BMC Medical Imaging

RESEARCH

Open Access

Computed tomography-based radiomics model for predicting station 4 lymph node metastasis in non-small cell lung cancer

Yanru Kang[1], Mei Li[1], Xizi Xing[1], Kaixuan Qian[1], Hongxia Liu[2], Yafei Qi[3], Yanguo Liu[4], Yi Cui[3*] and Hua Zhang[1*]

Abstract

Background This study aimed to develop and validate machine learning models for preoperative identification of metastasis to station 4 mediastinal lymph nodes (MLNM) in non-small cell lung cancer (NSCLC) patients at pathological N0-N2 (pN0-pN2) stage, thereby enhancing the precision of clinical decision-making.

Methods We included a total of 356 NSCLC patients at pN0-pN2 stage, divided into training ($n=207$), internal test ($n=90$), and independent test ($n=59$) sets. Station 4 mediastinal lymph nodes (LNs) regions of interest (ROIs) were semi-automatically segmented on venous-phase computed tomography (CT) images for radiomics feature extraction. Using least absolute shrinkage and selection operator (LASSO) regression to select features with non-zero coefficients. Four machine learning algorithms—decision tree (DT), logistic regression (LR), random forest (RF), and support vector machine (SVM)—were employed to construct radiomics models. Clinical predictors were identified through univariate and multivariate logistic regression, which were subsequently integrated with radiomics features to develop combined models. Models performance were evaluated using receiver operating characteristic (ROC) analysis, calibration curves, decision curve analysis (DCA), and DeLong's test.

Results Out of 1721 radiomics features, eight radiomics features were selected using LASSO regression. The RF-based combined model exhibited the strongest discriminative power, with an area under the curve (AUC) of 0.934 for the training set and 0.889 for the internal test set. The calibration curve and DCA further indicated the superior performance of the combined model based on RF. The independent test set further verified the model's robustness.

Conclusions The combined model based on RF, integrating radiomics and clinical features, effectively and non-invasively identifies metastasis to the station 4 mediastinal LNs in NSCLC patients at pN0-pN2 stage. This model serves as an effective auxiliary tool for clinical decision-making and has the potential to optimize treatment strategies and improve prognostic assessment for pN0-pN2 patients.

Clinical trial number Not applicable.

Keywords Non-small cell lung cancer, Mediastinal lymph node metastasis, Machine learning, Radiomics

*Correspondence:
Yi Cui
18560081621@163.com
Hua Zhang
21hua@sina.com
[1]School of Clinical and Basic Medicine, Shandong First Medical University, Jinan 250117, China
[2]Department of Radiology, The Affiliated of Shandong Traditional Medical University, Jinan 250011, China
[3]Department of Radiology, Qilu Hospital of Shandong University, Jinan 250012, China
[4]Department of Medical Oncology, Qilu Hospital of Shandong University, Jinan 250012, China

Background

Lung cancer is one of the malignant tumors with the highest incidence rate and mortality in the world. Non-small cell lung cancer (NSCLC) accounts for about 85% of lung cancer, and its 5-year survival rate is only about 19% [1]. Lung cancer mainly spreads through the lymphatic system, among which NSCLC has formed a certain pattern of lymph node metastasis (LNM). Generally, it first infiltrates the lymph nodes (LNs) in the lungs, then the LNs in the hilum, and finally the mediastinal LNs [2]. However, it can also directly skip metastasize to the mediastinal LNs [2]. Lymph node dissection (LND) has been shown to significantly enhance patient survival [3]. Because LNs in the lungs and hilum are often removed along with resection of the diseased lung, the thoroughness of LND is more fully reflected in the dissection of mediastinal LNs. Mediastinal lymph node metastasis (MLNM) determines the formulation of follow-up treatment for NSCLC [4]. For patients with pathological stage pN0 to pN2 NSCLC, surgery or surgery combined with neoadjuvant therapy can provide additional survival benefits. However, patients with pN3 stage NSCLC require systemic or palliative treatment [4]. Our study focuses on pN0-pN2 patients with the aim to predict metastasis early, devise appropriate anatomical and therapeutic strategies, and improve the survival rates of NSCLC patients.

Station 4, also known as the lower paratracheal LNs, is characterized by a high metastatic rate and a low resection rate [5]. It consists of the right lower paratracheal station 4R (S4R) and the left lower paratracheal station 4L (S4L), which are part of the superior mediastinum. S4R is surrounded by important blood vessels such as the superior vena cava, azygos vein, and aorta and is adjacent to the trachea and the upper lobe bronchus of the right lung. The risk and difficulty of puncture during clinical treatment are extremely high [5]. S4L is surrounded by the transverse ligament, the ligamentum arteriosum, the aortic arch, and the left main pulmonary artery [6–8]. S4L LND is not commonly undertaken due to the intricate anatomical structures in the vicinity and the risk of injury to the left recurrent laryngeal nerve and the thoracic duct [9–11].

The study found that in NSCLC, metastasis to the MLNM most frequently involves the S4R and station 7 LN [12]. There is debate regarding the impact of LNM at station 4 or 7 on prognosis, but it is widely acknowledged that the prognosis is poorer compared to other stations [5]. Currently, the primary challenge in precision therapy is determining which LND approach yields the greatest survival benefit for patients with NSCLC [12, 13, 14]. Most researches tend to favor systematic lymph node dissection (SLND) [13, 15]. According to the recommendations of the International Association for the Study of Lung Cancer, the minimal excision requirement for SLND should involve at least three mediastinal nodal stations, including station 7, the subcarinal LN [9]. However, specific requirements for removal of LNs in station 4 have not yet been established.

The pathological examination serves as the gold standard for diagnosing LNM in NSCLC. This method, although invasive, is contingent upon postoperative procedures for accurate assessment [16]. As an alternative, computed tomography (CT) serves as a crucial tool for assessing MLNM. The primary criterion for determining LNM is whether tumor cells induce alterations in imaging characteristics, including size, shape, and density. The short diameter of the station 4 LNs is larger than that at other stations, particularly in S4R [5]. Enlargement of mediastinal LNs can be associated with malignant tumor metastasis, hematological disorders, inflammation, and autoimmune diseases, among others [17]. However, MLNM do not necessarily exhibit enlargement [18, 19]. Therefore, relying solely on the size of LNs to determine MLNM may lead to higher false negative and false positive rates, as well as excessive radiation risk. The accuracy of detecting LNM by means of endoscopic ultrasonography (EUS) or positron emission tomography-computed tomography (PET-CT) still requires enhancement [20]. Therefore, there is a pressing need for a non-invasive and more accurate method to predict LNM in station 4.

Radiomics is a method of converting medical images into multidimensional quantitative data for clinical decision analysis [16, 21, 22]. In previous studies on predicting LNM, radiomics models have demonstrated strong predictive performance for both primary tumor and LN regions [23, 24]. However, due to variations in metastasis rates and to cleaning difficulties among different stations, inadequate or excessive cleaning of LNs can result in varying degrees of complications and adverse consequences. Although most studies concentrate on overall LNM to gauge patient prognosis rather than directing intraoperative LND [25, 26], these studies also highlight the potential of radiomics in predicting the status of individual LN stations. To our knowledge, there have been no reports on radiomics models specifically designed to predict metastasis in station 4 LNs. Consequently, this study aims to develop 12 preoperative predictive models using four algorithms based on radiomics features and clinical risk factors, targeting the prediction of MLNM in station 4 for NSCLC patients with pN0-pN2 stage. Through rigorous evaluation and comparison, the optimal model will be selected to assist clinicians in choosing the most appropriate treatment strategy, thereby improving clinical outcomes.

Patients and methods

Patients

This study was approved by the Ethics Committee of the hospital, and the individual informed consent requirement for this retrospective analysis was waived. This study was conducted in accordance with the CLEAR and CLEAR-E3 guidelines [27, 28], and the completed checklists were submitted as Supplementary Material 1 and 2. The detailed technology roadmap for this study is presented in Fig. 1 and is outlined below.

This retrospective study analyzed patients with NSCLC who were pathologically confirmed at two campuses of the hospital between January 2017 and October 2024. The data were derived from our hospital's private medical records, not public databases. The data filtering process is illustrated in Fig. 2. The inclusion criteria were as follows: (1) patients who underwent surgical resection and SLND in thoracic surgery, with postoperative pathological results confirming pN0-pN2 NSCLC and retaining complete pathological diagnostic data; (2) a complete enhanced chest CT scan conducted within two weeks prior to surgery; (3) image quality that met the analysis standards and complete clinical data; (4) pathological examination confirming that station 4 LNs were either completely metastatic or non-metastatic. The exclusion criteria included: (1) patients who had received preoperative radiotherapy, chemotherapy, or other treatments; (2) the presence of distant metastasis or other malignant tumors; (3) incomplete clinical data or the presence of image artifacts; (4) LNs at station 4 with incomplete site metastasis.

The sample size calculation for this study was performed using PASS 2021 software, based on the methods of sample size estimation for receiver operating characteristic (ROC) curve analysis, particularly focusing on differences in the area under the curve (AUC) [29, 30, 31]. To test the alternative hypothesis (H1), the study necessitates the recruitment of 289 patients to ensure that the model's AUC achieves a value of 0.85 under conditions of approximately 85% statistical power and a significance level (α) of 0.05 (two-sided). The calculation of the sample size was predicated on the following assumptions: (1) the AUC under the null hypothesis was set at 0.75; (2) the proportion of positive cases represented 38% of the total participant population.

This study included 356 patients with 591 LNs sourced from two campuses of the same hospital. The main campus contributed cases from January 2017 to December 2023, totaling 297 patients and 513 LNs, with 112 (38%) LNM-positive and 185 (62%) LNM-negative patients. Despite a slight class imbalance (1:1.6 ratio of LNM positive to negative), we refrained from resampling or balancing to preserve the data's natural distribution and avoid augmentation biases. An additional 59 cases (78 LNs) were collected from another campus from January to October 2024. In this study, we identified 146 patients without metastasis at S4R, accounting for 235 LNs, and 118 patients with metastasis, involving 208 LNs. For S4L,

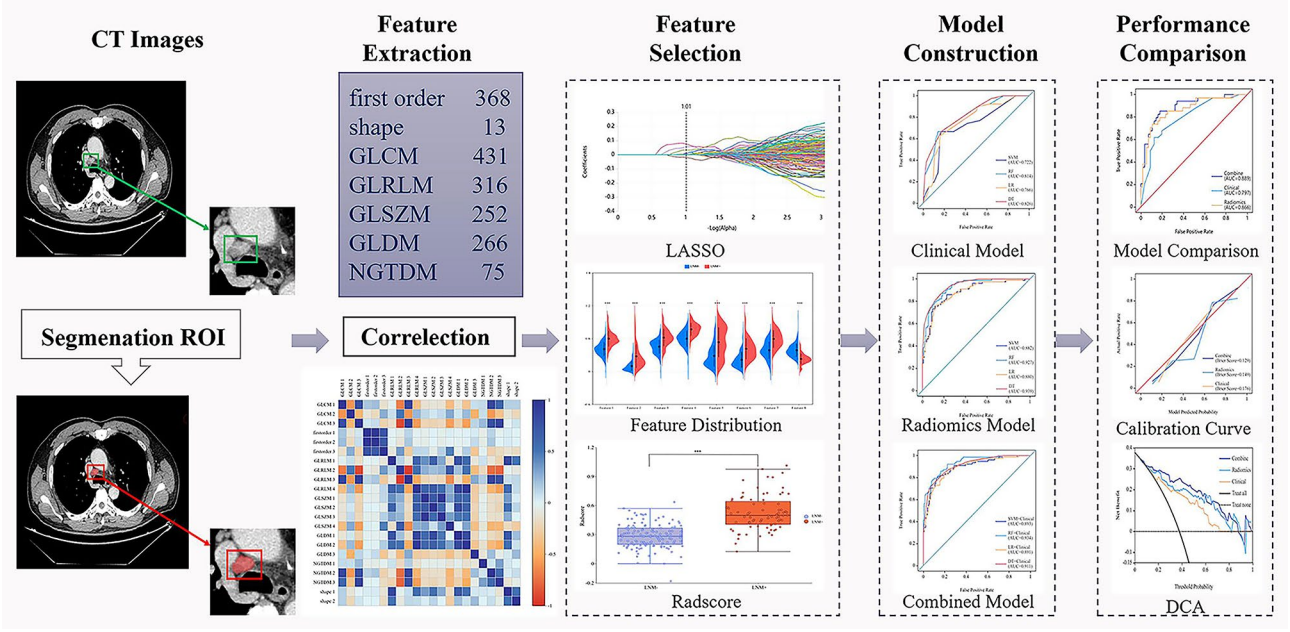

Fig. 1 Technology roadmap of radiomics analysis. DCA, decision curve analysis; GLCM, gray level co-occurrence matrix; GLDM, gray level dependence matrix; GLRLM, gray level run length matrix; GLSZM, gray level size zone matrix; LASSO, least absolute shrinkage and selection operator; NGTDM, neighboring gray tone difference matrix; ROI, region of interest

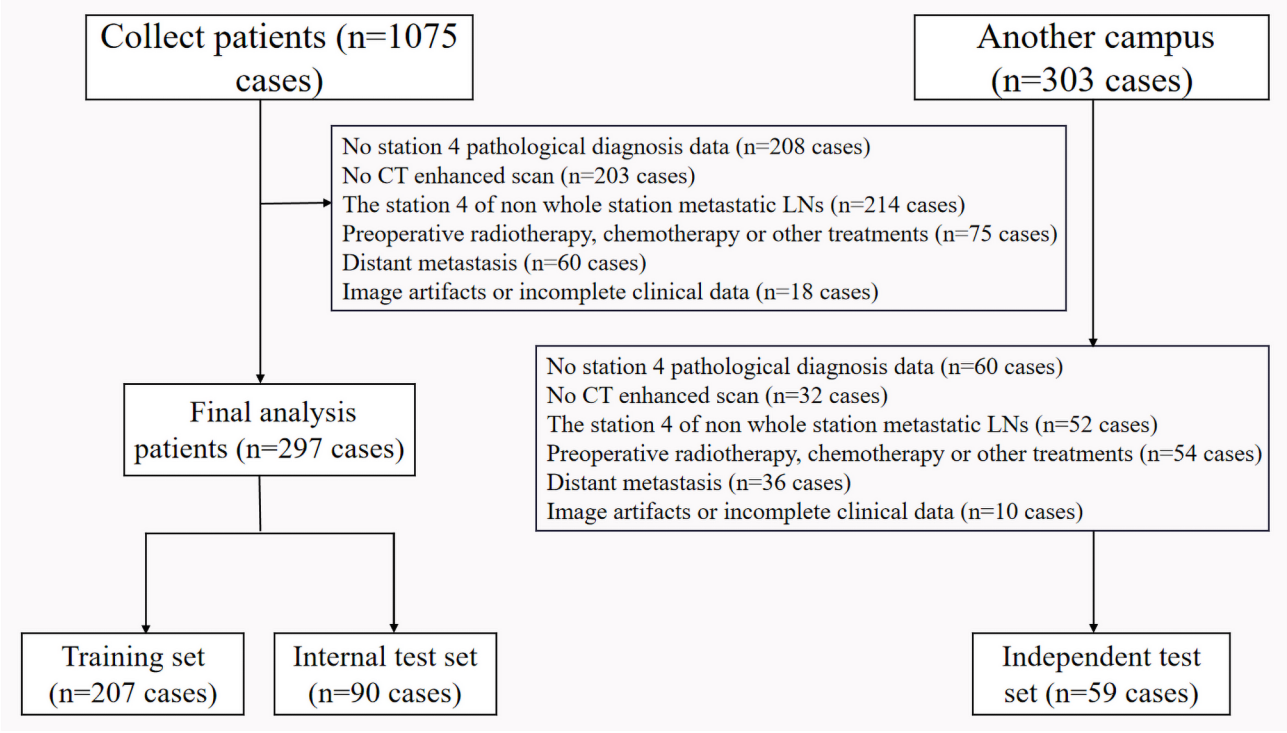

Fig. 2 Data filtering process

there were 62 patients without metastasis, with 100 LNs, and 30 patients with metastasis, with 48 LNs.

The main campus cases were randomly allocated to a training set (207 patients, 359 LNs) and an internal test set (90 patients, 154 LNs) in a 7:3 ratio. Additionally, the 59 cases (78 LNs) collected from another campus were used as an independent test set to further validate the model.

The training set had 78 patients with LN metastasis (LNs = 144) and 129 without (LNs = 215); The internal test set consisted of 34 patients with LN metastasis (LNs = 65) and 56 without (LNs = 89); The independent test set included 36 patients with LN metastasis (LNs = 47) and 23 without (LNs = 31).

To prevent information leakage between datasets, a hold-out test set was created prior to feature selection. The independent test set was derived from different devices and scanning protocols to ensure model robustness and generalizability.

The baseline clinical data were abstracted from medical records, including demographics such as age, sex, and smoking. Clinical information predominantly originated from preoperative CT reports and images, encompassing details such as tumor location, presence of lobulation sign, spiculation sign, vacuolation sign, air bronchogram sign, vascular convergence sign, pleural traction sign, and the short diameter of LNs.

Image acquisition

A thin-slice CT scan was performed on all patients. All patients were examined in the supine position, and the scan range was from the thorax entrance to the posterior costal angle. The single breath was continuously screened after end-inspiratory. All images were displayed at standard mediastinal window settings (width 350 HU, horizontal 50 HU). 1.2–1.5 ml/kg iodinated contrast agent was injected through the elbow vein at a speed of 3–3.5 ml/s, with a venous phase scan delay of 30 s.

The main campus Siemens SOMATOM Force CT scanning mode was spiral scanning, with a pitch between 0.50 and 0.80, a tube voltage of 120 kV, an effective tube current maintained at 250–350 mAs (via automatic mA technology), a tube rotation speed of 0.6–0.8 s/rotation, a detector width of 64 × 0.625 mm, and an acquisition matrix of 512 × 512.

The GE REVOLUTION CT scanning mode in another campus was spiral scanning, with a pitch ranging from 0.80 to 1.20, a tube voltage of 120 kV, an effective tube current maintained at 200–400 mAs (via automatic mA technology), a tube rotation speed of 0.6–0.8 s/rotation, a detector width of 64 × 0.625 mm, and an acquisition matrix of 512 × 512.

All images were exported in Digital Imaging and Communications in Medicine format (DICOM) and then uploaded to the uAI Research Portal system (United Imaging Intelligence, https://urp.united-imaging.com)

for LN region of interest (ROI) segmentation and extraction of imaging features.

Imaging evaluation

Two senior thoracic imaging diagnosticians independently assessed the following CT features: lobulation sign, spiculation sign, vacuolation sign, air bronchogram sign, vascular convergence sign, and pleural traction sign. In addition, they measured the short diameter of LNs at station 4. Lobulation sign refers to the irregular, wavy contour of a nodule's edge [25]. Spiculation sign is defined as linear projections extending from the surface of a nodule into the pulmonary parenchyma without reaching the pleural surface [25]. Vacuolation sign refers to the presence of small, round, oval, or streaky areas of reduced density within a dense mass or nodular lesion. The air bronchogram sign refers to the contrast between the consolidation of lung tissue near the hilum and larger air-containing bronchi, which results in the visualization of branch shadows of inflated bronchi in the merged lung area. Vascular convergence sign is defined by clustering of major vessels and bronchi around a nodule, which may either abut or penetrate the lesion. This radiographic feature is evident as blood vessels or bronchi converging on the nodule. Pleural traction sign is characterized by one or more lines extending from the nodule surface to the pleural surface due to the thickening of the interlobular septa, without reaching the pleural surface [26, 32]. The short diameter of the largest cross-section of the LN is used as the standard for measuring the short diameter, with a cutoff of 1 cm for grouping [5]. The pathological status of station 4 LNs in NSCLC was used as the gold standard, despite its invasiveness. All divergences in the description of imaging features were addressed and resolved through consensus.

Image segmentation

CT venous phase images were imported into the uAI Research Portal (https://urp.united-imaging.com) for preprocessing, To address potential limitations associated with each feature unit, we implemented the following preprocessing steps: grayscale discretization with a bin width of 25, window width/level normalization (window width 350 HU, window level 50 HU), and Z-score normalization of image intensity. These steps ensured consistent intensity information across images, thereby minimizing variations in gray-level texture. Additionally, the data were resampled to a consistent voxel size of $1 \times 1 \times 1$ mm^3 to standardize the image scale. The specific parameters for preprocessing were detailed in Supplementary Material 3.

The uAI Research Portal incorporates a variety of filters, including original, box mean, box sigma image, speckle noise, shot noise, log, and wavelet. These filters enhance model robustness and texture analysis capabilities, facilitating more effective feature extraction.

Two radiologists, each with over 15 years of experience, utilized the uAI Research Portal to meticulously semi-automatically outline ROI along the edges of LNs in venous phase CT scans. They took care to encompass the full thickness of the LN, carefully avoiding artifacts and adipose tissue that could obscure critical imaging details. Subsequently, a deputy chief physician or a higher-level radiologist conducted a thorough review and refinement of the LN delineation on the CT images. For LNs with ambiguous locations, radiologists accurately defined the ROI in accordance with pathological descriptions. To mitigate sampling bias within the study sample, a maximum of three LNs were selected from each patient [33].

Clinical imaging feature selection

We initially performed univariate analysis on baseline and clinical characteristics to identify clinical risk factors for MLNM, considering $p < 0.05$ as indicative of statistical significance. Following this, we incorporated the variables that were significant in the univariate analysis into a logistic regression model, with a significance level of $p < 0.05$ to identify the clinical risk factors associated with MLNM at station 4 for NSCLC patients in the pN0-pN2 stage.

Radiomics features extraction & selection

Using the uAI Research Portal (IBSI-compliant), We extracted 2221 handcrafted quantitative features, including texture and shape features, from 3D ROIs in venous phase images, which were labeled according to pathological LNM status. First order features (439) described pixel distributions like mean and standard deviation. Shape features (14) measured the geometry of ROIs, such as compactness. Gray level co-occurrence matrix (GLCM) features (515) quantified spatial relationships and texture. Gray level run length matrix (GLRLM) features (395) assessed texture uniformity. Gray level size zone matrix (GLSZM) features (393) analyzed the sizes of connected regions. Gray level dependence matrix (GLDM) features (346) captured complex patterns. Neighborhood gray tone difference matrix (NGTDM) features (119) measured gray-level differences. All other parameters remained as a default configuration.

To address the limitations associated with each feature unit, we performed Z-score normalization on the feature values across all images in the training set. This normalization was then independently applied to the internal and independent test sets. To confirm the robustness of radiomics features, we assessed the intra-observer consistency of feature extraction in the training set using the intraclass correlation coefficient (ICC) [34]. We randomly selected half the patients, and one month later, the

same radiologist performed ROI segmentation again on this subset of images. Any features with ICC < 0.75 were discarded. To eliminate redundant features, mitigate model overfitting, and enhance model interpretability, we applied least absolute shrinkage and selection operator (LASSO) regression analysis to the training dataset utilizing the selected features. L1 regularization was used to shrink the coefficients of less important features to zero by adding the absolute value of the magnitude of the coefficients as a penalty term to the loss function. Based on these selected features, we calculated the radiomics score (*Radscore*) for each patient. The *Radscore* of the model can be calculated as follows:

$$\text{Radscore} = \Sigma_i \text{Feature}_i \times \text{Coefficient} + b$$

where Coefficient is obtained by an iterative version of the LASSO algorithm and *i* represents the feature.

Models construction

In this study, we used four distinct algorithms to develop predictive models by hyperparameter method: decision tree (DT), logistic regression (LR), random forest (RF), and support vector machine (SVM). The specific parameters for each model were detailed in Supplementary Material 3. Clinical models were crafted by identifying independent risk factors through comprehensive multifactor analysis, whereas radiomics models were established utilizing *Radscore*. Given the unique characteristics of each algorithm, different results may emerge.

Construction of the combined models

We employed these four algorithms (DT, LR, RF, and SVM) to further construct combined models that use clinical factors together with radiomics features selected by LASSO. Subsequently, we compared the predictive capabilities of the combined models with those of the individual radiomics and clinical models.

Models performance evaluation and validation

The predictive performance of the models in the training and testing sets was evaluated using AUC. Sensitivity, specificity, accuracy, precision, and F1 score were employed to reflect the overall performance of the models. The calibration curve reflected the consistency between the predictive model and the observed outcomes. Furthermore, decision curve analysis (DCA) was used to quantify the net benefit at different probability threshold levels. "Treat All" assumed intervention for all, while "Treat None" assumed no intervention. Comparing a model's net benefit to these benchmarks assessed its clinical value. The selection of the optimal model was based on the training set and internal test set, with the independent test set from different scanning devices reserved for model validation. To enhance the interpretability of the optimal model, we utilized SHAP (Shapley additive explanations) plots for detailed analysis.

Statistical analysis

All statistical analyses were conducted using SPSS software version 26.0.0, and R software version 4.4.3, with additional support from PASS 2021 and G*Power software for sample size estimation and post hoc power analysis. The Mann-Whitney U test was used to compare continuous variables such as age that do not follow a normal distribution. The independent samples t-test was employed to analyze features that conform to a normal distribution, such as *Radscore*. Categorical variables, including sex, smoking, tumor location, lobulation sign, spiculation sign, vacuolation sign, air bronchogram sign, vascular convergence sign, pleural traction sign, and short diameter of LNs, were analyzed using the chi-squared test or Fisher's exact test. DeLong's test was employed to compare differences between ROC curves. The calibration curve was assessed using the Hosmer-Lemeshow goodness of fit test to evaluate the agreement between the predicted and observed outcomes. All statistical significance levels were set to be two-tailed, with $p < 0.05$ considered statistically significant.

Results

Clinical features and clinical models construction

Table 1 provied a summary of clinical characteristics for patients in the training set ($n = 207$), internal test set ($n = 90$), and external test set ($n = 59$). The detailed content of Table 1 is provided in Supplementary Material 4. Univariate analysis revealed significant differences ($p < 0.05$) between LNM and non-LNM groups in terms of smoking, tumor location, spiculation sign, and LN short diameter within the training set. In contrast, no significant differences ($p > 0.05$) were observed for age, sex, lobulation sign, vacuolation sign, air bronchogram sign, vascular convergence sign, and pleural traction sign. Following the adjustment for confounding factors by means of multivariate logistic regression, tumor location ($p < 0.05$; post hoc power, 73%), spiculation sign ($p < 0.05$; post hoc power, 59%), and LN short diameter were identified as independent risk factors for LNM at station 4 in patients with NSCLC at pN0-pN2 stages ($p < 0.001$, post hoc power, > 90%) (Fig. 3).

Among the clinical models constructed based on three independent risk factors using four algorithms, the RF model exhibited an AUC of 0.797 (95% confidence interval [CI] 0.704–0.890) (Fig. 4D), sensitivity 0.618, specificity 0.839, accuracy 0.756. An AUC of 0.814 (95% CI 0.757–0.872) was observed in the training set (Fig. 4A), with sensitivity 0.667, specificity 0.860, and accuracy 0.787 (Table 2).

Table 1 Features in development of clinical models

Variable	Training set (*n* = 207)			Internal test set(*n* = 90)			Independent test set (*n* = 59)		
	LNM (-) (*n* = 129)	LNM (+) (*n* = 78)	*p* value	LNM (-) (*n* = 56)	LNM (+) (*n* = 34)	*p* value	LNM (-) (*n* = 23)	LNM (+) (*n* = 36)	*p* value
Age (years)	59.0 (56.0, 65.5)	61.0 (54.0, 65.0)	0.737	63.0 (57.0, 67.0)	64.0 (55.8, 70.3)	0.614	68.0 (59.0, 72.0)	65.5 (57.0, 70.8)	0.619
Sex (%)									
Male	69 (53.5)	46 (59.0)	0.441	27 (48.2)	22 (64.7)	0.128	14 (60.9)	25 (69.4)	0.497
Female	60 (46.5)	32 (41.0)		29 (51.8)	12 (35.3)		9 (39.1)	11 (30.6)	
Smoking (%)									
NO	89 (69.0)	43 (55.1)	**0.044**	39 (69.6)	14 (41.2)	**0.008**	15 (65.2)	24 (66.7)	0.909
YES	40 (31.0)	35 (44.9)		17 (30.4)	20 (58.8)		8 (34.8)	12 (33.3)	
Tumor location (%)									
Right upper	51 (39.5)	45 (57.7)	**0.036**	27 (48.2)	17 (50.0)	**0.049**	8 (34.8)	15 (41.7)	0.642
Right middle	22 (17.1)	12 (15.4)		7 (12.5)	12 (35.3)		1 (4.3)	5 (13.9)	
Right lower	15 (11.6)	4 (5.1)		9 (16.0)	2 (5.9)		4 (17.4)	4 (11.1)	
Left upper	21 (16.3)	13 (16.7)		10 (17.9)	2 (5.9)		6 (26.1)	9 (25.0)	
Left lower	20 (15.5)	4 (5.1)		3 (5.4)	1 (2.9)		4 (17.4)	3 (8.3)	
Spiculation sign (%)									
NO	78 (60.5)	35 (44.9)	**0.029**	36 (64.3)	9 (26.5)	**0.001**	16 (69.6)	20 (55.6)	0.282
YES	51 (39.5)	43 (55.1)		20 (35.7)	25 (73.5)		7 (30.4)	16 (44.4)	
Lobulation sign (%)									
NO	63 (48.8)	39 (50.0)	0.871	27 (48.2)	16 (47.1)	0.915	13 (56.5)	20 (55.6)	0.942
YES	66 (51.2)	39 (50.0)		29 (51.8)	18 (52.9)		10 (43.5)	16 (44.4)	
Vacuolation sign (%)									
NO	113 (87.6)	61 (78.2)	0.074	51 (91.1)	26 (76.5)	0.109	20 (87.0)	28 (77.8)	0.589
YES	16 (12.4)	17 (21.8)		5 (8.9)	8 (23.5)		3 (13.0)	8 (22.2)	
Air bronchogram sign (%)									
NO	119 (92.2)	71 (91.0)	0.756	50 (89.3)	29 (85.3)	0.819	17 (73.9)	29 (80.6)	0.548
YES	10 (7.8)	7 (9.0)		6 (10.7)	5 (14.7)		6 (26.1)	7 (19.4)	
Vascular convergence sign (%)									
NO	110 (85.3)	72 (92.3)	0.132	48 (85.7)	30 (88.2)	0.983	17 (73.9)	30 (83.3)	0.586
YES	19 (14.7)	6 (7.7)		8 (14.3)	4 (11.8)		6 (26.1)	6 (16.7)	
Pleural traction sign (%)									
NO	71 (55.0)	47 (60.3)	0.462	34 (60.7)	15 (44.1)	0.125	12 (52.2)	23 (63.9)	0.372
YES	58 (45.0)	31 (39.7)		22 (39.3)	19 (55.9)		11 (47.8)	13 (36.1)	
Short diameter (%)									
Only < 1 cm	116 (89.9)	21 (26.9)	**< 0.001**	46 (82.2)	6 (17.6)	**< 0.001**	18 (78.3)	5 (13.9)	**< 0.001**
Only ≥ 1 cm	4 (3.1)	32 (41.0)		5 (8.9)	21 (61.8)		3 (13.0)	27 (75.0)	
All have	9 (7.0)	25 (32.1)		5 (8.9)	7 (20.6)		2 (8.7)	4 (11.1)	

The bold values represent p < 0.05

Feature selection and radiomics models construction

The training set consisted of 2,221 traditional radiomics features, which were assessed using an ICC threshold of greater than 0.75, resulting in the selection of 1,721 features. The intercorrelation among these features is graphically depicted in Fig. 5A. Subsequently, LASSO regression was used for further dimensionality reduction (Fig. 5B), and ultimately eight radiomics features with non-zero coefficients were retained for inclusion in the training set (Table 3; Fig. 5C). The distinct expression patterns of these eight features across groups with and without LNM are presented in Fig. 6A and B. The post hoc power analysis demonstrated that the statistical power for these eight features in the training set and internal test set was greater than 90%. These features were integrated into radiomics signatures and expressed through *Radscore* calculation functions. The formula for calculating the *RadScore* is:

Radscore = 0.07408811 * wavelet_glrlm_wavelet-HHH-Run Variance + 0.05581313 * box mean_gldm_Gray Level Non-Uniformity + 0.03211861 * shot noise_glrlm_Run Entropy + 0.01013453 * log glrlm_log sigma-1-0-mm-3D-Run Entropy + 0.00919728 * shot noise_gldm_Large Dependence Emphasis + 0.00762456

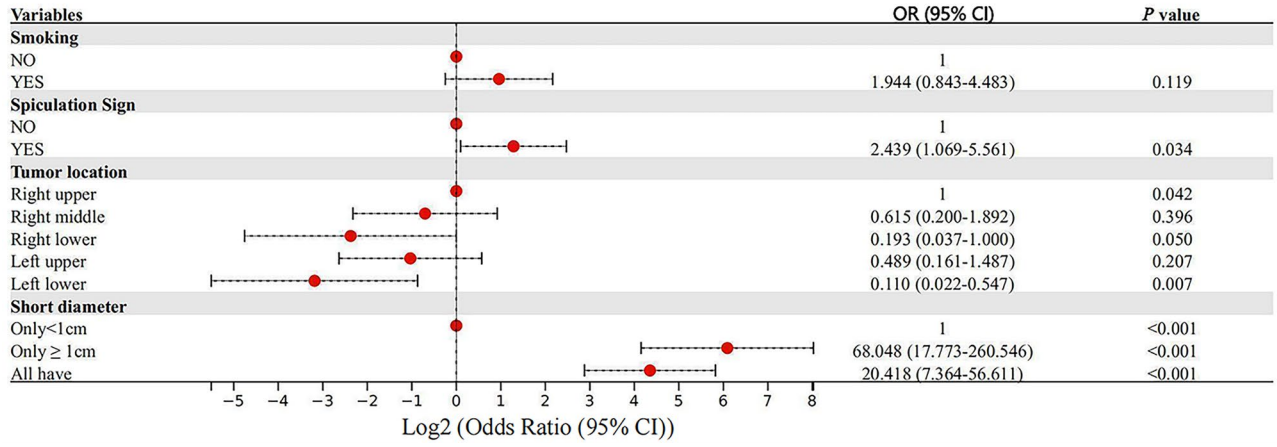

Fig. 3 Multivariate analysis results. CI, confidence interval

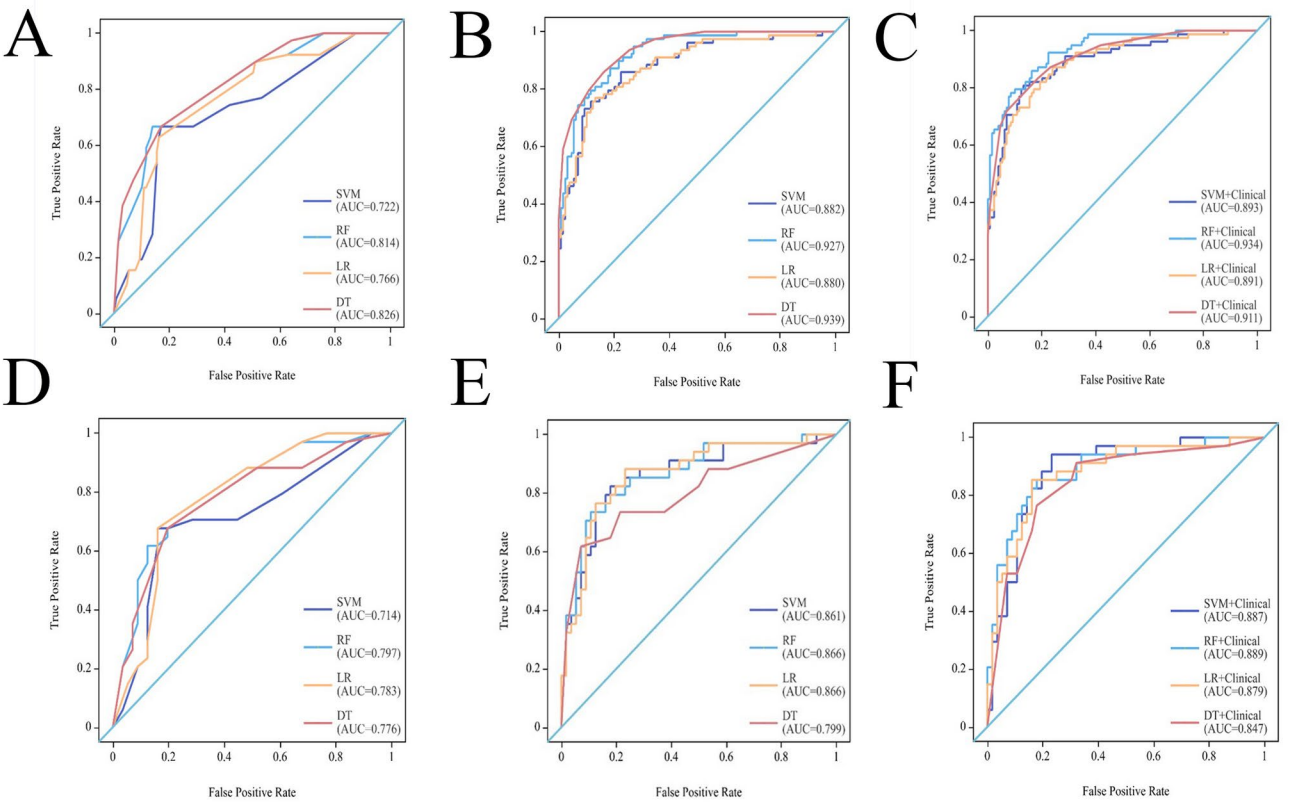

Fig. 4 ROC curves of models. ROC curves of clinical, radiomics, and combined models constructed by four algorithms in the training set (**A**, **B**, **C**) and the internal test set (**D**, **E**, **F**). DT, decision tree; LR, logistic regression; RF, random forest; ROC, receiver operating characteristic; SVM, support vector machine

* speckle noise_gldm_Dependence Non-Uniformity + 0.00345304 * log glrlm_log sigma-4-0-mm-3D-Run Entropy − 0.02049985 * original_shape_Surface Volume Ratio + 0.37681160.

Figure 6illustrates the distribution of *Radscore* across LNM and non-LNM groups within the two sets. In the training set, a significant difference in *Radscore* was observed between non-LNM and LNM groups (non-LNM: 0.29 ± 0.13; LNM: 0.53 ± 0.18; $p < 0.001$) (Fig. 6C). This finding was further validated in the internal test set, where non-LNM patients had a *Radscore* of 0.29 ± 0.13, and LNM patients had a *Radscore* of 0.52 ± 0.19 ($p < 0.001$) (Fig. 6D).

Four machine learning algorithms were then used to develop radiomics models based on the radiomics features selected by LASSO. The internal test set results indicated that both RF and LR models achieved an AUC of 0.866 (95% CI 0.785–0.946) (Fig. 4E). Sensitivity values

Table 2 Performance of clinical models on the training set and the internal test set

	Model	AUC (95% CI)	SEN	SPE	ACC	PRE	F1
Train	Clinical SVM	0.722 (0.650–0.794)	0.667	0.829	0.768	0.703	0.684
	Clinical RF	0.814 (0.757–0.872)	0.667	0.860	0.787	0.743	0.703
	Clinical LR	0.766 (0.701–0.832)	0.538	0.845	0.729	0.677	0.600
	Clinical DT	0.826 (0.772–0.881)	0.676	0.829	0.768	0.703	0.684
Internal test	Clinical SVM	0.714 (0.600-0.828)	0.676	0.804	0.756	0.676	0.676
	Clinical RF	0.797 (0.704–0.890)	0.618	0.839	0.756	0.700	0.656
	Clinical LR	0.783 (0.689–0.878)	0.618	0.839	0.756	0.700	0.656
	Clinical DT	0.776 (0.675–0.877)	0.676	0.804	0.756	0.676	0.676

ACC, accuracy; AUC, the area under the curve; CI: confidence interval; DT, decision tree; F1, F1 score; LR, logistic regression; PRE, precision; RF, random forest; SEN, sensitivity; SPE, specificity; SVM, support vector machine

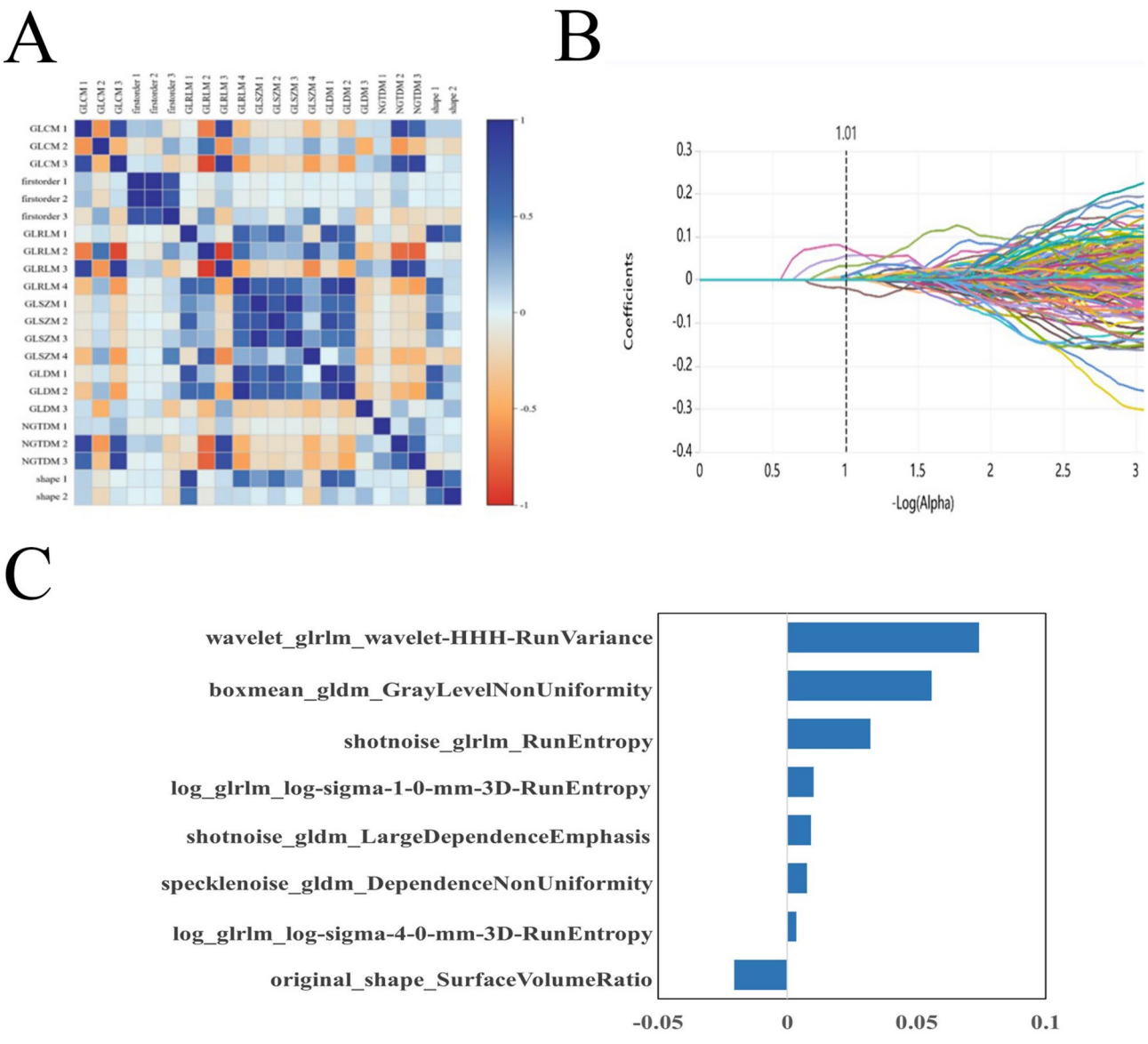

Fig. 5 Screening of radiomics features. Correlation analysis heatmap of radiomics features (**A**). LASSO regression feature selection process (**B**). Bar plot showing the regression coefficients for the hub feature of the radiomics model (**C**). LASSO is the least absolute shrinkage and selection operator

Table 3 Radiomics features selected in LASSO regression analysis

Feature Name	Category	Filter	Coefficient
Run Variance	GLRLM	wavelet-HHH	0.07408811
Gray Level Non-Uniformity	GLDM	box mean	0.05581313
Run Entropy	GLRLM	shot noise	0.03211861
Log sigma-1-0-mm-3D-Run Entropy	GLRLM	log	0.01013453
Large Dependence Emphasis	GLDM	shot noise	0.00919728
Dependence Non-Uniformity	GLDM	speckle noise	0.00762456
Log sigma-4-0-mm-3D-Run Entropy	GLRLM	log	0.00345304
Surface Volume Ratio	shape	original	-0.02049985

GLDM, gray level dependence matrix; GLRLM, gray level run length matrix

were 0.765 for RF and 0.706 for LR, specificity values were 0.821 for RF and 0.893 for LR, and accuracy values were 0.800 for RF and 0.822 for LR. Within the training set, RF obtained an AUC of 0.927 (95% CI 0.893–0.960) (Fig. 4B), with sensitivity of 0.808, specificity of 0.853, and accuracy of 0.836 (Table 4).

Development and test of the combined models

In the development of the models that combines the radiomics features selected by LASSO with three clinical characteristics, the RF algorithm demonstrated predictive performance as follows: The training set boasted an AUC of 0.934 (95% CI 0.901–0.966) with sensitivity of 0.821, specificity of 0.860, and accuracy of 0.845, whereas the internal test set recorded an AUC of 0.889 (95% CI 0.818–0.960) with sensitivity of 0.824, specificity of 0.821, and accuracy of 0.822 (Fig. 4C and F; Table 5). DeLong's test results for all models (clinical, radiomics,

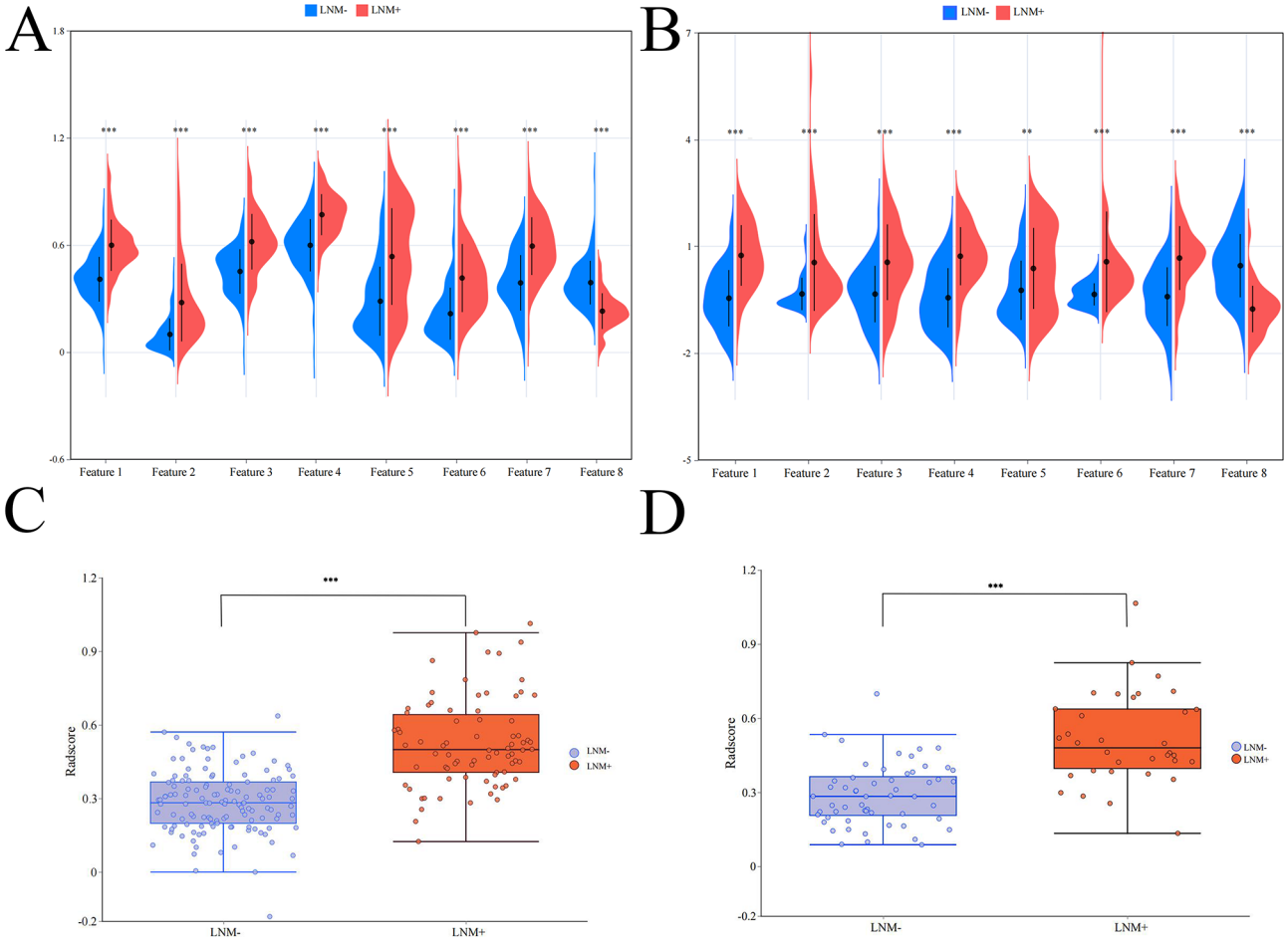

Fig. 6 Distribution of features. Violin plot comparison of characteristics between the non-LNM and LNM groups. Training set (**A**), internal test set (**B**). Distribution of *RadScore* across LNM and non-LNM groups within the training set (**C**) and the internal test set (**D**). **Feature 1** wavelet_glrlm_wavelet-HHH-Run Variance; **Feature 2** box mean_gldm_Gray Level Non-Uniformity; **Feature 3** shot noise_glrlm_Run Entropy; **Feature 4** log glrlm_log sigma-1-0-mm-3D-Run Entropy; **Feature 5** shot noise_gldm_Large Dependence Emphasis; **Feature 6** speckle noise_gldm_Dependence Non-Uniformity; **Feature 7** log glrlm_log sigma-4-0-mm-3D-Run Entropy; **Feature 8** original_shape_Surface Volume Ratio. LNM+, lymph node metastasis group; LNM-, non-lymph node metastasis group. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4 Performance of radiomics models on the training set and the internal test set

	Model	AUC (95% CI)	SEN	SPE	ACC	PRE	F1
Train	SVM	0.882 (0.834–0.930)	0.731	0.907	0.841	0.826	0.776
	RF	0.927 (0.893–0.960)	0.808	0.853	0.836	0.768	0.788
	LR	0.880 (0.832–0.928)	0.731	0.884	0.826	0.792	0.760
	DT	0.939 (0.910–0.969)	0.859	0.837	0.845	0.761	0.807
Internal test	SVM	0.861 (0.778–0.944)	0.647	0.875	0.789	0.759	0.698
	RF	0.866 (0.785 – 0.946)	0.765	0.821	0.800	0.722	0.743
	LR	0.866 (0.785–0.946)	0.706	0.893	0.822	0.800	0.750
	DT	0.799 (0.698–0.901)	0.735	0.625	0.667	0.543	0.625

ACC, accuracy; AUC, the area under the curve; CI: confidence interval; DT, decision tree; F1, F1 score; LR, logistic regression; PRE, precision; RF, random forest; SEN, sensitivity; SPE, specificity; SVM, support vector machine

Table 5 Performance of combined models on the training set and the internal test set

	Model	AUC (95% CI)	SEN	SPE	ACC	PRE	F1
Train	SVM + Clinical	0.893 (0.846–0.939)	0.846	0.752	0.787	0.673	0.750
	RF + Clinical	0.934 (0.901–0.966)	0.821	0.860	0.845	0.780	0.800
	LR + Clinical	0.891 (0.845–0.937)	0.782	0.829	0.812	0.735	0.758
	DT + Clinical	0.911 (0.873–0.950)	0.872	0.767	0.807	0.694	0.773
Internal test	SVM + Clinical	0.887 (0.817–0.957)	0.912	0.768	0.822	0.705	0.795
	RF + Clinical	0.889 (0.818–0.960)	0.824	0.821	0.822	0.737	0.778
	LR + Clinical	0.879 (0.803–0.954)	0.824	0.839	0.833	0.757	0.789
	DT + Clinical	0.847 (0.762–0.931)	0.912	0.679	0.767	0.633	0.747

ACC, accuracy; AUC, the area under the curve; CI: confidence interval; DT, decision tree; F1, F1 score; LR, logistic regression; PRE, precision; RF, random forest; SEN, sensitivity; SPE, specificity; SVM, support vector machine

and combined) are compiled in Supplementary Material 5.

The performance of the three models constructed using RF is compared in Fig. 7A and B. DeLong's test results for the ROC curves of the three RF-based models—the combined model, the radiomics model, and the clinical model (combined model vs. radiomics model [p-value = 0.112], combined model vs. clinical model [p-value = 0.084], radiomics model vs. clinical model [p-value = 0.230]) (Table 6).

Calibration curves demonstrated strong concordance between the actual LNM status and the predictive probabilities generated by the combined model based on RF (Fig. 7C). The Hosmer-Lemeshow goodness of fit test results were as follows: the combined model had a p-value of 0.779, the clinical model had a p-value of 0.636, and the radiomics model had a p-value of 0.227. DCA curve analysis showed that, within the majority of threshold probability ranges, the net benefit of the RF-based combined model in predicting MLNM was higher than that of the clinical and radiomics models (Fig. 7D).

In the SHAP plots for the training set (Fig. 8A) and the internal test set (Fig. 8B), the feature contributions of the RF-based combined model are displayed. The confusion matrix (Fig. 9A) shows the classification performance of the RF-based combined model in the internal test set.

The RF-based combined model achieved an AUC of 0.883 (95% CI 0.795–0.972) in the independent test set (Fig. 9B), with sensitivity of 0.829, specificity of 0.739, accuracy of 0.793, precision of 0.829, and FI score of 0.829. The calibration curve (Fig. 9C) and DCA (Fig. 9D) for the independent test set are presented, with the Hosmer-Lemeshow test yielding a p-value of 0.675.

Discussion

The status of LNs is a critical determinant of prognosis for patients with NSCLC, necessitating thorough removal of suspected metastatic LNs during surgery. In our research, we observed that LND at station 4 in patients with pN0-pN2 stage NSCLC, particularly S4L, has not received adequate attention. In the absence of guidelines for clearing station 4 of LNs in NSCLC, there is an urgent need to develop non-invasive LN assessment tools. While radiomics has shown promise in predicting LNM, prior studies have predominantly focused on binary LN status or single-station analyses [3, 10, 20]. These studies have laid the groundwork for evaluating individual LN stations and guiding decisions on LND extent, thus providing a solid foundation for our research.

This study developed 12 prediction models using four machine learning algorithms to assess the risk of metastasis to the station 4 in NSCLC patients at pN0-pN2 stages. The combined model significantly outperformed single models and demonstrated robust clinical applicability in an independent test set (AUC 0.883, 95% CI 0.795–0.972). Although DeLong's test didn't find statistically significant differences among the three models based on RF, this might be due to factors such as dataset

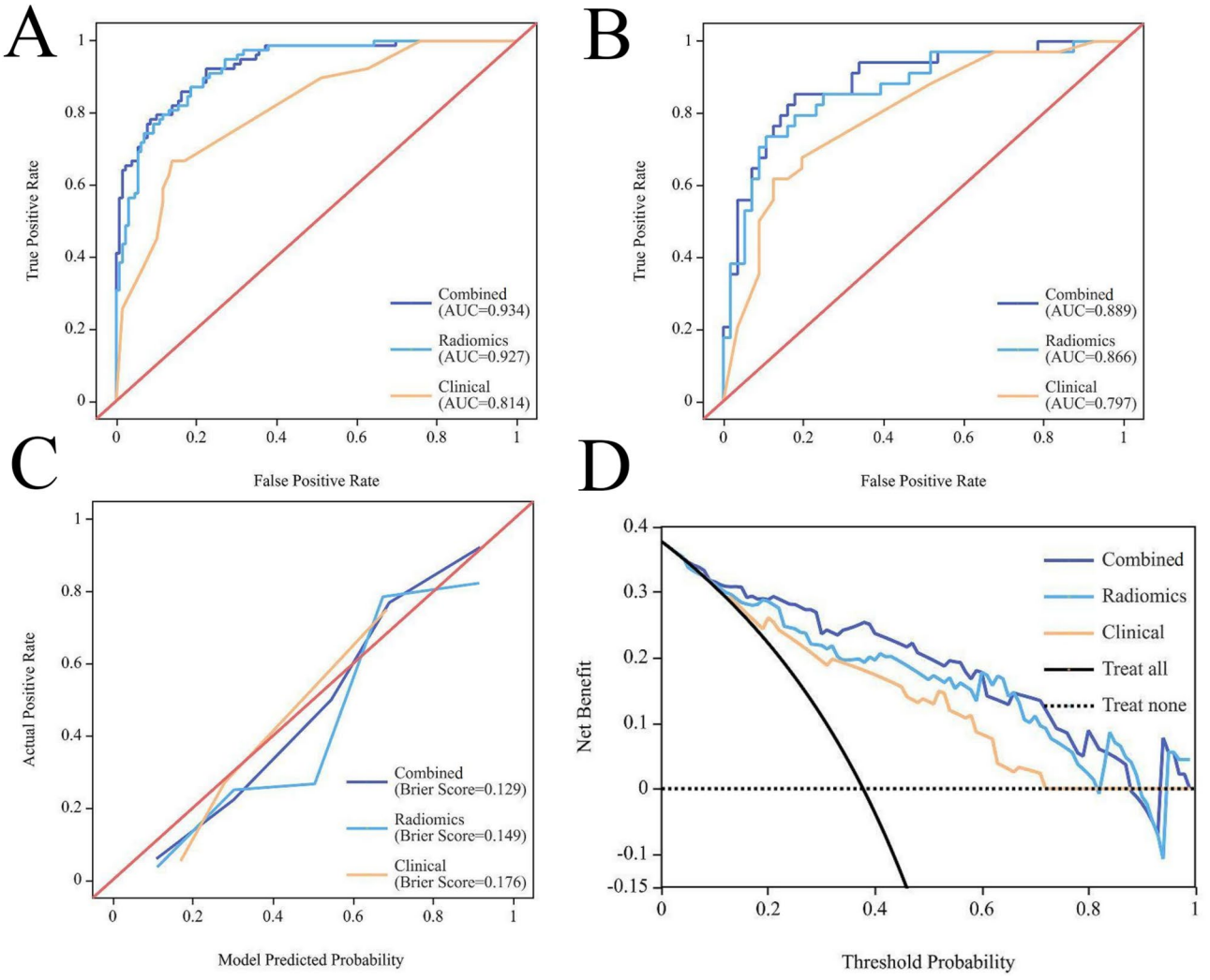

Fig. 7 ROC curves of three models constructed by RF. ROC curves of the combined model, radiomics model, and clinical model constructed by RF in the training set (**A**) and internal test set (**B**), respectively. Calibration curves (**C**), DCA curves (**D**) of the three models in the internal test set. DCA, decision curve analysis; ROC, receiver operating characteristic

Table 6 DeLong's test for radiomics model, clinical model, and combined model based on RF

	***p* value**
Radiomics VS Combined	0.112
Radiomics VS Clinical	0.230
Combined VS Clinical	0.084

complexity, test specifics, or model training refinement needs. Nonethless, the combined models consistent superiority across metrics highlights its potential for future enhancement. This work established a diverse model system, emphasizing the necessity of independent validation and the methodology for constructing statistically robust and clinically applicable predictive tools. As a clinical diagnostic evaluation study, our findings hold potential value for optimizing clinical decision-making pathways for NSCLC patients at pN0-pN2 stage.

Traditional imaging methods have exhibited suboptimal performance in preoperative MLNM assessment. While PET-CT exhibits promise in detecting metastatic LNs, its application in preoperative examinations is constrained by radioactivity, high cost, and low prevalence [35, 36]. Additionally, the specificity and accuracy of PET-CT in diagnosing NSCLC LNM can be compromised by factors such as tumor size and inflammation [37]. This underscores the imperative for non-invasive, cost-effective tools like radiomics to improve diagnostic accuracy and guide clinical decision-making.

Our predictive model enhances accuracy by integrating preoperative clinical data. Key findings indicate that tumor location, spiculation, and LN short diameter are crucial in forecasting MLNM at station 4 in patients with NSCLC at pN0-pN2 stages. Notably, the upper lobe shows a higher LNM risk than the middle and lower lobes

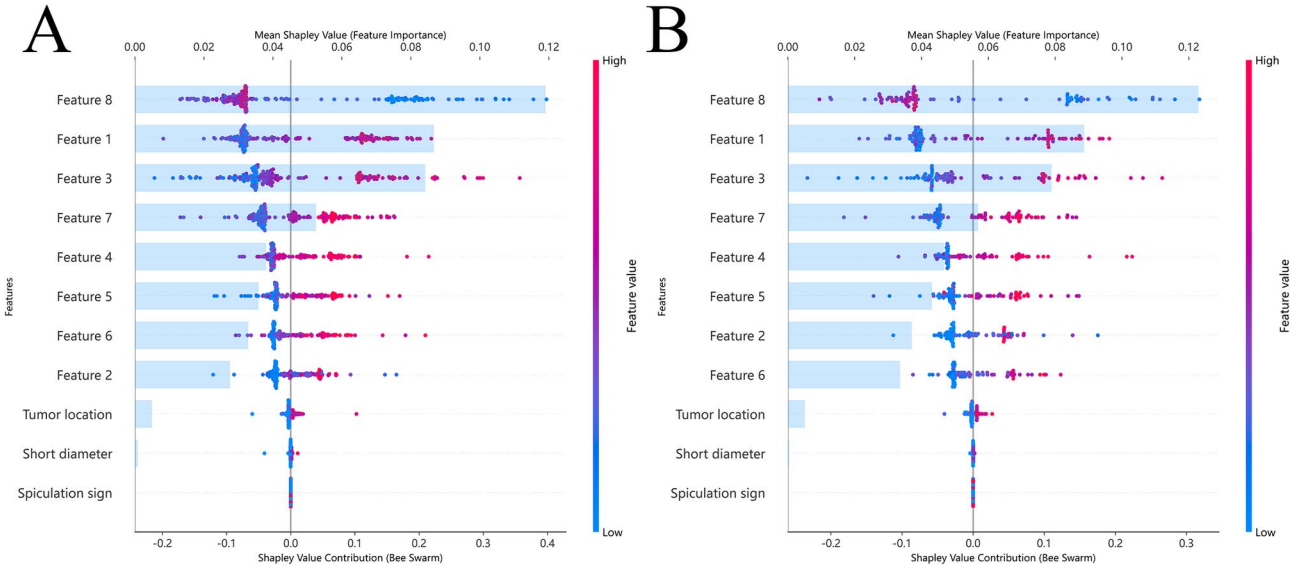

Fig. 8 SHAP plots of the RF-based combined model. Feature contribution analysis in the training set (**A**) and internal test set (**B**). **Feature 1** wavelet_glrlm_wavelet-HHH-Run Variance; **Feature 2** box mean_gldm_Gray Level Non-Uniformity; **Feature 3** shot noise_glrlm_Run Entropy; **Feature 4** log glrlm_log sigma-1-0-mm-3D-Run Entropy; **Feature 5** shot noise_gldm_Large Dependence Emphasis; **Feature 6** speckle noise_gldm_Dependence Non-Uniformity; **Feature 7** log glrlm_log sigma-4-0-mm-3D-Run Entropy; **Feature 8** original_shape_Surface Volume Ratio. Each point represents a patient sample. Color indicates the feature value (red: high, blue: low). Horizontal position shows the SHAP value direction and magnitude: rightward points (positive SHAP values) increase positive risk, leftward points (negative values) decrease risk. Features are ranked by their global importance (mean absolute SHAP value across all samples)

in station 4, aligning with anatomical lymphatic drainage patterns [3, 5]. However, the nomogram by Xiong et al. [3] for predicting LNM did not identify tumor location as a risk factor for station 4 MLNM, potentially due to their limited sample size. DuComb et al. [38] in a retrospective analysis of 332 T1 NSCLC patients also found LNM no association between LNM and tumor location, possibly due to differing criteria—our study was unrestricted in NSCLC tumor size, unlike DuComb et al.'s focus on T1 tumors (8–30 mm).

The presence of spiculation correlates with increased LNM risk, which is likely due to the tumor's aggressive infiltration of surrounding tissues [25]. In addition, increased LN short diameter emerged as an independent risk factor for MLNM at station 4 in patients with NSCLC at pN0-pN2 stages. Notably, 13.9–26.9% of NSCLC patients at pN0-pN2 stages exhibited metastasis in radiologically normal LNs at station 4, consistent with prior reports [39–41]. This finding underscores the necessity for comprehensive mapping of station 4 LNs to ensure thorough metastatic evaluation.

Pathological type or histological grade is a significant predictor of LNM, but it requires invasive biopsy or surgery to determine [23, 39, 42]. We posit that postoperative diagnoses should not be considered as preoperative parameters because physicians cannot make diagnoses as accurate as pathological results from CT images alone. This is also the reason why our study used all NSCLC patients to develop radiomics models.

Our clinical models, with its limited feature set, exhibit inferior predictive power and sensitivity compared to radiomics models, highlighting the richer information provided by radiomics on LNs and their microenvironments. GLRLM measures the continuity of pixel lengths, shedding light on the uniformity and complexity of LNs texture [43]. Texture features derived from GLRLM, including run variance and run entropy, effectively discriminated metastatic LNs by quantifying structural complexity. Multiscale filtering further demonstrated differential sensitivity to micro- versus macrostructural changes, with smaller sigma values capturing subtle density changes in early metastatic foci within LNs [44]. This contrasts with the limited sensitivity of clinical models in pN0-pN2 patients and indirectly highlights the advantages of radiomics in detecting subclinical lesions. GLDM delves into the interdependencies and spatial configurations among gray levels, revealing intricate patterns and structures [43]. Spatial heterogeneity, as measured by GLDM metrics, was markedly elevated in metastatic LNs, while a reduced surface volume ratio (trending toward sphericity) reflects the expansive growth pattern of metastatic LNs. This pathophysiological linkage corresponds to recognized LNs morphological transformations during metastatic dissemination [40].

SHAP analysis reveals that surface volume ratio is a dominant predictor, with lower SHAP values indicative of a lower likelihood of metastasis to station 4 LNs in NSCLC patients at pN0-pN2 stages. This parameter,

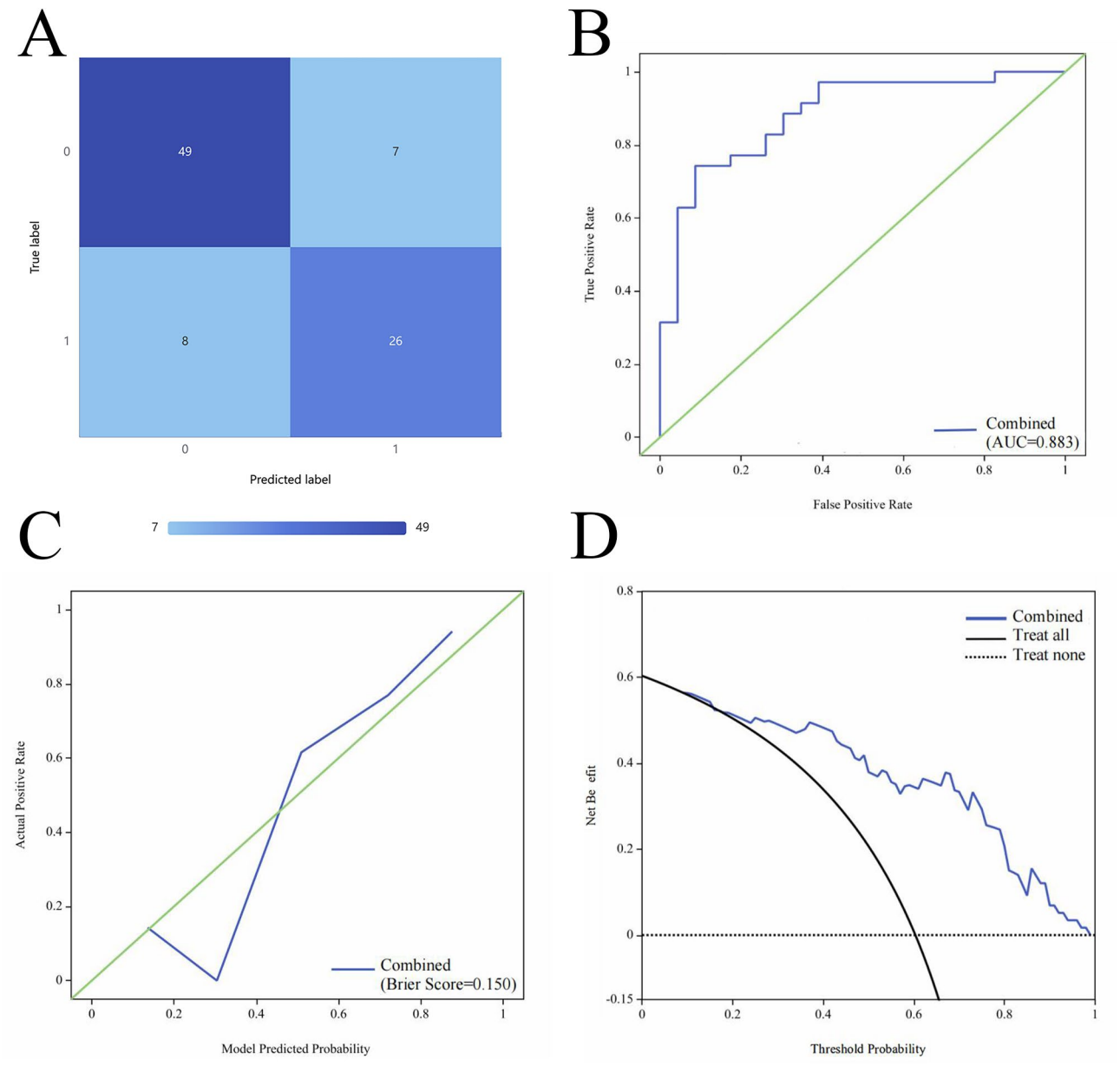

Fig. 9 Model evaluation. The confusion matrix of the RF-based combined model on the internal test set (**A**), and the ROC curve (**B**), calibration curve (**C**), and DCA curve (**D**) of the same model on the independent test set. **0** represents the absence of LNM, while **1** indicates the presence of LNM. DCA, decision curve analysis; ROC, receiver operating characteristic

which integrates 3D shape information, likely surpasses single morphological metrics like short diameter by inherently capturing multidimensional spatial characteristics. This reduces feature collinearity and measurement bias associated with traditional clinical models, thereby enhancing sensitivity.

This study demonstrates methodological rigor through the systematic development of 12 machine learning models using four distinct algorithms (DT, LR, RF, SVM), with the RF algorithm exhibiting superior computational efficiency and reliability in high-dimensional feature spaces [45]. Preprocessing techniques enhanced feature discriminability, while validation across diverse imaging platforms confirmed robust generalizability.

The limitations include constrained sample size and absence of external validation within Chinese populations, necessitating future large-scale, multi-ethnic studies. Secondly, semi-automatic segmentation, though reproducible, introduces operator-dependent variability, prompting planned integration of deep learning solutions. Thirdly, the exclusive use of venous-phase CT may overlook temporal heterogeneity, warranting multi-phase analyses. Fourthly, as a retrospective design limited to patients with uniformly positive or negative

station 4 LNs, there may be potential selection bias in this approach.

Conclusions

In conclusion, our combined model based on RF represents a promising non-invasive predictive tool, integrating radiomics features from CT images with independent clinical risk factors, which provides a non-invasive means to effectively identify metastasis to the station 4 mediastinal LNs in NSCLC patients at pN0-pN2 stage. It holds significant potential for practical application. Future studies should focus on conducting large-sample research and validation to further refine the model and enhance its predictive efficiency.

Abbreviations

MLNM	Mediastinal lymph node metastasis
NSCLC	Non-small cell lung cancer
CT	Computed tomography
DT	Decision tree
LR	Logistic regression
RF	Random forest
SVM	Support vector machine
ROC	Receiver operating characteristic
DCA	Decision curve analysis
AUC	Area under the curve
LNM	Lymph node metastasis
LNs	Lymph nodes
LND	Lymph node dissection
S4R	Station 4R
S4L	Station 4 L
SLND	Systematic lymph node dissection
EUS	Endoscopic ultrasonography
PET-CT	Positron emission tomography-computed tomography ROIs Regions of interest
DICOM	Digital imaging and communications in medicine format
ICC	Interclass correlation coefficient
LASSO	Least absolute shrinkage and selection operator
LASSO	Radscore Radiomics score
CI	Confidence interval
GLRLM	Gray level run length matrix
GLDM	Gray level dependency matrix
SHAP	Shapley additive explanations

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01686-1.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements

We sincerely thank all the participants involved in this study and the uAI Research Portal system for providing partial technical support for this study.

Author contributions

ZH, CY, and KYR designed the study. ZH and KYR reviewed the literature. KYR, LHX, and QYF carried out clinical data analysis. KYR, LM, XXZ, and QKX conducted statistical analysis. ZH, CY, and KYR drafted and revised the manuscript. All authors reviewed the manuscript, and ZH supervised the study. All authors contributed to the article and approved the submitted version.

Funding

Not applicable.

Data availability

The combined features datasets supporting the findings of this study are openly available in [Mendeley Data] at https://data.mendeley.com/datasets/vrpgfc99w6/1 and via DOI at http://doi.org/10.17632/vrpgfc99w6.1.

Declarations

Ethnics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. The Institutional Review Board of the Ethics Committee of Qilu Hospital, Shandong University (KYLL_202404_041) approved the research protocol for this retrospective study and waived the requirement for written informed consent from the patients.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 8 February 2025 / Accepted: 22 April 2025
Published online: 04 June 2025

References

1. Siegel RL, Miller KD, Wagle NS et al. Cancer statistics, 2023. CA: A Cancer journal for clinicians, 2023, 73(1): 17–48.https://doi.org/10.3322/caac.21763
2. Fourdrain A, Epailly J, Blanchard C, et al. Lymphatic drainage of lung cancer follows an intersegmental pathway within the visceral pleura. Lung Cancer. 2021;154:118–23. https://doi.org/10.1016/j.lungcan.2021.02.023.
3. Xiong L, Wei Y, Zhou X, et al. Development and validation of nomograms based on clinical characteristics and CT reports for the preoperative prediction of precise lymph node dissection in lung cancer. Lung Cancer. 2022;172:35–42. https://doi.org/10.1016/j.lungcan.2022.08.009.
4. Riely GJ, Wood DE, Ettinger DS, et al. Non–Small cell lung cancer, version 4.2024. J Natl Compr Canc Netw. 2024;22(4):249–74. https://doi.org/10.6004/jnccn.2204.0023.
5. Zhou D, Yue D, Zhang Z, et al. Prognostic significance of 4R lymph node dissection in patients with right primary non-small cell lung cancer. World J Surg Oncol. 2022;20(1):222. https://doi.org/10.1186/s12957-022-02689-w.
6. De Ruysscher DKM, Decaluwé H. 4L lymph node involvement in Left-Sided lung cancer: unique or not??. J Clin Oncol. 2018;36(29):2907–8. https://doi.org/10.1200/jco.2018.79.3299.
7. Lucchi M, Aprile V. Not one less, the role of 4L in left-sided lung cancer: the message from the Polish experience. Eur J Cardiothorac Surg. 2021;60(5):1210–1. https://doi.org/10.1093/ejcts/ezab358.
8. Wo Y, Li H, Zhang Y, et al. The impact of station 4L lymph node dissection on short-term and long-term outcomes in non-small cell lung cancer. Lung Cancer. 2022;170:141–7. https://doi.org/10.1016/j.lungcan.2022.06.018.
9. Peng L, Huang K, Shang Q, et al. The prognostic value of 4L lymph node dissection in left-side operable non-small-cell lung cancer: a meta-analysis. Updates Surg. 2023;21(10):23–32. https://doi.org/10.1007/s13304-023-01694-2.
10. Xu L, Guo J, Qi S, et al. Development and validation of a nomogram model for the prediction of 4L lymph node metastasis in thoracic esophageal squamous cell carcinoma. Front Oncol. 2022;12:887047. https://doi.org/10.3389/fonc.2022.887047.
11. Hanaoka J, Yoden M, Okamoto K, et al. Mediastinal lymph node evaluation, especially at station 4L, in left upper lobe lung cancer. J Thorac Disease. 2022;14(9):3321–34. https://doi.org/10.21037/jtd-22-537.
12. Gajra A, Newman N, Gamble GP, et al. Effect of number of lymph nodes sampled on outcome in patients with stage I Non–Small-Cell lung Cancer. J Clin Oncol. 2003;21(6):1029–34. https://doi.org/10.1200/jco.2003.07.010.

13. Jiang C, Zhang Y, Fu F, et al. A shift in paradigm: selective lymph node dissection for minimizing oversurgery in early stage lung Cancer. J Thorac Oncol. 2024;19(1):25–35. https://doi.org/10.1016/j.jtho.2023.09.1443.
14. Kuroda H, Sakao Y, Mun M, et al. Therapeutic value of lymph node dissection for right middle lobe non-small-cell lung cancer. J Thorac Disease. 2016;8(5):795–802. https://doi.org/10.21037/jtd.2016.03.30.
15. Zhao K, Mei J, Hu B, et al. Complete dissection of right paratracheal lymph nodes (stations 2R and 4R) is critical to improve the prognosis of lung cancer patients: A retrospective cohort study. Thorac Cancer. 2023;14(33):3309–16. https://doi.org/10.1111/1759-7714.15122.
16. Yan H, Zhao J, Zuo H, et al. Dual-Region computed tomography Radiomics-Based machine learning predicts subcarinal lymph node metastasis in patients with Non-small cell lung Cancer. Ann Surg Oncol. 2024;31(8):5011–20. https://doi.org/10.1245/s10434-024-15197-w.
17. Mascalchi M, Zompatori M. Mediastinal lymphadenopathy in lung Cancer screening: A red flag. Radiology. 2022;302(3):695–6. https://doi.org/10.1148/radiol.212501.
18. Lee GM, Stowell JT, Pope K, et al. Lymphatic pathways of the thorax: predictable patterns of spread. Am J Roentgenol. 2021;216(3):649–58. https://doi.org/10.2214/ajr.20.23523.
19. Dong M, Hou G, Li S, et al. Preoperatively estimating the malignant potential of mediastinal lymph nodes: A pilot study toward Establishing a robust radiomics model based on Contrast-Enhanced CT imaging. Front Oncol. 2021;10:558428. https://doi.org/10.3389/fonc.2020.558428.
20. Sun Z, Jiang Y, Chen C, et al. Radiomics signature based on computed tomography images for the preoperative prediction of lymph node metastasis at individual stations in gastric cancer: A multicenter study. Radiother Oncol. 2021;165:179–90. https://doi.org/10.1016/j.radonc.2021.11.003.
21. 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.
22. Singh S, Mohajer B, Wells SA, et al. Imaging Genomics and Multiomics: A Guide for Beginners Starting Radiomics-Based Research. Acad Radiol. 2024;31(6):2281–91. https://doi.org/10.1016/j.acra.2024.01.024.
23. Cong M, Yao H, Liu H, et al. Development and evaluation of a venous computed tomography radiomics model to predict lymph node metastasis from non-small cell lung cancer. Medicine. 2020;99(18):e20074. https://doi.org/10.1097/md.0000000000020074.
24. Sha X, Gong G, Qiu Q, et al. Discrimination of mediastinal metastatic lymph nodes in NSCLC based on radiomic features in different phases of CT imaging. BMC Med Imaging. 2020;20(1):12. https://doi.org/10.1186/s12880-020-0416-3.
25. Xie Z, Yang Y, Niu Z, et al. Preoperative computed tomography semantic features in predicting lymph node metastasis of part-solid nodules in non-small cell lung cancer: a multicenter retrospective study. Quant Imaging Med Surg. 2024;14(7):5151–63. https://doi.org/10.21037/qims-23-1631.
26. Hsu JS, Han IT, Tsai TH, et al. Pleural tags on CT scans to predict visceral pleural invasion of Non-Small cell lung Cancer that does not abut the pleura. Radiology. 2016;279(2):590–6. https://doi.org/10.1148/radiol.2015151120.
27. Kocak B, Baessler B, Bakas S, et al. CheckList for evaluation of radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights into Imaging. 2023;14(1):75. https://doi.org/10.1186/s13244-023-01415-8.
28. Kocak B, Borgheresi A, Ponsiglione A, et al. Explanation and elaboration with examples for CLEAR (CLEAR-E3): an EuSoMII radiomics auditing group initiative. Eur Radiol Experimental. 2024;8(1):72. https://doi.org/10.1186/s41747-024-00471-z.
29. Monti CB, Ambrogi F, Sardanelli F. Sample size calculation for data reliability and diagnostic performance: a go-to review. Eur Radiol Experimental. 2024;8(1):79. https://doi.org/10.1186/s41747-024-00474-w.
30. Zhong J, Liu X, Lu J, et al. Overlooked and underpowered: a meta-research addressing sample size in radiomics prediction models for binary outcomes. Eur Radiol. 2025;35(3):1146–56. https://doi.org/10.1007/s00330-024-11331-0.
31. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2018;38(7):1276–96. https://doi.org/10.1002/sim.7992.
32. Liu X, Xu Y, Shu J et al. Preoperative CT and radiomics nomograms for distinguishing bronchiolar adenoma and Early-Stage lung adenocarcinoma. Academic Radiology, 2024. https://doi.org/10.1016/j.acra.2024.08.047
33. Toker A, Tanju S, Ziyade S, et al. Alternative paratracheal lymph node dissection in left-sided hilar lung cancer patients: comparing the number of lymph nodes dissected to the number of lymph nodes dissected in right-sided mediastinal dissections. Eur J Cardiothorac Surg. 2011;39(6):974–80. https://doi.org/10.1016/j.ejcts.2010.09.013.
34. Wang P, Chen K, Han Y, et al. Prediction model based on radiomics and clinical features for preoperative lymphovascular invasion in gastric cancer patients. Future Oncol. 2023;19(23):1613–26. https://doi.org/10.2217/fon-2022-1025.
35. Sui H, Liu L, Li X, et al. CT-based radiomics features analysis for predicting the risk of anterior mediastinal lesions. J Thorac Disease. 2019;11(5):1809–18. https://doi.org/10.21037/jtd.2019.05.32.
36. Kim YK, Lee KS, Kim BT, et al. Mediastinal nodal staging of nonsmall cell lung cancer using integrated 18F-FDG PET/CT in a tuberculosis-endemic country. Cancer. 2007;109(6):1068–77. https://doi.org/10.1002/cncr.22518.
37. Schmidt-Hansen M, Baldwin DR, Hasler E, et al. PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer. Cochrane Database Syst Rev. 2014;11:CD009519. https://doi.org/10.1002/14651858.CD009519.
38. DuComb EA, Tonelli BA, Tuo Y, et al. Evidence for expanding invasive mediastinal staging for peripheral T1 lung tumors. Chest. 2020;158(5):2192–9. https://doi.org/10.1016/j.chest.2020.05.607.
39. Cong M, Feng H, Ren JL, 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;139:73–9. https://doi.org/10.1016/j.lungcan.2019.11.003.
40. Sugai K, Sekine Y, Kawamura T, et al. Sphericity of lymph nodes using 3D-CT predicts metastasis in lung cancer patients. Cancer Imaging. 2023;23(1):124. https://doi.org/10.1186/s40644-023-00635-x.
41. Suemitsu R, Takeo S, Matsuzawa H, et al. Can a thoracic surgeon identify lymph node metastases during surgery based on their size? Analysis of 844 metastatic and 10,462 nonmetastatic lymph nodes. J Thorac Oncol. 2010;5(3):349–53. https://doi.org/10.1097/JTO.0b013e3181c6b86b.
42. Dai M, Wang N, Zhao X, et al. Value of presurgical 18F-FDG PET/CT radiomics for predicting mediastinal lymph node metastasis in patients with lung adenocarcinoma. Cancer Biotherapy Radiopharmaceuticals. 2024;39(8):600–10. https://doi.org/10.1089/cbr.2022.0038.
43. Xie H, Song C, Jian L, et al. A deep learning-based radiomics model for predicting lymph node status from lung adenocarcinoma. BMC Med Imaging. 2024;24(1):121. https://doi.org/10.1186/s12880-024-01300-w.
44. Abranches DO, Maginn EJ, Colón YJ. Stochastic machine learning via Sigma profiles to build a digital chemical space. Proc Natl Acad Sci U S A. 2024;121(31):e2404676121. https://doi.org/10.1073/pnas.
45. Uddin S, Khan A, Hossain ME, et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inf Decis Mak. 2019;19(1):281. https://doi.org/10.1186/s12911-019-1004-8.

Publisher's note