



OPEN

Identifying severe community-acquired pneumonia using radiomics and clinical data: a machine learning approach

Tianning Yang¹, Ling Zhang², Siyi Sun², Xuexin Yao², Lichuan Wang^{1✉} & Yanlei Ge^{2✉}

Evaluating Community-Acquired Pneumonia (CAP) is crucial for determining appropriate treatment methods. In this study, we established a machine learning model using radiomics and clinical features to rapidly and accurately identify Severe Community-Acquired Pneumonia (SCAP). A total of 174 CAP patients were included in the study, with 64 cases classified as SCAP. Radiomic features were extracted from chest CT scans using radiomics techniques, and screened to remove irrelevant features. Additionally, clinical indicators of patients were similarly screened and constituted the clinical feature set. Subsequently, eight common machine learning models were employed to complete the SCAP identification task. Specifically, interpretability analysis was conducted on the models. In the end, we screened out 15 radiomic features (such as LeastAxisLength, Maximum2DDiameterColumn and ZonePercentage) and two clinical features: Lymphocyte ($p = 0.041$) and Albumin ($p = 0.044$). Using radiomic features as inputs in model predictions yielded the highest AUC of 0.85 on the test set. When using the clinical feature set as model inputs, the AUC was 0.82. Combining the two sets of features as model inputs, Ada Boost achieved the best performance with an AUC of 0.89. Our study demonstrates that combining radiomics and clinical data using machine learning methods can more accurately identify SCAP patients.

Keywords Radiomics, Machine learning, SCAP

SCAP is the most life-threatening form of CAP, characterized by intensive care unit (ICU) admission and high morbidity and mortality rates¹. Pneumonia severity scales, such as the CURB-65 (CUR-65) scale and the Pneumonia Severity Index (PSI), are available for clinical use^{2,3}. However, each scoring system has its advantages and disadvantages. The most widely accepted standard for SCAP currently comes from the 2007 Infectious Diseases Society of America (IDSA)/American Thoracic Society (ATS) consensus guidelines on the management of community-acquired pneumonia in adults⁴. It has always been difficult but essential in managing patients with CAP to recognize SCAP promptly and intervene aggressively⁵.

The place of CT in assessing pneumonia as a direct objective response to evidence of chest disease is unquestionable. A study correlates a correlation between CT signs (pleural effusion, pyothorax) and pneumonia severity⁶. Shifa and his team developed the chest severity score (CT-SS) to assess the severity of COVID-19 pneumonia patients, with a final positive predictive value of 90.6% and a negative predictive value of 94%⁷. However, these studies require experienced radiologists to review all the CT images and extract features from them. Yet, the assessment is highly subjective, with results varying clinically depending on the physician's experience and ability, and the process is time-consuming and labor-intensive.

Radiomics was first proposed by Lambin et al.⁸. This technique is based on the hypothesis that there is still untapped information in medical imaging that goes beyond visual interpretation. Radiomics can extract a large amount of feature information from medical images, which supplements doctors' qualitative analyses. Moreover, these feature data can be used to train models for predicting specific clinical tasks⁹. For example, some researchers have used radiomics to extract feature information from chest X-rays for diagnosing breast tumors¹⁰ and predicting the benign or malignant nature of lung nodules using non-invasive methods¹¹. These studies have achieved high prediction accuracy.

¹College of Science, North China University of Science and Technology, Tangshan, Hebei, China. ²Department of Respiratory Medicine, North China University of Science and Technology, Affiliated Hospital, Tangshan, Hebei, China. ✉email: wanglichuan@ncst.edu.cn; 495732196@qq.com

In addition, with the development of artificial intelligence, the combination of machine learning and radiomics has further broadened its application. During the COVID-19 pandemic, due to severe shortages of medical resources, numerous studies utilized artificial intelligence (AI) to assess the severity of COVID-19 patients^{12–14}. These studies trained deep learning models on a large amount of labeled image data, enabling the models to learn features within the images. Specifically, Dinh et al. trained five deep learning models using chest X-ray images of COVID-19 patients, achieving classification of the patients' lung infection status¹². A study trained a model based on the Siamese neural network-based algorithm to automatically measure the severity of chest CT in hospitalized patients¹⁵. However, the above studies used tens of thousands of high-quality annotated images. Because deep learning models have more parameters, they require many samples for practical training. When sufficient training samples are not available, overfitting can quickly occur.

Compared to machine learning methods, deep learning-based feature extraction requires a large amount of data and time to build the model, and the features learned by the model are not directly interpretable^{16,17}. However, using radiomics-based machine learning models for feature extraction and prediction can somewhat mitigate these issues. Consequently, some radiomics-based machine learning algorithms have been widely used. For example, Prinzi et al. extracted features from X-ray images using radiomics and employed SVM and random forest as classifiers to complete prognostic prediction tasks, achieving an AUC of 0.819 on the test set¹⁸.

Therefore, this study aims to use multiple machine learning algorithms for radiomics modeling and to compare the model performance in conjunction with clinical metrics to achieve fast and accurate identification of SCAP patients and, at the same time, to explore human-unrecognizable image features in chest CT based on artificial intelligence, and to analyze the correlation between the clinical metrics and the image features. Figure 1 shows the overall flow of this study.

Materials and methods

Study population

To determine the minimum sample size required for the study, power analysis was conducted using G*Power 3.1.9.7¹⁹. With a significance level of $\alpha = 0.05$, an effect size of 0.2, and a statistical power of 0.8, at least 150 participants were needed to achieve the desired statistical power. Considering the practical needs of the study, we ultimately included 174 participants.

Clinical and imaging data of patients diagnosed with CAP were retrospectively collected at the North China University of Science and Technology Affiliated Hospital. Among the inclusion criteria were: (1) Patients diagnosed with CAP; (2) CT scan was performed during admission; (3) There were lesions in the lungs on CT scan. Exclusion criteria were: (1) Patients who lack clinical data; (2) Patients who did not undergo CT scan during admission. A total of 174 patients were collected: 110 in the non-SCAP group and 64 in the SCAP group. The diagnosis of CAP and SCAP was made by superior physicians according to the Guidelines for the Diagnosis and Treatment of Community-Acquired Pneumonia in Chinese Adults (2016 edition) and IDSA/ATS standard Guidelines^{4,20}.

CT examinations

All patients were examined using a CT scanner with the exact specifications and parameters. Scanning parameters: 130 kV; 120 mA; scanning time 1s; matrix 512 × 512; in the supine position with breath held, scanning was performed sequentially from the lung apex to the diaphragm.

CT image segmentation

Currently, there are three main types of medical image segmentation models: manual segmentation, semi-automatic segmentation, and automatic segmentation. Among them manual segmentation needs to be done manually by a specialized radiologist. The segmentation process is time-consuming and labor-intensive, and manual segmentation also tends to lose the contour information of the lesion. Therefore, this study used automatic segmentation to assist manual segmentation methods in defining regions of interest (ROI).

No new network (nnU-Net) is a general and adaptive deep learning framework for medical image segmentation. It is developed based on the U-Net architecture, with its main advantage being the simplification and automation of the design and adjustment processes for medical image segmentation models²¹. Specifically, nnU-Net automatically generates suitable model network structures and training parameters according to the characteristics of the dataset (such as image resolution and image size). Additionally, the model can automatically handle tasks like data processing and data augmentation without requiring the user to process the data separately. During model training, nnU-Net manages the training process automatically, including learning rate scheduling and early stopping mechanisms. Furthermore, nnU-Net supports various loss functions, including Dice loss and cross-entropy loss, and can automatically select the appropriate loss function based on the specific task. Overall, nnU-Net is a practical and high-performance segmentation model.

We use publicly available and pre-trained models on lung CT images (<https://zenodo.org/records/4635822>). We perform inference on the image data to obtain a lung lesion mask, ROI. The final examination and processing are then carried out by an experienced radiologist, thus ensuring that the lung lesions can be accurately extracted.

Feature extraction

Radiomics feature extraction

To quantify the grey-scale features extracted for each ROI²², an open-source tool, Pyradiomics²³ (v3.0, compliant with the Biomarker Standardization Initiative Guidelines²⁴), was used to extract features from the CT scans automatically, and the extracted features were further analyzed. All radiomics features were classified into seven categories, namely: (i) shape-based features; (ii) first-order features; (iii) gray-level dependence matrix features;

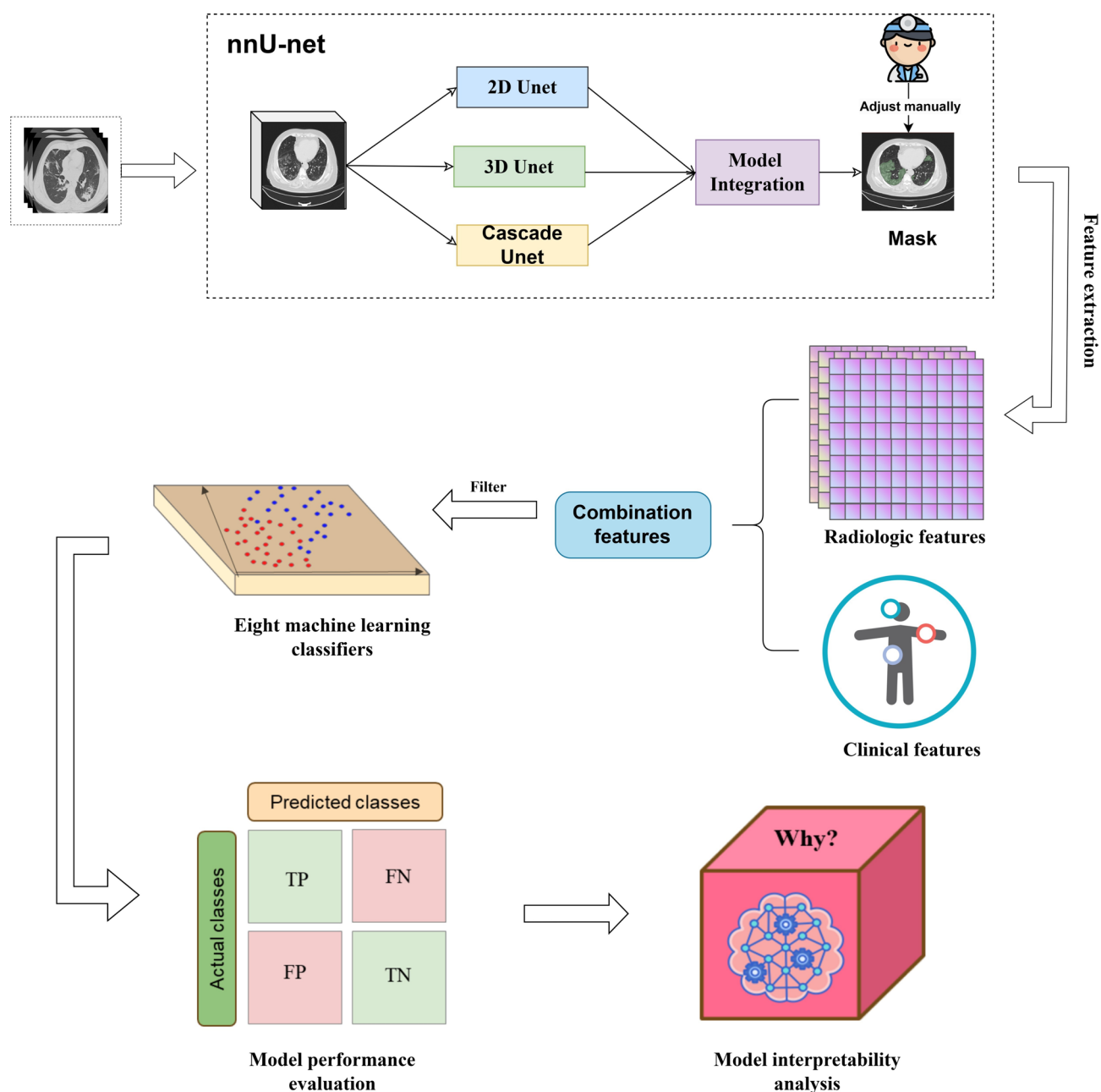


Fig. 1. The overall process of research.

(iv) gray-level size zone matrix features; (v) neighboring gray-tone difference matrix features; (vi) gray-level run-length matrix features; and (vii) gray-level co-occurrence matrix features.

We not only extracted features from the original CT scans but also from derived images processed through filters. Specifically, the original images were processed using a Laplacian of Gaussian filter. In the Laplacian of Gaussian filter, the parameter sigma defines the roughness of the texture intensity. A low sigma emphasizes fine textures, while a high sigma emphasizes coarse textures. We chose to set sigma to {1, 2, 3}. Ultimately, 126 features were extracted from the original images, and 258 features were extracted from the LoG-derived images, resulting in a total of 384 features.

Construction of radiomic feature set

The high number of features extracted using the feature extraction tool quickly increases the risk of overfitting the model during the subsequent statistical analysis and machine learning modeling process. Reducing the number of features is crucial to building effective and generalizable models.

To reduce the dimensionality of the features, we considered various feature selection methods. First, to eliminate highly similar redundant features, Pearson's correlation coefficient was used to measure the similarity between these features²⁵. The Pearson correlation coefficient measures the correlation between two features and

ranges from $[-1, 1]$. When it is close to -1 , it indicates a negative correlation between the two features; when it is close to 1 , it indicates a positive correlation; when it is 0 , it indicates no correlation. However, the threshold selection for the Pearson correlation coefficient often depends on the specific task. Generally, a threshold greater than 0.7 indicates a strong correlation, so we set it to 0.7 in our work²⁶.

Then, two filter-based methods were used: the Mann-Whitney U test and maximal relevance and minimal redundancy (mRMR). Both methods are based on statistical measures to select features and have the advantages of high computational efficiency and robustness. Specifically, the Mann-Whitney U test does not require the data to meet any distribution curve, making it more robust compared to the t-test, which compares means, as the Mann-Whitney U test compares medians²⁷. mRMR uses mutual information as a criterion to calculate the relevance of the feature subset to the output class and the redundancy among features.

To avoid the risk of overfitting during model training and to improve the generalization performance of the model, the construction of the feature set is therefore completed by evaluating the features and identifying those that have a significant impact on the target variable before training the classifier using the model-based feature screening methods: Random Forest, XGBoost, and Lasso model, respectively. From this, the importance scores of the features under the three models are obtained, the three scores are summed, and the features with the top 15 combined rankings are selected to be constructed as the imaging feature set.

Clinical feature set construction

Considering the influence of clinical characteristics on the target categorical variables, this study collated and collected clinical indicators from patients during their admission to the hospital as a clinical dataset. Patients' age, gender, whether they smoked, whether they drank alcohol, and CURB-65 scores were collected; laboratory test results: blood urea nitrogen (BUN), Calcitonin (PCT), white blood cell count (WBC), neutrophil (NEU), lymphocyte (LYM), Total Plasma Protein (TP), serum albumin (ALB), serum creatinine (Scr), alanine aminotransferase (ALT), aspartate transaminase (AST), fibrinogen, D-dimer. We performed a correlation analysis between clinical features and imaging features to investigate whether there is information that can be mined between clinical features and imaging features. Specifically, the potential relationship between the two types of features was further discovered by using Pearson's correlation coefficient to calculate the correlation coefficient between clinical and imaging features and visualization using heat maps. Univariate and multivariate logistic analyses were performed on the clinical features to identify statistically significant clinical features.

Establishment of models and performance comparison

Considering the small size of the data set and the high dimension of the features, the eight most mainstream machine learning models were selected. Including Ada Boost Classifier Logistic, Regression, Random Forest, SVM (Radial Kernel), XGBoost, KNN, Light Gradient Boosting, Naive Bayes, that use the image feature set clinical feature set, and a combination of the two as inputs to the model. For model validation, we employed 10-fold cross-validation. Specifically, the data were divided into 10 subsets, with 9 subsets used for model training and the remaining subset for testing. This process was repeated 10 times, and the results were averaged to obtain the final model performance.

In terms of model performance evaluation, AUC is defined as the area surrounded by the ROC curve and the lower coordinate axis, which is mainly used to evaluate the accuracy of binary classification problems. In the case of data imbalance, AUC still has good stability and can accurately evaluate the performance of the model. Therefore, we use AUC as the most important evaluation index to evaluate the model.

Interpretability analysis

Machine learning models are a black-box operation for medical professionals who do not understand the model's decision-making process. Interpretable analysis of machine learning models can provide a deep understanding of how the models work and can help identify incorrect decisions made by the models, which is particularly important in the medical field. In addition, interpretable analyses can provide valuable insights into medical research. For example, in identifying SCAP, which characteristic indicators are more important in the classification task, and which clinical or imaging indicators changes tend to worsen the patient's condition. Therefore, it is vital to perform interpretable analyses of the model using relevant tools.

Statistical analysis

The clinical characteristics of the patients were described using univariate analysis. Statistically significant indicators in the training set were subjected to multifactorial logistic regression analysis to screen for independent risk factors for SCAP. Statistical significance was defined as p value < 0.05 .

Ada Boost, Logistic Regression, Random Forest, SVM, XGBoost, and KNN machine learning models were constructed using the "sklearn" package. ROC curves were plotted, and AUC values were calculated to evaluate the models' discrimination. This section was done using Python (version 3.9.6).

Code availability

We open the core code, can be get by visiting <https://github.com/COOk921/Identifying-SCAP-Using-Radiomics>.

Results

Patient characteristics

A total of 174 people were enrolled in this study, with 110 patients in the non-SCAP group and 64 in the SCAP group. Randomly divide the training and test sets in the ratio of 7:3. Ultimately, the number of non-SCAP and

SCAP patients in the training set was 77 and 45, respectively, and the number of non-SCAP and SCAP patients in the test set was 33 and 19, respectively. Table 1 describes the demographic and clinical characteristics of the study population. Percentages were used to describe categorical variables. Continuous variables that fit a normal distribution are described using $X \pm S$. For constant variables that do not provide a normal distribution, they are defined using the median (IQR).

CT image segmentation

Figure 2a shows the original CT image slice. Figure 2b displays the result after automatic segmentation using the nnU-Net, where it is evident that the model can accurately identify the lesion area. Figure 2c is the adjusted result by a radiologist based on image (b).

Unlike X-ray films, CT scans consist of multiple slices, making ROI delineation more time-consuming. According to our statistics, it takes a radiologist approximately 15 min to independently delineate the ROI of a lung CT scan. However, if the nnU-Net model is used for initial segmentation, and then a radiologist makes

Characteristic	Train cohort			Test cohort		
	Non-SCAP (n = 77)	SCAP (n = 45)	P	Non-SCAP (n = 33)	SCAP (n = 19)	P
Age	62.89 ± 18.67	71.27 ± 14.95	0.015*	62 (34, 78)	76 (69.5, 83.5)	0.015*
Gender, n (%)						
Male	43 (56%)	28 (62%)	0.459	20 (61%)	11 (58%)	> 0.999
Female	34 (44%)	17 (38%)		13 (39%)	8 (42%)	
Smoking						
True	20 (26%)	9 (20%)	0.598	26 (79%)	16 (84%)	0.729
False	57 (74%)	36 (80%)		7 (21%)	3 (16%)	
Drinking						
True	17 (22%)	6 (13%)	0.341	29 (88%)	18 (95%)	0.641
False	60 (78%)	39 (87%)		4 (12%)	1 (5%)	
CURB-65						
0~2	69 (90%)	29 (64%)	0.009 *	28 (85%)	11 (58%)	0.047*
≥ 3	8 (10%)	15 (33%)		5 (15%)	8 (42%)	
BUN (mmol/L)	5.35 (3.82, 7.93)	5.75 (4.53, 10.74)	0.03*	5.03 (4.2, 9.83)	7.27 (6.26, 9.26)	0.073
PCT (ng/mL)	0.1 (0.05, 0.67)	0.34 (0.09, 1.51)	0.116	0.18 (0.05, 0.98)	0.12 (0.08, 0.92)	0.614
WBC (x 10^9)	8.1 (6.3, 11.9)	8.5 (5.9, 14.5)	0.924	7.2 (5.4, 12.1)	9.1 (5.9, 15.5)	0.337
Neutrophil (x 10^9)	5.87 (3.58, 9.58)	6.92 (4.03, 11.89)	0.421	5.8 (3.8, 9.18)	7.25 (4.27, 11.93)	0.345
LYM(x 10^9)	1.45 (1.01, 1.98)	0.87 (0.63, 1.2)	< 0.001*	1.23 (0.95, 1.91)	1.02 (0.56, 1.48)	0.082
TP (g/L)	68.4 ± 10.61	63.45 ± 11.18	0.018 *	69.9 (63.7, 75.7)	61.7 (55.9, 69)	0.01*
ALB (g/L)	37.1 (33.4, 46.1)	34.6 (28.7, 37.7)	< 0.001*	45.1 (32.3, 47)	33.8 (29.15, 36.9)	0.003*
SCr (umol/L)	67 (55, 83)	71 (56, 93)	0.199	67 (59, 88)	57 (53.5, 88.5)	0.776
ALT (U/L)	18 (12, 36)	18 (12, 27)	0.509	20 (13, 35)	26 (18, 31)	0.419
AST	20 (16, 31)	22 (19, 33)	0.287	21 (17, 36)	24 (20.5, 31.5)	0.682
Fibrinogen (g/L)	5.19 ± 1.65	5.48 ± 2.05	0.428	5.28 ± 1.44	4.73 ± 1.53	0.217
D-Dimer (ng/mL)	734.53 (434.26, 1706.89)	1138.64 (520.01, 2143.72)	0.133	669.24 (430.1, 1315.01)	973 (562.5, 2704)	0.162

Table 1. Statistical description of patients’ clinical characteristics. *P value < 0.05

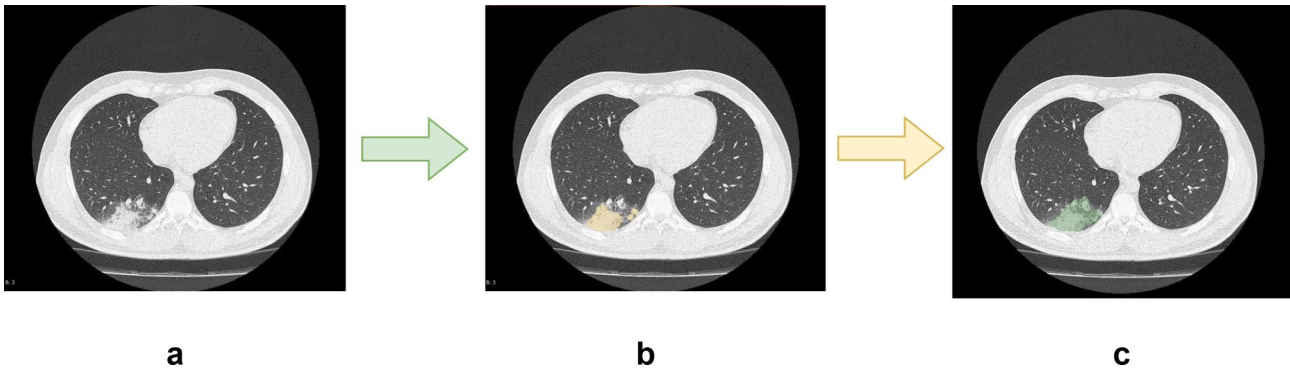


Fig. 2. CT image segmentation.

adjustments based on that, the time is reduced to 3 min. This significantly improves the efficiency of image segmentation. In terms of segmentation accuracy, since the radiologist further adjusts the ROI, the final segmentation accuracy is greatly ensured.

Image feature set construction

By automatically and manually segmenting the pneumonia mask, we first extracted 384 quantitative radiomics features. The number of features was initially reduced from 384 to 28 after an initial screening of image features by U-test, Pearson's correlation coefficient, and mRMR techniques. The features are then evaluated deeply using model-based feature screening methods: Random Forest, XGBoost, and Lasso. Figure 3 shows the combined scores for each feature under the three assessment methods. The final selection includes 15 high-scoring radiomics features, including 6 First Order Features that mainly describe the shape and grayscale characteristics of the image, 6 Gray Level Co-occurrence Matrix (GLCM) Features, 1 Gray Level Run Length Matrix (GLRLM) Features, 2 Gray Level Size Zone Matrix (GLSZM) Features.

Clinical feature set construction

The heat map was developed to investigate further the correlation between clinical features and clinical and imaging features (Fig. 4). Characteristics with correlation coefficients greater than 0.7 are summarised in Table 2. The strong correlations were mainly between imaging features. There were only two pairs of features with correlations greater than 0.7 between clinical features: neutrophils and leukocytes, CURB-65 score and age. However, the correlation coefficients for imaging and clinical features were relatively low. As shown in Fig. 4, the lower left part of the heat map (upper right part) has no too dark or too light color blocks. This means that there is no strong correlation between clinical features and imaging features.

A total of six clinical indicators in the training set with $P < 0.05$ on univariate analysis were Age, CURB-65, BUN, LYM, TP, and ALB. Multivariate analyses were performed on these characteristics, considering the interaction of multiple characteristic independent variables on the dependent variable. Two of them, LYM and ALB, had $P < 0.05$ and were combined into a clinical feature set (Table 3).

Establishment of models and performance comparison

Table 4 summarises the performance of the eight machine-learning models for the three feature sets. AUC was used as the primary evaluation index, and indexes such as accuracy, recall, and precision were used as secondary

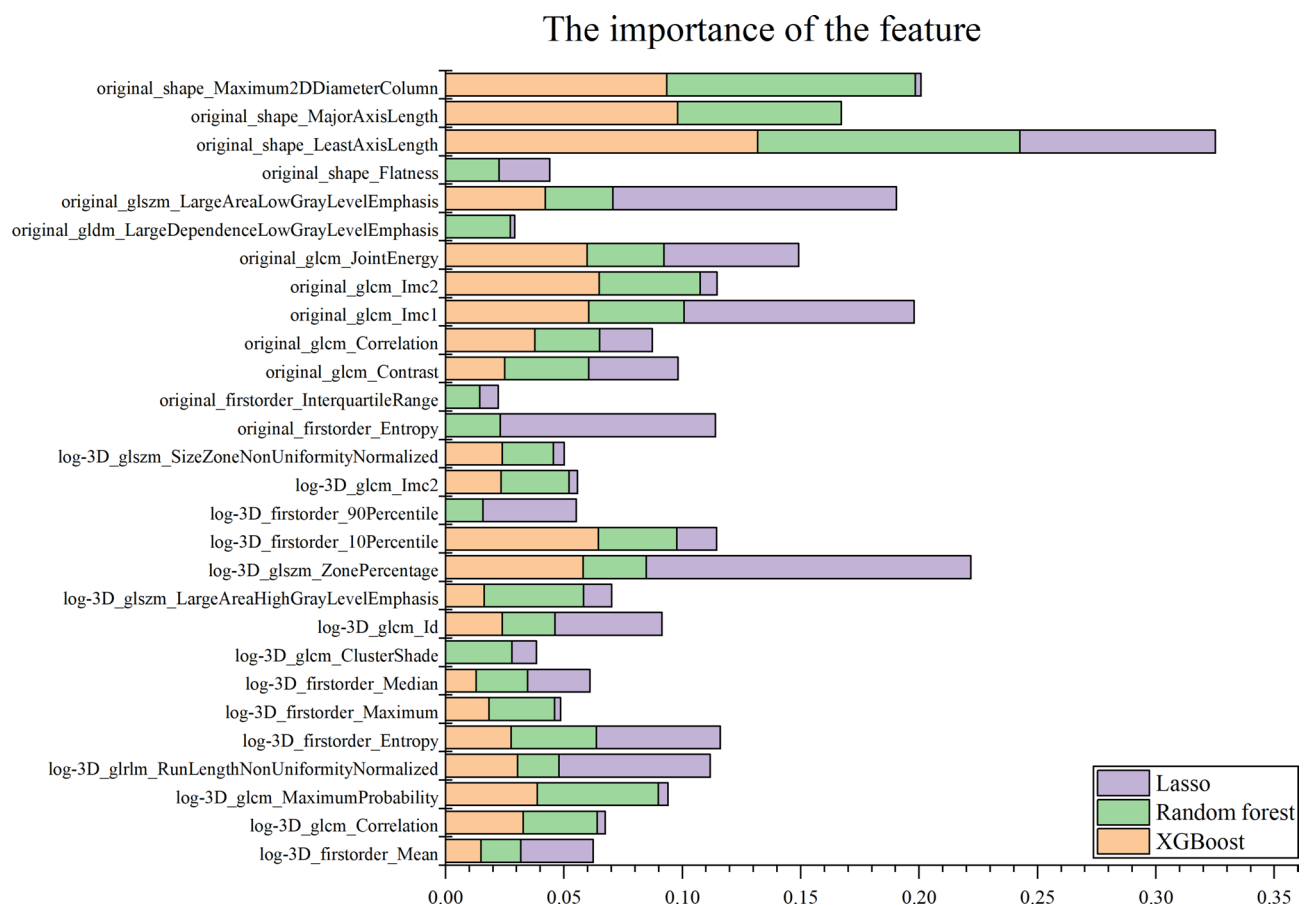


Fig. 3. Significance of features under the three screening methods.

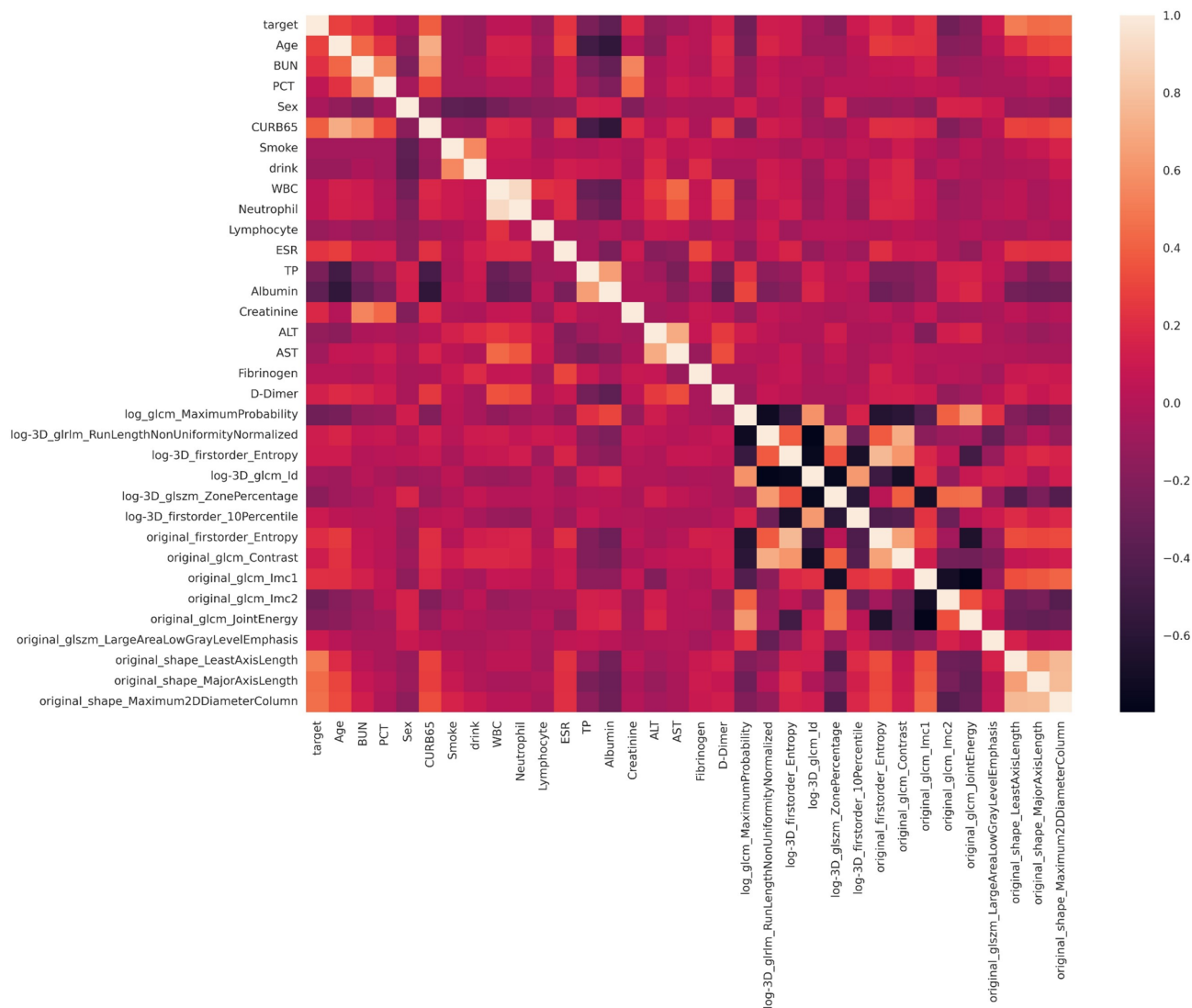


Fig. 4. Feature-related heat map.

Characteristic I	Characteristic II	Correlation coefficient
Neutrophil	WBC(white blood cell)	0.9123
original_glcm_JointEnergy	original_glcm_Imc1	−0.7906
log-3D_glcm_Id	log-3D_glrlm_RunLengthNonUniformityNormalized	−0.7984
original_firstorder_Entropy	log-3D_firstorder_Entropy	0.7628
log_glcm_MaximumProbability	log-3D_glrlm_RunLengthNonUniformityNormalized	0.7352
log-3D_glszm_ZonePercentage	log-3D_glcm_Id	−0.7631
original_shape_LeastAxisLength	original_shape_Maximum2DDiameterColumn	0.7636
original_shape_MajorAxisLength	original_shape_Maximum2DDiameterColumn	0.7739
original_glcm_Contrast	log-3D_glrlm_RunLengthNonUniformityNormalized	0.7080
CURB65	Age	0.7112

Table 2. Correlation between features.

evaluation indexes. It can be seen that higher AUC values and prediction accuracies are obtained for the imaging feature set compared to the clinical feature set. However, combination features performed the best in the classification task with an AUC value of 0.89 and a prediction accuracy of 0.84. Figure 5A–E illustrates the histograms of various evaluation metrics for different models with different feature sets, Fig. 5F ROC curves for the eight models under the combined feature set, and Fig. 6 shows its confusion matrix.

Characteristic						
Index	Age	CURB-65	BUN	LYM	TP	ALB
<i>P</i>	0.974	0.274	0.601	0.032*	0.758	0.030*
OR	0.999	1.926	1.025	0.542	1.008	0.908
OR lower 95%CI	0.969	0.599	0.935	0.298	0.956	0.828
OR upper 95%CI	1.031	6.421	1.127	0.901	1.063	0.983

Table 3. Multifactorial logistic regression analysis of clinical characteristics.

Feature set	Model name	AUC	Accuracy	Recall	Precision	F1
Radiologic features	Ada boost classifier	0.8517	0.7308	0.8421	0.5926	0.6957
	Logistic regression	0.8469	0.7885	0.6842	0.7222	0.7027
	Random forest	0.8293	0.75	0.4737	0.75	0.5806
	SVM (Radial kernel)	0.8158	0.7692	0.4737	0.8182	0.6
	XGBoost	0.7885	0.8293	0.6316	0.75	0.6857
	KNN	0.7584	0.7115	0.4211	0.6667	0.5161
	Light gradient boosting	0.7257	0.6346	0.3158	0.5	0.3871
	Naive bayes	0.7177	0.7115	0.4737	0.6429	0.5455
Clinical features	Ada boost classifier	0.8222	0.7115	0.5263	0.625	0.5714
	Logistic regression	0.7624	0.7115	0.5263	0.625	0.5714
	Random forest	0.7998	0.6923	0.5263	0.5882	0.5556
	SVM (Radial kernel)	0.7472	0.6923	0.4737	0.6	0.5294
	XGBoost	0.8132	0.75	0.7368	0.6364	0.6829
	KNN	0.8057	0.7115	0.5789	0.6111	0.5946
	Light gradient boosting	0.7616	0.6731	0.5263	0.5556	0.5405
	Naive bayes	0.7352	0.6731	0.6842	0.5417	0.6047
Combination features	Ada boost classifier	0.8947	0.8077	0.7368	0.7368	0.7368
	Logistic regression	0.8628	0.8462	0.6842	0.8667	0.7647
	Random forest	0.8844	0.8462	0.5789	0.9421	0.7333
	SVM (Radial kernel)	0.8612	0.7885	0.6316	0.75	0.6857
	XGBoost	0.8628	0.8077	0.6316	0.8	0.7059
	KNN	0.8158	0.7692	0.6316	0.7059	0.6667
	Light gradient boosting	0.823	0.6923	0.5789	0.5789	0.5789
	Naive bayes	0.8086	0.75	0.7895	0.625	0.6977

Table 4. Model performance evaluation. The bolded data is the optimal value of each evaluation index.

Interpretability analysis

We used the SHAP tool to perform interpretability analyses of the constructed models²⁸. SHAP is a tool for interpreting "black box models" by calculating the marginal contribution of features to the model. At the heart of SHAP is the SHAP value. In all the samples, there is a SHAP value for each feature. It indicates how much this feature affects the prediction results in this sample. When the SHAP value is 0, the features under the sample do not influence the prediction results; when the absolute value of the SHAP value is more significant, it means that the features under the sample have more influence on the prediction results. When the SHAP value is positive, it indicates a positive gain to the prediction, and vice versa; there is a negative gain.

Figure 7 plots the feature density scatter plot. In the graph, the horizontal coordinate indicates the SHAP value and the vertical coordinate indicates each feature (both clinical and imaging features) in the classification task. The dots in the graph represent each sample, and the dot's color indicates the magnitude of the value for that sample for that feature. It can be seen that among the image features, the absolute value of SHAP value of the midpoints of "MajorAxisLength", "Maximum2DDiameterColumn" and "shape_LeastAxisLength" are more significant, indicating that they play an essential role in the classification task. The clinical features "Lymphocyte" and "Albumin" were ranked fourth and fifth. In addition, more information can be extracted from it. For example, when the value of the "Albumin" feature is more significant, it is less likely to cause the patient to become a severe illness. This is because the red sample points in the "Albumin" feature have a SHAP value of less than 0, which negatively affects the prediction. This figure allows the clinical expert to gain further insight into the prediction process of the model.

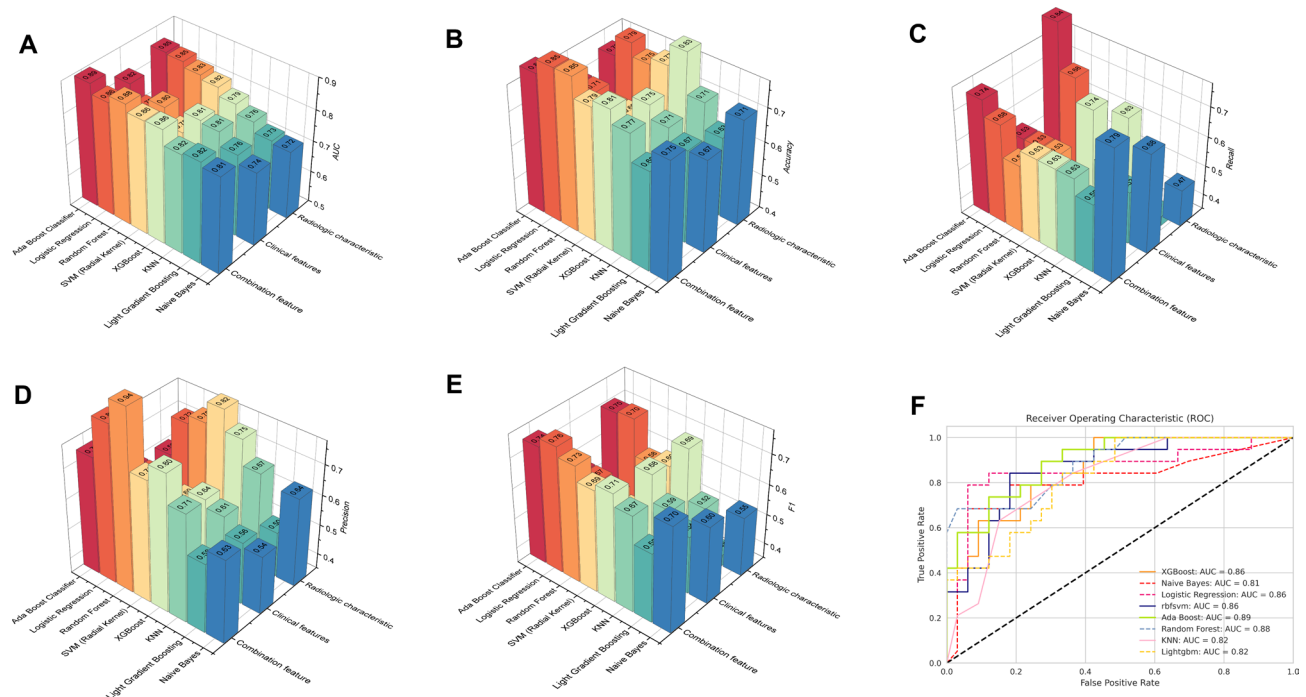


Fig. 5. (A–E) are the evaluation metrics under three feature sets and eight machine learning models, respectively; (F) is the ROC curve of the eight machine learning models under the combined features.

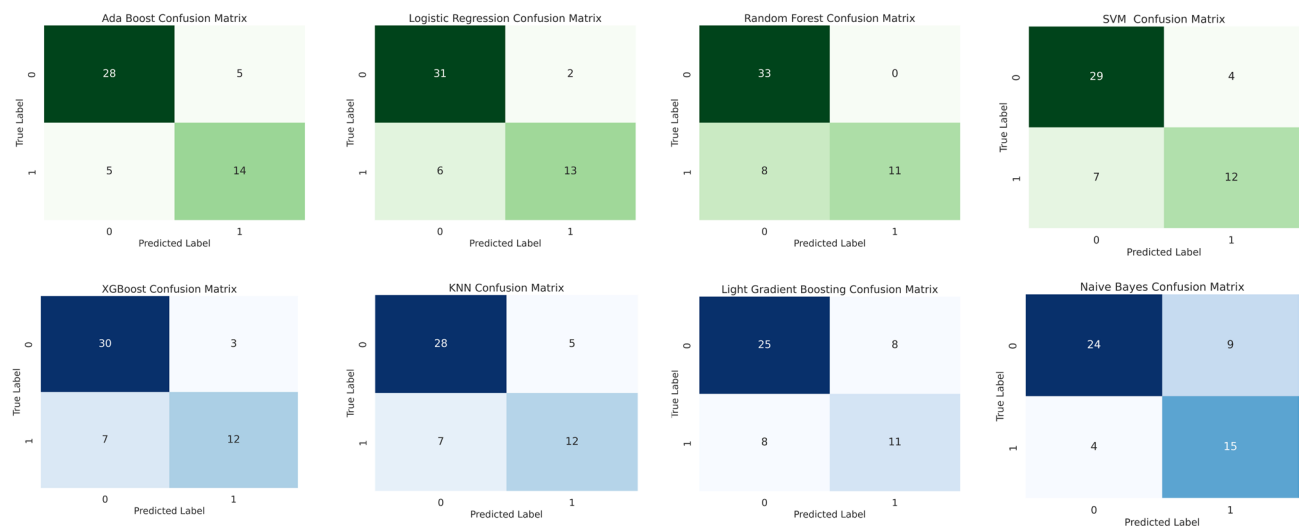


Fig. 6. Confusion matrix for eight machine learning models with combined features.

Discussion

The previous scoring of the pneumonia severity has focused more on physiological indicators such as oxygenation index, blood pressure, etc. Once a patient goes into shock or has severe respiratory failure, it is an extreme case. However, patients who appear to be non-SCAP and then deteriorate cannot be ignored. New biomarkers for the assessment of severity have also emerged over the years. Pathological changes in patients can be quickly and accurately localized to the lungs by chest CT features, and this study found that the radiomics model constructed using machine learning algorithms had a high accuracy in identifying patients with SCAP. Radiomics was used in our research to extract features hidden in lesions on CT images of patients' lungs to form a robust imaging feature set. In addition to this, the patient's clinical feature data is combined into a clinical feature set. The combination of the two together achieves the prediction of SCAP identification. Compared to traditional pneumonia severity scoring systems, such as CURB-65, this study uses radiomics to more fully consider the impact of hidden features in medical images on the extent of the condition.

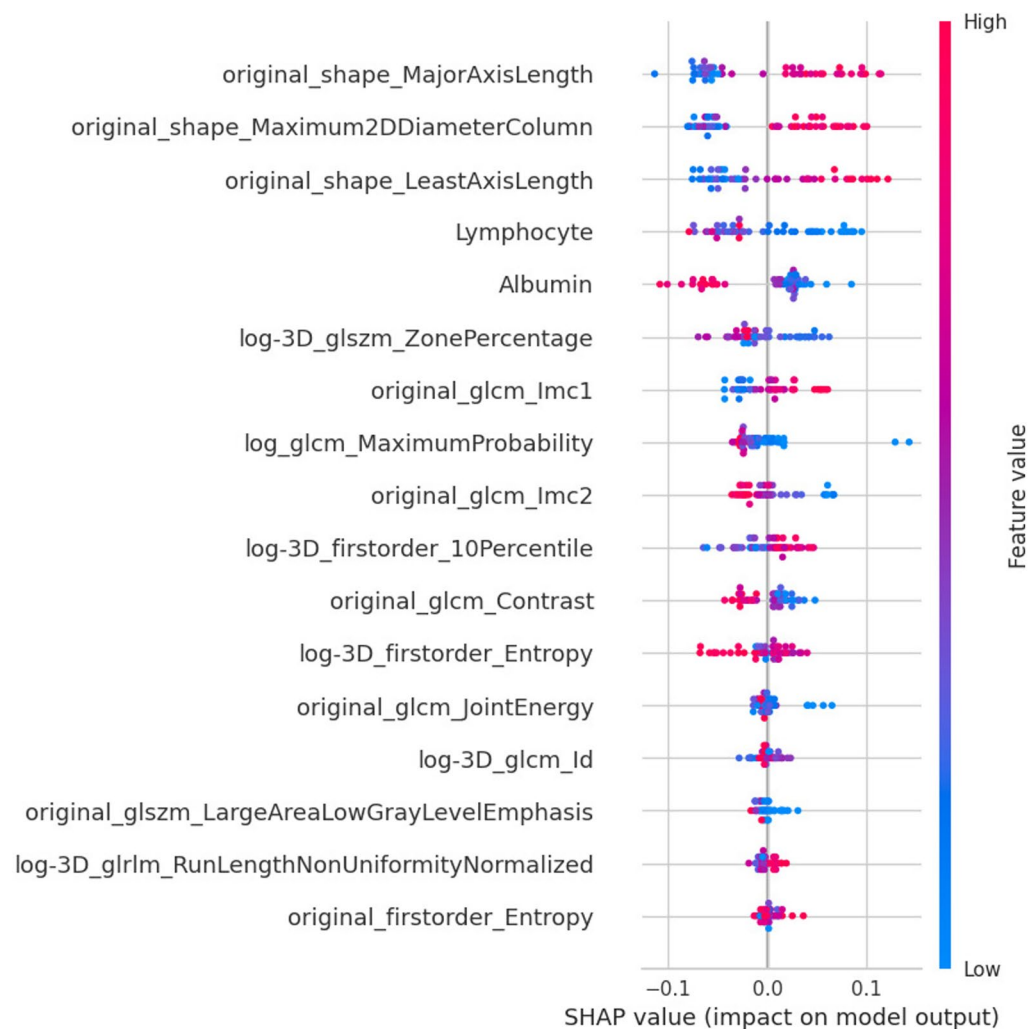


Fig. 7. Scatterplot of feature density.

The ROI in the image must be extracted before extracting the features in the image using the radiomics tools. The accuracy of the ROI extraction is crucial for the availability of the radiomics features. Currently, there are three main types of image segmentation: manual segmentation, semi-automatic segmentation, and automatic segmentation²⁹. Where manual segmentation means that the segmentation task is done manually by the relevant experts. This approach is very time-consuming, and there is disagreement among different experts on the description of ROI³⁰. Semi-automatic segmentation assists humans in segmentation through tools developed to characterize images³¹. Automatic segmentation has benefited from the recent development of image segmentation techniques in deep learning, and most medical image segmentation models are now developed and improved based on U-net³². Medical image segmentation based on deep learning can accomplish various segmentation tasks, such as tumor segmentation, organ segmentation, etc., with strong applicability. Automatic segmentation also avoids intra- and inter-observer variability of radiomics features. In our study, the extraction of ROI is accomplished using automatic segmentation in concert with manual segmentation. The deep learning segmentation model nnU-net is first used to make an initial sketch of the lesion in the image, which is then refined with the help of an experienced radiologist³⁰. Through such a series of operations, the lesion can be extracted more quickly and accurately, laying the foundation for subsequent image information extraction.

Radiomics is an emerging field that aims to extract information from medical images to support clinical decision-making accordingly³³. In the SCAP recognition task, traditional severity scoring systems focus on physiological features and use less of the image's information. Identifying patients with COVID-19 pneumonia was a significant challenge at the beginning of the COVID-19 epidemic. Investigators performed quantitative analyses of chest CTs to assess lung involvement in patients, ultimately demonstrating that quantitative CT scores are effective in diagnosing COVID-19 pneumonia, determining disease severity, and predicting mortality³⁴. This shows the need for radiomics features as an aid to clinical decision-making. In our study, 384 original images and their derived image features were mined using the radiomics feature extraction tool pyradiomics. Next, feature screening methods were used to filter out valuable features to the task, resulting in 15 imaging features.

In our study, a total of eight machine learning algorithms were selected for evaluation, mainly including integrated learning based Ada Boost, XGBoost, Random Forest, and LightGBM^{35–37}; Support vector machines based on kernel functions³⁸; Instance-based learning algorithm KNN³⁹; Naive Bayes based on Bayes' theorem and the assumption of conditional independence of features⁴⁰. Next, we used the imaging features, the clinical features, and the combination of the two as inputs to the machine learning models, respectively, and trained 24 sets of models. A comparative analysis revealed that when using a combination of features as inputs to the machine learning model, the model obtained the highest AUC value on the test set: 0.89. We also found that models trained using imaging features outperformed those using clinical features. This is further evidence of the importance of features in images.

This finding is also validated by other researchers. For example, Wang et al. found in their study on prognosis prediction in COVID-19 patients that the AUC was 0.843 when using radiomic features; 0.813 when using clinical features; and when using both combined features, the AUC was higher than the former, reaching 0.865⁴¹. Additionally, we found that some researchers used deep learning models to extract features from images, achieving better results than clinical features^{13,42}. Although we must acknowledge that this method can yield excellent results, the features extracted are difficult to understand and interpret. In contrast, the features we use for prediction are specific and meaningful. These multimodal features used for prediction offer both predictive performance and interpretability⁴³.

Finally, we performed an interpretable analysis of the model using XAI methods (SHAP). This method effectively identifies the specific features that drive the model's decisions. In contrast, the features obtained by deep learning models are often directly understandable and usable by humans⁴⁴. Moreover, in our work, using SHAP, we discovered the features that most significantly impact the model's classification results. Some clinical studies have validated our findings¹⁸:

- "Lymphocyte" indicates the number of white blood cells in a patient's blood. Studies have found that a decrease in lymphocyte count is a critical factor for patients transitioning to SCAP, which is consistent with our SHAP analysis results⁴⁵.
- "Albumin" is a routine indicator reflecting a patient's nutritional status and systemic inflammation. In our interpretable model, higher albumin levels in patients make them more likely to be predicted as having mild symptoms. Conversely, lower levels indicate a higher likelihood of deterioration. This aligns with existing research conclusions⁴⁶.

In addition to the clinical features mentioned above, the study of radiomic features is equally valuable. We found that the most important features belong to shape features, which describe the shape of the region of interest. These features abstract the shape of pulmonary CT lesions and are crucial for the model to make accurate predictions.

Our study also has some limitations. Firstly, our study involves a small number of samples, which is prone to overfitting during the training of machine learning models. Secondly, the model's predictive performance is not high enough; in the follow-up work, more data will be collected, the model will be optimized and adjusted, and better performance will be achieved.

Conclusions

In conclusion, we used radiomics tools to measure radiological features from chest CT, combined them with patient clinical features, and obtained good performance in identifying patients with SCAP. Meanwhile, our experimental study suggests that the research methodology of using radiomics tools to explore features hidden in medical images and unrecognizable by humans deserves to be explored and investigated in depth.

Data availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Code availability

We open the core code, can be get by visiting <https://github.com/COOk921/Identifying-SCAP-Using-Radiomics>.

Received: 29 February 2024; Accepted: 5 September 2024

Published online: 19 September 2024

References

1. Niederman, M. S. & Torres, A. Severe community-acquired pneumonia. *Eur. Respir. Rev.* **31**(166), 220123. <https://doi.org/10.1183/16000617.0123-2022> (2022).
2. Lim, W. S. *et al.* Defining community acquired pneumonia severity on presentation to hospital: An international derivation and validation study. *Thorax* **58**(5), 377–382. <https://doi.org/10.1136/thorax.58.5.377> (2003).
3. Fine, M. J. *et al.* A prediction rule to identify low-risk patients with community-acquired pneumonia. *N. Engl. J. Med.* **336**(4), 243–250. <https://doi.org/10.1056/nejm199701233360402> (1997).
4. Metlay, J. P. *et al.* Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American thoracic society and infectious diseases society of America. *Am. J. Respir. Crit. Care Med.* **200**(7), e45–e67. <https://doi.org/10.1164/rccm.201908-1581ST> (2019).
5. Chalmers, J. D. Identifying severe community-acquired pneumonia: Moving beyond mortality. *Thorax* **70**(6), 515–516. <https://doi.org/10.1136/thoraxjnl-2015-207090> (2015).

6. Pauls, S. *et al.* Assessment of pneumonia severity: Multidetector-row CT in comparison to clinical score CRB-65. *Clin. Imaging* **32**(5), 342–345. <https://doi.org/10.1016/j.clinimag.2007.12.008> (2008).
7. Younus, S., Maqsood, H., Sattar, A., Younas, A. & Shakeel, H. A. A novel chest CT severity score in COVID-19 and its correlation with severity and prognosis of the lung disease: A retrospective cohort study. *Ann. Med. Surg.* **82**, 104692. <https://doi.org/10.1016/j.amsu.2022.104692> (2022).
8. Lambin, P. *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**(4), 441–446. <https://doi.org/10.1016/j.ejca.2011.11.036> (2012).
9. Mayerhoefer, M. E. *et al.* Introduction to radiomics. *J. Nucl. Med.* **61**(4), 488–495. <https://doi.org/10.2967/jnumed.118.222893> (2020).
10. Wilson, R. & Devaraj, A. Radiomics of pulmonary nodules and lung cancer. *Transl. Lung Cancer Res.* **6**(1), 86–91. <https://doi.org/10.21037/tlcr.2017.01.04> (2017).
11. Mao, N. *et al.* Added value of radiomics on mammography for breast cancer diagnosis: A feasibility study. *J. Am. Coll. Radiol.* **16**(4), 485–491. <https://doi.org/10.1016/j.jacr.2018.09.041> (2019).
12. Le Dinh, T., Lee, S.-H., Kwon, S.-G. & Kwon, K.-R. COVID-19 chest X-ray classification and severity assessment using convolutional and transformer neural networks. *Appl. Sci.* **12**(10), 4861. <https://doi.org/10.3390/app12104861> (2022).
13. Guarraasi, V. & Sodan, P. Multi-objective optimization determines when, which and how to fuse deep networks: An application to predict COVID-19 outcomes. *Comput. Biol. Med.* **154**, 106625. <https://doi.org/10.1016/j.compbiomed.2023.106625> (2023).
14. Signoroni, A. *et al.* BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. *Med. Image Anal.* **71**, 102046. <https://doi.org/10.1016/j.media.2021.102046> (2021).
15. Li, M. D. *et al.* Automated assessment and tracking of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks. *Radiol. Artif. Intell.* **2**(4), e200079. <https://doi.org/10.1148/ryai.2020200079> (2020).
16. Wei, P. Radiomics, deep learning and early diagnosis in oncology. *Emerg. Top. Life Sci.* **5**(6), 829–835. <https://doi.org/10.1042/ETLS20210218> (2021).
17. Sze, V. *et al.* Efficient processing of deep neural networks: A tutorial and survey. *Proc. IEEE* **105**(12), 2295–2329. <https://doi.org/10.1109/JPROC.2017.2761740> (2017).
18. Francesco, P. *et al.* Explainable machine-learning models for COVID-19 prognosis prediction using clinical, laboratory and radiomic features. *IEEE Access* **11**, 121492–121510. <https://doi.org/10.1109/ACCESS.2023.3327808> (2023).
19. Faul, F. *et al.* G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**(2), 175–191. <https://doi.org/10.3758/BF03193146> (2007).
20. Mandell, L. A. *et al.* Infectious diseases society of America/American thoracic society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin. Infect. Dis.* **44**(2), S27–S72. <https://doi.org/10.1086/511159> (2007).
21. Isensee, F. *et al.* nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211. <https://doi.org/10.1038/s41592-020-01008-z> (2021).
22. van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H. & Baessler, B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging* **11**(1), 91. <https://doi.org/10.1186/s13244-020-00887-2> (2020).
23. Van Griethuysen, J. J. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**(21), e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339> (2017).
24. Zwanenburg, A. *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**(2), 328–338. <https://doi.org/10.1148/radiol.2020191145> (2020).
25. Duarte, F. S. *et al.* Decomposing time series into deterministic and stochastic influences: A survey. *Digit. Signal Process.* **95**, 102582. <https://doi.org/10.1016/j.dsp.2019.102582> (2019).
26. Sabilla, *et al.* Optimizing threshold using Pearson correlation for selecting features of electronic nose signals. *Int. J. Intell. Eng. Syst.* **12**(6), 81–90. <https://doi.org/10.22266/ijies2019.1231.08> (2019).
27. McKnight, P. E. & Najab, J. Mann-Whitney U test. In *The Corsini Encyclopedia of Psychology* 1–1 (Wiley, New York, 2010). <https://doi.org/10.1002/9780470479216.corpsy0524>.
28. Lundberg, S. M. & Su-In, L. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
29. Kumar, V. *et al.* Radiomics: The process and the challenges. *Magn. Reson. Imaging* **30**(9), 1234–1248. <https://doi.org/10.1016/j.mri.2012.06.010> (2012).
30. Yushkevich, P. A., Gao, Y., Gerig, G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3342–3345 (2016). <https://doi.org/10.1109/EMBC.2016.7591443>
31. O'Donnell, L. L. *Semi-automatic Medical Image Segmentation* (Massachusetts Institute of Technology, Cambridge, MA, 2001).
32. Yin, X.-X. *et al.* U-net-based medical image segmentation. *J. Healthcare Eng.* **202**, 1–16. <https://doi.org/10.1155/2022/4189781> (2022).
33. Rizzo, S. *et al.* Radiomics: The facts and the challenges of image analysis. *Eur. Radiol. Exp.* **2**(1), 1–8. <https://doi.org/10.1186/s41747-018-0068-z> (2018).
34. Topal, Ü., Yılmaz, G. & Şahin, A. S. Are the thorax computed tomography findings of ICU patients diagnosed with COVID-19 pneumonia related to the duration of hospital stay and mortality?. *J. Infect. Chemother.* **29**(5), 495–501. <https://doi.org/10.1016/j.jiac.2022.12.016> (2023).
35. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
36. Chen, T. *et al.* Xgboost: Extreme gradient boosting. *R Package Version 0.4-2* **1**(4), 1–4 (2015).
37. Schapire, R. E. Explaining adaboost. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik* 37–52 (Springer, Cham, 2013).
38. Huang, S. *et al.* Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* **15**(1), 41–51. <https://doi.org/10.21873/cgp.20063> (2018).
39. Zhang, M.-L. & Zhou, Z.-H. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.* **40**(7), 2038–2048. <https://doi.org/10.1016/j.patcog.2006.12.019> (2007).
40. Webb, G. I., Keogh, E. & Miikkulainen, R. Naïve Bayes. *Encycl. Mach. Learn.* **15**(1), 713–714 (2010).
41. Wang, D. *et al.* Study on the prognosis predictive model of COVID-19 patients based on CT radiomics. *Sci. Rep.* **11**(1), 11591. <https://doi.org/10.1038/s41598-021-90991-0> (2021).
42. Bijari, S. *et al.* Radiomics and deep features: Robust classification of brain hemorrhages and reproducibility analysis using a 3D autoencoder neural network. *Bioengineering* **11**(7), 643. <https://doi.org/10.3390/bioengineering11070643> (2024).
43. Mahboubisarighieh, A. *et al.* Assessing the efficacy of 3D Dual-CycleGAN model for multi-contrast MRI synthesis. *Egypt. J. Radiol. Nucl. Med.* **55**(1), 1–12. <https://doi.org/10.1186/s43055-024-01287-y> (2024).
44. Li, Z. *et al.* A deep-learning-based framework for severity assessment of COVID-19 with CT images. *Expert Syst. Appl.* **185**, 115616. <https://doi.org/10.1016/j.eswa.2021.115616> (2021).
45. Güell, E. *et al.* Impact of lymphocyte and neutrophil counts on mortality risk in severe community-acquired pneumonia with or without septic shock. *J. Clin. Med.* **8**(5), 754. <https://doi.org/10.3390/jcm8050754> (2019).
46. Chen, L., Ye-Lu, X. & Qing-Zhu, C. Prognostic value of albumin-red cell distribution width score in patients with severe community-acquired pneumonia. *Ann. Palliat. Med.* **9**(3), 75965–75765. <https://doi.org/10.21037/apm.2020.04.22> (2020).

Author contributions

T.Y. and L.Z. contributed equally to this work; Y.G. and L.W. conceived and led the study; T.Y. and L.Z. drafted the manuscript; S.S., X.Y, T.Y. and L.Z. participated in data collection and processing; X.Y. is responsible for the visualization of experimental results. All authors have read and approved the final manuscript.

Funding

This work was supported by the Hebei Province Science Development Program [20201246].

Competing interests

The authors declare no competing interests.

Ethics declarations

This study has been approved by the Institutional Review Committee of the Affiliated Hospital of North China University of Science and Technology (Approval No. 20220730005). The research was conducted following the 1964 Declaration of Helsinki principles and its later amendments or similar ethical standards. The Institutional Review Committee of the Affiliated Hospital of North China University of Science and Technology waived informed consent due to the retrospective nature of our study. Measures were implemented to ensure the confidentiality and anonymity of the data collected, in accordance with appropriate privacy protection protocols.

Additional information

Correspondence and requests for materials should be addressed to L.W. or Y.G.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024