BMC Cancer

RESEARCH

Open Access

Comparing radiomics, deep learning, and fusion models for predicting occult pleural dissemination in patients with non-small cell lung cancer: a retrospective multicenter study

Tao Bao[1†], Xiaoguang Li[2†], Yuanlin Deng[1†], Liang Chen[3†], Weijie Sun[4†], Mingjian Ge[5], Jigang Dai[3], Xiaolong Zhao[6], Xu Chen[1], Liang Zhang[1], Lei Bao[7*] and Wei Guo[1*]

Abstract

Background Occult pleural dissemination (PD) in non-small cell lung cancer (NSCLC) patients is likely to be missed on computed tomography (CT) scans, associated with poor survival, and generally contraindicated for radical surgery. This study aimed to develop and compare the performance of radiomics-based machine learning (ML), deep learning (DL), and fusion models to preoperatively identify occult PD in NSCLC patients.

Materials and methods A total of 326 NSCLC patients from three Chinese high-volume medical centers (2016–2023) were retrospectively collected and divided into training ($n=216$), internal test ($n=54$), and external test ($n=56$) cohorts. Ten radiomics-based ML models and eight DL models were trained using CT images at the maximum cross-sectional slice of the primary tumor. Moreover, another two fusion models (prefusion and postfusion) were developed using feature-based and decision-based methods. The receiver operating characteristic curve (ROC) and area under the curve (AUC) were mainly used to compare the predictive performance of the models.

Results The GBM (AUC: 0.821) and DenseNet121 (AUC: 0.764) models achieved the highest AUC among ML and DL models in the external test cohorts, respectively. The postfusion model, integrating the output probabilities from GBM and DenseNet121 models, showed superior performance (AUC: 0.828–0.978) compared to the prefusion model (AUC: 0.817–0.877). Moreover, the postfusion model demonstrated the highest degree of sensitivity (82.1–97.2%) among all models across the three cohorts.

†Tao Bao, Xiaoguang Li, Yuanlin Deng, Liang Chen and Weijie Sun are co-first author and contributed equally to this work.

*Correspondence:
Lei Bao
baolei1219@sina.cn
Wei Guo
gyguowei@hotmail.com

Full list of author information is available at the end of the article

Conclusions The postfusion model, which integrates radiomics-based ML and DL models, can serve as a sensitive diagnostic tool to predict occult PD in NSCLC patients, thereby helping to avoid unnecessary surgeries.

Keywords Non-small cell lung cancer, Occult pleural dissemination, Radiomics, Deep learning, Machine learning, Prediction

Introduction

Non-small cell lung cancer (NSCLC) with pleural dissemination (PD) is categorized as stage M1a in the 8th edition of the American Joint Committee on Cancer (AJCC) TNM staging system, with a median survival time (MST) of 4–11.5 months, and is generally contraindicated for radical surgery [1, 2]. For PD without pleural effusion, typical imaging features include multiple pleural nodules and uneven pleural thickening on CT scans, and fluorodeoxyglucose (FDG) uptake in pleural lesions at positron emission tomography (PET) [3–5]. Nevertheless, radiologists frequently encounter difficulties in detecting small pleural metastases or in differentiating them from benign conditions such as intrapulmonary lymph nodes and reactive pleurisy. A critical clinical challenge arises: approximately 0.9–6.2% of operable NSCLC patients were unexpectedly diagnosed with PD at thoracotomy, a condition known as occult PD [6–8]. For these affected cases, if PD can be accurately detected before surgery, unnecessary open-close surgery can be avoided [9, 10].

Prior studies on this topic remain scarce. Li et al. [11] identified that age ≤ 50, elevated carcinoembryonic antigen (CEA) levels, advanced N stage, adenocarcinoma and pleural invasion (PI) were independent risk factors for the occult PD. They developed a prediction model with an area under the curve (AUC) of 0.756. Recently, there have been notable developments in the application of radiomics-based machine learning (ML) and deep learning (DL) technology in the medical field, including the diagnosis and prognosis assessment of cancer [12–15]. In contrast to conventional experience-based radiological assessment, radiomics and DL facilitate the quantitative characterization of disease phenotypes through high-throughput feature extraction from medical images. The extracted features possess the capacity to capture underlying pathophysiological information, which is typically difficult to recognize through the naked eyes. Employing CT images radiomics features, Yang and colleagues [16] developed a predictive model for occult PD, exhibiting an AUC of 0.93. Nevertheless, to date, no studies have employed DL algorithms for the preoperative prediction of occult PD in NSCLC.

The objective of this study was to develop and validate a noninvasive tool for preoperative identification of occult PD in high-risk NSCLC patients to prevent unnecessary surgery. We hypothesized that a fusion model combining radiomics and DL algorithms would achieve superiority over single-modality strategies by integrating complementary information.

Materials and methods

Patient eligibility

From January 2016 to December 2023, we retrospectively enrolled 163 consecutive NSCLC patients with surgically confirmed occult PD from three high-volume centers in China. Inclusion criteria were: (1) pathologically confirmed primary NSCLC with malignant PD, (2) no preoperative treatment, and (3) clinicopathological data were complete. Patients meeting any exclusion criteria were omitted: (1) pleural effusion detected preoperatively, (2) preoperatively diagnosed with PD, and (3) poor CT quality or no CT scans within 1 month before surgery. Additionally, 163 NSCLC patients without PD and with postoperative pathology-confirmed PI were randomly selected from the above 3 hospitals during the same period to provide a control group of cases. In the present study, the training and internal test processes utilized data from two medical centers: the Daping Hospital of Army Medical University (DPAMU) and the First Affiliated Hospital of Chongqing Medical University (CQMU). The predictive performance of the models was then compared using the external test cohort from the Xinqiao Hospital of Army Medical University (XQAMU, Fig. 1a). Data were obtained by reviewing medical records. Disease stage was determined according to the general rules of the TNM classification of malignant tumors (8th edition). The study protocol was in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Daping Hospital (approval number: 2023 – 299; approval date: December 13, 2023), and all collaborative research centers also obtained permission from their respective Ethics Committees. All patients signed the informed consent form prior to the surgery. The need for individual consent to the utilization of their clinicopathological data for scientific research purposes was waived by the committee.

Reference standard for PD and PI

In the present study, the gold standard for diagnosis of PD and PI was based exclusively on surgical specimens obtained during thoracotomy or video-assisted thoracic surgery (VATS). In detail, the diagnosis of PD was confirmed intraoperatively by the presence of macroscopic tumor nodules on the parietal or visceral pleura. These lesions were meticulously resected and subjected to

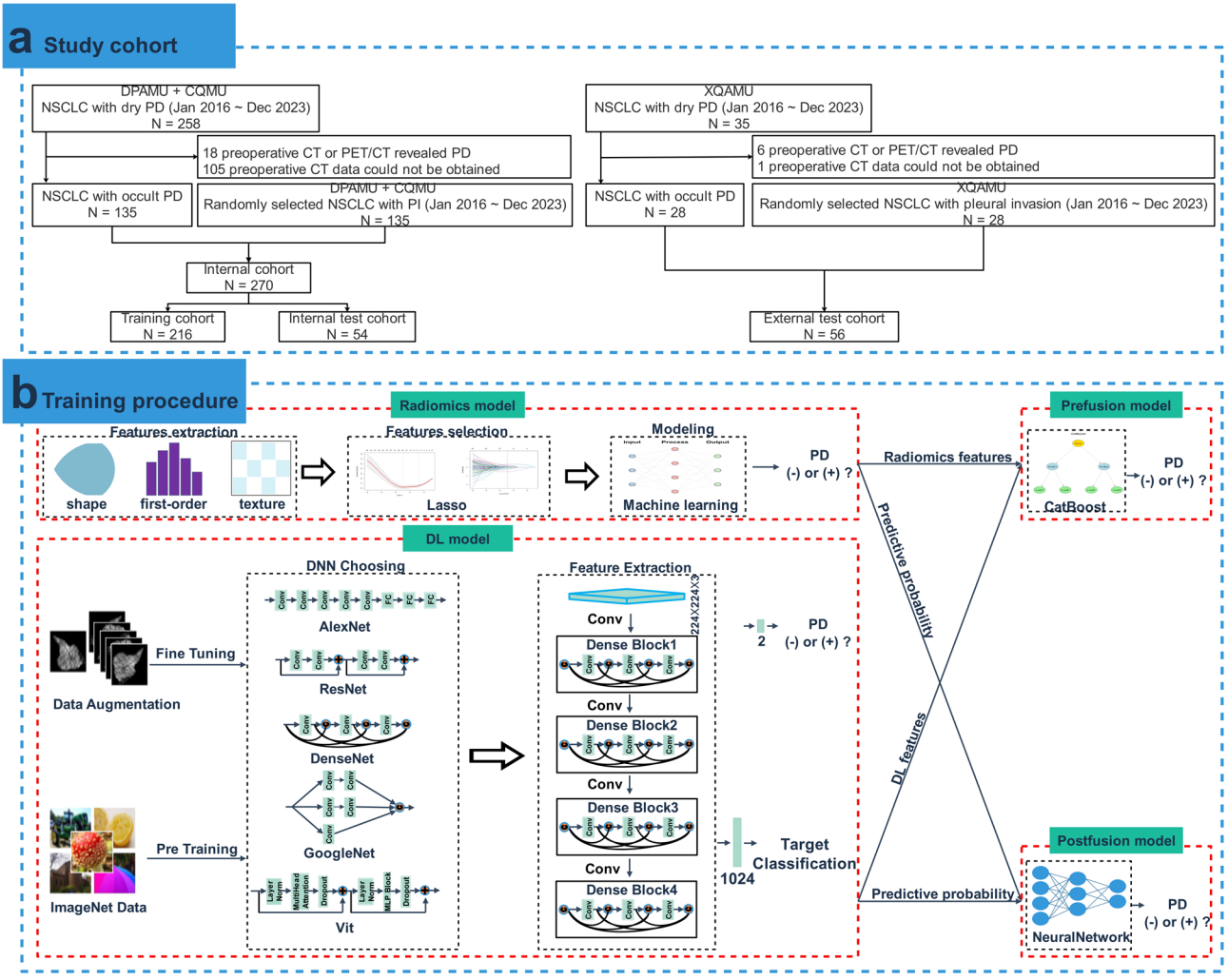

Fig. 1 Study flowchart and models construction. (a) Flowchart elucidating the selection process of 326 patients from three medical centers. (b) Feature extraction, selection and predictive models construction. For traditional machine learning (ML) models construction, radiomics features were extracted by using RIAS software and refined by least absolute shrinkage and selection operator (LASSO) regression. A total of 10 ML algorithms were used to develop the radiomics-based ML models. For deep learning (DL) models construction, 8 convolutional neural networks (CNNs) were pre-trained using the ImageNet dataset, and transfer learning was subsequently performed on the training cohort. DL features extracted by employing the DenseNet121 model. CatBoost was selected to train the prefusion model. The output probabilities from the GBM and DenseNet121 models were integrated through the implementation of a stacking ensemble learning strategy, and the NeuralNetwork classifier was selected to construct the postfusion model

intraoperative frozen section analysis, which were subsequently confirmed through definitive histopathological examination. With regard to PI, the diagnosis was determined by postoperative histopathological examination of the presence of primary tumor invasion beyond the elastic layer (PL1 or PL2). Elastic staining, as well as hematoxylin and eosin staining, were routinely performed [17]. The aforementioned reference test procedures were applied consistently for all patients across the three participating centers.

CT scans and image preprocessing

Chest CT scans were performed according to standardized scanning protocols within 1 month before surgery by using CT scanners. The primary CT scanner models and their corresponding acquisition parameters for each participating center are detailed in Supplementary Table S1. All tumors were preoperatively staged as resectable.

As demonstrated in Supplementary Figure S1, a thoracic radiologist (L.X.G., with 10 years of experience) manually delineated regions of interest (ROI) on two-dimensional (2D) CT plain scan images at the maximum cross-sectional slice of the primary tumor in a blinded manner by utilizing 3D Slicer software (version 4.11, https://www.slicer.org/). The ROI was delineated to include only the primary tumor within the lung parenchyma. The pleural surface or separate pleural lesions were not included in the segmentation. 6 months later, 30 patients from DPAMU were randomly selected for ROI delineation by L.X.G. and another thoracic surgeon (B.T.,

with 8 years of experience). This process was repeated to construct two resegmentation datasets for intraclass correlation coefficient (ICC) analysis.

All CT images were resampled to the identical voxel size of $1 \times 1 \times 1$ mm^3; and the pixel intensity was normalized by z-score.

Radiomics feature extraction, selection and ML models construction

RIAS (version 1.0.0, https://www.notion.so/RIAS-916ad7256e1e472985d4b11c8ebf0fe0) was employed to extract the radiomics features from the training and test cohorts. This process was conducted in accordance with the Image Biomarker Standardization Initiative (IBSI) guidelines [18]. A total of 1,427 radiomics features, comprising first-order statistics features, shape-based features, and texture features, were extracted from each ROI and then standardized using z-score normalization. Subsequently, intraobserver and interobserver ICC analysis was performed on the training cohort, and the stable features with ICCs >0.75 were preserved. Besides, Spearman's correlation analysis was conducted between each feature, and only one of the pairs with a correlation coefficient greater than 0.9 was retained to reduce redundancy. After that, least absolute shrinkage and selection operator (LASSO) regression algorithm with penalty parameter tuning via 10-fold cross-validation was employed. The lambda.1se criterion was utilized to further refine the optimal features. Furthermore, ten ML algorithms, including logistics, support vector machine (SVM), gradient boosting machine (GBM), neuralnetwork, randomforest (RF), eXtreme Gradient Boosting (XGBoost), k-nearest neighbour (KNN), Adaboost, lightGBM, and CatBoost, were used to develop the radiomics-based ML models. The predictive performance of above ML models was then compared using the internal and external test cohorts.

DL models development and feature extraction

A total of eight neural networks were pre-trained using the ImageNet dataset, including convolutional neural networks (AlexNet, ResNet, DenseNet121, GoogleNet, SqueezeNet, and Vgg) and transformer networks (SwinTransformer and VisionTransformer). Given the limited amount of data collected, data enhancement strategies were employed, including horizontal flipping, vertical flipping, and rotational transformation. Given the image patches span a range of sizes from 15 × 15 to 85 × 85, all the ROI patches were resized to 224 × 224 prior to being fed into the networks. The models had three input channels, that is the inputs are identical stacked grayscale images. After transitioning from a 1000-category classification task to a binary one, these pre-trained models were applied to our dataset using transfer learning techniques with fine-tuning of dropout, and fully connected layers. Early stopping with a reduced patience of 5 epochs was implemented to halt training once validation performance plateaued. The predictive performance of the models was evaluated using metrics such as accuracy, sensitivity, specificity, etc. The patient-level performance was then obtained by averaging all the CT patches for one patient.

The DenseNet121 architecture, which exhibited superior performance in comparison to other DL models on the dataset, was then employed to extract DL features. The features were extracted from the network following the last dense block and prior to the final fully-connected classification layer. Specifically, a 1,024-dimensional feature vector was obtained by applying global average pooling to the output of the final convolutional layer. The features were then standardized using z-score normalization, and furthermore compressed to a set of 115 features using principal component analysis (PCA).

Fusion models development

Two fusion models, namely prefusion and postfusion, were developed using feature-based and decision-based methods upon the aforementioned ML and DL models.

In detail, LASSO with penalty parameter tuning via 10-fold cross-validation was utilized to further select the optimal DL features. In combination with the selected radiomics features, the CatBoost algorithm was selected to train the prefusion model.

The postfusion model was constructed by integrating the output probabilities from the radiomics-based ML (GBM) and DL (DenseNet121) models through the implementation of a stacking ensemble learning strategy. The NeuralNetwork classifier was selected to construct the postfusion model. The overall workflow of this study is illustrated in Fig. 1b.

Finally, the predictive performance of the above ML (GBM), DL (DenseNet121), prefusion and postfusion models was compared using the internal and external test cohorts.

Statistical analysis

Continuous data were compared using the Mann-Whitney U test, and categorical data were compared using Fisher's exact test or X^2 test, as appropriate. The receiver operating characteristic curve (ROC) and the area under the curve (AUC) were mainly used to compare the predictive performance of different models. 95% confidence intervals (CI) for AUCs were estimated. The Delong's test was utilized to compare AUCs. Decision curve analysis (DCA) was performed to evaluate the clinical usefulness of the models. SHapley Additive exPlanations (SHAP) were used for interpretation of the contributions of each variable in the final model. P values of less than 0.05 were

considered statistically significant. All statistical analyses and graphical presentation were performed using Python (version 3.8.0, http://www.python.org), R (version 3.6.0, http://www.R-project.org), and SPSS (version 25.0, IBM, Inc.).

Results

Patient characteristics

As demonstrated in Table 1, a total of 163 consecutive NSCLC patients with surgically confirmed occult PD were identified, and another 163 NSCLC patients without PD but with postoperative pathology-confirmed PI were randomly selected to provide a control group of cases. All patients were divided into a training cohort ($n = 216$), an internal test cohort ($n = 54$), and an external test cohort ($n = 56$). There was no significant difference among all cohorts with the exception of age ($p = 0.020$).

Radiomics-based ML models construction

After ICC analysis was performed, a total of 637 stable features were preserved. Spearman's correlation analysis was then conducted, and 257 features were retained to reduce redundancy. Subsequently, LASSO regression algorithm was employed, and 12 radiomics features were finally identified for the modeling (Supplementary Figure S2, Table S2). The mean intra-/inter-observer ICC value across all 1,427 initially extracted radiomics features was 0.70 (95%CI: 0.68–0.71) and 0.71 (95%CI: 0.69–0.72). The mean intra-/inter-observer ICC value across the ultimately selected radiomics features was 0.87 (95%CI: 0.82–0.92) and 0.90 (95%CI: 0.86–0.94).

Among all 10 radiomics-based ML models, the GBM model attained the highest AUC (0.821) in the external test cohorts (Fig. 2a-c, Supplementary Table S3-4). The DCA indicated that the GBM model could serve as a good model for predicting occult PD (Supplementary Figure S3a-c). Furthermore, the SHAP analysis results demonstrated the most predictive radiomics feature was squareroot_firstorder_StandardDeviation. Lower values of this feature were associated with higher SHAP values (Fig. 2d-e).

Table 1 Patients' characteristics

Characteristics	Entire cohort ($n = 326$)	Internal cohort ($n = 270$)		External test cohort ($n = 56$)	*P* value
		Training cohort ($n = 216$)	Internal test cohort ($n = 54$)		
	PD+: 163; PD-: 163	PD+: 108; PD-: 108	PD+: 27; PD-: 27	PD+: 28; PD-: 28	
Age, years (IQR)	59 (53–67)	59 (53–66)	63 (54–69)	57 (50–67)	0.020
Sex					0.106
Female	168 (51.5)	110 (50.9)	23 (42.6)	35 (62.5)	
Male	158 (48.5)	106 (49.1)	31 (57.4)	21 (37.5)	
Histology subtype					0.872
Adenocarcinoma	294 (90.2)	195 (90.3)	48 (88.9)	51 (91.1)	
Squamous cell carcinoma	14 (4.3)	8 (3.7)	3 (5.6)	3 (5.4)	
Others	18 (5.5)	13 (6.0)	3 (5.6)	2 (3.6)	
Tumor size, cm (IQR)	2.6 (2.0–3.3)	2.5 (2.0–3.3)	2.8 (2.3–3.8)	2.4 (1.9–3.3)	0.061
Right-sided tumor	174 (53.4)	108 (50.0)	34 (63.0)	32 (57.1)	0.195
Tumor location					0.842
Periphery	313 (96.0)	208 (96.3)	52 (96.3)	53 (94.6)	
Center	13 (4.0)	8 (3.7)	2 (3.7)	3 (5.4)	
Clinical T stage					0.291
T1	216 (66.3)	146 (67.6)	30 (55.6)	40 (71.4)	
T2	98 (30.1)	61 (28.2)	21 (38.9)	16 (28.6)	
T3	10 (3.1)	8 (3.7)	2 (3.7)	0	
T4	2 (0.6)	1 (0.5)	1 (1.9)	0	
Clinical N stage					0.056
N0	227 (69.6)	151 (69.9)	37 (68.5)	39 (69.6)	
N1	23 (7.1)	9 (4.2)	5 (9.3)	9 (16.1)	
N2	72 (22.1)	52 (24.1)	12 (22.2)	8 (14.3)	
NX	4 (1.2)	4 (1.9)	0	0	
Serum tumor marker level*					0.297
High	108 (33.1)	71 (32.9)	22 (40.7)	15 (26.8)	
Normal or unknown	218 (66.9)	145 (67.1)	32 (59.3)	41 (73.2)	

Abbreviation: *IQR* interquartile range; SD, standard deviation

*Serum tumor markers include carcinoembryonic antigen and cytokeratin 19 fragment

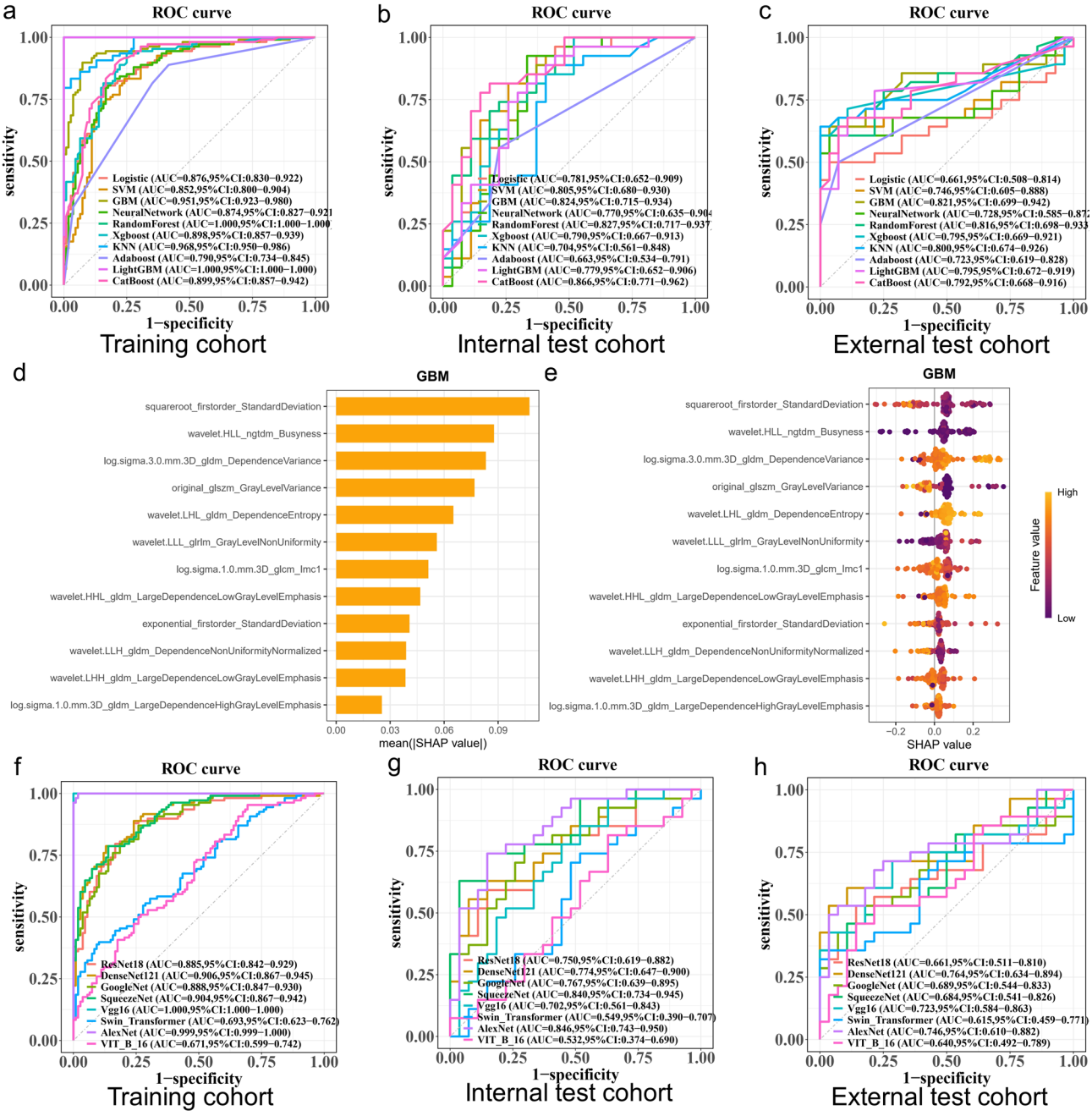

Fig. 2 Performance of radiomics-based machine learning (ML) and deep learning (DL) models. Receiver operating characteristic curves (ROC) of the ML models in the training (**a**), internal test (**b**), and external test (**c**) cohorts. SHAP analysis of the GBM model (**d**, **e**). ROC of the DL models in the training (**f**), internal test (**g**), and external test (**h**) cohorts

DL models construction

Among all DL models, the DenseNet121 attained the highest AUC (0.764) in the external test cohorts (Fig. 2f-h, Supplementary Table S5-6). The DCAs indicated that the DenseNet121 could serve as a good model for predicting occult PD (Supplementary Figure S4).

Prefusion and postfusion models construction

a total of 1,024 DL features were extracted through DenseNet121. The features were then compressed to a set of 115 features using PCA (Supplementary Figure S5a). Finally, 7 DL features were identified by utilizing LASSO regression algorithm (Supplementary Figure S5b-c and Table S7). These DL features, along with the previously extracted radiomics features, were then utilized to train the prefusion model employing the CatBoost algorithm.

The correlation heatmap demonstrated low cross-modality correlations between radiomics and DL features, supporting their complementary roles in PD prediction (Supplementary Figure S6). This model achieved an AUC of 0.817–0.846 in the test cohorts (Fig. 3a-c).

The postfusion model, which was constructed by integrating the output probabilities from the radiomics-based ML (GBM) and DL (DenseNet121) models, exhibited an AUC of 0.828–0.874 in the test cohorts, which was higher than those of the ML ($p = 0.218$–0.866), DL ($p = 0.082$–0.083), and prefusion model ($p = 0.590$–0.880; Fig. 3e-f; Table 2).

The calibration curves for the postfusion model across all three cohorts were shown in Supplementary Figure S7. The Brier scores for the training, internal test, and external test cohorts were 0.058, 0.148, and 0.171, respectively.

The DCA (Fig. 3h-i) demonstrated that the radiomics, prefusion and postfusion models could serve as effective

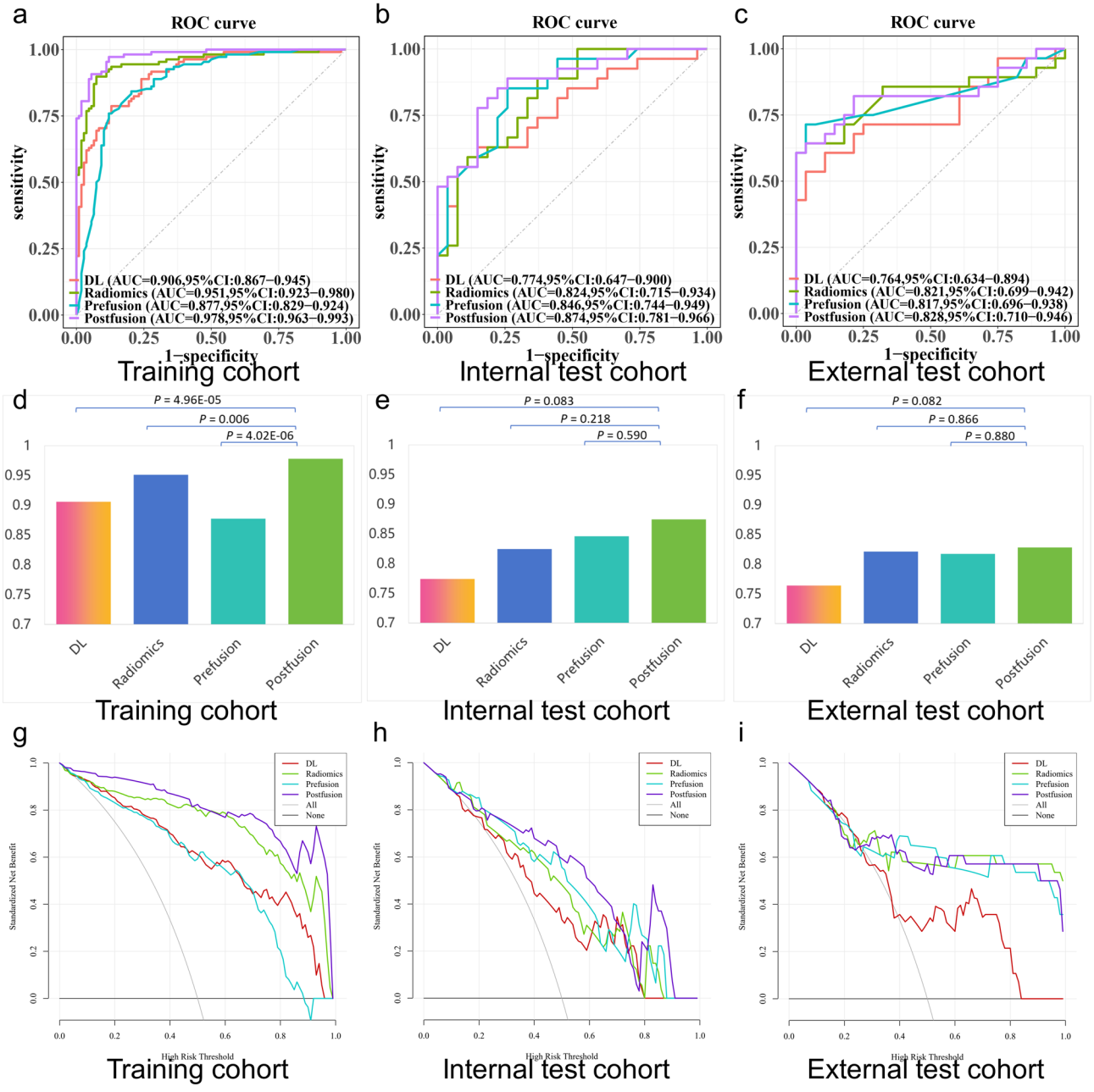

Fig. 3 Performance comparison of radiomics-based machine learning (ML), deep learning (DL), prefusion and postfusion models. Receiver operating characteristic curves (ROC) among the four models in the training (**a**), internal test (**b**), and external test (**c**) cohorts. The Delong's test among the four models in the training (**d**), internal test (**e**), and external test (**f**) cohorts. Decision curve analysis (DCA) of the four models in the training (**g**), internal test (**h**), and external test (**i**) cohorts

Table 2 Performances of the predictive models in the training and test sets

Models	AUC (95%CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 (%)
Training cohort ($n=216$)						
Deep learning	0.906 (0.867–0.945)	82.9	78.7	87.0	85.9	82.1
Radiomics	0.951 (0.923–0.980)	91.2	89.8	92.6	92.4	91.1
Prefusion	0.877 (0.829–0.924)	81.9	84.3	79.6	80.5	82.4
Postfusion	0.978 (0.963–0.993)	92.6	97.2	88.0	89.0	92.9
Internal test cohort ($n=54$)						
Deep learning	0.774 (0.647-0.900)	74.1	63.0	85.2	81.0	70.8
Radiomics	0.824 (0.715–0.934)	75.9	88.9	63.0	70.6	78.7
Prefusion	0.846 (0.744–0.949)	79.6	85.2	74.1	76.7	80.7
Postfusion	0.874 (0.781–0.966)	81.5	85.2	77.8	79.3	82.1
External test cohort ($n=56$)						
Deep learning	0.764 (0.634–0.894)	75.0	60.7	89.3	85.0	70.8
Radiomics	0.821 (0.699–0.942)	80.4	64.3	96.4	94.7	76.6
Prefusion	0.817 (0.696–0.938)	83.9	71.4	96.4	95.2	81.6
Postfusion	0.828 (0.710–0.946)	80.4	82.1	78.6	79.3	80.7

tools for predicting occult PD in the external test cohort. Furthermore, DCA with a prevalence adjustment set to 5% (based on the reported literature range of 0.9–6.2%) was conducted to accurately reflect the true occurrence of occult PD in the surgical NSCLC population. The findings indicated that the radiomics and postfusion models yielded a higher net benefit in comparison to the 'treat-all' and 'treat-none' strategies, when the threshold probability ranged from 0.025 to 0.075 (Supplementary Figure S8).

Discussion

In the present study, a total of ten radiomics-based ML models and eight DL models were trained using CT images to identify occult PD in NSCLC patients prior to surgery. Besides, the prefusion and postfusion models were developed using feature-based and decision-based methods upon the trained ML and DL models. Our findings demonstrated that both the radiomics model and the postfusion model exhibit strong predictive capability for occult PD. Notably, the postfusion model, which integrates the output probabilities from GBM and DenseNet121 models, demonstrated the highest degree of sensitivity (82.1–97.2%) among all models examined across the three cohorts. These findings highlight the model's particular efficacy as a sensitive diagnostic tool in a clinical setting.

The present study was conducted with several critical methodological considerations. Firstly, patients with pathology-confirmed PI were enrolled in the control group. As reported previously [3, 19], peripheral tumors generally metastasize to the pleural cavity through direct invasion, followed by the shedding of tumor cells and subsequent implantation. Alternatively, metastasis may occur via the subpleural lymphatic network. Consequently, PD predominantly occurs in patients whose primary tumor is adjacent to the pleura or fissures. In addition, PI has been verified as an independent risk factor for the development of PD [11]. Consequently, the comparison of PD + with PI-only controls enables the model to distinguish the subtle imaging signatures associated with metastatic progression beyond local invasion. Secondly, the patients presented with pleural effusion detected preoperatively were excluded. The present study focuses on occult PD, characterized by the absence of preoperative suspicion. In the event that a patient has been diagnosed with pleural effusion prior to surgery, we should consider whether the cancerous pleural effusion has arisen as a consequence of PD. Before surgery, thoracentesis and a subsequent cytological examination should be performed to exclude the possibility of PD, which contradicts our original research objective. In addition, pleural effusion has the potential to compress lung tissue and distort tumor anatomy, thereby introducing substantial confounders and compromising the accuracy of tumor segmentation and feature extraction. Future studies should explore the integration of multimodal models that incorporate pleural effusion imaging characteristics and tumor radiomics to enhance generalizability.

For PD without pleural effusion, typical imaging features include multiple pleural nodules and uneven pleural thickening on CT scans, and FDG uptake in pleural lesions at PET [3–5]. Nevertheless, the detection and diagnosis of occult PD prior to surgery have always been a challenge for radiologists and thoracic surgeons. In general, the disseminated nodules located in the interlobar fissures are relatively easily identifiable due to their surrounding air density of the lungs. Nevertheless, some relatively small nodules attached to the diaphragm can pose a diagnostic challenge on CT scanning due to respiratory interference. Murayama et al. [19] retrospectively reviewed the CT scans of 25 primary lung cancers with

occult PD that were not identified preoperatively, and revealed that 13 patients (52%) exhibited with pleural thickening and/or small pleural nodules. That is, nearly half of the patients did not exhibit any signs of PD on CT images, despite the radiologist's awareness of the condition and a meticulous review of the images. In addition, for some small lesions located on the chest wall or the interlobar fissures, manifesting as irregular pleural surfaces, differentiation from benign conditions, such as reactive pleurisy and intrapulmonary lymph nodes, is often challenging. As demonstrated in the study by Shim et al. [20], 25% (41/164) of the adenocarcinoma patients without PD exhibited nonmalignant pleural or fissural nodules, of which 21 (51.2%) patients demonstrated mild to high activity on PET scans. In contrast, among the eight patients with PD who exhibited pleural or fissural nodules, only two of them (25%) were considered to show positive uptake on PET scans. Consequently, there is an imperative need to develop a novel and accurate approach that can aid radiologists and thoracic surgeons in the preoperative identification of occult PD through CT or other diagnostic modalities.

To date, only a limited number of studies have addressed recognition of occult PD prior to surgery. Li et al. [11] developed a prediction model including 5 risk factors: age, CEA, N stage, adenocarcinoma and PI, which showed a diagnostic efficiency with an AUC of 0.756. However, the above pathological information can only be obtained by surgery, making their model inappropriate for detecting PD prior to surgery. Employing CT images radiomics features, Yang and colleagues [16] conducted a single-center retrospective study and developed a model for predicting occult PD. The developed model exhibited an AUC of 0.93. Nevertheless, they had relatively few cases to develop their model (64 PD-positive cases) and no other data to test the model's ability to generalize. Because cases of incidental intraoperative diagnosis of occult PD are rare, to develop and test our predictive model, we reviewed and collected case data from 3 high-volume hospitals, each with approximately 3,000 lung cancer operations per year. Finally, a total of 216 cases (108 PD-positive and 108 PD-negative) were collected for model training, and 54 internal and 56 external data for testing. It should be emphasized that the present study implemented rigorous inclusion and exclusion criteria. All cases with occult PD were defined by the absence of pleural metastases on preoperative imaging (e.g., CT or PET/CT), or by a misdiagnosis of pleural lesions as benign conditions, such as intrapulmonary lymph nodes and reactive pleurisy. Participants were excluded from the study if they had been diagnosed with pleural metastasis on preoperative imaging and underwent thoracoscopic exploration for pathological confirmation. This methodological decision ensured that the study was centered on the clinical problem of how to address the preoperative inability to definitively diagnose 'occult' PD, rather than simply comparing the differences between NSCLC patients with or without PD (including those who can easily be detected on CT scans).

The integration of radiomics-based ML and DL algorithms has been demonstrated to be capable of harnessing the advantages of both modeling strategies and enhancing the performance of models in disease diagnosis and prognosis [21]. In a multicenter, retrospective, diagnostic study for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging, the decision-based fusion model exhibited the best performance in comparison with DL and radiomics features [22]. In the present study, the Delong test revealed that the performance of the postfusion model was slightly superior to the radiomics, DL and prefusion models, but with no significant difference (p >0.05). It is conceivable that the dataset utilized for training the DL model was insufficient in size, thereby exerting an adverse impact on the model's performance; larger multi-institutional research is needed.

Our study had several limitations. First, selection bias was inevitable due to its retrospective nature (e.g., exclusion of patients without preoperative CT). Secondly, our study primarily included Chinese patients from high-volume centers, potentially limiting generalizability to other populations with distinct epidemiological characteristics. Multicenter international collaborations are needed to validate ethnic robustness. Thirdly, the model's perfect training accuracy contrasted with lower test performance suggests potential overfitting, likely due to the modest training cohort and high feature dimensionality. While data augmentation and cross-validation were employed, larger multicenter datasets are needed to confirm generalizability. Furthermore, while plain CT scans suffice for initial PD detection, incorporating contrast-enhanced CT or PET/CT, particularly for equivocal cases, could improve specificity by leveraging metabolic (FDG uptake) or perfusion features, albeit at increased cost and radiation exposure. Additionally, while SHAP analysis identified high-impact radiomics features, their biological underpinnings require further elucidation. Finally, it is important to acknowledge that the development of our models was conducted using a case-control design, with a 1:1 class ratio, to address the extreme class imbalance inherent in occult PD, a rare incidence rate of 0.9–6.2%. The predicted probabilities derived from these models are not calibrated to the low prevalence of occult PD in the general surgical population. For the purpose of clinical implementation, the model's outputs must undergo calibration to align with this prior probability, thereby converting its discriminative predictions into precise, patient-specific risk estimates.

In summary, the present research developed a prediction model that integrates radiomics-based ML and DL algorithms to predict occult PD for patients with NSCLC. This tool should be used to enhance the assessments conducted by radiologists. Specifically, a patient with radiologically suspected PI but a low prediction score from the model could be considered a strong candidate for direct curative-intent surgery. In contrast, a high prediction score would necessitate a more cautious and multidisciplinary approach, including consideration of diagnostic thoracoscopy for confirmation and a shift towards systemic therapy (e.g., targeted agents or immunotherapy) if PD is verified. This stratification has the potential to reduce futile thoracotomies and ensure that patients receive the most appropriate initial therapy.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-15121-9.

Supplementary Material 1.

Acknowledgements
The authors would like to thank the participating medical centers for providing the data and resources necessary for this study. We also acknowledge the contributions of all researchers and staff involved in the data collection and analysis process. Additionally, gratitude is extended Prof. L.S.H. (Chongqing Fudimai Digital Technology) for his technical support and assistance in code modification.

Financial support and sponsorship
None.

Authors' contributions
Conceptualization, G.W. and B.T.; Methodology & Formal analysis, B.T., B.L. and L.X.G.; Investigation, B.T., L.X.G., D.Y.L., C.L. and S.W.J.; Data Curation, Z.X.L., C.X., and Z.L.; Writing - Original Draft, B.T. and B.L.; Writing - Review & Editing, G.W.; Resources, G.M.J. and D.J.G.; Supervision, G.W. All authors have read and supervised the final manuscript and the process of research.

Funding
None.

Data availability
The codes used to develop the models have been uploaded to the following repository and are publicly accessible with the authorization of the corresponding author: https://doi.org/10.5281/zenodo.17135733.

Declarations

Ethics approval and consent to participate
The study was carried out in accordance with the Declaration of Helsinki and was approved by the ethics committee of Ethics Committee of Daping Hospital (approval number: 2023-299; approval date: December 13, 2023), and all collaborative research centers also obtained permission from their respective Ethics Committees. All patients signed the informed consent form prior to the surgery. The need for individual consent to the utilization of their clinicopathological data for scientific research purposes was waived by the committee.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Thoracic Surgery, Daping Hospital, Army Medical University, Changjiang Route 10#, Daping, Chongqing 400042, China
[2]Department of Radiology, Daping Hospital, Army Medical University, Chongqing, China
[3]Department of Thoracic Surgery, Xinqiao Hospital, Army Medical University, Chongqing, China
[4]Department of Emergency, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[5]Department of Cardiothoracic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[6]Department of Cardiothoracic Surgery, The First Affiliated Hospital of Chongqing Medical and Pharmaceutical College, Chongqing, China
[7]Computer Teaching and Research Office, Army Academy of Artillery and Air Defense, Huangshan Route 451#, Hefei 230031, PR China

Received: 31 May 2025 / Accepted: 28 September 2025
Published online: 29 October 2025

References

1. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. J Thorac Oncol. 2016;11(1):39–51.
2. Ou SH, Zell JA. Validation study of the proposed IASLC staging revisions of the T4 and M non-small cell lung cancer descriptors using data from 23,583 patients in the California cancer registry. J Thorac Oncol. 2008;3(3):216–27.
3. Leung AN, Müller NL, Miller RR. CT in differential diagnosis of diffuse pleural disease. AJR Am J Roentgenol. 1990;154(3):487–92.
4. Okutani D, Yamane M, Toyooka S, Oto T, Aoe M, Sano Y, et al. Dry small pleural dissemination of adenocarcinoma of the lung preoperatively detected by PET/CT: a report of two cases. Acta Med Okayama. 2008;62(1):55–8.
5. Kim YK, Lee HY, Lee KS, Han J, Ahn MJ, Park K, et al. Dry pleural dissemination in non-small cell lung cancer: prognostic and diagnostic implications. Radiology. 2011;260(2):568–74.
6. Li C, Kuo SW, Hsu HH, Lin MW, Chen JS. Lung adenocarcinoma with intraoperatively diagnosed pleural seeding: Is main tumor resection beneficial for prognosis? J Thorac Cardiovasc Surg. 2018;155(3):1238-1249.e1.
7. Ren YJ, She YL, Dai CY, Jiang GN, Fei K, Chen C. Primary tumor resection showed survival benefits for non-small-cell lung cancers with unexpected malignant pleural dissemination. Interact Cardiovasc Thorac Surg. 2016;22(3):321–6.
8. Okamoto T, Iwata T, Mizobuchi T, Hoshino H, Moriya Y, Yoshida S, et al. Pulmonary resection for lung cancer with malignant pleural disease first detected at thoracotomy. Eur J Cardiothorac Surg. 2012;41(1):25–30.
9. Yalong WANG, Yage WANG, Xiao HU, Yonggang WANG. Prognosis of 153 non-small cell lung cancer patients with pleural dissemination after surgery. Chin J Clin Oncol. 2017;44(14):712–6.
10. Bao T, Deng Y, Chen L, Sun W, Ge M, Zhao X, et al. Primary tumour resection in non-small cell lung cancer patients with pleural dissemination unexpectedly detected during operation: a two-centre retrospective cohort study. BMC Cancer. 2025;25(1):316.
11. Li S, Zhang S, Huang M, Ma Y, Yang Y. Management of occult malignant pleural disease firstly detected at thoracotomy for non-small cell lung cancer patients. J Thorac Dis. 2017;9(10):3851–8.
12. Mukherjee S, Patra A, Khasawneh H, Korfiatis P, Rajamohan N, Suman G, et al. Radiomics-based Machine-learning models can detect pancreatic cancer on prediagnostic computed tomography scans at a substantial lead time before clinical diagnosis. Gastroenterology. 2022;163(5):1435–1446.e3.
13. Gitto S, Annovazzi A, Nulle K, Interlenghi M, Salvatore C, Anelli V, et al. X-rays radiomics-based machine learning classification of atypical cartilaginous tumour and high-grade chondrosarcoma of long bones. EBioMedicine. 2024;101:105018.
14. She Y, He B, Wang F, Zhong Y, Wang T, Liu Z, et al. Deep learning for predicting major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer: A multicentre study. EBioMedicine. 2022;86:104364.

15. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. Lancet Digit Health. 2021;3(4):e250–9.
16. Yang M, Ren Y, She Y, Xie D, Sun X, Shi J, et al. Imaging phenotype using radiomics to predict dry pleural dissemination in non-small cell lung cancer. Ann Transl Med. 2019;7(12):259.
17. Travis WD, Brambilla E, Rami-Porta R, Vallières E, Tsuboi M, Rusch V, et al. Visceral pleural invasion: pathologic criteria and use of elastic stains: proposal for the 7th edition of the TNM classification for lung cancer. J Thorac Oncol. 2008;3(12):1384–90.
18. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for High-Throughput image-based phenotyping. Radiology. 2020;295(2):328–38.
19. Murayama S, Murakami J, Yoshimitsu K, Torii Y, Ishida T, Masuda K. CT diagnosis of pleural dissemination without pleural effusion in primary lung cancer. Radiat Med. 1996;14(3):117–9.
20. Shim SS, Lee KS, Kim BT, Choi JY, Shim YM, Chung MJ, et al. Integrated PET/CT and the dry pleural dissemination of peripheral adenocarcinoma of the lung: diagnostic implications. J Comput Assist Tomogr. 2006;30(1):70–6.
21. Demircioğlu A. Are deep models in radiomics performing better than generic models? A systematic review. Eur Radiol Exp. 2023;7(1):11.
22. Wang W, Liang H, Zhang Z, Xu C, Wei D, Li W, et al. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. EClinicalMedicine. 2024;67:102385.

Publisher's Note