Original Research

Radiomics nomograms of tumors and peritumoral regions for the preoperative prediction of spread through air spaces in lung adenocarcinoma

Yaoyao Zhuo [a,1], Mingxiang Feng [b,1], Shuyi Yang [a], Lingxiao Zhou [c,e], Di Ge [b], Shaohua Lu [d], Lei Liu [c,e], Fei Shan [a,e,*], Zhiyong Zhang [a,e,f,g,*]

[a] Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, No. 2901 Caolang Road, Jinshan, Shanghai 201508, China
[b] Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China
[c] Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China
[d] Department of Pathology, Zhongshan Hospital, Fudan University, Shanghai 200032, China
[e] Research Institude of Big Data, Fudan University, Shanghai 200032, China
[f] Headmaster's Office, Fudan University, Shanghai 200433, China
[g] Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai 200032, China

ARTICLE INFO

ABSTRACT

To evaluate the clinical features and radiomics nomograms of tumors and peritumoral regions for the preoperative prediction of the presence of spread through air spaces (STAS) in patients with lung adenocarcinoma. A total of 107 STAS-positive lung adenocarcinomas were selected and matched to 105 STAS-negative lung adenocarcinomas. Thin-slice CT imaging annotation and region of interest (ROI) segmentation were performed with semi-automatic in-house software. Radiomics features were extracted from all nodules and incremental distances of 5, 10, and 15 mm outside the lesion segmentation. A radiomics nomogram was established with multivariable logistic regression based on clinical and radiomics features. The maximum diameter of the solid component and mediastinal lymphadenectasis were selected as independent predictors of STAS. The radiomics nomogram of lung nodules showed especially good prediction in the training set [area under the curve (AUC), 0.98; 95% confidence interval (CI), 0.97–1.00] and test set (AUC, 0.99; 95% CI, 0.97–1.00). The radiomics nomogram of peritumoral regions also showed good prediction, but the fitting degrees of the calibration curves were not good. Our study may provide guidance for surgical methods in patients with lung adenocarcinoma.

Introduction

Lung cancer is the leading cause of cancer death worldwide and may be associated with a unique invasion pattern [1]. Spread through air spaces (STAS) was first designated as a possible pattern of lung tumor invasion by the Kadota study team [2] and was recognized as a new manifestation of tumor spread similar to visceral pleural and vascular invasion in the 2015 World Health Organization classification [3]. STAS is defined as "micropapillary clusters, solid nests, or single cells beyond the edge of the tumor into air spaces in the surrounding lung parenchyma", which is based on pulmonary air space anatomy in the alveolar interstitium [3,4]. Some studies have shown that the presence of STAS is closely related to lower survival and worse prognosis and could provide useful clinical treatment information for patients with lung adenocarcinoma [5,6]. Ren et al. reported that STAS-positive patients tended to have poor prognosis after undergoing sublobar resection, but not after undergoing lobectomy in stage IA lung adenocarcinoma [7]. STAS has been confirmed only by biopsy so far, but several previous studies suggest that some computed tomography (CT) characteristics of lung tumors may predict the existence of STAS, such as the diameter of the tumor and the percentage of solid components in the pulmonary nodule [8–10].

Radiomics, a process of converting radiographic images into quantifiable information, can potentially improve the accuracy of diagnosis, prognosis and prediction models [11,12]. The quantifiable radiomics features extracted from the region of interest (ROI) in medical images of lesions can provide more comprehensive and richer information than radiographic

* Corresponding authors at: Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, No. 2901 Caolang Road, Jinshan, Shanghai 201508, China.
E-mail addresses: feng.mingxiang@zs-hospital.sh.cn, (M. Feng), lingxiaoz@fudan.edu.cn, (L. Zhou), ge.di@zs-hospital.sh.cn, (D. Ge), shanfei@shphc.org.cn, (F. Shan), zhyzhang@fudan.edu.cn. (Z. Zhang).
[1] Co-first authors, equal contributions to this work.

images analyzed using machine learning methods and biostatistics [13,14]. Some studies on radiomics showed that radiomics features performed well for clinical decision making in patients with lung cancer [15,16]. To date, there have been two studies on the radiomics analysis of STAS, which predicted the existence of STAS by establishing different models [17,18]. Although the results showed that STAS could be predicted preoperatively, the methods and results of the two papers were not completely consistent.

In this study, we aimed to develop a prediction model based on clinical features and a radiomics nomogram for the preoperative prediction of the presence of STAS in patients with lung adenocarcinoma.

Materials and methods

This retrospective study was approved by the institutional review board, and the requirement for written informed consent was waived.

Patients

Patients with pathologically (paraffin section) confirmed STAS-positive lung cancer were selected from August 2016 to January 2019 at Zhongshan Hospital (Shanghai, China). Patients who met any one of the following criteria were excluded from the study: no continuous thin-slice CT images (thickness < 2 mm); no plain CT images; no lung adenocarcinoma pathologically; the maximum diameter of the lesion was greater than or equal to 3 cm; or history of pulmonary surgery. Finally, a total of 107 STAS-positive patients were included in the study and matched to 105 STAS-negative patients by using patient variables (including age and sex) in the same time period at Zhongshan Hospital (Fig. 1).

CT imaging acquisition and CT imaging annotation

All CT examinations were performed in the supine body position with arms up after deep inspiration. CT data were acquired from three scanners, including SOMATOM Force (SIEMENS, Germany), Aquilion One/320 (TOSHIBA, Japan), and uCT128 (UIH, China). The CT scan parameters of the above three devices were as follows: collimation, $160*0.75$ mm, $160*0.5$ mm, $64*0.625$ mm; tube voltage, 120–130 kVp; tube current, 100–150 mAs; rotation time, 0.5–0.75 s; pitch, 0.828–1.2; matrix, $512*512$; lung window settings (width/level), $1200/-600$ HU; and mediastinal window settings (width/level), $400/40$ HU. The lung algorithm was used to reconstruct 1/1.5-mm-thick sections of CT images.

CT imaging annotation was performed by two radiologists (Y Zhuo, a doctoral student with 2 years of chest radiological experience; and F Shan, a radiologist with 19 years of experience in chest radiology) with semi-automatic in-house software. Neither radiologist knew the patients' pathology results before performing imaging annotation, and consensus was obtained by discussion in any cases of disagreement. The CT morphological characteristics assessed in our study were as follows: nodule size (maximum, mean and minimum diameter), maximum diameter of the solid component, percent of solid component, nodule type (solid, part solid or ground glass), spiculated sign, cavity, vacuole, boundary (clear or unclear), lobulated sign, air bronchogram, pleural indentation, pulmonary vessel, and mediastinal lymph node size. The percentage of solid component was calculated according to the following formula: (maximum diameter of the solid component / maximum diameter of the nodule) $*$ 100%. Lobulated sign was defined as a nodule showing jagged edges with petal-like protrusions, and the ratio of arc-chord distance/chord length was greater than or equal to 0.2. Air bronchogram was defined as the tubular low-density bronchus reaching the edge of the nodule and may or may not enter the inside of the nodule. Mediastinal lymphadenectasis meant that the short diameter of the lymph nodes was >1 cm.

Nodule segmentation and radiomics feature extraction

Lung nodule segmentation in unenhanced chest CT images was performed with semi-automatic in-house software. The boundaries of the lung nodules were checked by the radiologist and manually adjusted if necessary. It should be noted that the parts that crossed the interlobar pleura, chest wall and mediastinum should be removed. Radiomics features were extracted from nodules after image processing with different filters (including LoG, Wavelet, Square, SquareRoot, Logarithm, Exponential, Gradient, local binary pattern in 2D, and local binary pattern in 3D filters) by using open-source PyRadiomics software (https://pyradiomics.readthedocs.io/en/latest/index.html). A total of 1526 radiomics features were extracted from each ROI, including first order statistics, shape-based (3D), shape-based (2D), gray level cooccurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray tone difference matrix (NGTDM) and gray level dependence matrix (GLDM) features.

Construction and assessment of the radiomics nomogram

The least absolute shrinkage and selection operator (LASSO) method, which compresses some independent variables with little or no influence to 0, was used to select nonzero coefficients from the 1526 radiomics features [19]. A total of 12 clinical variables were involved in this experiment,
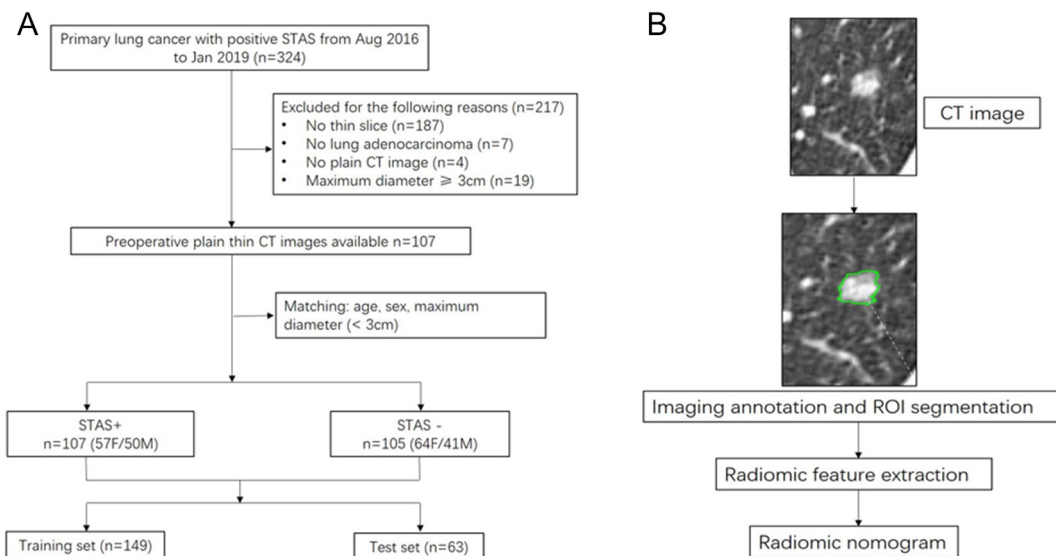

Fig. 1. A: Recruitment pathway in this study. B: Workflow of image processing. Aug: August; Jan: January; STAS + : positive STAS; STAS-: negative STAS; F: female; M: male.

which were combined with selected radiomics parameters into the multivariable logistic regression model for predicting the presence of STAS. For easier understanding, a nomogram was constructed after successfully establishing the prediction model. The experimental data of the STAS-negative group and the STAS-positive group were randomly divided into a training set and test set at a ratio of 7:3. The calibration curve was used to evaluate the calibration ability of the nomogram, and the Hosmer-Lemeshow test was used to evaluate the goodness of fit of the nomogram. The prediction accuracy of the radiomics model was represented by the receiver operating characteristic (ROC) curve and was quantified by the area under the ROC curve (AUC) in both the training and test sets. Finally, decision curve analysis was used to evaluate the clinical usefulness of the prediction model.

Peritumoral regions of lung nodules

The following 3 ROIs were extracted for each nodule using point positioning and region growing methods: incremental distances of 5, 10, and 15 mm outside the nodule segmentation. The center point of the lesion was determined according to CT imaging annotation, and a spherical shape was fitted with the maximum distance from the center point to the edge of the lesion as a radius. Finally, amplification was performed on the basis of this sphere (Fig. 2). The methods of radiomics feature extraction and radiomics nomogram construction were the same as above.

Statistical analysis

The LASSO method constructed a penalty function by adding constraint conditions, and a prediction model was constructed by performing 10-fold cross validation. Each prediction model included a clinical model and a radiomics nomogram model. DeLong's test was used to compare whether the ROC curves were different between the clinical and radiomics nomogram models.
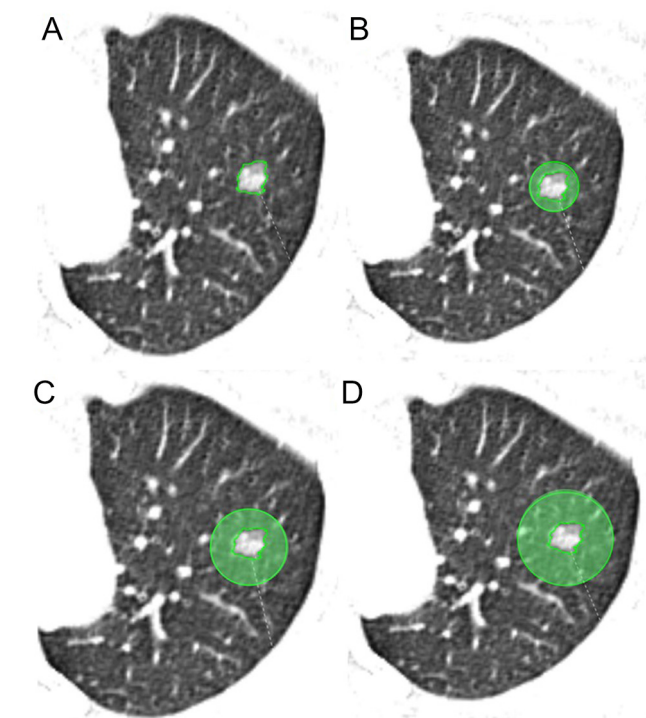

Fig. 2. A: Lung nodule segmentation. Extracting peritumoral regions: incremental distances of 5 mm (B), 10 mm (C), and 15 mm (D) outside the nodule segmentation. The center point of the lesion was determined according to CT imaging annotation, and a spherical shape was fitted with the maximum distance from the center point to the edge of the lesion as a radius. Finally, amplification was performed on the basis of this sphere.

The Wilcoxon rank sum test was used in the analysis of age, nodule size, diameter of solid component and percent of solid component because none of the above parameters were normally distributed. CT image features were compared by χ^2 tests between the STAS-positive and STAS-negative groups. SAS statistical software (version 8) was used for statistical analyses. The P value was analyzed by bilateral statistical analysis, and a P value < 0.05 indicated statistical significance. The measurement data were expressed as the mean $\pm$ standard deviation (SD).

Results

Patient clinical characteristics

There were 324 patients with pathologically confirmed STAS-positive lung cancer, but 217 patients were excluded for the following reasons: no continuous thin-slice CT images (n = 187), no plain CT images (n = 4), no lung adenocarcinoma pathologically (n = 7), and the maximum diameter of the lesion was greater than or equal to 3 cm (n = 19). The clinical characteristics of the participants in this study are shown in Table 1. In this study, there were 107 people in the STAS-positive group [50 males and 57 females; age 59.64 $\pm$ 9.54 years (mean $\pm$ SD)] and 105 people in the STAS-negative group [41 males and 64 females; age 58.05 $\pm$ 10.27 years (mean $\pm$ SD)]. Postoperative pathology showed no significant differences in pleural and bronchial invasion between the two groups (P = 0.0525 and 0.5916, respectively), but the STAS-positive group had more lymphatic metastasis than the STAS-negative group (P = 0.0069). In addition, the results of genetic testing indicated that epidermal growth factor receptor (EGFR) mutation and Napsin A expression were significantly different between the two groups (P = 0.0041 and 0.0000, respectively). Other indicators without statistical significance are shown in Table S1.

The CT image characteristics of the patients are shown in Table 2. There were significant differences in the maximum diameter of the nodule, maximum diameter of the solid component, percent of the solid component, nodule type (solid, part solid or ground glass), spiculated sign, boundary (clear or unclear), lobulated sign, air bronchogram and pleural indentation between the two groups (P = 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,

Table 1
The clinical characteristics of the participants.

	All patients (n = 212)	Negative for STAS (n = 105)	Positive for STAS (n = 107)	P value
Age (year)	58.84 $\pm$ 9.92	58.05 $\pm$ 10.27	59.64 $\pm$ 9.54	0.1960
Sex				0.2703
Female	121 (57.08)[b]	64 (60.95)	57 (53.27)	
Male	91 (42.92)	41 (39.05)	50 (46.73)	
Pleural invasion				0.0525
Present	46 (21.70)	13 (12.38)	33 (30.84)	
Absent	119 (56.13)	54 (51.43)	65 (60.75)	
NA	47 (22.17)	38 (36.19)	9 (8.41)	
Bronchial invasion				0.5916
Present	3 (1.42)	1 (0.95)	2 (1.87)	
Absent	153 (72.17)	84 (80.00)	69 (64.49)	
NA	56 (26.41)	20 (19.05)	36 (33.64)	
Lymphatic metastasis				0.0069[a]
Present	19 (8.96)	5 (4.76)	14 (13.08)	
Absent	137 (64.62)	82 (78.09)	55 (51.40)	
NA	56 (26.42)	18 (17.5)	38 (35.52)	
EGFR				0.0041[a]
Present	104 (49.06)	41 (39.05)	63 (58.88)	
Absent	108 (50.94)	64 (60.95)	44 (41.12)	
Napsin A				<0.001[a]
Present	141 (66.51)	41 (39.05)	100 (93.46)	
Absent	66 (31.13)	64 (60.95)	2 (1.87)	
NA	5 (2.36)	0	5 (4.67)	

Abbreviations: NA, not available.
[a] Statistically significant.
[b] Data are numbers of patients, with percentages in parentheses.

Table 2
The CT image characteristics of the participants.

	All patients (n = 212)	Negative for STAS (n = 105)	Positive for STAS (n = 107)	P value
Maximum diameter of nodule (mm)	15.31 ± 6.01	12.36 ± 3.88	18.21 ± 6.04	<0.001[a]
Minimum diameter of nodule (mm)	11.64 ± 4.71	9.67 ± 3.68	13.58 ± 4.70	<0.001[a]
Mean diameter of nodule (mm)	13.47 ± 5.53	11.01 ± 3.68	15.89 ± 5.20	<0.001[a]
Maximum diameter of solid component (mm)	12.12 ± 7.91	7.82 ± 5.91	16.34 ± 6.04	<0.001[a]
Percent of solid component (%)	75.64 ± 41.35	61.13 ± 37.13	90.24 ± 15.41	<0.001[a]
Nodule type				<0.001[a]
Solid	92 (43.40)[b]	34 (32.38)	58 (54.21)	
Part solid	100 (47.17)	51 (48.57)	49 (45.79)	
Ground glass	20 (9.43)	20 (19.05)	0 (0)	
Spiculated sign				<0.001[a]
Present	160 (75.47)	62 (59.05)	98 (91.59)	
Absent	52 (24.53)	43 (40.95)	9 (8.41)	
Cavity				0.7552
Present	9 (4.25)	4 (3.81)	5 (4.67)	
Absent	203 (95.75)	101 (96.19)	102 (95.33)	
Vacuole				0.0843
Present	38 (17.92)	14 (13.33)	24 (22.43)	
Absent	174 (82.08)	91 (86.67)	83 (77.57)	
Boundary				<0.001[a]
Clear	127 (59.91)	46 (43.81)	81 (75.70)	
Unclear	85 (40.09)	59 (56.19)	26 (24.30)	
Lobulated sign				0.0273[a]
Present	193 (91.04)	91 (86.67)	102 (95.33)	
Absent	19 (8.96)	14 (13.33)	5 (4.67)	
Air bronchogram				0.0013[a]
Present	132 (62.26)	54 (51.43)	78 (72.90)	
Absent	80 (37.74)	51 (48.57)	29 (27.10)	
Pleural indentation				<0.001[a]
Present	128 (60.38)	46 (43.81)	82 (76.64)	
Absent	84 (39.62)	59 (56.19)	25 (23.36)	
Pulmonary vessel				0.9773
Vessel convergence	182 (85.85)	86 (81.90)	96 (89.72)	
Vessel expansion	4 (1.89)	2 (1.91)	2 (1.87)	
Absent	26 (12.26)	17 (16.19)	9 (8.41)	
Mediastinal lymphadenectasis				0.4134
Present	24 (11.32)	10 (9.52)	14 (13.08)	
Absent	188 (88.68)	95 (90.48)	93 (86.92)	

[a] Statistically significant.
[b] Data are numbers of patients, with percentages in parentheses.

0.0000, 0.0273, 0.0013, and 0.0000, respectively). The remaining CT features, including cavity, vacuole, location, pulmonary vessel, and mediastinal lymph node size, were not statistically significant.

Radiomics feature extraction and radiomics signature construction

The radiomics parameters selected by using the LASSO method were different for each ROI. A total of seven features with nonzero coefficients were selected from the lung nodules, including first-order statistical, NGTDM, GLSZM, GLCM and GLDM features (Fig. 3).

The radscore was calculated by summing the selected features weighted by their coefficients and then adding a constant (−0.045) (Table S2). We compared the radscore from all patients on the training and test sets, and ROC analysis was used to evaluate the performance of the model. The results showed that the radscores of the STAS-negative group were lower than those of the STAS-positive group, and the differences were statistically significant (P = 0.0000 and 0.0000, respectively) (Fig. S1). ROC analysis showed good performance in the training set [AUC, 0.88; 95% confidence interval (CI), 0.82–0.93] and test set (AUC, 0.86; 95% CI, 0.77–0.95).

Construction, performance and validation of the radiomics nomogram

Univariate and multivariate logistic regression analyses were performed on the clinical data, and independent predictors of STAS were selected. The selected predictors related to STAS in the regression analysis results were the maximum diameter of the solid component and mediastinal lymph node size.

After obtaining multivariate logistic regression equations based on radiomics, a nomogram model was established to calculate the probability of STAS for each patient (Fig. 4A). ROC and decision curves were used to evaluate the clinical usefulness of the prediction model in both the training and test sets. The radiomics nomogram of lung nodules, consisting of seven selected radiomics parameters and clinical features, showed good prediction in the training set (AUC, 0.98; 95% CI, 0.97–1.00) and test set (AUC, 0.99; 95% CI, 0.97–1.00) (Fig. 4B–C). In addition, the results also showed that the AUC of the nomogram model was slightly larger than that of the clinical model in both the training and test sets, but there were no statistically significant differences between the nomogram and clinical model in either the training or test set (P = 0.2108 and 0.1324, respectively).

The calibration curve was resampled 1000 times using the self-service method to ensure the accuracy of the results. The Hosmer-Lemeshow test showed that the P value was 0.8209 in the training set, indicating that the fitting degree of the model was good. Similarly, the P value was 0.9703 in the test set, which also showed good calibration ability (Fig. 4D–E). The results above showed that the performance of the nomogram in both groups was good.

Clinical use

Decision curve analysis was used to evaluate the clinical usefulness of the prediction model (Fig. S2). Compared with the cases of treat-all and treat-none, both the clinical and nomogram models could bring net benefits to patients, of which the nomogram model added more benefits.
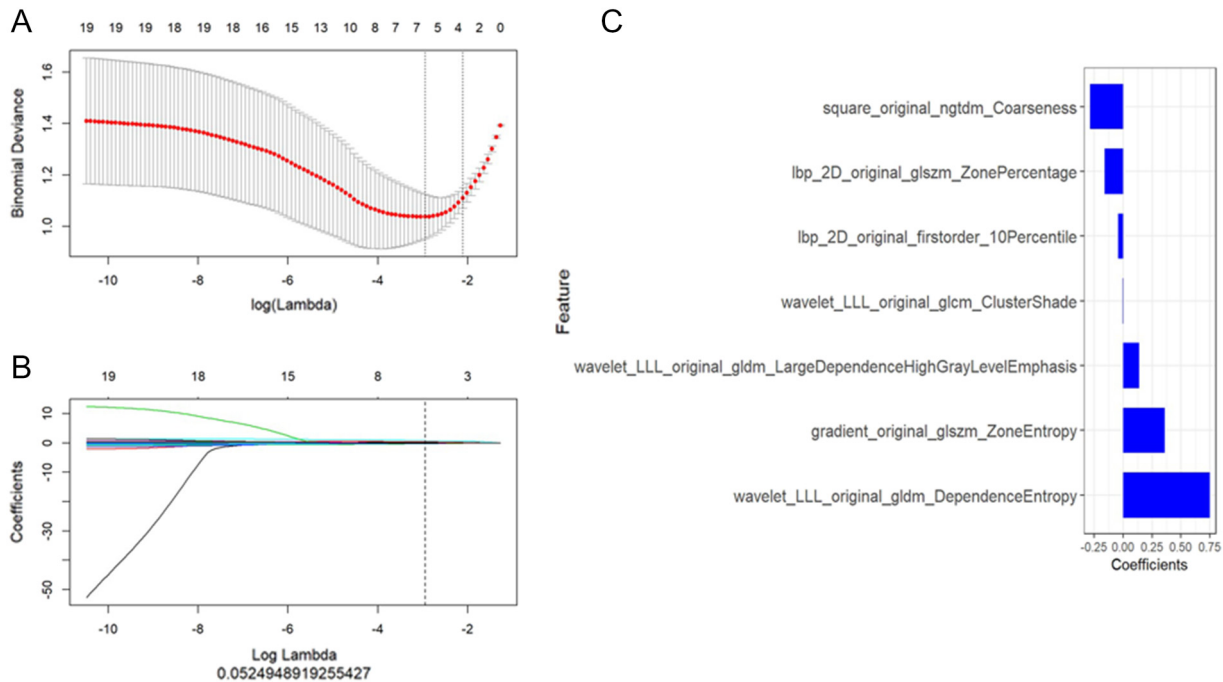
Fig. 3. Radiomics feature selection. The least absolute shrinkage and selection operator (A) included choosing the regularization parameter λ (B) and determining the number of features. A total of seven radiomics features were chosen (C).

Results of peritumoral regions

Intervals of 5, 10 and 15 mm outside the lung nodule created the peritumoral regions. Six features with nonzero coefficients were selected from the 5-mm peritumoral region, including first order statistics, GLCM, GLRLM and GLSZM. Seven features with nonzero coefficients were selected from the 10-mm peritumoral region, including first order statistics, GLSZM and GLCM. Ten features with nonzero coefficients were selected from the 15-mm peritumoral region, including first order statistics and GLSZM. The detailed data of the radiomics parameters selected from the three ROIs are shown in Fig. S3.

The prediction models of all peritumoral regions also achieved favorable prediction efficacy, but the Hosmer-Lemeshow test showed that the P values of the test set were 0.0426, 0.0000 and 0.0216, respectively, which indicated that the calibration curves were departed from good fit (Fig. S4).

Discussion

In our study, there were significant differences in the qualitative and quantitative CT characteristics between the STAS-negative and STAS-positive groups. Among these variables, the maximum diameter of the solid component and mediastinal lymphadenectasis were selected to build a prediction model for preoperatively predicting the presence of STAS. The radiomics nomogram of lung nodules, consisting of seven selected radiomics parameters and two clinical features, showed good prediction in both the training and test sets. However, there were no statistically significant differences between the nomogram model and clinical model in either the training set or test set. In addition, the discriminative ability of the radiomics nomogram of lung nodules was better than that of peritumoral regions.

STAS was reported to be an independent prognostic factor for poor recurrence-free survival and overall survival in previous studies, even in early-stage lung adenocarcinoma [6,20]. Therefore, the presence of STAS can influence clinical decision-making for surgery. The stepwise flowchart provided by Suh et al. showed that segmentectomy was proper for patients with negative STAS, and lobectomy was proper for patients with positive STAS [21]. In our study, postoperative pathology showed no significant differences in pleural and bronchial invasion between the two groups (P = 0.0525 and 0.5916, respectively). These results conflict with the findings reported by Sun et al. that the positive rate of STAS was significantly higher than that in lung tumors with pleural and bronchial invasion [22]. The STAS-positive group had more lymphatic metastasis than the STAS-negative group (P = 0.0069) in our study, which was compatible with the results of a previous study [6]. In addition, the results of genetic testing indicated that EGFR mutation and Napsin A expression were significantly different between the two groups (P = 0.0041 and 0.0000, respectively). The genetic mechanism of STAS in lung cancer has yet to be clarified. Because postoperative pathology cannot be used for preoperative prediction, these results were not included as factors of the prediction model.

Some CT imaging features were related to the presence of STAS in our study. The maximum diameter of the solid component and mediastinal lymphadenectasis were selected to be independent predictors of STAS. Kim et al. found that the percentage of solid components was independently related to STAS by building a prediction model using CT features [9]. They reported that the AUC in the model that used the percentage of the solid component (AUC = 0.77) was larger than that in the model that used the maximum diameter of the solid component (AUC = 0.66), which indicated that the percentage of the solid component was a more useful CT imaging predictor than the maximum diameter of the solid component [8]. Another study found that lobulated sign and nodule type were independent predictors of STAS by multivariable analysis, and the presence of lobulated sign and the absence of ground glass opacity were independently related to STAS [10]. Our results showed that solid nodules were most likely to be STAS positive, while pure ground glass nodules were least likely to be STAS positive, which was similar to the findings of previous research [8,9]. There was a significant difference in the lobulated sign between the two groups, but it was not selected to be an independent predictor of STAS.

The AUCs of the clinical model in the training and test sets were 0.98 and 0.97, respectively. The radiomics nomogram of lung nodules, combining clinical and radiomics features, showed better prediction in the training set (AUC, 0.98) and test set (AUC, 0.99). A total of seven features with non-zero coefficients were selected from the lung nodules, including first-order statistical, NGTDM, GLSZM, GLCM and GLDM features. First order statistics
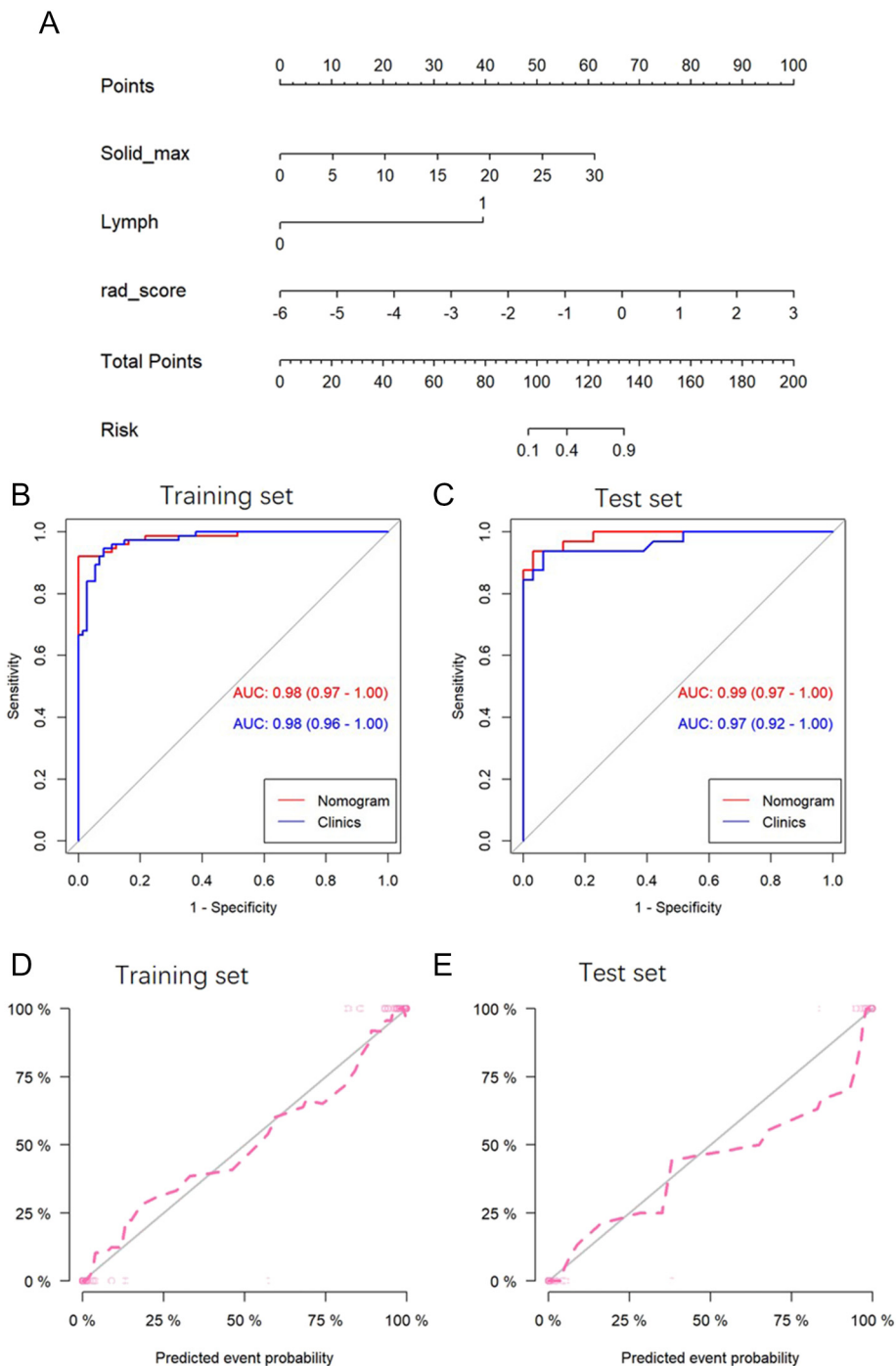
Fig. 4. Construction, performance and validation of the radiomics nomogram. A: The radiomics nomogram was developed using seven selected radiomics parameters and two clinical features. ROC curves of the nomogram and clinical model in the training (B) and test (C) sets. The calibration curves of the radiomics nomogram in the training (D) and test (E) sets.

are used to describe the distribution of voxel intensity in the image region, which is defined by the common mask and basic metrics. Second order statistics, including GLCM, GLRLM and GLSZM, need to involve the spatial position relationship with the voxel intensity. Higher order statistics, such as Fourier transform, wavelet decomposition, and other filtering, can be used for image preprocessing and feature extraction. Radiomics features are automatically extracted by the computer, which compensates for mistakes caused by subjective and manual measurements. Our results showed that the AUC of the nomogram model was slightly larger than that of the clinical model in both the training and test sets, but there were no statistically significant differences between the nomogram

and clinical models in either the training set or test set (P = 0.2108 and 0.1324, respectively), which indicated that CT image features could provide plenty of information for making a preliminary judgment on the existence of STAS. Jiang et al. developed a random forest model using CT-based radiomics features and achieved an AUC of 0.754 for predicting the existence of STAS [18]. Another study built a Naïve Bayes model using five radiomics features to predict STAS that showed good performance with an AUC of 0.63 in internal validation and an AUC of 0.69 in external validation [17]. The prediction effect of our nomogram model was better than the results of the above two studies. After further study, we speculated that this might be related to the

inclusion of many CT clinical features in the model building process, such as nodule size, nodule type, and other CT features. As shown in Fig. 4B–C, the clinical model of lung nodules showed good prediction in both the training set (AUC, 0.98) and test (AUC, 0.97) set. Before extracting parameters, we used 9 filters to process the CT images and then extracted 1526 radiomics parameters for each ROI. This method was far more involved than those of the above two studies but might provide richer information for building prediction models. In addition, STAS-positive patients accounted for approximately half of the total population in our study, while the STAS-positive rates in the above two studies were < 30%, which may mean that different data compositions may also influence the experimental results.

Dai et al. found that the maximum distances between the lung nodule edge and STAS were 1.35 cm and 0.87 cm in the study and validation cohorts, respectively [20]. Therefore, intervals of 5, 10 and 15 mm outside the lung nodule were selected to create the peritumoral regions. The prediction models of all peritumoral regions also achieved favorable prediction efficacy, but the Hosmer-Lemeshow test showed that the P values of the test set were 0.0426, 0.0000 and 0.0216, respectively, which indicated that the calibration curves were departed from good fit. The prediction model established by the radiomics parameters of lung tumors was better than the prediction model established by radiomics parameters in peritumoral regions.

There were several limitations in our study. First, this was a single-center retrospective study, and the predictions were unexpectedly good (although they were repeatedly confirmed), which may suggest that we need to validate this predictive model with an external validation set. Second, when the STAS-negative group and the STAS-positive group were matched, the smoking status was not matched because of incomplete data. There is a close relationship between smoking and lung tumors, which may influence the results. Third, the CT images in the study came from three different CT machines, and the images were not processed with consistency, which may affect the experimental results. Fourth, we only evaluated the relationship between STAS and lung adenocarcinoma, and other pathological types of lung tumors, such as lung squamous cell carcinoma, also needed further investigation.

In conclusion, the radiomics nomogram of lung nodules showed favorable prediction efficacy for the presence of STAS and may provide guidance for surgical methods in patients with lung adenocarcinoma.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2020.100820.

Author contributions

Yaoyao Zhuo: the acquisition of data, analysis of data, and drafting the article;

Mingxiang Feng: the acquisition of data, the conception of the study;

Shuyi Yang: the acquisition of data;

Lingxiao Zhou: the analysis of data;

Di Ge: the acquisition of data;

Shaohua Lu: the acquisition of data;

Lei Liu: the analysis and interpretation of data;

Fei Shan: the conception and design of the study, revising the article critically for important intellectual content;

Zhiyong Zhang: final approval of the version to be submitted.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] World Health Organization, Lung cancer, http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. (Accessed 11 August 2018).

[2] K. Kadota, J. Nitadori, C.S. Sima, H. Ujiie, N.P. Rizk, D.R. Jones, et al., Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas, J. Thorac. Oncol. 10 (2015) 806–814.

[3] W.D. Travis, E. Brambilla, A.G. Nicholson, Y. Yatabe, J.H.M. Austin, M.B. Beasley, et al., The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification, J. Thorac. Oncol. 10 (2015) 1243–1260.

[4] A.R. Shih, M. Mino-Kenudson, Updates on spread through air spaces (STAS) in lung cancer, Histopathology (2020). https://doi.org/10.1111/his.14062.

[5] Y. Liu, D. Chen, X. Qiu, S. Duan, Y. Zhang, F. Li, et al., Relationship between MTA1 and spread through air space and their joint influence on prognosis of patients with stage I-III lung adenocarcinoma, Lung Cancer 124 (2018) 211–218.

[6] Y. Terada, T. Takahashi, S. Morita, K. Kashiwabara, K. Nagayama, J.I. Nitadori, et al., Spread through air spaces is an independent predictor of recurrence in stage III (N2) lung adenocarcinoma, Interact. Cardiovasc. Thorac. Surg. 29 (2019) 442–448.

[7] Y. Ren, H. Xie, C. Dai, Y. She, H. Su, D. Xie, et al., Prognostic impact of tumor spread through air spaces in sublobar resection for 1A lung adenocarcinoma patients, Ann. Surg. Oncol. 26 (2019) 1901–1908.

[8] C. de Margerie-Mellon, A. Onken, B.H. Heidinger, P.A. VanderLaan, A.A. Bankier, CT manifestations of tumor spread through airspaces in pulmonary adenocarcinomas presenting as subsolid nodules, J. Thorac. Imaging 33 (2018) 402–408.

[9] S.K. Kim, T.J. Kim, M.J. Chung, T.S. Kim, K.S. Lee, J.I. Zo, et al., Lung adenocarcinoma: CT features associated with spread through air spaces, Radiology 289 (2018) 831–840.

[10] G. Toyokawa, Y. Yamada, T. Tagawa, T. Kamitani, Y. Yamasaki, M. Shimokawa, et al., Computed tomography features of resected lung adenocarcinomas with spread through air spaces, J. Thorac. Cardiovasc. Surg. 156 (2018) 1670–1676.

[11] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2016) 563–577.

[12] G.W. Ji, Y.D. Zhang, H. Zhang, F.P. Zhu, K. Wang, Y.X. Xia, et al., Biliary tract cancer at CT: a radiomics-based model to predict lymph node metastasis and survival outcomes, Radiology 290 (2019) 90–98.

[13] P. Giraud, P. Giraud, A. Gasnier, R. El Ayachy, S. Kreps, J.P. Foy, et al., Radiomics and machine learning for radiotherapy in head and neck cancers, Front. Oncol. 9 (2019) 174.

[14] T. De Perrot, J. Hofmeister, S. Burgermeister, S.P. Martin, G. Feutry, J. Klein, et al., Differentiating kidney stones from phleboliths in unenhanced low-dose computed tomography using radiomics and machine learning, Eur. Radiol. 29 (2019) 4776–4782.

[15] W. Shi, L. Zhou, X. Peng, H. Ren, Q. Wang, F. Shan, et al., HIV-infected patients with opportunistic pulmonary infections misdiagnosed as lung cancers: the clinicoradiologic features and initial application of CT radiomics, J. Thorac. Dis. 11 (2019) 2274–2286.

[16] Y. Huang, Z. Liu, L. He, X. Chen, D. Pan, Z. Ma, 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 281 (2016) 947–957.

[17] D. Chen, Y. She, T. Wang, H. Xie, J. Li, G. Jiang, et al., Radiomics-based prediction for tumour spread through air spaces in stage I lung adenocarcinoma using machine learning, Eur. J. Cardiothorac. Surg. (2020). https://doi.org/10.1093/ejcts/ezaa011.

[18] C. Jiang, Y. Luo, J. Yuan, S. You, Z. Chen, M. Wu, et al., CT-based radiomics and machine learning to predict spread through air space in lung adenocarcinoma, Eur. Radiol. (2020). https://doi.org/10.1007/s00330-020-06694-z.

[19] M.M. Vasquez, C. Hu, D.J. Roe, Z. Halonen, S. Guerra, Measurement error correction in the least absolute shrinkage and selection operator model when validation data are available, Stat. Methods Med. Res. 28 (2019) 670–680.

[20] C. Dai, H. Xie, H. Su, Y. She, E. Zhu, Z. Fan, et al., Tumor spread through air spaces affects the recurrence and overall survival in patients with lung adenocarcinoma >2 to 3 cm, J. Thorac. Oncol. 12 (2017) 1052–1060.

[21] J.W. Suh, Y.H. Jeong, A. Cho, D.J. Kim, K.Y. Chung, H.S. Shim, et al., Stepwise flow-chart for decision making on sublobar resection through the estimation of spread through air space in early stage lung cancer, Lung Cancer 142 (2020) 28–33.

[22] P.L. Sun, J.N. Liu, L.Q. Cao, M. Yao, H.W. Gao, To evaluate the clinicopathologic characteristics and outcome of tumor cells spreading through air spaces in patients with adenocarcinoma of lung, Zhonghua Bing Li Xue Za Zhi 46 (2017) 303–308.