BMC Medical Imaging

RESEARCH

Open Access

Combining computed tomography radiomics and clinical features to predict lymph node metastasis in patients with lung cancer

Peiqi Wang[1,2†], Hao Hu[3,4†], Yubo Wang[3,4†], Yadan Yin[3,4†], Yang Fu[3†], Bosen Xie[3], Jiageng Li[3], Mengxue Kong[3], Chunyuan Wei[3], Lei Yue[5], Duiming Yang[6*] and Bin Yang[3,4*]

Abstract

Background Accurate preoperative assessment of lymph node metastasis (LNM) is crucial for treatment planning and prognostic stratification in patients with lung cancer. This study aimed to develop and validate a predictive model for LNM using radiomic features derived from non-contrast computed tomography (CT) combined with clinical characteristics.

Methods A total of 403 patients with pathologically confirmed lung cancer were retrospectively enrolled and randomly divided into a training set ($n=282$) and an internal test set ($n=121$). In addition,30 lung cancer patients from other hospital were collected as an external test set. Clinical variables were collected, and radiomic features were extracted from non-contrast chest CT images using the Radiomics module of 3D Slicer. Feature selection was performed using least absolute shrinkage and selection operator (LASSO) regression. Multiple machine-learning models were constructed based on radiomic features and clinical features. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), and clinical utility was evaluated by decision curve analysis (DCA). Shapley additive explanations (SHAP) were applied to enhance model interpretability.

Results Lymph node metastasis was observed in 35.5% (143/403) of patients. 2 clinical features and 16 radiomic features most strongly associated with LNM were identified. Among the nine constructed models, the combined clinical–radiomic support vector machine (SVM) model demonstrated the best predictive performance, with AUCs of 0.927 in the training set, 0.852 in the internal test set, and 0.812 in the external test set. Decision curve analysis indicated that the combined model provided a favorable net clinical benefit across a wide range of threshold probabilities.

†Peiqi Wang, Hao Hu, Yubo Wang, Yadan Yin and Yang Fu contributed equally to this work.

*Correspondence:
Duiming Yang
17347083@qq.com
Bin Yang
yangbinapple@163.com

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

Conclusion The proposed clinical–radiomic model based on non-contrast CT achieved good performance in predicting lymph node metastasis in patients with lung cancer and may serve as a noninvasive tool to assist individualized clinical decision-making.

Keywords Lung cancer, Lymph node metastasis, Computed tomography, Radiomics

Background

Lung cancer remains the most frequently diagnosed malignancy and the leading cause of cancer-related mortality worldwide. The latest global cancer statistics estimate that in 2022, there were approximately 2.48 million new cases of lung cancer and 1.817 million associated deaths, accounting for 12.4% and 18.7% of all cancer cases and deaths, respectively, highlighting its substantial global burden [1]. Lung cancer is characterized by uncontrolled cell proliferation and an early tendency for metastasis, particularly to regional lymph nodes, which significantly influences prognosis and therapeutic decision-making [2].

Accurate preoperative staging is crucial for the optimal management of lung cancer. In the current Tumor-Node-Metastasis (TNM) classification system proposed by the American Joint Committee on Cancer (AJCC), lymph node involvement (N stage) is a key determinant of stage grouping and directly influences treatment planning, including the choice between surgical resection and combination therapy [3, 4]. For patients without lymph node metastasis (LNM), surgical resection (e.g., lobectomy) can preserve pulmonary function and yield a long survival time, with a reported 5-year survival rate of over 70% [4, 5]. Conversely, patients with confirmed LNM often require neoadjuvant or adjuvant chemotherapy and radiotherapy, yet their 5-year survival rates remain suboptimal (26–53%) [6–8]. These discrepancies confirm the importance of improving the accuracy of preoperative lymph node assessment.

The current gold standard for the preoperative diagnosis of LNM is mediastinoscopy and endobronchial ultrasound-guided transbronchial needle aspiration. However, because these procedures are invasive, they are not routinely recommended [9]. Computed tomography (CT) is the most commonly used noninvasive diagnostic tool for lung cancer, providing information on tumor location, size, and metastatic spread [10]. Typically, radiologists use a short-axis diameter > 10 mm in mediastinal lymph nodes as the criterion for metastasis, but some lymph nodes may enlarge owing to inflammatory changes, limiting diagnostic accuracy [11, 12]. In recent years, ^{18}F-FDG positron emission tomography (^{18}F-FDG PET) has been widely used in the preoperative evaluation of patients with lung cancer [13]. It provides both metabolic and anatomical information, with a sensitivity of 77% and specificity of 86% for mediastinal lymph node staging [14]. Nonetheless, some researchers have reported that ^{18}F-FDG PET still has a high false-positive rate (5–21%) for detecting malignancy in normal-sized lymph nodes and for excluding malignancy in patients with concurrent inflammatory conditions [15–17]. Moreover, its high cost, limited availability, and additional radiation exposure hinder its widespread use in patients with lung cancer. Therefore, a major clinical challenge remains: how to develop a reliable, noninvasive, and accessible method to improve preoperative prediction of lymph node metastasis in lung cancer.

Radiomics has emerged as a promising approach that extracts high-dimensional quantitative features from standard imaging modalities to characterize tumor heterogeneity and microenvironment beyond human visual interpretation. A recent meta-analysis has shown that CT-based radiomic models achieve a sensitivity of approximately 0.84 and a specificity of 0.82, with an area under the receiver operating characteristic curve (AUC) of 0.90 for predicting LNM in non-small cell lung cancer (NSCLC), indicating that radiomics can improve the accuracy of lymph node staging [18]. Additionally, systematic reviews have indicated that despite variations in the algorithms and the quality of external validation of existing models, various artificial intelligence models have exhibited promising application prospects in the diagnosis and treatment of LNM in patients with lung cancer [19]. In addition, advanced machine learning and deep learning radiomic models based on CT/PET-CT have demonstrated high predictive efficacy, integrated imaging and clinical variables, and improved interpretability through methods such as Shapley Additive Explanations (SHAP). [20, 21]. However, most existing studies rely on contrast-enhanced imaging or PET-CT, which may limit their applicability to patients who cannot receive contrast agents or undergo PET-CT examinations.

In this study, we aimed to develop and validate a machine learning–based radiomics model using non-contrast CT combined with readily available clinical features to predict lymph node metastasis in patients with lung cancer. Compared with previous work, our approach emphasizes a widely accessible imaging modality, integrates clinical and radiomic features, applies SHAP for model interpretability, and includes external validation to assess generalizability. By addressing these limitations, this model may provide a noninvasive and practical tool to support individualized clinical decision-making in preoperative staging.

Methods

Study design

This study aimed to develop a machine learning model that integrates common clinical characteristics and CT-based radiomics features of tumors to predict LNM in patients with lung cancer. Clinical information was collected, and radiomics features were extracted from non-contrast chest CT images. Least absolute shrinkage and selection operator regression analysis was employed to reduce the number of variables and establish a prediction model for lymph node metastasis in patients with lung cancer based on non-contrast CT scans. The predictive performance and clinical utility of the model were evaluated using AUC and decision curve analysis, and Shapley additive explanations analysis was applied to enhance interpretability (Fig. 1).

Participants

We retrospectively collected data from 403 patients with lung cancer treated at Panzhihua Traditional Chinese Medicine and Western Medicine Hospital between March 2020 and March 2025. A total of 270 patients underwent systematic lymph node dissection, and 133 patients underwent lymph node sampling. Among them, 143 had mediastinal lymph node metastases, whereas 260 had no lymph node metastases. Among the 403 patients with malignant tumors, the pathological subtypes were as follows: 369 cases of adenocarcinoma, 23 cases of squamous cell carcinoma, 7 cases of small cell lung cancer, and three cases not otherwise specified. The inclusion criteria were as follows: ① age > 18 years; ② presence of a single pulmonary nodular lesion; ③ confirmation of lung cancer via paraffin-embedded pathological section; ④ complete case data. The exclusion criteria were as follows: ① concomitant malignant tumors; ② concomitant diffuse lung diseases, such as pulmonary tuberculosis or silicosis; ③ history of preoperative neoadjuvant radiotherapy/chemotherapy or molecular targeted therapy; ④ history of previous malignant tumors. The patients were randomly divided into development ($n = 282$) and test ($n = 121$) sets at a ratio of 7:3 (Fig. 2). Basic information, laboratory data, and pathological data were collected from the patients' electronic medical records. Basic information included age, sex, smoking history, and family history. Laboratory parameters included alpha-fetoprotein, carcinoembryonic antigen (CEA), carbohydrate antigen CA125, carbohydrate antigen CA199, neuron-specific enolase, and squamous cell carcinoma-related antigen. Pathological examinations included tumor type and immunohistochemical indicators. In addition, we also collected and analyzed the information on the location of the patients' tumors and the patients' TNM stage(The 9th edition of the American Joint Committee on Cancer(AJCC) TNM staging system). In addition, data from 30 patients at Panzhihua Central Hospital were retrospectively collected as an external test set.

Analysis of clinicopathological factors and prediction model construction

For the statistical analysis of clinical data, categorical variables were described by frequency and constituent ratio. The description and comparison methods of continuous variables were determined by the normality of data distribution, which was tested using the Kolmogorov-Smirnov test. Normally distributed variables were expressed as mean ± standard deviation (±s) and compared using the independent-samples t-test; non-normally distributed variables were expressed as median [interquartile range] (Median [IQR]) and compared using the Mann-Whitney U test. Model performance was evaluated by the area under the curve (AUC) with a 95% confidence interval (95%CI). A two-sided $p < 0.05$ was considered statistically significant. All statistical analyses were performed using R software (Version 4.3.0; https://www.r-project.org) and Python (Version 3.10.0). Factors with a stronger association with LNM status ($p < 0.05$) were selected for multivariate analysis (Table 1).

CT image acquisition and feature extraction

All patients were placed in the supine position and scanned from the thoracic inlet to the diaphragmatic base. Scanning parameters were as follows: scanning slice thickness: 5.0 mm, tube voltage: 120 kVp, and tube current: 80–300 mAs,reconstruction kernel:B70f, reconstruct slice thickness: 1.0 mm. All images were displayed using the standard lung (window width: 1200 HU; window level: −600 HU).

All images were saved on a hard disk in DICOM format and imported into 3D Slicer software (version 5.8.1). We first standardized the voxel size via interpolation prior to feature extraction.A radiologist with 8 years of experience, Peiqi Wang, manually delineated the region of interest (ROI) of the tumor on a lung window. The entire tumor lesion area was outlined layer by layer. Once the delineation was completed, a radiology attending physician with 10 years of experience, Yang Fu, reviewed and corrected the delineated ROI, to ensure the Intraclass Correlation Coefficient(ICC) of radiomic features is ≥0.85. In total, 851 radiomics features were extracted using the SlicerRadiomics module, including: 1. 14 shape features describing tumor morphology and geometric structure, such as edge detection and contour description; 2. 93 texture features describing image texture information, such as the gray-level co-occurrence matrix, gray-level difference matrix, and gray-level run length matrix; 3. 744 frequency-domain features describing tumor characteristics in the frequency domain, such

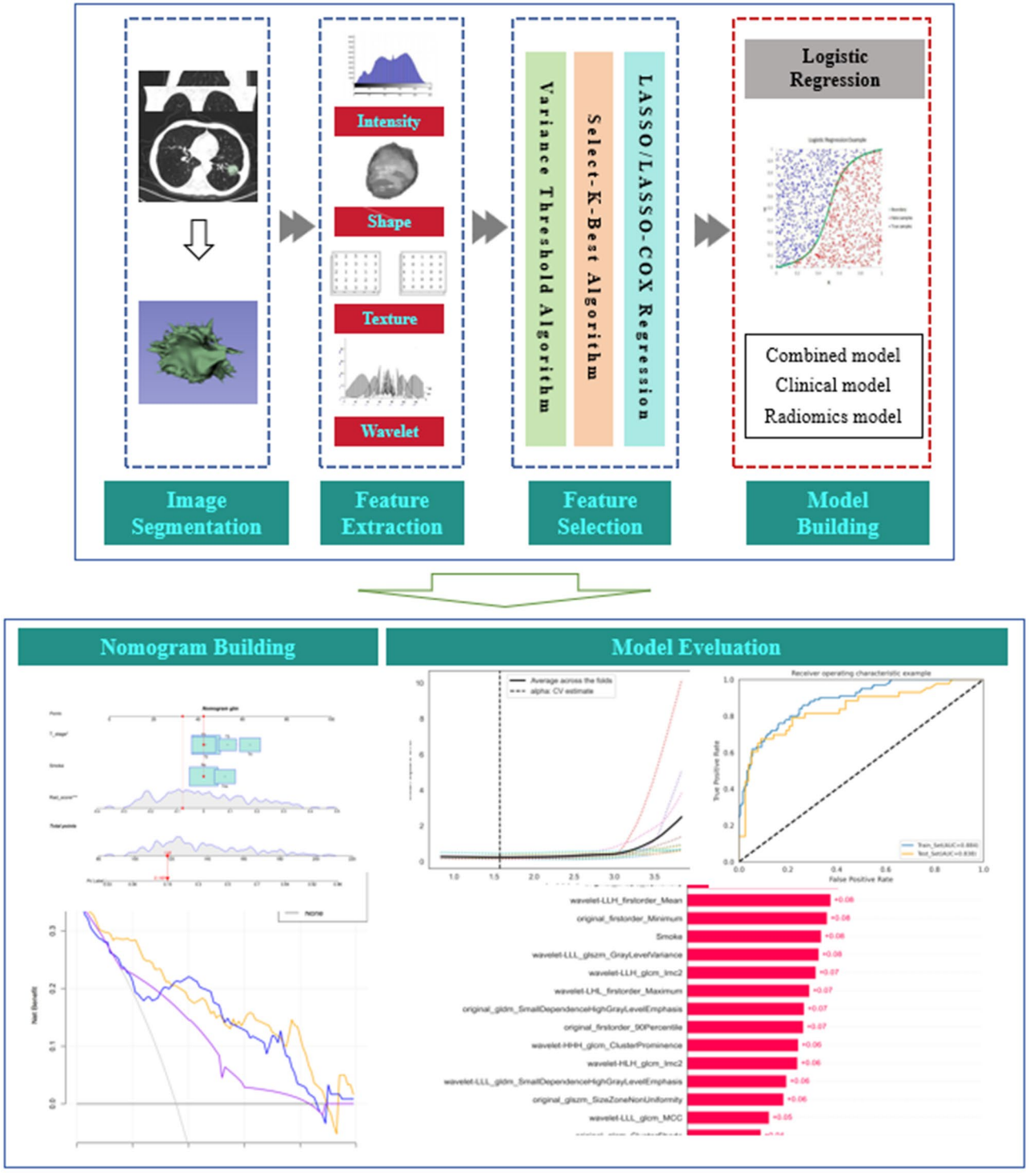

Fig. 1 Overall flowchart

as Fourier transform coefficients and wavelet transform coefficients.

Selection of radiomics features

Prior to feature selection, all radiomics features were standardized using Z-score normalization. The standardization of Z-score parameters (mean and standard deviation) was performed exclusively on the training folds. Owing to the large number of radiomics features, three methods were employed for dimensionality reduction to avoid issues, such as model overfitting and multicollinearity. First, variance thresholding was

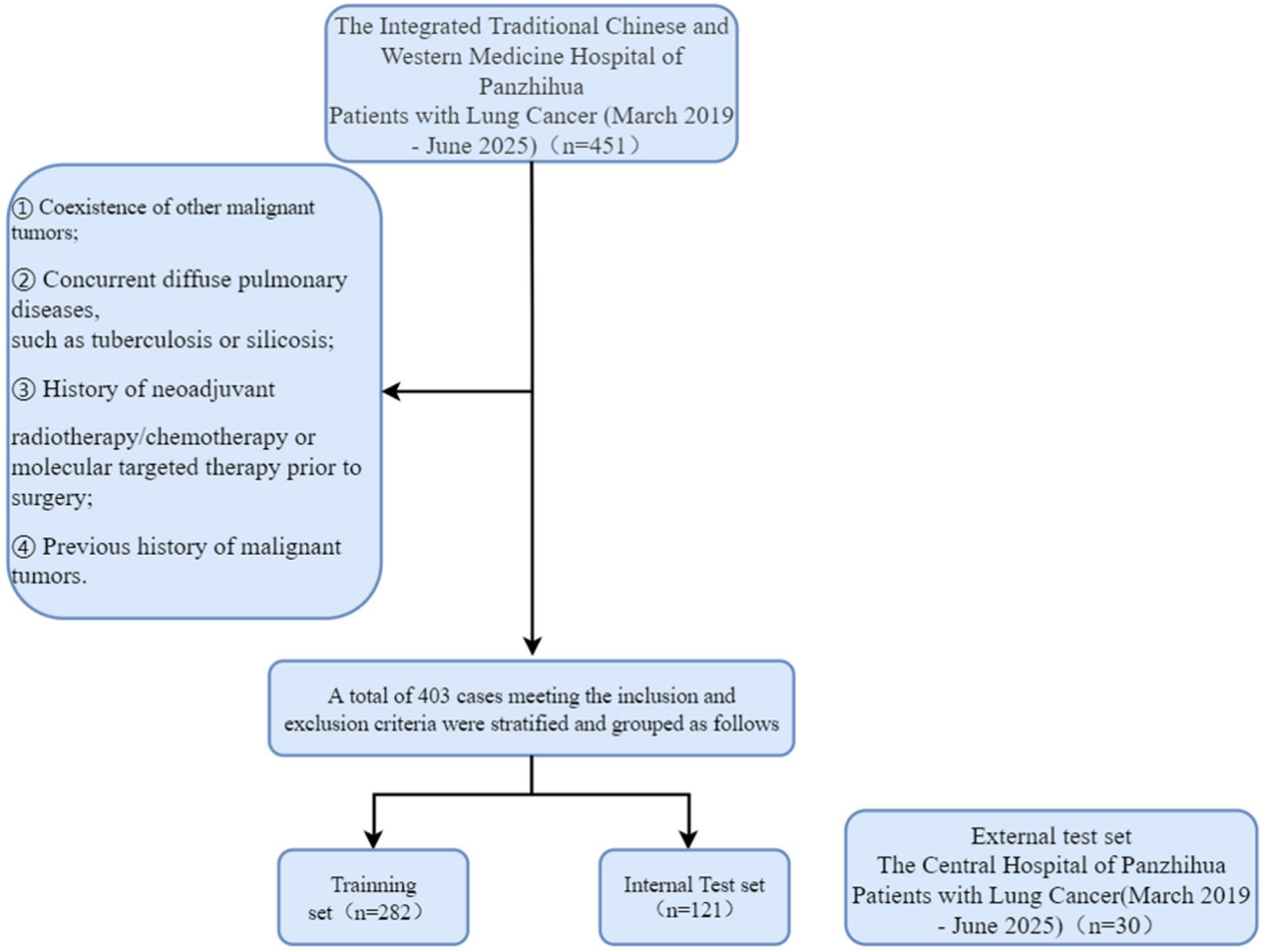

Fig. 2 Patient inclusion and exclusion criteria and enrollment flowchart

used to remove features with a variance < 0.8. Second, univariate selection was applied to filter out non-significant features ($p > 0.05$). Finally, the least absolute shrinkage and selection operator regression algorithm was used to sparsify the feature set, select indicators most relevant to the research objective, and obtain their weights. Key radiomics features were selected (Fig. 3). This study conducts a correlation analysis between the screened imaging features and T staging, excludes feature variables with a strong correlation with T staging (|r| >0.7), and finally retains features with high independence for subsequent modeling and analysis. Ultimately, 16 radiomics features were retained (Table 2), including: wavelet-LLH_glcm_Imc2, wavelet-HHH_glcm_ClusterProminence, original_shape_Sphericity, wavelet-LLL_glcm_MCC, wavelet-HLH_glcm_Imc2, original_glcm_ClusterShade, original_gldm_SmallDependenceHighGrayLevelEmphasis, original_glszm_ZoneEntropy, wavelet-LLL_gldm_SmallDependenceHighGrayLevelEmphasis, original_firstorder_90%ile, original_glszm_SizeZoneNonUniformity, wavelet-LHL_firstorder_Maximum, wavelet-LLH_firstorder_Mean, wavelet-LLL_glszm_GrayLevelVariance, and original_firstorder_Minimum. After dimensionality reduction, the radiomics score (RadScore) was calculated for each patient using the following formula. The RadScore, which comprehensively represents radiomic features, was incorporated into model construction.

The formula for calculating the RadScore is:

RadScore = wavelet-LLH_glcm_Imc2x（-0.025579894）+ wavelet-HHH_glcm_ClusterProminencex（-0.019328046）+ original_shape_Sphericityx(−0.013027585)+ wavelet-LLL_glcm_MCCx(−0.010390812)+ wavelet-HLH_glcm_Imc2x(−0.007564503)+ original_glcm_ClusterShadex(−0.007167992)+ original_gldm_SmallDependence HighGrayLevelEmphasisx(−0.00049788)+ original_glszm_ZoneEntropyx(0.0011592808893695907)+ wavelet-LLL_gldm_SmallDependenceHighGrayLevelEmphasisx(0.0027629811348519767)+ original_firstorder_90%ilex(0.009111013539256278)+

Table 1 Group statistics of clinical data in the test set and internal test set

Clinical pathological characteristics	Classification	Train	Internal Test	p
n		282	121	
Gender (%)	male	157 (55.7)	54 (44.6)	0.054
	female	125 (44.3)	67 (55.4)	
Age (median [IQR])		66 [55, 74]	68 [54, 74]	0.699
Smoke (%)	no	180 (63.8)	88 (72.7)	0.105
	yes	102 (36.2)	33 (27.3)	
Family_history (%)	no	257 (91.1)	110 (90.9)	1.000
	yes	25 (8.9)	11 (9.1)	
T_Stage (%)	T1	76 (27.0)	34 (28.1)	0.314
	T2	107 (37.9)	48 (39.7)	
	T3	42 (14.9)	10 (8.3)	
	T4	57 (20.2)	29 (24.0)	
M_stage (%)	M0	169 (59.9)	73 (60.3)	0.307
	M1	113 (40.1)	47 (38.8)	
	M2	0 (0.0)	1 (0.8)	
Clinical_stage (%)	IA	59 (20.9)	27 (22.3)	0.489
	IB	112 (39.7)	39 (32.2)	
	IIA	14 (5.0)	8 (6.6)	
	IIB	7 (2.5)	5 (4.1)	
	IIIA	1 (0.4)	0 (0.0)	
	IIIB	0 (0.0)	1 (0.8)	
	IV	89 (31.6)	41 (33.9)	
CEA (median [IQR])		2.90 [2.10, 4.10]	2.90 [2.30, 4.60]	0.765
Pathological_type (%)	adenocarcinoma	258 (91.5)	112 (92.6)	0.238
	squamous cell carcinom	19 (6.7)	4 (3.3)	
	small cell lung cancer	3 (1.1)	4 (3.3)	
	other	2 (0.7)	1 (0.8)	

original_glszm_SizeZoneNonUniformityx(0.03503650252922063)+ wavelet-LHL_firstorder_Maximumx(0.03734605667166642)+ wavelet-LLH_firstorder_Meanx(0.03901754470939138)+ wavelet-LLL_glszm_GrayLevelVariancex(0.05842791980640918)+ original_firstorder_Minimumx(0.0622995314014254).

Model construction

To predict LNM, filtered clinical and radiomics features were used as inputs to construct a radiomics model for the differential diagnosis of LNM. All data were randomly divided into development and internal test sets (split ratio: 7:3). The training set was used for model training, whereas the internal and external test sets were used to assess the generalization ability of the model. With roughly 35% positive lymph-node cases, the data set is class-imbalanced. We adopted two methods during data analysis to alleviate this issue: (1) The SMOTE method for oversampling the minority class; (2) Algorithm hyperparameter setting: class_weight='balanced'. Experimental results showed that setting the class_weight parameter to 'balanced' (which adjusts the model's attention to the minority or majority class by assigning different weights to different classes) achieved better performance（In the clinical–radiomic SVM model, the class_weight='balanced' method yielded an AUC that was 0.23 higher in the internal test set and 0.15 higher in the external test set compared with the SMOTE method）. Therefore, we chose this method to mitigate the class imbalance of the data.Logistic regression(LR), support vector machine (SVM), and random forest (RF) algorithms were employed to construct models for the differential diagnosis of LNM. For the three algorithms, we have conducted systematic hyperparameter tuning via GridSearch combined with internal 5-fold cross-validation. The search range for logistic regression includes penalty types (L1/L2), C (0.001, 0.01, 0.1, 1, 10, 100), and class weights (balanced/None); the search parameters for SVM cover kernel types ('rbf', 'linear', 'poly'), gamma values ('auto', 0.01, 0.1, 1, 10), and C (0.01, 0.1, 1, 10, 100); for random forest, we explored the number of n_estimators (10, 50, 100, 200, 300, 500), max_depth (5, 10, 20, 50), min_samples_split (2, 5, 10), and max_features ('sqrt', 'log2', None).Receiver operating characteristic curves were plotted, and AUC was calculated to evaluate accuracy, precision, sensitivity, and specificity.

Construction of the radiomics nomogram

A radiomics nomogram was constructed by combining RadScores and significant clinicopathological factors in a multivariate logistic regression model. Receiver operating characteristic curves were plotted and AUC values were calculated to evaluate accuracy, precision, sensitivity, and specificity. Additionally, a calibration curve was plotted to assess calibration and discrimination. Decision curve analysis was performed to quantify the net benefit of the nomogram and evaluate its clinical utility.

Results

Clinical characteristics

The demographic and clinicopathological characteristics of patients in the training and internal test sets are detailed in Table 1. The training set comprised 282 patients (median age: 66 years; interquartile range: 55–74 years; 157 men). Univariate and multivariate analyses revealed that LNM status was significantly associated with T stage and smoking history. However, no statistically significant differences were observed between patients with and without metastasis in the training set regarding sex, age, family history, metastasis

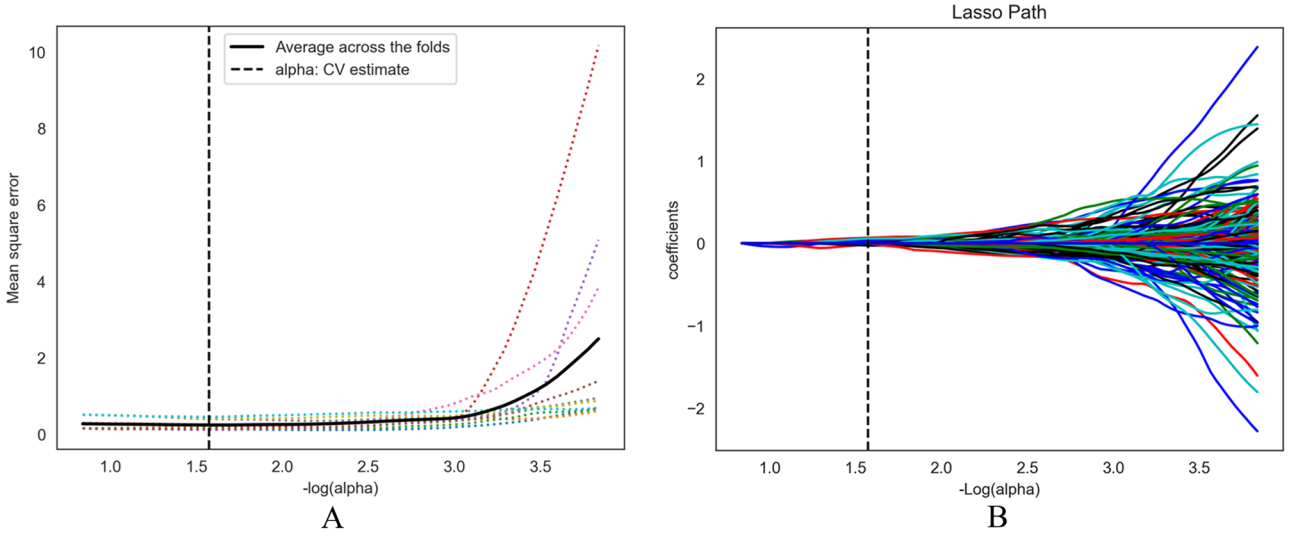

Fig. 3 Radiomic features selected using the LASSO regression model (**A**) shows the mean squared error of the radiomic features displayed by the lasso regression analysis in the model. The two vertical lines indicate the lambda values when the number of variables is reduced to the two lowest levels. (**B**) shows the trajectory of each radiomic feature, with the horizontal axis representing the log lambda for each radiomic feature and the vertical axis representing the coefficient of each radiomic feature

Table 2 LASSO feature selection table

Features	Coefficients
wavelet-LLH_glcm_Imc2	−0.025579894
wavelet-HHH_glcm_ClusterProminence	−0.019328046
original_shape_Sphericity	−0.013027585
wavelet-LLL_glcm_MCC	−0.010390812
wavelet-HLH_glcm_Imc2	−0.007564503
original_glcm_ClusterShade	−0.007167992
original_gldm_SmallDependenceHighGrayLevelEmphasis	−0.00049788
original_glszm_ZoneEntropy	0.0011592808893695907
wavelet-LLL_gldm_SmallDependenceHighGrayLevelEmphasis	0.0027629811348519767
original_firstorder_90%ile	0.009111013539256278
original_glszm_SizeZoneNonUniformity	0.03503650252922063
wavelet-LHL_firstorder_Maximum	0.03734605667166642
wavelet-LLH_firstorder_Mean	0.03901754470939138
wavelet-LLL_glszm_GrayLevelVariance	0.05842791980640918
original_firstorder_Minimum	0.0622995314014254

stage, clinical stage,tumor location or CEA levels (all $p > 0.05$, Table 3).

Model construction and predictive performance

Three machine learning algorithms—logistic regression, SVM, and RF—were selected to construct prediction models. Models were built using clinical features, radiomic features, and a combination of both (clinical–radiomics features). In the training set, the SVM model with combined clinical–radiomics features demonstrated the best performance (AUC = 0.927). The model based solely on radiomics features also performed well (AUC = 0.921), approaching the performance of the combined model. The model based solely on clinical features exhibited low performance (AUC = 0.7).

In the internal test set, the SVM model with combined clinical–radiomics features achieved the best performance (AUC = 0.852), whereas the RF model based solely

Table 3 Univariate and multivariate analyses

Clinical pathological characteristics	Univariate analysis				Multivariate analysis			
	OR	CI.lower	CI.upper	P value	OR	CI.lower	CI.upper	P value
Gender	0.614	0.404	0.926	0.021	0.982	0.568	1.712	0.948
Age	1.027	1.01	1.045	0.002	1.018	0.999	1.037	0.065
Smoke	1.961	1.279	3.01	0.002	1.959	1.115	3.481	**0.02**
Family_history	0.901	0.423	1.827	0.778				
T_stage	1.883	1.545	2.311	<0.001	1.726	1.365	2.197	**<0.001**
M_stage	2.255	1.494	3.42	<0.001	1.543	0.811	2.897	0.18
Clinical_stage	1.182	1.103	1.268	<0.001	1.025	0.911	1.154	0.679
CEA	0.994	0.985	0.999	0.092				
Pathological_type	0.614	0.404	0.926	0.021				
Tumor location	1.136	1.023	1.236	0.017	1.116	0.997	1.251	0.058

on radiomics features achieved an AUC of 0.844. In the external test set, the RF model with combined clinical–radiomics features performed best (AUC = 0.833) (Fig. 4, Table 4).

The Delong test was performed on the AUC of the internal test set. The test results showed that there were significant differences between the following model pairs: Clinical_LR model vs. Combined_LR model, Radiomic_SVM model vs. Clinical_SVM model, Clinical_SVM

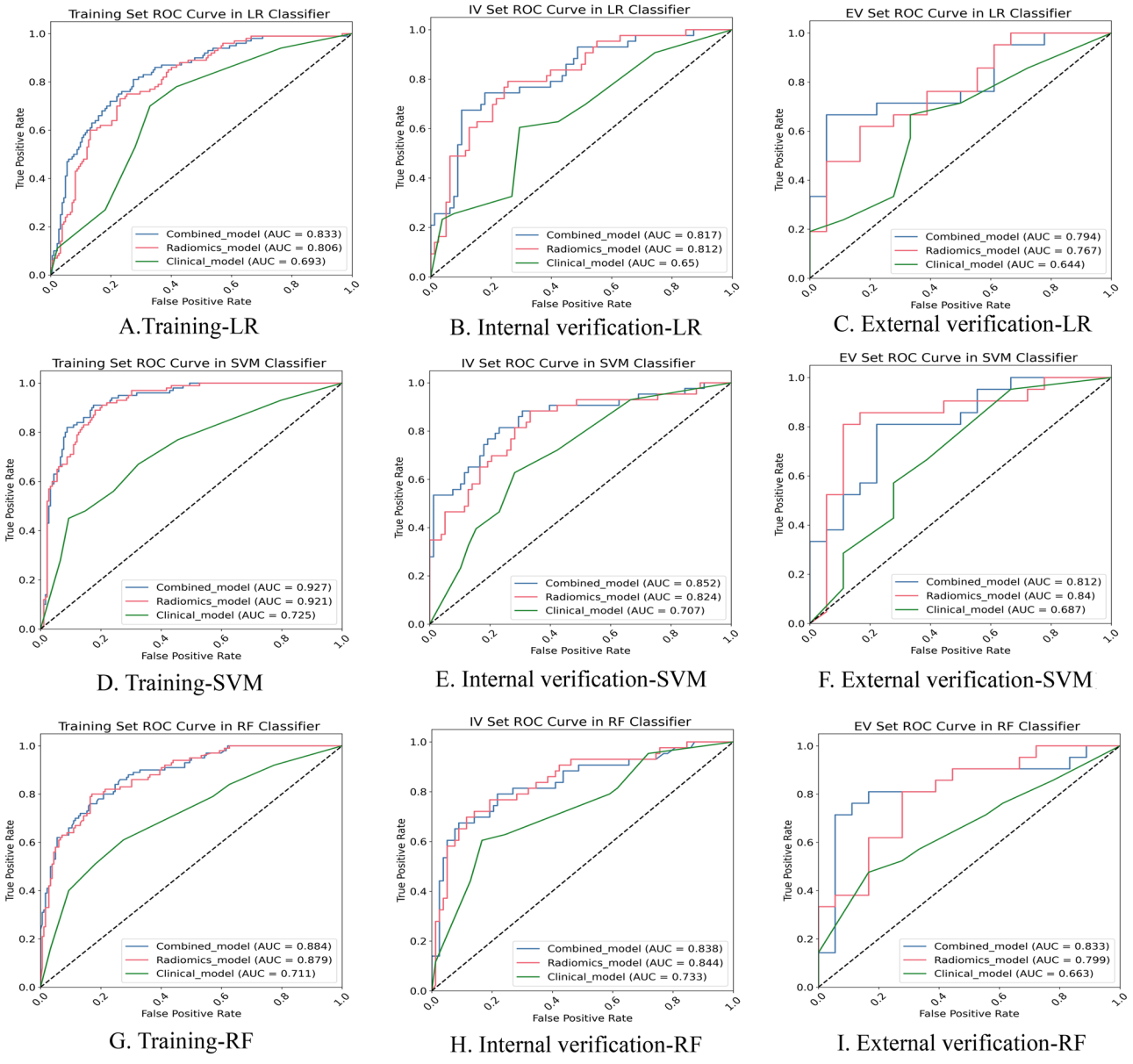

Fig. 4 Receiver operating characteristic (ROC) curves of the models. **A**. Development set: The LR model with clinical features alone (AUC = 0.693), the LR model with radiomic features alone (AUC = 0.806), and the LR model with combined features (AUC = 0.833). **B**. Internal test set: The LR model with clinical features alone (AUC = 0.650), the LR model with radiomic features alone (AUC = 0.812), and the LR model with combined features (AUC = 0.817). **C**. External test set: The LR model with clinical features alone (AUC = 0.644), the LR model with radiomic features alone (AUC = 0.767), and the LR model with combined features (AUC = 0.794). **D**. Development set: The SVM model with clinical features alone (AUC = 0.725), the SVM model with radiomic features alone (AUC = 0.921), and the SVM model with combined features (AUC = 0.927). **E**. Internal test set: The SVM model with clinical features alone (AUC = 0.707), the SVM model with radiomic features alone (AUC = 0.824), and the SVM model with combined features (AUC = 0.852). **F**. External test set: The SVM model with clinical features alone (AUC = 0.687), the SVM model with radiomic features alone (AUC = 0.840), and the SVM model with combined features (AUC = 0.812). **G**. Development set: The RF model with clinical features alone (AUC = 0.711), the RF model with radiomic features alone (AUC = 0.879), and the RF model with combined features (AUC = 0.884). **H**. Internal test set: The RF model with clinical features alone (AUC = 0.733), the RF model with radiomic features alone (AUC = 0.844), and the RF model with combined features (AUC = 0.838). **I**. External test set: The RF model with clinical features alone (AUC = 0.663), the RF model with radiomic features alone (AUC = 0.799), and the RF model with combined features (AUC = 0.833).

Table 4 Results of the models

Model	Classifier	wb_auc	wb_acc	wb_sen	wb_spe
Com	LR	0.794(0.897–0.659)	0.795	0.667	0.944
Com	RF	0.833(0.935–0.709)	0.821	0.714	0.944
Com	SVM	0.812(0.92–0.686)	0.795	0.81	0.778
RS	LR	0.767(0.874–0.621)	0.718	0.619	0.833
RS	SVM	0.84(0.946–0.719)	0.846	0.81	0.889
RS	RF	0.799(0.911–0.672)	0.769	0.81	0.722
Clic	LR	0.644(0.802–0.508)	0.667	0.667	0.667
Clic	RF	0.663(0.814–0.549)	0.641	0.476	0.833
Clic	SVM	0.687(0.809–0.518)	0.641	0.571	0.722

Table 5 Contrasted results of all models

Molde Comparation	*P*value in internal test set
Radiomic_LR model vs Clinical_LR model	0.21665
Radiomic_LR model vs Combined_LR model	0.17350
Clinical_LR model vs Combined_LR model	0.04469
Radiomic_SVM model vs Clinical_SVM model	0.03642
Radiomic_SVM model vs Combined_SVM model	0.89068
Clinical_SVM model vs Combined_SVM model	0.01418
Radiomic_RF model vs Clinical_RF model	0.02576
Radiomic_RF model vs Combined_RF model	0.76846
Clinical_RF model vs Combined_RF model	0.01432

model vs. Combined_SVM model, Radiomic_RF model vs. Clinical_RF model, and Clinical_RF model vs. Combined_RF model.(Table 5) Additionally, the Hosmer-Lemeshow test result ($p = 0.398$) is reported: a p-value > 0.05 indicates no statistically significant discrepancy between predicted and observed outcomes. As shown, the bias-corrected calibration curve closely aligns with the ideal curve, and the Hosmer-Lemeshow test result further confirms that the nomogram has satisfactory calibration—meaning its predicted probabilities are well-matched to the actual occurrence of the target outcome.

The clinical utility of the models was assessed using decision curve analysis curves. Within a reasonable threshold probability range, the net benefit of the combined model was higher than that of the clinical-only and radiomics-only models (Fig. 5).

Model interpretability analysis

SHapley Additive exPlanations (SHAP) plots were generated to visualize the impact of different features on individual predictions. Tumor stage was identified as the most critical variable, followed by radiomic features reflecting tumor gray-level distribution and shape and smoking history. All features exhibited positive SHAP values, indicating an overall tendency to increase the predicted probability of LNM. This demonstrates the synergistic effect of clinical and radiomic features, providing a basis for understanding model decisions, investigating metastatic mechanisms, and optimizing diagnostic and treatment strategies (Fig. 6, 7).

Nomograms

A nomogram was constructed to present the model visually (Fig. 8). The nomogram is based on the Combined-LR model, and it indeed represents a more interpretable but slightly less performant model, facilitating clinical understanding and popularization.This nomogram enables individualized prediction of lung cancer lymph node

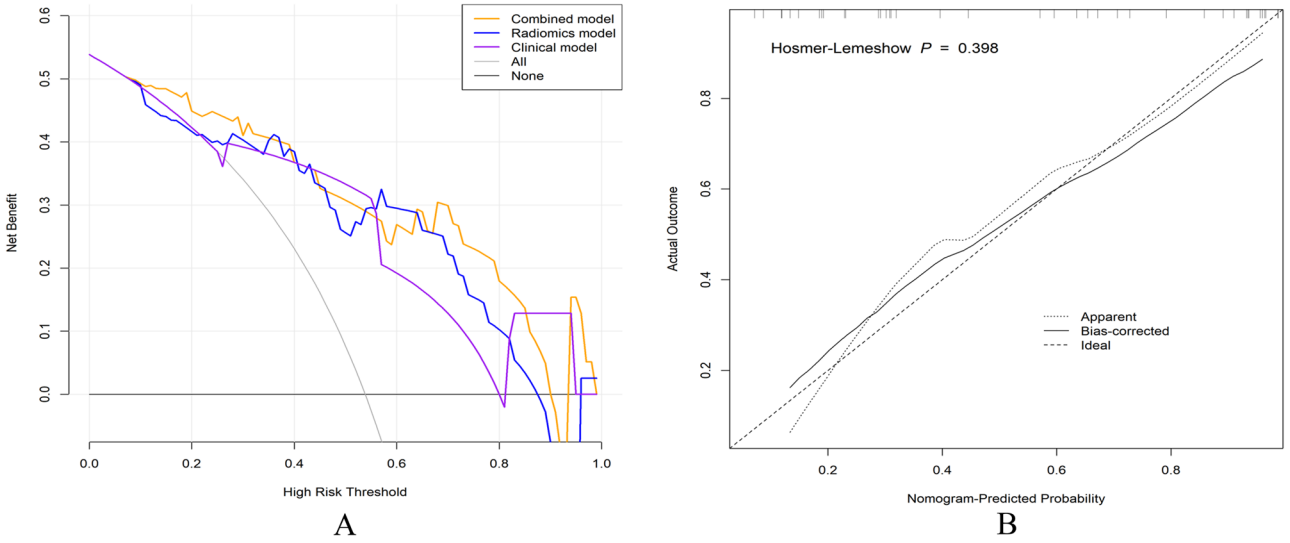

Fig. 5 Decision curve analysis (DCA). **A**. DCA curves for the clinical feature model, radiomic feature model, and clinical-radiomic combined feature model. Within the threshold probability range of 0.2–0.8, the combined model yielded a higher net benefit. **B**. The Hosmer-Lemeshow test result is ($p = 0.398$) :a p-value > 0.05 indicates no statistically significant discrepancy between predicted and observed outcomes. The bias-corrected calibration curve closely aligns with the ideal curve, the Hosmer-Lemeshow test result further confirms that the nomogram has satisfactory calibration

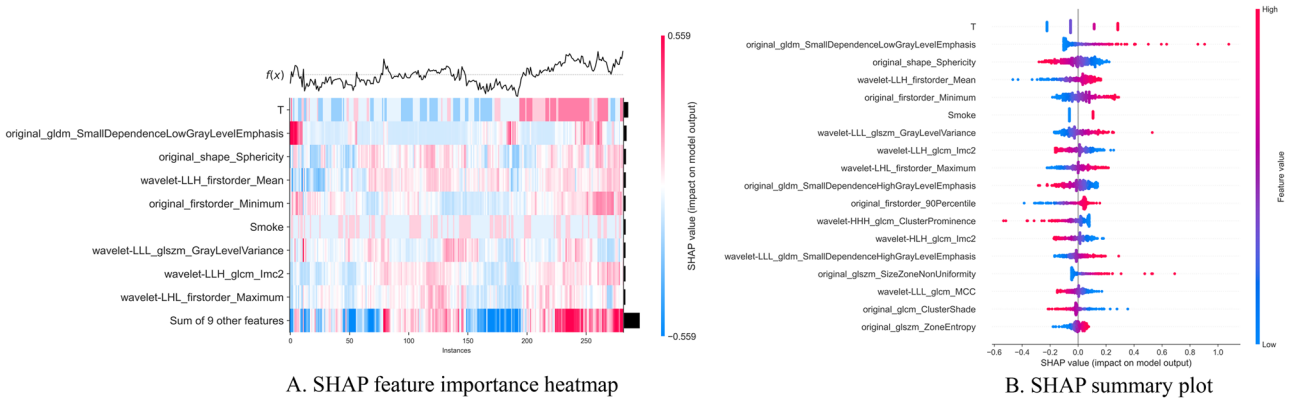

Fig. 6 SHAP plots for predicting lymph node metastasis. **A**: SHAP feature importance heatmap. The x-axis represents individual instances (samples), and the y-axis lists key features (including T stage, radiomic features, smoking history, etc.). Color represents the SHAP value (red: positive, blue: negative, indicating a promoting or inhibitory effect on LNM prediction, respectively). The curve at the top represents the model output. **B**: SHAP summary plot. The y-axis lists key features (including T stage, radiomic texture/shape features, smoking history, etc.). The x-axis represents the impact of the feature value on the model output (SHAP value). Color indicates the feature value (red: high, blue: low)

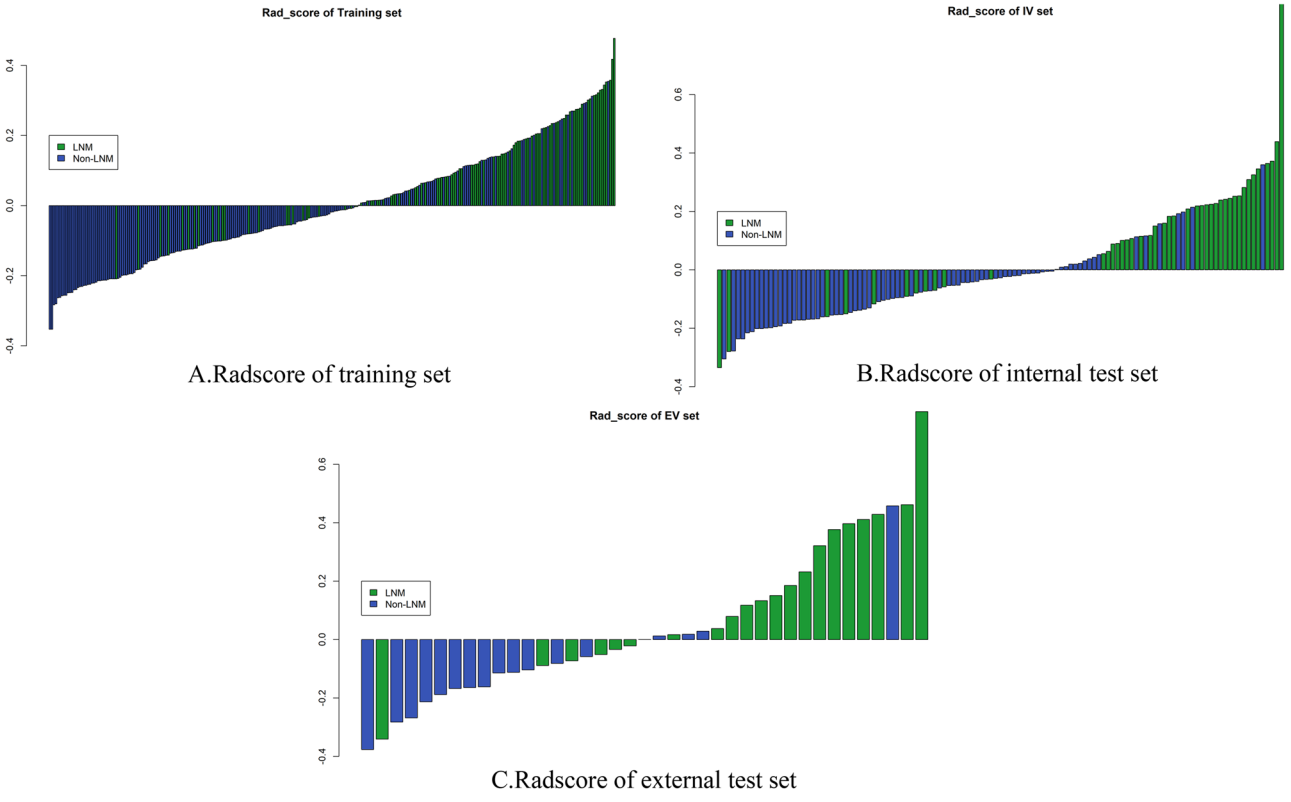

Fig. 7 Bar plot for predicting lymph node metastasis. **A**. Training set: The Rad_score distinction between the two groups of samples is the clearest (the blue group is concentrated in the low-score segment on the left, the green group in the high-score segment on the right, with little overlap);**B**. Internal test set: The distinction is slightly weaker than that of the training set, but the overall trend remains consistent (the blue group is still biased left, and the green group right);**C**. External test set: The distinction is relatively weaker (with some sample overlap), but the trend that "LNM corresponds to a high Rad_score and Non-LNM corresponds to a low Rad_score" still holds.

metastasis, with key clinical applications:①Risk stratification: Map T-stage, smoking status, and Rad_score to Points, sum for Total points, and derive metastasis probability (e.g., total 118 → 0.149) to group patients by risk.②Decision support: Guides preoperative evaluation (e.g., PET-CT for high-risk cases) or treatment planning, avoiding over-/under-intervention.③Improved communication: Visualizes feature-specific risk contributions, enhancing shared decision-making with patients.④Model interpretability: Explicitly displays the weight of

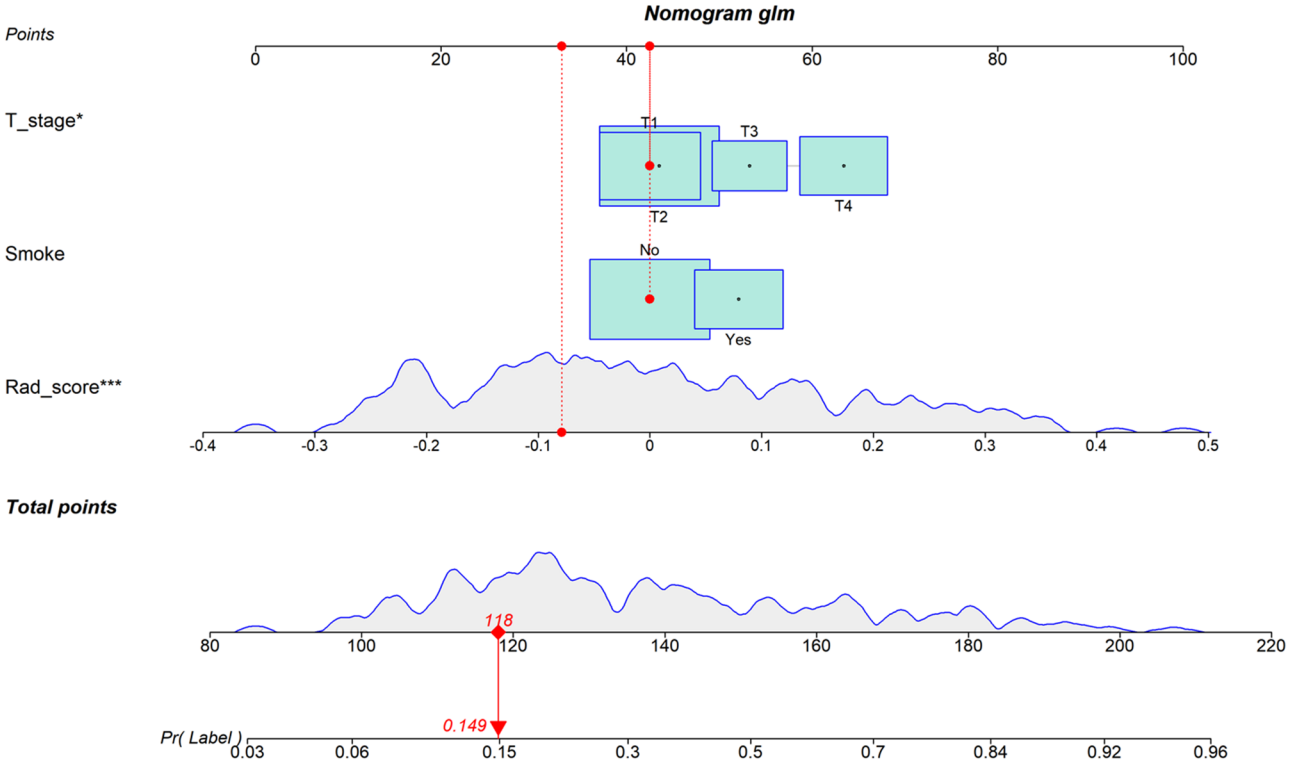

Fig. 8 Nomogram for predicting lymph node metastasis

significant predictors (T-stage, Rad_score), addressing "black-box" limitations of complex models.

Discussion

In this study, we retrospectively developed and validated machine learning models integrating clinical characteristics and radiomics features extracted from non-contrast chest CT to predict lymph node metastasis (LNM) in patients with lung cancer. Among the nine constructed models (three algorithms*three feature sets), the combined clinical–radiomic support vector machine (SVM) model demonstrated the best performance, with AUCs of 0.927 in the training set, 0.852 in the internal test set, and 0.812 in the external test set. In contrast, models based solely on clinical features showed limited discriminative ability, suggesting that radiomics features provide substantial incremental value beyond routine clinical variables.

Previous studies have reported associations between LNM and factors such as age, smoking history, tumor size, histological subtype, degree of differentiation, and serum tumor markers [22, 23]. Several clinical or clinicopathological prediction models have been proposed, with reported AUCs ranging from 0.75 to 0.90 [24–26]. However, the clinical-only model in our set exhibited relatively poor performance, and no significant differences were observed for several conventional clinical variables between patients with and without LNM. This discrepancy may be attributed to set heterogeneity, retrospective selection bias, and the limited discriminatory power of isolated clinical indicators in early nodal spread.

Radiomics enables the high-throughput extraction of quantitative imaging features that reflect intratumoral heterogeneity and spatial complexity beyond visual assessment [27]. Prior CT-based radiomics studies have demonstrated moderate predictive performance for occult LNM, with reported AUCs generally between 0.78 and 0.86 [28, 29]. Deep learning–based approaches and PET/CT-derived radiomics models have shown further performance gains but often rely on contrast agents, metabolic imaging, or complex acquisition protocols [13, 30, 31], which may limit accessibility and increase cost and radiation exposure.

A notable strength of the present study is the exclusive use of non-contrast CT, which is routinely available, cost-effective, and associated with lower radiation burden. Compared with conventional CT(sensitivity: 0.81, specificity: 0.55) and ^{18}F-FDG PET(sensitivity: 0.77, specificity: 0.86) [14], our clinical-radiomics combined model demonstrated a significantly higher specificity (0.93) with a comparable sensitivity (0.81), indicating a superior discriminatory ability relative to traditional size-based assessment methods for lung cancer lymph node metastasis. This finding underscores the potential of radiomics-based models to overcome the well-known limitations of morphological CT criteria, which are prone to false

positives from reactive lymphadenopathy and false negatives in micrometastatic disease.

The radiomics features identified in this study, including first-order, shape, and multi-scale texture features derived from GLCM, GLDM, and GLSZM matrices with wavelet decomposition, provide biologically plausible imaging surrogates of tumor heterogeneity and aggressiveness in lung cancer. First-order intensity features reflect variations in tissue density and composition, potentially corresponding to differences in cellularity, necrosis, and proliferative activity. Shape features, such as sphericity, characterize macroscopic growth patterns, where irregular tumor morphology is indicative of invasive behavior. Texture features quantify spatial complexity and regional non-uniformity of gray-level intensities, capturing intratumoral heterogeneity that has been linked to genomic instability, elevated tumor mutation burden (TMB), and unfavorable prognosis [32, 33]. GLDM- and GLSZM-based features further emphasize localized high-density regions and structural disorder, which may reflect heterogeneous cell packing, stromal remodeling, and variations in immune cell infiltration. Notably, increasing evidence suggests that radiomic heterogeneity correlates with immune-related biomarkers, including PD-L1 expression and tumor-infiltrating lymphocytes (TILs), highlighting the potential of radiomics to non-invasively characterize the tumor immune microenvironment [27, 34]. Wavelet-based features enable multi-scale assessment of tumor architecture, enhancing sensitivity to subtle microstructural patterns associated with proliferation, immune–stromal interactions, and molecular alterations [35]. Collectively, these findings support the role of radiomics as a non-invasive approach for capturing imaging phenotypes that reflect underlying molecular and immunological processes in lung cancer. Consistent with prior reports, tumor T stage emerged as a high-impact predictor, reflecting the increased likelihood of nodal dissemination with advancing local tumor burden [25].

Decision curve analysis further demonstrated that the combined model provided a higher net clinical benefit than treat-all or treat-none strategies across a wide range of threshold probabilities, supporting its potential utility in preoperative risk stratification and treatment planning.

Nevertheless, several limitations should be acknowledged. First, the retrospective design may introduce selection bias. Second, although external validation was performed, the external cohort size was relatively limited, which may influence the precision of generalizability estimates. Importantly, formal radiomics feature stability assessment—such as test–retest analysis or phantom-based reproducibility evaluation—was not conducted in this study. Although imaging was acquired using standardized protocols, feature reproducibility cannot be assumed from scanner uniformity alone, as physiological factors (e.g., respiratory motion), technical variations (e.g., contrast timing and dosage), and patient positioning may influence feature values. Moreover, external validation evaluates model-level performance rather than feature-level robustness, and therefore does not substitute for dedicated stability testing. Third, manual tumor segmentation may introduce interobserver variability despite standardized delineation procedures. Future prospective, multicenter studies incorporating repeat acquisitions or phantom-based analyses in accordance with the image biomarker standardisation initiative(IBSI) methodological recommendations are warranted to ensure feature reproducibility and to further validate and refine the proposed model prior to routine clinical implementation.

Conclusion

This study demonstrates that a machine learning model combining non-contrast CT radiomics and clinical features can outperform conventional CT criteria in predicting lymph node metastasis in lung cancer. With further validation, this approach may serve as a practical, noninvasive tool to assist individualized treatment decision-making.

Abbreviations

AUC	Area under the curve
CEA	Carcinoembryonic antigen
CT	Computed tomography
LNM	Lymph node metastasis
PET	^{18} F-FDG positron emission tomography
ICC	Intraclass Correlation Coefficients
RadScore	Radiomics score
RF	Random forest
SHAP	SHapley Additive exPlanations
SVM	Support vector machine
TMB	Tumor mutation burden
TILs	Tumor-infiltrating lymphocytes
IBSI	Image biomarker standardisation initiative

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-026-02262-x.

Supplementary material 1

Supplementary material 2

Acknowledgements

We thank Panzhihua Traditional Chinese Medicine and Western Medicine Hospital and Panzhihua Central Hospital for providing data records.

Author contributions

PQW, HH, YBW ,YDY and YF contributed equally to the study concept and design, acquisition of subjects and data, analysis and interpretation of data, and preparation of the manuscript and need therefore be granted both first authorship.BSX and JGL were involved in the study design, data analysis and interpretation, and manuscript review. MXK contributed to the study design and manuscript review. CYW participated in the study design, data analysis and interpretation, and manuscript preparation. LY assisted in data acquisition and reviewed the manuscript. DMY contributed to the revision of

the article and the communication and collaboration within the team.BY was involved in the study design, data collection, data analysis and interpretation, and manuscript preparation. All the authors read and approved the final manuscript.

Funding
This work was supported by the Program for 2023 Kunming Health Science and Technology Talent Training Project (Medical Technology Center) (No.2023-SW-13). National Natural Science Foundation of China (NSFC) (No.82160348), 2024 Senior Health Technology and Medical Discipline Leader of Yunnan Provincial Health Commission (No. D-2024056).

Data availability
The data that supports the findings of this study are available in the supplementary material of this article.

Declarations

Ethics approval and consent to participate
The research protocol was approved by the Ethics Committee of Panzhihua Traditional Chinese Medicine and Western Medicine Hospital (PCWEC-C-003). Given the retrospective nature of this study, the requirement for informed consent was waived. All methods were carried out in accordance with relevant guidelines and regulations. Our study adhered to the principles of the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Medical Imaging, Panzhihua Traditional Chinese Medicine and Western Medicine Hospital, Panzhihua 617000, China
[2]Department of Clinical Medical, Dali University, Dali 671003, China
[3]Department of Medical Imaging, Kunming Medical University Affiliated Calmette Hospital, Kunming 650051, China
[4]Kunming Medical University, Kunming 650500, China
[5]Department of Medical Imaging, Panzhihua Central Hospital, Panzhihua 617000, China
[6]Department of Medical Imaging, Baoshan Second People's Hospital, Baoshan 678000, Yunnan, China

Received: 9 October 2025 / Accepted: 2 March 2026
Published online: 16 March 2026

References

1. Freddie B, Mathieu L, Hyuna S, Jacques F. Rebecca LSI.Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.CA. Cancer J Clin. 2024, May-Jun;74(3):229–63. https://doi.org/10.3322/caac.21834.
2. Zhang Y, Li ZJ, Chen YX, Xiao H, Zhou YK, Du SS, et al. Outcomes of image-guided moderately hypofractionated radiotherapy for stage III non-small-cell lung Cancer. J Oncol. 2021, 2721261. https://doi.org/10.1155/2021/2721261.
3. Wanyin L, Carole AR, Andrew GN, Saeed M. The 8(th) lung cancer TNM classification and clinical staging system: review of the changes and clinical implications. Quant imaging Med Surg. 2018, Aug;8(7):709–18. https://doi.org/10.21037/qims.2018.08.02.
4. Zhang ZR, Feng HX, Zhao H, Hu J, Liu LX, Liu Y, et al. Sublobar resection is associated with better perioperative outcomes in elderly patients with clinical stage I non-small cell lung cancer: a multicenter retrospective set study. J Thorac Dis. 2019, May;11(5):1838–48. https://doi.org/10.21037/jtd.2019.05.20.
5. Babatunde AYK, Chi-Fu JY, Brian CG, Li XC, Michael SM, Gu L, et al. A national analysis of wedge resection versus stereotactic body radiation therapy for stage IA non-small cell lung cancer. J Thorac Cardiovasc Surg. 2017, Aug;154(2):675–86.e4. https://doi.org/10.1016/j.jtcvs.2017.02.065.
6. Megan ED, Navneet S, Nofisat I, Mara BA, Douglas AA, Jeffrey B, et al. Management of stage III non-small-cell lung cancer: ASCO guideline. J Clin Oncol. 2022, Apr, 20;40(12):1356–84. https://doi.org/10.1200/JCO.21.02528.
7. Postmus PE, Kerr KM, Oudkerk SS M, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2017, Jul, 1;28iv1–21. https://doi.org/10.1093/annonc/mdx222.
8. Hwayoung S, Sung H Y, Junghoon K, Jihang K, Kyoung W L, Woojoo L, et al. Application of N descriptors proposed by the International association for the study of lung cancer in clinical staging. Radiology. 2021, Aug;300(2):450–57. https://doi.org/10.1148/radiol.2021204461.
9. Noeul K, Sun H S, Hongseok Y, Byung W J, Kyungjong L, Sang-Won U, et al. Infectious complications of EBUS-TBNA: a nested case-control study using 10-year registry data. Lung Cancer. 2021, Nov;1611–8. https://doi.org/10.1016/j.lungcan.2021.08.016.
10. Annalori P, Paolo S. Lung cancer and radiological imaging. Curr Radiopharm. 2020;13(3):238–42. https://doi.org/10.2174/1874471013666200523161849.
11. Cong MD, Feng H, Ren JL, Xu Q, Cong LL, Hou ZZ, et al. Development of a predictive radiomics model for lymph node metastases in pre-surgical CT-based stage IA non-small cell lung cancer. Lung Cancer. 2020, Jan;13973–79. https://doi.org/10.1016/j.lungcan.2019.11.003.
12. Fieke H, David J H, Johannes M D, Naomi B, Rob A E M T, Michel W J M W, et al. Variation between multidisciplinary tumor boards in clinical staging and treatment recommendations for patients with locally advanced non-small cell lung cancer. Chest. 2020, Dec;158(6):2675–87. https://doi.org/10.1016/j.chest.2020.07.054.
13. Michael KG, Ware GK, Chara ER, Courtney CM, Anita ND, Hidenobu S, et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. Ann Intern Med. 2003, Dec, 2;139(11):879–92. https://doi.org/10.7326/0003-4819-139-11-200311180-00013.
14. Sobic-Saranovic D, Petrusic I, Artiko V, Pavlovic S, Subotic D, Saranovic D, et al. Comparison of 18F-FDG PET/CT and MDCT for staging/restaging of non-small cell lung cancer. Neoplasma. 2015;62(2):295–301. https://doi.org/10.4149/neo_2015_035.
15. Lee SW, Kim S-J. Is delayed image of 18F-FDG PET/CT necessary for mediastinal lymph node staging in non-small cell lung cancer patients? Clin Nucl Med. 2022, May, 1;47(5):414–21. https://doi.org/10.1097/RLU.0000000000004110.
16. Huang M, Li SL, Li N, Lai YM, Zhang SY, Lu FL, et al. Occult mediastinal lymph node metastasis in FDG-PET/CT node-negative lung adenocarcinoma patients: risk factors and histopathological study. Thorac Cancer. 2019, Jun;10(6):1453–60. https://doi.org/10.1111/1759-7714.13093.
17. Sarah JG, Anthony WK, Jonathan TP, Kyle B, Frank C D, Daniel J B, et al. Indications for invasive mediastinal staging in patients with early non-small cell lung cancer staged with PET-CT. Lung Cancer. 2017, Jul;10936–41. https://doi.org/10.1016/j.lungcan.2017.04.018.
18. Li YP, Deng JY, Ma XL, et al. Diagnostic accuracy of CT and PET/CT radiomics in predicting lymph node metastasis in non-small cell lung cancer: a systematic review and meta-analysis. Eur radiol. 2025, Apr;35(4):1966–79. https://doi.org/10.1007/s00330-024-11036-4.
19. Wu T, Gao C, Lou XJ, et al. Predictive value of radiomic features extracted from primary lung adenocarcinoma in forecasting thoracic lymph node metastasis: a systematic review and meta-analysis. BMC pulm Med. 2024, May, 18;24(1):246. 10.1186/s12890–024–03020–x.
20. Duan FR, Zu HY, Zhang R, et al. Deep learning radiomics and 18F-FDG PET/CT imaging: mediastinal lymph node characteristics as predictors of metastasis in NSCLC. Transl Lung Cancer Res. 2025, Oct, 31;14(10):4301–14. https://doi.org/10.21037/tlcr-2025-650.
21. Duan FY, Zhang MH, Yang CY, et al. Non-invasive prediction of lymph node metastasis in NSCLC combining clinical, radiomics, and deep learning features with interpretable machine learning. Acad radiol, 2025, Mar;32(3):1645–55. https://doi.org/10.1016/j.acra.2024.11.037.
22. Chen B, Xia WJ, Wang ZQ, Zhao H, Li XF, Liu LX, et al. Risk analyses of N2 lymph-node metastases in patients with T1 non-small cell lung cancer: a multi-center real-world observational study in China. J Cancer Res Clin Oncol. 2019, Nov;145(11):2771–77. https://doi.org/10.1007/s00432-019-03006-x.
23. Shan LY, Zhang L, Zhu XL, Wang ZL, Fang SH, Lin JF, et al. Chinese never smokers with adenocarcinoma of the lung are younger and have fewer lymph node metastases than smokers. Respir Res. 2022, Oct, 29;23(1):293. https://doi.org/10.1186/s12931-022-02199-z.

24. Ryuichi Y, Yoshitaka E, Takuya A, Hiroyuki D, Makoto T, Wataru S, et al. Diagnostic artificial intelligence model predicts lymph node status in non-small cell lung cancer using simplified examination. J Thorac Dis. 2024, Nov;16(11):7320–28. https://doi.org/10.21037/jtd-24-1067.
25. Hu SL, Luo M, Li YL. Machine learning for the prediction of lymph node micrometastasis in patients with non-small cell lung cancer: a comparative analysis of two practical prediction models for gross target volume delineation. Cancer Manag Res. 2021, 13:4811–20. https://doi.org/10.2147/CMAR.S313941.
26. Tang YB, Garu A, Chen X, Guan ZY, Cai XD, Huang HX, et al. Risk factors for mediastinal lymph node metastases in early-stage non-small-cell lung cancer and prediction model establishment. Am J Cancer Res. 2024;14(12):5952–64. https://doi.org/10.62347/DIZG4944.
27. Gillies RJ, Kinahan PE. Hricak H.Radiomics: images are more than pictures, they are data. Radiology. 2016 278(2):563–77. https://doi.org/10.1148/radiol.2015151169.
28. Zhang R, Zhang RR, Luan T, Liu BW, Zhang YM, Xu YP, et al. A radiomics nomogram for preoperative prediction of clinical occult lymph node metastasis in cT1-2N0M0 solid lung adenocarcinoma. Cancer Manag Res. 2021, 13:8157–67. https://doi.org/10.2147/CMAR.S330824.
29. Geng MF, Geng MS, Wei R, Chen MW. Artificial intelligence neural network analysis and application of CT imaging features to predict lymph node metastasis in non-small cell lung cancer. J Thorac Dis. 2022, Nov;14(11):4384–94. https://doi.org/10.21037/jtd-22-1511.
30. Kyongmin SB, Bomi G, Sae JN, Ji HH, Sang HC, Ho JA, et al. DeepCUBIT: predicting lymphovascular invasion or pathological lymph node involvement of clinical T1 stage non-small cell lung cancer on chest CT using deep cubical nodule transfer learning. Front Oncol. 2021, 11:661244. https://doi.org/10.3389/fonc.2021.661244.
31. Boris G, Philipp HB, Alexander DJB, Alexander H, Julian MMR, Christian F, et al. Inferring FDG-PET positivity of lymph node metastases in proven lung cancer from contrast-enhanced CT using radiomics and machine learning. Eur radiol exp. 2022, 6(1):44. https://doi.org/10.1186/s41747-022-00296-8.
32. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014, 5:4006. https://doi.org/10.1038/ncomms5006.
33. Ganeshan B. Miles KA.Quantifying tumour heterogeneity with CT. Cancer imaging. 2013, Mar, 26;13(1):140–49. https://doi.org/10.1102/1470-7330.2013.0015.
34. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine.Nat. Rev Clin Oncol. 2017, Dec;14(12):749–62. https://doi.org/10.1038/nrclinonc.2017.141.
35. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med biol. 2016, Jul, 7;61(13):R150–66. https://doi.org/10.1088/0031-9155/61/13/R150.

Publisher's Note