



# Preoperatively Estimating the Malignant Potential of Mediastinal Lymph Nodes: A Pilot Study Toward Establishing a Robust Radiomics Model Based on Contrast-Enhanced CT Imaging

OPEN ACCESS

**Edited by:**

Meiyappan Solaiyappan,  
Johns Hopkins University,  
United States

**Reviewed by:**

Guihua Jiang,  
Guangdong Second Provincial  
General Hospital, China  
Long Jiang Zhang,  
Nanjing General Hospital of Nanjing  
Military Command, China  
Zaiyi Liu,  
Guangdong Provincial People's  
Hospital, China  
Dong Li,  
Tianjin Medical University General  
Hospital, China

**\*Correspondence:**

Lina Zhang  
zhanglinda@163.com  
Ke Xu  
kxu@cmu.edu.cn

**Specialty section:**

This article was submitted to  
Cancer Imaging and  
Image-directed Interventions,  
a section of the journal  
Frontiers in Oncology

**Received:** 02 May 2020

**Accepted:** 18 November 2020

**Published:** 08 January 2021

**Citation:**

Dong M, Hou G, Li S, Li N, Zhang L  
and Xu K (2021) 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.* 10:558428.  
doi: 10.3389/fonc.2020.558428

Mengshi Dong<sup>1</sup>, Gang Hou<sup>2</sup>, Shu Li<sup>1</sup>, Nan Li<sup>1</sup>, Lina Zhang<sup>1\*</sup> and Ke Xu<sup>1\*</sup>

<sup>1</sup> Department of Radiology, The First Affiliated Hospital of China Medical University, Shenyang, China, <sup>2</sup> Institute of Respiratory Disease, The First Affiliated Hospital of China Medical University, Shenyang, China

**Purpose:** To establish and validate a radiomics model to estimate the malignancy of mediastinal lymph nodes (LNs) based on contrast-enhanced CT imaging.

**Method:** In total, 201 pathologically confirmed mediastinal LNs from 129 patients were enrolled and assigned to training and test sets. Radiomics features were extracted from the region of interest (ROI) delineated on venous-phase CT imaging of LN. Feature selection was performed with least absolute shrinkage and selection operator (LASSO) binary logistic regression. Multivariate logistic regression was performed with the backward stepwise elimination. A model was fitted to associate mediastinal LN malignancy with selected features. The performance of the model was assessed and compared to that of five other machine learning algorithms (support vector machine, naive Bayes, random forest, decision tree, K-nearest neighbor) using receiver operating characteristic (ROC) curves. Calibration curves and Hosmer-Lemeshow tests were used to assess the calibration degree. Decision curve analysis (DCA) was used to assess the clinical usefulness of the logistic regression model in both the training and test sets. Stratified analysis was performed for different scanners and slice thicknesses.

**Result:** Among the six machine learning methods, the logistic regression model with the eight strongest features showed a significant association with mediastinal LN status and the satisfactory diagnostic performance for distinguishing malignant LNs from benign LNs. The accuracy, sensitivity, specificity and area under the ROC curve (AUC) were 0.850/0.803, 0.821/0.806, 0.893/0.800, and 0.922/0.850 in the training/test sets, respectively. The Hosmer-Lemeshow test showed that the P value was > 0.05, indicating good calibration, and the calibration curves showed good agreement between the classifications and actual observations. DCA showed that the model would obtain more benefit when the threshold probability was between 30% and 90% in the test set. Stratified analysis showed that the performance was not affected

by different scanners or slice thicknesses. There was no significant difference (DeLong test,  $P > 0.05$ ) between any two subgroups, which showed the generalization of the radiomics score across different factors.

**Conclusion:** The model we built could help assist the preoperative estimation of mediastinal LN malignancy based on contrast-enhanced CT imaging, with stability for different scanners and slice thicknesses.

**Keywords:** mediastinal lymph nodes, malignant, benign, radiomics, machine learning, classification

## INTRODUCTION

Mediastinal lymph nodes (LNs) are the most common metastatic targets of lung cancer, esophageal cancer and other malignant tumors. LN metastasis has been suggested to be very important for tumor staging, treatment plan selection and prognosis prediction (1–4). Preoperative staging of the mediastinum is an essential task (5). Preoperative staging is helpful for clinicians to fully understand the patient's condition and to make better diagnosis and treatment decisions. However, it is difficult to estimate mediastinal LN metastasis before surgery in the clinic. Metastasis does not necessarily cause lymphadenopathy. Some LNs less than 1 cm are finally confirmed as metastasis, while some with obvious enlargement are confirmed as chronic lymphadenitis or reactive hyperplasia.

Endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) is the most effective diagnostic method for estimating the malignancy of mediastinal LNs (6–9). However, EBUS-TBNA is not popular enough in developing countries. It is also an invasive procedure with the possibility of complications (10–12) and false negative results (13–16). It is meaningful to find a method to assist clinicians in determining the status of mediastinal LNs preoperatively.

It is worth noting that CT scan can be performed before surgery. Some researchers point out that radiomics is very promising in preoperative staging of the mediastinum (5). “Radiomics” refers to an approach that can extract innumerable quantitative features using advanced feature analysis with high throughput from digital medical images, including computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET) images (17). Radiomics has been indicated to be effective for the diagnosis, treatment evaluation and prognosis prediction of tumors (18–20), as well as tumor phenotype decoding (18, 21). Recent evidence suggests that radiomics features have the capability to predict the distant metastasis of lung adenocarcinoma (22) and LN metastasis of colorectal cancer (23). Several researchers have attempted to estimate the malignancy of primary lung lesions with radiomics (22, 24) or mediastinal LNs with texture (25, 26). The results were promising. A recent research confirms that the CT radiomics models show good performances for the diagnosis of LN metastasis in non-small cell lung cancer patients (27). However, with different equipment and scanning protocols in different studies, the repeatability and applicability of the model is relatively low.

In our study, a robust model with validation by stratified analysis (28) was established for malignant estimation based on the radiomics analysis of mediastinal LNs on contrast-enhanced CT. We hypothesized that the robust model could be effective in distinguishing the malignancy of LNs, and it could be generally applied on different scanners or slice thicknesses with stable performance.

## MATERIALS AND METHODS

### Patients

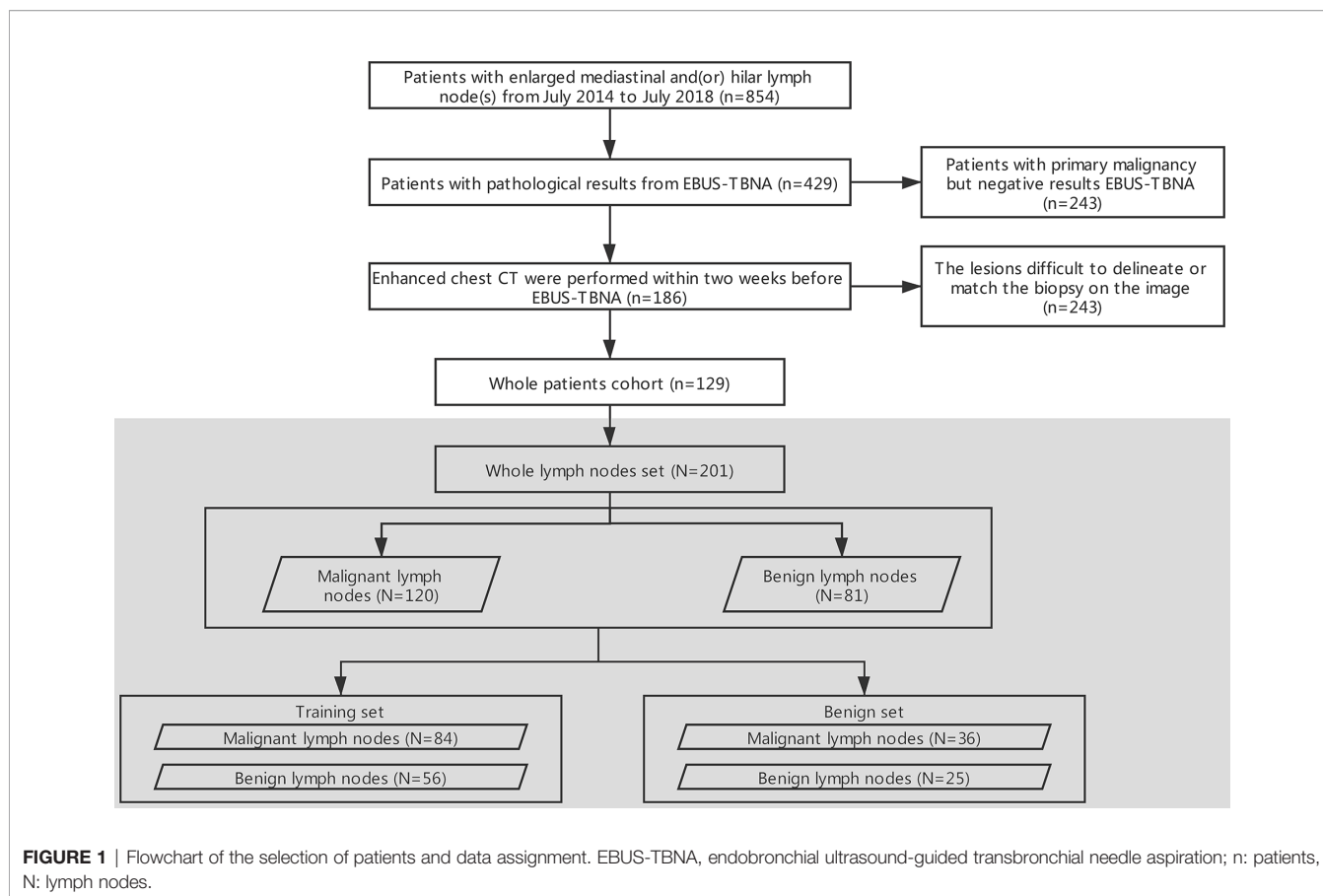
The retrospective study was approved by the institutional review board of the First Affiliated Hospital of China Medical University, and the requirement for informed consent was waived. A review of medical records was performed in accordance with the guidelines of the ethical review committee of the institution. One hundred and twenty-nine patients (73 men, 56 women, age: 18–93, mean age: 56, median: 57) with both EBUS-TBNA and contrast-enhanced chest CT from July 2014 to July 2018 were included. The inclusion criteria were as follows: (i) patients with pathological results from EBUS-TBNA; and (ii) contrast-enhanced chest CT performed within two weeks before EBUS-TBNA. The exclusion criteria were as follows: (i) patients with primary malignancy but negative EBUS-TBNA results; (ii) lesions difficult to delineate; and (iii) lesions difficult to match with the biopsy on the image. Most of the biopsied LNs were at 4R, 4L and 7 (4R, right lower paratracheal; 4L, left lower paratracheal; 7, subcarina). LNs were classified as benign or malignant with pathology as the gold standard.

All LNs ( $N = 201$ ) were randomly divided into 2 subsets in a 7:3 ratio (29–31). Seventy percent ( $N = 140$ ) were assigned to the training set by stratified sampling, including 84 malignant cases (60%, 84/140) and 56 benign cases (40%, 56/140). The remaining 30% ( $N = 61$ ) were assigned to the test set, including 36 malignant cases (59%, 36/61) and 25 benign cases (41%, 25/61).

The flowchart of the enrollment of patients and data assignment is shown in **Figure 1**.

### CT Image Acquisition

CT scans were performed on Siemens (SOMATOM Force, 192-section dual-source CT), GE (Discovery CT 750 HD, 64 multidetector CT system), Toshiba (Aquilion One, 320 slice) and Philips (Brilliance iCT, 256 slice) systems. The patients were maintained in the supine position, their arms were raised, and



they held breath after deep inhalation for acquisition. The scan parameters were as follows: 120 kVp, 100–200 mA, 350 mm×350 mm field of view, and matrix of 512 × 512. Venous-phase contrast-enhanced CT was performed after a 30–40 s delay following the intravenous administration of 70–90 mL of nonionic contrast-enhanced medium (**Iproamine**) and 35–45 mL of normal saline at a rate of 2.5–3.0 mL/s with a power injector. Images were retrieved separately from the Picture Archiving and Communication System (PACS) (IMPAX, AGFA, Belgium).

## Feature Extraction

The study workflow is shown in **Figure 2**. CT images were extracted from PACS and imported into ITK-SNAP software (version 3.8.0; <http://www.itksnap.org>). A thoracic radiologist (reader 1) with 10 years' experience, who was blinded to the pathological results, reviewed the CT images on mediastinal window settings (W: 350, L: 50) and drew the region of interest (ROI) on the venous-phase contrast-enhanced images by using ITK-SNAP. The ROI was manually drawn along the boundaries of the LN on the maximal 2D axial slice.

Six months after the initial segmentation by reader 1, 50 LNs were randomly selected and resegmented by reader 2 with 7 years' experience, who was blinded to the pathological results, in

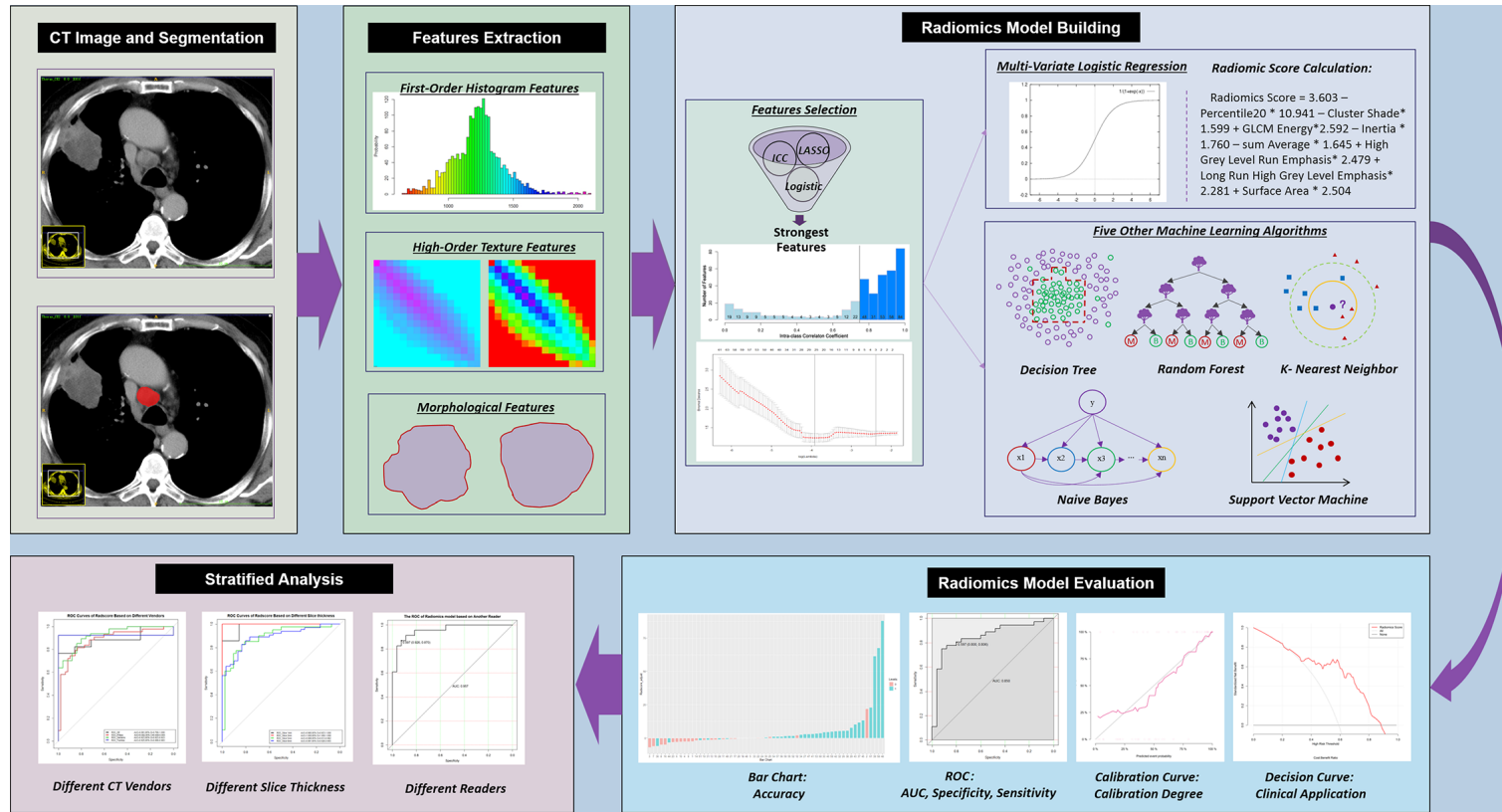
order to assess the interreader agreement in the extracted radiomics features. The intraclass correlation coefficient (ICC) was calculated.

Both the raw digital imaging and communications in medicine (DICOM) images and the corresponding ROIs were imported into A.K. software (Artificial Intelligence Kit, GE Healthcare, China) for the automatic quantitation of radiomics features compliant with the Image Biomarker Standardization Initiative (IBSI) (32).

## Radiomics Model Building

To reduce overfitting, least absolute shrinkage and selection operator (LASSO) regression analysis with 10-fold cross-validation was conducted and repeated 100 times before modeling to identify related features with nonzero coefficients to differentiate malignant and benign LNs in the training set. Multivariate logistic regression with the backward stepwise elimination method was used to select the strongest features associated with the probability of malignancy of LNs.

The radiomics score was calculated for each LN *via* a linear combination of the selected strongest features that were weighted by their respective coefficients. To compare different machine learning algorithms in distinguishing malignant LNs from benign LNs, five other machine learning models, including support vector



**FIGURE 2 |** Workflow of the current study. CT, computed tomography; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the ROC curve.

machine (SVM), K-nearest neighbors (KNN), random forest (RF), decision tree (DT) and naive Bayes (NB), were built.

## Radiomics Model Evaluation

The diagnostic performance of different machine learning models for LN metastasis was evaluated by using receiver operating characteristic (ROC) curves and calculating the diagnostic accuracy in both the training and test sets. The performance of the established radiomics score was evaluated with discrimination, calibration and clinical application in both the training and test sets.

- 1. Discrimination.** The discrimination degree (which measures the ability of model to distinguish malignant and benign LNs) of the model was assessed by the area under the ROC curve (AUC). The optimal cutoff of the radiomics score calculated from the training set was applied in the test set.
- 2. Calibration.** The calibration degree (which measures the agreement between the observed outcome frequencies and classified probabilities) of the model was evaluated by plotting the calibration curve. The Hosmer-Lemeshow test was used to determine the goodness of fit of the models, and P values of more than 0.05 were considered well calibrated.
- 3. Clinical application.** To evaluate the potential clinical application of the estimation model, decision curve analysis (DCA) was conducted by quantifying the net benefits at different threshold probabilities.

We used the same machine learning process as above to analyze the non-enhanced images. The process included drawing ROI, features extraction, modeling and model evaluation.

## Stratified Analysis

Stratified analysis was performed on subsets of different CT scanners and slice thicknesses. Based on different scanners, 201 LNs were divided into four subsets: GE (n = 24; malignant: 17, benign: 7), Philips (n = 91; malignant: 43, benign: 48), Siemens (n = 14; malignant: 13, benign: 1) and Toshiba (n = 72; malignant: 47, benign: 25). Based on different slice thicknesses, 61 LNs were divided into four subsets: 1 mm (n = 14; malignant: 7, benign: 7), 2 mm (n = 10; malignant: 8, benign: 2), 5 mm (n = 82; malignant: 40, benign: 42) and 8 mm (n = 95; malignant: 30, benign: 65). The DeLong test was used to compare the differences in the AUC values of the established radiomics score between these subpopulations, and a P value greater than 0.05 indicated that there was no significant difference in diagnostic performance in different subpopulations.

## Statistical Analysis

All statistical analyses were performed using R studio software (version 1.2.1335). LASSO regression was performed using the “glmnet” package. ROC curves were plotted using the “pROC” package. DCA was performed using the “DecisionCurve” function. P values of less than 0.05 (two-sided) were considered statistically significant. The differences in clinical factors and malignancy-related features between the malignant group and benign group in both the training and test sets were

assessed by independent t tests according to the distribution type of the data. The chi-squared test was used to compare the significance of the differences between categorical variables. The DeLong test was used to compare the differences in the AUC values of different models.

## RESULTS

### Pathological Results and Clinical Information

The pathological results and clinical information of 201 mediastinal LNs in 129 patients are shown in **Table 1** and **Table 2**. There was no significant difference ( $P > 0.05$ ) in malignant and benign LNs between the training set and test set. Within the training set, there were no significant differences ( $P > 0.05$ ) in gender or age between the benign and malignant groups. Within the test set, the malignant group was older ( $P = 0.001$ ), and the proportion of males was higher ( $P < 0.001$ ).

### Feature Extraction

A total of 396 quantitative radiomics features were extracted, including first-order histogram features (N = 42), high-order texture features, including haralick features (N = 10), gray level size zone matrix (GLSZM, N = 11) features, gray level cooccurrence matrix (GLCM, N = 144) features and gray level run length matrix (GLRLM, N = 180) features, and form-factor features (N = 9). Details are reported in **Supplementary Figure S4**.

### Radiomics Model Building

To build a robust radiomics model, only the features that showed high stability remained for the next step. Two hundred and seventy-four features with an ICC greater than 0.75 remained (**Figure 3**).

For feature selection, 26 radiomics features with a nonzero coefficient after the LASSO regression method remained (shown in **Figure 4**). The eight strongest features among the 26 features were identified after further stepwise elimination in multivariate logistic regression. Then, the radiomics score was established based on the eight features (the formula and interpretation are shown in **Supplementary A1**), and the calculation was as follows:

**TABLE 1** | Demographic data of training set and test set.

	Training Set	Test Set	P
Age (m, [Q1, Q3])	56.00 (50.00, 63.00)	55.00 (47.70, 61.30)	0.276
Sex (No, [%])			0.744
Male	74 (52.86)	30(49.18)	
Female	66 (47.14)	31(50.82)	
Smoking			0.557
Nonsmoker	74	34	
Ex-Smoker	11	8	
Current Smoker	35	12	
Unknown	20	7	

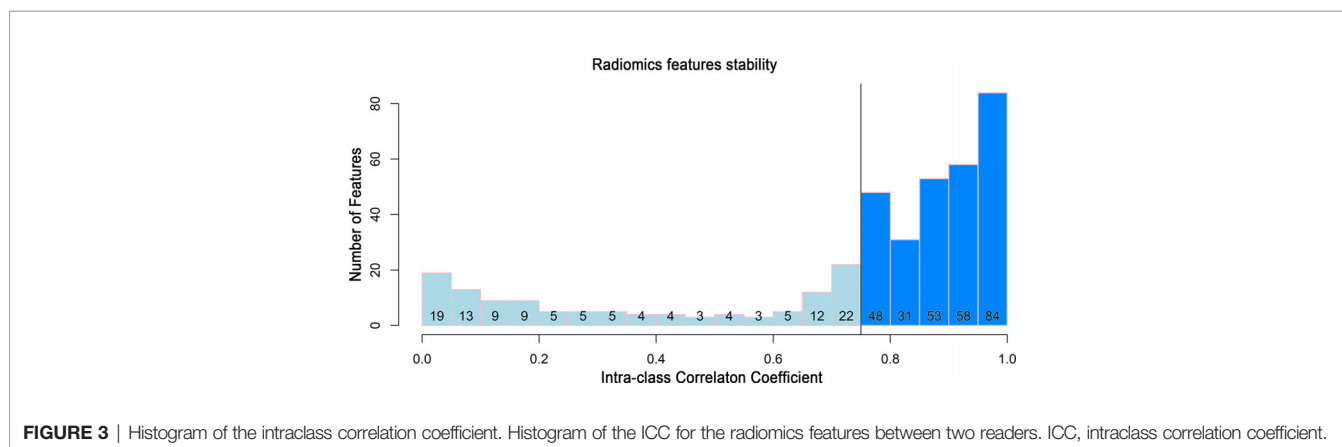
m, median; No, number.



**TABLE 2 |** Demographic data of benign and malignant group in training set and test set.

Characteristic	Training Set		P	Test Set		P
	Benign	Malignant		Benign	Malignant	
Age (years, mean ± SD)	55.14 ± 11.85	56.92 ± 11.99	0.390	47.12 ± 12.36	58.33 ± 11.42	0.001
Sex (No, [%])			0.128			<0.001
Male	34 (60.71)	40 (47.62)		2 (0.08)	28 (0.78)	
Female	22 (39.29)	44 (52.38)		23 (0.92)	8 (0.22)	
Pathological results			N/A			N/A
Tuberculosis	3	–		3	–	
Sarcoidosis	46	–		19	–	
Inflammation	7	–		3	–	
LAC	–	32		–	16	
LSCC	–	8		–	2	
SCLC	–	39		–	17	
Breast cancer	–	3		–	1	
Kidney cancer	–	2		–	0	
Smoking status			0.074			<0.001
Nonsmoker	31	43		24	10	
Ex-smoker	5	6		0	8	
Current smoker	17	18		1	11	
Unknown	3	17		0	7	
Contact history of harmful or toxic substances			1.000			1.000
Yes	3	5		2	4	
No	53	79		23	32	
Family history			0.528			1.000
Pulmonary tumor	0	3		0	1	
Extra-pulmonary tumor	3	7		1	3	

SD, standard deviation; No, number; LSCC, lung squamous cell carcinoma; LAC, lung adenocarcinoma; SCLC, small cell lung cancer; N/A, not available.



**FIGURE 3 |** Histogram of the intraclass correlation coefficient. Histogram of the ICC for the radiomics features between two readers. ICC, intraclass correlation coefficient.

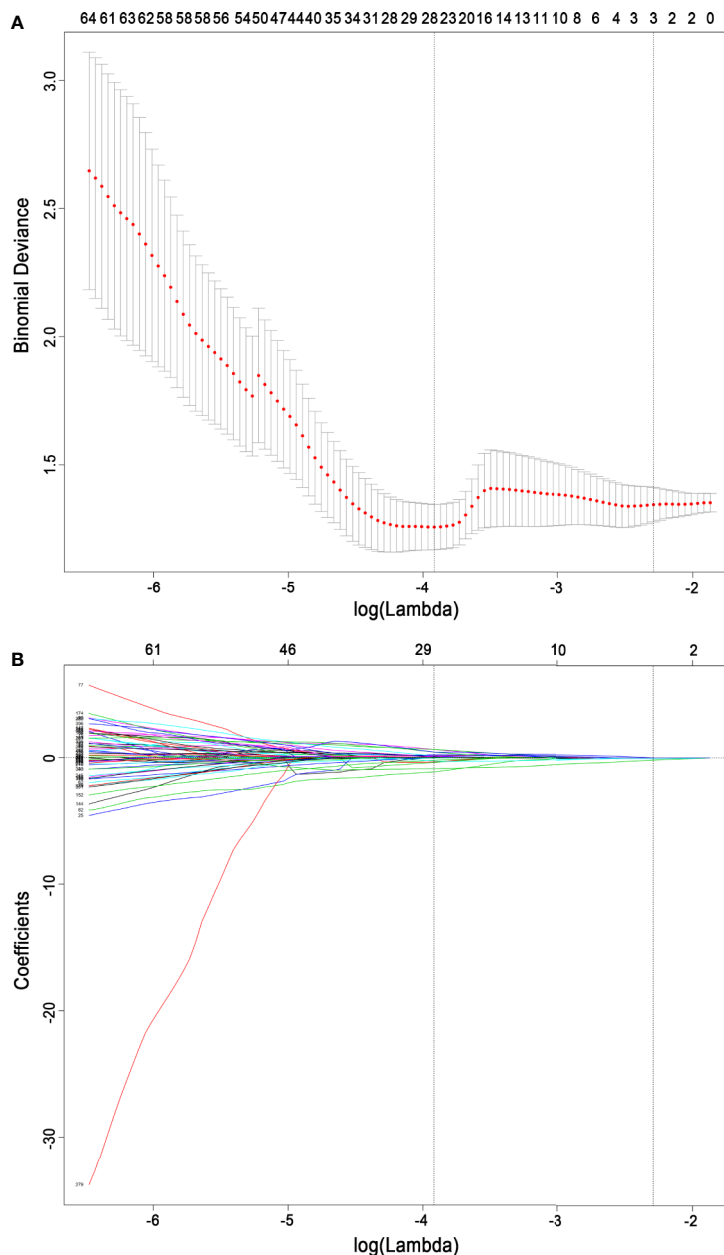
*Radiomics score* = 3.603 – Percentile20

$$\begin{aligned}
 &\times 10.941 - ClusterShade\_AllDirection\_offset1\_SD \\
 &\times 1.599 + GLCMEnergy\_angle135\_offset7 \\
 &\times 2.592 - Inertia\_AllDirection\_offset1\_SD \times 1.760 \\
 &- sumAverage \times 1.645 \\
 &+ HighGrayLevelRunEmphasis\_AllDirection\_offset7\_SD \\
 &\times 2.479 \\
 &+ LongRunHighGrayLevelEmphasis\_angle135\_offset4 \\
 &\times 2.281 + SurfaceArea \times 2.504
 \end{aligned}$$

We additionally used two other methods to select radiomics features. They were Spearman correlation analysis with univariate logistic regression, and Spearman correlation analysis with Gradient Boost Decision Tree (GBDT) and multivariate logistic regression. The classification performances of the two methods were similar to those of LASSO regression. Details are shown in **Supplementary A2**.

### Radiomics Model Evaluation

For all six machine learning algorithms, the AUCs were greater than 0.65 in the test set (**Table 3**). We used DeLong test to compare the classification performances of the six machine learning algorithms (please see **Supplementary Tables S4, S5**).



**FIGURE 4 |** Radiomics feature selection based on the least absolute shrinkage and selection operator binary logistic regression model. **(A)** Ten-fold cross-validation *via* minimum criteria was used for tuning the parameter ( $\lambda$ ) in the least absolute shrinkage and selection operator model. The relationship between the binomial deviance and  $\log(\lambda)$  was plotted. The left dotted vertical line was at the optimal  $\lambda$  value point by using the minimum criteria, and the right line was at the optimal  $\lambda$  value point by using one standard error of the minimum criteria (the 1-SE criteria). A  $\lambda$  value of 0.01987634, with  $\log(\lambda) -3.918$ , was chosen (minimum) according to 10-fold cross-validation. **(B)** The least absolute shrinkage and selection operator coefficient profiles of the 396 radiomics features. A coefficient profile plot was plotted versus the  $\log(\lambda)$  sequence. A vertical line was drawn at the value selected using 10-fold cross-validation, where the optimal  $\lambda$  resulted in 26 nonzero coefficients.

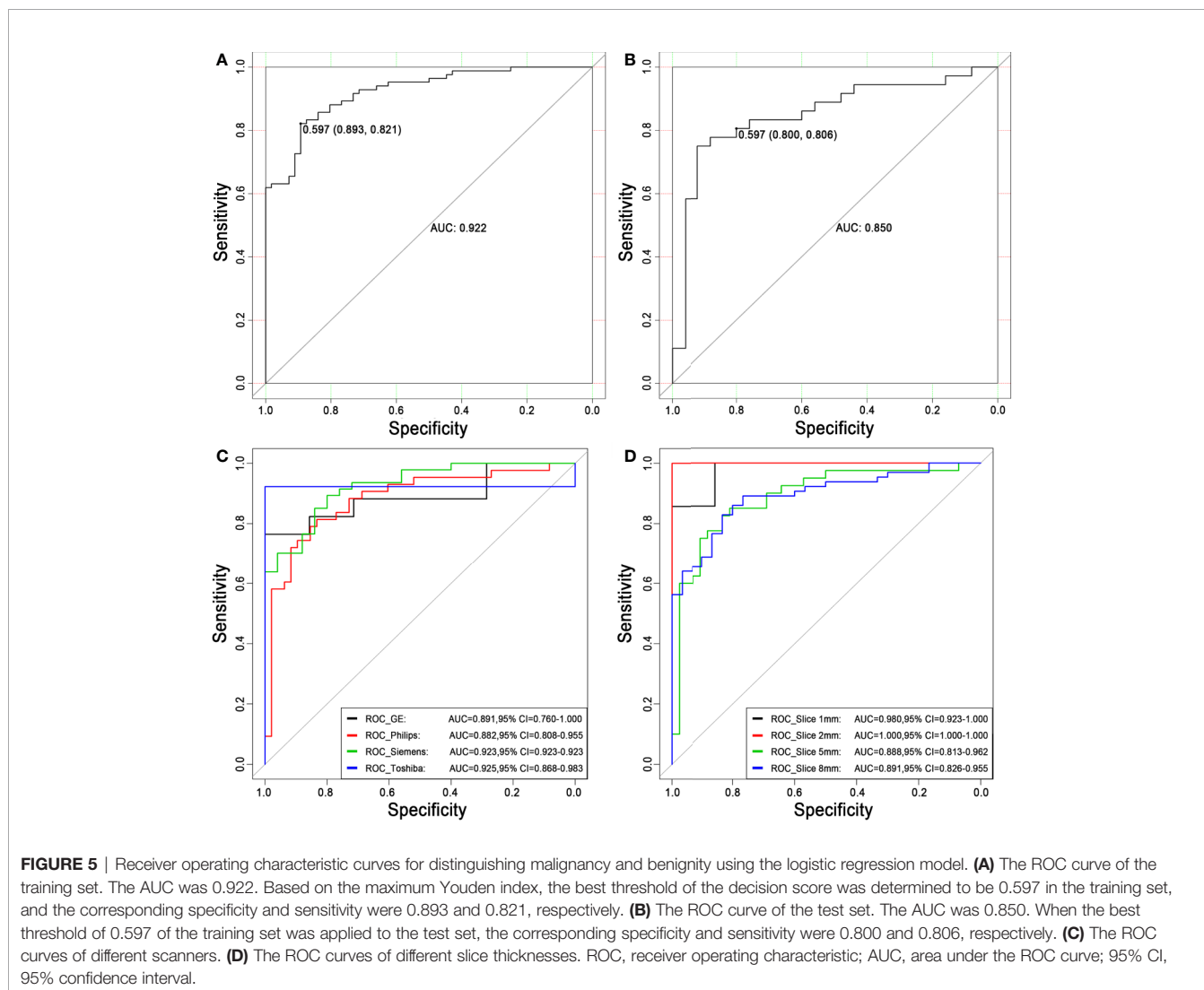
Both SVM and logistic regression are generalized linear models, and their performances were similar. Considering interpretability and generalizability, logistic regression model was selected as the main result in this study. The performance of the radiomics score in estimating the status of LNs was as follows.

- 1. Discrimination.** The ROC curves of the radiomics score in the training set and the test set are shown in **Figures 5A, B**. The accuracy, sensitivity, specificity and AUC were 0.850, 0.821, 0.893 and 0.922 [95% confidence interval (95% CI), 0.880 to 0.964], respectively, for the radiomics model in the training set. The accuracy, sensitivity, specificity and AUC

**TABLE 3** | The classification performance of different algorithms.

Set	Logistic Regression	SVM	Naive Bayes	RF	DT	KNN
Training set						
AUC	0.922	0.907	0.808	0.999	0.765	0.914
Accuracy	0.850	0.843	0.771	0.986	0.757	0.836
Sensitivity	0.821	0.869	0.810	0.976	0.726	0.845
Specificity	0.893	0.804	0.714	1.000	0.804	0.821
Test set						
AUC	0.850	0.853	0.751	0.789	0.661	0.743
Accuracy	0.803	0.803	0.738	0.672	0.672	0.705
Sensitivity	0.806	0.833	0.778	0.833	0.722	0.722
Specificity	0.800	0.760	0.680	0.440	0.600	0.680

AUC, area under the receiver operating characteristic curve; SVM, support vector machine; RF, random forest; DT, decision tree; KNN, k-nearest neighbors.

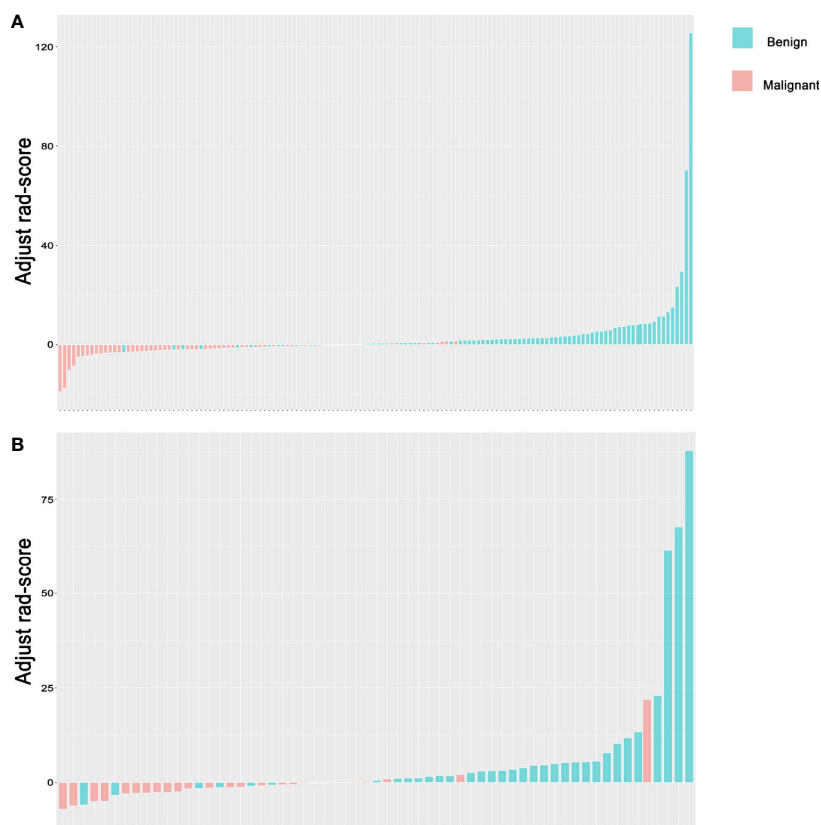


were 0.803, 0.806, 0.800 and 0.850 (95% CI, 0.747 to 0.953), respectively, in the test set. The accuracy, sensitivity, specificity and AUC were 0.900, 0.870, 0.926 and 0.957 (95% CI, 0.906 to 1.000), respectively, for reader 2. The

distributions of the radiomics score in the training and test sets are shown in **Figure 6**.

- Calibration.** The radiomics score showed good agreement between the actual observations and classifications in both





**FIGURE 6** | Rad-score for lymph nodes. **(A)** Rad-score for every LN in the training set. **(B)** Rad-score for every LN in the test set. The y-axis refers to the Rad-score minus the optimal cutoff value. Up and down bars refer to the classified malignant and benign LNs, respectively. Blue and red bars refer to actual malignant and benign LNs, respectively. LN, lymph node.

the training and test sets (**Figures 7A, B**). Nonsignificant statistics were achieved in the Hosmer-Lemeshow test in the training set ( $P = 0.6898$ ) and test set ( $P = 0.1762$ ).

- Clinical application.** DCA showed that the radiomics model would add more benefit in distinguishing malignant and benign LNs when the threshold probability was between 10% to 100% in the training set or between 30% to 90% in the test set (**Figures 7C, D**).

The classification performances of the model of non-enhanced images are reported in **Supplementary A3**. The classification performances of the model of venous-phase images were better than those of model of non-enhanced images.

### Stratified Analysis

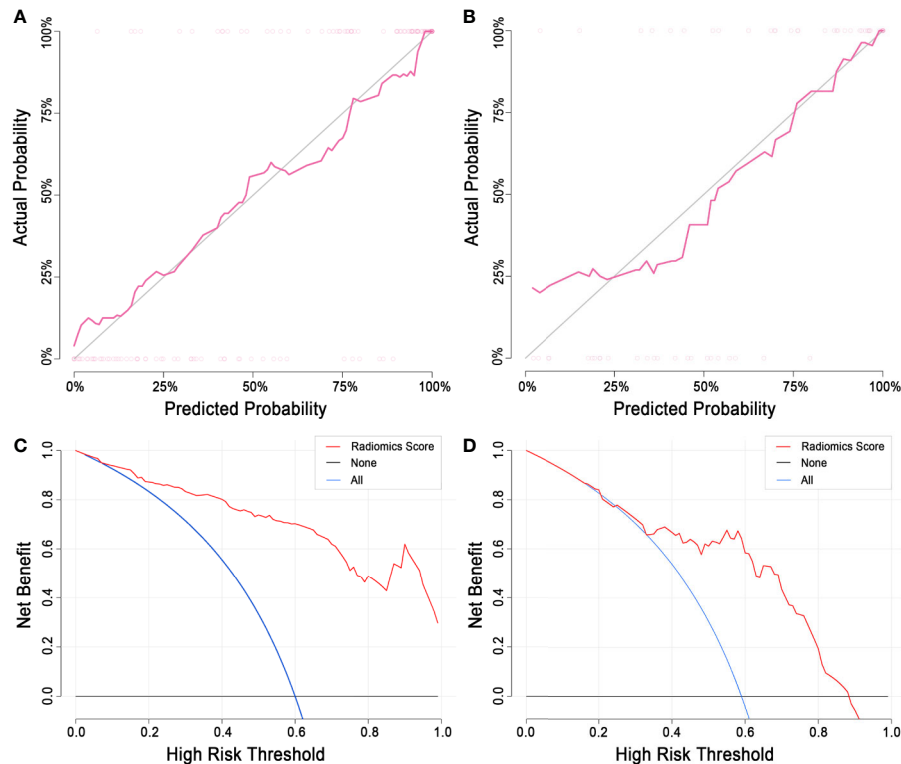
The stratified analysis results are shown in **Figures 5C, D**. The AUCs for the GE, Philips, Siemens and Toshiba subsets were 0.891 (95% CI, 0.760 to 1.000), 0.882 (95% CI, 0.808 to 0.955), 0.923 (95% CI, 0.923 to 0.923) and 0.925 (95% CI, 0.868 to 0.983), respectively. The AUCs for the 1 mm, 2 mm, 5 mm and 8 mm subsets were 0.980 (95% CI, 0.923 to 1.000), 1.000 (95% CI, 1.000 to 1.000), 0.888 (95% CI, 0.813 to 0.962) and 0.891 (95% CI, 0.826 to 0.955), respectively. There was no significant

difference (DeLong test,  $P > 0.05$ ) between any two subgroups, which showed the generalization of the radiomics score across different factors.

## DISCUSSION

In the present study, we established a robust radiomics model for distinguishing malignant from benign mediastinal LNs based on contrast-enhanced CT imaging. Stratified analysis demonstrated the stable performance of the radiomics model, which was not affected by different scanners or slice thicknesses. Calibration curve and DCA suggested good fitness and the potential prospect of clinical application.

The accurate evaluation of mediastinal LNs before surgery plays a pivotal role in staging and treatment planning for lung cancer. Ultrasound, CT and PET/CT are the main methods, with size, shape, internal structure and contrast uptake as the main criteria for evaluation. The accuracy for mediastinal LN evaluation is not satisfactory at present. The reasons are as follows. (1) Size is not the gold standard, as smaller LNs might be early metastases. (2) No objective and accurate criteria for morphological evaluation exist. (3) The visual inspection of the



**FIGURE 7** | Calibration curve and decision curve analysis of the radiomics model. **(A)** The calibration curve of the training set. **(B)** The calibration curve of the test set. The diagonal line represents a perfect classification. **(C)** DCA of the training set. **(D)** The DCA of the test set. The y-axis indicates the net benefit. The net benefit is determined by calculating the difference between the expected benefit and the expected harm associated with each proposed model [net benefit = true positive rate – (false positive rate × weighting factor); weighting factor = threshold probability/(1–threshold probability)]. The blue line represents the assumption that all LNs are malignant. The black line represents the assumption that all LNs are benign. DCA, decision curve analysis; LN, lymph node.

internal structure is limited. (4) The enhancement of metastatic LNs varies. Heterogeneity is an important feature of malignancy, presumably reflecting infiltration, angiogenesis and necrosis (33–35). Studies of perfusion CT suggested that the blood flow might be more heterogeneous in metastatic nodes than in non-metastatic nodes (36). However, it was difficult to tell subtle differences by the naked eye, even for experienced radiologists. For further evaluation of the distinct information of the image, including the heterogeneity of the contrast inside the nodes, contrast-enhanced CT was included, and radiomics features were extracted in our study.

Eight potential features remained in our model: a first-order histogram feature (Percentile20), a Haralick feature (Sum Average), GLCM features (GLCM Energy, Cluster Shade, Inertia), GLRLM features [High Gray Level Run Emphasis (HGLRE), Long Run High Gray Level Emphasis (LRHGLE)] and a morphological feature (Surface Area). Surface Area was the only selected morphological feature with a relatively high coefficient of 2.504. Surface Area is an approximation of the surface of the ROI in  $\text{mm}^2$ , which suggests that the larger the maximum cross-sectional area of the LN, the more likely it is to be malignant. The other features are calculated based solely on pixel or pixel pairs and provide meaningful information that is invisible to the naked eye, which reflects the characteristics of heterogeneity in addition

to morphological features. Percentile20 showed the strongest negative correlation, with a coefficient of -10.941. It was one of the first-order histogram features used in statistics, indicating that the image intensity value below 20% of observations in a group of observations falls. The GLCM features were the most important radiomics category for the malignancy of LNs, with 3 features included in the model. They represent the joint probability of certain sets of pixels having certain gray-level values. GLCM Energy was the strongest positive feature with a coefficient of 2.592. Energy measures the homogeneous patterns in the image. A greater energy implies that there are more instances of intensity value pairs in the image that neighbor each other at higher frequencies. Cluster Shade measures the skewness and uniformity of the GLCM, and a higher cluster shade implies greater asymmetry about the mean. Inertia reflects the clarity of the image and the depth of the texture groove, the contrast is proportional to the texture groove, and high values of the groove produce more clarity; in contrast, small values of the groove would result in small contrast and fuzzy images. In our study, inertia showed a negative relation with malignancy.

To stabilize the performance of the radiomics model, several steps were performed in our study. (1) Strict inclusion and exclusion criteria were followed to avoid potential bias. Nine patients with primary malignancy but negative EBUS-TBNA

were excluded considering the probability of false negative results of the EBUS-TBNA. Occasionally, small LNs could be biopsied during EBUS, but they are difficult to match or delineate on chest CT. Fifteen patients were excluded for this reason. (2) A.K. software, which has been used and verified in many studies, was applied for radiomics feature extraction. (3) The ICC was calculated to verify the reproducibility of feature extraction between two readers. (4) Multiple evaluations of the efficiency and stability of the model were performed. The AUC, classification probability and DCA were calculated, representing the discrimination, calibration and clinical application, respectively. (5) For a robust radiomics model fit for different scanners or slice thicknesses, the stratified analysis was conducted in the study. The radiomics model was applied to each subset divided by scanners and slice thicknesses and shown to be slightly affected by different scanners or slice thicknesses, suggesting robustness.

Our study had several limitations. First, it was a retrospective single-center study, and the sample size was relatively small, although the images were derived from various scanners, and the stratified analysis was performed among different scanners and different slice thicknesses. The robustness of the classification model needs to be tested in a multicenter dataset. Second, for reproducibility, we manually drew ROIs on one 2D slice instead of 3D slices. It is easy to process and analyze, but it may not represent all the features and characteristics of LNs. Although the ICC suggested reliable interreader agreement for the extracted radiomics features, further study based on 3D images should be performed. Third, no clinical factors were integrated into the model, which might further improve the performance. Fourth, we didn't use deep learning to investigate our data. In future research, we will try to use deep learning to investigate our data. Fifth, we had no external validation set. In order to verify the clinical value of the classification performances, we need external validation set. Sixth, in this work, the distribution of age and gender in the training set and the test set was inconsistent. But there was no statistical difference in the distribution between the training set and the test set. Additionally, within the training set, there were no significant differences in gender and age between the benign and malignant groups. It largely avoided that the trained model was a model for diagnosing gender or age, although there were statistical differences in gender and age within the test set.

In conclusion, we established a robust radiomics model for distinguishing malignant from benign mediastinal LNs based on contrast-enhanced CT imaging, with stability for different scanners and slice thicknesses. The radiomics model has the potential for clinical application.

## REFERENCES

- Guo D, Ni Y, Lv X, Zhang Z, Ye P. Distribution and prognosis of mediastinal lymph node metastases of nonsmall cell lung cancer. *J Cancer Res Ther* (2016) 12(5):120. doi: 10.4103/0973-1482.191613
- Hegde PV, Liberman M. Mediastinal staging: endosonographic ultrasound lymph node biopsy or mediastinoscopy. *Thoracic Surg Clinics* (2016) 26(3):243–9. doi: 10.1016/j.thorsurg.2016.04.005
- Alberts WM. Diagnosis and management of lung cancer executive summary: ACCP evidence-based clinical practice guidelines. *Chest* (2007) 132(3):1S–19S. doi: 10.1378/chest.07-1860
- Li L, Zhang H, Wang L, Xie C, Zhou Y, Zhong Y. A retrospective analysis on metastatic rate of the internal mammary lymph node and its clinical significance in adjuvant radiotherapy of breast cancer patients. *BMC Cancer* (2020) 20(1):1–7. doi: 10.1186/s12885-020-6642-9
- Sanz-Santos J, Call S. Preoperative staging of the mediastinum is an essential and multidisciplinary task. *Respirology* (2020) 25:37–48. doi: 10.1111/resp.13901

## DATA AVAILABILITY STATEMENT

Derived data supporting the findings of this study are available from the corresponding author LZ on request. Requests to access the datasets should be directed to LZ, zhanglinda@163.com.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of the First Affiliated Hospital of China Medical University. This is a retrospective study, and the requirement for informed consent was waived.

## AUTHOR CONTRIBUTIONS

MD and LZ mainly wrote the paper. GH, SL and NL were in charge of designing the method and data acquisition. KX and LZ designed the study. The manuscript was revised by all authors and the final version of the manuscript was approved by all authors. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by National Financial Appropriation Research Project (2017YFC1309100) and National Scientific Foundation of China (81971695).

## ACKNOWLEDGMENT

We sincerely appreciate Yan Guo from the Department of Core Imaging, Life Science, GE healthcare, Shanghai, China, for her assistance in the data processing and statistical analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2020.558428/full#supplementary-material>

6. Gu P, Zhao YZ, Jiang LY, Zhang W, Xin Y, Han BH. Endobronchial ultrasound-guided transbronchial needle aspiration for staging of lung cancer: a systematic review and meta-analysis. *Eur J Cancer* (2009) 45(8):1389–96. doi: 10.1016/j.ejca.2008.11.043
7. Silvestri GA, Gonzalez AV, Jantz MA, Margolis ML, Gould MK, Tanoue LT, et al. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* (2013) 143(5 Suppl):e211S–e50S. doi: 10.1378/chest.12-2355
8. Varela-Lema L, Fernandez-Villar A, Ruano-Ravina A. Effectiveness and safety of endobronchial ultrasound-transbronchial needle aspiration: a systematic review. *Eur Respir J* (2009) 33(5):1156–64. doi: 10.1183/09031936.00097908
9. Yasufuku K, Chiyo M, Koh E, Moriya Y, Iyoda A, Sekine Y, et al. Endobronchial ultrasound guided transbronchial needle aspiration for staging of lung cancer. *Lung Cancer* (2005) 50(3):347–54. doi: 10.1016/j.lungcan.2005.07.013
10. Haas AR. Infectious complications from full extension endobronchial ultrasound transbronchial needle aspiration. *Eur Respir J* (2009) 33(4):935–8. doi: 10.1183/09031936.00025708
11. Asano F, Aoe M, Ohsaki Y, Okada Y, Sasada S, Sato S, et al. Complications associated with endobronchial ultrasound-guided transbronchial needle aspiration: a nationwide survey by the Japan Society for Respiratory Endoscopy. *Respir Res* (2013) 14:50. doi: 10.1186/1465-9921-14-50
12. Steinfort DP, Johnson DF, Irving LB. Infective complications from endobronchial ultrasound-transbronchial needle aspiration. *Eur Respir J* (2009) 34(2):524–5; author reply 5. doi: 10.1183/09031936.00057009
13. Tian Q, Chen LA, Wang RT, Yang Z, An Y. The reasons of false negative results of endobronchial ultrasound-guided transbronchial needle aspiration in the diagnosis of intrapulmonary and mediastinal malignancy. *Thoracic Cancer* (2013) 4(2):186–90. doi: 10.1111/1759-7714.12010
14. Medford AR, Bennett J, Free C, Agrawal S. Endobronchial ultrasound guided transbronchial needle aspiration. *J Thorac Dis* (2010) 86(1012):106–15. doi: 10.1136/pgmj.2009.089391
15. Kemp S, El Batrawy S, Harrison R, Skwarski K, Munavvar M, Roselli A, et al. Learning curves for endobronchial ultrasound using cusum analysis. British: BMJ Publishing Group Ltd (2010).
16. Medford A. False negative results and endobronchial ultrasound-guided transbronchial needle aspiration. *Thoracic Cancer* (2013) 4(4):484–. doi: 10.1111/1759-7714.12057
17. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* (2012) 30(9):1234–48. doi: 10.1016/j.mri.2012.06.010
18. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
19. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
20. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
21. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* (2016) 6:23428. doi: 10.1038/srep23428
22. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* (2015) 114(3):345–50. doi: 10.1016/j.radonc.2015.02.015
23. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128
24. Cong M, Feng H, Ren JL, Xu Q, Cong L, Hou Z, 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* (2019) 139:73–9. doi: 10.1016/j.lungcan.2019.11.003
25. Bayanati H, ET R, Souza CA, Sethi-Virmani V, Gupta A, Maziak D, et al. Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? *Eur Radiol* (2015) 25(2):480–7. doi: 10.1007/s00330-014-3420-6
26. Shin SY, Hong IK, Jo YS. Quantitative computed tomography texture analysis: can it improve diagnostic accuracy to differentiate malignant lymph nodes? *Cancer Imaging* (2019) 19(1):25. doi: 10.1186/s40644-019-0214-8
27. Sha X, Gong G, Qiu Q, Duan J, Li D, Yin Y. Discrimination of mediastinal metastatic lymph nodes in NSCLC based on radiomic features in different phases of CT imaging *J BMC Med Clin* (2020) 20(1):12. doi: 10.1186/s12880-020-0416-3
28. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol* (2019) 30(3):431–8. doi: 10.1093/annonc/mdz001
29. Xu X, Li H, Wang S, Fang M, Zhong L, Fan W, et al. Multiplanar MRI-Based Predictive Model for Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer. *Front Oncol* (2019) 9:1007. doi: 10.3389/fonc.2019.01007
30. Cong M, Feng H, Ren J-L, Xu Q, Cong L, Hou Z, 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. doi: 10.1016/j.lungcan.2019.11.003
31. Zhang Y, Yu S, Zhang L, Kang L. Radiomics Based on CECT in Differentiating Kimura Disease From Lymph Node Metastases in Head and Neck: A Non-Invasive and Reliable Method. *Front Oncol* (2020) 10:1121. doi: 10.3389/fonc.2020.01121
32. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv preprint arXiv:161207003* (2016). doi: 10.1148/radiol.2020191145
33. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol* (2012) 22(4):796–802. doi: 10.1007/s00330-011-2319-8
34. Lunt SJ, Chaudary N, Hill RP. The tumor microenvironment and metastatic disease. *Clin Exp Metastasis* (2009) 26(1):19–34. doi: 10.1007/s10585-008-9182-2
35. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging* (2012) 3(6):573–89. doi: 10.1007/s13244-012-0196-6
36. Spira D, Wecker M, Spira SM, Hetzel J, Spengler W, Sauter A, et al. Does volume perfusion computed tomography enable differentiation of metastatic and non-metastatic mediastinal lymph nodes in lung cancer patients? A feasibility study. *Cancer Imaging* (2013) 13(3):323–31. doi: 10.1102/1470-7330.2013.0033

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

Copyright © 2021 Dong, Hou, Li, Li, Zhang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.