



OPEN ACCESS

EDITED BY

Chen Shanxiang,
Southwest University, China

REVIEWED BY

Chuanming Li,
Chongqing University, China
Akanksha Soni,
University Institute of Technology RGPV
Shivpuri, India

*CORRESPONDENCE

Qian Li
✉ 2430151@cqmpc.edu.cn
Li Zhang
✉ 2430150@cqmpc.edu.cn
Wutao Xie
✉ 2430153@cqmpc.edu.cn

†These authors have contributed equally to
this work and share first authorship

RECEIVED 29 July 2025

REVISED 27 October 2025

ACCEPTED 05 November 2025

PUBLISHED 03 December 2025

CITATION

Li X, Wang W, Li X, Liu Q, Liu Y, Wang L, Li Q,
Zhang L and Xie W (2025) Differential
diagnosis of pneumoconiosis mass shadows
and peripheral lung cancer using CT
radiomics and the AdaBoost machine
learning model.
Front. Med. 12:1675840.
doi: 10.3389/fmed.2025.1675840

COPYRIGHT

© 2025 Li, Wang, Li, Liu, Liu, Wang, Li, Zhang
and Xie. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Differential diagnosis of pneumoconiosis mass shadows and peripheral lung cancer using CT radiomics and the AdaBoost machine learning model

Xiaobing Li^{1,2,3,4,5†}, Wei Wang^{6†}, Xuemei Li^{7†}, Qianqian Liu^{1,2,3,4},
Yongsheng Liu^{1,2,3,4}, Li Wang^{1,2,3,4}, Qian Li^{1,2,3,4*}, Li Zhang^{1,2,3,4*}
and Wutao Xie^{6*}

¹Science and Technology Industry Development Center, Chongqing Medical and Pharmaceutical College, Chongqing, China, ²Laboratory of Toxicology, The First Affiliated Hospital of Chongqing Medical and Pharmaceutical College, Chongqing, China, ³Chongqing Key Laboratory of Prevention and Treatment for Occupational Diseases and Poisoning, The First Affiliated Hospital of Chongqing Medical and Pharmaceutical College, Chongqing, China, ⁴Department of Occupational Disease and Poisoning Medicine, The First Affiliated Hospital of Chongqing Medical and Pharmaceutical College, Chongqing, China, ⁵College of Public Health, Chongqing Medical University, Chongqing, China, ⁶Department of Radiology, The First Affiliated Hospital of Chongqing Medical and Pharmaceutical College, Chongqing, China, ⁷NHC Key Laboratory of Diagnosis and Treatment on Brain Functional Diseases, Department of Neurology, First Affiliated Hospital of Chongqing Medical University, Chongqing, China

Objective: To develop a differential diagnostic prediction model for distinguishing large opacities in pneumoconiosis from peripheral lung cancer based on CT radiomics.

Methods: A total of 103 cases of large opacities in pneumoconiosis and 85 cases of peripheral lung cancer were retrospectively collected from routine CT scans at the First Affiliated Hospital of Chongqing Medical and Pharmaceutical College between March 2021 and June 2025. Diagnosis was confirmed by an expert panel, clinical evaluations, and pathological examinations. Patients were randomly assigned to a training set ($n = 132$) and a test set ($n = 56$). Lesions were delineated by at least two pneumoconiosis experts using ITK-SNAP software. Radiomic features were extracted from CT images of lung lesions in the training set, including first-order features, shape features (2D and 3D), texture features (gray-level co-occurrence matrix, gray-level run-length matrix, gray-level size-zone matrix, gray-level dependence matrix), and wavelet transform filters. Feature dimensionality reduction was applied to construct morphological biomarkers. Diagnostic prediction models were built using machine learning algorithms. Model performance was evaluated using the ROC curve and the area under the curve (AUC) in the test set.

Results: A total of 108 features were extracted from 110 large opacity regions and 85 peripheral lung cancer regions of interest (ROIs). Dimensionality reduction identified a subset of eight most significant features. LR, SVM, and AdaBoost algorithms were implemented using Python to build the models. In the training set, the accuracies of the LR, SVM, and AdaBoost models were 79.4, 84.0, and 80.9%, respectively; the sensitivities were 74.1, 74.1, and 81.0%, respectively; the specificities were 83.6, 91.8, and 80.8%, respectively; and the AUC values

were 0.837, 0.886, and 0.900, respectively. In the test set, the accuracies of the LR, SVM, and AdaBoost models were 80.7, 82.5, and 86.0%, respectively; the sensitivities were 89.3, 89.3, and 82.1%, respectively; the specificities were 72.4, 75.9, and 89.7%, respectively; and the AUC values were 0.825, 0.855, and 0.900, respectively. The AUC of the AdaBoost ROC curve was significantly superior to those of the LR and SVM models. The AdaBoost model demonstrated the optimal predictive performance in both the training and test sets.

Conclusion: The AdaBoost-based prediction model, developed using CT radiomic features, effectively differentiates large opacities of stage III occupational pneumoconiosis from peripheral lung cancer.

KEYWORDS

pneumoconiosis, large opacities, CT radiomics, AdaBoost, machine learning, diagnostic model

1 Introduction

Pneumoconiosis is a progressive lung disease caused by the long-term inhalation of harmful mineral dusts, typically encountered in occupational settings such as mining, construction, and manufacturing (1). The dust particles become lodged in the pulmonary tissue, leading to inflammation, fibrosis, and the eventual development of severe pulmonary dysfunction (2). The disease primarily manifests as diffuse pulmonary fibrosis, which can progress to more severe forms, including progressive massive fibrosis (PMF) in advanced stages (3). PMF is characterized by large opacities in the lungs, often exceeding 10 mm in diameter, as defined by the National Occupational Health Standard of the People's Republic of China (4). In stage III pneumoconiosis, these large opacities present with a long diameter of at least 20 mm and a short diameter greater than 10 mm. Clinically, patients with advanced pneumoconiosis, particularly those with PMF, experience progressive respiratory symptoms such as dyspnea, chronic cough, and a significant decline in lung function (5). These clinical manifestations are not only debilitating but also share a striking resemblance to those seen in patients with peripheral lung cancer, which complicates the differentiation between the two diseases (6).

Lung cancer, especially in its peripheral form, is one of the most common and lethal malignancies worldwide. Peripheral lung cancer typically originates from the distal bronchioles or alveolar epithelium and often presents as solitary pulmonary nodules or masses located near the pleura (7). In its early stages, the disease may be asymptomatic or manifest with non-specific respiratory symptoms such as cough or chest pain (8). As the tumor progresses, features such as pleural indentation, vascular convergence, spiculation, and local invasion become more apparent. These radiologic characteristics, however, can closely mimic those of PMF, particularly when large opacities are present (9). In addition, patients with pneumoconiosis who are at risk due to occupational exposure may concurrently develop lung cancer, further complicating diagnosis (9). Unlike PMF, which generally follows a chronic, fibrotic course, peripheral lung cancer tends to grow rapidly and metastasize early, leading to poor prognosis if not identified and treated in a timely manner (10).

The overlapping clinical and radiologic features between pneumoconiosis and lung cancer present significant diagnostic challenges. Imaging studies, especially chest X-rays and CT scans, often reveal large opacities that can mimic lung cancer masses in terms of their size, shape, and density (11). As a result, distinguishing between the two conditions solely based on traditional imaging techniques becomes increasingly difficult (10). Therefore, there is an urgent need for accurate diagnostic tools that can help clinicians differentiate between these two conditions effectively (12).

Radiomics, a rapidly evolving field in medical imaging, offers significant promise in addressing this diagnostic dilemma. Radiomics involves the extraction of a wide range of quantitative features from medical images, such as CT scans, to capture the complex spatial and textural characteristics of tissue. These features go beyond traditional imaging analysis and provide detailed insights into the underlying pathological changes in the tissue (13). Lambin et al. first proposed the concept of radiomics in 2012, highlighting its potential to extract clinically relevant information from imaging data that may not be immediately apparent to the human eye (14). By applying advanced machine learning (ML) algorithms to radiomic features, it is possible to develop predictive models that can aid in non-invasive disease diagnosis, prognosis, and treatment planning (15).

In the context of pneumoconiosis and lung cancer, radiomics offers substantial diagnostic value owing to the remarkable similarity in imaging characteristics between progressive massive fibrosis (PMF) and peripheral lung tumors (16). The primary objective of this study was to develop a differential diagnostic prediction model based on radiomic features extracted from CT images of pulmonary lesions. By identifying distinct quantitative patterns that differentiate pneumoconiosis-related large opacities from malignant nodules, we aimed to enhance diagnostic precision and assist clinical decision-making (17).

Leveraging the capacity of machine learning algorithms to analyze high-dimensional data and capture complex, non-linear relationships, our approach provides a robust and scalable framework for improving diagnostic discrimination between these two clinically overlapping conditions (18). This is particularly

important given that misdiagnosis can lead to inappropriate management, delayed treatment, or unnecessary invasive procedures (19). Overall, this study addresses a critical unmet need in the field of occupational lung disease and oncologic imaging. By integrating radiomics with advanced computational modeling, our work contributes to the development of a standardized, non-invasive, and reproducible diagnostic tool capable of supporting accurate differentiation between pneumoconiosis and peripheral lung cancer, ultimately improving patient outcomes and guiding more informed clinical practice.

2 Materials and methods

2.1 General data

Chest CT images of patients with pneumoconiosis-related large opacities and peripheral lung cancer were retrospectively collected from the First Affiliated Hospital of Chongqing Medical and Pharmaceutical College between March 2021 and June 2025. The study cohort was divided into two groups based on clinical and diagnostic criteria: Pneumoconiosis large opacity group: Patients diagnosed with stage III occupational pneumoconiosis presenting with large opacities, as confirmed by an expert panel in accordance with the GBZ70–2015 Diagnostic Criteria for Pneumoconiosis. Lesions were required to have a long-axis diameter ≥ 20 mm and a short-axis diameter > 10 mm. Peripheral lung cancer group: patients with peripheral lung cancer confirmed by histopathological examination, exhibiting mass lesions with a maximum diameter ≥ 30 mm.

Inclusion criteria were as follows: (1) For the pneumoconiosis group: confirmed diagnosis of stage III pneumoconiosis with large opacities meeting the size thresholds (long-axis ≥ 20 mm, short-axis > 10 mm) (Figure 1; Supplementary Figure S1). (2) For the lung cancer group: confirmed diagnosis of peripheral lung cancer by both imaging and pathology, with lesions ≥ 30 mm in diameter (Figure 2; Supplementary Figure S2).

Exclusion criteria were defined to ensure image quality and diagnostic specificity: (1) For the pneumoconiosis group: cases with suboptimal CT image quality that failed to meet diagnostic imaging standards, or cases with coexisting pulmonary tuberculosis, metastatic lesions, or other respiratory comorbidities that could confound radiomic analysis. (2) For the lung cancer group: patients with concurrent tuberculosis, pulmonary metastases, or other comorbidities potentially interfering with imaging interpretation.

2.2 Methods

2.2.1 CT examination

All subjects underwent CT scans using one of the following devices: GE 128-slice 256-layer Revolution ES Spiral CT (United States), GE Optima CT680, or United Imaging uMI Panorama 860 160-slice CT. Scans were performed from the bilateral lung apices to the lung bases for all patients. The scanning procedure was as follows: ① verify patient ID; ② prepare the patient; ③ position the patient supine; ④ instruct the patient

to hold their breath during the scan; ⑤ perform the scan; ⑥ end the scan.

Scanning parameters: ① tube voltage: 120 kV; ② rotation time: 0.5 s/rotation; ③ pitch: 0.991:1; ④ tube current: Auto mA; ⑤ scan slice thickness and interval: 5 mm ⑥ reconstruction slice thickness: 0.625 mm; ⑦ reconstruction interval: 0.625 mm. Images were transmitted to the GE ADW4.7 workstation for further processing and analysis.

2.2.2 CT image processing and analysis

2.2.2.1 ROI image segmentation

CT images were imported into ITK-SNAP software¹ for image segmentation. Two radiologists with over 5 years of experience, both qualified for pneumoconiosis diagnosis, manually delineated the ROIs for each case. For large shadow lesions, the long diameter was ≥ 20 mm and the short diameter > 10 mm; for peripheral lung cancer lesions, the size was ≥ 3 cm. Each slice was carefully outlined to capture the entire lesion. Completed ROIs were saved in the nii.gz format in the “images” and “mask” directories. A total of 110 ROIs were outlined from the CT images of 103 pneumoconiosis patients, and 85 ROIs were outlined from 85 peripheral lung cancer patients.

To assess the reproducibility of ROI delineation, the inter-observer consistency of the radiomics features was evaluated using the Intraclass Correlation Coefficient (ICC). An ICC value > 0.75 indicated good consistency between the two radiologists (20).

2.2.2.2 Feature selection

Radiomic features were extracted using the Pyradiomics software.² The following feature extraction methods were applied: ① Neighborhood Gray Tone Difference Matrix (NGTDM); ② Shape features; ③ First-order features; ④ Gray-Level Co-occurrence Matrix (GLCM); ⑤ Gray-Level Dependence Matrix (GLDM); ⑥ Gray-Level Run-Length Matrix (GLRLM); ⑦ Gray-Level Size-Zone Matrix (GLSZM). Features for large shadows and peripheral lung cancer were labeled as 0 and 1, respectively.

2.2.2.3 Data processing

To minimize the impact of feature dimensionality and improve model performance, the data were standardized using Z-scores for normalization.

2.2.2.4 Feature screening

For normally distributed features, a t-test was applied, while for non-normally distributed features, a U-test was used ($p < 0.05$ was considered statistically significant). Following screening, 85 features were retained for further analysis.

2.2.2.5 Feature dimensionality reduction

To reduce the dimensionality of the initially extracted 85 radiomic features, Pearson's correlation analysis was first performed to assess inter-feature relationships. Features exhibiting high pairwise correlation ($|r| > 0.9$) were removed to minimize multicollinearity and redundancy. Subsequently, Least Absolute Shrinkage and Selection Operator (LASSO) regression was

¹ <http://www.itksnap.org/>

² <http://pyradiomics.readthedocs.io>

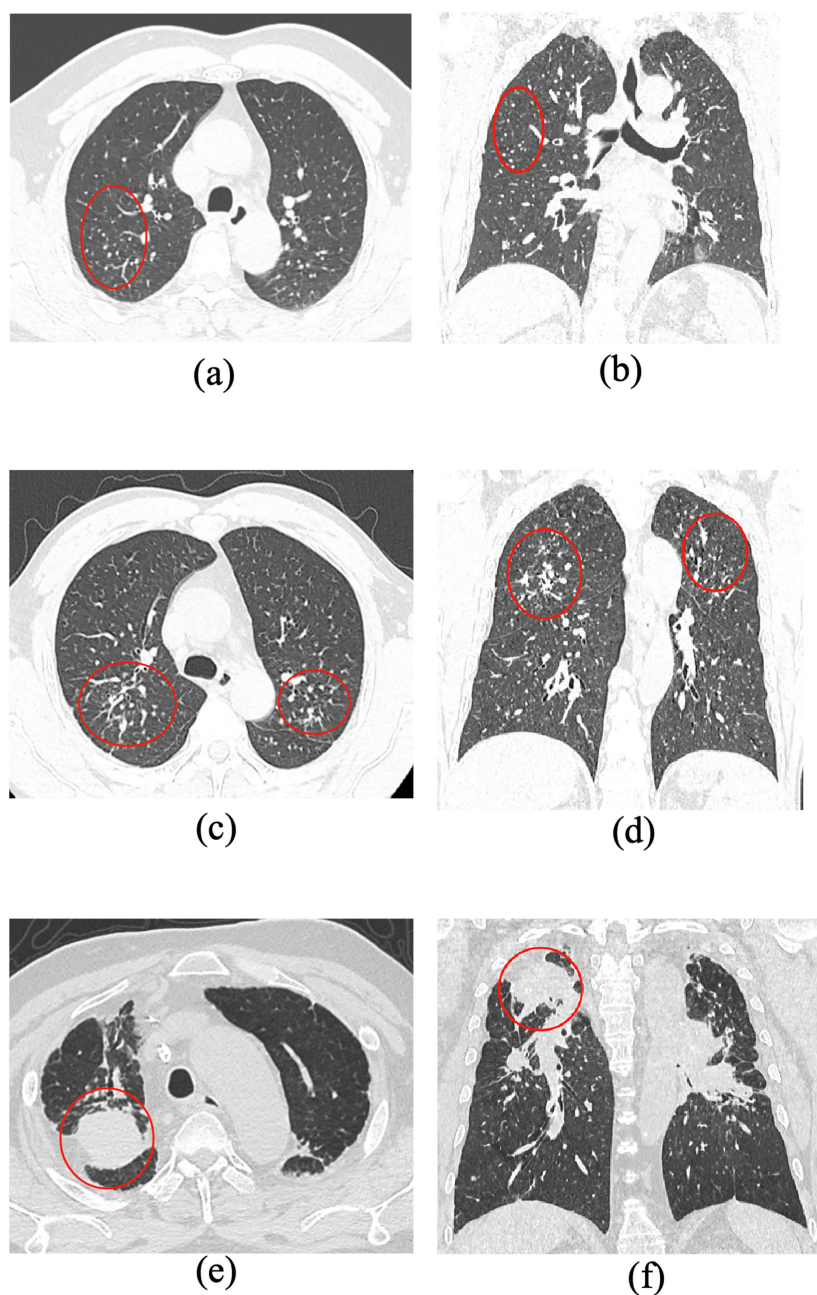


FIGURE 1

Representative CT imaging features across different stages of pneumoconiosis (axial and coronal views). **(a,b)** Axial and coronal chest CT images from a patient with Stage I pneumoconiosis, revealing multiple small, round, high-density nodules predominantly distributed in the upper lobes of both lungs, especially in the right upper and middle fields. **(c,d)** Correspond to Stage II pneumoconiosis, characterized by an increased number of opacities involving both upper lung zones and the dorsal segment of the lower lobes. **(e,f)** Illustrate stage III pneumoconiosis, showing a homogeneous mass-like opacity in the apicoposterior segment of the right upper lobe, accompanied by scattered calcifications, pleural thickening, adjacent localized emphysema, and nodular interstitial markings. Notably, large opacities (long-axis diameter > 20 mm, short-axis diameter > 10 mm) in advanced pneumoconiosis tend to display asymmetric distribution, suggesting progressive fibrotic remodeling of lung parenchyma.

employed to perform penalized feature selection and identify the most informative predictors. The optimal regularization parameter (λ) was determined using 10-fold cross-validation, where the value minimizing the mean binomial deviance was selected to balance model simplicity and predictive performance. The resulting optimal λ value ($\lambda = 0.0222$) achieved the lowest deviance and ensured a parsimonious and stable model (Table 1).

As shown in Figure 3a, the X-axis represents the penalty coefficient (λ) and the Y-axis denotes the corresponding mean binomial deviance. The vertical dashed line marks the optimal λ value ($\lambda = 0.0222$), indicating the point of minimal deviance. Figure 3b displays the trajectories of regression coefficients for all features as λ increases, illustrating how stronger penalization gradually shrinks less relevant coefficients toward zero. Figure 3c

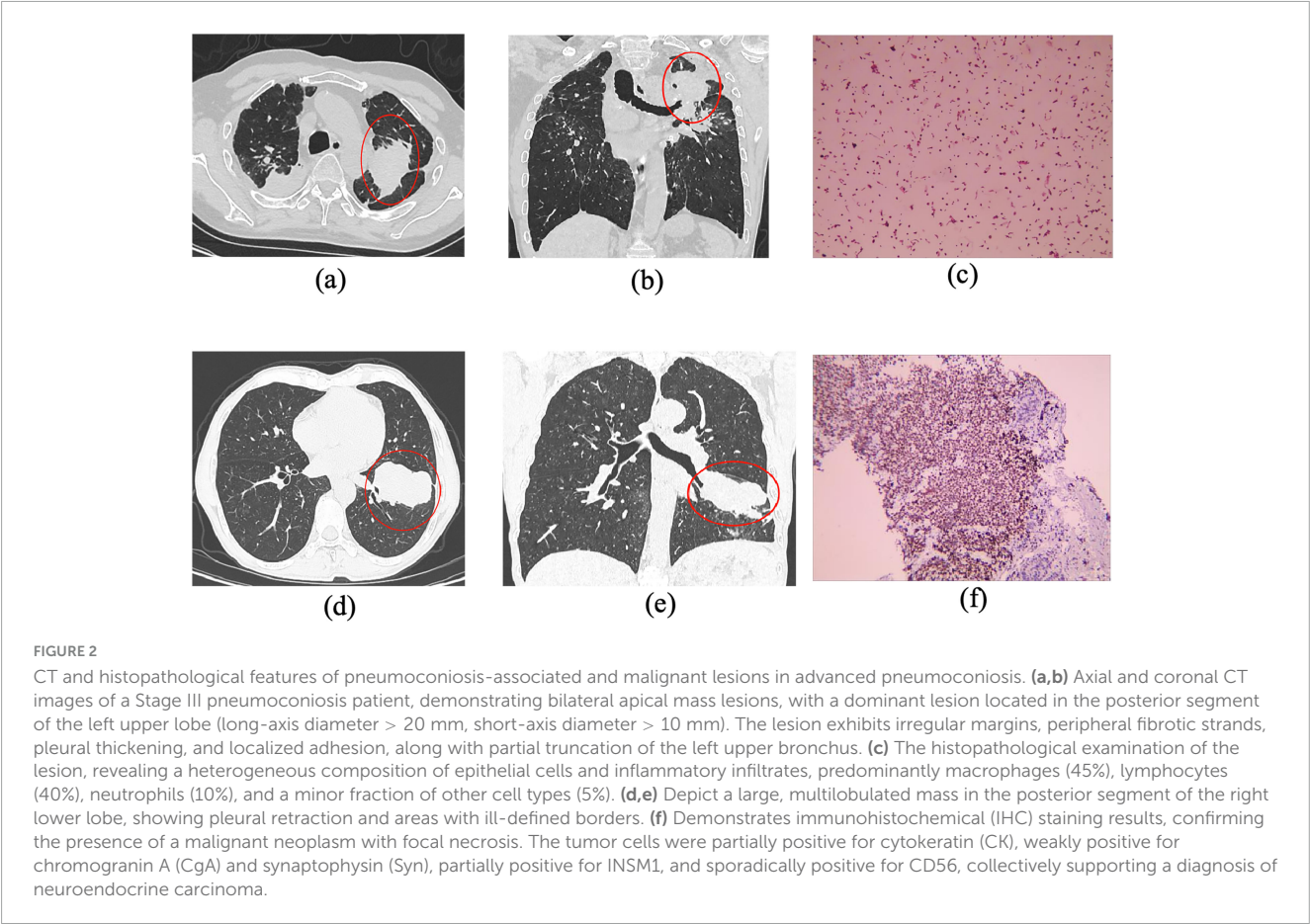


TABLE 1 Selected radiomic features and corresponding regression coefficients after dimensionality reduction using the LASSO regression model.

Type	Feature name
Selected gray-level non-uniformity texture features (Ngt dm)	original_ngt dm_Busyness, original_ngt dm_Complexity, original_ngt dm_Contrast
Shape features (shape)	original_shape_Maximum3DDiameter, original_shape_Sphericity
Gray-level co-occurrence matrix (Glc m)	original_glc m_ClusterShade, original_glc m_MCC
Gray-level size-zone matrix (Glszm)	original_glszm_SmallAreaLowGrayLevelEmphasis

summarizes the final eight selected radiomic features and their standardized regression coefficients, where positive coefficients (bars extending to the right) and negative coefficients (bars extending to the left) indicate their respective contributions to the predictive model.

2.3 Statistical processing

Data analysis was performed using the onekey V5.1.25 platform for imaging feature extraction. The Kolmogorov-Smirnov test was

used to assess the normality of measurement data. Data following a normal distribution were expressed as means $\pm$ standard deviation ($x \pm s$), while non-normally distributed data were reported as medians. The independent *t*-test or Mann-Whitney U test was used for group comparisons, as appropriate. Categorical data were expressed as percentages, and the chi-square test (χ^2 test) was employed for inter-group comparisons. A *P* < 0.05 was considered statistically significant. The MedCalc version 20.101 software was used to evaluate the efficacy of the diagnostic model by calculating the area under the Receiver Operating Characteristic (ROC) curve (AUC).

3 Results

3.1 General patient data

A total of 188 patients were enrolled in this study, consisting of 103 patients in the pneumoconiosis large shadow group (all male, with a mean age of 58.37 ± 7.34 years) and 85 patients with peripheral lung cancer (male-to-female ratio approximately 2.6:1, with a mean age of 70.95 ± 10.82 years). Among the 103 pneumoconiosis patients with large shadows, the pathological diagnoses included 32 cases of lung adenocarcinoma, 15 cases of squamous cell carcinoma, 1 case of neuroendocrine carcinoma, 1 case of adenosquamous carcinoma, 6 cases of small cell lung cancer, 2 cases of non-small cell lung cancer, and 28 cases where the

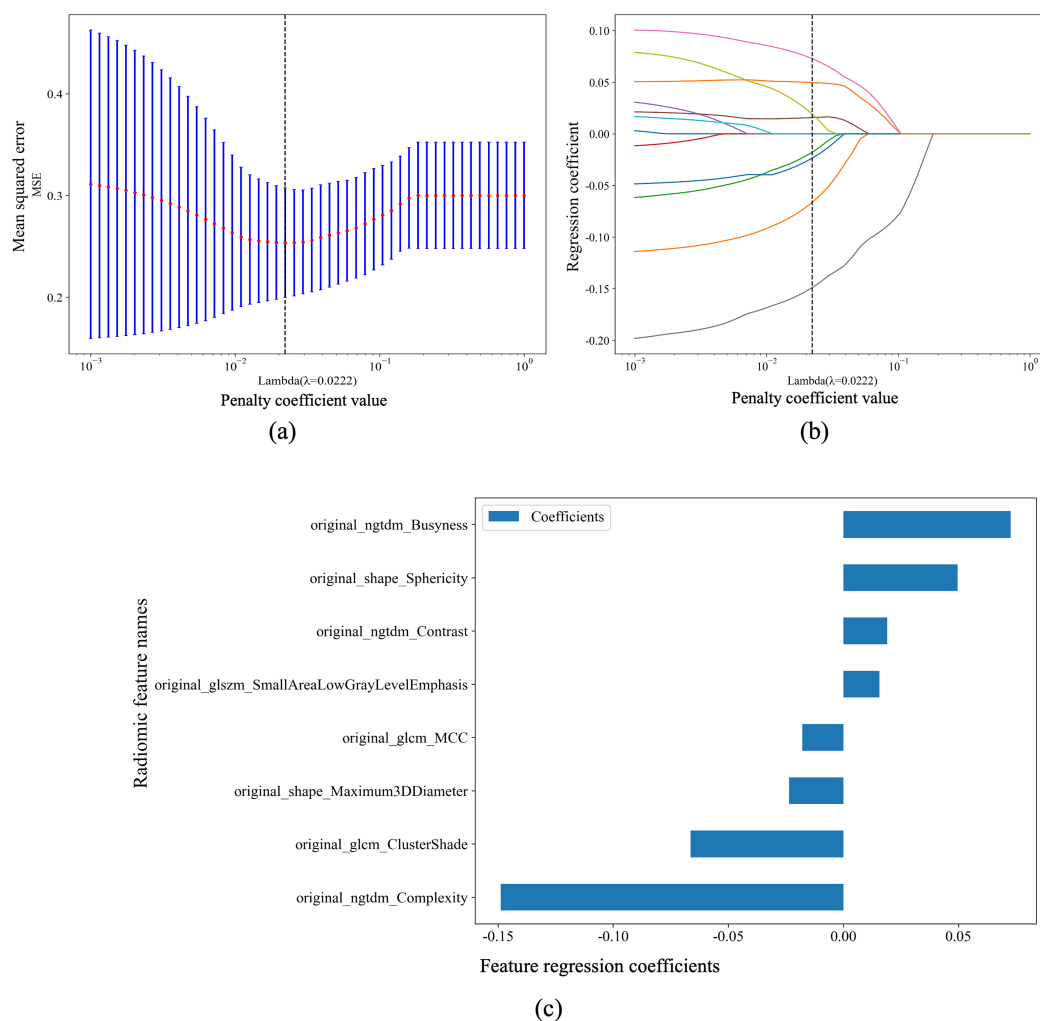


FIGURE 3

Penalized feature selection and regression coefficients for pneumoconiosis large opacities and peripheral lung cancer classification. **(a)** The relationship between the penalty coefficient value (λ) and the binomial deviance (Y-axis), with the X-axis showing the corresponding number of features. The vertical dashed line indicates the optimal λ value ($\lambda = 0.0222$). **(b)** The regression coefficients for each feature plotted against the penalty coefficient value (λ), illustrating the shrinkage effect of the penalty term and the evolution of feature coefficients with increasing λ . **(c)** A horizontal bar chart displaying the selected features with their corresponding regression coefficients. The Y-axis shows the radiomic feature names, and the X-axis represents the magnitude and direction of the coefficients, where bars to the right indicate positive coefficients and those to the left indicate negative coefficients.

pathological type remained unknown. These patients had all been diagnosed with stage III pneumoconiosis, and the large shadows were confirmed as massive fibrosis by expert clinical evaluation and imaging analysis. For the peripheral lung cancer group, all cases were pathologically confirmed after biopsy and surgery. The study population exhibited a range of underlying conditions, with some patients presenting with comorbidities such as chronic obstructive pulmonary disease (COPD) or cardiovascular disease, which were considered during the analysis to prevent bias in the feature extraction process.

3.2 Construction of classifiers

The 188-patient dataset was randomly divided into a training set and a validation set at a ratio of 7:3, resulting in 132 patients for the training set and 56 for the validation set. Several

common machine learning algorithms were employed to construct predictive models, using the onekey AI V5.1.25 platform: (1) Logistic Regression (LR): a generalized linear model that maps the linear regression output to a probability space through a logistic function, which is used for classification purposes. LR is widely used for its interpretability and simplicity, making it a strong baseline method for binary classification tasks. (2) Support Vector Machine (SVM): a supervised learning algorithm that performs classification by finding the optimal hyperplane that maximizes the margin between different classes. SVM is effective for classifying complex, non-linear data and is particularly useful for high-dimensional feature spaces, such as the radiomic features extracted from CT images. (3) Adaptive Boosting (AdaBoost): an ensemble learning method that combines multiple weak classifiers to form a strong classifier. By giving higher weights to misclassified samples in each iteration, AdaBoost improves the overall model accuracy. Notably,

AdaBoost is known for its ability to reduce overfitting and enhance the generalization capability of the model, even when training errors approach zero.

3.3 Model performance results

The performance of each model was evaluated based on various metrics, including accuracy, AUC (Area Under the Curve), sensitivity, specificity, precision, recall, F1 score, and predictive values. The detailed results for each model are as follows:

LR Model: ① accuracy: 79.4%; ② AUC: 0.837 (95% CI: 0.7687–0.9059); ③ sensitivity: 74.1%; ④ specificity: 83.6%; ⑤ positive Predictive Value (PPV): 78.2%; ⑥ precision: 78.2%; ⑦ Negative Predictive Value (NPV): 80.3%; Recall: 74.1%; ⑧ F1 Score: 76.1%.

SVM Model: ① accuracy: 84.0%; ② AUC: 0.886 (95% CI: 0.8277–0.9437); ③ sensitivity: 74.1%; ④ specificity: 91.8%; ⑤ PPV: 87.8%; ⑥ precision: 87.8%; ⑦ NPV: 81.7%; ⑧ recall: 74.1%; ⑨ F1 Score: 80.4%.

AdaBoost Model: ① accuracy: 80.9%; ② AUC: 0.900 (95% CI: 0.8498–0.9497); ③ sensitivity: 81.0%; ④ specificity: 80.8%; ⑤ PPV: 77.0%; ⑥ precision: 77.0%; ⑦ NPV: 84.3%; ⑧ recall: 81.0%; ⑨ F1 Score: 79.0% (Table 2).

The AdaBoost model achieved the highest performance across all metrics, with an accuracy of 80.9% and an AUC of 0.900, suggesting its superior ability to distinguish between pneumoconiosis large opacities and peripheral lung cancer. The SVM model also demonstrated robust performance with an accuracy of 84.0% and an AUC of 0.886, while the LR model provided a solid baseline, achieving an accuracy of 79.4% and an AUC of 0.837.

3.4 Net benefit and clinical relevance

Further evaluation of the models’ clinical relevance was performed by calculating the net benefit within different probability threshold ranges. In the low threshold probability range (0–0.4), the AdaBoost model showed a significantly higher net benefit compared to the “treat all” strategy, indicating its capacity to accurately identify patients who would benefit from early intervention. In the moderate threshold probability range (0.4–0.6), the net benefit of the AdaBoost model remained higher than both the “treat all” and “treat none” strategies, suggesting its substantial clinical predictive value.

Figures 4, 5 illustrate the detailed net benefit curves for each model across different probability thresholds, providing a

comprehensive view of their decision-making potential. These figures demonstrate the AdaBoost model’s superior performance in maximizing clinical benefit while minimizing unnecessary treatments.

4 Discussion

Radiomics enables the quantitative extraction and analysis of high-dimensional features from medical imaging data, providing a non-invasive, reproducible, and multidimensional assessment of disease characteristics (21). By integrating multimodal and multiparametric information, radiomics facilitates early diagnosis, individualized treatment planning, and disease monitoring, thus serving as a cornerstone of precision medicine (22).

In the context of pneumoconiosis and lung cancer, the diagnostic challenge arises from the striking similarity in their radiologic manifestations, particularly in Stage III occupational pneumoconiosis with progressive massive fibrosis (PMF) (23). Advanced fibrotic lesions in PMF often exhibit imaging characteristics, e.g., abnormal CT attenuation, calcification, satellite nodules, spiculation, pleural thickening, and cavitation—that closely mimic those of peripheral lung malignancies (24). The situation becomes even more complex when pneumoconiosis coexists with lung cancer, which may lead to misinterpretation and delayed diagnosis (25). Epidemiological data suggest that approximately 3.2% of patients with simple pneumoconiosis progress to PMF over an 8-year follow-up, and these patients have higher mortality rates than those with uncomplicated pneumoconiosis (26). Therefore, accurate differentiation between PMF-related large opacities and lung cancer is crucial for guiding clinical management and preventing inappropriate or delayed interventions (27).

Previous studies have explored the use of supplementary imaging modalities, such as MRI, including T2-weighted and diffusion-weighted imaging, to distinguish PMF from malignancies (28). However, MRI is often impractical in patients with advanced pneumoconiosis, who typically exhibit severe respiratory impairment (29). This limitation underscores the urgent need for non-invasive, efficient, and clinically applicable diagnostic tools (30). In this regard, artificial intelligence (AI)-based radiomics and machine learning approaches have shown significant promise in decoding complex imaging patterns that exceed human perceptual capability (31). Prior investigations, such as those by Warkentin et al. and Dong et al., demonstrated the potential of ML algorithms in predicting lung nodule malignancy and assessing pneumoconiosis risk, respectively

TABLE 2 Performance metrics of LR, SVM, and AdaBoost models in the training and test sets.

Model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	F1
LR (train)	0.794	0.837	0.7687–0.9059	0.741	0.836	0.782	0.803	0.761
LR (test)	0.807	0.825	0.7121–0.9381	0.893	0.724	0.758	0.875	0.82
SVM (train)	0.84	0.886	0.8277–0.9437	0.741	0.918	0.878	0.817	0.804
SVM (test)	0.825	0.855	0.7547–0.9547	0.893	0.759	0.781	0.88	0.833
AdaBoost (train)	0.809	0.900	0.8498–0.9497	0.810	0.808	0.770	0.843	0.790
AdaBoost (test)	0.860	0.869	0.7634–0.9755	0.821	0.897	0.885	0.839	0.852

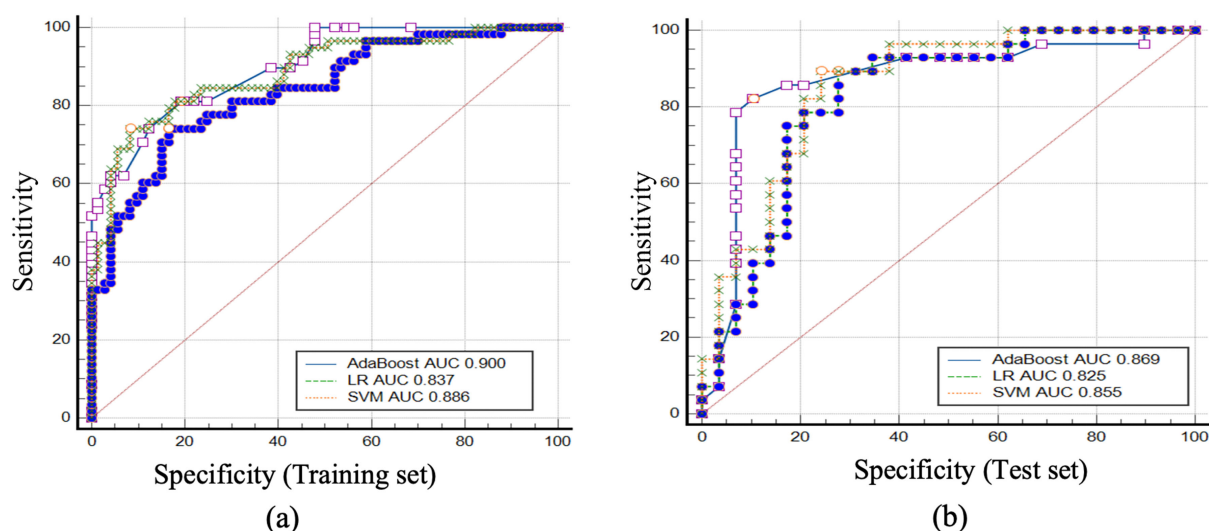


FIGURE 4

ROC curves and AUC values for the AdaBoost, LR, and SVM models in predicting pneumoconiosis large opacities and peripheral lung cancer in the training and test sets. (a) the ROC curves and corresponding AUC values for the AdaBoost, LR, and SVM models in the training set, with the AUC values for AdaBoost, LR, and SVM being 0.900, 0.837, and 0.886, respectively. (b) the ROC curves and AUC values for the models in the test set, with the AUC values for AdaBoost, LR, and SVM being 0.869, 0.825, and 0.855, respectively. These results demonstrate the performance of each model in differentiating between pneumoconiosis large opacities and peripheral lung cancer, highlighting AdaBoost as the top performer in both the training and test sets.

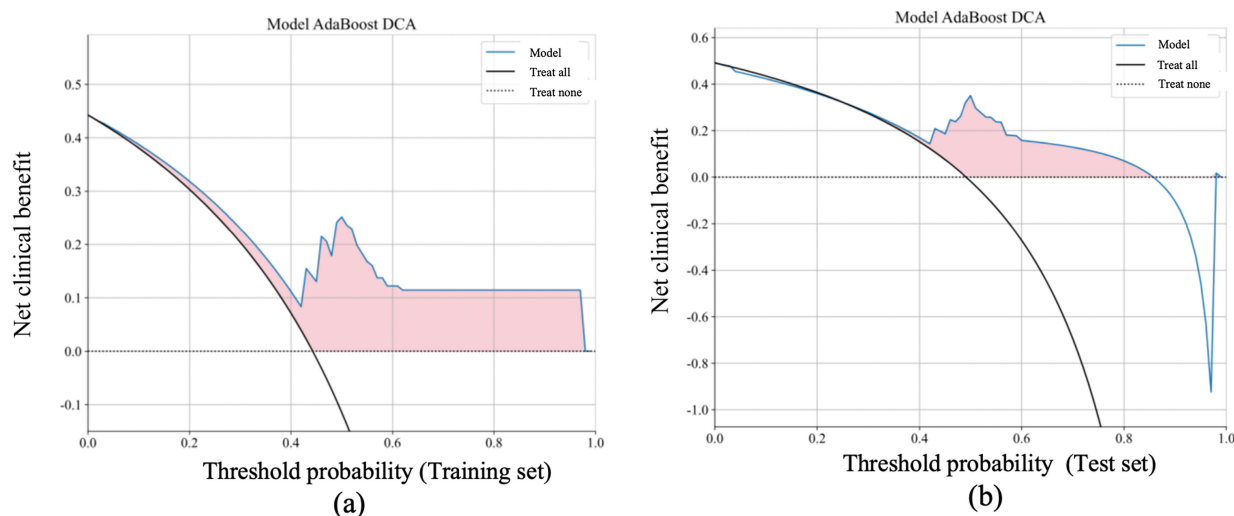


FIGURE 5

Net benefit of the AdaBoost model in the training and test sets based on DCA Curves, compared with "Treat all" and "Treat none" strategies. The DCA curves for the AdaBoost model, showing the net benefit in both the training (a) and test sets (b) at various threshold probabilities. The net benefit is calculated as the correct decision gain minus the loss from incorrect decisions. The horizontal axis represents the threshold probability, which indicates the minimum probability threshold for classifying a sample as positive. The vertical axis represents the net benefit. The model's performance (blue line) is compared with the "treat all" (black line) and "treat none" (dashed black line) strategies. The shaded pink area indicates the range where the AdaBoost model provides a higher net benefit compared to both "treat all" and "treat none" strategies. The AdaBoost model demonstrates significantly higher net benefit in the low threshold probability range (0–0.4), indicating its ability to effectively identify patients who require treatment at an early stage. In the moderate threshold range (0.4–0.6), the model still outperforms both the "treat all" and "treat none" strategies, suggesting its predictive value for clinical decision-making.

(32, 33). Collectively, these findings establish a foundation for applying AI-driven radiomics to differential diagnosis in occupational lung diseases.

The present study developed and validated multiple ML-based diagnostic models—including AdaBoost, LR, and SVM—to differentiate pneumoconiosis-associated large opacities from

peripheral lung cancer. Among these, the AdaBoost model demonstrated superior performance, achieving an accuracy of 80.9%, sensitivity of 81.0%, specificity of 80.8%, and an area under the ROC curve (AUC) of 0.900. These results affirm the feasibility of integrating radiomic features with ML algorithms to construct clinically valuable, non-invasive predictive

models. Importantly, DCA further verified the net clinical benefit of the proposed models, surpassing both “treat-all” and “treat-none” strategies across a broad range of threshold probabilities (34).

From a methodological standpoint, our study highlights the synergistic potential of combining traditional statistical models (e.g., LR) with more advanced ensemble learning techniques (e.g., AdaBoost and SVM) (35). This hybrid approach leverages both interpretability and predictive power, offering a balanced framework for clinical implementation. The inclusion of LASSO regression and Pearson correlation for dimensionality reduction ensured that only the most relevant, non-redundant features were retained, optimizing model performance and generalizability (36).

Nevertheless, several limitations warrant acknowledgment. Firstly, this was a single-center, retrospective study, which may introduce selection bias and limit external generalizability (37). Secondly, although the models achieved satisfactory accuracy, external validation using independent, multicenter datasets is essential to confirm robustness and clinical applicability (38). Thirdly, the current analysis relied solely on imaging-derived features and basic clinical parameters, without incorporating molecular, genomic, or proteomic biomarkers (39). Integrating multi-omics data in future studies could further enhance diagnostic precision and biological interpretability (39). Additionally, the implementation of explainable AI (XAI) frameworks may help elucidate model decision pathways, thereby increasing clinician confidence and promoting real-world adoption (40).

5 Conclusion

In conclusion, this study demonstrates that machine learning-based radiomic models, particularly the AdaBoost algorithm, can effectively differentiate between pneumoconiosis with large opacities and peripheral lung cancer, achieving robust diagnostic accuracy and measurable clinical benefit. With further multicenter validation and model interpretability enhancements, such AI-driven approaches hold great potential as practical, non-invasive decision-support tools in occupational and oncologic respiratory medicine, ultimately contributing to more accurate diagnosis, optimized treatment strategies, and improved patient outcomes (41).

Data availability statement

The original contributions presented in this study are included in this article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.

Ethics statement

This study was conducted following the ethical guidelines outlined in the Declaration of Helsinki. Ethical approval was obtained from the Institutional Review Board (IRB)

of Chongqing Medical and Pharmaceutical College, with approval number KYLLSC20240730015. All procedures involving human participants were performed in compliance with the institutional and national research committees' ethical standards. Informed consent was obtained from all participants involved in this study based on the Declaration of Helsinki. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

XiL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. WW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. XuL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. QLi: Conceptualization, Funding acquisition, Project administration, Validation, Writing – original draft, Writing – review & editing. Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Visualization. LZ: Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. WX: Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. Data curation, Formal analysis, Funding acquisition, Resources, Software, Visualization.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This study was supported by the Chongqing Health and Science Joint Project (Joint Key Laboratory Open Project) (No. 2026KFXM051), Natural Science Foundation of Chongqing (No. CSTB2025NSCO-GPX1116), Technological Innovation Project of Shapingba District, Chongqing (No. 2025016), 2024 Scientific Research

Project of Chongqing Medical and Pharmaceutical College (No. ygzrc2024101), Chongqing Municipal Education Commission Youth Project (Nos. KJQN202402821 and KJQN202502819), 2024 Chongqing Medical and Pharmaceutical College Innovation Research Group Project (No. ygz2024401), Science and Health Joint Medical Research Project of Shapingba District, Chongqing (No. 2024SQKWLHMS051), 2025 Scientific Research Project of Chongqing Medical and Pharmaceutical College (No. YGZZK2025116), 2025 Technological Innovation Project of Shapingba District, Chongqing (No. 2025031), Chongqing Municipal Education Commission Youth Project (Nos. KJQN202402821 and KJQN202302811), Chongqing Health and Science Joint Project (No. 2024MSXM115), In-house Pilot Research Grant of the First Affiliated Hospital of Chongqing Medical and Pharmaceutical College (No. 2022-2023MS04), 2023 Chongqing Medical Scientific Research Project (Joint Project of Chongqing Health Commission and Science and Technology Bureau) (No. 2023GGXM006), Postdoctoral Research Funding of the First Affiliated Hospital of Chongqing Medical University (No. 0303020203P0597, Grant to Xuemei Li), Foundation of Chongqing Key Laboratory of Prevention and Treatment for Occupational Diseases and Poisoning (No. 2021ZYBKF01), Medical Research Program of Chongqing Municipal Health, Health Committee (No. 2023WJWYX12), Key Projects of Science and Technology Research Program of Chongqing Municipal Education Commission (No. KJZD-K202302802), Chongqing Municipal Health Commission Medical Research Project (No. 2026WSJK013), and Chongqing Municipal Education Commission Youth Project (No. KJQN202302820), respectively.

Acknowledgments

We would like to extend my heartfelt appreciation to Zhenjun Xi, and Li Wang from the Chongqing Medical and Pharmaceutical College for their invaluable contributions to the data collection and experimental process. Their meticulous efforts ensured the accuracy and reliability of our research findings. Additionally, we grateful to Mei Yu, Jiemei Jiang, Lvsu Ye from the First Affiliated Hospital of Chongqing Medical and Pharmaceutical College for their assistance in data analysis and interpretation, which greatly enhanced the depth and quality of our research outcomes.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1675840/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Representative CT images showing different stages of pneumoconiosis (axial and coronal views). **(a,b)** Axial and coronal chest CT images of a Stage I pneumoconiosis patient demonstrate multiple scattered, small, round, high-density nodules distributed in both lungs, predominantly in the right middle lung field. **(c,d)** Stage II pneumoconiosis is characterized by an increased number of small nodules and linear opacities, primarily clustered in the posterior segment of the right upper lobe. **(e,f)** Stage III pneumoconiosis exhibits more extensive and symmetrical large opacities (long diameter > 20 mm, short diameter > 10 mm) in the apical and posterior segments of both upper lobes. The lesions are accompanied by pleural thickening, traction, and partial calcification within the masses, along with multiple peripheral nodular opacities. *Interpretation:* as pneumoconiosis progresses from Stage I to Stage III, CT imaging reveals a gradual increase in nodule size and density, evolving into confluent fibrotic masses with pleural involvement and architectural distortion, consistent with advanced PMF.

SUPPLEMENTARY FIGURE 2

CT imaging and immunohistochemical (IHC) profiles of different histological types of lung cancer. **(a,b)** Axial and coronal CT images of squamous cell carcinoma in the left upper lobe reveal a soft-tissue mass located in the anterior segment, exhibiting short spiculation and well-defined margins. IHC profile: CK (+), CK7(+), CK5/6(few +), TTF-1(–), NapsinA(–), CgA(–), Syn(–), P40(+), P63(+), CDX-2(–), CK20(–), Ki-67 (~80% +), Villin(–). Diagnosis: Squamous cell carcinoma of the lung. **(c)** Axial CT image of poorly differentiated adenocarcinoma in the right middle lobe shows an ill-defined mass with partial truncation of the middle bronchus. IHC profile: CK (+), CK7(+), TTF-1(+), NapsinA(–), CgA(–), Syn(–), CD56(–), INSM-1(+), CK5/6(–), P40(–), P63(–), Ki-67 (~70% +). Diagnosis: Poorly differentiated adenocarcinoma. **(d)** Axial CT image of a poorly differentiated small cell carcinoma in the right lower lobe demonstrates a large, irregular mass with unclear margins, bronchial obstruction, and significant mass effect. IHC profile: CK (+), Ki-67 (~60% +), CK7(+), P63(–), P40(–), TTF-1(–), NapsinA(–), Syn(weak +), CgA(–), CK5/6(weak +), INSM-1(–), CD56(weak +), HMB-45(–), LCA(–), INI-1(+), S100(–), NUT(–), CD20(–). Diagnosis: Poorly differentiated small cell carcinoma confirmed by immunohistochemistry. *Interpretation:* Distinct CT and immunohistochemical patterns are evident among different lung cancer subtypes. While squamous cell carcinoma typically presents as a well-circumscribed lesion with peripheral spiculation, adenocarcinoma and small cell carcinoma often exhibit irregular borders, bronchial involvement, and aggressive radiologic behavior. Combined radiologic and IHC assessment remains essential for accurate histopathological classification.

References

- Gong X, Cui K, Liu Z, Tang Y, Li J, Chen Y, et al. The disease burden of pneumoconiosis in China from 1990 to 2021 and projection to 2050 based on GBD 2021: a cross-sectional study. *Front Public Health*. (2021) 13:1583895. doi: 10.3389/fpubh.2025.1583895
- Guo LW, Lyu ZY, Liu Y, Meng QC, Zheng LY, Chen Q, et al. The probability of lung cancer in patients with pulmonary nodules detected via low-dose computed tomography screening in China. *BMC Cancer*. (2025) 25:1058. doi: 10.1186/s12885-025-14449-6
- Machineni NVK, Gullapalli D, Tangutoori S, Tippiarathi A, Lahori S, Khater F, et al. Dust to disease: cavitory lung lesion with *Mycobacterium kansasii* in a black lung patient. *J Investig Med High Impact Case Rep*. (2025) 13:23247096251348906. doi: 10.1177/23247096251348906
- Bao Z, Xu B, Zhang X, Yin Y, Yang X, Niu Q. Potential pneumoconiosis patients monitoring and warning system with acoustic signal. *Sensors*. (2025) 25:1874. doi: 10.3390/s25061874
- Çakmakçı Karakaya S, Hasanlı YS, Demir AUA. welder case with secondary haemochromatosis, which rarely accompanies pneumoconiosis. *Occup Med*. (2025) 75:256–60. doi: 10.1093/occmed/kqaf030
- Caton T, Pescatore J, Criner G. The importance of asking a thorough social history; a case report of aluminum pneumoconiosis. *BMC Pulm Med*. (2025) 25:200. doi: 10.1186/s12890-025-03662-5
- Chau NK, Park EK, Choi S. Longitudinal analysis of coal workers' pneumoconiosis using enhanced resolution-computed tomography images: unveiling patterns in lung structure, function, and clinical correlations. *Front Physiol*. (2025) 16:1578058. doi: 10.3389/fphys.2025.1578058
- Chen Y, Liu D, Ji H, Li W, Tang Y. Global and regional burden of pneumoconiosis, 1990–2021: an analysis of data from the global burden of disease study 2021. *Front Med*. (2021) 12:1559540. doi: 10.3389/fmed.2025.1559540
- Zhou J, Xu H, Du W, Peng L. The first case report of inactive nontuberculous mycobacterial pulmonary disease (NTM-PD) in a pneumoconiosis patient caused by *Mycobacterium europaeum* in China. *Infect Drug Resist*. (2024) 17:1515–21. doi: 10.2147/IDR.S448805
- Yamano S, Umeda Y. Fibrotic pulmonary dust foci is an advanced pneumoconiosis lesion in rats induced by titanium dioxide nanoparticles in a 2-year inhalation study. *Part Fibre Toxicol*. (2025) 22:7. doi: 10.1186/s12989-025-00623-y
- Xue B, Cao G, Ren L, Zhao Y, Lv X, Li Y, et al. Brain functional alterations in early stage of coal workers' pneumoconiosis with alcoholism: insights from a resting-state fMRI investigation. *Front Neurosci*. (2025) 19:1610657. doi: 10.3389/fnins.2025.1610657
- López-Guillén A, León-Jiménez A, Delgado-García D, López-López A. [Proposal for an International Pneumoconiosis Observatory in 2025]. *Arch Prev Riesgos Labor*. (2025) 28:89–98. doi: 10.12961/aprl.2025.28.01.08
- Wei Y, Jia Z, Ma J, Zhang W, Li H, Wu J, et al. Proteomics and metabolomics analyses reveal a dynamic landscape of coal workers' pneumoconiosis: an insight into disease progression. *J Proteome Res*. (2025) 24:1715–31. doi: 10.1021/acs.jproteome.4c00715
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
- Wang J, Min C, Yu F, Chen K, Mao L. Cough recognition in pneumoconiosis patients based on a flexible patch with an embedded ACC sensor for remote monitoring. *BMC Med Inform Decis Mak*. (2025) 25:41. doi: 10.1186/s12911-025-02879-y
- Vella F, Filetti V, Matera S, Spinello S, Vinnikov D, Muscato G, et al. Risk of pneumoconiosis in workers exposed to crystalline silica from lava rock dust from mount etna. *J Clin Med*. (2025) 14:3781. doi: 10.3390/jcm14113781
- Sood VA, Rishel Brakey H, Myers O, Shore X, Sood A. Interprofessional discussion for knowledge transfer in a digital “Community of Practice” for managing pneumoconiosis: mixed methods study. *JMIR Form Res*. (2025) 9:e67999. doi: 10.2196/67999
- Shi Y, Li C, Zhou S, Li C, Meng F, Wang X, et al. Pneumoconiosis and chronic diseases: a narrative review. *Iran J Public Health*. (2025) 54:521–9. doi: 10.18502/ijph.v54i3.18245
- Akgündüz B. The relationship between connective tissue disease and autoantibody positivity in pneumoconiosis cases. *Medicine*. (2025) 104:e41702. doi: 10.1097/MD.00000000000041702
- Ruiz ER, Arellano CA, Archila CA, Llobet C, Carrasco G, Pinochet F. Clinical validation of an AI system for pneumoconiosis detection using chest X-rays. *J Occup Environ Med*. (2025) 67:e250–6. doi: 10.1097/JOM.00000000000003329
- Ren X, Ji G, Chu S, Yoshida S, Zhao J, Jia B, et al. A multimodal similarity-aware and knowledge-driven pre-training approach for reliable pneumoconiosis diagnosis. *J Xray Sci Technol*. (2025) 33:229–48. doi: 10.1177/08953996241296400
- Pakbaz Y, Moodi F. Coal worker's pneumoconiosis mimicking lung cancer in a 75-year-old woman with a history of breast cancer: a clinical case report. *Clin Case Rep*. (2025) 13:e70049. doi: 10.1002/ccr3.70049
- Liu L, Tian X, Guo Y, Yu Y, Wang Y, Wang W, et al. Association between interleukin gene polymorphisms and the risk of pneumoconiosis: a systematic review and meta-analysis. *Front Med*. (2025) 12:1479730. doi: 10.3389/fmed.2025.1479730
- Ma H, Dong Z, Zhang X, Li N, Liu C, Zhou X, et al. Microbial dysbiosis in the lung and gut in response to inhalable particulate matters in pneumoconiosis patients and animals. *Environ Sci Technol*. (2025) 59:10826–40. doi: 10.1021/acs.est.5c00798
- Yu XW, Wang Q, Liu L, Zhou ZJ, Cai T, Yuan HM, et al. Detection and genomic characterization of torque teno virus in pneumoconiosis patients in China. *Viruses*. (2024) 16:1059. doi: 10.3390/v16071059
- Yu X, Xiong T, Yu L, Liu G, Yang F, Li X, et al. Gut microbiome and metabolome profiling in coal workers' pneumoconiosis: potential links to pulmonary function. *Microbiol Spectr*. (2024) 12:e0004924. doi: 10.1128/spectrum.00049-24
- Fortarezza F, Perilli M, Della Barbera M, Pezzuto F, Faccioli E, Coconcelli E, et al. Giant cell interstitial pneumonia: case series with comprehensive ultrastructural analyses of “not only” hard metal pneumoconiosis. *Histopathology*. (2025) 86:450–9. doi: 10.1111/his.15335
- Yang L, Deng H, Chen H, Wu M, Li J, Zhang T. Correlations of mucin 5B gene polymorphisms and expression levels with the risk of onset of coal workers' pneumoconiosis. *Medicine*. (2024) 103:e41088. doi: 10.1097/MD.00000000000041088
- Li X, Li Q, Xie X, Wang W, Li X, Zhang T, et al. Integrating CT radiomics and clinical data with machine learning to predict fibrosis progression in coalworker pneumoconiosis. *Front Med*. (2025) 12:1599739. doi: 10.3389/fmed.2025.1599739
- Xiong L, Liu X, Qin X, Li W. Accurate pneumoconiosis staging via deep texture encoding and discriminative representation learning. *Front Med*. (2024) 11:1440585. doi: 10.3389/fmed.2024.1440585
- Dong H, Zhu B, Kong X, Su X, Liu T, Zhang X. Deep learning-based algorithm for classifying high-resolution computed tomography features in coal workers' pneumoconiosis. *Biomed Eng Online*. (2025) 24:7. doi: 10.1186/s12938-025-01333-4
- Warkentin MT, Al-Sawaihey H, Lam S, Liu G, Diergaarde B, Yuan JM, et al. Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches. *Thorax*. (2024) 79:307–15. doi: 10.1136/thorax-2023-220226
- Dong H, Zhu B, Kong X, Zhang X. Efficient clinical data analysis for prediction of coal workers' pneumoconiosis using machine learning algorithms. *Clin Respir J*. (2023) 17:684–93. doi: 10.1111/crj.13657
- Hu X, Liu XD, Gao T, Guo Q. The application value of HRCT quantitative analysis in early pneumoconiosis and concomitant diseases. *Medicine*. (2025) 104:e42374. doi: 10.1097/MD.00000000000042374
- Zhou Z, Li X, Ji H, Xu X, Chang Z, Wu K, et al. Application of improved Unet network in the recognition and segmentation of lung CT images in patients with pneumoconiosis. *BMC Med Imaging*. (2024) 24:220. doi: 10.1186/s12880-024-01377-3
- Jia J, Huang J, Cui Y, Zhang D, Li H, Wang S, et al. Machine learning prediction of coal workers' pneumoconiosis classification based on few-shot clinical data. *Digit Health*. (2025) 11:20552076251359498. doi: 10.1177/20552076251359498
- He R, Mao L. Diffuse pulmonary ossification with pneumoconiosis confirmed on transbronchial lung biopsy. *Pol Arch Intern Med*. (2025) 135:16919. doi: 10.20452/pamw.16919
- Zhu H, Zhou L, Zhou J, Han L, Wu M. Analysis of mortality and life expectancy determinants among 5,791 deceased pneumoconiosis patients - Jiangsu Province, China, 2011–2023. *China CDC Wkly*. (2024) 6:1417–24. doi: 10.46234/ccdcw2024.280
- Zhang Y, Zheng B, Zeng F, Cheng X, Wu T, Peng Y, et al. Potential of digital chest radiography-based deep learning in screening and diagnosing pneumoconiosis: an observational study. *Medicine*. (2024) 103:e38478. doi: 10.1097/MD.00000000000038478
- Zhang Y, Sun D, Song Y, Ye Q. Candidate gene polymorphisms associated with silicosis and coal workers' pneumoconiosis: a systematic review and meta-analysis. *BMC Pulm Med*. (2024) 24:580. doi: 10.1186/s12890-024-03392-0
- Zhang J, Li Y, Zhu F, Guo X, Huang Y. Time-/dose- series transcriptome data analysis and traditional Chinese medicine treatment of pneumoconiosis. *Int J Biol Macromol*. (2024) 267(Pt 2):131515. doi: 10.1016/j.ijbiomac.2024.131515