European Journal of Radiology 112 (2019) 161–168

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Development and validation of a radiomics nomogram for identifying invasiveness of pulmonary adenocarcinomas appearing as subcentimeter ground-glass opacity nodules

Wei Zhao[a,b,1], Ya'nan Xu[d,1], Zhiming Yang[a], Yingli Sun[a], Cheng Li[a], Liang Jin[a], Pan Gao[a], Wenjie He[a], Peijun Wang[e], Hongli Shi[d,**], Yanqing Hua[a,**], Ming Li[a,b,c,*]

[a] Department of Radiology, Huadong Hospital Affiliated to Fudan University, Shanghai 200040, China
[b] Diagnosis and Treatment Center of Small Lung Nodules of Huadong Hospital, China
[c] Institute of Functional and Molecular Medical Imaging, Fudan University, China
[d] School of Biomedical Engineering, Capital Medical University, Beijing 100069, China
[e] Department of Radiology, Tongji Hospital, School of Medicine, Tongji University, Shanghai 200065, China

ARTICLE INFO

Keywords:
Radiomics
Nomogram
Subcentimeter
Pulmonary ground-glass nodules

ABSTRACT

The aim of the present study was to develop and validate a radiomics-based nomogram for differentiation of pre-invasive lesions from invasive lesions that appearing as ground-glass opacity nodules (GGNs) ≤10 mm (subcentimeter) in diameter at CT. A total of 542 consecutive patients with 626 pathologically confirmed pulmonary subcentimeter GGNs were retrospectively studied from October 2011 to September 2017. All the GGNs were divided into a training set (n = 334) and a validation set (n = 292). Researchers extracted 475 radiomics features from the plain CT images; a radiomics signature was constructed with the least absolute shrinkage and selection operator (LASSO) based on multivariable regression in the training set. Based on the multivariable logistic regression model, a radiomics nomogram was developed in the training set. The performance of the nomogram was evaluated with respect to its calibration, discrimination, and clinical-utility and this was assessed in the validation set. The constructed radiomics signature, which consisted of 15 radiomics features, was significantly associated with the invasiveness of subcentimeter GGNs ($P < 0.0001$ for both training set and validation set). To build the nomogram model, radiomics signature and mean CT value were used. The nomogram model demonstrated good discrimination and calibration in both training set (C-index, 0.716 [95% CI, 0.632 to 0.801]) and validation set (C-index, 0.707 [95% CI, 0.625 to 0.788]). Decision curve analysis (DCA) indicated that radiomics-based nomogram was clinically useful. A radiomics-based nomogram that incorporates both radiomics signature and mean CT value is constructed in the study, which can be conveniently used to facilitate the preoperative individualized prediction of the invasiveness in patients with subcentimeter GGNs.

1. Introduction

With the popularization of low-dose computed tomography (LDCT) screening for lung cancer, a frequently reported incidence of small-sized carcinomas appearing as pulmonary ground-glass nodules (GGNs), are increasingly encountered in routine clinical practice [1]. Stage I lung cancers account for 60% to 70% of detected lung cancers in screening programs, of which 56% are less than 1 cm in diameter (subcentimeter lesions) [2]. However, controversy and uncertainty regarding the management of subcentimeter lesions still remain. Given that pre-invasive lesions [atypical adenomatous hyperplasias (AAHs) and adenocarcinomas in situ (AIS)] usually show stable or very slow growth on follow-up CT and that a favorable prognosis is indicated after safely treating with limited resection [3,4], accurately differentiating these

Abbreviations: GGN, ground-glass nodule; LDCT, low-dose computed tomography; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinomas in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive pulmonary adenocarcinoma; CEA, carcinoembryonic antigen; SCC, squamous cell carcinoma antigen; VOI, volume of interest; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; DCA, decision curve analysis; VIF, variance inflation factor

* Corresponding author at: Department of Radiology, Huadong Hospital Affiliated to Fudan University, Shanghai 200040, China.
** Co-Corresponding authors.
E-mail addresses: shl@ccmu.edu.cn (H. Shi), huayq007@163.com (Y. Hua), minli77@163.com (M. Li).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ejrad.2019.01.021
Received 16 August 2018; Received in revised form 19 January 2019; Accepted 21 January 2019

preinvasive lesions from minimally invasive adenocarcinomas (MIAs) and invasive pulmonary adenocarcinomas (IPAs) is a priority for thoracic radiologists and surgeons. Although the previous studies reported that a 10 mm diameter is used as a cutoff value to distinguish preinvasive and invasive lesions [5,6], some subcentimeter GGNs, including pure GGNs, have finally been pathologically confirmed as MIA or IPA [7,8]. Therefore, a definitive diagnosis of subcentimeter GGNs may not only help the surgeons to make personalized clinical decisions and thus avoid inappropriate overtreatment or conservative treatment but can also reduce the patient's mental press and financial burden.

Radiographic features such as air bronchogram, the solid portion presenting in GGNs or a larger lesion size on LDCT can indicate the malignancy of GGNs [9,10]. Nevertheless, the essential accurate measurement of visual radiographic features based on CT images between observers and considerable overlapping radiographic features among four subtypes is observed as an inevitable discrepancy, remaining as a dilemma for both radiologists and thoracic surgeons. Moreover, the aforementioned radiographic features may be absent in subcentimeter GGNs due to the early stage of the tumor, which could limit its diagnostic performance. Numerous researchers have realized that morphological characteristics presented on CT images are just the tip of the iceberg, while a nearly limitless mineable high-dimensional and valuable data remain behind the images, which are unfortunately invisible to us [11–13]. In this context, "Radiomics", an emerging data-driven strategy firstly proposed by Philippe [14] in 2012, can extract either a set of predefined engineered features that describe radiographic aspects of shape, intensity, and texture, or alternatively features, and incorporates several important disciplines, including radiology (eg, imaging interpretation), computer vision (eg, quantitative feature extraction), and machine learning (eg, classifier evaluation) [15]. In clinical practice, it could aid in disease detection, diagnosis, evaluation of prognosis, and prediction of treatment response and facilitate improved clinical decision-making and personalized medicine [11–13]. Although CT texture assessments have been applied and demonstrated to be useful for preoperative differentiation of the invasiveness of lesions in patients with GGNs [16,17], an optimal approach that combines multiple imaging biomarkers as a predictive signature is yet to be developed. To the best of our knowledge, there are no radiomics-based studies for the preoperative identifying invasiveness of subcentimeter GGNs.

Hence, the aim of this study was to construct a preoperative nomogram, incorporating radiomic features, clinical characteristics, and radiographic features, to differentiate invasive lesions from preinvasive ones in patients with subcentimeter GGNs.

2. Materials and methods

This study was approved by the institutional review board of our institution (Grant NO.2017K062), which waived the requirement for patients' informed consent of this retrospective study. The workflow was presented in Fig. 1.

2.1. Study population

One author reviewed the pathology lung cancer surgical database and the radiology information systems in our institution from October 2011 to September 2017. A total of 542 patients (mean age, 53.9 years ± 12.2 [standard deviation]; range, 16–82 years) with 626 subcentimeter nodules were enrolled in the study. The inclusion criteria were the following: (1) the presence of chest CT scan with thin-slice thickness (1–1.5 mm) before surgical treatment; (2) lesions appearing as ground-glass opacity nodules (including part-solid and no solid nodules) on lung window images (level, −600 Hounsfield unit [HU]; width, 1500 HU) with a diameter ≤10 mm; (3) no treatment prior surgery; (4) no distant metastasis and tumor history.

Researchers divided the patients into two training and validation sets with a ratio over 1:1 according to the surgical time of included patients: 293 patients (85 males and 208 females; mean age, 54.5 years ± 11.8 [standard deviation]; rage, 22–78 years) with 334 lesions treated between October 2011 and December 2015 constituted the training set, whereas 249 patients (86 males and 163 females; mean age, 53.1 years ± 12.7 [standard deviation]; range, 16–82 years) with 292 lesions treated between January 2016 and September 2017 constituted the validation set. The unbalanced distribution of gender in our study was determined by a unique characteristic of female predominance in this case [4,18]. The nodules were surgically removed with segmentectomy or lobectomy via video-assisted thoracic surgery or thoracotomy surgery in our institution depending on size, growth of GGNs, the probability of malignance and patients' decision. Clinical data, such as age, sex, smoking history, carcinoembryonic antigen (CEA) and squamous cell carcinoma antigen (SCC) were obtained by reviewing the medical records. Clinical and demographic features of our study population were summarized in Table 1.

2.2. CT image acquisition, retrieval procedure and radiomics feature extraction methodology

All CT examinations were performed with one of the four scanners: GE Discovery CT750 HD, 64-slice LightSpeed VCT (GE Medical Systems); Somatom Definition flash, Somatom Sensation-16 (Siemens Medical Solutions). The details of the scanning parameters and the number of nodules performed by each scanner were showed in Table 2. After non-contrast enhanced CT scanning, 252 patients received 80–100 mL of IV contrast medium (350 mg I/ mL; Optiray, Mallinckrodt), which was injected at a rate of 3–4 mL/s with the use of a power injector via an 18- or 20-gauge cannula in an antecubital vein. The contrast-enhanced CT scan was acquired 35 and 60 s after the administration of contrast medium. In all patients, CT images were acquired in the supine position at full inspiration. In this study, only plain CT images were utilized. When more than one CT examination was available for analysis, the last CT examination before surgery was selected. The mean interval between CT examination and surgery was 13 days (range, 1–96 days; median, 7 days). Plain CT images (thickness: 1–1.5 mm) were retrieved from the picture archiving and communication system (PACS) in the institution for image feature extraction.

The process of extracting radiomic features was described in the **Supplementary data**. Inter- and intra-observer reproducibility of radiomics feature extraction was initially analyzed with 60 randomly chosen images for volume of interest (VOI)-based texture feature generation in a blind fashion by two radiologists with 6 (reader 1) and 15 years (reader 2) of experience in chest CT interpretation. Reader 1 then repeated the same procedure one month later. The intra- and inter-observer agreement of feature extraction was assessed by inter- and intra-class correlation coefficients (ICCs). An ICC greater than 0.75 indicated good agreement of the feature extraction. The remaining image delineation was performed by reader 1.

2.3. The assessment of radiographic characteristics, surgical resection and histological evaluation

The assessment of radiographic characteristics, surgical resection and histological evaluation were described in the **Supplementary data**. All histological preparations and analyses were performed by two senior pathologists. All GGNs were divided into two groups: a preinvasive lesion group (AAH, AIS) and an invasive group (MIA, IPA).

2.4. Feature selection, radiomics signature building and diagnostic validation of radiomics signature

In the study, the least absolute shrinkage and selection operator (LASSO) method, which is suitable for the regression of high-dimensional data [19], was used to select the most useful predictive features

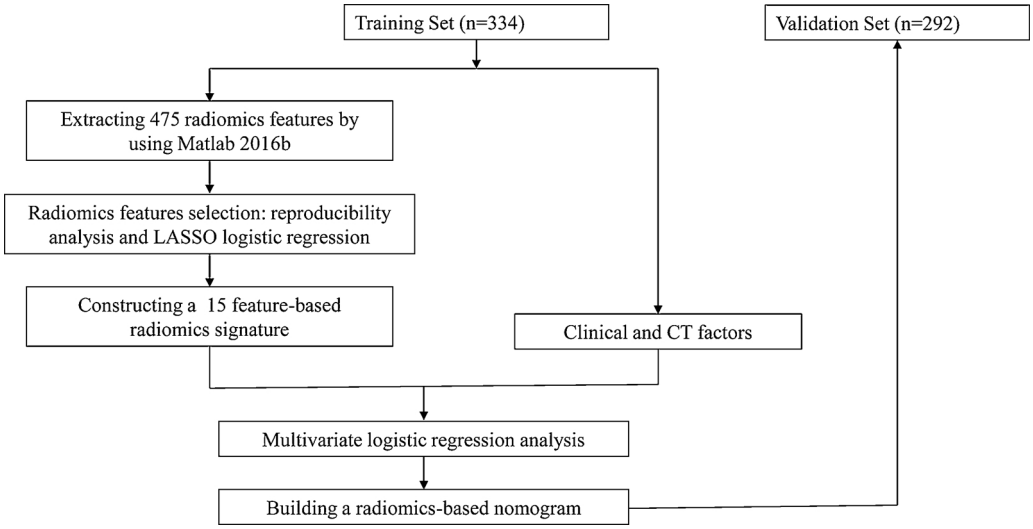

Fig. 1. The workflow of the study.

from the training set. A radiomics signature was calculated for each patient via a linear combination of selected features that were weighted by their respective coefficients. The potential association of the radiomics signature with invasiveness of GGNs was first assessed in the training set and then validated in the validation set using a Mann-Whitney U test.

2.5. Construction of the radiomics nomogram

A multivariate logistic regression model was performed to identify the independent factors among radiomics signature, clinical variables and radiographic features for differentiating preinvasive lesions from invasive ones in the training set. Before performing the multivariate logistic regression analysis, the collinearity diagnosis was performed using the variance inflation factor (VIF). A radiomics nomogram was then constructed on the basis of the multivariate logistic regression model.

2.6. Performance of the radiomics nomogram in the training set and validation set

A radiomics signature was calculated for each patient using the formula constructed in the training set. Independent validation was performed using the validation set. Internal validation was performed using the data set of the second measurement by reader 1, as well as the measurement by reader 2 of the 60 patients. The calibration of the nomogram was measured by a calibration curve. The Hosmer-Lemeshow test was performed to assess the goodness-of-fit of the nomogram.

2.7. Clinical utility of the radiomics nomogram

To estimate the clinical utility of the nomogram, decision curve analysis (DCA) was performed in the combined training and validation set by calculating the net benefits for a range of threshold probabilities.

2.8. Statistical analysis

Statistical analysis was performed using R software (version 3.4.3; http://www.Rproject.org). Lasso binary logistic regression was done using the "glmnet" package, multivariate binary logistic regression, nomograms, and calibration plots were done with the "rms" package. The C-index calculation was performed using the "Hmisc" package. Internal validation was performed using the "rms" package. DCA was performed using the "rmda" package. A two-sided P value < 0.05 was considered statistically significant.

3. Results

3.1. Baseline characteristics

Of the 626 nodules, 202 (32.3%) nodules were pathologically confirmed as preinvasive lesions (AAH, n = 39; AIS, n = 163), whereas 424 (67.7%) were confirmed as invasive lesions (MIA, n = 312; IPA, n = 112). There were no significant differences between the two sets in the prevalence of invasive nodules [70.7% (236/334) vs. 64.4% (188/292), $P = 0.094$, chi-square test].

Except for smoking, there were no significant differences in the other clinical and CT characteristics between the two sets, justifying their use as training and validation data sets. However, smoking was not identified as a risk factor in the univariate and multivariate analysis. The patient basal characteristics in two sets are shown in Table 1 and the **Supplementary data**.

3.2. Feature selection, radiomics signature building and diagnostic validation of radiomics signature

One hundred and twenty-seven robust radiomic features were selected after assessing their inter- and intra-observer reproducibility (see **Supplementary data**); fifteen potential predictors with nonzero coefficients were finally selected from the 127 robust radiomic features based on 334 nodules in the training set (Fig. 2**A** and **B**) using a LASSO-logistic regression model. The 15 selected features were consequently conducted into a radiomics signature by using the calculation formula:

Radiomics signature = 2.249224 + 0.104342*mean_50_0 - 0.00091*homogeneity 0_45_0 + 6.30E-08*correlation 90_135_0 - 0.0000525*contrast 90_135_1.0 + 0.000495*SD_25_1.5 - 0.00536*contrast 0_45_1.5 - 0.02875*contrast 0_135_1.5 2.93E-05*contrast 90_45_1.5 - 0.00017*contrast 90_135_1.5 + 3.57E-05*SRE_135_1.5 - 0.000016*contrast 0_135_2.0 + 0.000354 * entropy 45_45_2.0 + 2.94E-05*contrast 90_90_2.0–1.639969E-06*contrast 90_135_2.0 + 0.000135*entropy 45_45_2.5

The details about the formula were presented in the **Supplementary data**. Distributions of the radiomics signature and invasiveness of subcentimeter GGNs in two sets are shown in the **Supplementary data**.

The invasive group had a significantly higher radiomics signature than the preinvasive groups in the training set (1.191 ± 0.729 vs.

Table 1
Characteristics of Patients in the Training and Validation Set.

Characteristics	Training Set (n = 334)		*P*	Validation Set (n = 292)		*P*
	Preinvasive Group (n = 98)	Invasive Group (n = 236)		Preinvasive Group (n = 104)	Invasive Group (n = 188)	
Demographic and Clinical Characteristics						
Gender			0.849			0.222
Male	28(28.6)	65(27.5)		40(38.5)	59(31.4)	
Female	70(71.4)	171(72.5)		64(61.5)	129(68.6)	
Age (year)	53.65 ± 12.08	55.00 ± 11.41	0.337	52.49 ± 12.21	54.14 ± 13.35	0.299
Location			0.252			0.324
Right upper lobe	47(48.0)	84(35.6)		40(38.5)	59(31.4)	
Right middle lobe	9(9.2)	19(8.1)		7(6.7)	22(11.7)	
Right lower lobe	14(14.2)	40(16.9)		20(19.2)	30(16.0)	
Left upper lobe	19(19.4)	65(27.5)		25(24.0)	49(26.0)	
Left lower lobe	9(9.2)	28(11.9)		12(11.6)	28(14.9)	
SCC (ng/mL)	0.811 ± 0.424	0.757 ± 0.411	0.283	0.769 ± 0.430	0.815 ± 0.377	0.342
CEA (ng/mL)	1.809 ± 1.100	2.014 ± 1.369	0.352	2.308 ± 3.268	2.277 ± 3.304	0.938
Smoke			0.656			0.06
Never Smoker	90(91.8)	220(93.2)		104(100)	182(96.8)	
Current or Former Smoker	8(8.2)	16(6.8)		0(0)	6(3.2)	
Radiographic Characteristics						
Size (cm)	0.667 ± 0.204	0.798 ± 0.225	**< 0.001**	0.623 ± 0.179	0.799 ± 0.390	**0.003**
Type			**< 0.001**			**0.015**
Pure GGN	71(72.4)	111(47.0)		69(66.3)	97(51.6)	
Part-solid GGN	27(27.6)	125(53.0)		35(33.7)	91(48.4)	
Margin			**0.008**			**0.035**
Easily Differentiated	72(73.5)	13,758.1)		68(65.4)	99(52.7)	
Uneasily Differentiated	26(26.5)	99(41.9)		36(34.6)	89(47.3)	
Shape			**0.002**			**< 0.001**
Round	35(35.7)	49(20.8)		25(24.0)	31(16.5)	
Oval	40(40.8)	88(37.3)		54(52.0)	54(28.7)	
Irregular	23(23.5)	99(41.9)		25(24.0)	103(54.8)	
Pleural Retraction			0.151			0.815
Present	20(20.4)	66(28.0)		22(21.2)	42(22.3)	
Absent	78(79.6)	170(72.0)		82(78.2)	146(77.7)	
Bubble luceccy			0.516			0.707
Present	24(38.5)	66(28.0)		27(38.5)	46(24.5)	
Absent	74(38.5)	170(72.0)		77(38.5)	142(75.5)	
Vascular change			**< 0.001**			**< 0.001**
Present	13(13.3)	84(35.6)		16(15.4)	86(45.7)	
Absent	85(86.7)	152(64.4)		88(84.6)	102(54.3)	
Bronchiole change			0.566			0.305
Present	5(5.1)	16(6.8)		6(5.8)	24(12.8)	
Absent	93(94.9)	220(93.2)		98(94.2)	164(87.2)	
Mean CT value (HU)	−576.051 ± 116.491	−461.729 ± 153.517	**< 0.001**	−565.490 ± 129.685	−512.335 ± 141.048	**0.002**
Radiomics Signature	0.555 ± 0.729	1.191 ± 0.822	**< 0.001**	0.541 ± 0.678	0.988 ± 0.728	**< 0.001**

NOTE: Values are presented as no. (%) or mean ± SD. P value is derived from the univariable association analyses between each of the basal variables and the invasive extent. Abbreviations: CEA, carcinoembryonic antigen; CT, computed tomography; SCC, squamous cell carcinoma antigen; GGN, ground-glass nodules. Bold P values < 0.05.

0.555 ± 0.822, $P < 0.0001$). This difference was confirmed in the validation set (0.988 ± 0.728 vs. 0.541 ± 0.678, $P < 0.0001$). The radiomics signature yielded a C-index of 0.751 (95%CI, 0.672 to 0.831) in training set, and 0.713 (95%CI, 0.632 to 0.795) in the validation set.

3.3. Construction of the radiomics nomogram

In univariate analysis, 7 factors were significantly associated with invasive extent (Table 1). There was no multicollinearity between the

Table 2
CT scanning parameters and number of nodules performed by each scanner.

	GE Discovery CT750 HD	LightSpeed VCT	Somatom Definition Flash	Somatom Sensation-16
Parameters				
Tube voltage	120 kVp	120 kVp	120 kVp	120 kVp
Tube current	200mA	200mA	110mAs	110mAs
Pitch	0.984:1	0.984:1	1.0	0.8
Collimation	0.625 mm*64	0.625 mm*64	0.6mm*64	0.75 mm*16
Rotation time	0.5 s/rot	0.5 s/rot	0.33 s/rot	0.35 s/rot
SFOV	50cm	50cm	50cm	50cm
Slice thickness of reconstruction	1.25 mm	1.25 mm	1mm	1/1.5 mm
Slice interval of reconstruction	1.25 mm	1.25 mm	1mm	1/1.5 mm
Reconstruction algorithm	STND	STND	Medium sharp	Medium sharp
Number of nodules	137	192	145	152

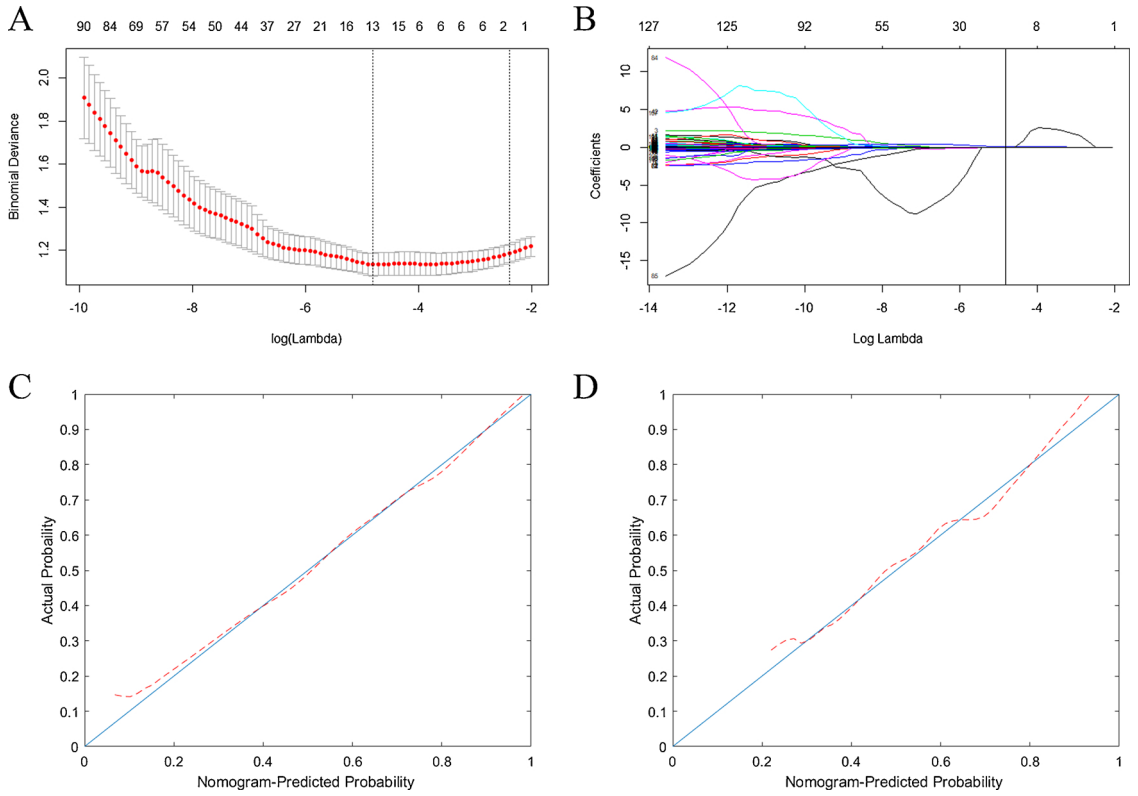

Fig. 2. Texture feature selection using least absolute shrinkage and selection operator (LASSO) binary logistic regression model. **(A)** Selection of the tuning parameter (λ) in the LASSO model via 10-fold cross-validation based on minimum criteria. Binomial deviances from the LASSO regression cross-validation procedure were plotted as a function of log (λ). The y-axis indicated binomial deviances. The lower x-axis indicated the log (λ). Numbers along the upper x-axis represented the average number of predictors. Red dots indicated average deviance values for each model with a given λ, and vertical bars through the red dots showed the upper and lower values of the deviances. The vertical black dotted lines defined the optimal values of λ, where the model provided its best fit to the data. The optimal λ value of 0.68 with log (λ) = - 4.82 was selected. **(B)** LASSO coefficient profiles (y-axis) of the 127 texture features. The upper and lower x-axis indicated the same meaning as in Fig. 2A. The black vertical line was drawn at the value selected using 10-fold cross-validation in Fig. 2A. The 15 resulting features with nonzero coefficients were indicated in the plot. **(C)** and **(D)**: Calibration curves of the radiomics nomogram in the training set and the validation set. **(C)** Calibration curve of the radiomics nomogram in the training set. **(D)** Calibration curve of the radiomics nomogram in the validation set.

Table 3
Risk Characteristics for Predicting the Invasiveness of Subcentimeter GGNs in the Training Set.

	Univariate Logistic Regression		Multivariate Logistic Regression	
Characteristic	OR (95% CI)	*P*	OR (95% CI)	*P*
Size	0.241(-0.236-0.719)	< 0.0001	NA	NA
Type	0.156(-0.414-0.664)	< 0.0001	NA	NA
Margin	−0.148(-0.96-0.664)	0.008	NA	NA
Shape	0.083(-0.894-1.06	0.002	NA	NA
Vascular change	0.379(-0.404-1.163)	< 0.0001	NA	NA
Mean CT value (HU)	1.093(0.574-1.612)	< 0.0001	1.160(0.697-1.622)	< 0.0001
Radiomics Signature	0.818(0.381-1.255)	< 0.0001	0.951(0.582-1.319)	< 0.0001

NOTE: Abbreviations: GGN, ground-glass nodules; OR, odds ratio; CI, confidence interval; CT, computed tomography; NA, not available.

potential factors identified by univariate analysis and radiomics signature. According to the multivariate analysis, mean CT value and radiomics signature were statistically significant independent differentiators of preinvasive lesions from invasive ones (Table 3) and they were incorporated to develop the nomogram (Fig. 3).

3.4. Performance of the radiomics nomogram in the training set and validation set

The calibration curve of the radiomics nomogram for the probability of invasive of subcentimeter GGNs demonstrated good agreement between prediction and observation in the two sets (Fig. 2**C** and **D**). The Hosmer-Lemeshow test yielded a nonsignificant statistic in the training and validation set ($P = 0.908$, $P = 0.191$, respectively), which suggested that there was no departure from a perfect fit. The C-index for the prediction nomogram was 0.716 (95%CI, 0.632 to 0.801) in the training set and 0.707 (95% CI, 0.625 to 0.788) in the validation set. With regard to the internal validation, the nomogram model yielded a C-index of 0.721 (95%CI, 0.534 to 0.907) for reader 1 and 0.713 (95%CI, 0.527 to 0.899) for reader 2.

3.5. Clinical use

The DCA for the radiomics nomogram was presented in Fig. 4. The decision curve showed that if the threshold probability of a patient or doctor was > 10%, using the radiomics nomogram to predict invasiveness of subcentimeter GGNs would add more benefit with respect to the treat-all-patients scenario or the treat-none scenario.

4. Discussion

In this study, a diagnostic radiomics nomogram, incorporating radiomics signature and mean CT value, was developed and validated

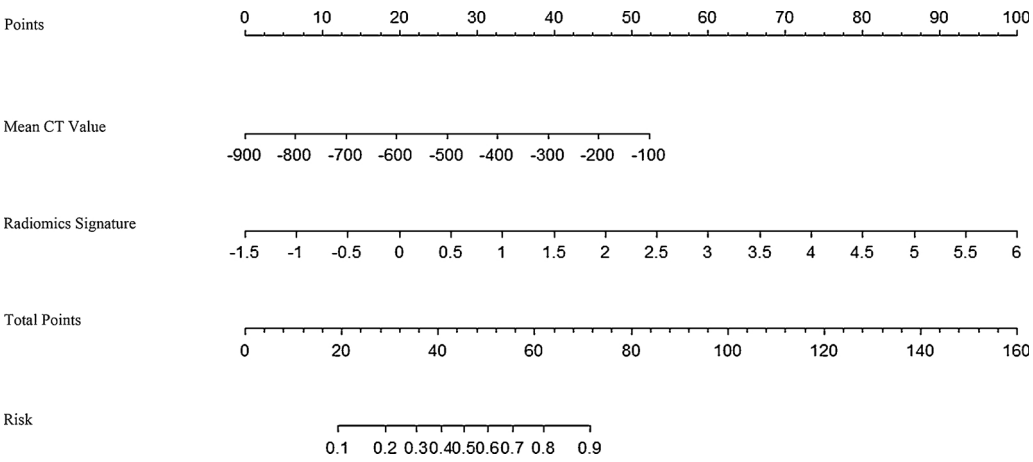

Fig. 3. Developed radiomics nomogram. The radiomics based nomogram was developed in the training set, with the radiomics signature and mean CT value incorporated.

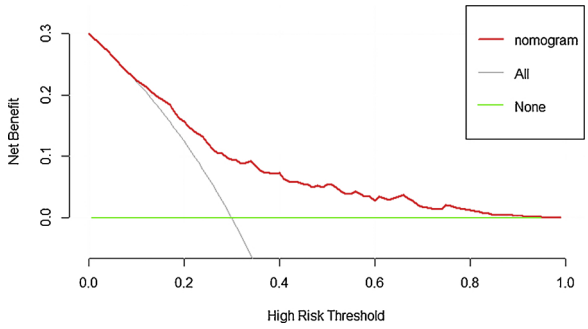

Fig. 4. DCA for the radiomics-based nomogram. The y-axis represented the net benefit. The red line represented the radiomics-based nomogram. The grey line represented the hypothesis that all patients with subcentimeter GGNs were invasive lesions. The green line represented the hypothesis that all patients with subcentimeter GGNs were preinvasive lesions. The x-axis represented the threshold probability. The threshold probability was where the expected benefit of treatment was equal to the expected benefit of avoiding treatment. The decision curve demonstrated that if the threshold probability of a patient or doctor is > 10%, using the radiomics-based nomogram in the current study to predict the invasiveness of subcentimeter GGNs would add more benefit than the treat-all-patients scenario or the treat-none scenario.

for the preoperative individualized prediction of the invasiveness of subcentimeter GGNs. The identified radiomics signature successfully classified the subcentimeter GGNs of patients into preinvasive lesions and invasive ones. Incorporating the radiomics signature and mean CT value into an easy-to-use nomogram, facilitated the preoperative individualized prediction of invasiveness of subcentimeter GGNs. In addition, the constructed nomogram demonstrated adequate discrimination in the training set (C-index, 0.716) and the validation set (C-index, 0.707).

For the construction of the radiomics signature, 15 reliable potential predictors were carefully selected among 475 candidate radiomic features. According to the radiomics hypothesis, radiomics can be used to assess the overall tumor spatial heterogeneity, tumor microenvironment, as well as reflect tumor gene patterns, which could aid in conducting personalized clinical scenarios [13]. Yet, interpreting the association between radiomic features and the complex biological processes remains an intractable challenge. Radiomics-based studies that incorporate multiple individual imaging biomarkers into a radiomics signature were similar to the construction of multi-factor panels [20–23], making multi-marker analyses less complicated to use in clinical work. In this study, researchers identified 15 potential predictors that were combined into a radiomics signature and were used to stratify subcentimeter GGNs into preinvasive lesions and invasive ones.

The clinicopathological characteristics of subcentimeter lung adenocarcinomas has been already well-defined [7,24,25]; however, its radiographic features has been rarely reported. Wu et al. investigated the CT features of subcentimeter pure GGNs (pGGNs) and concluded that vessel changes, unsmooth margin, and clear lung-tumor interface may all indicate the invasiveness of subcentimeter pGGNs [26]; yet, his study included only pGGNs and a small sample size (150 lesions). In clinical context, part-solid GGNs consist of an important proportion of subcentimeter pulmonary adenocarcinomas [7,25]. To comprehensively investigated the characteristics of subcentimeter GGNs, both of pGGNs and part-solid GGNs were included in this study. Researchers found that a higher mean CT value was associated with invasive lesions, which is consistent with previous studies [27,28]. Increased mean CT value within GGN is believed to reflect the increased heterogeneity, caused by the infiltration of invasive tumor cells [29]. Note that several suggested risk characteristics in previous studies, including lesion size, shape, margin, vascular changes [6,9,27,30], failed to have a significant association with the invasiveness of subcentimeter GGNs in multivariate analysis. CT morphological features like irregular shape, ill-defined margin, and vascular changes were suggested to have the ability to discriminate invasive lesions and preinvasive lesions appearing as GGNs [10,27,30,31]. However, these conclusions remain controversial [27,32], especially in studies involving pGGNs [6,10,26,33]. Several previous studies reported a pGGN lesion with an oval appearance suggesting a malignant nature and a polygonal shape, therefore, indicating a benign nature [34]. Lobulated or spiculated margin can be seen in both organizing pneumonia/fibrosis and tumor, histologically caused by irregular interstitial fibrosis or an infiltrative tumor growth [32]. Therefore, the aforementioned morphological CT features may be not reliable for identifying the invasiveness of GGNs, especially for subcentimeter GGNs. In the early stage of tumor (i.e., subcentimeter cancer), those morphological features that indicate malignancy may be absent. With the tumor progression, vasculature remodeling or neoangiogenesis subsequently occur to facilitate the sustainable growth. Therefore, the hypothesis that vascular changes like distortion and concentration were more often seen in more aggressive lesions was proposed and proved in a few studies [26,30]. Nonetheless, the probability of this feature was lower in our data set (31.8%) than previous studies (35.3% 40.8%). Data selection bias may be a plausible interpretation for the discrepancy between the results and others', as only subcentimeter GGNs were investigated in the current study. Lesion size was frequently suggested as an independent risk factor to indicate malignancy [6,8] and such conclusion has been deeply incorporated into the Fleischner Society recommendations for management of lung nodules [35]. However, with regard to subcentimeter GGNs, researchers concluded that the predictive performance of size may

compromise, as the overlap of four subtype lesions in terms of size may be more pronounced. Wu and his team confirmed the limited diagnostic performance of size for discriminating invasive lesions from preinvasive ones in subcentimeter pGGNs [26].

Nomogram, a statistic model which could generate an individual numerical probability of a clinical event by integrating multiple variables, fulfills the drive towards personalized medicine [36]. A recent multi-institutional study successfully constructed a nomogram to preoperatively predict the invasiveness of pGGNs with an excellent C-Index of 0.94 by incorporating multiple clinical factors and CT features [27]. Therefore, an easy-to-use radiomics-based nomogram was also constructed based on the multivariate analysis in the study. The nomogram, combining the radiomics signature and mean CT value, demonstrated good calibration and discrimination in the training and validation sets. Like previous radiomics studies, the application of radiomic-based nomogram requires computational power, which may be a little complicated for clinicians. However, when the experimental results really come to clinical use, the radiomics signature can be automatically and conveniently calculated by software. To assess the clinical utility of our nomogram, a clinical decision curve analysis was performed instead of the multi-institutional prospective validation. Because prospective validation of a nomogram prior to use is largely impractical and the progress may be subject to the heterogeneous CT acquisition parameters and tedious clinical data collection. The decision curve demonstrated that if the threshold probability of a patient or doctor is > 10%, the radiomics nomogram adds more benefit than the treat-all-patients scenario or the treat-none scenario. This conclusion indicated that this tool could be useful for clinical decision-making and might facilitate the individualized precision medical treatment.

However, although the conclusions were encouraging, several limitations should be discussed. First, this was a single-center retrospective study that only included the pathologically confirmed subcentimeter GGNs, which means that it is subject to potential data selection bias. Second, researchers included several different CT scanners' images in the current study, potential confounding variability caused by heterogeneous parameters [37]. However, to minimize these variabilities, all images included in the current study were thin-slice CT images (1–1.5 mm), which were suitable for radiomic features analysis [38]. Moreover, imaging normalization and reproducibility study were performed in preprocessing phase (see **Supplementary data**). Third, the constructed nomogram lacks of external validation. However, the decision curve analysis, which is an appropriate alternative method to assess the clinical utility, was performed in the current study and justified that the constructed radiomics signature and radiomics-based nomogram hold great potential for clinical application in personalized risk stratification for invasiveness in patients with subcentimeter GGNs. Therefore, a prospective multicenter study with standardized CT scanning parameters is needed to evaluate the generalizability and to further verify the reported findings.

5. Conclusions

A radiomics-based nomogram incorporating both radiomics signature and mean CT value was constructed in our study, which was conveniently used to facilitate the preoperative individualized prediction of invasiveness in patients with subcentimeter GGNs.

Declaration of interest

None.

Acknowledgments

This study was supported by the Research Program of Shanghai Hospital Development CenterSHDC22015025 (Ming Li), the National Key Research and Development Program of China2017YFC0112800 (Peijun Wang), 2017YFC0112905 (to M. Li) and the Medical Imaging Key Program of Wise Information Technology of 120, Health Commission of Shanghai2018ZHYL0103 (Ming Li).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2019.01.021.

References

[1] National Lung Screening Trial Research T, D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, et al., Reduced lung-cancer mortality with low-dose computed tomographic screening, N. Engl. J. Med. 365 (5) (2011) 395–409.
[2] R. Rami-Porta, V. Bolejack, J. Crowley, D. Ball, J. Kim, G. Lyons, et al., The IASLC lung Cancer Staging project: proposals for the revisions of the t descriptors in the forthcoming eighth edition of the TNM classification for lung Cancer, J. Thorac. Oncol. 10 (7) (2015) 990–1003.
[3] P.E. Van Schil, H. Asamura, V.W. Rusch, T. Mitsudomi, M. Tsuboi, E. Brambilla, et al., Surgical implications of the new IASLC/ATS/ERS adenocarcinoma classification, Eur. Respir. J. 39 (2) (2012) 478–486.
[4] W.D. Travis, E. Brambilla, M. Noguchi, A.G. Nicholson, K.R. Geisinger, Y. Yatabe, et al., International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma, J. Thorac. Oncol. 6 (2) (2011) 244–285.
[5] M. Nakata, S. Sawada, H. Saeki, S. Takashima, H. Mogami, N. Teramoto, et al., Prospective study of thoracoscopic limited resection for ground-glass opacity selected by computed tomography, Ann. Thorac. Surg. 75 (5) (2003) 1601–1605 discussion 05-6.
[6] S.M. Lee, C.M. Park, J.M. Goo, H.J. Lee, J.Y. Wi, C.H. Kang, Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features, Radiology 268 (1) (2013) 265–273.
[7] H. Sakurai, K. Nakagawa, S. Watanabe, H. Asamura, Clinicopathologic features of resected subcentimeter lung cancer, Ann. Thorac. Surg. 99 (5) (2015) 1731–1738.
[8] H. Mao, K. Labh, F. Han, S. Jiang, Y. Yang, X. Sun, Diagnosis of the invasiveness of lung adenocarcinoma manifesting as ground glass opacities on high-resolution computed tomography, Thorac. Cancer 7 (1) (2016) 129–135.
[9] M. Yanagawa, T. Johkoh, M. Noguchi, E. Morii, Y. Shintani, M. Okumura, et al., Radiological prediction of tumor invasiveness of lung adenocarcinoma on thin-section CT, Medicine 96 (11) (2017) e6331.
[10] H.J. Lee, J.M. Goo, C.H. Lee, C.M. Park, K.G. Kim, E.A. Park, et al., Predictive CT findings of malignancy in ground-glass nodules on thin-section chest CT: the effects on radiologist performance, Eur. Radiol. 19 (3) (2009) 552–560.
[11] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2) (2016) 563–577.
[12] G. Lee, H.Y. Lee, H. Park, M.L. Schiebler, E.J.R. van Beek, Y. Ohno, et al., Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art, Eur. J. Radiol. 86 (2017) 297–307.
[13] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[14] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, et al., Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (4) (2012) 441–446.
[15] M. Zhou, J. Scott, B. Chaudhury, L. Hall, D. Goldgof, K.W. Yeom, et al., Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches, AJNR Am. J. Neuroradiol. 39 (2) (2018) 208–216.
[16] H.D. Chae, C.M. Park, S.J. Park, S.M. Lee, K.G. Kim, J.M. Goo, Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas, Radiology 273 (1) (2014) 285–293.
[17] I.P. Hwang, C.M. Park, S.J. Park, S.M. Lee, H.P. McAdams, Y.K. Jeon, et al., Persistent pure ground-glass nodules larger than 5 mm: differentiation of invasive pulmonary adenocarcinomas from preinvasive lesions or minimally invasive adenocarcinomas using texture analysis, Invest. Radiol. 50 (11) (2015) 798–804.
[18] A. Jemal, K.D. Miller, J. Ma, R.L. Siegel, S.A. Fedewa, F. Islami, et al., Higher lung Cancer incidence in young women than young men in the United States, N. Engl. J. Med. 378 (21) (2018) 1999–2009.
[19] W. Sauerbrei, P. Royston, H. Binder, Selection of important variables and determination of functional form for continuous predictors in multivariable model building, Stat. Med. 26 (30) (2007) 5512–5528.
[20] S. Wu, J. Zheng, Y. Li, H. Yu, S. Shi, W. Xie, et al., A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer, Clin. Cancer Res. 23 (22) (2017) 6904–6911.
[21] L. Zhang, B. Chen, X. Liu, J. Song, M. Fang, C. Hu, et al., Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell lung Cancer, Transl. Oncol. 11 (1) (2018) 94–101.
[22] Y.Q. Huang, C.H. Liang, L. He, J. Tian, C.S. Liang, X. Chen, et al., Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal Cancer, J. Clin. Oncol. 34 (18) (2016) 2157–2164.
[23] 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 (3) (2016) 947–957.
[24] F. Kato, M. Hamasaki, Y. Miyake, A. Iwasaki, H. Iwasaki, K. Nabeshima, Clinicopathological characteristics of subcentimeter adenocarcinomas of the lung, Lung cancer 77 (3) (2012) 495–500.
[25] H. Asamura, K. Suzuki, S. Watanabe, Y. Matsuno, A. Maeshima, R. Tsuchiya, A clinicopathological study of resected subcentimeter lung cancers: a favorable prognosis for ground glass opacity lesions, Ann. Thorac. Surg. 76 (4) (2003) 1016–1022.
[26] F. Wu, S.P. Tian, X. Jin, R. Jing, Y.Q. Yang, M. Jin, et al., CT and histopathologic characteristics of lung adenocarcinoma with pure ground-glass nodules 10 mm or less in diameter, Eur. Radiol. 27 (10) (2017) 4037–4043.
[27] Y. She, L. Zhao, C. Dai, Y. Ren, J. Zha, H. Xie, et al., Preoperative nomogram for identifying invasive pulmonary adenocarcinoma in patients with pure ground-glass nodule: a multi-institutional study, Oncotarget 8 (10) (2017) 17229–17238.
[28] Q.J. Zhou, Z.C. Zheng, Y.Q. Zhu, P.J. Lu, J. Huang, J.D. Ye, et al., Tumor invasiveness defined by IASLC/ATS/ERS classification of ground-glass nodules can be predicted by quantitative CT parameters, J. Thorac. Dis. 9 (5) (2017) 1190–1200.
[29] Y.P. Zhang, M.A. Heuvelmans, H. Zhang, M. Oudkerk, G.X. Zhang, X.Q. Xie, Changes in quantitative CT image features of ground-glass nodules in differentiating invasive pulmonary adenocarcinoma from benign and in situ lesions: histopathological comparisons, Clin. Radiol. (2018).
[30] F. Gao, M. Li, X. Ge, X. Zheng, Q. Ren, Y. Chen, et al., Multi-detector spiral CT study of the relationships between pulmonary ground-glass nodules and blood vessels, Eur. Radiol. 23 (12) (2013) 3271–3277.
[31] M.J. Si, X.F. Tao, Cai L.L. Du GY, H.X. Han, X.Z. Liang, et al., Thin-section computed tomography-histopathologic comparisons of pulmonary focal interstitial fibrosis, atypical adenomatous hyperplasia, adenocarcinoma in situ, and minimally invasive adenocarcinoma with pure ground-glass opacity, Eur. J. Radiol. 85 (10) (2016) 1708–1715.
[32] H.Y. Kim, Y.M. Shim, K.S. Lee, J. Han, C.A. Yi, Y.K. Kim, Persistent pulmonary nodular ground-glass opacity at thin-section CT: histopathologic comparisons, Radiology 245 (1) (2007) 267–275.
[33] X. Jin, S.H. Zhao, J. Gao, D.J. Wang, J. Wu, C.C. Wu, et al., CT characteristics and pathological implications of early stage (T1N0M0) lung adenocarcinoma with pure ground-glass opacity, Eur. Radiol. 25 (9) (2015) 2532–2540.
[34] L. Felix, G. Serra-Tosio, S. Lantuejoul, J.F. Timsit, D. Moro-Sibilot, C. Brambilla, et al., CT characteristics of resolving ground-glass opacities in a lung cancer screening programme, Eur. J. Radiol. 77 (3) (2011) 410–416.
[35] H. MacMahon, D.P. Naidich, J.M. Goo, K.S. Lee, A.N.C. Leung, J.R. Mayo, et al., Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017, Radiology 284 (1) (2017) 228–243.
[36] V.P. Balachandran, M. Gonen, J.J. Smith, R.P. DeMatteo, Nomograms in oncology: more than meets the eye, Lancet Oncol. 16 (4) (2015) e173–80.
[37] D. Mackin, X. Fave, L. Zhang, D. Fried, J. Yang, B. Taylor, et al., Measuring computed tomography scanner variability of radiomics features, Invest. Radiol. 50 (11) (2015) 757–765.
[38] L. He, Y. Huang, Z. Ma, C. Liang, C. Liang, Z. Liu, Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule, Sci. Rep. 6 (2016) 34921.