BJR

https://doi.org/10.1259/bjr.20211048

Received: 10 September 2021 | Revised: 16 December 2021 | Accepted: 22 December 2021

Cite this article as:
Li Y, Liu J, Yang X, Xu H, Qing H, Ren J, et al. Prediction of invasive adenocarcinomas manifesting as pure ground-glass nodules based on radiomic signature of low-dose CT in lung cancer screening. *Br J Radiol* (2022) 10.1259/bjr.20211048.

FULL PAPER

Prediction of invasive adenocarcinomas manifesting as pure ground-glass nodules based on radiomic signature of low-dose CT in lung cancer screening

YONG LI, MS, JIEKE LIU, MD, XI YANG, MS, HAO XU, MS, HAOMIAO QING, MD, JING REN, BS and PENG ZHOU, BS

Department of Radiology, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

Address correspondence to: Dr Peng Zhou
E-mail: *penghyzhou@126.com*

The authors Yong Li and Jieke Liu contributed equally to the work.

Objective: To develop a radiomic model based on low-dose CT (LDCT) to distinguish invasive adenocarcinomas (IAs) from adenocarcinoma *in situ*/minimally invasive adenocarcinomas (AIS/MIAs) manifesting as pure ground-glass nodules (pGGNs) and compare its performance with conventional quantitative and semantic features of LDCT, radiomic model of standard-dose CT, and intraoperative frozen section (FS).
Methods: A total of 147 consecutive pathologically confirmed pGGNs were divided into primary cohort (43 IAs and 60 AIS/MIAs) and validation cohort (19 IAs and 25 AIS/MIAs). Logistic regression models were built using conventional quantitative and semantic features, selected radiomic features of LDCT and standard-dose CT, and intraoperative FS diagnosis, respectively. The diagnostic performance was assessed by area under curve (AUC) of receiver operating characteristic curve, sensitivity, and specificity.
Results: The AUCs of quantitative-semantic model, radiomic model of LDCT, radiomic model of standard-dose CT, and FS model were 0.879 (95% CI, 0.801-0.935), 0.929 (95% CI, 0.862-0.971), 0.941 (95% CI, 0.876-0.978), and 0.884 (95% CI, 0.805-0.938) in the primary cohort and 0.897 (95% CI, 0.768-0.968), 0.933 (95% CI, 0.815-0.986), 0.901 (95% CI, 0.773-0.970), and 0.828 (95% CI, 0.685-0.925) in the validation cohort. No significant difference of the AUCs was found among these models in both the primary and validation cohorts (all $p > 0.05$).
Conclusion: The LDCT-based quantitative-semantic score and radiomic signature, with good predictive performance, can be pre-operative and non-invasive biomarkers for assessing the invasive risk of pGGNs in lung cancer screening.
Advances in knowledge: The LDCT-based quantitative-semantic score and radiomic signature, with the equivalent performance to the radiomic model of standard-dose CT, can be pre-operative predictors for assessing the invasiveness of pGGNs in lung cancer screening and reducing excess examination and treatment.

INTRODUCTION

Low-dose CT (LDCT) has been widely and routinely used in lung cancer screening,[1,2] leading to an increase in the detection of pulmonary ground-glass nodules (GGNs), particularly in Asian.[3,4] About 90% of pathologically confirmed GGNs were adenocarcinomas in China.[5] The pulmonary adenocarcinoma is categorized to pre-invasive lesions including atypical adenomatous hyperplasia (AAH) and adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IA).[6] Previous studies found the 5-year disease-free survival (DFS) rate of AIS/MIA was close to 100%, while that of IA ranged from 38 to 93%.[7–9] Given the different prognosis, AIS/MIA might be treated with limited resection, while lobectomy should be performed for IA.[10–12] Although pure GGNs (pGGNs) usually tends to be AIS/MIA and a solid component tends to be an indicator of invasiveness,[13,14] a substantial number of pGGNs have been eventually diagnosed as IAs, with an incidence rate of 18.0 to 40.4%.[15–19] Therefore, precision differentiation between IA and AIS/MIA in pGGNs is a priority and also a challenge for thoracic radiologists and surgeons as it is essential to guide clinical decision.

Although radiologists could assess the invasive probability of pGGNs using quantitative and semantic CT features,

such as diameter, mean attenuation, spiculation, and lobulation, the diagnostic criteria were inconsistent with various sensitivity and specificity because of subjectivity.[15,16,19] Radiomics, via high-throughput extraction of features from imaging data,[20–22] has been recently applied in differential diagnosis between IA and AIS/MIA.[23–29] However, all these radiomic models were built using standard-dose CT data rather than LDCT data.

With the prosperities of radiomics, the reproducibility of quantitative radiomic features might be affected by scan acquisition parameters, which aroused the concerns of researchers. The scanner, radiation dose, reconstruction method, and slice thickness could affect the quantification of radiomic features.[30–34] Our previous study demonstrated that the diagnostic performance of radiomic models based on LDCT and standard-dose CT was equivalent to differentiate adenocarcinomas from benign lesions in solid pulmonary nodules.[35] However, the comparison of radiomic models based on LDCT and standard-dose CT for differentiating IA from AIS/MIA in pGGNs has not been addressed.

Therefore, this study is aimed to develop and validate the radiomic model of LDCT to distinguish IA from AIS/MIA manifesting as pGGNs, and compare its diagnostic performance with conventional quantitative and semantic features of LDCT, radiomic model of standard-dose CT, and intraoperative frozen section (FS) diagnosis.

METHODS AND MATERIALS

Patients

The retrospective analysis was approved by the Institutional Review Board of Sichuan Cancer Hospital, and the informed consent was waived. The inclusion criteria for this study were as follows: 1) patients with pulmonary pGGNs detected by LDCT scanning; 2) standard-dose CT scan from pre-operative examination within 1 week; 3) the interval between LDCT scan and standard-dose CT scan less than 1 week (4.64 ± 1.69 days, ranged from 2 to 7); 4) pathologically confirmed by surgery. The exclusion criteria were as follows: 1) pGGNs larger than 3 cm; 2) history of cancer in previous 5 years; 3) images of poor quality with respiratory and movement artifacts. The flowchart of nodules selection was shown in Figure 1.

From July 2018 to December 2020, 132 consecutive patients with 147 pGGNs were included from our hospital. All pGGNs were divided into two independent cohorts according to a ratio of 7:3 and the date of scan. 43 IAs and 60 AIS/MIAs enrolled between July 2018 and January 2020 constituted the primary cohort, and 19 IAs and 25 AIS/MIAs enrolled between January 2020 and December 2020 constituted the validation cohort. The clinical characteristics of age, gender, T stage, pathology of intraoperative FS diagnosis, and final histologic diagnosis were recorded.

Histopathological assessment

Both intraoperative FS diagnosis and paraffin pathology diagnosis were followed by the 2011 International Association for the Study of Lung Cancer, American Thoracic Society, and European Respiratory Society (IASLC/ATS/ERS) classification and the 2015 World Health Organization (WHO) classification for lung adenocarcinoma.[6,13] Accordingly, lung adenocarcinoma was classified as AIS, MIA, and IA. AIS was defined as an adenocarcinoma lesion less than 3 cm with pure lepidic pattern. MIA had a predominant lepidic pattern with an invasive component of less than 5 mm. IA was further classified by the predominant growth pattern using comprehensive histological subtyping (lepidic, acinar, papillary, micropapillary, and solid) and the percentage of each histological component was recorded in 5% increments. The predominant pattern of IA was defined as the pattern with the greatest percentage.

Figure 1. The flowchart of nodules selection

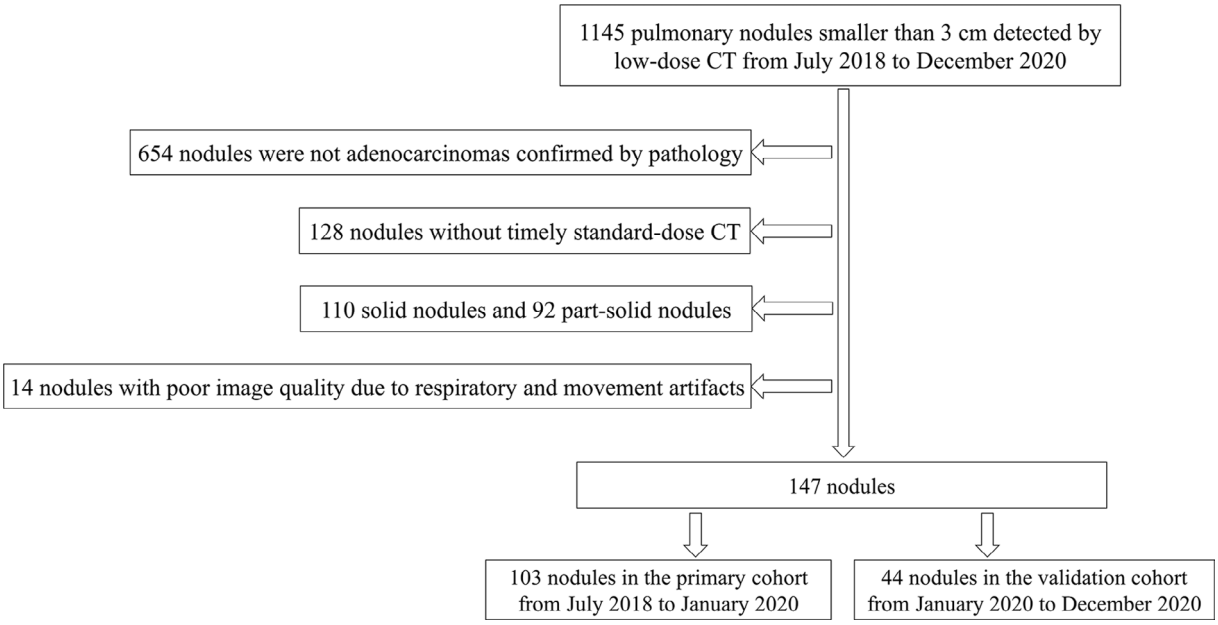


Figure 2. Representative images and segmentation results of nodules. (a, b, c, d) A 53-year-old male with minimally invasive adenocarcinoma. (e, f, g, h) A 41-year-old female with invasive adenocarcinoma. (a, e) Original images and (b, f) segmentation results of low-dose CT. (c, g) Original images and (d, h) segmentation results of standard-dose CT.

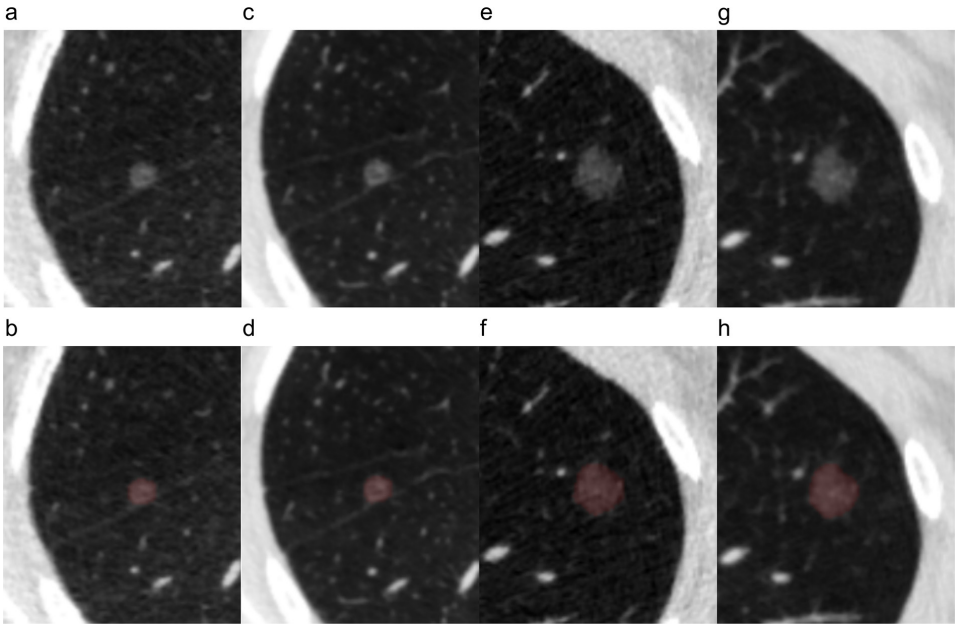


Image acquisition and segmentation

All included patients underwent chest LDCT and standard-dose CT scans on a 256-slice multidetector CT scanner (Brilliance iCT, Philips Healthcare, Amsterdam, Netherlands). All patients were told to hold their breath as soon as possible at the end of inspiration to ensure the quality of the images. The estimated effective dose of LDCT and standard-dose CT for all subjects were 0.69 ± 0.10 mSv and 9.32 ± 2.51 mSv, respectively. More details of acquisition parameters were provided in Supplementary Material 1.

All pGGNs were automatically detected and segmented in three-dimension using uAI platform (United Imaging Healthcare, Shanghai, China), an artificial intelligence (AI) software basing on deep learning method.[36,37] The segmentation results were assessed by two thoracic radiologists (JL and HQ, with 5 years and 10 years of experience) in the lung window (window - 500 HU, width 1500 HU). As all the segmentation results were satisfactory to both radiologists, no manual adjustments of the segmentation results were conducted to avoid inter- and intraobserver variability. The representative segmentation results were shown in Figure 2.

Quantitative-semantic model construction

The conventional quantitative features of LDCT and standard-dose CT including average diameter, volume, and mean attenuation were recorded from the clinical reporting module of uAI platform. Average diameter was the mean of long- and short-axis diameters according to the Fleischner Society guideline.[14] The intraclass correlation coefficient (ICC) was used to evaluate the consistency of the conventional quantitative features between LDCT and standard-dose CT.

Two thoracic radiologists (JL and HQ) were blinded to histopathological results and evaluated the semantic features on LDCT images. The semantic features included nodule shape (round/oval or irregular), margin (smooth or coarse), nodule-lung interface (clear or blurry), lobulation, spiculation, pleural indentation, air bronchogram, and vacuole sign. The cases of disagreement for semantic features between the two radiologists were resolved by consulting a third thoracic radiologist with 25 years of experience (PZ). To evaluate the consistency of the semantic features between LDCT and standard-dose CT, the ICCs were calculated from 40 randomly selected nodules. The semantic features of these nodules on standard-dose CT images were evaluated by a thoracic radiologist (JL) at a 1-week interval.

The quantitative-semantic model was built by multivariable logistic regression of conventional quantitative and semantic features of LDCT with backward stepwise selection, using likelihood ratio test with Akaike's information criterion as stopping rule.[38]

Radiomic feature extraction

The region of interest (ROI) of each pGGN was extracted from the research module of uAI platform after segmentation. A total of 104 original radiomic features including shape, first order, and texture features were then extracted from the ROIs using an open-source Python package (PyRadiomics, v. 3.0, https://pyradiomics.readthedocs.io),[39] which was basically in compliance with the Imaging Biomarker Standardization Initiative (IBSI).[40] A fixed bin width of 25 HU was used to discretize voxel intensity values to efficiently compute texture features and reduce noise,[41,42] which was the same method employed by previous studies.[27,28,43] As the voxel size of all original images was same

and nearly anisotropic ($0.684 \times 0.684 \times 0.625\,mm^3$), no interpolation and resegmentation algorithm was used to resample the original images. Details of radiomic features were described in the Supplementary Material 1.

Radiomic feature selection and signature construction

Radiomic features were standardized with Z-score normalization, and then selected to construct radiomic signatures in the primary cohort using LDCT and standard-dose CT dataset, respectively. Mann-Whitney *U* test was firstly applied to select features which were significantly different between two groups ($p < 0.05$). Secondly, minimum redundancy-maximum relevance (mRMR) was conducted to exclude redundant radiomic features,[44] and the top 10 ranked features were reserved. Thirdly, the least absolute shrinkage and selection operator (LASSO) was utilized to select the most predictive features.[45] Finally, multivariable logistic regression with backward stepwise selection and Akaike's information criterion was applied to construct radiomic signature. Radiomic signature of each patient was calculated according to a linear combination of the selected features and weighted by the respective non-zero coefficients. Radiomic models were built by univariate logistic regression of radiomic signatures of LDCT and standard-dose CT, respectively.

Statistical analysis, model performance, and TRIPOD

The statistical analysis was performed with R software (v. 4.0.3; https://www.r-project.org/), Medcalc (v. 18.2.1; https://www.medcalc.org/), and SPSS software (v. 25.0; https://www.ibm.com). χ^2 test or Fisher exact test was used to compare group difference in gender, T stage, semantic features, and intraoperative FS diagnosis, and independent sample *t*-test was performed in age and conventional quantitative features of LDCT including average diameter, volume, and mean attenuation. $p < 0.05$ was considered statistically significant.

The FS model was built by univariate logistic regression of intraoperative FS diagnosis, which was a binary variable (IA or AIS/MIA). Four models for differentiating IA from AIS/MIA including quantitative-semantic model of LDCT, radiomic model of LDCT, radiomic model of standard-dose CT, and FS model were trained on the primary cohort and tested on the validation cohort. The clinical variables including age and gender were further added to all models except for FS model using multivariable logistic regression.

The discrimination performance of each model in both the primary and validation cohorts was evaluated using the area under the curve (AUC) of receiver operating characteristic (ROC) curve. The optimal threshold value (cut-off point) was delimited according to Youden Index of ROC analysis. The corresponding sensitivity, specificity, and accuracy were also calculated from the confusion matrix. Binomial exact method was used to determine the confidence intervals (CIs). The comparisons of AUCs among models were conducted using Delong test in the primary and validation cohorts.[46]

The Hosmer-Lemeshow test was performed to estimate the goodness-of-fit of each model. The calibration curves were plotted for each model to assess the consistency between predicted and actual probability in both the primary and validation cohorts.

This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.[47]

RESULTS

Demographic characteristics

The clinical characteristics, conventional quantitative features, semantic features, and FS diagnosis in the primary and validation cohorts were listed in Table 1. Significant differences were found in age, T stage, lobulation, air bronchogram, average diameter, volume, and FS diagnosis between IA and AIS/MIA groups in both the primary and validation cohorts (all $p < 0.05$). As for gender, margin, vacuole sign, and mean attenuation, group difference was found in the primary ($p = 0.037$, 0.041, 0.002, and 0.004) but not in the validation cohort ($p = 0.355$, 0.333, 0.074, and 0.096). The shape ($p = 0.009$) were significantly different in the validation cohort but not in the primary cohort ($p = 0.079$). No significant difference was found in nodule-lung interface, speculation, and pleural indentation in the primary or validation cohort (all $p > 0.05$).

Quantitative-semantic model

All the quantitative and semantic features showed good agreement between LDCT and standard-dose CT. The value of ICCs ranged from 0.888 to 0.995. Therefore, the quantitative and semantic features of LDCT were selected to build quantitative-semantic model. The volume, mean attenuation, and vacuole sign were finally selected to build the quantitative-semantic score by multivariable logistic regression analysis (Supplementary Material 1). The calculation formula for quantitative-semantic score was as follows: $\ln(P/1\text{-}P) = 5.717058 + 0.002319 \times \text{volume} + 0.013133 \times \text{mean attenuation} + 1.176054 \times \text{vacuole sign}$, where P is the probability of IA (cut-off > 0.461937). The distributions of the quantitative-semantic score of each patient in the primary and validation cohorts were shown in the Supplementary Material 1.

Radiomic models

The process of radiomic features selection was presented in the Supplementary Material 1. Finally, three features in LDCT and four features in standard-dose CT were selected in the primary cohort (Supplementary Material 1). To conduct z-score normalization before calculation, the mean and standard deviation of these features were provided in Supplementary Material 1.

The calculation formula for radiomic signature of LDCT was as follows: $\ln(P/1\text{-}P) = -0.139131 - 0.854518 \times \text{Shape\_Flatness} + 2.759479 \times \text{First order\_Interquartile range} + 3.160011 \times \text{NGTDM\_Busyness}$, where P is the probability of IA (cut-off > 0.404114).

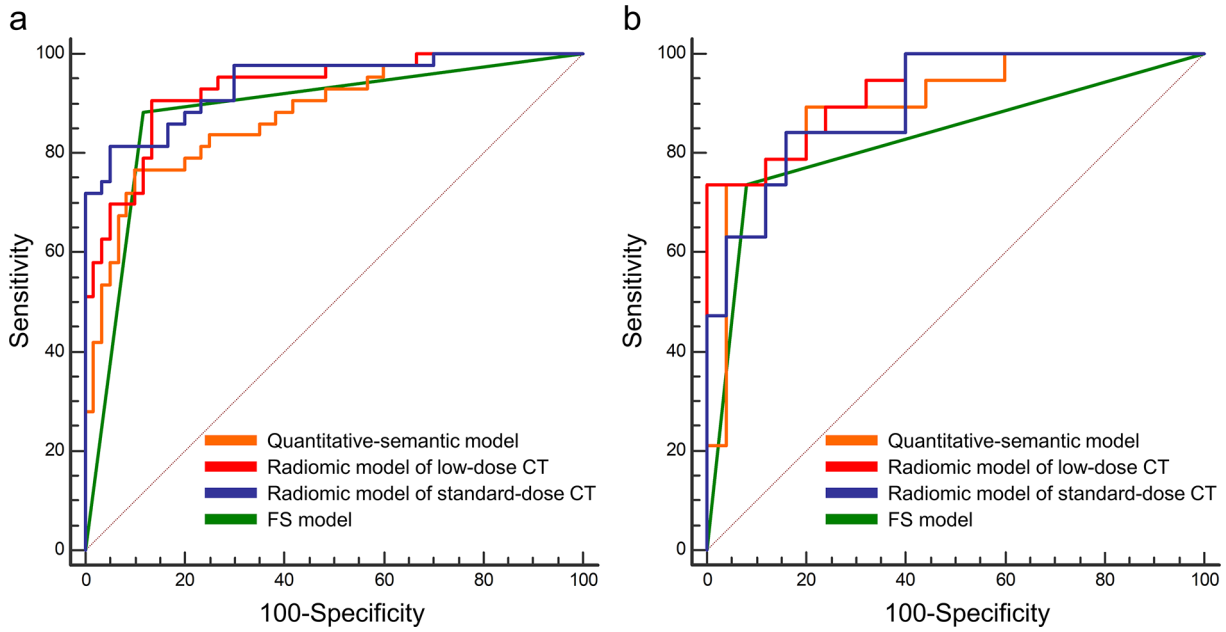
The calculation formula for radiomic signature of standard-dose CT was as follows: $\ln(P/1\text{-}P) = -0.094977 - 1.290115 \times$

Table 1. Demographic characteristics of patients in the primary and validation cohorts

Characteristic	Primary cohort (n = 103)			Validation cohort (n = 44)		
	AIS/MIA (n = 60)	IA (n = 43)	P	AIS/MIA (n = 25)	IA (n = 19)	P
Gender			0.037*			0.355
Female	16	20		6	7	
Male	44	23		19	12	
Age (years)	51.55 ± 9.59	55.91 ± 9.92	0.027*	46.68 ± 12.58	54.37 ± 9.58	0.026*
T stage			<0.001*			0.002*
T1a	36	9		18	4	
T1b	24	29		7	13	
T1c	0	5		0	2	
Shape			0.079			0.009*
Round/oval	37	19		19	7	
Irregular	23	24		6	12	
Margin			0.041*			0.333
Smooth	17	5		5	1	
Coarse	43	38		20	18	
Nodule-lung interface			0.803			0.405
Clear	28	19		15	9	
Blurry	32	24		10	10	
Lobulation			0.017*			0.006*
No	55	32		24	11	
Yes	5	11		1	8	
Spiculation			0.769			0.999
No	59	41		23	17	
Yes	1	2		2	2	
Pleural indentation			0.183			0.054
No	45	27		20	10	
Yes	15	16		5	9	
Air bronchogram			0.007*			0.011*
No	57	32		25	14	
Yes	3	11		0	5	
Vacuole sign			0.002*			0.074
No	49	23		16	7	
Yes	11	20		9	12	
Average diameter (mm)	7.95 ± 2.78	12.55 ± 4.39	<0.001*	7.32 ± 2.52	11.97 ± 4.52	<0.001*
Volume (mm^3)	494.96 ± 398.68	1727.05 ± 1936.51	<0.001*	438.18 ± 395.12	1669.83 ± 2003.81	0.016*
Mean attenuation (HU)	- 657.39 ± 66.24	- 616.29 ± 73.90	0.004*	- 662.99 ± 65.31	- 621.91 ± 94.88	0.096
FS			<0.001*			<0.001*
AIS/MIA	53	7		23	5	
IA	7	38		2	14	

AIS/MIA, adenocarcinoma *in situ*/minimally invasive adenocarcinoma; IA, invasive adenocarcinoma; FS, frozen section. Age, average diameter, volume, mean attenuation are shown in mean ± standard deviation, and other data are the number of nodules. *p < 0.05.

Figure 3. Receiver operating characteristic curves of the quantitative-semantic model, radiomic models, and FS model for differentiating invasive adenocarcinomas from adenocarcinoma *in situ*/minimally invasive adenocarcinomas. (a) Primary cohort. (b) Validation cohort. FS, frozen section.



Shape_Flatness + 1.795128 × First order_Interquartile range + 1.575698 × GLRLM_High grey level run emphasis + 2.881956 × NGTDM_Busyness, where P is the probability of IA (cut-off > 0.559019).

Distributions of the radiomic signature of each patient in the primary and validation cohorts were shown in the Supplementary Material 1.

Model performance and TRIPOD

Age and gender were not significant when they were inputted in the multivariable logistic regression in quantitative-semantic and radiomic models (Supplementary Material 1).

The ROC curves of quantitative-semantic model, radiomic model of LDCT, radiomic model of standard-dose CT, and FS model in the primary and validation cohorts were shown in Figure 3. The AUC, sensitivity, specificity, and accuracy of each model were shown in Table 2. The AUCs of quantitative-semantic model of LDCT, radiomic model of LDCT, radiomic model of standard-dose CT, and FS model were 0.879 (95% CI, 0.801-0.935), 0.929 (95% CI, 0.862-0.971), 0.941 (95% CI, 0.876-0.978), and 0.884 (95% CI, 0.805-0.938) in the primary cohort and 0.897 (95% CI, 0.768-0.968), 0.933 (95% CI, 0.815-0.986), 0.901 (95% CI, 0.773-0.970), and 0.828 (95% CI, 0.685-0.925) in the validation cohort. According to the DeLong test, no significant difference of the AUCs was found among the

Table 2. Diagnostic performance of the quantitative-semantic model, radiomic models, and frozen section model for differentiating invasive adenocarcinomas from adenocarcinoma *in situ*/minimally invasive adenocarcinomas in the primary and validation cohorts

	AUC	Sensitivity	Specificity	Accuracy
Primary cohort				
Quantitative-semantic model	0.879 (0.801-0.935)	0.767 (0.614-0.882)	0.900 (0.795-0.960)	0.845 (0.760-0.909)
Radiomic model of low-dose CT	0.929 (0.862-0.971)	0.907 (0.779-0.974)	0.867 (0.754-0.941)	0.883 (0.805-0.938)
Radiomic model of standard-dose CT	0.941 (0.876-0.978)	0.814 (0.666-0.916)	0.950 (0.861-0.990)	0.893 (0.817-0.945)
FS model	0.884 (0.805-0.938)	0.884 (0.749-0.961)	0.883 (0.774-0.952)	0.883 (0.805-0.938)
Validation cohort				
Quantitative-semantic model	0.897 (0.768-0.968)	0.737 (0.488-0.909)	0.920 (0.740-0.990)	0.841 (0.699-0.934)
Radiomic model of low-dose CT	0.933 (0.815-0.986)	0.842 (0.604-0.966)	0.800 (0.593-0.932)	0.818 (0.673-0.918)
Radiomic model of standard-dose CT	0.901 (0.773-0.970)	0.789 (0.544-0.939)	0.840 (0.639-0.955)	0.818 (0.673-0.918)
FS model	0.828 (0.685-0.925)	0.737 (0.488-0.909)	0.920 (0.740-0.990)	0.841 (0.699-0.934)

AUC, area under curve; FS, frozen section. Numbers in the parentheses are 95% confidence interval

Table 3. Comparisons of area under the curves among the quantitative-semantic model, radiomic models, and frozen section model in the primary and validation cohorts

Pairwise comparison	Primary cohort		Validation cohort	
	Z	*P*	*Z*	*P*
Quantitative-semantic model *vs* Radiomic model of low-dose CT	1.535	0.125	0.944	0.345
Quantitative-semantic model *vs* Radiomic model of standard-dose CT	1.748	0.081	0.080	0.936
Quantitative-semantic model *vs* FS model	0.085	0.932	0.839	0.402
Radiomic model of low-dose CT *vs* Radiomic model of standard-dose CT	0.543	0.587	0.936	0.350
Radiomic model of low-dose CT *vs* FS model	1.191	0.234	1.618	0.106
Radiomic model of standard-dose CT *vs* FS model	1.723	0.085	0.983	0.326

FS, frozen section.

aforementioned models in both the primary and validation cohorts (all $p > 0.05$) (Table 3).

The calibration curves of quantitative-semantic model, radiomic model of LDCT, and radiomic model of standard-dose CT for predicting the risk of IA showed good consistency between prediction and observation in the primary and validation cohorts (Supplementary Material 1). The Hosmer-Lemeshow test showed no significant difference in both the primary and validation cohorts of quantitative-semantic model model (p = 0.892 and 0.623), radiomic model of LDCT (p = 0.992 and 0.830), and radiomic model of standard-dose CT (p = 0.409 and 0.679), which indicated no departure from the perfect fit.

We concluded the type of this study could be Type 2b, and the TRIPOD checklist is presented in Supplementary Material 1.

DISCUSSION

In the present study, we developed the quantitative-semantic model and radiomic model of LDCT, radiomic model of standard-dose CT, and FS model to distinguish IA from AIS/MIA appearing as pGGNs. The novel finding of this study was that both the quantitative-semantic model and radiomic model of LDCT had equivalent AUC to radiomic model of standard-dose CT and FS model, suggesting that the LDCT-based quantitative-semantic score and radiomic signature could be the pre-operative and non-invasive biomarkers for assessing the invasive risk of pGGNs in lung cancer screening.

Basing on the indolent nature, pGGNs are usually followed up for 3 to 5 years according to the most guidelines of lung cancer screening.[14,48–50] Resections are considered until pGGNs grows or new solid component emerges, namely developing to IAs. It should be noticed that 58% of pathologically confirmed pGGNs in this study were AIS/MIAs, and this ratio ranged from 34 to 93% in previous Asian studies,[15,16,19,23,24,51] which indicating the positive attitudes towards pGGNs and the great fear of missing IAs for both surgeons and patients. The medical environment of China tending to favor cautiousness may exacerbate the over-treatment,[50] especially in the cancer hospital. Besides, the number of detected pGGNs is increasing with the popularization of lung cancer screening and development of AI approaches, which may cause undue anxiety and lead to unnecessary and excess examinations such as standard-dose CT, contrast-enhanced CT, and positron emission tomography. Hence, we developed LDCT-based quantitative-semantic and radiomic models to distinguish IAs from AIS/MIAs in pGGNs. More importantly, we demonstrated that the quantitative-semantic and radiomic models of LDCT had equivalent AUC to radiomic model of standard-dose CT. This novel finding suggested LDCT-based quantitative-semantic score and radiomic signature could be pre-operative biomarkers for assessing the invasive risk of pGGNs in lung cancer screening.

Thoracic radiologists usually stratify the risk of pulmonary nodules by morphology, size, and attenuation in lung cancer screening. However, morphological characteristics such as spiculation and lobulation are uncommon and atypical in early-stage lung cancer appearing as pGGNs.[16,19,24] In our quantitative-semantic model, the vacuole sign was the only independent variable among the semantic features. The vacuole, also known as bubble-like appearance, was associated with increasing nodule size[52] and had histological characteristics of collapse and dilated bronchioles.[53] Previous study also found that the vacuole sign could be an independent predictive factor of invasive adenocarcinoma after adjustment for multiple covariates.[51] Compared with the semantic features, the quantitative features had better reproducibility with the application of advanced AI software regardless of the subjectivity and experience of radiologists. Previous studies found that quantitative measures including size, attenuation, and mass can identify the pathological invasiveness of lung adenocarcinomas, and the most common quantitative measure was the diameter. The optimal cutoff diameter ranged from 10.0 to 16.4 mm, resulting in various sensitivity (53.3%-89.5%) and specificity (70.9%-100%).[15,16,19,51] Although the diameter was significant different between AIS/MIA and IA in our primary and validation cohorts, the volume instead of the diameter was included in the final quantitative-semantic model, which might be due to the collinearity between the volume and diameter. The mean attenuation was another significant quantitative feature in predicting invasiveness of pGGNs as the amount of alveolar airspace and the thickness of alveolar walls might cause the difference of attenuation between AIS/MIA and IA.[14]

Several previous studies of standard-dose CT compared the performance of radiomic features with conventional quantitative

and semantic features in predicting invasiveness of pGGNs, but the results were inconsistent. Some reported that the radiomic model had higher AUC than the conventional radiological model,[23,54] while others reported no significant difference.[25,26,55] In this study of LDCT, the combined quantitative-semantic model consisting of vacuole sign, volume, and attenuation showed comparable diagnostic performance (AUC = 0.879 and 0.897 in the primary and validation cohorts) compared with the radiomic model (AUC = 0.929 and 0.933 in the primary and validation cohorts). One possible reason of the limited value for radiomic features was that the relatively obscure variability of gray value in pGGNs,[25] which might be aggravated by radiation dose reduction and noise increase. Another possible reason was that the semantic features in this study were evaluated by thoracic radiologists experienced with the diagnosis of pulmonary GGNs.[26]

Four radiomic features were selected to construct radiomic models of LDCT and standard-dose CT, which were analyzed as follows: The Shape_Flatness, First order_Interquartile range, and NGTDM_Busyness were common significant features in both radiomic model of LDCT and CT. High value of Shape_Flatness indicates a ROI is sphere-like. It suggested that IAs were flatter while AIS/MIAs were more spherical. First order_Interquartile range is the difference between 75th and 25th percentile of the image grey level intensity. High value of NGTDM_Busyness represents an image with rapid changes of intensity between pixels and its neighborhood. Our radiomic models indicated that IAs were more heterogeneous than AIS/MIAs. GLRLM_High gray level run emphasis measures the distribution of the higher gray level intensities. It suggested that IAs had greater concentration of high gray level intensities than AIS/MIAs.

Currently, FS is a common method for intraoperative pathological classification of pulmonary nodules. However, the diagnosis of an invasive component cannot be confirmed without entire histologic sampling of the lesion.[6] Our results showed that the intraoperative FS diagnosis was not completely consistent with final histologic diagnosis, which was similar to previous studies.[56–59] It should be noticed nine AIS/MIAs were overdiagnosed as IAs in FS in our study (seven in primary cohort and two in validation cohort). This may result from the subjective overestimation of AIS/MIAs by pathologists to avoid IA underestimation. Further efforts should be made to improve the accuracy of FS in distinguishing IA from AIS/MIA to guide intraoperative decisions of pGGNs. Our study found no difference in diagnostic performance between the radiological models and FS model, suggesting that the radiological features could be non-invasive biomarkers to assist the diagnosis of invasive component in pGGNs.

There are several limitations in our study. First, potential selection bias cannot be avoided due to the inherent nature of retrospective study. Second, this is a single-center study with relatively small sample. The advantages are the standardization of acquisition parameters and the use of automatic segmentation approach, avoiding potential confounding biases caused by heterogeneous acquisition parameters and manual contouring.[30,34] Third, different cohorts, scanners, and reconstruction methods are also known to affect the radiomic features.[30–34] Further external validation data sets are needed to test the reliability of our radiological models. Fourth, previous studies using standard-dose CT have demonstrated improved diagnostic performance in predicting invasiveness of GGNs by adding perinodular radiomic features.[28,29] This still needs further study in LDCT.

In conclusion, the quantitative-semantic and radiomic models of LDCT had equivalent diagnostic performance to radiomic model of standard-dose CT and FS model in differentiating IA from AIS/MIA in pGGNs. The LDCT-based quantitative-semantic score and radiomic signature can be preoperative and non-invasive biomarkers for assessing the invasive risk of pGGNs in lung cancer screening and reducing excess examination and treatment.

FUNDING

This study has received funding by Sichuan Science and Technology Program (grant numbers 2021YFS0075, 2021YFS0225, 2019YJ0585).

REFERENCES

1. Patz EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in national lung screening trial participants who underwent low-dose ct prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. *Lancet Oncol* May 2016; **17**: 590–99. https://doi.org/10.1016/S1470-2045(15)00621-X
2. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume ct screening in a randomized trial. *N Engl J Med* February 6, 2020; **382**: 503–13. https://doi.org/10.1056/NEJMoa1911793
3. Zhang Y, Jheon S, Li H, Zhang H, Xie Y, Qian B, et al. Results of low-dose computed tomography as a regular health examination among chinese hospital employees. *J Thorac Cardiovasc Surg* 2020; **160**: 824–31. https://doi.org/10.1016/j.jtcvs.2019.10.145
4. Lee CT. What do we know about ground-glass opacity nodules in the lung? *Transl Lung Cancer Res* 2015; **4**: 656–59. https://doi.org/10.3978/j.issn.2218-6751.2015.04.05
5. Yang W, Sun Y, Fang W, Qian F, Ye J, Chen Q, et al. High-resolution computed tomography features distinguishing benign and malignant lesions manifesting as persistent solitary subsolid nodules. *Clin Lung Cancer* 2018; **19**: e75-83. https://doi.org/10.1016/j.cllc.2017.05.023
6. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, 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* 2011; **6**: 244–85. https://doi.org/10.1097/JTO.0b013e318206a221

7. Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. Does lung adenocarcinoma subtype predict patient survival?: a clinicopathologic study based on the new international association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary lung adenocarcinoma classification. *J Thorac Oncol* 2011; **6**: 1496–1504. https://doi.org/10.1097/JTO.0b013e318221f701
8. Yoshizawa A, Motoi N, Riely GJ, Sima CS, Gerald WL, Kris MG, et al. Impact of proposed iaslc/ats/ers classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage i cases. *Mod Pathol* 2011; **24**: 653–64. https://doi.org/10.1038/modpathol.2010.232
9. Zhang J, Wu J, Tan Q, Zhu L, Gao W. Why do pathological stage ia lung adenocarcinomas vary from prognosis?: a clinicopathologic study of 176 patients with pathological stage ia lung adenocarcinoma based on the iaslc/ats/ers classification. *J Thorac Oncol* 2013; **8**: 1196–1202. https://doi.org/10.1097/JTO.0b013e31829f09a7
10. Blasberg JD, Pass HI, Donington JS. Sublobar resection: a movement from the lung cancer study group. *J Thorac Oncol* 2010; **5**: 1583–93. https://doi.org/10.1097/jto.0b013e3181e77604
11. Van Schil PE, Asamura H, Rusch VW, Mitsudomi T, Tsuboi M, Brambilla E, et al. Surgical implications of the new iaslc/ats/ers adenocarcinoma classification. *Eur Respir J* 2012; **39**: 478–86. https://doi.org/10.1183/09031936.00027511
12. Cho JH, Choi YS, Kim J, Kim HK, Zo JI, Shim YM. Long-term outcomes of wedge resection for pulmonary ground-glass opacity nodules. *Ann Thorac Surg* 2015; **99**: S0003-4975(14)01659-2: 218–22: . https://doi.org/10.1016/j.athoracsur.2014.07.068
13. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, 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* 2015; **10**: S1556-0864(15)33571-1: 1243–60: . https://doi.org/10.1097/JTO.0000000000000630
14. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on ct images: from the fleischner society 2017. *Radiology* July 2017; **284**: 228–43. https://doi.org/10.1148/radiol.2017161659
15. Lee SM, Park CM, Goo JM, Lee H-J, Wi JY, Kang CH. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using ct features. *Radiology* 2013; **268**: 265–73. https://doi.org/10.1148/radiol.13120949
16. Lim H-J, Ahn S, Lee KS, Han J, Shim YM, Woo S, et al. Persistent pure ground-glass opacity lung nodules ≥ 10 mm in diameter at ct scan: histopathologic comparisons and prognostic implications. *Chest* 2013; **144**: S0012-3692(13)60675-3: 1291–99: . https://doi.org/10.1378/chest.12-2987
17. Liang J, Xu X-Q, Xu H, Yuan M, Zhang W, Shi Z-F, et al. Using the ct features to differentiate invasive pulmonary adenocarcinoma from pre-invasive lesion appearing as pure or mixed ground-glass nodules. *Br J Radiol* 2015; **88**: 20140811: 1053: . https://doi.org/10.1259/bjr.20140811
18. Jin X, Zhao S, Gao J, Wang D, Wu J, Wu C, et al. CT characteristics and pathological implications of early stage (t1n0m0) lung adenocarcinoma with pure ground-glass opacity. *Eur Radiol* 2015; **25**: 2532–40. https://doi.org/10.1007/s00330-015-3637-z
19. Chu Z-G, Li W-J, Fu B-J, Lv F-J. CT characteristics for predicting invasiveness in pulmonary pure ground-glass nodules. *AJR Am J Roentgenol* August 2020; **215**: 351–58. https://doi.org/10.2214/AJR.19.22381
20. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48**: 441–46. https://doi.org/10.1016/j.ejca.2011.11.036
21. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; **278**: 563–77. https://doi.org/10.1148/radiol.2015151169
22. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749–62. https://doi.org/10.1038/nrclinonc.2017.141
23. Hwang I-P, Park CM, Park SJ, Lee SM, McAdams HP, Jeon YK, 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* 2015; **50**: 798–804. https://doi.org/10.1097/RLI.0000000000000186
24. Fan L, Fang M, Li Z, Tu W, Wang S, Chen W, et al. Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. *Eur Radiol* 2019; **29**: 889–97. https://doi.org/10.1007/s00330-018-5530-z
25. Xu F, Zhu W, Shen Y, Wang J, Xu R, Qutesh C, et al. Radiomic-based quantitative ct analysis of pure ground-glass nodules to predict the invasiveness of lung adenocarcinoma. *Front Oncol* 2020; **10**: 872. https://doi.org/10.3389/fonc.2020.00872
26. Sun Y, Li C, Jin L, Gao P, Zhao W, Ma W, et al. Radiomics for lung adenocarcinoma manifesting as pure ground-glass nodules: invasive prediction. *Eur Radiol* 2020; **30**: 3650–59. https://doi.org/10.1007/s00330-020-06776-y
27. Wu G, Woodruff HC, Sanduleanu S, Refaee T, Jochems A, Leijenaar R, et al. Preoperative ct-based radiomics combined with intraoperative frozen section is predictive of invasive adenocarcinoma in pulmonary nodules: a multicenter study. *Eur Radiol* 2020; **30**: 2680–91. https://doi.org/10.1007/s00330-019-06597-8
28. Wu G, Woodruff HC, Shen J, Refaee T, Sanduleanu S, Ibrahim A, et al. Diagnosis of invasive lung adenocarcinoma based on chest ct radiomic features of part-solid pulmonary nodules: a multicenter study. *Radiology* November 2020; **297**: 451–58. https://doi.org/10.1148/radiol.2020192431
29. Wu L, Gao C, Xiang P, Zheng S, Pang P, Xu M. CT-imaging based analysis of invasive lung adenocarcinoma presenting as ground glass nodules using peri- and intra-nodular radiomic features. *Front Oncol* 2020; **10**: 838. https://doi.org/10.3389/fonc.2020.00838
30. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring computed tomography scanner variability of radiomics features. *Invest Radiol* 2015; **50**: 757–65. https://doi.org/10.1097/RLI.0000000000000180
31. Lo P, Young S, Kim HJ, Brown MS, McNitt-Gray MF. Variability in ct lung-nodule quantification: effects of dose reduction and reconstruction methods on density and texture based features. *Med Phys* 2016; **43**(8): 4854. https://doi.org/10.1118/1.4954845
32. Solomon J, Mileto A, Nelson RC, Roy Choudhury K, Samei E. Quantitative features of liver lesions, lung nodules, and renal stones at multi-detector row ct examinations. *Dependency on Radiation Dose and Reconstruction Algorithm Radiology* 2016; **279**: 185–94. https://doi.org/10.1148/radiol.2015150892
33. Zhao B, Tan Y, Tsai W-Y, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 24, 2016; **6**: 23428. https://doi.org/10.1038/srep23428

34. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of ct features may be nonreproducible and redundant: influence of ct acquisition parameters. *Radiology* August 2018; **288**: 407–15. https://doi.org/10.1148/radiol.2018172361
35. Liu J, Xu H, Qing H, Li Y, Yang X, He C, et al. Comparison of radiomic models based on low-dose and standard-dose ct for prediction of adenocarcinomas and benign lesions in solid pulmonary nodules. *Front Oncol* 2020; **10**: 634298. https://doi.org/10.3389/fonc.2020.634298
36. Wang Q, Zhou X, Wang C, Liu Z, Huang J, Zhou Y, et al. WGAN-based synthetic minority over-sampling technique: improving semantic fine-grained classification for lung nodules in ct images. *IEEE Access* 2019; **7**: 18450–63. 08629907 . https://doi.org/10.1109/ACCESS.2019.2896409
37. Suk H-I, Liu M, Yan P, Lian C. Machine Learning in Medical Imaging. *Relu Cascade of Feature Pyramid Networks for CT Pulmonary Nodule Detection*. Cham: Springer International Publishing; 2019., pp. 444–52. RESOURCE> . https://doi.org/10.1007/978-3-030-32692-0
38. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics* 2001; **57**: 120–25. https://doi.org/10.1111/j.0006-341x.2001.00120.x
39. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* November 1, 2017; **77**: e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339
40. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* May 2020; **295**: 328–38. https://doi.org/10.1148/radiol.2020191145
41. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol* July 7, 2016; **61**: R150-66. https://doi.org/10.1088/0031-9155/61/13/R150
42. Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, et al. Vulnerabilities of radiomic signature development: the need for safeguards. *Radiother Oncol* January 2019; **130**: S0167-8140(18)33551-5: 2–9: . https://doi.org/10.1016/j.radonc.2018.10.027
43. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 3, 2014; **5**: 4006. https://doi.org/10.1038/ncomms5006
44. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005; **27**: 1226–38. https://doi.org/10.1109/TPAMI.2005.159
45. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 30, 2007; **26**: 5512–28. https://doi.org/10.1002/sim.3148
46. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45. https://doi.org/10.2307/2531595
47. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): explanation and elaboration. *Ann Intern Med* 6, 2015; **162**: W1-73. https://doi.org/10.7326/M14-0698
48. Bai C, Choi C-M, Chu CM, Anantham D, Chung-Man Ho J, Khan AZ, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for asia. *Chest* 2016; **150**: S0012-3692(16)01308-8: 877–93: . https://doi.org/10.1016/j.chest.2016.02.650
49. Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 3.2018, nccn clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2018; **16**: 412–41. https://doi.org/10.6004/jnccn.2018.0020
50. Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. Lung cancer ldct screening and mortality reduction - evidence, pitfalls and future perspectives. *Nat Rev Clin Oncol* March 2021; **18**: 135–51. https://doi.org/10.1038/s41571-020-00432-6
51. Ding H, Shi J, Zhou X, Xie D, Song X, Yang Y, et al. Value of ct characteristics in predicting invasiveness of adenocarcinoma presented as pulmonary ground-glass nodules. *Thorac Cardiovasc Surg* 2017; **65**: 136–41. https://doi.org/10.1055/s-0036-1587592
52. Takahashi S, Tanaka N, Okimoto T, Tanaka T, Ueda K, Matsumoto T, et al. Long term follow-up for small pure ground-glass nodules: implications of determining an optimum follow-up period and high-resolution ct findings to predict the growth of nodules. *Jpn J Radiol* 2012; **30**: 206–17. https://doi.org/10.1007/s11604-011-0033-8
53. Saito H, Yamada K, Hamanaka N, Oshita F, Ito H, Nakayama H, et al. Initial findings and progression of lung adenocarcinoma on serial computed tomography scans. *J Comput Assist Tomogr* 2009; **33**: 42–48. https://doi.org/10.1097/RCT.0b013e3181633509
54. Xiong Z, Jiang Y, Che S, Zhao W, Guo Y, Li G, et al. Use of ct radiomics to differentiate minimally invasive adenocarcinomas and invasive adenocarcinomas presenting as pure ground-glass nodules larger than 10 mm. *Eur J Radiol* 2021; **141**: S0720-048X(21)00253-9: 109772: . https://doi.org/10.1016/j.ejrad.2021.109772
55. Song L, Xing T, Zhu Z, Han W, Fan G, Li J, et al. Hybrid clinical-radiomics model for precisely predicting the invasiveness of lung adenocarcinoma manifesting as pure ground-glass nodule. *Acad Radiol* September 2021; **28**: S1076-6332(20)30280-4: e267–77: . https://doi.org/10.1016/j.acra.2020.05.004
56. Yeh Y-C, Nitadori J, Kadota K, Yoshizawa A, Rekhtman N, Moreira AL, et al. Using frozen section to identify histological patterns in stage i lung adenocarcinoma of ≤ 3 cm: accuracy and interobserver agreement. *Histopathology* 2015; **66**: 922–38. https://doi.org/10.1111/his.12468
57. Liu S, Wang R, Zhang Y, Li Y, Cheng C, Pan Y, et al. Precise diagnosis of intraoperative frozen section is an effective method to guide resection strategy for peripheral small-sized lung adenocarcinoma. *J Clin Oncol* 1, 2016; **34**: 307–13. https://doi.org/10.1200/JCO.2015.63.4907
58. Zhu E, Xie H, Dai C, Zhang L, Huang Y, Dong Z, et al. Intraoperatively measured tumor size and frozen section results should be considered jointly to predict the final pathology for lung adenocarcinoma. *Mod Pathol* September 2018; **31**: 1391–99. https://doi.org/10.1038/s41379-018-0056-0
59. Wang B, Tang Y, Chen Y, Hamal P, Zhu Y, Wang T, et al. Joint use of the radiomics method and frozen sections should be considered in the prediction of the final classification of peripheral lung adenocarcinoma manifesting as ground-glass nodules. *Lung Cancer* January 2020; **139**: S0169-5002(19)30712-3: 103–10: . https://doi.org/10.1016/j.lungcan.2019.10.031