European Journal of Nuclear Medicine and Molecular Imaging (2021) 48:231–240
https://doi.org/10.1007/s00259-020-04924-6

ORIGINAL ARTICLE

Value of ^{18}F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis

Yujing Hu[1,2] · Xinming Zhao[1] · Jianyuan Zhang[3] · Jingya Han[1] · Meng Dai[1]

Received: 27 February 2020 / Accepted: 9 June 2020 / Published online: 25 June 2020

Abstract

Purpose To develop a predictive model by ^{18}F-FDG PET/CT radiomic features and to validate the predictive value of the model for distinguishing solitary lung adenocarcinoma from tuberculosis.

Methods A total of 235 ^{18}F-FDG PET/CT patients with pathologically or follow-up confirmed lung adenocarcinoma ($n = 131$) or tuberculosis ($n = 104$) were retrospectively and randomly divided into a training ($n = 163$) and validation ($n = 72$) cohort. Based on the Transparent Reporting of Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD), this work was belonged to TRIPOD type 2a study. The Mann-Whitney U test and least absolute shrinkage and selection operator (LASSO) algorithm were used to select the optimal predictors from 92 radiomic features that were extracted from PET/CT, and the optimal predictors were used to build the radiomic model in the training cohort. The meaningful clinical variables comprised the clinical model, and the combination of the radiomic model and clinical model was a complex model. The performances of the models were assessed by the area under the receiver operating characteristic curve (AUC) in the training and validation cohorts.

Results In the training cohort, 9 radiomic features were selected as optimal predictors to build the radiomic model. The AUC of the radiomic model was significantly higher than that of the clinical model in the training cohort (0.861 versus 0.686, $p < 0.01$), and this was similar in the validation cohort (0.889 versus 0.644, $p < 0.01$). The AUC of the radiomic model was slightly lower than that of the complex model in the training cohort (0.861 versus 0.884, $p > 0.05$) and validation cohort (0.889 versus 0.909, $p > 0.05$), but there was no significant difference.

Conclusion ^{18}F-FDG PET/CT radiomic features have a significant value in differentiating solitary lung adenocarcinoma from tuberculosis.

Keywords ^{18}F-FDG · PET/CT · Radiomic feature · Pulmonary tuberculosis · Lung adenocarcinoma

Introduction

Lung cancer (LC) is still the main cause of cancer worldwide. Non-small-cell lung cancer (NSCLC) accounts for more than 85% of lung cancer incidence, and NSCLC is the most common type of LC. Moreover, lung adenocarcinoma is the most common pathological subtype of NSCLC [1]. Since some benign lesions are similar to LC in terms of clinical and radiological aspects, those lesions are easily misdiagnosed as LC. Tuberculosis is the most common misdiagnosed cause in the region with a high incidence of tuberculosis, which invades the lung and has affected a quarter of the world's population [2]. Therefore, distinguishing lung cancer, especially adenocarcinoma, from tuberculosis has become a common and challenging issue.

A large number of imaging techniques, including dynamic contrast-enhanced CT [3] and MRI [4], PET/CT with ^{18}F-

This article is part of the Topical Collection on Oncology – Chest

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-020-04924-6) contains supplementary material, which is available to authorized users.

✉ Xinming Zhao
xinm_zhao@163.com

[1] Department of Nuclear Medicine, The Fourth Hospital of Hebei Medical University, 12 Jiankang Road, Shijiazhuang 050011, Hebei, China

[2] Department of Nuclear Medicine, Hebei General Hospital, Shijiazhuang 050051, Hebei, China

[3] Department of Nuclear Medicine, Baoding No. 1 Central Hospital, Baoding 071000, Hebei, China

FDG [5] or other radioactive tracers [6] and serological biomarkers [7] have been applied to distinguish the two diseases. However, no single method has satisfactory performance. Currently, biopsy or resection as an invasive technique is the most accurate method to distinguish between the two diseases. Nevertheless, patients with tuberculosis should avoid invasive techniques and are used to being treated with drugs rather than surgical resection [8]. Thus, there is a pressing need for non-invasive, accurate, convenient and specific techniques to differentiate between solitary lung adenocarcinoma and tuberculosis.

Radiomics is a promising and noninvasive approach that can extract high-throughput quantitative features from images and convert the information into mineable databases. It reflects the lesion's biological information, for instance, cell morphology and molecular and gene expression, which can improve the diagnostic, prognostic and predictive accuracy of the disease [9]. ^{18}F-FDG PET/CT has been proven to be a crucial method for detection, clinical staging, therapeutic evaluation and recurrence in patients with NSCLC [10, 11]. Recently, there have been some studies about radiomics based on CT to distinguish benign pulmonary nodules from lung cancer [12, 13]. However, there have been few studies on radiomics based on ^{18}F-FDG PET/CT to discriminate between the two diseases. We hypothesized that ^{18}F-FDG PET/CT radiomic features may possess valuable information to improve the accuracy of distinguishing solitary lung adenocarcinoma from tuberculosis. Therefore, the aim of this study was to develop a predictive model by ^{18}F-FDG PET/CT radiomic features and to validate the predictive value of the model to distinguish solitary lung adenocarcinoma from tuberculosis.

Materials and methods

Patient selection

We retrospectively screened and collected ^{18}F-FDG PET/CT and clinical data about patients in The Fourth Hospital of Hebei Medical University and Hebei General Hospital from January 2015 to October 2019. The institutional ethics committee approved all the data in the study for retrospective analysis and waived the demand for informed consent.

The enrolled patients met the following criteria: (1) patients were diagnosed with lung adenocarcinoma or pulmonary tuberculosis by surgery, biopsy pathology or follow-up; (2) ^{18}F-FDG PET/CT examination performed before surgery or biopsy; (3) density of observed lesions without characteristic calcification or fat; and (4) patients with the result of T-SPOT.TB. The following was determined as the positive result: the number of spots in negative control holes was less than 5, and the difference in the number of spots between antigen A or B and negative control holes was more than 6, or the number of spots in negative control holes was 6–10, and the number of spots in antigen A or B was more than twice that in negative control holes [14]. The exclusion criteria were as follows: (1) a lesion without the accumulation of ^{18}F-FDG, (2) maximum diameter of the lesion < 1 cm and (3) other malignancies on or before the performance of PET/CT. The clinical data included age, gender and the result of T-SPOT.TB.

According to the above criteria, 235 patients (138 males and 97 females, mean age ± SD, 62.79 ± 12.14 years, range 17–90) were enrolled. The data of patients were randomly divided into two cohorts by the ratio of 7:3. The ratio was based on hold-out method which is a model evaluation method, and the 7:3 is the commonest ratio to divide training cohort and verification cohort. The grouping is shown in Fig. 1. According to the guidelines of Transparent Reporting of Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [15], the data was randomly split into the training cohort and validation cohort, and this work was TRIPOD 2a study.

PET/CT examination

^{18}F-FDG PET/CT examination was performed on either PHILIPS GEMINI GXL16 (Eindhoven, Netherlands, 102 patients, 43.4%) or General Electric Discovery Elite (Waukesha, WI, 133 patients, 56.6%). According to the reporting guidelines of the Image Biomarker Standardisation Initiative (IBSI) [16], all the acquisition parameters are shown in electronic supplementary material 1. Patients fasted for at least 6 h before being injected with ^{18}F-FDG (3.7–5.55 MBq/kg), and image acquisition started 60 ± 5 min later from the bottom of the skull to the midthigh on the basis of the guidelines of the European Association of Nuclear Medicine [17]. PET images were CT-based attenuation corrected.

Radiomic feature extraction

PET and CT images with the DICOM format were successively input into LIFEx freeware (v4.00, http://www.lifexsoft.org) and automatically fused by the freeware. When respiratory movements led to a mismatch between the two types of images, CT images could be manually adjusted to the correct position of the lesion. Compared with the spatial resampling of 1 mm and 4 mm, 2 mm had a slight influence on the calculation of radiomic feature [18]. Spatial resampling was 2 mm in spacing X, Y and Z both on PET and CT images in all the patients. Intensity discretization for CT data was processed by decreasing the continuous scale to 400 bins with absolute scale bounds between − 1000 and 3000 HU, while that of PET data was done with 64 bins between 0 and 25. Two well-experienced nuclear medicine physicians used

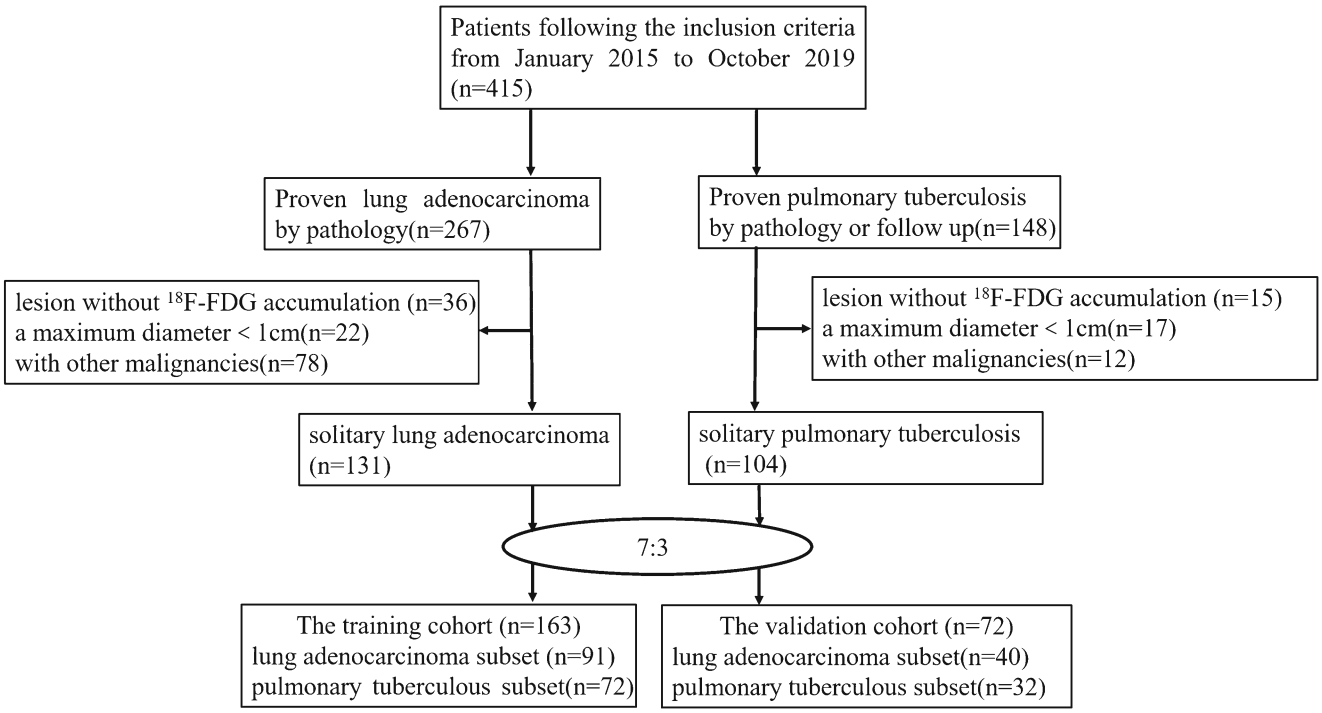

Fig. 1 Selection of patients. The flowchart shows the process of selection and grouping

3D drawing tools to manually delineate the volume of interest (VOI) on every slice of the PET images and then used 40% of the maximum standardized uptake value (SUVmax) as a threshold to optimize the VOI [19]. Because of the good matching of PET and CT images, radiomic features were extracted from the two types of images within the same VOI. According to the VOI, LIFEx freeware automatically processed and extracted 92 radiomic features: 47 features based on PET and 45 features from CT. The 92 features are shown in electronic supplementary material 2.

Radiomic feature screening and model building

In the training cohort, the optimum features were screened from 92 radiomic features to build a radiomic model, and clinical variables were selected to develop a clinical model. Additionally, the complex model was the combination of the radiomic model and the clinical model by multivariate logistic regression, which included the optimum radiomic features and clinical variables with significant differences. The predictive performance of the models was assessed by receiver operating characteristic (ROC) curves and the area under the ROC curve (AUC) in the training and validation cohorts, respectively. A nomogram was established to obtain the predictive probability of every patient, and calibration curves were drawn by the predicted probability against the actual probability, which were used to analyse the performance of the nomogram.

Statistical analysis

The Mann-Whitney *U* test was applied to screen the candidate features from the 92 radiomic features between the two diseases in the training cohort. The clinical data, which included continuous variables (age and gender) and categorical variables (T-SPOT.TB), were compared by the Mann-Whitney *U* test and Chi-square test, respectively. And the statistically different variables were used to develop the clinical model by logistic regression. The above analyses used SPSS 21.0 Statistics for Windows.

To avoid multicollinearity and overfitting phenomena, LASSO (least absolute shrinkage and selection operator) algorithm was employed to further screen the optimal subset from the 61 candidate features and develop a formula by logistic regression in the training set to calculate the score of radiomic features (RAD-score) [20]. Ten-fold cross-validation was performed to select the λ in the LASSO algorithm, and the 1 standard error of λmin (the min of λ) was the optimal λ, which could choose the optimal subset of radiomic features. The radiomic model was developed by multivariate logistic regression with the optimal subset of radiomic features. The LASSO algorithm was carried out by the "glmnetcr" package of R software 3.4.3 (http://www.Rproject.org). A heat map was applied to visually show the difference of every feature in the optimal features by the "pheatmap" package of R software. Waterfall plots were performed to demonstrate the RAD-score for every enrolled patient in the training and validation cohorts. The complex model was the combination of

the radiomic model and clinical model by multivariate logistic regression in SPSS 21.0.

The performances of the models were assessed by the ROC curve and AUC in the training and validation cohorts, respectively. MedCalc statistical software 19.0 was used to determine the difference between the AUC values. To appraise the robustness of the models, ten-fold cross-validation with 100 random cycles was performed by the "caret" and "AUC" packages of R software, and the uncertainty of the AUC, sensitivity, specificity, accuracy and Youden index of those models in the training cohort was assessed as the average of 100 means ± standard deviation (SD). The Youden index, also known as the correct index, is a method to evaluate the authenticity of screening tests. If the false-negative and false-positive results have the same significance, the Youden index can be applied. The Youden index is the sum of sensitivity and specificity minus 1, which represents the total ability of screening methods to find real patients and nonpatients. The meaning of the Youden index is that the larger the index is, the better the effect of the screening experiment and the greater the authenticity.

The dependence of the selected radiomic features and RAD-score on the type of scanner used (PHILIPS or GE) was compared with the Mann-Whitney *U* test.

Nomograms and calibration curves were generated using the "rms" package in R software. The goodness of fit was examined by the Hosmer-Lemeshow test in SPSS. The Hosmer-Lemeshow test is a goodness of fit (GOF) test that indicates whether the predicted probability is close to the actual probability. When the *p* value was > 0.05, we concluded that it was a good fit. The Mann-Whitney *U* test, Chi-square test and Hosmer-Lemeshow test as well as the difference between the AUC values obtained a two-sided *p*, and $p < 0.05$ was considered statistically significant.

Results

Clinical characteristics of patients

The clinical characteristics of the enrolled patients are shown in Table 1. T-SPOT.TB in the tuberculosis subset was much higher than that in the lung adenocarcinoma subset in the training cohort ($p < 0.0001$) and the validation cohort ($p = 0.0188$). In terms of age and gender, there were no significant differences not only in the two cohorts ($p = 0.180$ and 0.281) but also in the tuberculosis and lung adenocarcinoma subsets ($p = 0.621$ and 0.505). Since T-SPOT.TB was significantly different but age and gender were not different between the two cohorts, T-SPOT.TB was used to develop a clinical model by univariate logistic regression.

Feature extraction and selection

The value of − 3.093 was eventually chosen for the optimal λ, and then 61 radiomic candidate features were reduced to 9 with nonzero coefficients in the training cohort (electronic supplementary material 3). The 9 selected radiomic features were Conventional_SUVstandard deviation (SUVstd), Histogram_Skewness (HISTO_Skewness), Histogram_Kurtosis (HISTO_Kurtosis), GreyLevel Co-occurrence Matrix_Correlation (GLCM_Correlation), maxValue, Histogram_Skewness (HISTO_Skewness CT), SHAPE_Sphericity (only for 3D ROI nZ > 1) (SHAPE_Sphericity), GreyLevel Co-occurrence Matrix_Contrast (GLCM_Contrast) and GreyLevel Co-occurrence Matrix_Entropy_log10 (GLCM_Entropy_log10). Among the 9 features, the first 4 were derived from PET, while the 5 remaining features were related to CT. The distinction of the 9 radiomic features between the lung adenocarcinoma and tuberculosis subsets is shown in

Table 1 Characteristics of patients in the training and validation cohorts

Characteristic	Training cohort no. (%)			Validation cohort no. (%)		
	Pulmonary tuberculosis	Lung adenocarcinoma	*p*	Pulmonary tuberculosis	Lung adenocarcinoma	*p*
Age (mean ± SD, years)	60.57 ± 13.29	62.03 ± 9.24	0.4283	63.34 ± 9.40	62.38 ± 8.94	0.6567
Gender			0.0574			0.132
Male	46 (63.89)	44 (48.35)		18 (56.25)	30 (75.00)	
Female	26 (36.11)	47 (51.65)		14 (43.75)	10 (25.00)	
T-SPOT.TB			< 0.0001			0.0188
Positive	45 (62.50)	23 (25.27)		22 (68.75)	16 (40.00)	
Negative	27 (37.50)	68 (74.73)		10 (31.25)	24 (60.00)	
PET/CT scanner			0.5277			0.8126
PHIL IPSGEMINI GXL16	33 (45.83)	37 (41.66)		15 (46.88)	17 (42.50)	
GE Discovery Elite	39 (54.17)	54 (59.34)		17 (53.13)	23 (57.50)	

SD standard deviation

the heat map (electronic supplementary material 4) in the training cohort, and HISTO_Skewness from PET had a more obvious distinction than the remaining features. The RAD-score of every enrolled patient was calculated according to the formula from logistic regression as follows: $-1.1490 + 2.9686 \times SHAPE\_Sphericity + 2.4531 \times GLCM\_Correlation - 1.3451 \times HISTO\_Skewness - 0.5425 \times GLCM\_Entropy\_log10 - 0.3090 \times HISTO\_Kurtosis - 0.0415 \times HISTO\_Skewness\ CT + 0.0147 \times SUVstd - 0.0031 \times maxValue - 0.0003 \times GLCM\_Contrast$.

There was a significant difference in the 9 features between the tuberculosis and lung adenocarcinoma subsets in the two cohorts (Table 2). The RAD-score in the lung adenocarcinoma subset was much higher than that in the tuberculosis subset both in the training cohort (0.904 versus −0.417) and validation cohort (0.830 versus −0.638), and there was a significant difference between the two subsets in the training and validation cohorts (all $p < 0.0001$). The RAD-score for every enrolled patient is exhibited by the waterfall plot in Fig. 2.

Performances of the radiomic, clinical and complex models

The AUC of the radiomic model (0.861 and 0.889) was higher than that of the clinical model (0.686 and 0.644) in the training ($p = 0.0002$) and the validation ($p = 0.0005$) cohorts. The AUC of the complex model was the highest (0.884 and 0.909) among the three models either in both the training and validation cohorts. In terms of the AUC, there were significant differences between the complex model and the clinical model in the two cohorts (both $p < 0.0001$), while there was no significant difference between the complex model and the radiomic model in the training ($p = 0.1181$) and validation ($p = 0.3219$) cohorts. The ROC curves of the three models are shown in Fig. 3. The sensitivity, specificity, accuracy and Youden index of those models in the validation cohort and the uncertainties of those parameters in the training cohort are summarized in Table 3.

Scanner comparability

The 9 radiomic features, the RAD-score and the AUC of the radiomic model were stratified by the different scanners (PHILIPS or GE) and showed statistical compatibility in both the training and validation cohorts. The 9 radiomic features and RAD-score were not significantly different (all $p > 0.05$), and the details are shown in electronic supplementary material 5. The AUCs of the radiomic model classified by PHILIPS and GE were 0.91 versus 0.84 in the training cohort and 0.96 versus 0.86 in the validation cohort. There

Table 2 Comparison of the 9 selected radiomic features and radiomic signature score (RAD-score) between pulmonary tuberculosis subset and lung adenocarcinoma subset in training cohort and validation cohort

Characteristic	Training cohort Median (interquartile range)		*p*	Validation cohort Median (interquartile range)		*p*
	Pulmonary tuberculosis	Lung adenocarcinoma		Pulmonary tuberculosis	Lung adenocarcinoma	
RAD-score	−0.417 (−1.290, 0.319)	0.904 (0.447, 1.356)	< 0.0001	−0.638 (−1.876, 0.120)	0.830 (0.381, 1.244)	< 0.0001
Features PET						
SUVstd	0.800 (0.525, 1.200)	1.200 (0.600, 1.800)	0.0065	0.700 (0.400, 1.100)	1.200 (0.700, 1.775)	0.0017
HISTO_Skewness	0.510 (0.293, 0.778)	0.170 (−0.090, 0.380)	< 0.0001	0.515 (0.207, 0.817)	0.200 (0.060, 0.423)	0.0007
HISTO_Kurtosis	2.810 (2.538, 3.268)	2.500 (2.300, 2.780)	< 0.0001	2.940 (2.463, 3.535)	2.535 (2.410, 2.660)	0.0016
GLCM_Correlation	0.692 (0.593, 0.761)	0.729 (0.639, 0.805)	0.0175	0.704 (0.617, 0.757)	0.756 (0.674, 0.798)	0.0071
Features CT						
maxValue	148.500 (69.200, 444.950)	71.300 (52.000, 89.500)	< 0.0001	207.35 (82.967, 537.500)	73.950 (61.750, 117.975)	0.0020
HISTO_Skewness CT	−0.260 (−1.113, 0.248)	−0.880 (−1.980, −0.390)	0.0006	−0.055 (−1.048, 0.487)	−1.210 (−2.005, −0.608)	0.0002
SHAPE_Sphericity	0.960 (0.930, 1.000)	0.980 (0.960, 1.010)	0.0138	0.940 (0.920, 0.975)	0.980 (0.970, 1.000)	0.0007
GLCM_Contrast	276.007 (148.517, 384.650)	198.974 (102.523, 280.696)	0.0017	258.351 (199.265, 342.006)	156.368 (98.377, 274.634)	0.0022
GLCM_Entropy_log10	2.730 (2.443, 2.968)	2.590 (2.140, 2.790)	0.0040	2.855 (2.537, 3.098)	2.420 (2.173, 2.750)	0.0003

SUVstd Conventional_SUVstandard deviation, *HISTO_Skewness* Histogram_Skewness, *HISTO_Kurtosis* Histogram_Kurtosis, *GLCM_Correlation* GreyLevel Co-occurrence Matrix_Correlation, *SHAPE_Sphericity* SHAPE_Sphericity (only for 3D ROI nZ>1), *GLCM_Contrast* GreyLevel Co-occurrence Matrix_Contrast, *GLCM_Entropy_log10* GreyLevel Co-occurrence Matrix_Entropy_log10

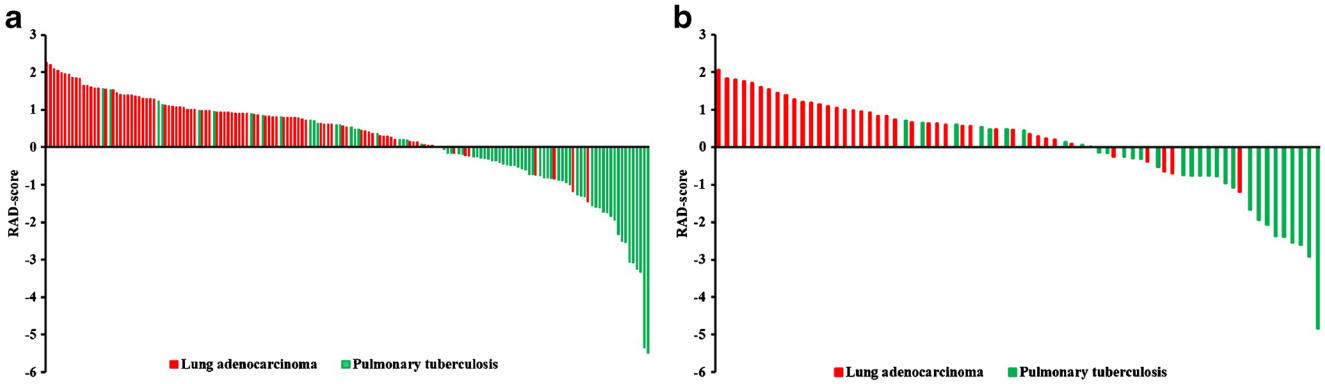

Fig. 2 Score of radiomic features (RAD-score) for patients. Waterfall plots were drawn according to the RAD-score for every patient in the training cohort (**a**) and validation cohort (**b**)

were minor absolute discrepancies but no significant difference ($p = 0.4114$ and 0.4617) in the two cohorts.

Individualized nomogram construction and validation

According to the good predictive ability of the complex model, a nomogram that could demonstrate the proportion of every factor and predictive value of every patient was generated from the training cohort (Fig. 4a).

The calibration curves were drawn, and they showed that the predicted probability was in line with the actual probability in the two cohorts (Fig. 4b and c). The goodness of fit examined by the Hosmer-Lemeshow test showed no significant difference, which demonstrated good fits for the predicted and actual probabilities in both the training cohort ($\chi 2 = 6.855$, $p = 0.552$) and validation cohort ($\chi 2 = 8.492$, $p = 0.387$).

Discussion

In the present study, radiomics based on ^{18}F-FDG PET/CT was first used for differentiating solitary lung adenocarcinoma from tuberculosis. The radiomic model showed good predictive performance (AUC > 0.86), and the performance was higher than that of the clinical model and was similar to that of the complex model.

The 9 radiomic features are composed of 4 PET-derived features and 5 CT-derived features. The SUVstandard deviation is the only feature related to the standard intake value (SUV), which is the basic parameter of PET. SUVstandard deviation abbreviated as SUVstd assesses the variability of the metabolic activity, which is relatively sensitive to statistical noise [21]. The region of high receptor density shows more apparent statistical noise than the region of low or no receptor density [22]. In this study, SUVstd in the lung adenocarcinoma subset was slightly higher

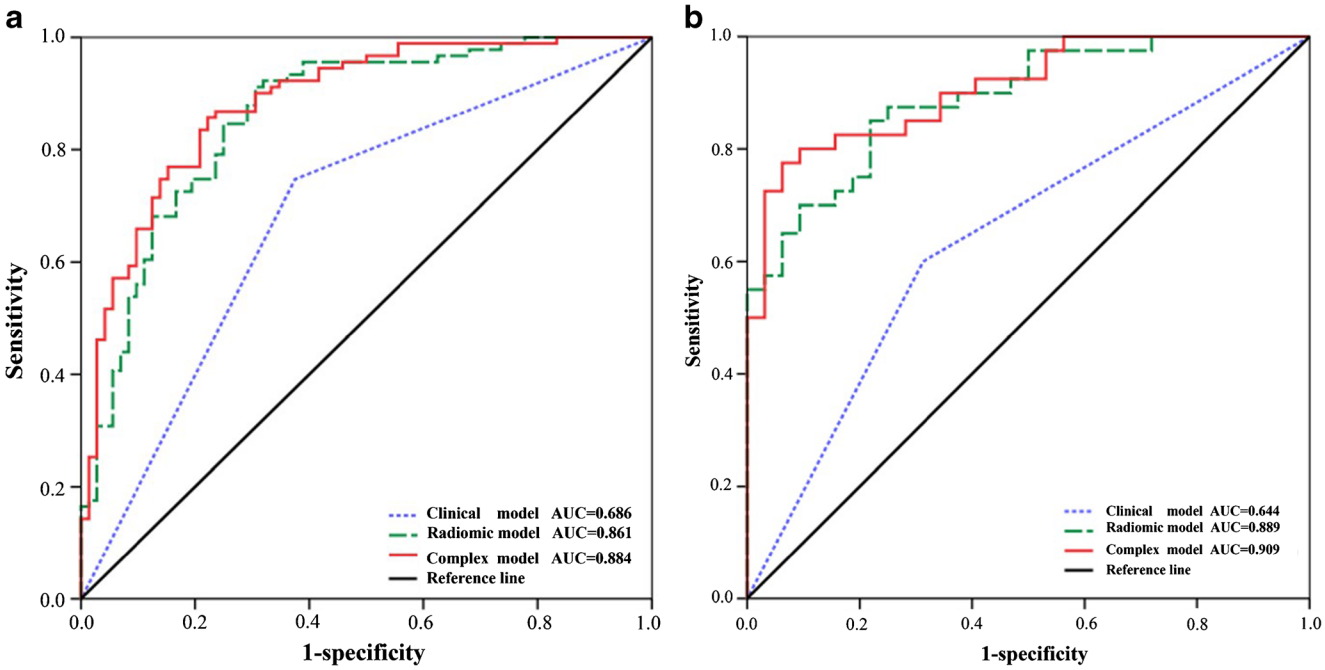

Fig. 3 Area under the receiver operating characteristic (ROC) curve of the models **a** in the training cohort and **b** in the validation cohort

Table 3 Predictive performance of three models in the training and validation cohorts

Model	Training cohort (10-fold cross-validation*, mean ± SD)					Validation cohort				
	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	Youden index	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	Youden index
Clinical model	0.686 ± 0.011	72.10 ± 0.57	62.09 ± 0.83	65.55 ± 0.28	0.342 ± 0.014	0.644	60.00	68.75	65.75	0.288
Radiomic model	0.861 ± 0.012	81.78 ± 0.51	69.07 ± 0.88	77.48 ± 0.60	0.509 ± 0.007	0.889	85.00	78.12	79.53	0.6313
Complex model	0.884 ± 0.009	82.50 ± 0.59	73.97 ± 0.80	78.03 ± 0.84	0.565 ± 0.006	0.909	77.50	93.75	92.53	0.7125

AUC area under receiver operating characteristic curve
SD standard deviation
*Random 100 cycles

than that in the tuberculosis subset in both cohorts, which reflected that the receptor density in lung adenocarcinoma was slightly higher than that in tuberculosis.

HISTO_Skewness derived from either PET or CT and HISTO_Kurtosis from PET are 3 features from the histogram that represent the grey-level distribution within the volume of interest. HISTO_Skewness represents the asymmetric distribution of grey level in the histogram, while HISTO_Kurtosis shows the shape (peaked or flat) of the distribution of the grey level, and those features are relatively simple parameters that describe the heterogeneity of lesions [23]. Chandarana et al. [24] found that clear cell renal cell cancer (ccRCC) had significantly lower kurtosis and skewness than papillary renal cell cancer (pRCC), which reflected that ccRCC was less heterogeneous than pRCC. In our study, the 3 radiomic features in the tuberculosis subset were significantly higher than those in the lung adenocarcinoma subset in the two cohorts. This may be related to caseous necrosis in tuberculosis.

There were still 3 radiomic features related to the GreyLevel Co-occurrence Matrix (GLCM): one PET feature (GLCM_Correlation) and two CT features (GLCM_Contrast, GLCM_ Entropy_log10). The GLCM quantifies the intensity distribution of the grey level at a given offset to extract information about tone homogeneity, linear connection, contrast and boundaries adjacent to grey zones, as well as complicacy of the distribution [25]. According to the introduction of LIFEx (https://www.lifexsoft.org/index.php/resources/19-texture/radiomic-features), the offset in this study was applied to 1. GLCM_Correlation shows the linear correlation, which is related to the grey level in the GLCM, and it was slightly higher in lung adenocarcinoma than in tuberculosis in both cohorts. GLCM_Contrast, which is sometimes named Variance or Inertia, expresses the local varieties of the grey level, while GLCM_Entropy_log10 represents the randomicity of voxel pairs at the grey level. GLCM_Contrast and GLCM_Entropy_log10 in lung adenocarcinoma were lower than those in tuberculosis in the two cohorts. Tuberculous granuloma is composed of macrophages, including T lymphocytes, B lymphocytes, dendritic cells, fibroblasts and extracellular matrix components [26]. The performance of the 3 radiomic features of the GLCM is related to the spatial distribution of cells inside the lesion.

The remaining 2 features were maxValue and SHAPE_Sphericity from CT. MaxValue in tuberculosis was significantly higher than that in lung adenocarcinoma. Tuberculosis is more likely to form calcification and fibrosis than lung cancer [27], which is the main reason for the high maxValue. SHAPE_Sphericity has a similar degree of sphericity regarding the volume of interest. In previous studies, the 3D shape features PET/CT-derived had high repeatability and accurately offered morphological parameters about the lesions [28]. In this study, lung adenocarcinoma had higher sphericity and was more compact than tuberculosis.

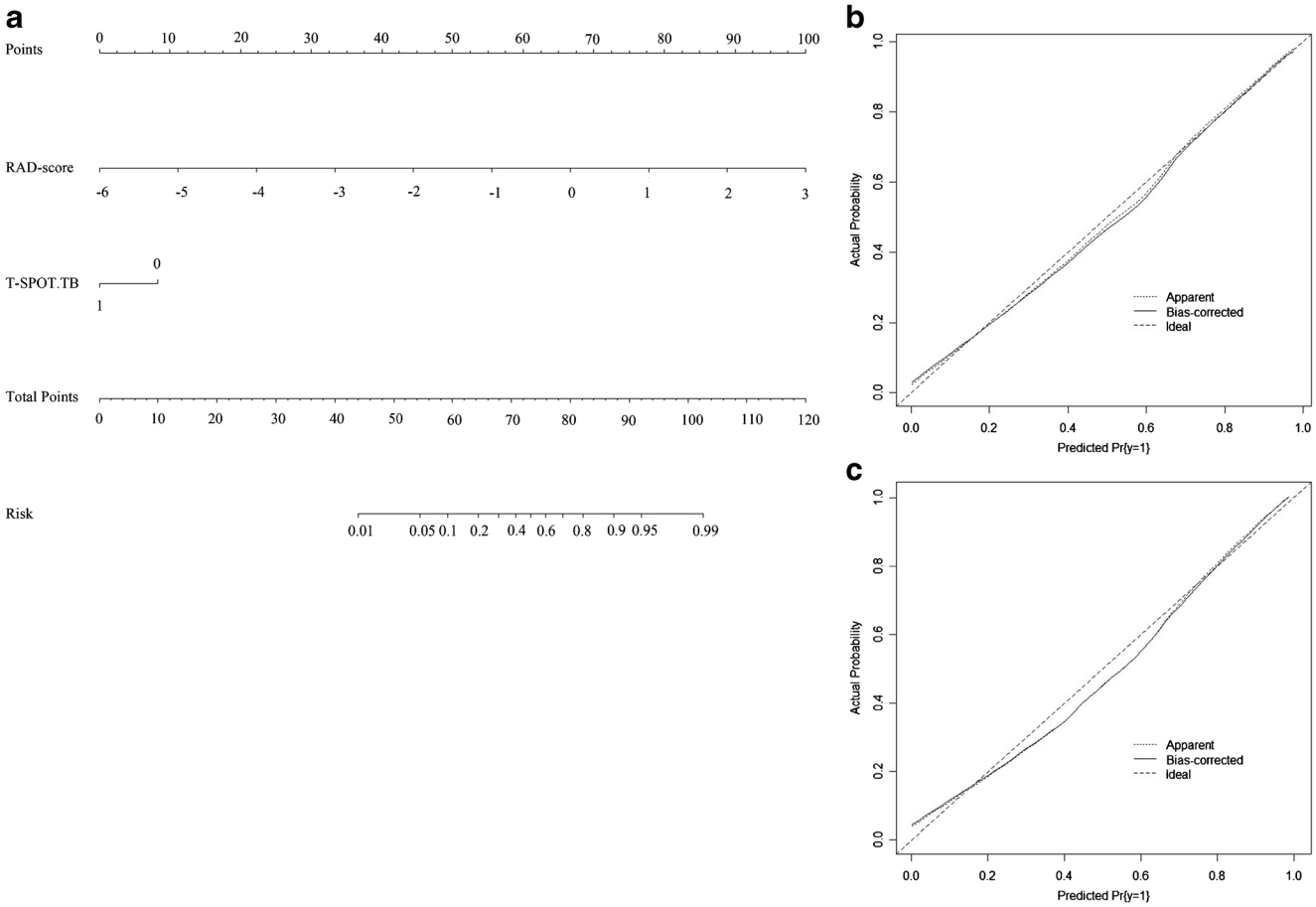

Fig. 4 Development of the nomogram and calibration curves. The nomogram was based on the RAD-score and clinical factor (T-SPOT.TB) (**a**). Calibration curves showed the predicted versus actual probability for solid lung adenocarcinoma in the training cohort (**b**) and the validation cohort (**c**)

T-SPOT.TB is an in vitro interferon gamma release assay used to diagnose tuberculosis by the specific cellular immune response to *Mycobacterium tuberculosis*, which was of higher sensitivity and specificity than traditional means such as tuberculin skin tests [14]. Large-scale studies have indicated that the sensitivity of T-SPOT.TB in patients with suspected TB was more than 80%, while the specificity was higher than 70% [29]. In our study, the sensitivity and specificity were slightly lower than those in previous studies, which might be related to most enrolled patients with older age [30].

Yang et al. [12] showed that combining CT-based radiomic features and clinical variables could improve the predictive performance to differentiate solitary granulomatous nodules, including tuberculosis, from lung adenocarcinoma. However, this study only considered CT-based radiomic features but did not take PET-based radiomic features into account. In our study, there were 4 PET-derived features among the 9 selected features, which added metabolic information of the lesions and improved diagnostic performance. Therefore, we believe that the radiomic model based on PET/CT has the ability to differentiate solid lung adenocarcinoma and tuberculosis.

Several limitations should be considered in our study. First, this study was a retrospective analysis, and inherent selection bias existed. In future research, a larger number of patients should be recruited to limit the bias as much as possible. Second, our research only discussed the predictive performance of the model for adenocarcinoma, but whether this model can be applied to other subtypes of lung cancer should be further studied. Third, tumour markers of lung adenocarcinoma, such as CEA, were not involved in our study. Future studies could examine the value of tumour markers in the complex model.

In conclusion, ^{18}F-PET/CT-based radiomic features have a significant value in differentiating solitary lung adenocarcinoma from tuberculosis. Large-scale multicentre studies should be carried out to further confirm the preliminary results so that this noninvasive, effective and convenient technique can be applied in routine clinical practice.

Acknowledgements The authors would like to acknowledge all the coworkers who participated in this study.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Xinming Zhao, Yujing Hu, Jianyuan Zhang, Jingya Han and Meng Dai. The first draft of the manuscript was written by Yujing Hu, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding information This study was supported by the Foundation of Science and Technology Department of Hebei Province, China (grant number 15277776D).

Date availability All authors made sure that all data and materials as well as the software application supported the published claims and complied with field standards.

Code availability All authors made sure that all the custom codes supported the published claims and complied with field standards.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethics approval All procedures performed in the study and involving human participants were carried out in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study was approved by the Institutional Review Board of the Fourth Hospital of Hebei Medical University.

Informed consent This retrospective analysis was approved by the Institutional Review Board of the Fourth Hospital of Hebei Medical University, and the requirement of informed consent was waived.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Hsu WH, Yang JCH, Mok TS, et al. Overview of current systemic management of EGFR-mutant NSCLC. Ann Oncol. 2018;29:i3–9.
2. World Health Organization. Global Tuberculosis Report 2019 WHO. 2019; https://www.who.int/tb/publications/global_report/en/; Accessed 17 Oct 2019
3. Wang XL, Shan W. Application of dynamic CT to identify lung cancer, pulmonary tuberculosis, and pulmonary inflammatory pseudotumor. Eur Rev Med Pharmacol Sci. 2017;21:4804–9.
4. Qi LP, Chen KN, Zhou XJ, et al. Conventional MRI to detect the differences between mass-like tuberculosis and lung cancer. J Thorac Dis. 2018;10:5673–84.
5. Lim CG, Shin KM, Lim JS, et al. Predictors of conversion to thoracotomy during video-assisted thoracoscopic surgery lobectomy in lung cancer: additional predictive value of FDG-PET/CT in a tuberculosis endemic region. J Thorac Dis. 2017;9:2427–36.
6. Kang F, Wang S, Tian F, et al. Comparing the diagnostic potential of 68Ga-Alfatide II and 18F-FDG in differentiating between non-small cell lung cancer and tuberculosis. J Nucl Med. 2016;57:672–7.
7. Haas MK, Belknap RW. Diagnostic tests for latent tuberculosis infection. Clin Chest Med. 2019;40:829–37.
8. Tiberi S, du Plessis N, Walzl G, et al. Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies. Lancet Infect Dis. 2018;18:e183–98.
9. Hatt M, Le Rest CC, Tixier F, et al. Radiomics: data are also images. J Nucl Med. 2019;60:38S–44S.
10. Castello A, Rossi S, Toschi L, et al. Hyper-progressive disease in patients with non-small cell lung cancer treated with checkpoint inhibitors: the role of ^{18}F-FDG PET/CT. J Nucl Med. 2019. https://doi.org/10.2967/jnumed.119.237768.
11. Dissaux G, Visvikis D, Da-Ano R, et al. Pre-treatment ^{18}F-FDG PET/CT radiomics predict local recurrence in patients treated with stereotactic radiotherapy for early-stage non-small cell lung cancer: a multicentric study. J Nucl Med. 2019. https://doi.org/10.2967/jnumed.119.228106.
12. Yang X, He J, Wang J, et al. CT-based radiomics signature for differentiating solitary granulomatous nodules from solid lung adenocarcinoma. Lung Cancer. 2018;125:109–14.
13. Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res. 2017;6:86–91.
14. Feng B, Li Y, Guo D, et al. Research on the value of the T cell spot test for tuberculosis for the diagnosis of lung cancer combined with pulmonary tuberculosis. Thorac Cancer. 2018;9:1231–4.
15. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Eur Urol. 2015;67:1142–51.
16. Zwanenburg A, Leger S, Vallières M, Löck S, for the Image Biomarker Standardisation Initiative (IBSI). Image biomarker standardisation initiative-feature definitions. 2019 [Current version V11 2019]. https://arxiv.org/abs/1612.07003. Accessed 17 Dec 2019.
17. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54.
18. Nioche C, Orlhac F, Boughdad S, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res. 2018;78:4786–9.
19. Bashir U, Azad G, Siddique MM, et al. The effects of segmentation algorithms on the measurement of F-FDG PET texture parameters in non-small cell lung cancer. EJNMMI Res. 2017;7:60.
20. She Y, Zhang L, Zhu H, et al. The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules. Eur Radiol. 2018;28:5121–8.
21. Ter Voert E, Muehlematter UJ, Delso G, et al. Quantitative performance and optimal regularization parameter in block sequential regularized expectation maximization reconstructions in clinical 68Ga-PSMA PET/MR. EJNMMI Res. 2018;8:70.
22. Slifstein M, Laruelle M. Effects of statistical noise on graphic analysis of PET neuroreceptor studies. J Nucl Med. 2000;41:2083–8.
23. Wagner F, Hakami YA, Warnock G, et al. Comparison of contrast-enhanced CT and [F]FDG PET/CT analysis using kurtosis and skewness in patients with primary colorectal cancer. Mol Imaging Biol. 2017;19:795–803.
24. Chandarana H, Rosenkrantz AB, Mussi TC, et al. Histogram analysis of whole-lesion enhancement in differentiating clear cell from papillary subtype of renal cell cancer. Radiology. 2012;265:790–8.
25. Pantic I, Dacic S, Brkic P, et al. Application of fractal and grey level co-occurrence matrix analysis in evaluation of brain corpus callosum and cingulum architecture. Microsc Microanal. 2014;20:1373–81.
26. Cosma CL, Sherman DR, Ramakrishnan L. The secret lives of the pathogenic mycobacteria. Annu Rev Microbiol. 2003;57:641–76.
27. Lee JW, Kim BS, Lee DS, et al. ^{18}F-FDG PET/CT in mediastinal lymph node staging of non-small-cell lung cancer in a tuberculosis-endemic country: consideration of lymph node calcification and distribution pattern to improve specificity. Eur J Nucl Med Mol Imaging. 2009;36:1794–802.
28. Zhang J, Zhao X, Zhao Y, et al. Value of pre-therapy ^{18}F-FDG PET/CT radiomics in predicting EGFR mutation status in patients

with non-small cell lung cancer. Eur J Nucl Med Mol Imaging. 2020;47:1137–46.

29. Wu J, Bai J, Wang W, et al. ATBdiscrimination: an in silico tool for identification of active tuberculosis disease based on routine blood test and T-SPOT.TB detection results. J Chem Inf Model. 2019;59: 4561–8.

30. Kang W, Wu M, Yang K, et al. Factors associated with negative T-SPOT.TB results among smear-negative tuberculosis patients in China. Sci Rep. 2018;8:4236.