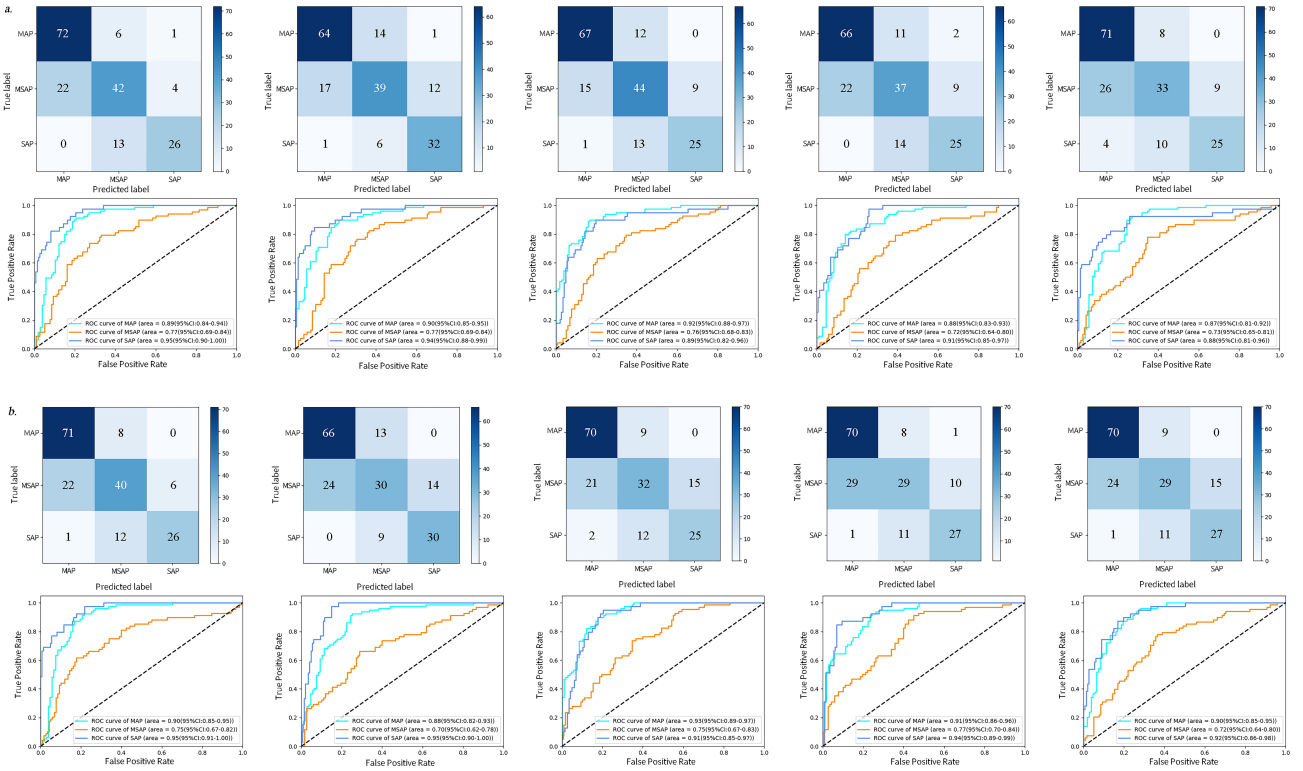


**Fig. 6. Confusion matrices and receiver operating characteristic (ROC) curves of three-classification models with clinical indexes as input.** (a) Confusion matrices and ROC curves of MAP/MSAP/SAP three-classification DNN model with clinical indexes as input (5-fold cross-validation (CV)). (b) Confusion matrices and ROC curves of MAP/MSAP/SAP three-classification support vector machine (SVM) model with clinical indexes as input (5-fold CV).

89.98%) and 75.59% (95% CI: 71.80%–79.37%), respectively. The overall accuracy of the SVM comparison model was 78.23% (95% CI: 74.70%–81.76%), with a macro-sensitivity of 76.00% (95% CI: 74.18%–81.35%), a micro-sensitivity of 78.23% (95% CI: 74.70%–81.76%), a macro-specificity of 77.76% (95% CI: 74.17%–81.34%), a micro-specificity of 78.23% (95% CI: 74.70%–81.76%), and an AUC of 0.85 (95% CI: 0.82–0.88). Specifically, the accuracy of MAP prediction was 84.05% (95% CI: 79.80%–88.30%), and the accuracy of MSAP prediction was 71.47% (95% CI: 66.05%–76.89%) (Fig. 7).

As for MSAP/SAP binary classification models, the overall accuracy of the model based on DNN in the validation set was 82.43% (95% CI: 78.82%–86.04%), with a macro-sensitivity of 79.83% (95% CI: 75.04%–84.62%), a micro-sensitivity of 82.43% (95% CI: 78.82%–86.04%), a macro-specificity of 79.83% (95% CI: 75.04%–84.62%), a micro-specificity of 82.43% (95% CI: 78.82%–86.04%), an AUC of 0.87 (95% CI: 0.83–0.91). The prediction accuracy for MSAP was 88.24% (95% CI: 84.80%–91.68%), while the prediction accuracy for SAP was 68.77% (95% CI: 62.38%–75.16%). As for comparison model, the overall accuracy of SVM was 79.25% (95% CI: 74.23%–84.27%), with a macro-sensitivity of 76.02% (95% CI: 71.17%–80.88%), a micro-sensitivity of 79.25% (95% CI: 74.24%–84.27%), a macro-specificity of 76.02% (95% CI: 71.17%–80.88%), a micro-specificity of 79.25% (95% CI: 74.24%–84.27%), and an AUC of 0.88 (95% CI: 0.83–0.92). The prediction accuracy for MSAP was 87.94% (95% CI: 80.37%–95.51%), whereas the prediction accuracy for SAP was 70.26% (95% CI: 59.09%–81.42%) (Fig. 8).

### *Performance of DL Models with Imaging Data as Input*

In this part of analysis, 1311 cases with valid admission CT scans were divided into training and validation sets in a ratio of 4:1 to establish DL models with or without attention mechanism.

For VGG-based attention mechanism models, the overall accuracy of the three-classification model in the validation set was 71.19% (95% CI: 68.78%–73.60%), with the macro-sensitivity, micro-sensitivity, macro-specificity, and micro-specificity being 64.29% (95% CI: 62.81%–65.78%), 71.19% (95% CI: 68.78%–73.60%), 83.75% (95% CI: 82.14%–85.36%), and 85.59% (95% CI: 84.39%–86.80%), respectively. The prediction accuracies for MAP, MSAP, and SAP were 84.96% (95% CI: 76.38%–93.54%), 62.07% (95% CI: 44.19%–79.95%), and 45.85% (95% CI: 36.42%–55.29%), respectively. The sensitivities of MAP,

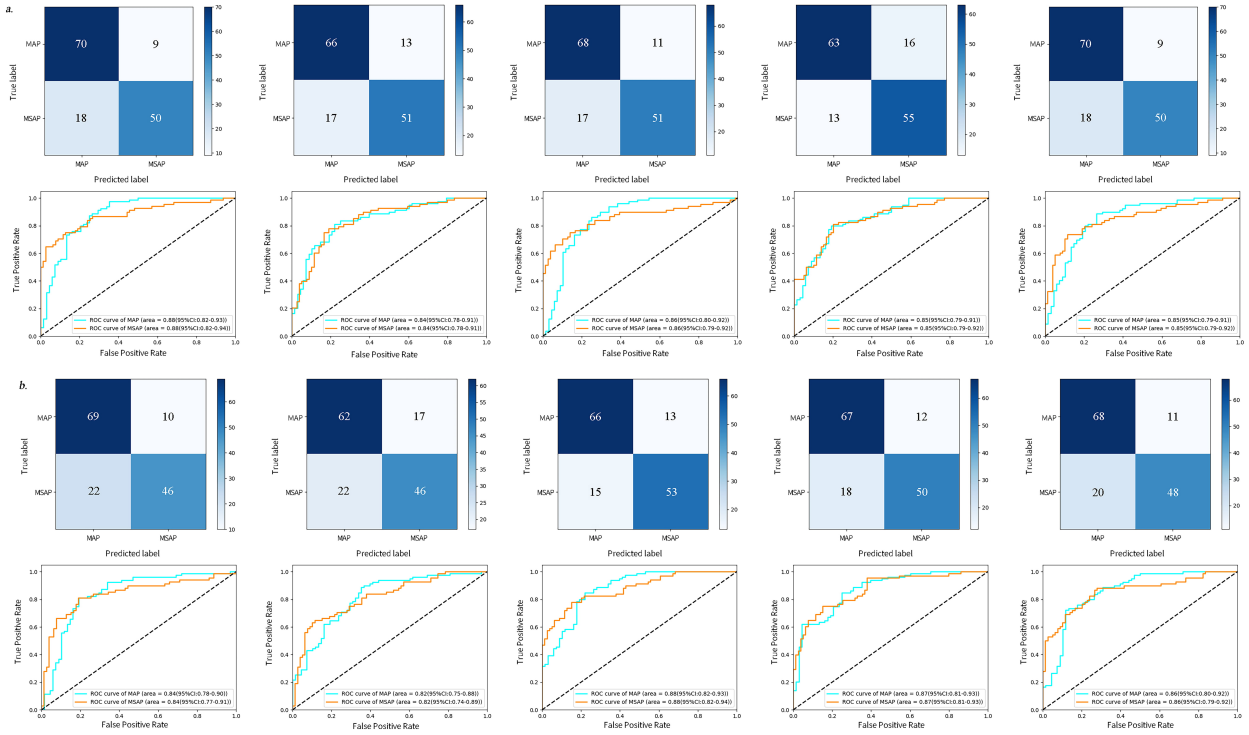


**Fig. 7. Confusion matrices and ROC curves of MAP/MSAP binary classification models with clinical indexes as input.** (a) Confusion matrices and ROC curves of MAP/MSAP binary classification DNN model with clinical indexes as input (5-fold CV). (b) Confusion matrices and ROC curves of MAP/MSAP binary classification SVM model with clinical indexes as input (5-fold CV).

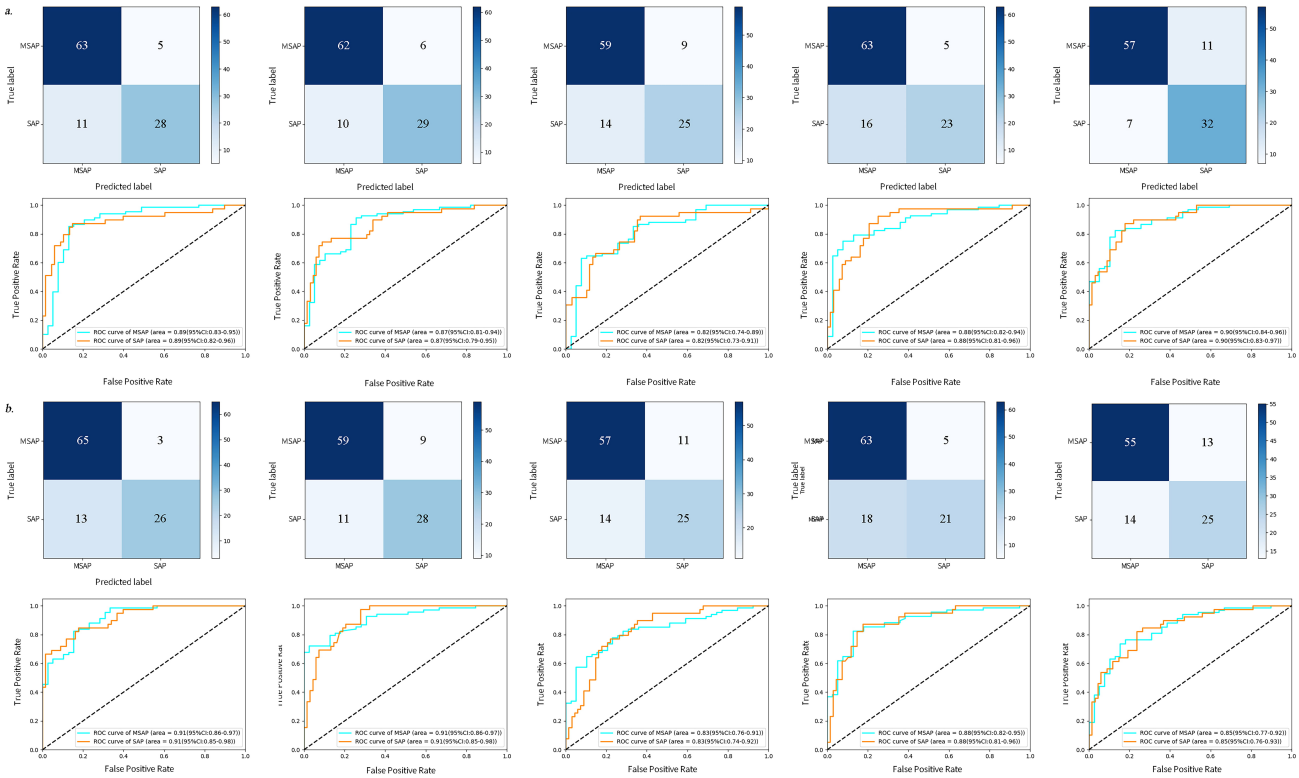


**Fig. 8. Confusion matrices and ROC curves of MSAP/SAP binary classification models with clinical indexes as input.** (a) Confusion matrices and ROC curves of MSAP/SAP binary classification DNN model with clinical indexes as input (5-fold CV). (b) Confusion matrices and ROC curves of MSAP/SAP binary classification SVM model with clinical indexes as input (5-fold CV).

MSAP, and SAP were 84.96% (95% CI: 76.38%–93.54%), 62.07% (95% CI: 44.19%–79.95%), and 45.85% (95% CI: 36.42%–55.29%), respectively. The specificities of MAP, MSAP, and SAP were 75.78% (95% CI: 65.13%–86.43%), 78.74% (95% CI: 70.84%–86.63%), and 96.73% (95% CI: 94.97%–98.48%). On the other hand, the AUCs of MAP, MSAP, and SAP were 0.88 (95% CI: 0.87–0.90), 0.77 (95% CI: 0.76–0.79), and 0.88 (95% CI: 0.84–0.92).

The overall accuracy of MAP/MSAP binary classification model in the validation set was 78.82% (95% CI: 74.43%–83.20%), with a macro-sensitivity of 77.47% (95% CI: 71.43%–83.51%), a micro-sensitivity of 78.82% (95% CI: 74.43%–83.20%), a macro-specificity of 77.47% (95% CI: 71.43%–83.51%), a micro-specificity of 78.82% (95% CI: 74.43%–83.20%), and an AUC of 0.86 (95% CI: 0.83–0.88). The prediction accuracy for MAP was 83.91% (95% CI: 77.49%–90.33%), while the prediction accuracy for MSAP was 71.03% (95% CI: 55.13%–86.94%). For MSAP/SAP binary classification model, the overall accuracy in the validation set was 78.91% (95% CI: 76.85%–80.96%), with a macro-sensitivity of 71.46% (95% CI: 68.43%–74.49%), a micro-sensitivity of 78.91% (95% CI: 76.85%–80.97%), a macro-specificity of 71.46% (95% CI: 68.43%–74.49%), a micro-specificity of 78.91% (95% CI: 76.85%–80.97%), and an AUC of 0.75 (95% CI: 0.67–0.83). The prediction accuracies for MSAP and SAP were 92.18% (95% CI: 88.22%–96.14%) and 50.73% (95% CI: 42.61%–58.86%) (Fig. 9).

For VGG-based attention-free models, the overall accuracy, macro-sensitivity, micro-sensitivity, macro-specificity, and micro-specificity of the three-classification model in the validation set were 69.66% (95% CI: 68.17%–71.14%), 60.39% (95% CI: 56.19%–64.59%), 69.66% (95% CI: 68.17%–71.14%), 82.50% (95% CI: 82.03%–82.97%), and 84.83% (95% CI: 84.09%–85.57%). The prediction accuracies for MAP, MSAP, and SAP were 87.22% (95% CI: 76.51%–97.92%), 59.31% (95% CI: 44.46%–74.16%), and 34.63% (95% CI: 14.61%–54.65%), respectively. The sensitivities of MAP, MSAP, and SAP were 87.22% (95% CI: 76.51%–97.92%), 59.31% (95% CI: 44.46%–74.16%), and 34.63% (95% CI: 14.62%–54.65%), respectively. The specificities of MAP, MSAP, and SAP were 71.25% (95% CI: 61.71%–80.79%), 78.62% (95% CI: 66.71%–90.54%), and 97.64% (95% CI: 94.61%–100.00%). In addition, the AUCs of MAP, MSAP, and SAP were 0.87 (95% CI: 0.85–0.89), 0.63 (95% CI: 0.54–0.72), and 0.88 (95% CI: 0.85–0.91).

The overall accuracy, macro-sensitivity, micro-sensitivity, macro-specificity, micro-specificity, and AUC of MAP/MSAP binary classification model in the validation set were 78.45% (95% CI: 75.25%–81.66%), 77.21% (95% CI: 73.69%–80.73%), 78.45% (95% CI: 75.25%–81.66%), 77.21% (95% CI: 73.69%–80.73%), 78.45% (95% CI: 75.25%–81.66%), and 0.82 (95% CI: 0.73–0.90). The prediction accuracies for MAP and MSAP were 83.16% (95% CI: 74.59%–91.73%) and 71.26% (95% CI: 59.80%–82.73%), respectively. The overall accuracy, macro-sensitivity, micro-sensitivity, macro-specificity, micro-specificity, and AUC of MSAP/SAP binary classification model in the validation set were 75.78% (95% CI: 71.08%–80.48%), 67.10% (95% CI: 57.67%–76.52%), 75.78% (95% CI: 71.08%–80.48%), 67.10 (95% CI: 57.67%–76.52%), 75.78% (95% CI: 71.08%–80.48%), and 0.77 (95% CI: 0.66–0.88). The accuracy of MSAP prediction was 91.26% (95% CI: 82.20%–100.00%), whereas the accuracy of SAP prediction was 42.93% (95% CI: 17.35%–68.50%) (Fig. 10).

### *Performance of DL Models with Clinical Indexes and Imaging Data as Input*

Fusion multi-view models with attention mechanism, harmonizing both clinical indexes and imaging data as input, were developed in this study.

The overall accuracy, macro-sensitivity, micro-sensitivity, macro-specificity, and micro-specificity of the three-classification model were 80.26% (95% CI: 79.58%–80.94%), 77.38% (95% CI: 75.76%–79.01%), 80.26% (95% CI: 79.58%–80.94%), 89.48% (95% CI: 88.67%–90.29%), and 90.13% (95% CI: 89.79%–90.47%). The accuracy rates of MAP, MSAP, and SAP were 91.69% (95% CI: 87.80%–95.57%), 64.90% (95% CI: 58.85%–70.95%), and 75.56% (95% CI: 68.58%–82.53%). The sensitivities of MAP, MSAP, and SAP were 91.69% (95% CI: 87.80%–95.58%), 64.90% (95% CI: 58.85%–70.95%), and 75.56% (95% CI: 68.58%–82.53%). The specificities of MAP, MSAP, and SAP were 86.05% (95% CI: 79.66%–92.45%), 88.27% (95% CI: 85.55%–90.99%), and 94.13% (95% CI: 90.96%–97.29%). The AUCs of MAP, MSAP, and SAP were 0.96 (95% CI: 0.96–0.96), 0.76 (95% CI: 0.72–0.81), and 0.96 (95% CI: 0.95–0.98).

The results revealed that this model is more accurate than the two previously described single-view models (Fig. 11). The heatmap of attention for this model was plotted (Fig. 12).

An MAP/MSAP binary classification neural network was also built in this study. Its overall accuracy, macro-sensitivity, micro-sensitivity, macro-specificity, mico-specificity, and AUC of the multi-view binary classification model in the validation set were 87.14% (95% CI: 84.69%–89.60%), 86.44% (95% CI: 83.79%–89.09%), 87.14% (95% CI: 84.68%–89.60%), 86.44% (95% CI: 83.79%–89.09%), 87.14% (95% CI: 84.68%–89.60%), and 0.92 (95% CI: 0.91–0.94). The prediction accuracy for MAP was 89.61% (95% CI: 85.83%–93.39%), and the prediction accuracy for MSAP was 83.27% (95% CI: 77.77%–88.76%) (Fig. 13). Meanwhile, an MSAP/SAP binary classification neural network was also established. Its overall accuracy, macro-sensitivity, micro-sensitivity, macro-specificity, micro-specificity, and AUC of the multi-view binary classification model in the validation set

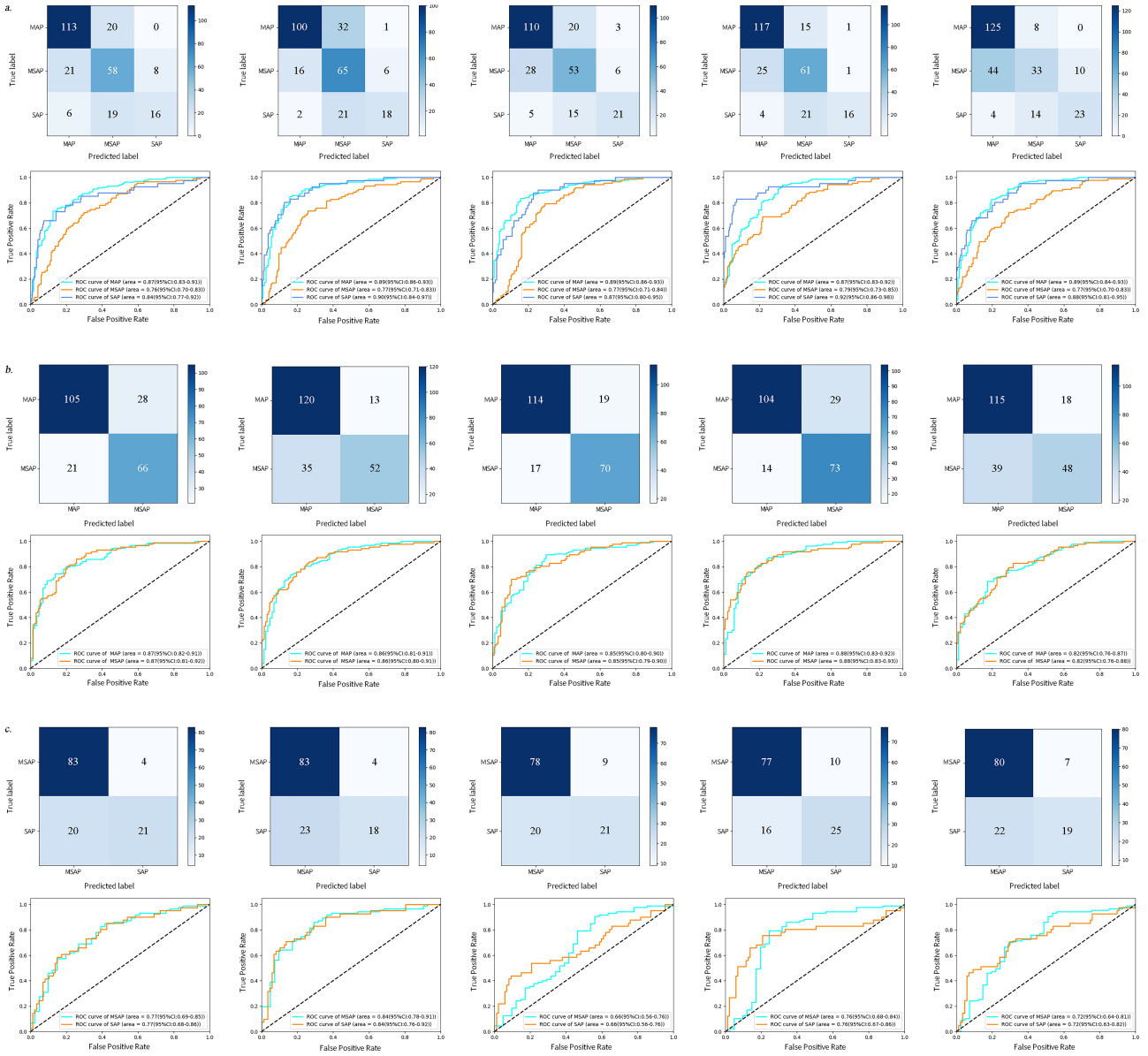

**Fig. 9. Confusion matrices and ROC curves of attention mechanism models with imaging data as input.** (a) Confusion matrices and ROC curves of MAP/MSAP/SAP three-classification VGG-based attention mechanism model with imaging data as input (5-fold CV). (b) Confusion matrices and ROC curves of MAP/MSAP binary classification VGG-based attention mechanism model with imaging data as input (5-fold CV). (c) Confusion matrices and ROC curves of MSAP/SAP binary classification VGG-based attention mechanism model with imaging data as input (5-fold CV).

were 84.74% (95% CI: 80.48%–89.00%), 81.68% (95% CI: 76.17%–87.19%), 84.74% (95% CI: 80.48%–89.00%), 81.68% (95% CI: 76.17%–87.19%), 84.74% (95% CI: 80.48%–89.00%), and 0.91 (95% CI: 0.87–0.94). The prediction accuracies for MSAP and SAP were 92.24% (95% CI: 85.21%–99.28%) and 71.11% (95% CI: 57.55%–84.68%) (Fig. 13).

Calibration plots for the three-classification DL models with imaging data, clinic indexes and merged data as input are shown in Fig. 14. Among them, the MAP's average integrated discrimination improvement (IDI) of multi-view model compared with imaging data model and clinic indexe model were 0.33 and 0.22. The MSAP's average IDI of multi-view model compared with imaging data model and clinic indexes model were 0.27 and 0.19. The SAP's average IDI of multi-view model compared with imaging data model and clinic indexes model were 0.36 and 0.15. These results showed that multi-view DL model was more accurate at predicting severity classification of AP.

## Discussion

At present, the commonly used scoring systems for assessing the severity and prognosis of AP in clinical prac-

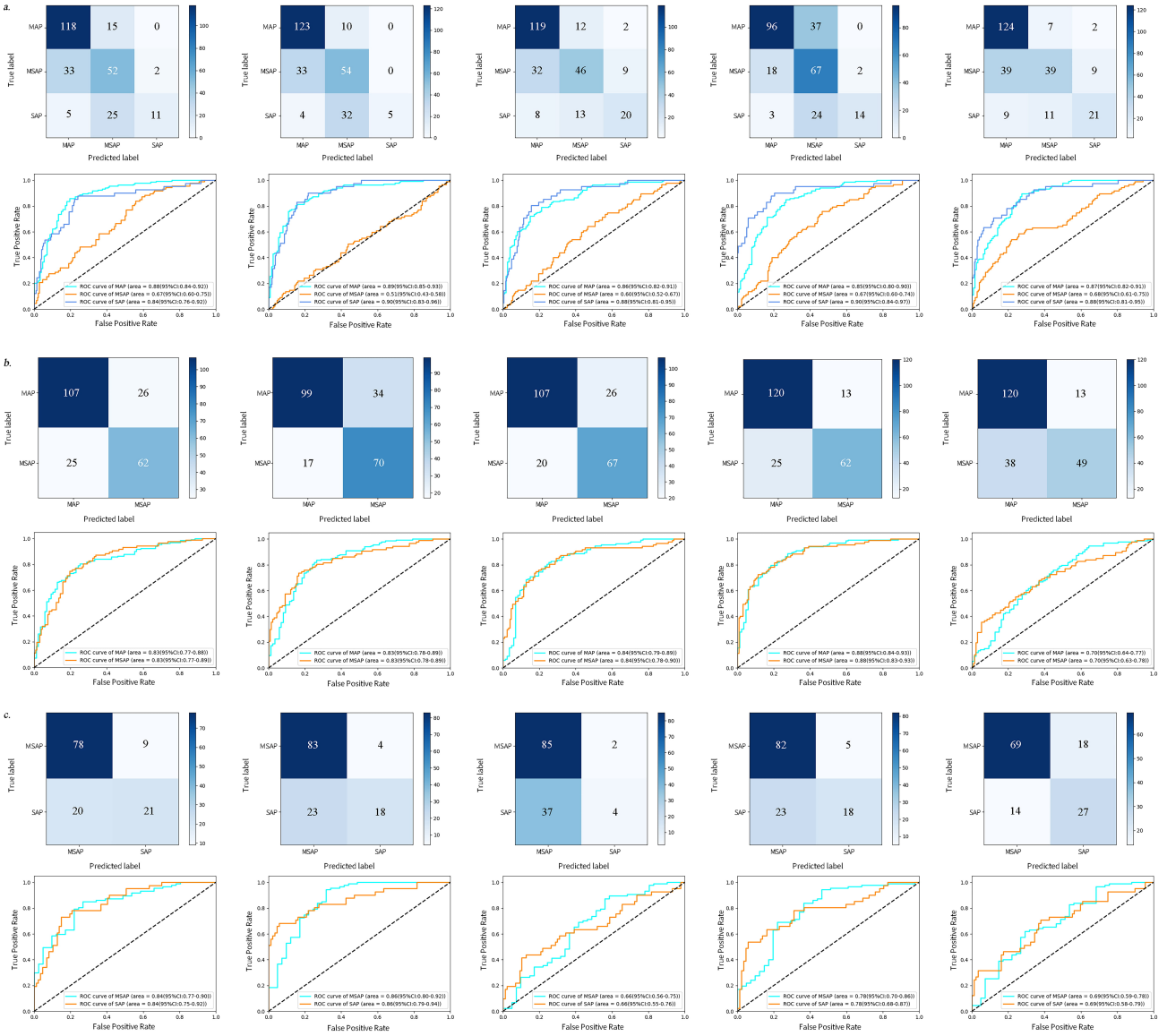


**Fig. 10. Confusion matrices and ROC curves of attention-free models with imaging data as input.** (a) Confusion matrices and ROC curves of MAP/MSAP/SAP three-classification VGG-based attention-free model with imaging data as input (5-fold CV). (b) Confusion matrices and ROC curves of MAP/MSAP binary classification VGG-based attention-free mechanism model with imaging data as input (5-fold CV). (c) Confusion matrices and ROC curves of MSAP/SAP binary classification VGG-based attention-free mechanism model with imaging data as input (5-fold CV).

tice include Ranson score, APACHE-II score, BISAP score, Balthazar score, CTSI score, and Sequential Organ Failure Assessment (SOFA) score [37–41]. Ranson score, which is the first AP scoring system created, assesses AP based on patient condition changes in the first 48 hours after admission and is determined by 11 objective indexes, including five indexes at admission (age, WBC, GLU, serum lactate dehydrogenase, and AST) and six indexes at 48 hours after admission ($Ca^{2+}$, HCT, BUN, arterial oxygen saturation, acid-base balance, and fluid loss) [42]. Several limitations of Ranson score should be acknowledged. First, it is of no considerable value in the dynamic monitoring of patients' health status. Secondly, the Ranson score does not take into account previous health status of patients, undermining the accuracy of AP prognosis determined by patients' overall health status [43–45]. The APACHE-II score is a widely used scoring system in intensive care units and can also be utilized to assess the severity of AP [46,47]. This scoring system is built upon an acute physiology score, an age score, and a chronic health score, providing a comprehensive and accurate assessment for a relatively accurate prediction of AP prognosis, including survival rate and length of hospital stay [48–53]. Although such assessment enables the making of more informed treatment decisions by healthcare professionals, the necessity to gather data for multiple indexes poses an inconvenience in clinical practice, espe-

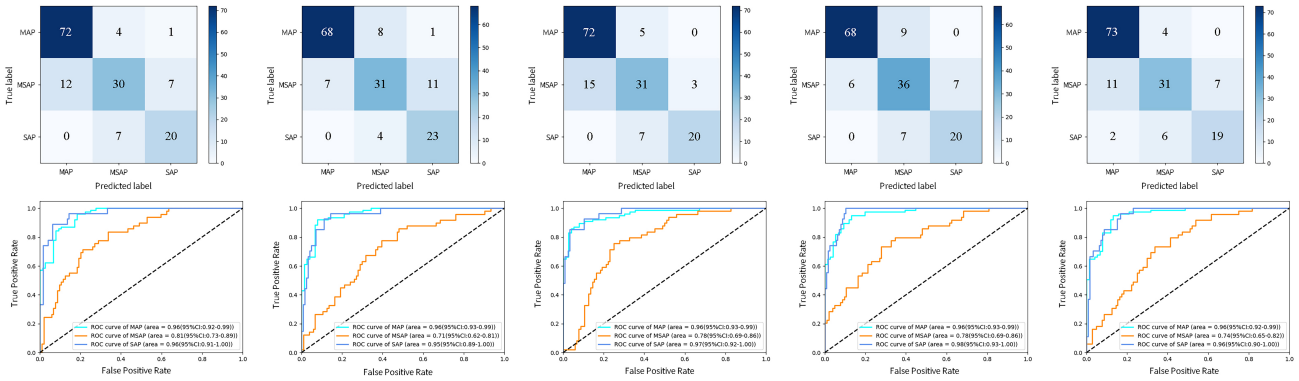


**Fig. 11. Confusion matrices and ROC curves of MAP/MSAP/SAP three-classification multi-view model with clinical indexes and imaging data as input (5-fold CV).**

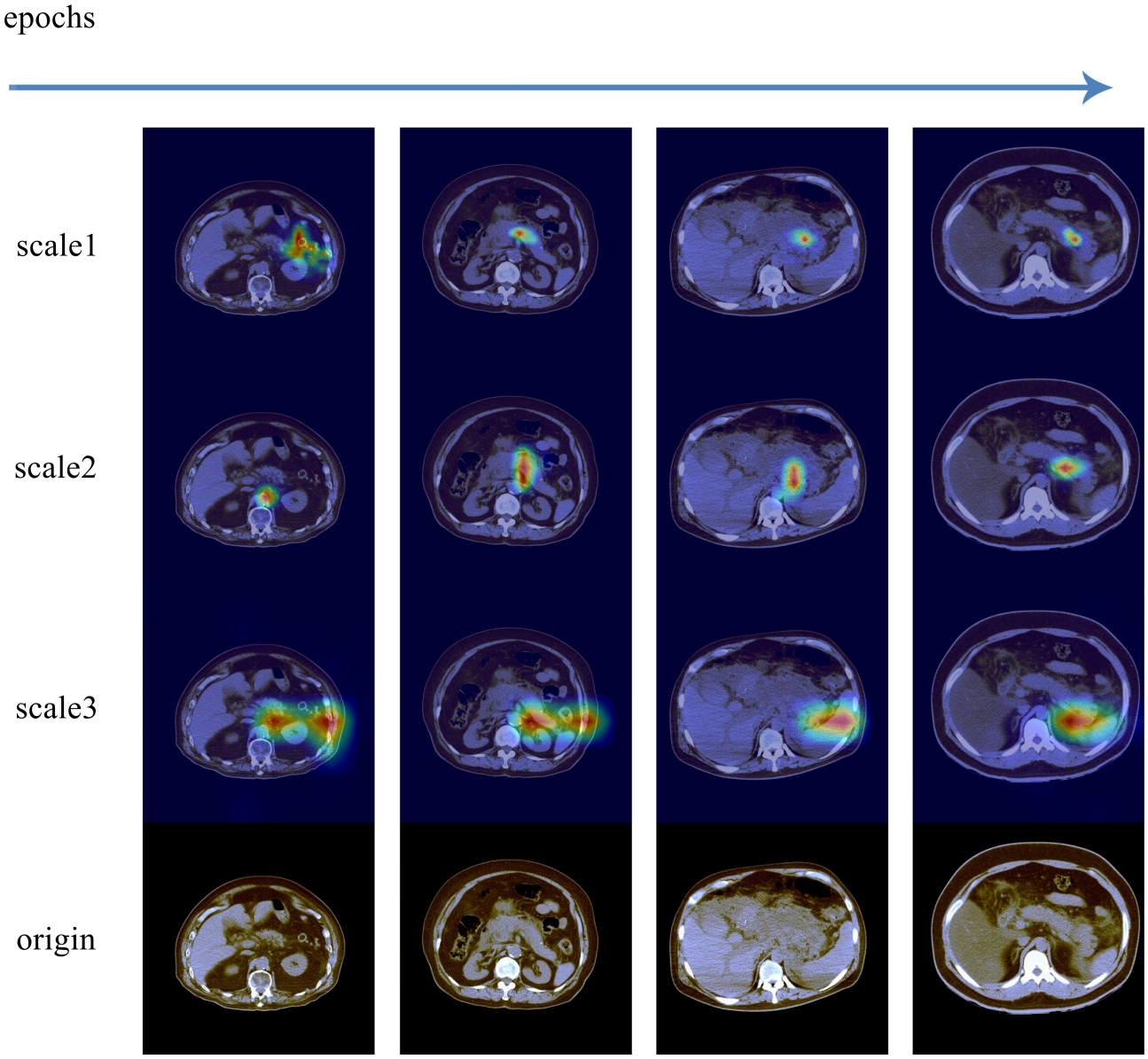


**Fig. 12. Heatmap of attention.** The heatmap of attention for CNN model was drawn on three different scales, and the attention of network was fixed on some local area with the increase of the number of training epochs.

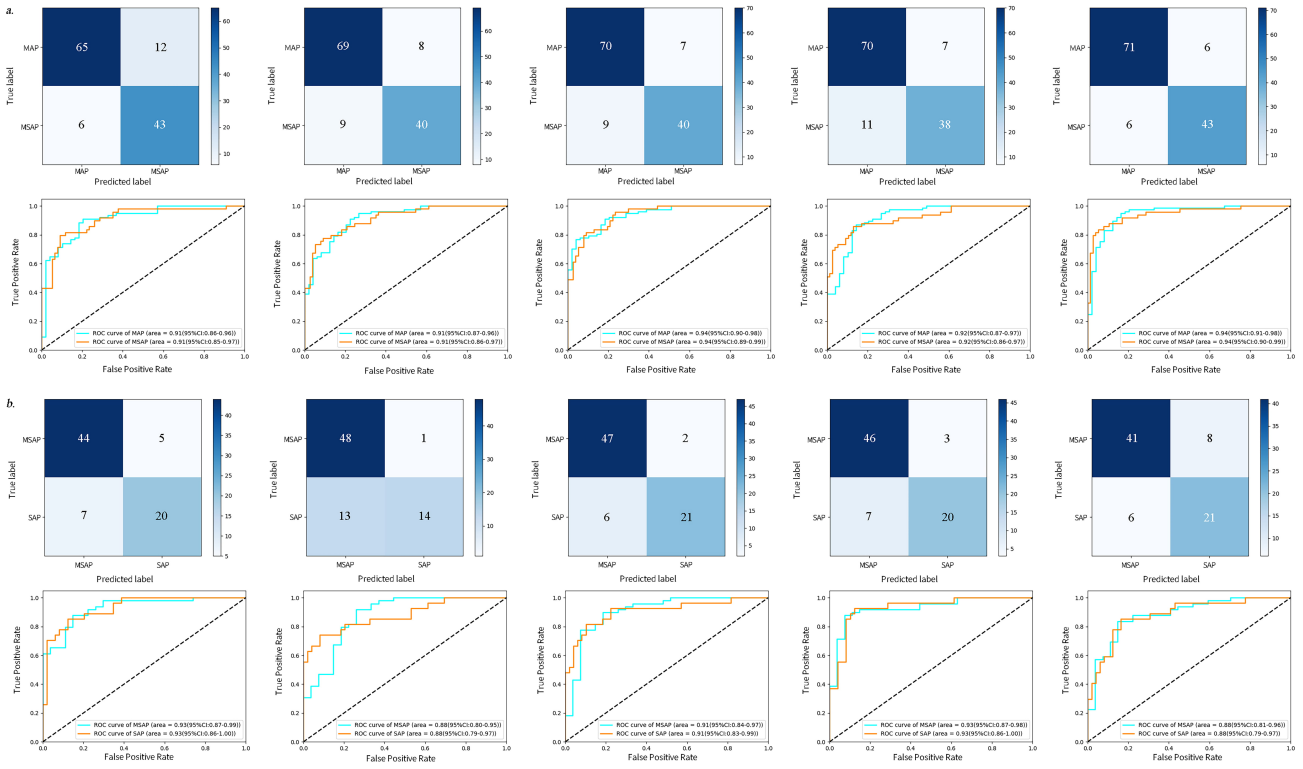


**Fig. 13. Confusion matrices and ROC curves of binary classification multi-view models with clinical indexes and imaging data as input.** (a) Confusion matrices and ROC curves of MAP/MSAP binary classification multi-view model with merged data as input (5-fold CV). (b) Confusion matrices and ROC curves of MSAP/SAP binary classification multi-view model with merged data as input (5-fold CV).

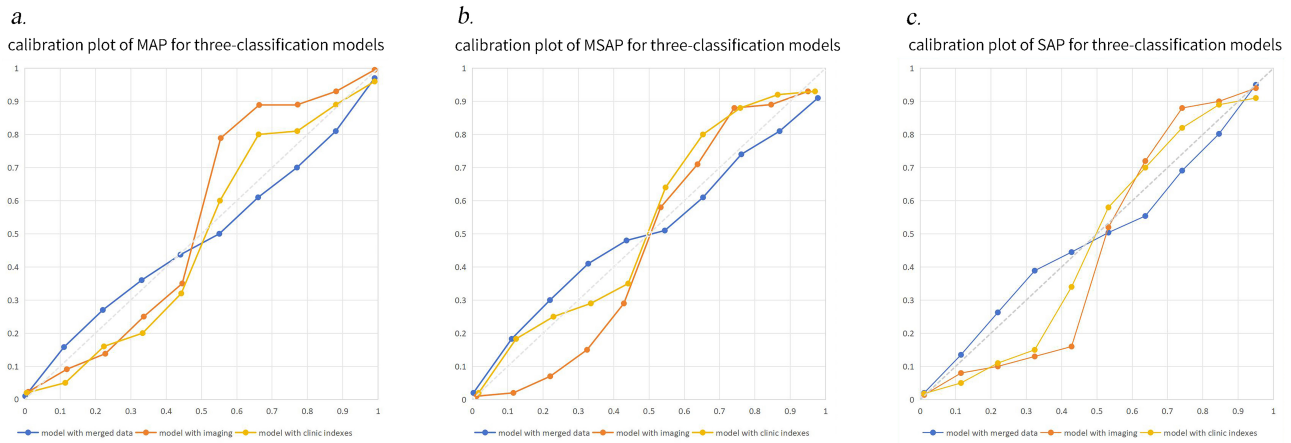


**Fig. 14. Average calibration plots of different AP severities in three-classification models with imaging data, clinic indexes and merged data as input.** (a) Average calibration plots of MAP in three-classification models with imaging data, clinic indexes and merged data as input (5-fold CV). (b) Average calibration plots of MSAP in three-classification models (5-fold CV). (c) Average calibration plots of SAP in three-classification models (5-fold CV).

cially for patients who do not need to be admitted to the intensive care unit. The APACHE-II score is a general assessment tool, but not specifically designed for assessing AP. Therefore, some AP-specific items are left out and not assessed if APACHE-II score is used for evaluation. The CTSI score is composed of Balthazar score and pancreatic necrosis score, whereas the modified CTSI (MCTSI) score is a revised imaging evaluation method based on the CTSI, which not only improves the scoring of pancreatic morphology and pancreatic necrosis, but also incorporates the assessment of extra-pancreatic complications. Both CTSI and MCTSI scores are straightforward imaging scores that

can reasonably quantify some of the imaging information. However, it is important to note that they are incapable of offering an evaluation regarding the overall condition of AP patients and their scoring results may be predisposed to subjective errors [54–60]. The BISAP score is a relatively new method, encompassing five indexes, namely age, BUN, GCS score, systemic inflammatory response syndrome (SIRS), and pleural effusion in its scoring evaluation. With the limited number of indicators being factored in, this score cannot fully capture the clinical condition of AP patients. Therefore, attempts to modify this scoring system to incorporate other factors, such as function of vital organs, nutritional status, and complications, for a more comprehensive assessment are necessary [61–65].

Although widely used in clinical practice, from an ML perspective, these scoring methods are clearly inadequate in the assessment of patient data features. By principle, most of them simply aggregate scores from various indexes. Nevertheless, technological progress in recent years has enabled significant development of the clinical feature extraction methods and feature processing methods of AP. For example, for non-imaging clinical data, classification and regression algorithms based on classical ML algorithms such as SVM and RF have been extensively researched and implemented in the AP context. These algorithms can be applied to the processing of multi-dimensional data input, feature selection, and fitting of the complex interaction relationships within data input [66–68]. On the other hand, the field of medical imaging saw a vigorous development in computer-aided diagnosis systems combining computer vision and ML. Radiomics methods have been used to extract features from AP imaging data, reducing the reliance on subjective judgment of physicians. These features are quantified by thousands of preset feature filters [69–75]. Meanwhile, some of the classical ML algorithms mentioned above can also be integrated into downstream tasks, optimizing the utilization of raw medical data to a great extent. However, the application of these algorithms is fraught with certain constraints. Firstly, in AP image processing, most of the existing ML processing methods focus on classifying the pancreatic parenchyma as a region of interest (ROI) and conducting subsequent feature extraction on the pancreas itself. The assessment of only pancreatic characteristics by these methods precludes the analysis of peripancreatic manifestations such as exudation, fluid collection, necrosis around the pancreas, which are critical findings used for assessment. Moreover, the manual delineation of ROI is laborious, costly and time-consuming [76]. Secondly, the thousands of manually preset feature filters based on radiomics have drawbacks such as feature redundancy and high computational cost. Additionally, it is challenging to design specific feature filters for certain tasks. In addition to the lesions in the pancreas, the inflammatory factors released by AP could also affect various organ systems of a patient.

Relying on either clinical indexes or imaging manifestations of a patient provides a unilateral view of AP severity classifications. The clinical symptoms and imaging findings of AP patients are not always compatible with each other. For instance, in some cases, severe clinical symptoms presented by patients are contradicted by the relatively mild imaging abnormalities detected in the pancreatic area [77]. Similar kind of discrepancies are also exemplified by patients with pancreatitis experiencing mild pain or discomfort in the initial stage despite imaging evidence of severe pancreatic edema and peripancreatic exudation. Therefore, to address the aforementioned problems, technical methods that can reasonably improve the assessment of AP are urgently needed.

The application of AI in the medical field is continuously expanding. In the context of AP, AI lends valuable support to the severity assessment of AP. AI models can predict the likelihood of a patient developing SAP and recommend appropriate therapeutic interventions based on patient characteristics and clinical indexes. The most contemporary AI models are capable of analyzing medical images, including those from ultrasonography, computed tomography (CT) or magnetic resonance imaging (MRI), streamlining the detection of inflammation level, fluid accumulation, and tissue necrosis. The utilization of AI models provides clinical assistance in assessing and predicting the severity of AP. At the same time, AI systems can contribute to disease monitoring, progression assessment, and prognosis evaluation, based on the utilization of patients' clinical data.

In this study, we applied AI methods to establish AP severity prediction models. For single-view models with imaging data and clinic indexes as input, the overall accuracies of DL models were higher than those of traditional models of comparison. And multi-view DL model was more accurate at predicting severity classification of AP.

At the same time, while establishing the AP multi-view model, we observed that the attention heatmap of AP tended to shift with the severity of the disease. Compared with MAP, the attention of MSAP tended to focus on the left anterior renal fascia and the tail of the pancreas. Additionally, a small amount of accumulated fluid could often be observed in these areas on the original CT scans. Compared with MSAP, the focal point of attention heatmap of SAP moved upward and to the right, and was fixed on the pancreas or scattered around it. The original CT images often revealed signs of pancreatic necrosis or a significant accumulation of fluid and necrosis around the pancreas. This body of observations corroborate the potential function of AI in the identification of disease characteristics to a certain extent.

Despite the great potential of AI in medicine, we should be aware that the professional clinical judgment and diagnosis made by physicians is irreplaceable. AI models can provide critical insights to physicians to support final diagnosis and treatment decisions as a myriad of factors

need to be considered while assessing the severity of AP. The application of AI systems should be governed by medical ethics and regulations, especially in the aspects of patient privacy and data security.

## Conclusion

The multi-view models utilizing clinical indexes and imaging data as input are superior to single-view models in AP severity prediction.

## Availability of Data and Materials

The data and code were available on GitHub (https://github.com/hellenecc/acute-pancreatitis).

## Author Contributions

Data collection, material preparation and first draft writing were performed by KX. Project design was performed by DS. Both authors contributed significantly to editorial changes of important content. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

## Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of the First Affiliated Hospital of Dalian Medical University with an approval number of PJ-KS-KY-2024-11. Informed consent was obtained from all patients.

## Acknowledgment

Not applicable.

## Funding

This research received no external funding.

## Conflict of Interest

The authors declare no conflict of interest.

## References

[1] Mederos MA, Reber HA, Girgis MD. Acute Pancreatitis: A Review. JAMA. 2021; 325: 382–390.

[2] Greenberg JA, Hsu J, Bawazeer M, Marshall J, Friedrich JO, Nathens A, *et al*. Clinical practice guideline: management of acute pancreatitis. Canadian Journal of Surgery. Journal Canadien De Chirurgie. 2016; 59: 128–140.

[3] Foster BR, Jensen KK, Bakis G, Shaaban AM, Coakley FV. Revised Atlanta Classification for Acute Pancreatitis: A Pictorial Essay. Radiographics: a Review Publication of the Radiological Society of North America, Inc. 2016; 36: 675–687.

[4] 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: 102–111.

[5] Lankisch PG, Apte M, Banks PA. Acute pancreatitis. Lancet (London, England). 2015; 386: 85–96.

[6] 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. Medicine. 2019; 98: e15275.

[7] Ge P, Luo Y, Okoye CS, Chen H, Liu J, Zhang G, *et al*. Intestinal barrier damage, systemic inflammatory response syndrome, and acute lung injury: A troublesome trio for acute pancreatitis. Biomedicine & Pharmacotherapy. 2020; 132: 110770.

[8] Ortiz Morales CM, Girela Baena EL, Olalla Muñoz JR, Parlorio de Andrés E, López Corbalán JA. Radiology of acute pancreatitis today: the Atlanta classification and the current role of imaging in its diagnosis and treatment. Radiologia. 2019; 61: 453–466.

[9] Yin M, Zhang R, Zhou Z, Liu L, Gao J, Xu W, *et al*. Automated Machine Learning for the Early Prediction of the Severity of Acute Pancreatitis in Hospitals. Frontiers in Cellular and Infection Microbiology. 2022; 12: 886935.

[10] Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF. SVM and SVM Ensembles in Breast Cancer Prediction. PloS One. 2017; 12: e0161501.

[11] Zhu G, Hu J, Xi R. The cellular niche for intestinal stem cells: a team effort. Cell Regeneration (London, England). 2021; 10: 1.

[12] Computer-aided detection and diagnosis/radiomics/machine learning/deep learning in medical imaging. Medical Physics. 2023; 50: 50–53.

[13] Harrison K, Pullen H, Welsh C, Oktay O, Alvarez-Valle J, Jena R. Machine Learning for Auto-Segmentation in Radiotherapy Planning. Clinical Oncology (Royal College of Radiologists (Great Britain)). 2022; 34: 74–88.

[14] Gu Y, Li R, Wang X, Zhou Z. Automatic Medical Report Generation Based on Cross-View Attention and Visual-Semantic Long Short Term Memorys. Bioengineering (Basel, Switzerland). 2023; 10: 966.

[15] Saba T. Computer vision for microscopic skin cancer diagnosis using handcrafted and non-handcrafted features. Microscopy Research and Technique. 2021; 84: 1272–1283.

[16] Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular Diversity. 2021; 25: 1315–1360.

[17] Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. Annals of the Royal College of Surgeons of England. 2004; 86: 334–338.

[18] Nensa F, Demircioglu A, Rischpler C. Artificial Intelligence in Nuclear Medicine. Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine. 2019; 60: 29S–37S.

[19] Zhou LQ, Wang JY, Yu SY, Wu GG, Wei Q, Deng YB, *et al*. Artificial intelligence in medical imaging of the liver. World Journal of Gastroenterology. 2019; 25: 672–682.

[20] Adak A, Khan MR. An insight into gut microbiota and its functionalities. Cellular and Molecular Life Sciences: CMLS. 2019; 76: 473–493.

[21] Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics. 2022; 23: bbab569.

[22] Renganathan V. Overview of artificial neural network models in the biomedical domain. Bratislavske Lekarske Listy. 2019; 120: 536–540.

[23] Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. Journal of Medical Imaging and Radiation Sciences. 2019; 50: 477–487.

[24] Matsuo K, Purushotham S, Jiang B, Mandelbaum RS, Takiuchi T, Liu Y, *et al*. Survival outcome prediction in cervical cancer:

Cox models vs deep-learning model. American Journal of Obstetrics and Gynecology. 2019; 220: 381.e1–381.e14.

[25] Liang H, Wang M, Wen Y, Du F, Jiang L, Geng X, *et al*. Predicting acute pancreatitis severity with enhanced computed tomography scans using convolutional neural networks. Scientific Reports. 2023; 13: 17514.

[26] Li ZQ, Pan HQ, Liu Q, Song H, Wang JM. Comparing the performance of time series models with or without meteorological factors in predicting incident pulmonary tuberculosis in eastern China. Infectious Diseases of Poverty. 2020; 9: 151.

[27] Pang T, Li P, Zhao L. A survey on automatic generation of medical imaging reports based on deep learning. Biomedical Engineering Online. 2023; 22: 48.

[28] Umapathy C, Gajendran M, Mann R, Boregowda U, Theethira T, Elhanafi S, *et al*. Pancreatic fluid collections: Clinical manifestations, diagnostic evaluation and management. Disease-a-month: DM. 2020; 66: 100986.

[29] Thoeni RF. Imaging of Acute Pancreatitis. Radiologic Clinics of North America. 2015; 53: 1189–1208.

[30] Sun L, Wang Z, Pu H, Yuan G, Guo L, Pu T, *et al*. Attention-embedded complementary-stream CNN for false positive reduction in pulmonary nodule detection. Computers in Biology and Medicine. 2021; 133: 104357.

[31] Working Group IAP/APA Acute Pancreatitis Guidelines. IAP/APA evidence-based guidelines for the management of acute pancreatitis. Pancreatology. 2013; 13: e1–15.

[32] Li Y, Zhao YQ, Zhang F, Liao M, Yu LL, Chen BF, *et al*. Liver segmentation from abdominal CT volumes based on level set and sparse shape composition. Computer Methods and Programs in Biomedicine. 2020; 195: 105533.

[33] Sitaula C, Hossain MB. Attention-based VGG-16 model for COVID-19 chest X-ray image classification. Applied Intelligence (Dordrecht, Netherlands). 2021; 51: 2850–2863.

[34] Cao Z, Huang J, He X, Zong Z. BND-VGG-19: A deep learning algorithm for COVID-19 identification utilizing X-ray images. Knowledge-based Systems. 2022; 258: 110040.

[35] Li M, Hsu W, Xie X, Cong J, Gao W. SACNN: Self-Attention Convolutional Neural Network for Low-Dose CT Denoising with Self-Supervised Perceptual Loss Network. IEEE Transactions on Medical Imaging. 2020; 39: 2289–2301.

[36] Jiang J, Elguindi S, Berry SL, Onochie I, Cervino L, Deasy JO, *et al*. Nested block self-attention multiple resolution residual network for multiorgan segmentation from CT. Medical Physics. 2022; 49: 5244–5257.

[37] Ong Y, Shelat VG. Ranson score to stratify severity in Acute Pancreatitis remains valid - Old is gold. Expert Review of Gastroenterology & Hepatology. 2021; 15: 865–877.

[38] Chauhan R, Saxena N, Kapur N, Kardam D. Comparison of modified Glasgow-Imrie, Ranson, and Apache II scoring systems in predicting the severity of acute pancreatitis. Polski Przeglad Chirurgiczny. 2022; 95: 6–12.

[39] Karabuga B, Gemcioglu E, Konca Karabuga E, Baser S, Ersoy O. Comparison of the predictive values of CRP, CRP/albumin, RDW, neutrophil/lymphocyte, and platelet/lymphocyte levels in determining the severity of acute pancreatitis in patients with acute pancreatitis according to the BISAP score. Bratislavske Lekarske Listy. 2022; 123: 129–135.

[40] Alberti P, Pando E, Mata R, Vidal L, Roson N, Mast R, *et al*. Evaluation of the modified computed tomography severity index (MCTSI) and computed tomography severity index (CTSI) in predicting severity and clinical outcomes in acute pancreatitis. Journal of Digestive Diseases. 2021; 22: 41–48.

[41] Tee YS, Fang HY, Kuo IM, Lin YS, Huang SF, Yu MC. Serial evaluation of the SOFA score is reliable for predicting mortality in acute severe pancreatitis. Medicine. 2018; 97: e9654.

[42] Kapadia NN, Siddiqui E. Bedside index (BISAP) v/s Ranson scores in predicting mortality and severity in patients with acute pancreatitis. JPMA. the Journal of the Pakistan Medical Association. 2021; 71: 1988–1991.

[43] Kuo DC, Rider AC, Estrada P, Kim D, Pillow MT. Acute Pancreatitis: What's the Score? The Journal of Emergency Medicine. 2015; 48: 762–770.

[44] Acehan F, Tez M, Kalkan C, Akdogan M, Altiparmak E, Doganay M, *et al*. Revisiting the Ranson score in acute pancreatitis: Is the drop in hematocrit a worrisome sign? Journal of Hepatobiliary-pancreatic Sciences. 2023; 30: 315–324.

[45] Capurso G, Ponz de Leon Pisani R, Lauri G, Archibugi L, Hegyi P, Papachristou GI, *et al*. Clinical usefulness of scoring systems to predict severe acute pancreatitis: A systematic review and meta-analysis with pre and post-test probability assessment. United European Gastroenterology Journal. 2023; 11: 825–836.

[46] Kaur H, Chandran VP, Rashid M, Kunhikatta V, Poojari PG, Bakkannavar SM, *et al*. The significance of APACHE II as a predictor of mortality in paraquat poisoning: A systematic review and meta-analysis. Journal of Forensic and Legal Medicine. 2023; 97: 102548.

[47] Tang W, Zha ML, Zhang WQ, Hu SQ, Chen HL. APACHE scoring system and pressure injury risk for intensive care patients: A systematic review and meta-analysis. Wound Repair and Regeneration: Official Publication of the Wound Healing Society [and] the European Tissue Repair Society. 2022; 30: 498–508.

[48] Bao Y, Ge W. Correlation between serum levels of PTX-3, SIL-2R, inflammatory markers, and APACHE II scores in patients with severe acute pancreatitis. Medicine. 2022; 101: e31252.

[49] Lankisch PG, Warnecke B, Bruns D, Werner HM, Grossmann F, Struckmann K, *et al*. The APACHE II score is unreliable to diagnose necrotizing pancreatitis on admission to hospital. Pancreas. 2002; 24: 217–222.

[50] Escobar-Arellano R, Guraieb-Barragán E, Mansanares-Hernández A, Sánchez-Valdivieso EA. Sensitivity, specificity and reliability of the POP score vs. APACHE II score as predictors of severe acute biliary pancreatitis. Cirugia Y Cirujanos. 2019; 87: 402–409.

[51] Pando E, Alberti P, Mata R, Gomez MJ, Vidal L, Cirera A, *et al*. Early Changes in Blood Urea Nitrogen (BUN) Can Predict Mortality in Acute Pancreatitis: Comparative Study between BISAP Score, APACHE-II, and Other Laboratory Markers-A Prospective Observational Study. Canadian Journal of Gastroenterology & Hepatology. 2021; 2021: 6643595.

[52] van den Berg FF, de Bruijn AC, van Santvoort HC, Issa Y, Boermeester MA. Early laboratory biomarkers for severity in acute pancreatitis; A systematic review and meta-analysis. Pancreatology. 2020; 20: 1302–1311.

[53] Xu C, Wang J, Jin X, Yuan Y, Lu G. Establishment of a predictive model for outcomes in patients with severe acute pancreatitis by nucleated red blood cells combined with Charlson complication index and APACHE II score. The Turkish Journal of Gastroenterology: the Official Journal of Turkish Society of Gastroenterology. 2020; 31: 936–941.

[54] Gupta P, Dawra S, Chandel K, Samanta J, Mandavdhare H, Sharma V, *et al*. Fat-modified computed tomography severity index (CTSI) is a better predictor of severity and outcome in patients with acute pancreatitis compared with modified CTSI. Abdominal Radiology (New York). 2020; 45: 1350–1358.

[55] Alhajeri A, Erwin S. Acute pancreatitis: value and impact of CT severity index. Abdominal Imaging. 2008; 33: 18–20.

[56] 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. The American Journal of Gastroenterology. 2010; 105: 435–41; quiz 442.

[57] Tahir H, Rahman S, Habib Z, Khan Y, Shehzad S. Compari-

son of the Accuracy of Modified CT Severity Index Score and Neutrophil-to-Lymphocyte Ratio in Assessing the Severity of Acute Pancreatitis. Cureus. 2021; 13: e17020.

[58] Yang L, Liu J, Xing Y, Du L, Chen J, Liu X, *et al*. Comparison of BISAP, Ranson, MCTSI, and APACHE II in Predicting Severity and Prognoses of Hyperlipidemic Acute Pancreatitis in Chinese Patients. Gastroenterology Research and Practice. 2016; 2016: 1834256.

[59] Cucuteanu B, Negru D, Gavrilescu O, Popa IV, Floria M, Mihai C, *et al*. Extrapancreatic necrosis volume: A new tool in acute pancreatitis severity assessment? World Journal of Clinical Cases. 2021; 9: 9395–9405.

[60] Sahu B, Abbey P, Anand R, Kumar A, Tomer S, Malik E. Severity assessment of acute pancreatitis using CT severity index and modified CT severity index: Correlation with clinical outcomes and severity grading as per the Revised Atlanta Classification. The Indian Journal of Radiology & Imaging. 2017; 27: 152–160.

[61] Gao W, Yang HX, Ma CE. The Value of BISAP Score for Predicting Mortality and Severity in Acute Pancreatitis: A Systematic Review and Meta-Analysis. PLoS ONE. 2015; 10: e0130412.

[62] Hagjer S, Kumar N. Evaluation of the BISAP scoring system in prognostication of acute pancreatitis - A prospective observational study. International Journal of Surgery (London, England). 2018; 54: 76–81.

[63] Coluoglu I, Coluoglu E, Binicier HC, Binicier OB. The role of the BISAP score in predicting acute pancreatitis severity according to the revised Atlanta classification: a single tertiary care unit experience from Turkey. Acta Gastro-enterologica Belgica. 2021; 84: 571–576.

[64] Dancu GM, Popescu A, Sirli R, Danila M, Bende F, Tarta C, *et al*. The BISAP score, NLR, CRP, or BUN: Which marker best predicts the outcome of acute pancreatitis? Medicine. 2021; 100: e28121.

[65] Valverde-López F, Matas-Cobos AM, Alegría-Motte C, Jiménez-Rosales R, Úbeda-Muñoz M, Redondo-Cerezo E. BISAP, RANSON, lactate and others biomarkers in prediction of severe acute pancreatitis in a European cohort. Journal of Gastroenterology and Hepatology. 2017; 32: 1649–1656.

[66] Zhou Y, Ge YT, Shi XL, Wu KY, Chen WW, Ding YB, *et al*. Machine learning predictive models for acute pancreatitis: A systematic review. International Journal of Medical Informatics. 2022; 157: 104641.

[67] Liu F, Yao J, Liu C, Shou S. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis. BMC Surgery. 2023; 23: 267.

[68] Hameed MAB, Alamgir Z. Improving mortality prediction in Acute Pancreatitis by machine learning and data augmentation. Computers in Biology and Medicine. 2022; 150: 106077.

[69] Yan G, Yan G, Li H, Liang H, Peng C, Bhetuwal A, *et al*. Radiomics and Its Applications and Progress in Pancreatitis: A Current State of the Art Review. Frontiers in Medicine. 2022; 9: 922299.

[70] Tong T, Gu J, Xu D, Song L, Zhao Q, Cheng F, *et al*. Deep learning radiomics based on contrast-enhanced ultrasound images for assisted diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis. BMC Medicine. 2022; 20: 74.

[71] Zhong J, Hu Y, Xing Y, Ge X, Ding D, Zhang H, *et al*. A systematic review of radiomics in pancreatitis: applying the evidence level rating tool for promoting clinical transferability. Insights into Imaging. 2022; 13: 139.

[72] Li J, Liu F, Fang X, Cao K, Meng Y, Zhang H, *et al*. CT Radiomics Features in Differentiation of Focal-Type Autoimmune Pancreatitis from Pancreatic Ductal Adenocarcinoma: A Propensity Score Analysis. Academic Radiology. 2022; 29: 358–366.

[73] Chen Y, Chen TW, Wu CQ, Lin Q, Hu R, Xie CL, *et al*. Radiomics model of contrast-enhanced computed tomography for predicting the recurrence of acute pancreatitis. European Radiology. 2019; 29: 4408–4417.

[74] Tang L, Ma L, Chen Y, Hu Y, Chen X, Huang X, *et al*. Radiomics analysis of contrast-enhanced T1W MRI: predicting the recurrence of acute pancreatitis. Scientific Reports. 2023; 13: 2762.

[75] Zhou T, Xie CL, Chen Y, Deng Y, Wu JL, Liang R, *et al*. Magnetic Resonance Imaging-Based Radiomics Models to Predict Early Extrapancreatic Necrosis in Acute Pancreatitis. Pancreas. 2021; 50: 1368–1375.

[76] Ma C, Li J, Boukar MB, Yang P, Wang L, Chen L, *et al*. Optimized ROI size on ADC measurements of normal pancreas, pancreatic cancer and mass-forming chronic pancreatitis. Oncotarget. 2017; 8: 99085–99092.

[77] Tsuji Y, Takahashi N, Fletcher JG, Hough DM, McMenomy BP, Lewis DM, *et al*. Subtraction color map of contrast-enhanced and unenhanced CT for the prediction of pancreatic necrosis in early stage of acute pancreatitis. AJR. American Journal of Roentgenology. 2014; 202: W349–W56.