Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of Pulmonary Embolism in Patients With Chronic Obstructive Pulmonary Disease Using Clinical and Non-contrast CT Radiomic Features

Keke Zhou
The Fourth Affiliated Hospital of School of Medicine, Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University

Yun Gong
The Fourth Affiliated Hospital of School of Medicine, Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University

Doudou Chen
The Fourth Affiliated Hospital of School of Medicine, Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University

Jinxiao Zhang
The Fourth Affiliated Hospital of School of Medicine, Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University

Yu Liu
The Fourth Affiliated Hospital of School of Medicine, Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University

Zhihao Xu
xuzhihao@zju.edu.cn

The Fourth Affiliated Hospital of School of Medicine, Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University

Research Article

Keywords: Chronic obstructive pulmonary disease, Pulmonary embolism, Radiomics, Machine learning, Prediction model

Posted Date: September 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10841580/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Identification of Pulmonary Embolism in Patients With Chronic Obstructive Pulmonary Disease Using Clinical and Non-contrast CT Radiomic Features

Keke Zhou[1], Yun Gong[1], Doudou Chen[1], Jinxiao Zhang[1] , Yu Liu[1] and Zhihao Xu [1,*]

[1] Department of Respiratory, The Fourth Affiliated Hospital of School of Medicine, and Zhejiang University School of Medicine, International Institutes of Medicine, Zhejiang University, *Yiwu,China, 322000.* 19106858418@163.com; 22418647@zju.edu.cn ,22418626@zju.edu.cn , 22318334@zju.edu.cn, 15290858150@163.com,xuzhihao@zju.edu.cn.

* Correspondence: xuzhihao@zju.edu.cn

Abstract

Pulmonary embolism (PE) is a potentially life-threatening complication in patients with chronic obstructive pulmonary disease (COPD), but early recognition is challenging because of substantial overlap in clinical manifestations. We aimed to develop and internally validate a multimodal model integrating clinical variables and radiomic features derived from non-contrast chest computed tomography (CT) to identify concomitant PE in patients with COPD. In this retrospective study, 415 patients with COPD were included, of whom 116 had concomitant PE confirmed by computed tomography pulmonary angiography (CTPA). Radiomic features were extracted from non-contrast chest CT images. Feature selection was performed using a multistep procedure incorporating univariable screening, correlation-based redundancy reduction, minimum redundancy maximum relevance, and least absolute shrinkage and selection operator regression. Clinical, radiomics, and combined models were developed and internally validated. Model performance was assessed using receiver operating characteristic curve analysis, calibration analysis, and decision curve analysis. The combined model showed the highest discriminative performance, with an area under the receiver operating characteristic curve of 0.987 in the internal test cohort, compared with 0.959 for the clinical model and 0.937 for the radiomics model. The combined model also demonstrated good calibration and yielded greater net clinical benefit across a broad range of threshold probabilities. A multimodal model integrating routinely available clinical variables with radiomic

features from non-contrast chest CT showed excellent performance for identifying concomitant PE in patients with COPD. This approach may enable opportunistic PE risk stratification during routine chest CT examinations and help identify patients who may benefit from further CTPA evaluation. External validation is warranted before clinical implementation.

Keywords: Chronic obstructive pulmonary disease; Pulmonary embolism; Radiomics; Machine learning; Prediction model

1.Introduction

Chronic obstructive pulmonary disease (COPD) is a common chronic inflammatory airway disease associated with high morbidity, disability, and mortality. It is characterized by persistent respiratory symptoms and airflow limitation. The Global Burden of Disease Study estimated that COPD will become the third leading cause of death worldwide by 2030[1]. Patients with COPD often have multiple comorbidities. Pulmonary embolism (PE) is a potentially life-threatening thrombotic complication and is not uncommon in patients with COPD, particularly in those with acute exacerbation of COPD (AECOPD). However, the reported prevalence of PE in patients with COPD varies considerably across studies. A meta-analysis reported a pooled prevalence of PE of 17.82% among hospitalized patients with AECOPD[2]. A prospective study conducted in Turkey reported an even higher prevalence of 29.1%[3]. A population-based prospective cohort study from Norway demonstrated that COPD, particularly severe COPD, was independently associated with an increased risk of venous thromboembolism (VTE) and post-VTE mortality[4]. The prognosis of patients with COPD deteriorates significantly after VTE. Systemic hypoxia, endothelial injury, chronic inflammation, and hypercoagulability associated with COPD create a prothrombotic milieu and increase the risk of PE. COPD has therefore been recognized as an independent risk factor for PE[5].

However, early identification of PE in patients with COPD remains challenging. First, AECOPD and PE share many clinical manifestations. Both may present with acute dyspnea, chest tightness, and hypoxemia[6]. The lack of specific clinical signs makes PE easily misdiagnosed as an acute exacerbation of the underlying COPD. Second, some patients with PE do not present with typical manifestations, such as chest pain, hemoptysis, or signs of deep vein thrombosis (DVT). This further increases the risk of missed or delayed diagnosis. A Spanish cohort study showed that concomitant PE was associated with a 2.84-fold increase in the risk of in-hospital mortality among hospitalized patients with COPD[7]. Meta-analyses have also demonstrated that PE is associated with adverse outcomes and increased mortality in patients with

AECOPD[8,9]. Current clinical assessment of PE mainly relies on the Wells score[10], revised Geneva score[11], D-dimer testing, and computed tomography pulmonary angiography (CTPA). Diagnostic studies and meta-analyses have shown that D-dimer has a high negative predictive value for excluding PE[12]. However, its specificity is limited by acute inflammatory conditions, which can lead to false-positive results and reduce its diagnostic utility when used alone[13]. COPD has a distinct pathophysiological background. Acute exacerbations can themselves cause dyspnea, tachycardia, hypoxemia, and elevated D-dimer levels. These changes may compromise the discriminatory performance of conventional PE risk assessment tools. A cross-sectional study showed that both the Wells and Geneva scores had limited diagnostic accuracy for PE in patients with COPD[14]. More than 70% of patients still required CTPA for definitive diagnosis[15], suggesting low diagnostic efficiency and limited applicability in this population. Although CTPA is an important imaging modality for diagnosing PE, it is associated with contrast-related risks, radiation exposure, and increased healthcare costs[16]. Therefore, indiscriminate CTPA screening is not appropriate for all patients with COPD. Early identification of patients at high risk of PE is essential to guide further imaging and timely intervention and may improve clinical outcomes.

Risk prediction models integrate multiple disease-related predictors to estimate an individual's probability of developing a specific outcome. They provide quantitative information for clinical risk stratification and decision-making. With advances in statistical methods and machine learning, prediction models have increasingly been applied to the early identification of PE[17,18]. However, the development of models based on clinically accessible predictors specifically for patients with COPD remains an area of active investigation[19]. A single clinical indicator is insufficient for accurate risk stratification in this population. COPD-specific prediction models integrating multidimensional information are therefore needed. Several important gaps remain in the prediction of PE among patients with COPD. First, conventional clinical prediction models rely mainly on demographic characteristics, medical history, symptoms, and blood gas parameters. These macroscopic variables cannot quantify microscopic pathological changes, such as airway remodeling and pulmonary vascular injury. The highest reported area under the receiver operating characteristic curve (AUC) of existing clinical models for predicting PE in ICU patients with COPD is only 0.829[19]. Second, most existing combined models were developed for the general population with acute PE[20]. They do not adequately account for COPD-specific pathophysiological features, including airway abnormalities, chronic hypoxia, and persistent hypercoagulability. A COPD-specific multimodal prediction framework is still lacking. Third, several thrombotic and inflammatory biomarkers, including D-dimer, red blood cell distribution width (RDW), and systemic immune-inflammation index (SII), reflect hypercoagulability and systemic inflammation. Previous studies reported an AUC of 0.737 for red blood cell distribution width-standard deviation (RDW-SD) in predicting PE among patients with COPD[21]. The combination of SII and D-dimer improved the AUC to 0.841[22]. However, individual laboratory

biomarkers are susceptible to confounding by infection and chronic hypoxia. Their utility as standalone screening tools is therefore limited. Radiomics studies of PE have shown that imaging-derived features can improve the prediction of thrombotic events[23]. However, radiomics alone cannot capture systemic coagulation status or the underlying severity of COPD. Radiomics-only models may therefore have limited generalizability. Recent multicenter studies have demonstrated that multimodal models integrating clinical characteristics, serum biomarkers, and radiomic features can achieve better discrimination and greater clinical net benefit than single-modality models[24]. Multimodal integration has consequently emerged as an important direction for risk prediction in thrombotic diseases.

To address these limitations, a COPD-specific prediction model integrating multidimensional information is needed. In this study, we collected clinical data from patients with COPD and compared demographic characteristics, comorbidities, clinical manifestations, laboratory parameters, and radiomic features between patients with and without PE. Independent predictors of PE were identified. Based on these predictors, we developed and validated a risk prediction model for PE in patients with COPD. Model performance was assessed in terms of discrimination, calibration, and clinical utility. This study aims to provide evidence for the early identification of high-risk patients, optimize indications for CTPA, and ultimately improve clinical outcomes in patients with COPD.

2.Methods

2.1Study Patients

This study was a single-center retrospective cohort study designed to develop and validate a risk prediction model for pulmonary embolism (PE) in patients with chronic obstructive pulmonary disease (COPD), and was conducted in accordance with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement. The study protocol was approved by the Ethics Committee of the Fourth Affiliated Hospital of Zhejiang University School of Medicine (approval number: K2026033), and the requirement for informed consent was waived due to the retrospective nature of the study. We consecutively enrolled COPD patients who were hospitalized between January 2017 and December 2025. The inclusion criteria were: (1) age ≥ 40 years; (2) diagnosis of COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2026 report and the Chinese Guidelines for Primary Care and Management of Chronic Obstructive Pulmonary Disease (2024); (3) completion of computed tomography pulmonary angiography (CTPA) within 7 days of admission; and (4) availability of complete clinical and imaging data. The exclusion criteria were: (1) prior diagnosis of PE or deep vein thrombosis before admission; (2) advanced malignancy or severe hepatic/renal insufficiency; (3) pregnancy or lactation; and (4) severe imaging artifacts or poor image quality precluding radiomic feature extraction.
Based on the events per variable (EPV) principle (at least 10 events per predictor

variable), it was estimated that approximately 8–10 predictor variables would be included in the final model, requiring at least 80–100 PE events. Patients were divided into PE and non-PE groups according to CTPA findings. A random number generator was used to allocate all included patients to the training set (70%) and the internal validation set (30%). (See **Figure 1** for details).

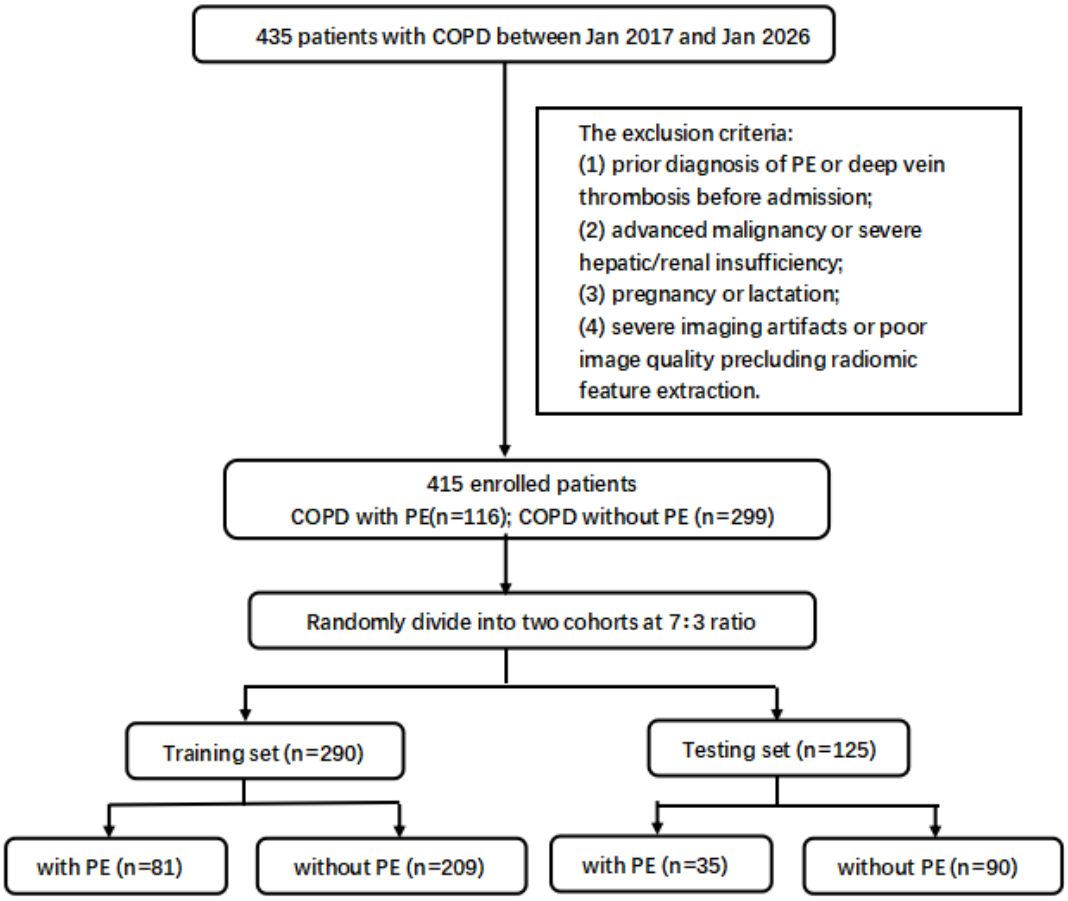


Figure 1. Flow chart for selecting the study population

2.2 Data Collection

2.2.1 Clinical Information Collection

This study was a retrospective cohort study based on hospital electronic medical records. We systematically collected data on patients' demographic characteristics, comorbidities, COPD-related clinical features, admission symptoms, laboratory indicators, and bedrest status. The outcome variable was pulmonary embolism (PE) confirmed by computed tomography pulmonary angiography (CTPA). Demographic characteristics included age, sex, body mass index (BMI), and smoking history. Comorbidities comprised hypertension, diabetes mellitus, coronary artery disease, and prior surgery history. COPD-related features included acute exacerbation of COPD (AECOPD) at admission, disease duration, and GOLD grade. Admission symptoms mainly documented worsening dyspnea, chest pain, lower extremity swelling or pain. Laboratory indicators collected within 24 hours of admission included pulmonary

function parameters (FEV1/FVC, FEV1% predicted), thrombotic and inflammatory biomarkers (D-dimer, fibrinogen, NT-proBNP, CRP), and oxygenation-related indicators (SaO_2, oxygenation index). Bedrest duration (days) was also recorded. All laboratory and imaging examinations were performed using standardized and uniformly calibrated equipment to ensure data reliability. PE was diagnosed according to the European Society of Cardiology (ESC) guidelines, with CTPA showing one or more filling defects or vascular occlusion in the pulmonary arteries or branches. All data were independently extracted and cross-verified by two researchers, with abnormal or missing data rechecked, and cases with severe missing key variables excluded, ensuring a complete and accurate dataset for subsequent feature screening and prediction model development.

2.2.2 Imaging Acquisition

The spiral CT images were acquired with a 16-row spiral CT scanner (GE Optima, USA) with the following parameters: current, 200380 mA (automatic exposure control); potential, 120–140 kV; pitch, 1.625; field of view, 16 cm; and reconstruction thickness, 1.25 mm. The DICOM-format was used as exported for the CT data. Two experienced radiologists with 5 and 10 years of diagnostic expertise independently utilized ITK-SNAP software (version 3.8.0, http:// www. itksnap.org/) to manually delineate the region of interest (ROI). A complete diagram is shown in **Figure 2.**

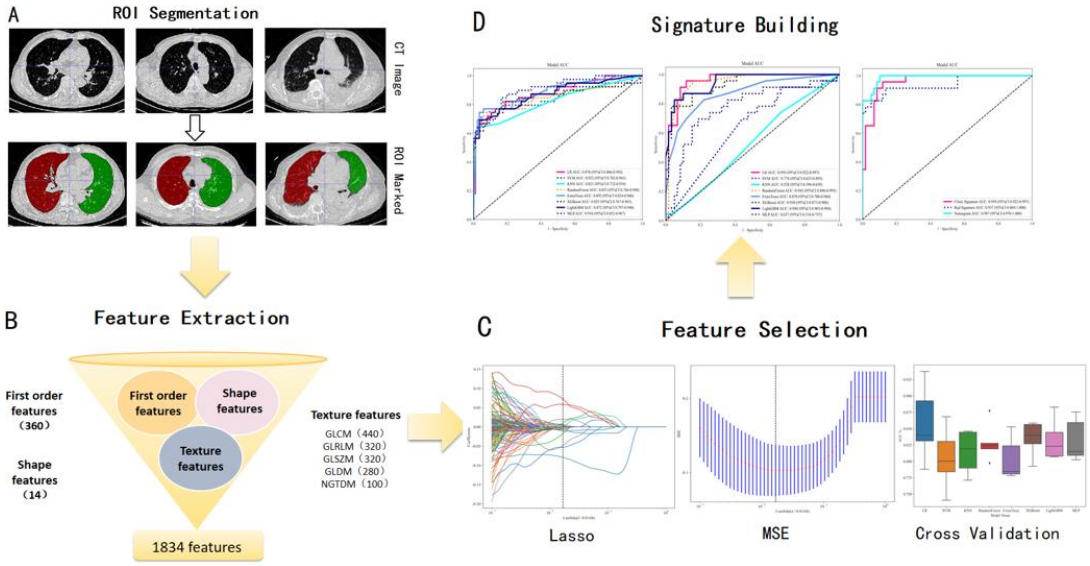


Figure 2 The process of radiomics analysis.

2.3 Procedure of Radiomics Analysis

2.3.1 Image Segmentation and Feature Extraction

This study performed radiomic feature extraction based on chest CT images. Prior to feature extraction, all CT images were standardized following the guidelines of the Imaging Biomarker Standardization Initiative (IBSI), including isotropic resampling

to a uniform voxel size and gray-level discretization, to reduce image heterogeneity arising from variations in scanning equipment and reconstruction parameters, thereby ensuring feature reproducibility and comparability. All original CT images in DICOM format were imported into ITK-SNAP software. Two radiologists with over 5 years of experience in chest imaging independently delineated the regions of interest (ROIs) of the bilateral lung parenchyma using a threshold-based segmentation method combined with manual correction. The segmentation encompassed the entire bilateral lung fields, while carefully excluding the trachea, main bronchi, and hilar great vessels; slice-by-slice review and adjustment of segmentation boundaries were performed to ensure ROI accuracy. To assess inter-observer agreement, 20% of the samples were randomly selected and independently segmented by the two radiologists, and the intraclass correlation coefficient (ICC) was calculated. An ICC > 0.75 was considered indicative of good agreement, and only samples meeting this consistency threshold were retained for subsequent feature extraction. Radiomic feature extraction was performed using the PyRadiomics open-source package (version 3.1.0) within the Python environment (version 3.7.12). To comprehensively capture the gray-level distribution, morphological structure, textural heterogeneity, and spatial information of the lung parenchyma, multiple preprocessing transformations were applied to the original images, including square, exponential, gradient, square root, logarithm, stationary wavelet transform, 2D local binary pattern (lbp-2D), and 3D local binary pattern (lbp-3D). The wavelet transform employed multi-directional stationary wavelet decomposition; lbp-3D further derived several subclasses including kurtosis map (lbp-3D-k), lbp-3D-m1, and lbp-3D-m2, corresponding to different orders of spherical harmonics computation. Radiomic features were extracted from each ROI, encompassing first-order statistical features, shape-based features, texture features, and higher-order spatial distribution features, yielding a total of 1,834 original radiomic features. All features were then standardized using Z-score normalization to eliminate inter-feature scale differences, providing a standardized, high-dimensional dataset for subsequent feature selection and model construction.

2.3.2 Feature Selection

In this study, a multi-level stepwise screening strategy was applied to progressively refine the entire set of radiomic features, sequentially incorporating differential analysis, correlation-based redundancy elimination, minimum redundancy maximum relevance (mRMR) selection, and least absolute shrinkage and selection operator (LASSO) regularization, ultimately yielding a stable, parsimonious radiomic feature subset with optimal predictive performance.

First, all original radiomic features underwent statistical differential screening. The non-parametric Mann-Whitney U test was employed to identify features with significant inter-group differences, and only those with a p-value < 0.05 were retained for subsequent analysis, excluding ineffective parameters lacking discriminative capacity between groups.

Subsequently, Spearman's rank correlation analysis was conducted to eliminate high collinearity and redundant information among features. Pairwise correlation testing was performed on candidate features. When the correlation coefficient between any two features exceeded 0.9, they were deemed highly redundant, and one of the two was preferentially retained. To maximally preserve the discriminative information of the features, a recursive iterative elimination strategy was adopted: in each iteration, the single most redundant feature in the current set was removed, and the correlations among the remaining features were recalculated. This process was iterated until a well-balanced feature subset characterized by low redundancy and high independence was achieved. Following this correlation-based iterative screening, a total of 290 radiomic features were retained.

To further reduce model dimensionality and mitigate the risk of overfitting, the 290 preliminarily screened features were subjected to minimum redundancy maximum relevance (mRMR) feature selection. Based on the comprehensive ranking of feature relevance and redundancy generated by the mRMR algorithm, the top-ranked candidate features were selected for subsequent LASSO regression.

Finally, the LASSO L1-regularized regression model was employed for ultimate dimensionality reduction and radiomic signature construction on the 20 candidate features. The LASSO algorithm compresses feature weights through a regularization penalty coefficient, automatically nullifying and eliminating the coefficients of features without predictive value, thereby effectively simplifying the model structure and enhancing generalization capability. In this study, the optimal regularization coefficient λ was determined by 10-fold cross-validation using the minimum-error criterion, with the λ value corresponding to the lowest cross-validation error selected as the optimal model parameter. Under the optimal λ, all features with non-zero coefficients were retained as the core features for the final radiomic signature. Based on the selected features and their corresponding regression coefficients, an individualized radiomic score (Rad-score) was calculated for each patient via linear weighted fitting. All LASSO regression modeling procedures in this study were implemented using the scikit-learn package within the Python environment.

2.3.3 Signature building

2.3.3.1 Radiomics Signature

The final radiomic feature subset was obtained through LASSO regression-based screening. These selected features were then used as inputs to construct multiple machine learning classifiers, including logistic regression (LR), support vector machine (SVM), random forest, and extreme gradient boosting (XGBoost). Five-fold cross-validation was performed within the training set to optimize the hyperparameters of each model, yielding a set of generalizable radiomic models. Among these, the

linear weighted score derived from the LASSO-logistic regression model was defined as the radiomic signature (Rad-score).

2.3.3.2 Clinical Signature

For clinical signature construction, all pre-defined clinical variables were enrolled as candidate predictors, including demographic characteristics, comorbidities, COPD-related clinical indexes, admission symptoms, thrombotic and inflammatory biomarkers, and oxygenation parameters. Univariate logistic regression was initially performed to screen potential predictive factors, and only variables with $P < 0.05$ were retained. Significant clinical variables were further entered into multivariate logistic regression with backward stepwise elimination to screen out independent clinical risk factors. Finally, the identified independent clinical predictors were imported into the same machine learning algorithm panel (logistic regression, SVM, random forest, and XGBoost) as utilized for radiomic modeling to establish clinical prediction models.

2.3.3.3 Radiomic nomogram

The radiomic nomogram was generated based on the optimal radiomic features screened by LASSO regression. In the test cohort, the diagnostic efficacy of the radiomic nomogram was assessed via receiver operating characteristic (ROC) curves. Calibration curves were plotted to evaluate the prediction accuracy, and Hosmer–Lemeshow test was performed to quantitatively verify calibration consistency. Furthermore, decision curve analysis (DCA) was conducted to evaluate the clinical net benefit of the radiomic model.

2.4 Statistical analysis

All data processing and statistical analyses in this study were performed using Python and R software. Continuous variables that did not conform to a normal distribution (Shapiro–Wilk test, all $P < 0.05$) were summarized as medians with interquartile ranges (M [Q1–Q3]) and compared between groups using the Mann–Whitney U test. Categorical variables were expressed as percentages (%) and analyzed using the Pearson χ^2 test; Fisher's exact test was applied when the expected frequency was less than 5. Missing data were minimal (BMI: 6 cases [1.4%]; smoking history: 2 cases [0.5%]) and were handled using complete-case analysis, with missing data excluded from all analyses.

Eight machine learning algorithms—logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), k-nearest neighbors (KNN), extremely randomized trees (ExtraTrees), and multi-layer perceptron (MLP)—were employed to construct the radiomic, clinical, and combined prediction models. Five-fold cross-validation was conducted within the training set for hyperparameter optimization and model stability assessment.

Model performance was comprehensively evaluated from three dimensions: discrimination, calibration, and clinical net benefit. Discrimination performance was

assessed using the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals. Pairwise comparisons of AUC values between different models were performed using the DeLong test implemented in the pROC package. Additionally, accuracy, sensitivity, specificity, positive and negative predictive values, F1 score, and Kappa coefficient were calculated for each model. The optimal probability cut-off value for each model was determined in the training cohort via 5-fold cross-validation using the Youden index (maximizing sensitivity + specificity − 1). The obtained cut-off was then fixed and applied to the internal hold-out test cohort to calculate sensitivity, specificity, accuracy, positive predictive value, negative predictive value, F1 score, and Kappa coefficient. Calibration curves were plotted to assess the agreement between predicted risks and observed outcomes, and decision curve analysis (DCA) was performed to quantify the clinical net benefit and practical applicability of each prediction model. The intraclass correlation coefficient (ICC) based on a two-way random-effects absolute agreement model, combined with 1000 bootstrap iterations, was used to evaluate inter-observer agreement in radiomic feature delineation. All statistical tests were two-sided, and a P-value < 0.05 was considered statistically significant.

3.Results

We randomly partitioned all the data, allocating 70% as the training set and the remaining 30% as the testing set. For the training set, we employed 5-fold cross-validation to fine-tune the hyperparameters.

3.1Radiomics Analysis

In total, 1834 handcrafted features were extracted across seven c-ategories for each modality, encompassing 360 first-order features, 14 shape features, and a concluding set of texture features. As depicted in **Figure 3**a,b, which presents features of ratios and p-value distribution of handcrafted radiomic features; Following the application of mRMR selection, we employed the t-SNE method for visualization of all samples. As depicted in **Figure 3**c , our omics features could effectively distinguish COPD patients with pulmonary embolism from those without, indicating a robust dis criminative capacity.

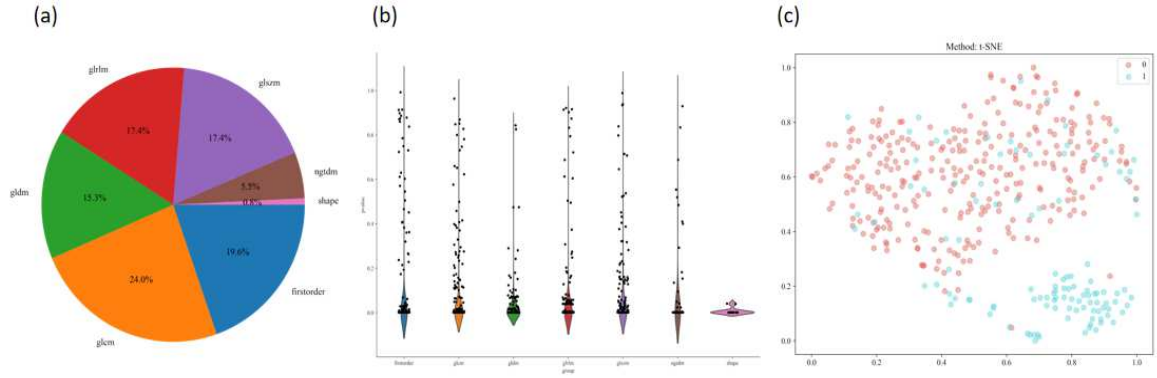


Figure 3 Ratio (a) and P-value distribution (b) of handcrafted radiomic features; (c) the t-SNE method for visualization of all samples. firstorder, first order features; glcm, gray level co-occurrence matrix features; gldm, gray level dependence matrix features; glrlm, gray level run length matrix features; glszm, gray level size zone matrix features; ngtdm, neighbouring gray tone difference matrix features; shape, shape features; t-SNE, t-distributed stochastic neighbor embedding.

To identify predictive factors and mitigate the risk of overfitting inherent to high-dimensional radiomic features, we performed penalized dimensionality reduction using LASSO regression (**Figure 4**a), The optimal penalty coefficient $\lambda = 0.0168$ was determined via 10-fold cross-validation. The coefficient profile curves demonstrated that as the penalty strength increased, the coefficients of most features progressively shrank toward zero, suggesting that the original feature set contained numerous redundant variables with limited predictive utility. The mean-squared-error curve (**Figure 4**b) revealed a marked increase in model error when λ exceeded the optimal value, indicating that excessive penalization would eliminate crucial predictive features and consequently impair model performance. Ultimately, 20 features with non-zero coefficients were retained at the optimal λ, encompassing wavelet-transformed, first-order, textural, and morphological features. This observation suggests that the discriminative information relevant to the outcome is distributed across multidimensional imaging signatures rather than concentrated within a single feature category. The histogram of the radiomics score based on the selected features (**Figure 4**c), reflecting divergent directional effects—namely, risk-promoting and protective factors. This systematic procedure effectively eliminated collinear and noisy features, yielding a reliable feature subset for subsequent construction of a robust prediction model.

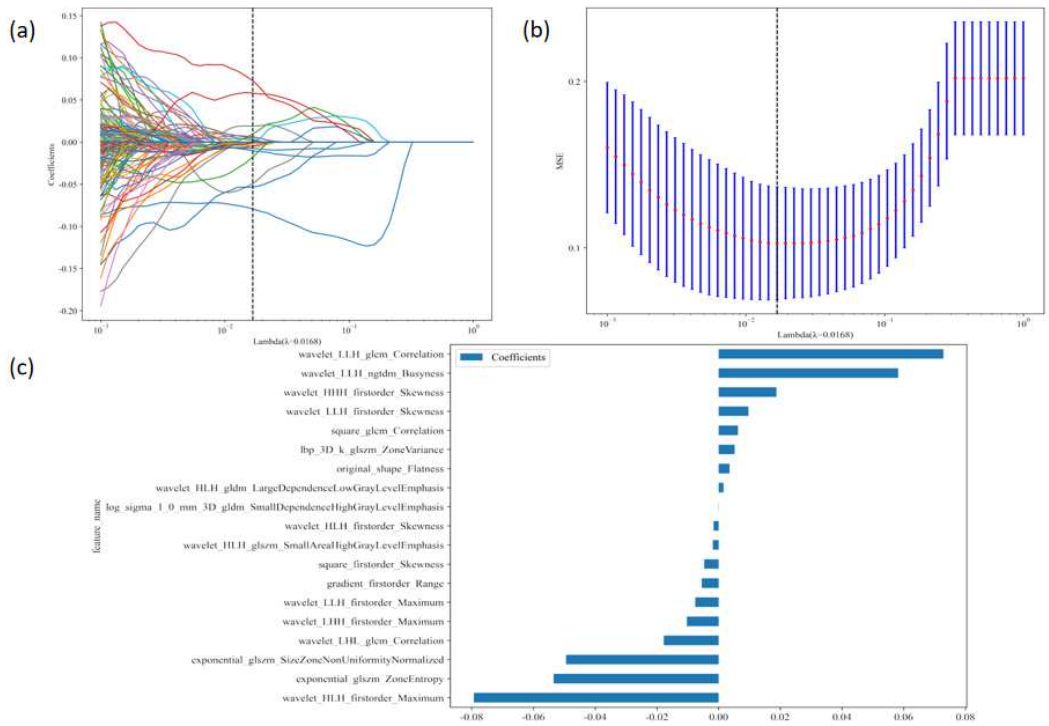


Figure 4 (a) LASSO coefficient profile plot with different λ was shown.,The vertical dashed lines represent radiomics features with nonzero coefficients selected with the optimal λ value (0.0168); (b) Mean square error of 10-fold cross-validation of the LASSO model. The vertical dashed lines indicate the optimal λ value (0.0168); (c) The histogram of the radiomics score based on the selected features. firstorder, first order features; glcm, gray level co-occurrence matrix features; gldm, gray level dependence matrix features; glrlm, gray level run length matrix features; glszm, gray level size zone matrix features; ngtdm, neighbouring gray tone difference matrix features; shape, shape features; wavelet, wavelet transform filtered features; square, square filtered features; exponential, exponential filtered features; log_sigma, Laplacian of Gaussian filtered features; gradient, gradient filtered features; lbp, 3D local binary pattern filtered features; MSE, mean squared error.

To comprehensively compare the predictive performance of different machine learning algorithms, we constructed eight prediction models, as shown in **Table 1**, The discrimination performance of each model was visualized by ROC curves in the training set (**Figure 5a**)and testing set (**Figure 5b**). ()In the training cohort, the XGBoost model achieved the highest AUC of 0.976 (95%CI 0.955–0.996), followed by KNN (AUC=0.949), SVM (AUC=0.947).The LR model exhibited the lowest predictive efficacy among all algorithms (AUC=0.888). In the test cohort, all models showed a mild decline in AUC values compared with the training set. The MLP model yielded the optimal discrimination with an AUC of 0.910 (95%CI 0.852–0.967), whereas KNN remained the worst-performing model (AUC=0.821).
The stability of model performance was further assessed via 5-fold cross-validation, and the distribution of cross-validated AUC values for each model was presented in the box plot (**Figure 5c**).Consistent with the findings from the training ROC curves, MLP and LightGBM maintained relatively high median AUC values and narrow value

distribution, indicating robust and stable predictive ability. By contrast, the KNN model displayed the lowest median AUC and wider data dispersion, demonstrating inferior and less stable discrimination capacity.

Table 1 Radiomics model performance of different machine learning algorithms in the training and testing sets.

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.875	0.888	0.8397 - 0.9365	0.707	0.934	0.791	0.900	0.791	0.707	0.746	0.272	label-train
LR	0.839	0.878	0.8060 - 0.9496	0.769	0.871	0.732	0.892	0.732	0.769	0.750	0.245	label-test
SVM	0.948	0.947	0.9048 - 0.9887	0.933	0.953	0.875	0.976	0.875	0.933	0.903	0.120	label-train
SVM	0.879	0.852	0.7624 - 0.9425	0.744	0.941	0.853	0.889	0.853	0.744	0.795	0.175	label-test
KNN	0.892	0.949	0.9260 - 0.9722	0.813	0.920	0.782	0.933	0.782	0.813	0.797	0.400	label-train
KNN	0.863	0.821	0.7324 - 0.9096	0.641	0.965	0.893	0.854	0.893	0.641	0.746	0.600	label-test
RandomForest	0.927	0.945	0.9099 - 0.9811	0.840	0.958	0.875	0.944	0.875	0.840	0.857	0.230	label-train
RandomForest	0.879	0.853	0.7660 - 0.9402	0.744	0.941	0.853	0.889	0.853	0.744	0.795	0.265	label-test
ExtraTrees	0.892	0.879	0.8265 - 0.9310	0.693	0.962	0.867	0.899	0.867	0.693	0.770	0.198	label-train
ExtraTrees	0.887	0.892	0.8235 - 0.9605	0.769	0.941	0.857	0.899	0.857	0.769	0.811	0.179	label-test
XGBoost	0.962	0.976	0.9548 - 0.9964	0.907	0.981	0.944	0.967	0.944	0.907	0.925	0.271	label-train
XGBoost	0.823	0.855	0.7670 - 0.9425	0.821	0.824	0.681	0.909	0.681	0.821	0.744	0.246	label-test
LightGBM	0.930	0.945	0.9121 - 0.9781	0.853	0.958	0.877	0.949	0.877	0.853	0.865	0.252	label-train
LightGBM	0.879	0.872	0.7972 - 0.9476	0.692	0.965	0.900	0.872	0.900	0.692	0.783	0.322	label-test
MLP	0.906	0.882	0.8340 - 0.9309	0.640	1.000	1.000	0.887	1.000	0.640	0.780	0.701	label-train
MLP	0.839	0.910	0.8523 - 0.9673	0.846	0.835	0.702	0.922	0.702	0.846	0.767	0.176	label-test

Abbreviations: AUC, area under curve; CI, confidence interval; LightGBM, light gradient boosting machine; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; SVM, support vector machine; XGBoost, extreme gradient boosting; KNN, k-nearest neighbor; RandomForest, random forest; ExtraTrees, extremely randomized trees; MLP, multi-layer perceptron; F1, F1-score.

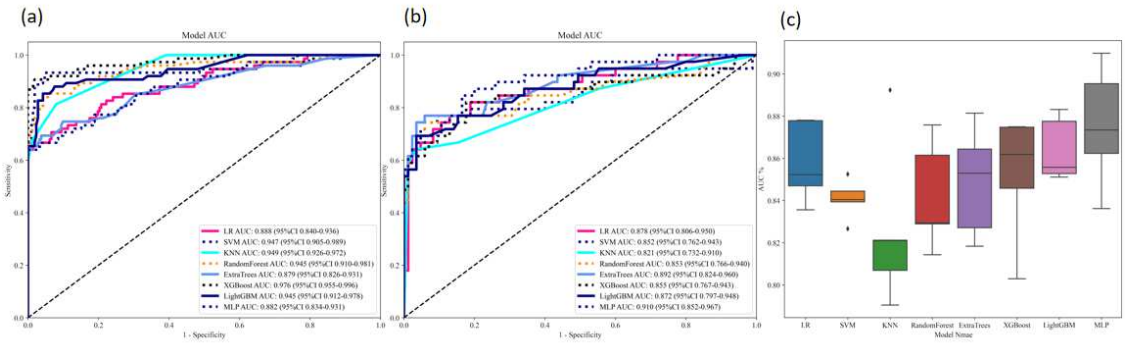


Figure 5 The receiver operating characteristic curves of different machine learning models based on radiomic features in the training cohort (a) and independent testing cohort (b); (c) Box plot of AUC values derived from 5-fold cross-validation on the training dataset. AUC, area under the receiver operating characteristic curve; LightGBM, light gradient boosting machine; LR, logistic regression; SVM, support vector machine; XGBoost, extreme gradient boosting; KNN, k-nearest neighbor; RandomForest, random forest; ExtraTrees, extremely randomized trees; MLP, multi-layer perceptron.

3.2 Baseline Characteristics

Demographics and clinical characteristics of all enrolled COPD patients stratified by pulmonary embolism (PE) are detailed in **Table 2**. The study population consisted of 415 COPD patients (median age 75.0 years, 80.0% male)，among whom 116 patients were complicated with PE (27.95%). The COPD-with-PE group exhibited significantly older age (P=0.005)，and distinct distribution of GOLD grades (P<0.001), with a higher proportion of grade 2 and lower proportion of grade 3 cases. Patients with PE presented markedly higher incidence of chest pain (P=0.003) and lower limb swelling/pain (P=0.039), significantly elevated D-dimer (P<0.001) and NT-proBNP levels (P<0.001), increased FEV1/FVC and FEV1%pred values (P<0.001 and P=0.002, respectively), longer bedridden days (P=0.008), alongside reduced fibrinogen (P=0.016), oxygenation index and arterial oxygen saturation (both P<0.001). Conversely, no significant intergroup differences were observed in sex distribution, BMI, smoking history, hypertension, diabetes mellitus, coronary artery disease, surgical history, dyspnea exacerbation, and C-reactive protein levels (all P>0.05).

Table 2 Baseline characteristics of all patients stratified by PE status

Characteristic	N	Overall[1]	COPD without PE[1]	COPD with PE[1]	P value[2]
Sex	415				0.189
Male		332 (80%)	244 (82%)	88 (76%)	
Female		83 (20%)	55 (18%)	28 (24%)	
Age (years)	415	75.0 (70.0 - 80.0)	75.0 (69.0 - 80.0)	77.0 (72.0 - 82.5)	0.005
BMI (kg/m²)	409	22.0 (19.6 - 24.4)	22.0 (19.5 - 24.4)	22.2 (20.3 - 24.4)	0.326
Missing		6	1	5	
Smoking history (years)	413	30.0 (0.0 - 40.0)	30.0 (0.0 - 40.0)	20.0 (0.0 - 40.0)	0.189
Missing		2	0	2	
Hypertension	415	213 (51%)	152 (51%)	61 (53%)	0.749
Diabetes mellitus	415	72 (17%)	58 (19%)	14 (12%)	0.077
Coronary artery disease	415	150 (36%)	110 (37%)	40 (34%)	0.661
Surgery history	415	120 (29%)	90 (30%)	30 (26%)	0.393
Acute exacerbation of COPD	415	329 (79%)	245 (82%)	84 (72%)	0.032
COPD duration (years)	415	10.0 (5.0 - 20.0)	10.0 (6.0 - 20.0)	10.0 (3.0 - 17.5)	<0.001
GOLD grade	415				<0.001
1		13 (3.1%)	5 (1.7%)	8 (6.9%)	
2		116 (28%)	69 (23%)	47 (41%)	
3		203 (49%)	167 (56%)	36 (31%)	
4		83 (20%)	58 (19%)	25 (22%)	
Dyspnea exacerbation	415	378 (91%)	273 (91%)	105 (91%)	0.801
Chest pain	415	59 (14%)	33 (11%)	26 (22%)	0.003
Lower limb swelling/pain	415	87 (21%)	55 (18%)	32 (28%)	0.039
FEV1/FVC (%)	415	50.2 (41.3 - 61.9)	47.9 (41.0 - 58.2)	59.1 (43.7 - 65.3)	<0.001
FEV1% predicted (%)	415	42.4 (31.9 - 55.3)	40.8 (31.7 - 49.6)	49.2 (33.3 - 62.9)	0.002
D-dimer (mg/L)	415	0.9 (0.4 - 2.8)	0.6 (0.4 - 1.1)	6.6 (2.7 - 14.3)	<0.001
Fibrinogen (g/L)	415	4.0 (3.1 - 5.2)	4.2 (3.1 - 5.3)	3.7 (3.0 - 4.7)	0.016
NT-proBNP (pg/mL)	413	127.0 (34.2 - 549.3)	95.7 (26.2 - 347.1)	310.5 (117.5 - 1,467.7)	<0.001
Missing		2	0	2	
C-reactive protein (mg/L)	415	35.4 (9.9 - 92.5)	35.4 (7.3 - 92.3)	35.8 (13.9 - 97.8)	0.464
Oxygenation index	415	287.0 (225.0 - 349.0)	296.0 (239.0 - 355.0)	245.0 (203.5 - 322.5)	<0.001
Oxygen saturation (%)	415	94.3 (92.3 - 96.0)	94.9 (93.0 - 96.3)	93.5 (89.8 - 95.0)	<0.001
Bedridden days	415	3.0 (2.0 - 7.0)	3.0 (2.0 - 7.0)	4.5 (2.0 - 9.0)	0.008

[1] Median (Q1-Q3) or n (%). Comparisons: Wilcoxon rank-sum test for continuous variables; Chi-square test (or Fisher's exact test when expected cell count <5) for categorical variables. Missing data were excluded from analysis.

[2] Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test

We first analyzed the correlation between candidate clinical predictors using Spearman correlation analysis (**Figure 6c**). Moderate to strong correlations were observed between several pulmonary function indicators (FEV1/FVC and FEV1% pred) as well as oxygenation-related indicators (OI and SaO2), whereas most remaining clinical variables exhibited weak pairwise correlations, indicating limited multicollinearity.Subsequently, eight machine learning algorithms including LR, SVM, KNN, Random Forest, Extra Trees, XGBoost, LightGBM and MLP were constructed based on clinical features. The receiver operating characteristic curves of all models in the training cohort and independent testing cohort are presented in **Figure 6a** and **Figure 6b**, and corresponding quantitative predictive metrics were summarized in the **Table 3**. In the training cohort, Extra Trees and Random Forest, XGBoost achieved outstanding discriminative performance with AUC values reaching 1.000, 0.999 and 0.999, respectively. In the testing cohort, LR yielded the highest discriminative performance with an AUC of 0.959, followed by LightGBM (AUC = 0.948), Random Forest (AUC = 0.945), and XGBoost (AUC = 0.930). By contrast, KNN and MLP displayed poor generalization ability, with obvious performance decline from training to testing datasets. Overall, tree-based ensemble models (LightGBM, RandomForest, XGBoost) demonstrated favorable discrimination and generalization for identifying the risk of pulmonary embolism.

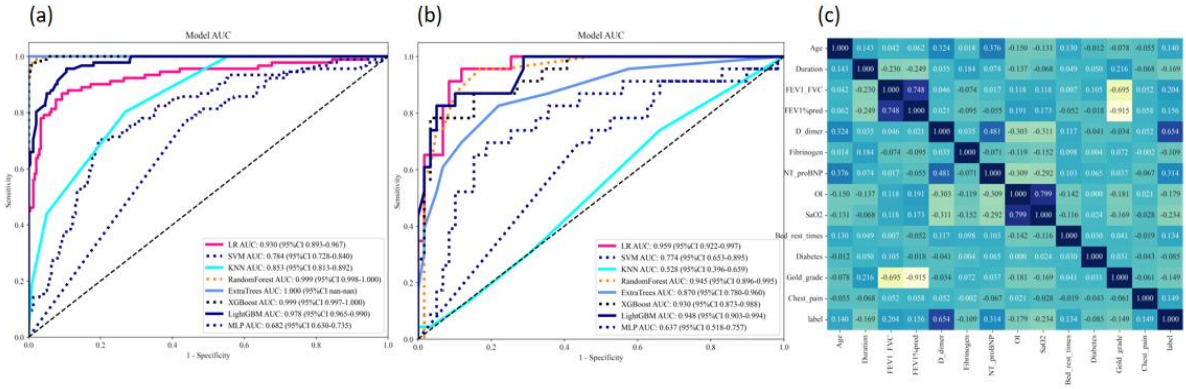


Figure 6 The receiver operating characteristic curves of different machine learning models based on clinical features in the training cohort (a) and independent testing cohort (b); (c) Spearman correlation coefficients among all clinical variables. AUC, area under the receiver operating characteristic curve; LightGBM, light gradient boosting machine; LR, logistic regression; SVM, support vector machine; XGBoost, extreme gradient boosting; RandomForest, random forest; ExtraTrees, extremely randomized trees; KNN, k-nearest neighbor; MLP, multi-layer perceptron.

Table 3 Clinical model performance of different machine learning algorithms in the training and testing sets.

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.897	0.930	0.8928 - 0.9669	0.868	0.908	0.782	0.947	0.782	0.868	0.823	0.219	label-train
LR	0.902	0.959	0.9218 - 0.9971	0.957	0.881	0.759	0.981	0.759	0.957	0.846	0.233	label-test
SVM	0.778	0.784	0.7276 - 0.8402	0.703	0.807	0.582	0.877	0.582	0.703	0.637	0.256	label-train
SVM	0.780	0.774	0.6527 - 0.8949	0.696	0.814	0.593	0.873	0.593	0.696	0.640	0.256	label-test
KNN	0.751	0.853	0.8133 - 0.8918	0.802	0.731	0.533	0.906	0.533	0.802	0.640	0.400	label-train
KNN	0.451	0.528	0.3961 - 0.6591	0.739	0.339	0.304	0.769	0.304	0.739	0.430	0.200	label-test
RandomForest	0.988	0.999	0.9979 - 1.0000	0.989	0.987	0.968	0.996	0.968	0.989	0.978	0.500	label-train
RandomForest	0.866	0.945	0.8964 - 0.9945	0.957	0.831	0.687	0.980	0.687	0.957	0.800	0.300	label-test
ExtraTrees	1.000	1.000	1.0000 - 1.0000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	label-train
ExtraTrees	0.793	0.870	0.7802 - 0.9604	0.826	0.780	0.594	0.920	0.594	0.826	0.691	0.300	label-test
XGBoost	0.985	0.999	0.9973 - 1.0000	0.978	0.987	0.967	0.992	0.967	0.978	0.973	0.409	label-train
XGBoost	0.915	0.930	0.8732 - 0.9875	0.783	0.966	0.900	0.919	0.900	0.783	0.837	0.646	label-test
LightGBM	0.912	0.978	0.9647 - 0.9905	0.956	0.895	0.777	0.982	0.777	0.956	0.857	0.239	label-train
LightGBM	0.915	0.948	0.9026 - 0.9943	0.826	0.949	0.864	0.933	0.864	0.826	0.844	0.333	label-test
MLP	0.620	0.682	0.6299 - 0.7349	0.780	0.559	0.403	0.869	0.403	0.780	0.532	0.553	label-train
MLP	0.585	0.637	0.5178 - 0.7571	0.739	0.525	0.378	0.838	0.378	0.739	0.500	0.487	label-test

Abbreviations: AUC, area under curve; CI, confidence interval; LightGBM, light gradient boosting machine; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; SVM, support vector machine; XGBoost, extreme gradient boosting; KNN, k-nearest neighbor; RandomForest, random forest; ExtraTrees, extremely randomized trees; MLP, multi-layer perceptron; F1, F1-score.

3.3 Signature comparison

The ROC curves illustrated the discriminative performance of the clinical signature, radiomic signature and combined nomogram in the training and testing cohorts (**Figure 7**a,b). The nomogram achieved the highest AUC in both datasets, verifying that integrating clinical and radiomic information improved predictive power. The standalone radiomic signature showed relatively inferior performance, indicating that imaging features alone were insufficient for satisfactory discrimination. The more prominent inter-model difference observed in the test set further reflected that complementary clinical information enhanced the generalizability of the combined model.

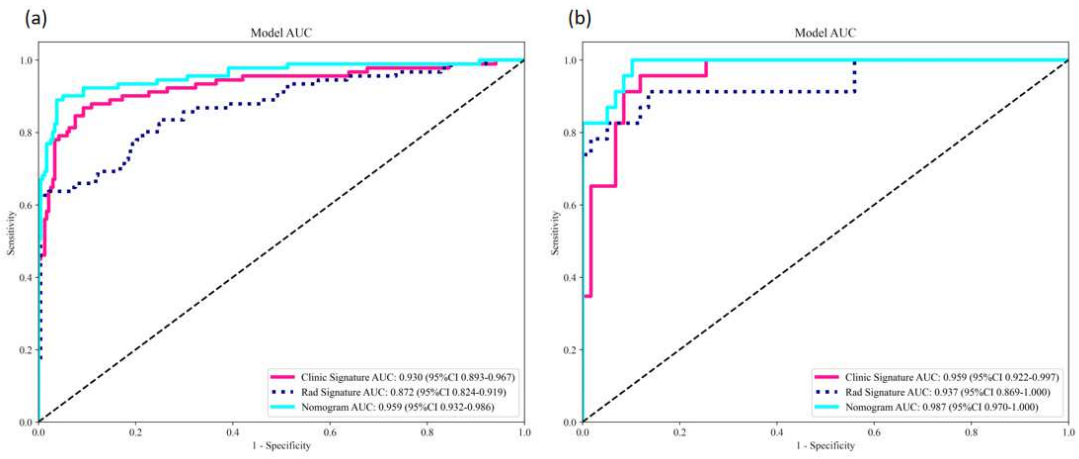


Figure 7 Receiver operating characteristic curves for the radiomic model, clinical model, and combined model in training cohort (a) and test cohort (b). AUC, area under curve.

The calibration plots (**Figure 8**a) revealed favourable concordance between predicted probabilities and observed outcomes for the three models in the test cohort, among which the nomogram showed the best agreement with the ideal calibration reference, while the radiomic-only signature presented obvious deviation. Hosmer-Lemeshow (HL) test was performed to evaluate model goodness-of-fit, and the nomogram achieved satisfactory fitting performance. Decision-curve analysis (**Figure 8**b) further illustrated that the nomogram maintained higher net benefit across most threshold probabilities ranging from 0 to 0.85, outperforming both the clinical-only and radiomic-only models. These results suggested that high AUC alone cannot guarantee reliable clinical practicability. Even if a model obtains satisfactory discriminative ability, poor calibration may lead to biased risk estimation. By fusing clinical variables and radiomic features, the nomogram achieved well-calibrated risk prediction and yielded superior net clinical benefit, which supported its practical value for individualized risk-stratified clinical decision-making.

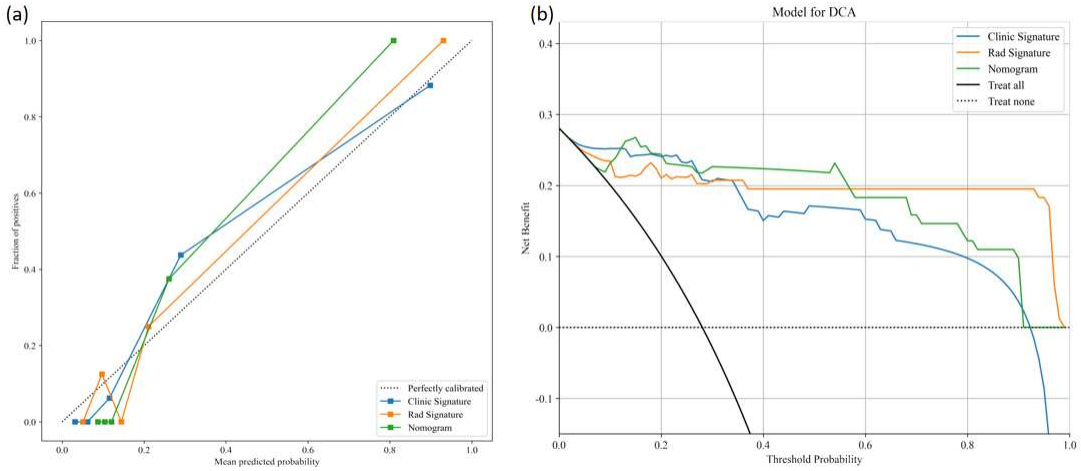


Figure 8 Calibration curve in the test cohort (a). Decision curve analysis (DCA) for three models in the test cohort (b), which shows that the nomogram yields superior net clinical benefit across a wide range of threshold probabilities.

4.Discussion

This study developed and validated clinical, radiomic, and combined clinical-radiomic models based on clinical data and routine non-contrast chest CT radiomic features for individualized prediction of concomitant pulmonary embolism (PE) in patients with chronic obstructive pulmonary disease (COPD). It should be noted that only internal hold-out validation was performed, and no independent external cohort validation was conducted. Therefore, the generalizability and cross-institutional stability of the models require further confirmation. In the internal test cohort, the combined model

demonstrated the best predictive performance, with an AUC of 0.987 (95% CI: 0.970–1.000), outperforming the clinical model alone (AUC=0.959) and the radiomic model alone (AUC=0.937). It also showed favorable calibration and higher theoretical net clinical benefit across a range of threshold probabilities as assessed by decision curve analysis. These findings suggest that clinical characteristics and quantitative CT imaging phenotypes may provide complementary information, and their combination may improve risk identification of PE in COPD patients. However, given the limited test sample size and number of PE events, and the fact that the models were only internally validated, the notably high AUC should be interpreted with caution. Its performance may be influenced by single-center case composition, random data splitting, and cohort homogeneity. Future evaluation of performance attenuation in independent external populations remains essential.

Previous studies on COPD and PE have focused primarily on PE prevalence and clinical risk factors. Yang et al. conducted a prospective cohort study in a plateau region and reported a relatively high PE prevalence in patients with acute exacerbation of COPD (AECOPD)[25], suggesting that high-altitude hypoxia may further increase thromboembolic risk. A retrospective study of elderly hospitalized COPD patients reported a PE prevalence of 12.63% and identified multiple independent risk factors, including prolonged bed rest, deep vein thrombosis (DVT), and hypoxemia[26]. In addition, a multicenter European study also indicated a non-negligible PE detection rate in the AECOPD population[27], and that clinical symptoms alone were insufficient for effective differentiation. Our findings are generally consistent with these observations, supporting the plausibility of the clinical association patterns identified in our cohort. However, notable differences in inclusion criteria, geographic altitude, and CTPA examination strategies across studies should be acknowledged. Therefore, PE prevalence rates and risk factors should not be directly compared across cohorts.

In terms of machine learning model comparison, although complex ensemble algorithms (Extra Trees, Random Forest, and XGBoost) achieved excellent performance in the training set with near-perfect AUCs, their discriminative ability declined substantially in the independent test set. This suggests that highly flexible nonlinear models are prone to overfitting cohort noise in high-dimensional imaging features. Given that the imaging feature dimension (1,834) substantially exceeded the sample size (415) in this study, the generalizability of conventional machine learning algorithms was clearly limited. Therefore, we ultimately employed LASSO-regularized logistic regression to construct the radiomic signature (Rad-score). Through sparse constraint, this approach eliminated redundant and low-signal-to-noise features, significantly improving model extrapolation stability. Of note, the LASSO-derived Rad-score in our study (test AUC=0.937) outperformed the best MLP model among eight machine learning algorithms (AUC=0.910). This difference does not reflect inherent algorithmic superiority, but rather differences in feature input and selection strategies between the two modeling approaches. MLP was built on all

candidate features retained after preliminary screening, which inevitably retained some noise, whereas LASSO performed a secondary precise sparse selection, retaining only 20 highly stable predictive features. This finding suggests that in high-dimensional radiomic studies with limited sample size, appropriate feature regularization and dimensionality reduction strategies contribute more to model robustness than simply increasing algorithmic complexity.

Current PE radiomic studies have primarily focused on prognostic prediction and risk stratification in patients with established PE, mostly based on radiomic features extracted from targeted thrombus regions within the pulmonary artery[28], with a specific focus on the thrombus itself and lacking attention to the underlying COPD population. Gotta et al. performed PE risk stratification based on dual-energy CT radiomics[29]. Currently, there is a lack of PE radiomic prediction models specifically targeting populations with underlying lung diseases such as COPD[30], suggesting that our study provides certain complementary value. Unlike previous studies, our approach did not rely on targeted segmentation of thrombi or pulmonary arteries, but rather completed feature extraction and modeling based on the entire lung parenchyma. The underlying biological mechanism may be that occult PE can lead to focal pulmonary microcirculatory disturbances, ventilation-perfusion mismatch, subtle interstitial edema, and altered lung texture heterogeneity. These microstructural changes can be captured by quantitative radiomics, thereby enabling indirect assessment of underlying thrombus burden. This modeling strategy does not require fine targeted annotation, is more compatible with routine clinical CT workflows, and offers advantages in automation and scalability. Importantly, our study used routine non-contrast CT rather than CTPA images for model construction. Compared with contrast-enhanced CTPA, non-contrast chest CT is more accessible in hospitalized COPD patients and does not require intravenous contrast agents. Therefore, it theoretically allows PE risk assessment using existing imaging information. Based on this characteristic, our findings propose a potentially clinically applicable scenario that requires further validation: using routine chest CT combined with clinical data to identify patients at high risk of PE, who may then be selected for confirmatory CTPA. If this strategy maintains adequate sensitivity, calibration, and clinical net benefit in independent prospective cohorts, it may in the future help optimize the use of CTPA resources. However, it must be emphasized that this study has not directly evaluated whether this strategy can reduce unnecessary CTPA examinations, lower missed diagnosis rates, shorten diagnostic time, or improve patient outcomes. Therefore, at present, this cannot be regarded as an established screening pathway. The intended role of this model is as a potential adjunctive risk stratification tool, and it cannot replace CTPA in the confirmatory diagnosis of PE.

Although the discriminative performance of the radiomic model alone was slightly inferior to that of the clinical model, the combined nomogram integrating both approaches significantly improved overall discriminative performance (test AUC=0.987). Clinical indicators primarily reflect macro-phenotypes such as

demographic characteristics, systemic inflammation, coagulation activation, oxygenation status, and clinical symptoms, whereas radiomics can quantify microscopic lung texture heterogeneity and structural disorganization that are not perceptible to the naked eye, thereby providing imaging-based predictive value independent of clinical information. DCA results demonstrated that the combined model provided higher net clinical benefit across a wide range of clinical threshold probabilities, supporting its potential clinical decision-making advantage. However, DCA reflects model utility under specific assumptions and threshold probabilities, and cannot directly demonstrate that the model can improve patient management or clinical outcomes in real-world practice.

Several limitations warrant consideration. First, the sample size and number of PE events in this study were relatively limited, while the initial radiomic feature dimension was high. Although feature selection, regularization, and cross-validation were employed to mitigate overfitting, it is not possible to completely exclude spurious associations and model optimism bias in high-dimensional modeling. In particular, the notably high AUC of the combined model in the internal test cohort requires replication in larger samples. Second, this was a single-center retrospective study with only internal hold-out validation. External validation across multiple centers, different scanner types, and varying acquisition parameters is lacking, and model generalizability remains to be rigorously tested. Third, radiomic features are susceptible to image reconstruction algorithms, slice thickness, and segmentation accuracy, and their cross-device and cross-center reproducibility still requires systematic evaluation. Fourth, some clinical variables in the retrospective data had missing values that were handled with multiple imputation, so residual information bias may still exist. Additionally, the retrospective inclusion method may introduce selection bias, and the study cohort may differ in disease spectrum from the unselected real-world COPD hospitalized population. Fifth, we have not systematically compared the combined model against existing standardized clinical prediction rules for PE or D-dimer-based diagnostic strategies regarding incremental value. Therefore, it remains uncertain how much additional useful information the radiomic model can provide beyond existing clinical diagnostic workflows. Finally, this study only validated the predictive accuracy and decision benefits of the models. We have not demonstrated that applying the model can reduce unnecessary CTPA examinations, lower PE missed diagnosis rates, alter treatment decisions, or improve patient outcomes. Future prospective interventional studies are needed to further validate its clinical value.

The occurrence of PE in COPD patients is closely associated with multidimensional clinical risk factors and quantitative CT imaging phenotypes of the lung. This study differs from traditional clinical risk factor analyses and existing PE prognostic radiomic studies, and provides an important supplement to the existing literature. However, given that the current evidence originates only from retrospective single-center data and internal validation, our findings should be considered as evidence from

the model development and preliminary validation stage, rather than final evidence supporting clinical implementation. Future large-sample, multicenter, cross-device independent external validation and prospective studies are needed to further clarify the stability, incremental value, and impact on real-world clinical decision-making and patient outcomes of the model.

5.Conclusions

In summary, PE is a non-negligible comorbidity in patients with AECOPD. Its clinical manifestations overlap substantially with those of AECOPD. This overlap increases the difficulty of early identification and diagnosis. In this study, we developed and validated a multimodal risk prediction model for PE in patients with COPD. The model integrated clinical data and CT radiomic features. Compared with the clinical-only and radiomics-only models, the combined model achieved an AUC of 0.987 in the test cohort. It also demonstrated favorable calibration and significant net benefit. These findings confirm that clinical and imaging information provide complementary value. The model may serve as an auxiliary tool to identify high-risk PE patients during routine CT examinations. However, the current results are based solely on single-center retrospective internal validation. Multicenter external validation is still needed to further evaluate its generalizability and clinical utility. This model may provide supportive reference for clinical identification of high-risk PE patients with COPD. However, it cannot replace CTPA as the definitive diagnostic modality.

6.Declarations

Ethics approval and consent to participate

This retrospective study was reviewed and approved by the Human Research Ethics Committee of the Fourth Affiliated Hospital, Zhejiang University School of Medicine (Approval No. K2026033). The requirement for obtaining informed consent from the participants was waived by the Ethics Committee due to the retrospective nature of the study. This study was conducted in accordance with the Declaration of Helsinki.

Funding

This research received no external funding.

Consent to Publish

Consent to Publish declaration: not applicable.

Conflict of Interest

All authors declare that they have no conflicts of interest.

Clinical trial number

Clinical trial number: not applicable.

7.References

[1] Burney PGJ, Patel J, Newson R, et al. Global and regional trends in COPD mortality, 1990-2010. Eur Respir J 2015;45:1239–47. https://doi.org/10.1183/09031936.00142414.

[2] Li M, Jiang Y, Xu Y, et al. The prevalence and risk factors of pulmonary embolism in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Thromb J 2025;23:42. https://doi.org/10.1186/s12959-025-00728-6.

[3] Akpinar EE, Hoşgün D, Akpinar S, et al. Incidence of pulmonary embolism during COPD exacerbation. J Bras Pneumol 2014;40:38–45. https://doi.org/10.1590/S1806-37132014000100006.

[4] Børvik T, Brækkan SK, Enga K, et al. COPD and risk of venous thromboembolism and mortality in a general population. Eur Respir J 2016;47:473–81. https://doi.org/10.1183/13993003.00402-2015.

[5] Poulsen SH, Noer I, Møller JE, et al. Clinical outcome of patients with suspected pulmonary embolism. A follow-up study of 588 consecutive patients. J Intern Med 2001;250:137–43. https://doi.org/10.1046/j.1365-2796.2001.00866.x.

[6] Liu X, Jiao X, Gong X, et al. Prevalence, Risk Factor and Clinical Characteristics of Venous Thrombus Embolism in Patients with Acute Exacerbation of COPD: A Prospective Multicenter Study. Int J Chron Obstruct Pulmon Dis 2023;18:907–17. https://doi.org/10.2147/COPD.S410954.

[7] de-Miguel-Diez J, Albaladejo-Vicente R, Jiménez-García R, et al. The effect of COPD on the incidence and mortality of hospitalized patients with pulmonary embolism: A nationwide population-based study (2016-2018). European Journal of Internal Medicine 2021;84:18–23. https://doi.org/10.1016/j.ejim.2020.11.009.

[8] Sato R, Hasegawa D, Nishida K, et al. Prevalence of pulmonary embolism in patients with acute exacerbations of COPD: A systematic review and meta-analysis. The American Journal of Emergency Medicine 2021;50:606–17. https://doi.org/10.1016/j.ajem.2021.09.041.

[9] Aleva FE, Voets LWLM, Simons SO, et al. Prevalence and Localization of Pulmonary Embolism in Unexplained Acute Exacerbations of COPD: A Systematic Review and Meta-analysis. CHEST 2017;151:544–54. https://doi.org/10.1016/j.chest.2016.07.034.

[10] Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. Thromb Haemost 2000;83:416–20.

[11] Le Gal G, Righini M, Roy P-M, et al. Prediction of pulmonary embolism in the emergency department: the revised Geneva score. Ann Intern Med 2006;144:165–71. https://doi.org/10.7326/0003-4819-144-3-200602070-00004.

[12] Patel P, Patel P, Bhatt M, et al. Systematic review and meta-analysis of test accuracy for the diagnosis of suspected pulmonary embolism. Blood Advances 2020;4:4296–311. https://doi.org/10.1182/bloodadvances.2019001052.

[13] Sadeghi S, Emami Ardestani M, Raofi E, et al. Diagnostic Value of D-dimer in Detecting Pulmonary Embolism in Patients with Acute COPD Exacerbation. Tanaffos 2020;19:371–9.

[14] Maritano Furcada J, Castro HM, De Vito EL, et al. Diagnosis of pulmonary embolism in patients

with acute exacerbations of chronic obstructive pulmonary disease: A cross-sectional study. Clin Respir J 2020;14:1176–81. https://doi.org/10.1111/crj.13257.

[15] Rambaud G, Mai V, Motreff C, et al. Pulmonary embolism diagnostic strategies in patients with COPD exacerbation: Post-hoc analysis of the PEP trial. Thrombosis Research 2023;231:58–64. https://doi.org/10.1016/j.thromres.2023.09.013.

[16] Sharpe A, Hartley R, Antunes G. Characteristics and outcomes of patients who undergo multiple computer tomographic pulmonary angiograms (CTPAs) for suspected pulmonary embolism (PE). European Respiratory Journal 2015;46. https://doi.org/10.1183/13993003.congress-2015.PA3745.

[17] Zhou Q, Huang R, Xiong X, et al. Prediction of pulmonary embolism by an explainable machine learning approach in the real world. Sci Rep 2025;15:835. https://doi.org/10.1038/s41598-024-75435-9.

[18] Cao Z, Yang L, Han J, et al. Development of a predictive nomogram for early identification of pulmonary embolism in hospitalized patients: a retrospective cohort study. BMC Pulmonary Medicine 2024;24:594. https://doi.org/10.1186/s12890-024-03377-z.

[19] Li H, Ning Q, Liu Y, et al. Construction and verification of risk prediction model of pulmonary embolism in ICU patients with COPD in acute exacerbation based on age, SAPSII score, braking state, and mechanical ventilation. Front Med (Lausanne) 2025;12:1564220. https://doi.org/10.3389/fmed.2025.1564220.

[20] Cahan N, Klang E, Marom EM, et al. Multimodal fusion models for pulmonary embolism mortality prediction. Sci Rep 2023;13:7544. https://doi.org/10.1038/s41598-023-34303-8.

[21] Wang J, Wan Z, Liu Q, et al. Predictive Value of Red Blood Cell Distribution Width in Chronic Obstructive Pulmonary Disease Patients with Pulmonary Embolism. Anal Cell Pathol (Amst) 2020;2020:1935742. https://doi.org/10.1155/2020/1935742.

[22] Zhao X, Li J, Wang Y, et al. The predictive value of D-dimer combined with systemic immune-inflammation index for the presence of pulmonary thromboembolism in AECOPD patients. Front Med 2025;12. https://doi.org/10.3389/fmed.2025.1582913.

[23] Song J, Yu Y, Wang M, et al. Prediction of risk stratification in acute pulmonary embolism using a combined CTPA radiomics and machine learning model. Medicine (Baltimore) 2026;105:e49227. https://doi.org/10.1097/MD.0000000000049227.

[24] Song J, Chen R, Yu Y, et al. Predictive value of a combined clinical, serum biomarker, and radiomic model for poor short-term prognosis in acute pulmonary embolism. Int J Cardiovasc Imaging 2026;42:579–90. https://doi.org/10.1007/s10554-026-03615-x.

[25] Yang C, Tuo Y, Shi X, et al. Prevalence, risk factors, and clinical characteristics of pulmonary embolism in patients with acute exacerbation of COPD in Plateau regions: a prospective cohort study. BMC Pulm Med 2024;24:102. https://doi.org/10.1186/s12890-024-02915-z.

[26] Jin C, Hu Y, Liu F. Clinical characteristics, clinical laboratories, and risk factors of elderly patients with chronic obstructive pulmonary disease complicated with pulmonary embolism. Front Med (Lausanne) 2025;12:1421169. https://doi.org/10.3389/fmed.2025.1421169.

[27] Dentali F, Pomero F, Micco PD, et al. Prevalence and risk factors for pulmonary embolism in patients with suspected acute exacerbation of COPD: a multi-center study. European Journal of Internal Medicine 2020;80:54–9. https://doi.org/10.1016/j.ejim.2020.05.006.

[28] Ehrhardt L, Fiedler P, Surov A, et al. Prognostic potential of radiomics evaluation of lung artery thrombus for pulmonary embolism patients. Int J Comput Assist Radiol Surg 2026;21:307–18.

https://doi.org/10.1007/s11548-025-03530-x.
[29] Gotta J, Koch V, Geyer T, et al. Imaging-based risk stratification of patients with pulmonary embolism based on dual-energy CT-derived radiomics. Eur J Clin Invest 2024;54:e14139. https://doi.org/10.1111/eci.14139.
[30] Casado-Suela MÁ, Torres-Macho J, Prada-Alonso J, et al. Use of Radiomics to Predict Adverse Outcomes in Patients with Pulmonary Embolism: A Scoping Review of an Unresolved Clinical Challenge. Diagnostics 2025;15:2022. https://doi.org/10.3390/diagnostics15162022.