Original Article

Radiomics-based predictive risk score: A scoring system for preoperatively predicting risk of lymph node metastasis in patients with resectable non-small cell lung cancer

Lan He[1], Yanqi Huang[1], Lixu Yan[2], Junhui Zheng[1], Changhong Liang[1], Zaiyi Liu[1]

[1]Department of Radiology; [2]Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China

Correspondence to: Zaiyi Liu, PhD. Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, China. Email: zyliu@163.com; Changhong Liang, PhD. Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, China. Email: cjr.lchh@vip.163.com.

Abstract

Objective: To develop and validate a radiomics-based predictive risk score (RPRS) for preoperative prediction of lymph node (LN) metastasis in patients with resectable non-small cell lung cancer (NSCLC).

Methods: We retrospectively analyzed 717 who underwent surgical resection for primary NSCLC with systematic mediastinal lymphadenectomy from October 2007 to July 2016. By using the method of radiomics analysis, 591 computed tomography (CT)-based radiomics features were extracted, and the radiomics-based classifier was constructed. Then, using multivariable logistic regression analysis, a weighted score RPRS was derived to identify LN metastasis. Apparent prediction performance of RPRS was assessed with its calibration, discrimination, and clinical usefulness.

Results: The radiomics-based classifier was constructed, which consisted of 13 selected radiomics features. Multivariate models demonstrated that radiomics-based classifier, age group, tumor diameter, tumor location, and CT-based LN status were independent predictors. When we assigned the corresponding score to each variable, patients with RPRSs of 0–3, 4–5, 6, 7–8, and 9 had distinctly very low (0%–20%), low (21%–40%), intermediate (41%–60%), high (61%–80%), and very high (81%–100%) risks of LN involvement, respectively. The developed RPRS showed good discrimination and satisfactory calibration [C-index: 0.785, 95% confidence interval (95% CI): 0.780–0.790]. Additionally, RPRS outperformed the clinicopathologic-based characteristics model with net reclassification index (NRI) of 0.711 (95% CI: 0.555–0.867).

Conclusions: The novel clinical scoring system developed as RPRS can serve as an easy-to-use tool to facilitate the preoperatively individualized prediction of LN metastasis in patients with resectable NSCLC. This stratification of patients according to their LN status may provide a basis for individualized treatment.

Keywords: Lymph node; radiomics; risk score; CT; non-small cell lung cancer

Submitted Dec 07, 2018. Accepted for publication Apr 28, 2019.
doi: 10.21147/j.issn.1000-9604.2019.04.08
View this article at: https://doi.org/10.21147/j.issn.1000-9604.2019.04.08

Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of such deaths (1,2). Precise staging is the key to appropriate prognosis and treatment strategy decision (3). For patients with resectable (stage I–IIIA) NSCLC, surgical resection remains the primary and preferred approach to the treatment with the

best chance for cure (4). As recommended by National Comprehensive Cancer Network (NCCN) guidelines, patients with resectable NSCLC should receive N1 and N2 node resection and a minimum of 3 N2 stations sampled or complete lymph node (LN) dissection during pulmonary resection (5). However, patients who had negative nodes by complete mediastinal LN dissection did not have improved survival compared with systematic LN sampling (6,7), and patients with positive LN involvement have a higher risk of disease recurrence. Therefore, accurate identification of LN involvement is crucial for prognosis and treatment strategy decision in patients with resectable NSCLC.

Although LN sampling or dissection plays an important role in precise nodal staging by identifying LN, they are available only invasively and postoperatively. The accurate preoperative assessment of LN involvement in patients with resectable NSCLC has a crucial effect on therapy planning, thus aiding in pretreatment decision making (8,9). As an alternative, imaging modalities like computed tomography (CT), positron emission tomography (PET)-CT and magnetic resonance (MR) are the most widely used for preoperative work-up, and are very important in detecting LN enlargement; however, a previous study has shown that they have limitations for their predictive value to differentiate benign nodes from malignant ones (9).

Recent advances in radiomics, which extract quantitative descriptors from routinely acquired medical images noninvasively, has provided deep insights into different fields of personalized medicine in oncologic practice, including tumor detection, subtypes classification, and therapy response assessment (10-14). It is worth noting that the recent studies have indicated that radiomics predictive models have been accepted as reliable tools to quantify risk by incorporating and illustrating important factors for oncologic prognosis and prediction (15-17). Although CT-based radiomics features assessments have been applied and demonstrated to be useful for prognosis prediction in patients with NSCLC (14,16,18), to our knowledge, no published study has determined whether the individual prediction of LN metastasis in resectable NSCLC could be achieved by a radiomics clinical scoring system based on preoperative CT images.

Therefore, this study aims to develop and validate a radiomics-based predictive risk score (RPRS) that preoperatively predicts risk of LN metastasis in patients with resectable NSCLC.

Materials and methods

Patients

The Institutional Review Board of Guangdong General Hospital approved this retrospective study, with informed consent waived. We retrospectively analyzed medical records from 717 consecutive patients who underwent surgical resection for primary NSCLC with systematic mediastinal lymphadenectomy from October 2007 to July 2016. The patient recruitment process was presented in *Figure 1*. All patients underwent contrast-enhanced chest CT images within two weeks before surgical resection. Between October 2007 and December 2013, patients were included to form the development cohort, and 423 cases were enrolled, while between January 2014 and July 2016, other patients were evaluated to form an independent validation cohort, and 294 cases were enrolled. Of these 717 patients in total, 277 patients additionally underwent PET-CT before resection. We selected these patients for subgroup analysis.

At baseline, clinical features on primary NSCLC (age, gender, smoking history, tumor location), and the acquisition date of CT imaging were recorded.

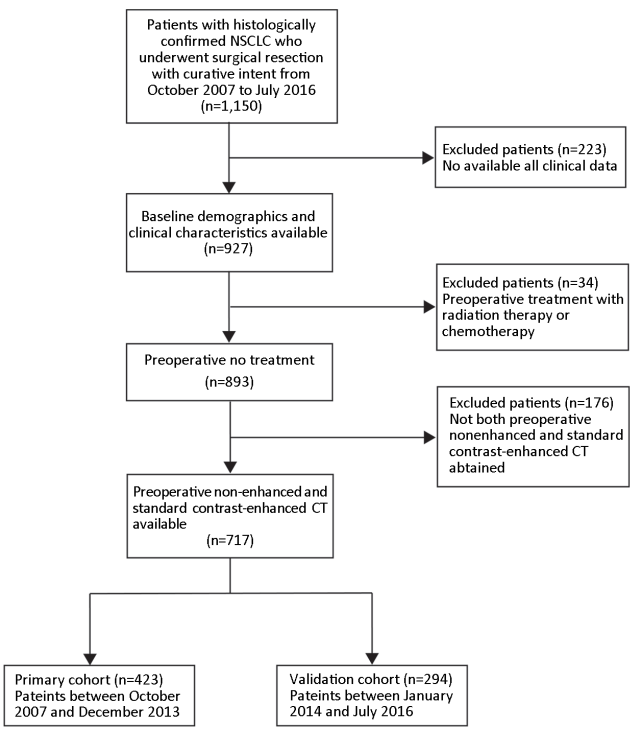

Figure 1 Patient recruitment process. NSCLC, non-small cell lung cancer; CT, computed tomography.

LN metastasis evaluation

LNs were dissected during surgical resection of NSCLC. The International Association for the Study of Lung Cancer (IASLC) has defined adequate LN sampling as: at least 3 mediastinal LN stations, station 7 in all case, station 5/6 with left upper lobe tumors and station 9 with lower lobe tumors; the sampling of at least 3 hilar LN stations is also recommended. The tissue was examined and the number of LNs was identified by pathologist postoperatively. Note that LN status in this study was defined by case.

CT image achievement

All patients underwent contrast-enhanced (reconstruction thickness of 1.25 mm) chest CT with one of two multi-detector row CT (GE Light-speed Ultra 8; GE Healthcare, Hino, Japan; 64-slice LightSpeed VCT, GE Medical systems, Milwaukee, WI, USA). The CT images acquisition parameters were as follows: 120 kV; 160 mAs; 0.5- or 0.4-second rotation time; detector collimation: 8×2.5 mm or 64×0.625 mm; field of view, 350 mm × 350 mm; matrix, 512×512. The contrast-enhanced CT image was performed after a 25-second delay following intravenous administration of 85 mL of iodinated contrast material (Ultravist 370, Bayer Schering Pharma, Berlin, Germany) at a rate of 3 mL/s with a pump injector (Ulrich CT Plus 150, Ulrich Medical, Ulm, Germany) after routine non-enhanced CT. We then retrieved the images in DICOM format from the picture archiving and communication system (PACS; Carestream, Canada).

Assessment of tumor diameter and determination of conventional CT-based LN status

Tumor diameter of primary NSCLC on preoperative radiologic finding was measured by the maximum diameter of the primary tumor on the trans-axial image of CT image. Then, the conventional CT-based LN status was established preoperatively according to European Society of Thoracic Surgeons (ESTS) guidelines (19). In these guidelines, an LN with a diameter smaller than 1 cm in the short axis is defined as negative; otherwise, it is defined as positive. All these measurements were done by an experienced radiologist [15 years of clinical experience in chest CT interpretation (Reader 1)] at a GE AW 4.3 workstation.

Radiomics features extraction

Figure 2 showed the flowchart of radiomics analysis. The corresponding algorithms for the region-of-interest (ROI) settings and calculation of all radiomics features were described in *Supplementary Materials*. In total, 591 quantitative radiomics features were extracted from each patient's contrast-enhanced chest CT images. The details of features are summarized in *Supplementary Table S1*.

To ensure the reproducibility and accuracy, 80 patients were randomly selected for reproducibility analysis. The intra- and inter- observer agreement of the radiomics feature extraction were initially analyzed using inter- and intra-class correlation coefficients (ICCs). For the assessment of the interobserver agreement of the radiomics feature extraction, two radiologists with 15 years (Reader 1)

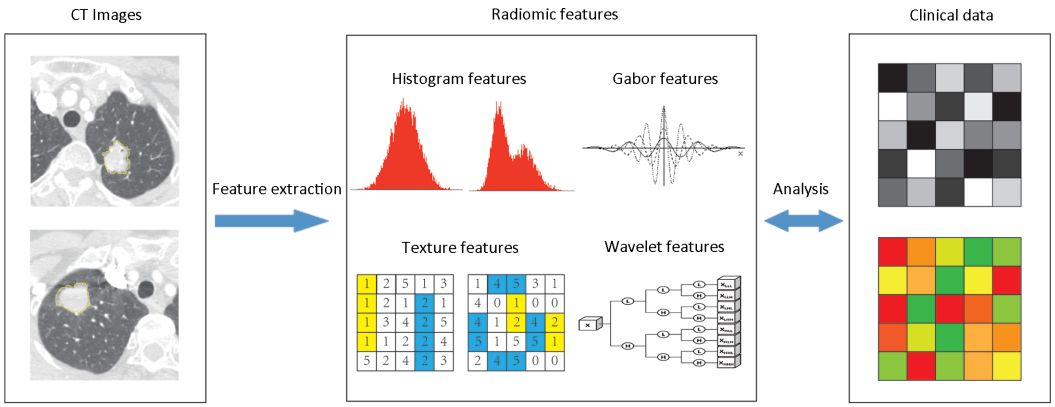

Figure 2 Flowchart of radiomics analysis pipeline. Image segmentation is performed on contrast-enhanced computed tomography (CT) images. Region of interest (ROI) is delineated initially around the tumor outline for the largest cross-sectional area by experienced radiologists. Radiomics features are extracted from the defined ROI of tumor on CT images, including tumor intensity, texture features, Gabor features, and wavelet features. For the analysis, radiomics features are compared with clinical data.

and 12 years (Reader 2) of experience in chest CT interpretation performed the ROI-based radiomics feature extraction procedure, in a blind fashion, respectively. Then, Reader 1 repeated the procedure with an interval of 1 week for the assessment of the interobserver agreement of the radiomics feature extraction. An ICC greater than 0.75 indicated good agreement (20).

Statistical analysis

Statistical analysis was performed by R software (Version 3.0.1; http://www.Rproject.org). Packages used in this study are listed in *Supplementary materials*. Two-sided P<0.05 indicated a significant difference.

Baseline continuous variables were compared between development and validation cohort using the Mann-Whitney *U* test, and categorical variables were compared using the Chi-square test. The univariate analysis of the correlation between LN positive and LN negative was conducted by using the univariate logistic analysis. Nonlinear effects of continuous variables (age, tumor diameter) were evaluated using restricted cubic splines (21,22); the relationship with LN status was not statistically significantly nonlinear (Wald P=0.049 for age; Wald P<0.001 for tumor diameter), of which age and tumor diameter could be broken as dichotomous variable (age: ≤60 years old and >60 years old; tumor diameter: ≤3 cm and >3 cm).

Radiomics-based classifier construction

To make the features values have properties of a standard normal distribution, radiomics features were normalized with z-score normalization (23). Then, we filtered features based on their independence from other features as determined by Pearson's correlation coefficient among the features (cut-off value of 0.9) (24). After that, feature selection was done in the development cohort using the least absolute shrinkage and selection operator method (LASSO) logistic regression analysis (25,26). The candidate predictive features with a zero-fit weight were selected. Thereafter, radiomics signature was built through the combination of the selected features weighted by their respective coefficients. The radiomics-based classifier was constructed with optimal cutoff point on the basis of maximum Youden index value determined by a receiver operating characteristic (ROC) analysis, of which patients were classified into high-risk or low-risk group according to the score of radiomics signature.

Derivation of fitting model for RPRS

Potential predictor variables, including age group, gender, smoking history, tumor diameter, tumor location, CT-based LN status, and radiomics-based classifier, were involved in the development of the prediction model. Multivariable logistic regression analysis with a stepwise selection procedure was used to develop the prediction model with candidate predictors based on development cohort. During modeling, backward-elimination procedures were used to confirm the significant covariates with the stopping rule of Akaike information criterion (AIC) (significance level α is set to 0.157) (27,28).

On the basis of the results of the final fitting model, a weighted scoring system (i.e., RPRS) incorporating parameters for prediction of LN metastasis was devised. Each independent predictor was a score criterion and was assigned a value based on the coefficients derived from the multivariable regression coefficients, which were multiplied by two and rounded (21).

Apparent performance assessment of RPRS

The predictive ability of RPRS was assessed in terms of discrimination and calibration. Discrimination was assessed with the C-index (28,29), and calibration was assessed graphically by plotting observed outcome against the predicted probability and tested with the Hosmer-Lemeshow goodness-of-fit test (P<0.05 implies that the model does not calibrate perfectly) (29). The RPRS system was then studied to predict LN metastasis using Chi-square test. P<0.05 was considered to be statistically significant.

Internal and external validation of RPRS

The RPRS was first validated using 1,000 resampled bootstrapping techniques based on the development cohort. Furthermore, we also assessed the validation of the RPRS using the independent validation cohort with C-index and calibration curve.

Incremental predictive value of radiomics-based classifier in RPRS for LN metastasis

The incremental predictive value of radiomics-based classifier in RPRS, compared with clinical model which was established only with clinicopathological risk factors for predicting LN metastasis, was assessed with the net reclassification improvement (NRI) (30,31). NRI is a novel prospective measure which quantifies the correctness of upward and downward reclassification or movement of

predicted probabilities as a result of adding a new marker. A value of negative percentage indicates a net worsening in risk classification for patients with LN metastases. To demonstrate the incremental value of RPRS to the PET-CT node staging, conventional CT-based LN status, and other clinicopathological risk factors for individualized assessment of LN metastasis, we selected those 277 patients with PET-CT data for comparing the predictive value. The discrimination performance was assessed with C-index.

Results

The clinical and laboratory data of development and validation cohort are listed in *Table 1*. In total, 423 patients with median age of 61 (range: 21–83) years old formed the development cohort, while another 294 patients with median age of 61 (range: 19–83) years old formed independent validation cohort. Median numbers of resected nodes were 20 (range: 5–69) in the development cohort and 18 (range: 7–50) in the independent validation cohort. Positive LN metastasis was 31.0% and 32.0% in the development and validation cohorts, respectively. There were no significant differences between two cohorts in clinical characteristics (P>0.05, *Table 1*) and LN prevalence (P=0.776).

In univariate analyses, factors that were significantly

Table 1 Baseline demographics and clinical characteristics of study population

Characteristics	All (N=717)	Development cohort (n=423)				Validation cohort (n=294)				P**
		All	LN-positive	LN-negative	P*	All	LN-positive	LN-negative	P*	
Age (year)										
Median (IQR)	61 (53, 67)	61 (54, 67)	59 (50, 65)	62 (55, 68)	0.002	61 (52, 67)	58 (52, 64)	62 (53, 69)	0.023	0.829
>60 [n (%)]	365 (50.9)	214 (50.6)	51 (38.9)	163 (55.8)	0.001	151 (51.4)	38 (40.4)	113 (56.5)	0.010	0.839
Gender [n (%)]										
Male	425 (59.3)	258 (61.0)	79 (60.3)	179 (61.3)	0.846	167 (56.8)	59 (62.8)	108 (54.0)	0.157	0.261
Female	292 (40.7)	165 (39.0)	52 (39.7)	113 (38.7)		127 (43.2)	35 (37.2)	92 (46.0)		
Smoking history [n (%)]										
Yes	246 (34.3)	152 (35.9)	43 (32.8)	109 (37.3)	0.372	94 (32.0)	36 (38.3)	58 (29.0)	0.111	0.272
No	471 (65.7)	271 (64.1)	88 (67.2)	183 (62.7)		200 (68.0)	58 (61.7)	142 (71.0)		
Tumor diameter (cm)										
Median (IQR)	25 (18, 38)	26 (19, 37)	35 (26, 48)	23 (17, 32)	<0.0001	24 (17, 38)	34 (22, 44)	23 (14, 32)	<0.0001	0.146
>3 [n (%)]	265 (37.0)	160 (37.8)	78 (59.5)	82 (28.1)	<0.0001	105 (35.7)	51 (54.3)	54 (27.0)	<0.0001	0.565
Tumor location [n (%)]										
Peripheral	575 (80.2)	340 (80.4)	73 (55.7)	267 (91.4)	<0.0001	235 (79.9)	63 (67.0)	172 (86.0)	<0.0001	0.883
Central	142 (19.8)	83 (19.6)	58 (44.3)	25 (8.6)		59 (20.1)	31 (33.0)	28 (14.0)		
CT-based LN status [n (%)]										
Negative	392 (54.7)	231 (54.6)	44 (33.6)	187 (64.0)	<0.0001	161 (54.8)	21 (22.3)	140 (70.0)	<0.0001	0.968
Positive	325 (45.3)	192 (45.4)	87 (66.4)	105 (36.0)		133 (45.2)	73 (77.7)	60 (30.0)		

IQR, inter-quartile range; CT, computed tomography; LN, lymph node; *, P value was derived from univariable association analyses between each of clinicopahologic variables and LN status; **, P value was derived from comparison of patient characteristics between development cohort and validation cohort.

associated with LN metastasis were age group, tumor diameter, tumor location, and CT-based LN status (P<0.05, *Table 1*), except for gender and smoking history (both P>0.05, *Table 1*).

Inter- and intra-observer reproducibility of radiomics features extraction

There were no statistically significant differences in the radiomics features between Reader 1's twice features extractions, nor between Reader 1's first and Reader 2's extraction (P values ranged from 0.753 to 0.807). The interobserver ICCs calculated on the basis of Reader 1's and Reader 2's first-extracted features ranged from 0.751 to 0.921, and the intraobserver ICCs calculated based on Reader 1's twice feature extractions ranged from 0.773 to 0.861. Therefore, all data analysis was based on the features extracted by Reader 1.

Radiomics-based classifier construction and corresponding discrimination performance

After filtering by correlation, 92 radiomics features which showed independence from other features were used for subsequent analysis. A correlation mapping for this radiomics features was showed in *Figure 3*. Then 13 radiomics features with non-zero coefficients in LASSO logistic regression model were selected. The radiomics signature was constructed based on the regression analysis with a radiomics score calculated for each patient (*Table 2*).

After radiomics signature construction, Mann-Whitney *U* test showed statistically significant association between the LN status and the radiomics signature both in the training and validation cohort (all P<0.0001). For the

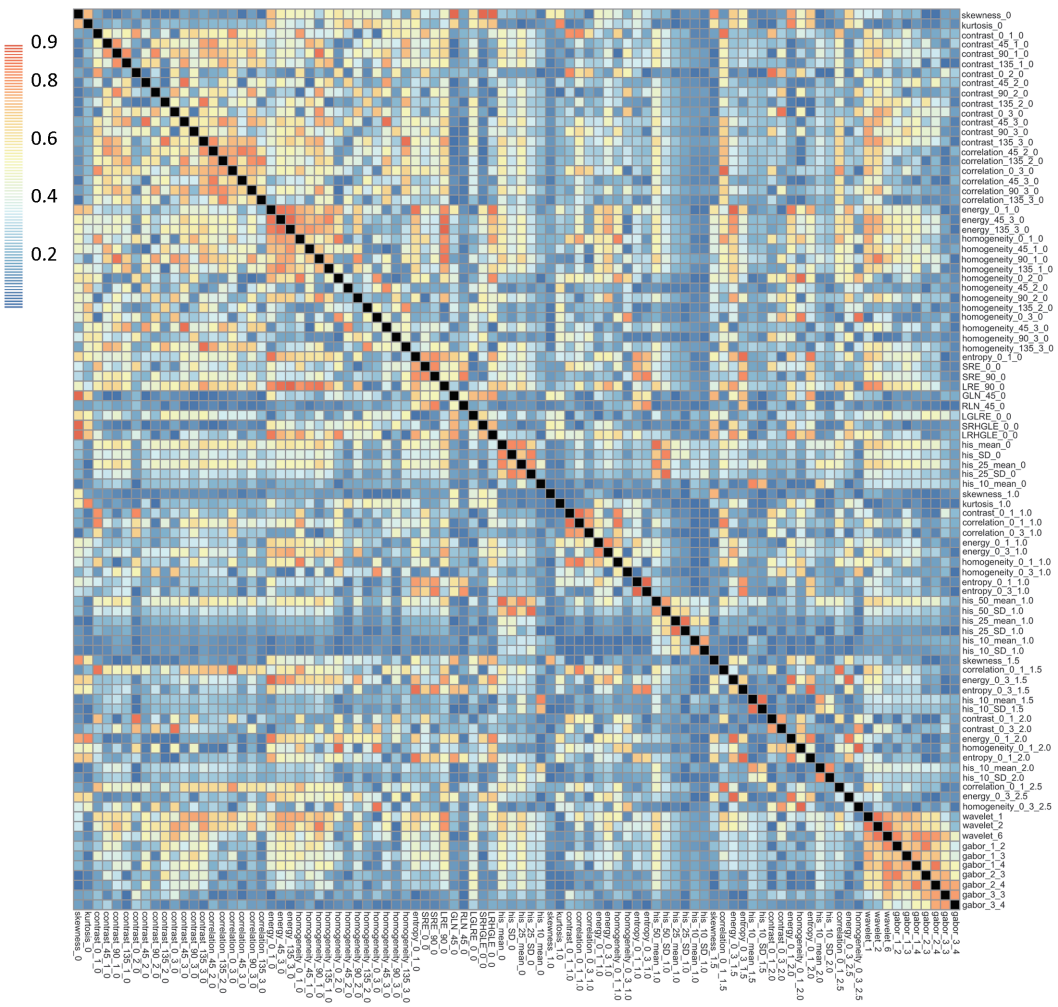

Figure 3 Correlation mapping for those radiomics features which were independent from other features. All Pearson's correlation coefficients among the features were not higher than 0.9.

Table 2 Radiomics features selected in LASSO regression analysis

Intercept and radiomics features	Coefficients
Intercept	−0.893
Contrast_135_1_0	0.099
Contrast_90_2_0	0.011
Correlation_45_3_0	−0.178
Homogeneity_135_2_0	0.241
Homogeneity_90_3_0	0.076
Gln_45_0	−0.145
His_10_Mean_0	0.075
Entropy_0_3_1.0	−0.007
His_25_SD_1.0	−0.060
Correlation_0_1_1.5	−0.186
Entropy_0_3_1.5	−0.015
His_10_SD_2.0	0.091
Gabor_2_4	−0.095

Thirteen radiomics features with non-zero coefficients in the least absolute shrinkage and selection operator method (LASSO) logistic regression model were selected. The radiomics signature was constructed based on the regression analysis with a radiomics score calculated for each patient. The formula to calculate the score of radiomics signature is *Score = Intercept + Coefficient × Radiomics features*. The details of radiomics features were described in *Supplementary Table S1*.

discrimination performance, radiomics signature yielded a C-index of 0.734 [95% confidence interval (95% CI): 0.729−0.739], Additionally, the C-index in the internal validation was 0.734. In the independent validation cohort, the discrimination performance of radiomics signature yielded a C-index of 0.734 (95% CI: 0.727−0.741).

The optimum cutoff of radiomics signature scores was generated by ROC with value of −1.048, then the radiomics-based classifier was constructed. We included those patients with a score lower than the cutoff into the low-risk group and those with a score higher than the cutoff into the high-risk group.

Derivation of RPRS

Table 3 shows the results of multivariate analysis performed in the development cohort. By multivariable logistic analysis, radiomics-based classifier and the following four independent clinical variables were found to be significant in predicting risk of LN metastasis: age group, tumor diameter, tumor location, and CT-based LN status. These four clinical variables remained independent factors in the clinical model (all P<0.05).

On the basis of findings in *Table 3*, a working score RPRS was devised. RPRS was the sum of scores for five parameters: radiomics-based classifier (low-risk =0, high-risk =3), age group (≤60 years old =1, >60 years old =0), tumor diameter (≤3 cm =0, >3 cm =1), tumor location (peripheral =0, central =3), and CT-based LN status (negative=0, positive=1) (*Table 4*). The summary score corresponding to a predicted probability of LN metastasis is shown in *Figure 4A*. The sum of all values yielded a score for each patient that ranged from 0 to 9. We applied RPRS for classifying patients into five risk categories (*Table 4*). RPRSs of 0−3, 4−5, 6, 7−8, and 9 were associated with very low (0−20%), low (21%−40%), intermediate (41%−60%), high (61%−80%), and very high (81% to 100%) risks of LN involvement, respectively.

Performances of RPRS

The calibration curves of RPRS and clinical model both in

Table 3 Logistic regression prediction models developed in development cohort

Variables	RPRS			Clinical model		
	Coefficient	OR (95% CI)	P	Coefficient	OR (95% CI)	P
Age group	−0.659	0.517 (0.316, 0.846)	0.009	−0.622	0.537 (0.333, 0.866)	0.011
Tumor diameter	0.581	1.788 (1.077, 2.969)	0.025	0.750	2.117 (1.294, 3.463)	0.003
Tumor location	1.624	5.073 (2.835, 9.079)	<0.001	1.742	5.710 (3.219, 10.128)	<0.001
CT-based LN status	0.678	1.970 (1.184, 3.276)	0.009	0.923	2.517 (1.543, 4.107)	<0.001
Radiomics-based classifier	1.514	4.546 (2.347, 8.806)	<0.001	NA	NA	NA

CT, computed tomography; LN, lymph node; RPRS, radiomics-based predictive risk score; OR, odds ratio; 95% CI, 95% confidence interval; NA, not applicable.

Table 4 Working RPRS for predicting risk of LN involvement

Score component	Score
Age group (year)	
≤60	+1
>60	0
Tumor diameter (cm)	
≤3	0
>3	+1
Tumor location	
Peripheral	0
Central	+3
CT-based LN status	
Negative	0
Positive	+1
Radiomics-based classifier	
Low-risk	0
High-risk	+3
Total (sum of score for each component)	9 (maximum)
Total RPRS	Risk of LN metastasis
0–3	Very low (0–20%)
4–5	Low (21%–40%)
6	Intermediate (41%–60%)
7–8	High (61%–80%)
9	Very high (81%–100%)

Sum score is obtained by adding scores. RPRS, radiomics-based predictive risk score; LN, lymph node; CT, computed tomography.

the development and validation cohort were satisfactory (*Figure 5*), which showed good agreement between prediction and actual observation. The Hosmer-Lemeshow test yielded a nonsignificant statistic both in the development and validation cohort (RPRS: P=0.622 for development cohort and P=0.086 for validation cohort; clinical model: P=0.546 for development cohort and P=0.084 for validation cohort).

For the discrimination performance, RPRS showed improved discrimination compared with clinical model [C-index: 0.825 (95% CI: 0.821–0.829) *vs*. 0.785 (95% CI: 0.780–0.790); P=0.005] with NRI of 0.711 (95% CI: 0.555–0.867), which was confirmed in the internal validation (C-index: 0.817 *vs*. 0.777). As for the independent validation, performance of RPRS still outperformed the clinical model [C-index: 0.810 (95% CI: 0.804–0.816) *vs*. 0.777 (95% CI: 0.771–0.783); P=0.001] with NRI of 0.425 (95% CI: 0.187–0.664).

Subgroup analysis for patients with PET-CT data

Compared with either the conventional CT-based LN status (C-index: 0.729; 95% CI: 0.723–0.35), clinical model (C-index: 0.802; 95% CI: 0.796–0.808), or PET-CT status (C-index: 0.745; 95% CI: 0.739–0.751), RPRS showed a better discrimination capability (C-index: 0.822; 95% CI: 0.816–0.828) in cohort of 277 patients with PET-CT data (P<0.05 for each comparison).

RPRS for application in clinical practice

For application in clinical practice, RPRS was presented as a chart in *Figure 4B*. The cells of the chart were classified into five groups according to ranges of the risk of LN metastasis. For each cell of the chart, the risk for a patient with values of each predictor was estimated. For example, a 56-year-old patient with a radiomics signature score of 0.136, tumor diameter of 2 cm, positive value of CT-based LN status, and central of tumor location has a 61%–80% probability of LN involvement.

Discussion

In this study, we developed and validated a simple weighted novel score RPRS composed of radiomics-based classifier and clinicopathological variables for preoperative individualized prediction of LN metastasis in patients with NSCLC. To our knowledge, this is the first study of CT-based radiomics clinical scoring system using a novel approach for evaluating risk of LN metastasis in NSCLC. Patients with RPRSs of 0–3, 4–5, 6, 7–8, and 9 had distinctly very low (0–20%), low (21%–40%), intermediate (41%–60%), high (61%–80%), and very high (81%–100%) risks of LN involvement, respectively. RPRS had successfully stratified patients into a corresponding risk of LN metastasis. Additionally, the RPRS outperformed the traditional clinicopathological variables, conventional CT-based LN status, and PET-CT status.

For the field of radiomics, intratumoral heterogeneity has been suggested to correlate with the clinical outcomes and the construction of a radiomics signature combining a panel of individual features as a prognostic imaging marker and may be a superior choice as it has now been regarded as a potential powerful method to facilitate better clinical decision making (10,12,13,15-17,32). In this study, we

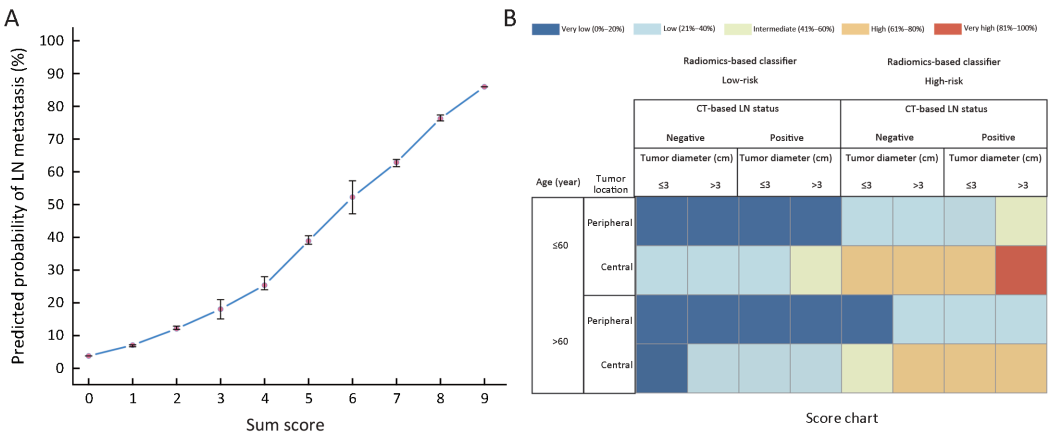

Figure 4 Presentation and visualization of radiomics-based predictive risk score (RPRS). (A) Probability of lymph node (LN) metastasis assessment in resectable non-small cell lung cancer (NSCLC) based on RPRS. X-axis presents the sum score; Y-axis presents the predicted probability of LN metastasis corresponded to the sum score from *Table 4*. After calculating a sum score for a patient, this graph can be used to determine the corresponding predicted probability with the 95% confidence interval (95% CI); (B) A graphical chart to obtain predicted individual probability for LN metastasis based on RPSR. X-axis presents the values of tumor diameter, computed tomography (CT)-based LN status, and radiomics-based classifier; Y-axis presents the values of tumor location and age group status.

analyzed 591 radiomics features to assess intratumoral heterogeneity quantitatively based on the contrast-enhanced CT image with the corresponding radiomics-based classifier constructed. The identified classifier consisted of 13 selected radiomics features, which are consisted with results of recent studies of risk stratification (10,15,16). The function of the score of the radiomics signature used in our classifier has been investigated in previous studies (15,16,32,33). The developed radiomics-based classifier showed significant association with LN metastasis, which successfully stratified patients into low-risk and high-risk group (P<0.0001).

Although no published studies found age to be an independent factor for LN status of NSCLC, in our study, age group was found to be significant association with LN status in both univariate analyses and multivariate analyses. Therefore, we kept age group as a candidate factor in the process of model development. In addition, preoperative CT-based LN status which can be easily obtained, was showed to be associated with actual LN status and was identified as an independent risk factor for the prediction of LN status.

The strengths of this study are that the final RPRS system includes radiomics-based classifier and clinicopathological parameters, which stratifies patients into categories at differing risk of LN metastasis. The observed proportion of LN positivity was satisfactorily consisted with the predicted probability according to RPRS in both the development and validation cohort in clinical practices. Unlike the traditional preoperative N staging methods, radiomics approach offers a robust way to characterize the intra-tumor heterogeneity noninvasively. Our result showed that the RPRS outperformed the clinical model not only in the development cohort, but also in the validation cohort. Both physicians and patients could perform a preoperative individualized prediction of the risk of LN metastasis using this easy-to-use clinical scoring system.

The present study has its limitations, as even though the analysis examines independent validation, it was performed in a single institute in one country. Risk factors for LN metastasis may be different in other countries because of the heterogeneity in CT image acquisition and clinical data collection in different institutions. However the available development data set is sufficiently large, splitting by time and developing a model using data from one period and evaluating its performance using the data from the other period (temporal validation) is a stronger approach in our study (28,29). Temporal validation is a prospective evaluation of a model, independent of the original data and development process according to the TRIPOD statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) (29), which can be considered external in time and thus deemed as a better alternative way to assess the generalizability of the prediction model. Furthermore, the RPRS constructed in our study was remaining a satisfactory prediction performance in the independent validation cohort, which

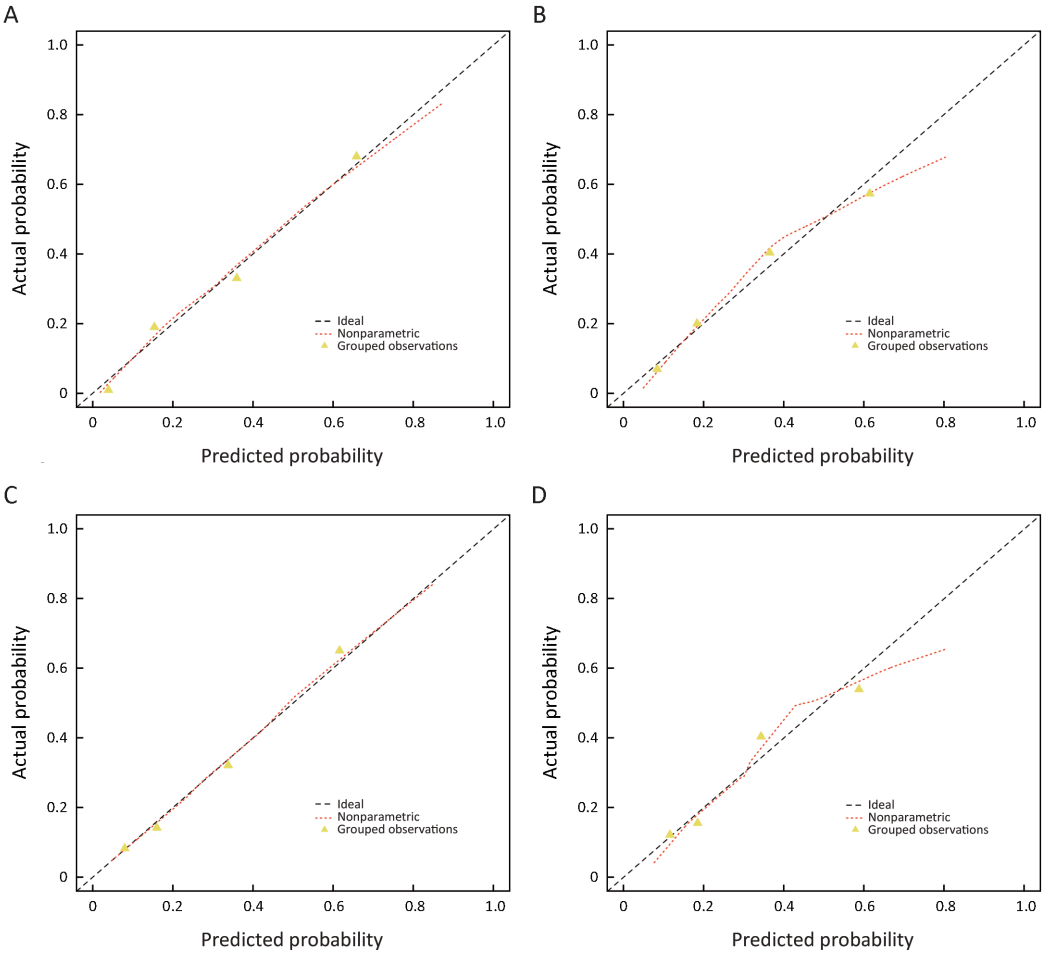

Figure 5 Calibration curves of radiomics-based predictive risk score (RPRS) and clinical model in development cohort and the independent validation cohort. RPRS (A) (P=0.622) and clinical model (C) (P=0.546) in development cohort; RPRS (B) (P=0.086) and clinical model (D) (P=0.084) in independent validation cohort. Diagonal line represents a perfect prediction.

justified that the RPRS hold great potential for clinical application in prediction of LN metastasis in patients with resectable NSCLC.

A further limitation is that whole-tumor analysis was not performed in this study. Instead, radiomics features were extracted from the largest cross-sectional area of tumor. Theoretically, whole-tumor analysis may represent more diverse component of tumor heterogeneity by avoiding sampling errors that may result from single slice selection. However, the 3D whole tumor analysis is computationally more complex and time-consuming. Previous studies have reported that there was no significant difference between texture result of 2D and 3D analysis, with the conclusion that a single slice analysis being adequate (34). Therefore, we only delineated the largest cross-sectional area for features extraction. Further investigations are warranted to exploring the potential usefulness of 3D radiomics analysis for the prediction of LN metastasis in patient with resectable NSCLC.

Conclusions

RPRS integrated radiomics-based classifier and clinicopathological factors, can serve as a novel noninvasive and easy-to-use tool to facilitate the preoperatively individualized prediction of LN metastasis in patient with resectable NSCLC.

Acknowledgements

This study was supported by the National Key Research and Development Plan of China (No. 2017YFC1309100), the National Natural Scientific Foundation of China (No. 81771912, 81901910, and 81701782), and the Provincial

Science and Technology Plan Project of Guangdong Province (No. 2017B020227012).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
2. Miller KD, Siegel RL, Lin CC, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016;66:271-89.
3. Tsim S, O'Dowd CA, Milroy R, et al. Staging of non-small cell lung cancer (NSCLC): a review. Respir Med 2010;104:1767-74.
4. Howington JA, Blum MG, Chang AC, et al. Treatment of stage I and II 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:e278S-313S.
5. Ettinger DS, Wood DE, Akerley W, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 4.2016. J Natl Compr Canc Netw 2016; 14:255-64.
6. Allen MS, Darling GE, Pechet TT, et al. Morbidity and mortality of major pulmonary resections in patients with early-stage lung cancer: initial results of the randomized, prospective ACOSOG Z0030 trial. Ann Thorac Surg 2006;81:1013-9.
7. Darling GE, Allen MS, Decker PA, et al. Randomized trial of mediastinal lymph node sampling versus complete lymphadenectomy during pulmonary resection in the patient with N0 or N1 (less than hilar) non-small cell carcinoma: results of the American College of Surgery Oncology Group Z0030 Trial. J Thorac Cardiovasc Surg 2011;141:662-70.
8. van Tinteren H, Hoekstra OS, Smit EF, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet 2002;359:1388-93.
9. Silvestri GA, Gonzalez AV, Jantz MA, et al. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd edition: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e211S-50S.
10. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
11. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. Eur J Radiol 2017;86:297-307.
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
13. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012; 30:1234-48.
14. van Timmeren JE, Leijenaar RTH, van Elmpt W, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol 2017;123:363-9.
15. Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 2016;34:2157-64.
16. Huang Y, Liu Z, He L, et al. Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. Radiology 2016;281:947-57.
17. Li H, Zhu Y, Burnside ES, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. Radiology 2016;281:382-91.
18. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114:345-50.
19. De Leyn P, Lardinois D, Van Schil PE, et al. ESTS guidelines for preoperative lymph node staging for non-small cell lung cancer. Eur J Cardiothoracic Surg 2007;32:1-8.
20. Huang Y, He L, Dong D, et al. Individualized prediction of perineural invasion in colorectal cancer: development and validation of a radiomics prediction model. Chin J Cancer Res 2018;30:40-50.
21. Frank E HJ. Regression Modeling Strategies: With

Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd edition. New York: Springer, 2015.
22. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 2008;5:e165.
23. Acharya UR, Dua S, Du X, et al. Automated diagnosis of glaucoma using texture and higher order spectra features. IEEE Trans Inf Technol Biomed 2011;15:449-55.
24. Segal E, Sirlin CB, Ooi C, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol 2007;25:675-80.
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1-22.
26. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J Royal Statistical Society 1996;58:267-88.
27. Sauerbrei W, Boulesteix AL, Binder H. Stability investigations of multivariable regression models derived from low- and high-dimensional data. J Biopharm Stat 2011;21:1206-31.
28. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73.
29. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63.
30. Pencina MJ, D'Agostino RB. Sr., D'Agostino RB. Jr., et al Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72.
31. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. JAMA 2011;305:1553-9.
32. Liang C, Huang Y, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. Oncotarget 2016;7:31401-12.
33. He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 2016;6:34921.
34. Lubner MG, Stabo N, Lubner SJ, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. Abdom Imaging 2015;40:2331-7.

Cite this article as: He L, Huang Y, Yan L, Zheng J, Liang C, Liu Z. Radiomics-based predictive risk score: A scoring system for preoperatively predicting risk of lymph node metastasis in patients with resectable non-small cell lung cancer. Chin J Cancer Res 2019;31(4):641-652. doi: 10.21147/j.issn.1000-9604.2019.04.08

Supplementary materials

Algorithm for calculation of radiomics features

All computed tomography (CT) images were gathered for tumor segmentation. The contrast-enhanced CT image data were retrieved from the institution archive and were loaded to personal laptop for further analysis. The radiomics features analysis was applied to the pretreatment CT using in-house radiomics analysis software with algorithms implemented in Matlab 2014a (Mathworks, Natick, USA). Primary tumors of all eligible patients were manually segmented by experienced radiologists with more than 10 years of clinical experience in chest CT interpretation, thus a region of interest (ROI) was delineated initially around the tumor outline for the largest cross-sectional area. The ROI was further refined by excluding air area with a thresholding procedure that removed from analysis any pixels with attenuation values below -50 HU and beyond 300 HU.

A total of 591 quantitative radiomics features were extracted from each chest contrast-enhanced CT images, which included tumor intensity, texture features, wavelet features, and Gabor features. The features of tumor intensity and texture features were extracted without/after a filtration of the Laplacian of Gaussian filter (filter parameter = 1.0, 1.5, 2.0, 2.5, respectively) from the CT image. Wavelet features and Gabor features focus the features on different frequency ranges within the tumor pixels.

Laplacian of Gaussian filtration for gray-level histogram features and texture features

The Laplacian of Gaussian filter ($\nabla^2 G$) distribution is given by

$$\nabla^2 G(x,y) = \frac{-1}{\pi\sigma^4}\left(1-\frac{x^2+y^2}{2\sigma^2}\right)e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}$$

x,y denote the spatial coordinates of the pixel and σ is the value of filter parameter.

Gray-level histogram features

1) Mean

$$mean = \frac{1}{N}\sum_{i=1}^{N} X(i)$$

2) SD

$$SD = \frac{1}{N}\sum_{i=1}^{N}\left(X(i)-\bar{X}\right)^2$$

3) Percentile mean and Percentile SD

$$\beta\_mean = \frac{1}{N-M}\sum_{i=M}^{N} X(i)$$

$$\beta\_SD = \frac{1}{N-M}\sum_{i=M}^{N}\left(X(i)-\bar{X}\right)^2$$

4) Kurtosis

$$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(X(i)-\bar{X}\right)^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X(i)-\bar{X}\right)^2}\right)^4}$$

5) Skewness

$$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(X(i)-\bar{X}\right)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X(i)-\bar{X}\right)^2}\right)^3}$$

X(i) indicates the intensity of gray level i, N denotes the sum of pixels in the image. β indicates the top percentage of the histogram curve, which could be 50%, 25%, and 10%, M denotes the number of pixels in the histogram on the percentage of (1-β).

Gray-level co-occurrence matrix features

1) Contrast

$$contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|^2 P(i,j)$$

2) Correlation

$$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP(i,j) - \mu_i(i)\mu_j(j)}{\sigma_x(i)\sigma_y(j)}$$

3) Entropy

$$entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P(i,j)\log[P(i,j)]$$

4) Energy

$$energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[P(i,j)]^2$$

5) Homogeneity

$$homo\ geneity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+|i-j|^2}$$

A matrix P(i,j) to indicate the relative frequency with intensity values of two pixels (i and j) at the three distances (δ=1,2,3) and in four directions (0°, 45°, 90°, 135°). N_g is the number of discrete intensity levels in the image;x,y denote the spatial coordinates of the pixel. μ, $\mu_x(i)$, $\mu_y(j)$ is the mean of P(i,j),P_x(i),P_y(j), and σ_x(i), σ_y(j)is the standard deviation of P_x(j), P_y(j), respectively.

Gray-level run-length matrix features

1) Short run emphasis (SRE)

$$SRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)}{j^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

2) Long run emphasis (LRE)

$$LRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2 p(i,j|\theta)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

3) Gray level non-uniformity (GLN)

$$GLN = \frac{\sum_{i=1}^{N_g}\left[\sum_{j=1}^{N_r}p(i,j|\theta)\right]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

4) Run length non-uniformity (RLN)

$$RLN = \frac{\sum_{j=1}^{N_r}\left[\sum_{i=1}^{N_g} p(i,j|\theta)\right]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

5) Run percentage (RP)

$$RP = \sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j|\theta)}{N_p}$$

6) Low gray level run emphasis (LGLRE)

$$LGLRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)}{i^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

7) High gray level run emphasis (HGLRE)

$$HGLRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} i^2 p(i,j|\theta)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

8) Short run low gray level emphasis (SRLGLE)

$$SRLGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)}{i^2 j^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

9) Short run high gray level emphasis (SRHGLE)

$$SRHGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)i^2}{j^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

10) Long run low gray level emphasis (LRLGLE)

$$LRLGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)j^2}{i^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

11) Long run high gay level emphasis (LRHGLE)

$$LRHGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)i^2 j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)}$$

A gray level run is defined as the length in number of pixels, of consecutive pixels that have the same gray level value. In a gray level run length matrix $p(i,j|\theta)$, the (i,j)th element describes the number of times j a gray level i appears consecutively in the direction specified by θ (0°, 45°, 90°, 135°), and N_g is the number of discrete gray level intensities. N_r is the number of different run lengths, and N_p is the number of pixels in the image.

Wavelet features

Wavelet transform effectively decouples textural information by decomposing the original image, in a similar manner as Fourier analysis, in low- and high-frequencies. In this study, three-level discrete wavelet transform was to applied to each CT image, which decomposes the original image X into 9 decompositions.

Gabor features

Gabor filter, named after Dennis Gabor, is a linear filter used for edge detection, which is usually used in the field of face recognition. It could select valuable image information in different directions and different scales. In this study, we used four directions and three scales to extract Gabor features. Mean was used to construct the Gabor feature group.

$$Mean\ (Gabor) = \frac{1}{N}\sum_{i=1}^{N} X\ (i)$$

Nindicates the sum of image pixels and the $X(i)$ presents the intensity i of on the Gabor image.

R packages used

The LASSO logistic regression model was done using the "glmnet" package. The multivariable logistic regression analysis and calibration plots were done with the "rms" package. The calculation of C-index and net reclassification index (NRI) were performed with the "Hmisc" package. Internal validation of the C-index was performed with the "rms" package.

Table S1 Radiomics features extracted

Type	Detail	
Gray-level histogram	skewness_σ his_mean_σ his_β _ mean _σ	kurtosis_σ his_SD_σ his_β _SD _σ
Gray-level co-occurrence matrix	contrast_α_γ_σ energy_α_γ_σ entropy_α_γ_σ	correlation_α_γ_σ homogeneity_α_γ_σ
Gray-level run-length matrix	SRE_α_σ GLN_α_σ RP_α_σ HGLRE_α_σ SRHGLE_α_σ LRHGLE_α_σ	LRE_α_σ RLN_α_σ LGLRE_α_σ SRLGLE_α_σ LRLGLE_α_σ
Wavelet	wavelet_ε	
Gabor	gabor_γ_α	

σ represents the filter value applied, which could be 1.0, 1.5, 2.0 and 2.5. When σ=0, features were extracted without filtration; α represents the considered direction, which could be 0°, 45°, 90° and 135°; β represents the top percentage of the histogram curve, which could be 50%, 25% and 10%; γ represents the considered distances, which could be 1, 2, 3.

