Clinical Radiology 74 (2019) 570.e1–570.e11

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

# Differentiating minimally invasive and invasive adenocarcinomas in patients with solitary sub-solid pulmonary nodules with a radiomics nomogram

B. Feng [a,b], X. Chen [a], Y. Chen [b], Z. Li [b], Y. Hao [a,c], C. Zhang [a], R. Li [d], Y. Liao [d], X. Zhang [d], Y. Huang [e], W. Long [a,*]

[a] *The Department of Radiology, The Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen City, Guangdong Province, China*
[b] *School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin City, Guangxi Province, China*
[c] *The Department of Radiology, Guangzhou First People's Hospital, Guangzhou City, Guangdong Province, China*
[d] *The Department of Pathology, The Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen City, Guangdong Province, China*
[e] *The Department of Respiratory Medicine, The Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen City, Guangdong Province, China*

ARTICLE INFORMATION

*Article history:*
Received 1 October 2018
Accepted 22 March 2019

AIM: To evaluate the preoperative differentiation between the minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC) in patients with sub-solid pulmonary nodules using a radiomics nomogram.

MATERIALS AND METHODS: A total of 100 patients with sub-solid pulmonary nodules who had pathologically confirmed MIA (43 patients, 13 male and 30 female) or IAC (57 patients, 26 male and 31 female) were recruited retrospectively. Radiomics features were extracted from computed tomography (CT) images. A radiomics signature was constructed by the least absolute shrinkage and selection operator (LASSO) algorithm. Solid presence, lesion size, shape regularity, and margins of pulmonary nodules were assessed to construct a subjective finding model. An integrated model of radiomics signatures and CT-based subjective findings, which was presented as a radiomics nomogram, was developed based on a multivariate logistic regression. The nomogram performance was assessed by its calibration, discrimination, and clinical usefulness.

RESULTS: The radiomics signature, which consisted of 11 radiomics features, showed good discrimination accuracy. The radiomics nomogram showed good calibration and discrimination in the training set (AUC [area under the curve] 0.943; 95% confidence interval [CI]: 0.895–0.991) and validation set (AUC 0.912; 95% CI: 0.780–1.000). The radiomics nomogram was determined to be clinically useful in the decision curve analysis (DCA).

* Guarantor and correspondent: W. Long, Department of Radiology, The Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen City, Guangdong Province, China. Tel.: +8613902882776.
*E-mail address:* jmlws2@163.com (W. Long).

https://doi.org/10.1016/j.crad.2019.03.018

CONCLUSION: The proposed radiomics nomogram has the potential to preoperatively differentiate MIA and IAC in patients with sub-solid pulmonary nodules.


## Introduction

Due to the increased popularity of low-dose computed tomography (CT) for lung cancer screening, the detection rates of sub-solid pulmonary nodules have dramatically improved.[1] In contrast to solid pulmonary nodules, sub-solid pulmonary nodules do not completely obscure bronchovascular bundles that could be classified as non-solid or part-solid nodules.[2] According to the International Early Lung Cancer Action Project (IELACP), the detection rates of non-solid and part-solid nodules are approximately 4.2% and 5%, respectively.[3,4]

Although a large number of sub-solid nodules have resolved on follow-up CT, persistent sub-solid nodules identified on CT images have a higher likelihood of malignancy compared to solid nodules.[5,6] In 2011, a joint working group of the International Association for the Study of Lung Cancer, American Thoracic Society, and the European Respiratory Society (IASLC/ATS/ERS) developed a new classification scheme for lung adenocarcinomas. The four subtypes of lung adenocarcinoma are atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IAC), which were also issued in the 4$^{th}$ Edition of the World Health Organization (WHO) *Classification of Tumors of the Lungs*.[7–9] The grading of tumours was associated with the disease prognosis and crucial for patient management.[10] Regular follow-up visits were recommended for patients with AAH and AIS. MIA should be resected using a sublobectomy or wedge resection approach without removing any surrounding lymph nodes. For the MIA patients who underwent complete resection, the 5-year survival rate could be as high as 100%.[11,12] Lastly, IAC should be resected with a lobectomy accompanied by removal of the surrounding lymph nodes; however, the 5-year survival rate for these patients was significantly lower at 74.6%.[13]

Although differentiating between MIA and IAC at the preoperative assessment can provide vital insight into treatment planning, differentiating between MIA and IAC remains difficult as the visual assessment of morphological structures at CT is limited by considerable overlapping.[7,14]

Radiomics is a high-throughput image extraction process that computes high-dimensional and mineable data.[15] Using advanced machine learning methods, high-dimensional information, including texture, shape, and intensity features can be extracted from medical images. The subsequent data analysis can be used to relate specific features to the tumour phenotypes.[16] Preoperative assessment of sub-solid pulmonary nodules primarily relies on texture analysis to differentiate between pre-invasive and IAC, while the differentiation of MIA and IAC relies on more high-throughput image features.[17,18] Compared with texture analysis, radiomics combines multiple imaging features to develop predictive signatures to provide a more comprehensive view of the tumour. Previous studies have shown the potential of radiomics for identifying potential imaging biomarkers that may aid cancer detection, diagnosis, and the assessment of prognosis.[19–21] The radiomics approach has been used for differentiation of malignant lesions,[22] tumour subtypes,[23] and prediction of disease-free survival[24] in patients with lung cancer. To date, only one study has utilised traditional CT-based subjective findings for preoperative differentiation of MIA and IAC with sub-solid nodules.[25]

The present study investigated a nomogram that integrated the radiomics signature and CT subjective findings for the differentiation of MIA and IAC with sub-solid pulmonary nodules.

## Materials and methods

### *Patients*

This retrospective study was approved by the Institutional Review Board with a waiver of informed consent. From January 2014 to December 2017, a total of 1,211 cases with histopathologically confirmed pulmonary cancer were collected. Inclusion criteria for this study included: (1) the presence of sub-solid nodules on chest CT; (2) the diameter of nodules was $\leq$30 mm; and (3) the disease was histopathologically confirmed as pulmonary MIA or IAC after surgical resection. The exclusion criteria for this study included: (1) the presence of solid pulmonary nodules at CT; (2) CT section thickness of $>$3 mm; (3) patients who received CT enhancement; (4) the disease was histopathologically confirmed to comprise pre-invasive adenocarcinoma lesions, such as AAH and AIS, or other pathological types such as squamous cell carcinoma or small cell lung cancer; and (5) patients with multiple nodules. After applying the inclusion and exclusion criteria, 100 cases were included in the study. Patients were further divided into two independent sets, in which 70 patients who were seen between January 2014 and June 2017 were the training set and 30 patients who were seen between July 2017 and Nov 2017 were the validation set. Two experienced pathologists (with 12 and 10 years of pulmonary pathology experience, respectively) reviewed all pathological specimens and classified the lung lesions as AAH, AIS, MIA, or IAC according to the new lung adenocarcinoma classification guidelines.[8]

### *CT image acquisition*

CT examinations were performed using either a Somatom Sensation 16 (Siemens, Munich, Germany) or Aquilion

64 (Toshiba, Tokyo, Japan) CT machine. The patients were in the supine position with scan coverage from the thorax inlet to the bilateral adrenal glands level. Thoracic CT was acquired with a deep inhaled breath-hold. CT parameters were as follows: 120 kV, automatic tube current, 0.75–1.5 pitch, 0.625–1.25 mm collimation, 200×200 scan field, 512×512 pixel. Scan dose was 50–150 mGy·cm depending on the patient's body mass index. Both standard and high-resolution algorithms were reconstructed (1–3 mm section thickness; 0.8–3 mm intersection gap). Neither of the CT examinations was performed using contrast medium.

### *Evaluation of CT-based subjective findings*

CT images were reviewed independently by two chest radiologists (with 30 and 10 years of experience, respectively), who were blinded to the histopathology diagnosis. The scans were analysed in the lung window setting (1,500 HU window width; −600 HU window level) and mediastinal window (300 HU window width; 40 HU window level). The CT-based subjective findings were based on the followings: (1) solid presence (non-solid nodule manifests as hazy with increased attenuation in the lung that does not obliterate the bronchial and vascular margins; part-solid nodule consists of both ground-glass and solid soft-tissue attenuation components); (2) lesion size (the average of major diameter and vertical short diameter measuring on the largest section in lung window); (3) margin (spiculate or non-spiculate); (4) border (lobulate or non-lobulate). CT-based subjective findings were recorded in consensus.

### *Volume of interest (VOI) segmentation and radiomics feature extraction*

A semi-automatic in-house software was used for volume of interest (VOI) segmentation that was mainly implemented by MATLAB 2016 (Mathworks, Natick, MA, USA). A radiologist provided the initial position of the pulmonary nodule. The boundary curve of the pulmonary nodules in the selected rectangular region was calculated automatically and the final VOI was contoured on the two-dimensional (2D) image sections of the entire tumour volume. A sample segmentation is shown in Fig 2.

The radiomics features used in this study contained 10,329 three-dimensional (3D) features. The extracted radiomics features included first-order statistics, shape and size-based features, texture-based features, and wavelet-

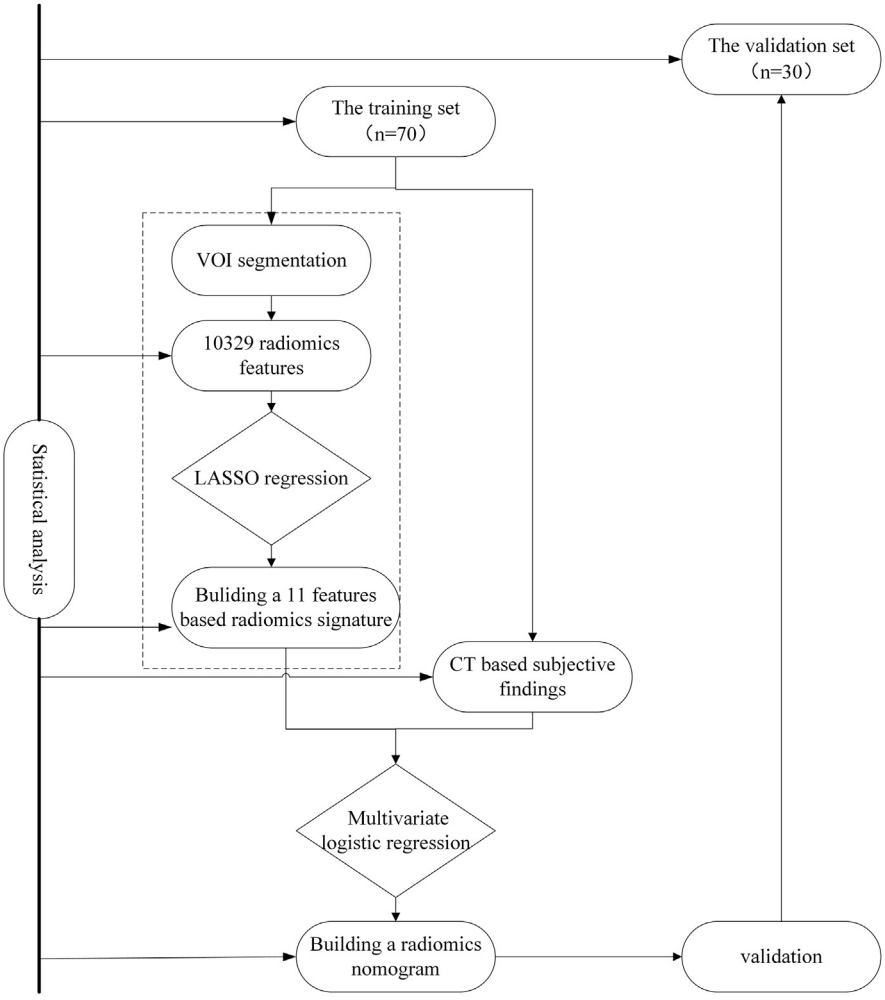


**Figure 1** Study flowchart detailing the radiomics nomogram.

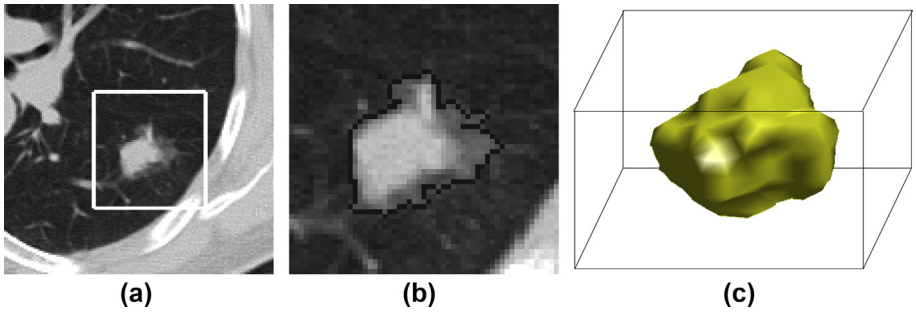


**Figure 2** VOI segmentation of the sub-solid pulmonary nodule. (a) Original CT image of the part-solid pulmonary nodule. (b) 2D segmentation result of the pulmonary nodule. (C) 3D volumetric reconstruction of the pulmonary nodule.

based features. This extraction step was performed in MATLAB 2016.

*Statistical analysis*

The differences in age and gender between the MIA and IAC groups were assessed by the Fisher's exact test and Pearson chi-squared test, respectively. The inter-reader agreements of the subjective findings were assessed by Cohen's kappa test. The interpretations were as follows, $\leq$0 as no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as nearly perfect agreement. The CT-based subjective findings model was developed in two steps. First, the differences in the four CT-based subjective findings were assessed using Fisher's exact test or the chi-square test. Second, the subjective findings were combined with the significant variables from the univariate analysis by multivariate logistic regression. From the training group, 20 pulmonary sub-solid nodules were randomly selected. VOI groups were obtained using two different types of in-house segmentation software. The radiomics features were then extracted from both of the groups. A Mann–Whitney *U*-test was used to evaluate each radiomics feature for differentiation of the MIA and IAC groups. The radiomics features, which had an ICC >0.75 and were significantly different between the MIA and IAC groups, were selected for the least absolute shrinkage and selection operator (LASSO) logistic regression with a 10-fold cross validation.

Coefficients from the LASSO logistic regression for each radiomics features were calculated. The radiomics features with non-zero coefficients were selected as the most valuable feature for differentiation of MIA and IAC groups. A radiomics signature was built by the linear combination of the LASSO-selected features and their corresponding coefficients. Differences of the radiomics signatures between the MIA and IAC groups were evaluated using the Mann–Whitney *U*-test.

The integrated prediction model, which incorporated the radiomics signature and CT subjective findings, was carried out by multivariable logistic regression analysis. It used the significant predictors from the univariate analysis as inputs. Backward step-wise selection with the stopping rule of the likelihood ration test was performed with Akaike's information. A radiomics nomogram based on the integrated model was developed with multivariable logistic regression to provide a quantitative tool for preoperative differentiation between MIA and IAC.

The calibration of the nomogram was accomplished with a calibration curve analysis. The Hosmer–Lemeshow test was used to calculate the goodness of fit. The performance of the independent validation was tested. The formula obtained in the training set under 10-fold cross validation was used to compute a radiomics score in the validation set. Receiver operating curve (ROC) analysis was performed to evaluate the performance of the subjective findings model, the radiomics signature, and the radiomics nomogram in the training and validation sets. The area under the curve (AUC), sensitivity, specificity, and accuracy were calculated.

The classification generalisability of the CT subjective findings model, radiomics signature, and the combined model were evaluated in terms of AUC, sensitivity, specificity, and accuracy in the combined training and validation sets. The cut-off value was selected to maximise the sum of the sensitivity and specificity values. The improvements of introducing subjective CT findings were quantified by net reclassification improvement (NRI).

All statistical tests were performed using R3.0.1 (http://www.rproject.org) and MATLAB. LASSO was done through the "glmnet" package, and the ROC was done via "pROC". The nomogram was completed by "rms" and DCA was completed by "dca.r." The two-tailed *t*-test was used to compare the parametric difference between MIA and IAC and $p<0.05$ was considered statistically significant.

## Results

*Clinical characteristics of sub-solid pulmonary nodules*

The patient characteristics in the training and validation sets are shown in Table 1. There were no significant differences in gender or age between groups in the training set (gender: $p=0.3333$; age $p=0.2385$) or validation set (gender $p=0.2599$; age $p=0.1931$). The study flowchart is shown in Fig 1.

**Table 1**
Clinical characteristics and computed tomography (CT) features of the training data set and the testing data set.

| | Training data set ($n$=70) | | | Testing data set ($n$=30) | | |
|---|---|---|---|---|---|---|
| | MIA | IAC | $p$-Value | MIA | IAC | $p$-Value |
| Gender | | | | | | |
| Male | 10 | 17 | 0.3333 | 3 | 9 | 0.2599 |
| Female | 21 | 22 | | 9 | 9 | |
| Age (years) | | | | | | |
| <60 | 21 | 21 | 0.2385 | 11 | 12 | 0.1931 |
| ≥ 60 | 10 | 18 | | 1 | 6 | |
| Lesion Size | | | | | | |
| <10 mm | 27 | 5 | <0.0001 | 8 | 4 | 0.0243 |
| ≥ 10 mm | 4 | 34 | | 4 | 14 | |
| Solid presence | | | | | | |
| Non-solid | 28 | 16 | <0.0001 | 10 | 6 | 0.0106 |
| Part-solid | 3 | 23 | | 2 | 12 | |
| Margin | | | | | | |
| Non-spiculate | 10 | 29 | 0.0004 | 4 | 15 | 0.0086 |
| Spiculate | 21 | 10 | | 8 | 3 | |
| Border | | | | | | |
| Non-lobulate | 25 | 14 | 0.0002 | 12 | 5 | 0.0001 |
| Lobulate | 6 | 25 | | 0 | 13 | |
| Rad-score, media[a] | −1.335 (−1.946 to-0.754) | 1.786 (0.278−3.072) | <0.0001 | −0.816 (−1.735 to -0.462) | 3.290 (1.750−4.088) | 0.0003 |

$p$-Value is derived from the univariate association analysis between the clinical factor and the different subgroup of lung adenocarcinoma. $p<0.05$ indicates statistical significance.
MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma.
[a] Rad-score is denoted as interquartile range.

### *Inter-observer agreement of CT subjective findings of sub-solid pulmonary nodules*

Lesion size and solid presence showed excellent agreement with kappa values of 0.840 (95% CI: 0.753−0.938) and 0.854 (95% CI: 0.756−0.959), respectively. In addition, margin and border showed good agreement with kappa values of 0.754 (95% CI: 0.615−0.897) and 0.701 (95% CI: 0.562−0.82), respectively. Lesion size, solid presence, margin, and border were all significantly different between the MIA and IAC groups in the training set ($p<0.0001$, $p<0.0001$, $p=0.0004$, and $p=0.0002$, respectively), as shown in Table 1.

For the multivariable logistic regression analysis, lesion size (odds ration [OR], 24.649; 95% CI: 5.990−101.438) and solid presence (OR, 4.201; 95% CI: 1.389−37.323) were determined to be independent predictors in the subjective findings model.

### *Radiomics feature selection and radiomics signature building*

There were 7,434 radiomics features with ICC values >0.75 with significant differences between MIA and IAC (p=0.00001−0.05) in the training set that were involved in the LASSO logistic regression analysis. From these, 11 radiomics features with non-zero coefficients in the LASSO logistic regression model were selected to build a radiomics score calculation formula (Fig 3). The radiomics score calculation formula and the selected radiomics features are presented in Electronic Supplementary Material Table S1.

In the selected 11 features, six cluster-size-related features of MIA groups, including Global_Variance_0.50_0.8_8, GLSZM_ZSN_0.5_1_8, GLSZM_SZLGE_0.67_1_8, GLRLM _LRHGE _0.67_1.2 _64, GLRLM_SRE_1_0.8_8, and GLRLM_LRHGE_1.50_1.2 _32, have smaller values than those of IAC groups, as shown in Table 2; however, MIA groups were larger than those of IAC groups in terms of GLRLM_RLV_0.5_1.2_8, GLSZM_SZE _0.5_1_32, NGT DM_Coarseness_0.67_1.2_8, GLSZM_HGZE_2_2_16, and GLCM_ Dissimilarity_2_2_8, which described the texture units.

The radiomics scores revealed a significant difference between the groups in the training set ($p<0.0001$) and validation set ($p=0.0003$), as shown in Table 1. With the cut-off value of 0.921, derived by ROC analysis, the radiomics scores showed high discrimination efficiency in the training set with an AUC=0.935 (95% CI: 0.882−0.987) and validation set with an AUC=0.903 (95% CI: 0.774−1.000), as shown in Fig 3. The waterfall plot for the distribution of the radiomics score of each patient is presented in Fig 4.

Long Run High Gray Level Emphasise (LRHG LE_0.67_1.2_64, OR 3.469 [95%CI: 0.566−7.680]) based on Gray Level Run Length Matrix (GLRLM) and the Coarseness_0.67_1.2_8 (OR 0.122 [95% CI 0.040−0.375]) based on the Neighborhood Gray-Tone Difference Matrix were used as the core feature in the radiomics label model of 11 image radiomics features by multi-factor logistic regression (Fig 5).

### *Development of an individualised prediction model*

The radiomics signature (OR 3.953; 95% CI: 1.864−8.383) and solid presence (OR 5.348; 95% CI: 0.903−31.668) were identified as independent predictors from multivariable logistic regression analysis (Table 3). The integrated model that incorporated these two

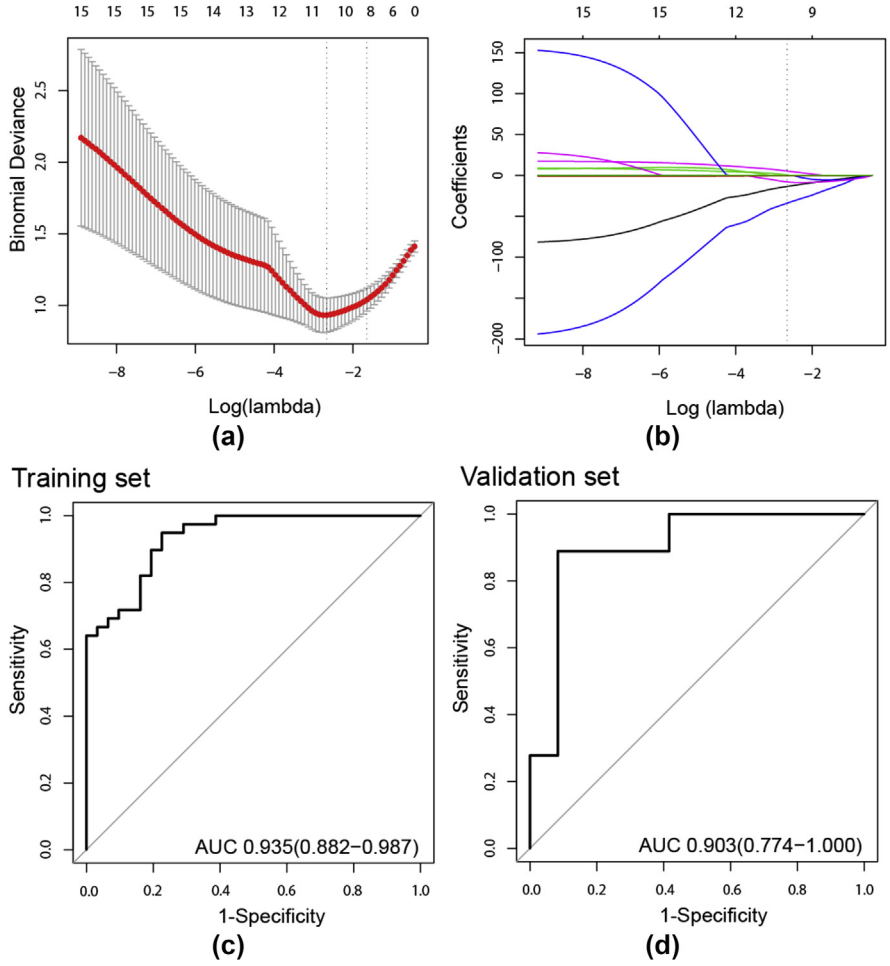


**Figure 3** Feature selection of the LASSO logistic regression and the predictive accuracy of the radiomics signature. (a) Tuning parameter $\lambda$ selection by 10-fold cross-validation with minimum criteria. Binomial deviances ($y$-axis) was plotted versus log($\lambda$) ($x$-axis). The dotted vertical lines were drawn at the optimal value of $\lambda$, where the model providing best fit to the data. (b) LASSO coefficient profiles of the whole features. The dotted vertical line was plotted at the value selected with 10-fold cross-validation, where 11 optimal features with non-zero coefficients were indicated in the plot. (c,d) ROC curves of the radiomics signature in the training and validation sets, respectively.

**Table 2**
The selected 11 radiomics features of MIA and IAC groups, respectively.

| Characteristic | MIA(n=31) | IAC(n=39) | *p*-Value |
|---|---|---|---|
| Global_Variance_0.50_0.8 _8 | 369.0±101.3 | 373.5±66.1 | 0.0015 |
| GLSZM_ZSN_0.5_1_8 | 22.8±21.7 | 50.2±28.9 | 0.0001 |
| GLSZM_SZE_0.5_1_32 | 0.79±0.06 | 0.76±0.04 | 0.0375 |
| GLRLM_RLV_0.5_1.2_8 | 0.014±0.010 | 0.006±0.004 | 0.0001 |
| GLSZM_SZLGE_0.67_1_8 | 0.134±0.044 | 0.148±0.030 | 0.0414 |
| NGTDM_Coarseness_0.67_1.2 _8 | 0.038±0.028 | 0.015±0.013 | 0.0001 |
| GLRLM_LRHGE_0.67_1.2 _64 | 1136.0±449.6 | 1513.0±472.4 | 0.0025 |
| GLRLM_SRE_1_0.8 _8 | 0.8668±0.0491 | 0.8671±0.0270 | 0.0279 |
| GLRLM_LRHGE_1.50_1.2 _32 | 294.8±106.9 | 411.8±116.3 | 0.0001 |
| GLCM_Dissimilarity_2_2 _8 | 1.936±0.522 | 1.629±0.273 | 0.0007 |
| GLSZM_HGZE_2_2 _16 | 88.7±25.6 | 84.7±14.7 | 0.0393 |

*p*-Value is derived by the Mann–Whitney *U*-test.
MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma.

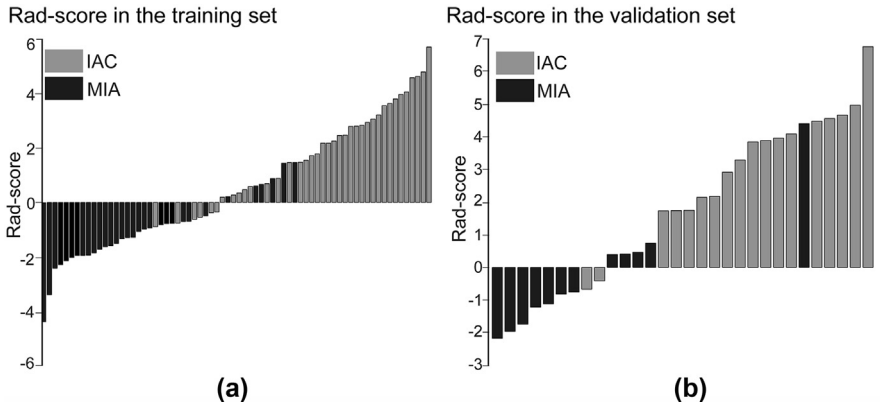


**Figure 4** Waterfall plot for distribution of the radiomics score of each patient in the (a) training set and (b) validation set, respectively. The best cut-off values of the radiomics score were defined based on the maximum Youden's index from the training set. Below the cut-off value, Patients were classified in the MIA groups below the cut-off value and the IAC group above the cut-off value. The cut-off value for discrimination was 0.921.

independent predictors was developed and presented as a radiomics nomogram (Fig 6a).

*Independent validation of the radiomics nomogram*

The calibration curve showed good connection between the predicted data and actual data in the training set (Fig 6b and c). The Hosmer–Lemeshow test ($p$=0.703) showed no departure from the perfect fit. The subjective findings model showed a lower AUC (0.899; 95% CI: 0.821–0.977) in the training set than the AUC of the radiomics nomogram (0.943; 95% CI: 0.895–0.991; Fig 7a). This favourable discrimination ability of the radiomics nomogram was confirmed using the validation set with the AUC (0.912; 95% CI: 0.780–1; Fig 7b).

*Performance comparison of the subjective findings model, the radiomics signature, and the combined model*

The radiomics nomogram achieved the highest sensitivity (0.840), specificity (0.880), accuracy (0.860), and AUC value (0.930, 95% CI: 0.879–0.982); however, the subjective findings model showed lower sensitivity (0.772), specificity (0.832), accuracy (0.800), and AUC value (0.876, 95% CI: 0.798–0.953), as shown in Table 4.

Introducing subjective CT findings into the combined model did not show significant improvement in performance (NRI=−0.057, $p$=0.280, event NRI 0.143; non-event NRI −0.200).

## Discussion

A radiomics signature-based nomogram was constructed to distinguish MIA and IAC in patients with sub-solid pulmonary nodules preoperatively. By extracting high-throughput quantitative descriptors from routinely acquired CT images, radiomics may provide a non-invasive tool to improve routine lung cancer diagnosis and speed up the implementation of personalised medicine. The interobserver agreement for CT findings of sub-solid nodules was substantial to perfect. In particular, classification of sub-solid into non-solid and part-solid showed very good interobserver agreement (kappa=0.84). These results can support the reliability and applicability of image processing according to nodule type.

A previous study showed that age, sex, family history of lung cancer, nodule size, location of the nodule, solid proportion, and spiculation were predictors for evaluating malignant pulmonary nodules.[26] Sub-solid nodules with lobulate shape and spiculate margins were found to be significantly more frequent in the IAC group in the univariate analyses, but not in the multivariate analyses. Only nodule size and solid proportion were significant to the risk of IAC, which is similar to prior studies.[27–30]

Through multivariable logistic regression, the current study showed that LRHGE based on GLRLM and the coarseness based on NGTDM were two important features for the differentiation of MIA and IAC groups. The scale of voxel clusters with high-intensity values was described by LRHGE-GLRLM. The IAC group had a larger LRHGE-GLRLM value relative to MIA that may be interpreted as invasive lesions with a higher degree of accumulation of cancer cells and its strong association with pathological invasion. The coarseness of the NGTDM describes the size of texture units, which increased with coarseness of the NGTDM. Interestingly, the IAC group had a smaller value of coarseness for the NGTDM when compared with the MIA group.

For other cluster-size-related radiomics features (including Global_Variance_0.50_0.8_8, GLSZM_ZSN_0.5_1_8,GLSZM_SZLGE_0.67_1_8, GLRLM_SRE_1_0.8_8 and GLRLM_LRHGE_1.50_1.2_32), the IAC group showed larger values than the MIA group. For other texture-related radiomics features (including GLRLM_RLV_0.5_1.2_8, GLSZM_SZE_0.5_1_32, GLSZM_HGZE_2_2_16 and GLCM_Dissimilarity_2_2_8), the IAC group had smaller values when

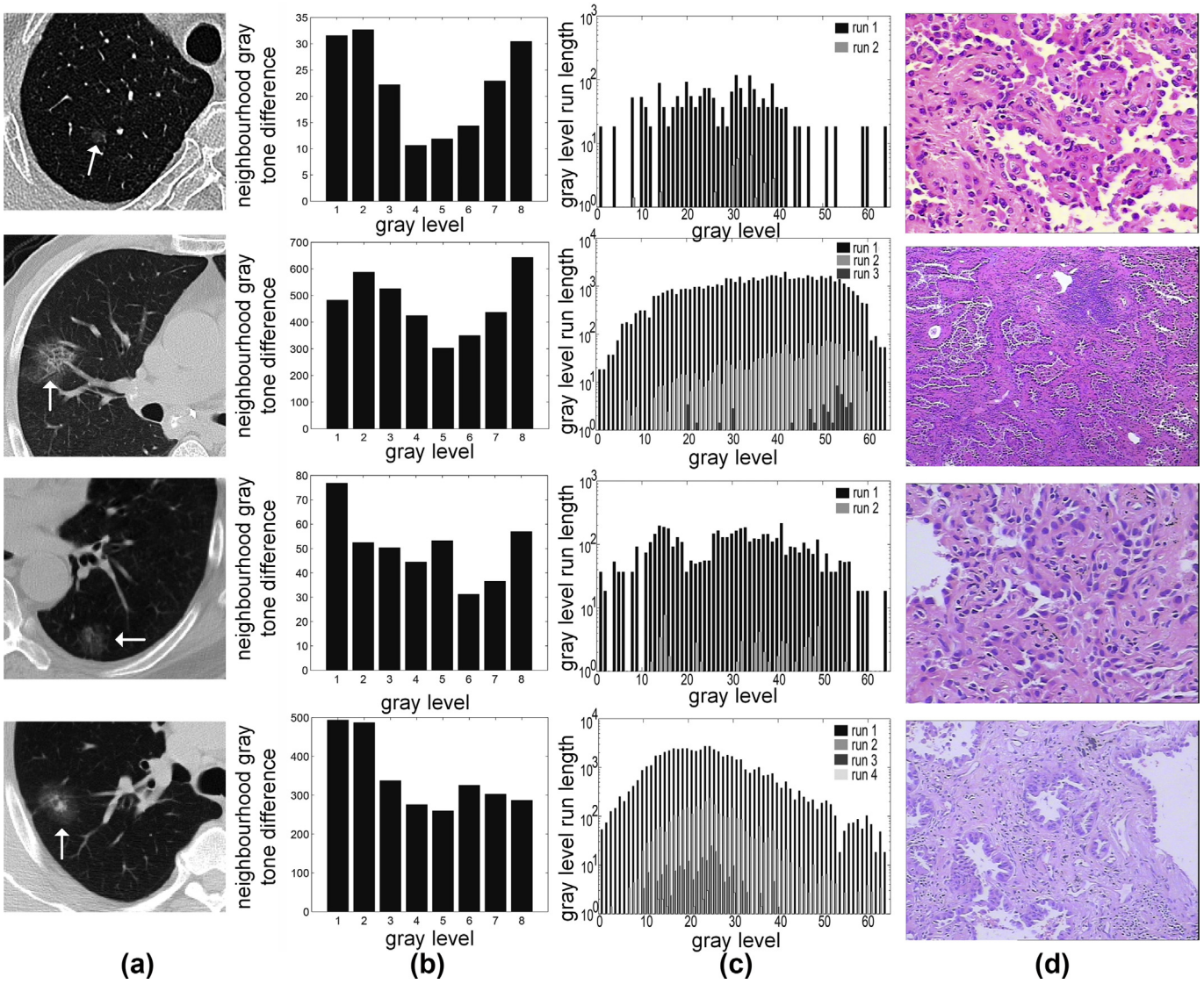


**Figure 5** CT images and radiomics feature maps of sub-solid pulmonary nodules with pathologic diagnosis. From left to right: (a) the original plain CT images; (b) the feature maps of coarseness based on NGTDM; (c) the feature map of LRHGLE based on GLRLM; (d) the pathological images. The first row: CT image shows a 6.8 mm non-solid nodule (white arrow) in the right upper lobe in a 67-year-old man. The value of coarseness based on NGTDM was 0.0425. The value of LRHGLE based on GLRLM was 974.123. This nodule was confirmed as MIA at histopathological analysis (haematoxylin and eosin staining [H&E], ×400). The second row: CT image show a 26.3 mm non-solid nodule (white arrow) in the right upper lobe in a 67-year-old man. The value of coarseness based on NGTDM was 0.0021. The value of LRHGLE based on GLRLM was 1706.739. This nodule was confirmed as IAC at histopathological analysis (H&E, ×100). The third row: CT image shows a 15.2 mm part-solid nodule (white arrow) in the left lower lobe in a 71-year-old man. The value of coarseness based on NGTDM was 0.012. The value of LRHGLE based on GLRLM was 1149.180. This nodule was confirmed as MIA at histopathological analysis (H&E, ×200). The fourth row: CT image shows a 22.4 mm part-solid nodule (white arrow) in the right lower lobe in a 69-year-old man. The value of coarseness based on NGTDM was 0.014. The value of LRHGLE based on GLRLM was 2018.646. This nodule was confirmed as IAC at histopathological analysis (H&E, ×100).

**Table 3**
Multivariate analysis for building the radiomics nomogram.

| Intercept and variable | $\beta$ | Odds ratio (95% CI) | $p$-Value |
|---|---|---|---|
| Intercept | −2.259 | | 0.054 |
| Radiomics signature | 1.366 | 3.953 (1.864–8.383) | <0.001[a] |
| Solid presence | 1.755 | 5.348 (0.903–31.668) | 0.041[a] |

$\beta$, the regression coefficient; CI, confidence interval.
[a] $p<0.05$.

compared with the MIA group. These results were consistent with the two core radiomics features mentioned above. This may be attributed to an increased tumour tissue component and thickening of the alveolar septa in invasive adenocarcinomas. The smaller texture units were built to describe the complex proportions in the IAC group. It reflected the alteration from homogeneous to heterogeneous due to the mixture of solid portions.

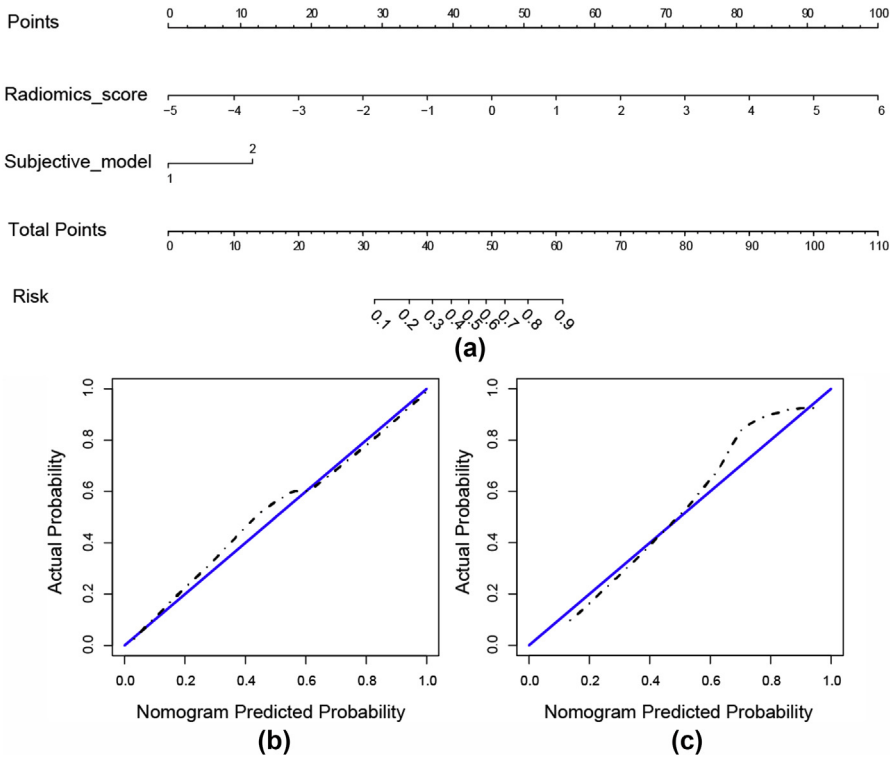


**Figure 6** (a) The developed radiomics nomogram. (b,c) Calibration curves of the radiomics nomogram in the training and validation set, respectively.

The current study was complementary to other studies. The invasiveness of sub-solid pulmonary nodules was evaluated preoperatively using 3D radiomics features. The study demonstrated that 3D features provided more details about the lesions when compared with 2D features.[31] Moreover, reproducibility of radiomics features should be validated according to the radiomics-based guideline.[32] The extracted features with an ICC >0.75 was selected for LASSO logistic regression. Lastly, the proposed radiomics nomogram was easy to use and might provide quantitative risk factors for subsolid nodule.

Several limitations in this study should be addressed in the future. As a retrospective study, there might be selection bias. The performance of the proposed nomogram may be overestimated due to limited number of patients

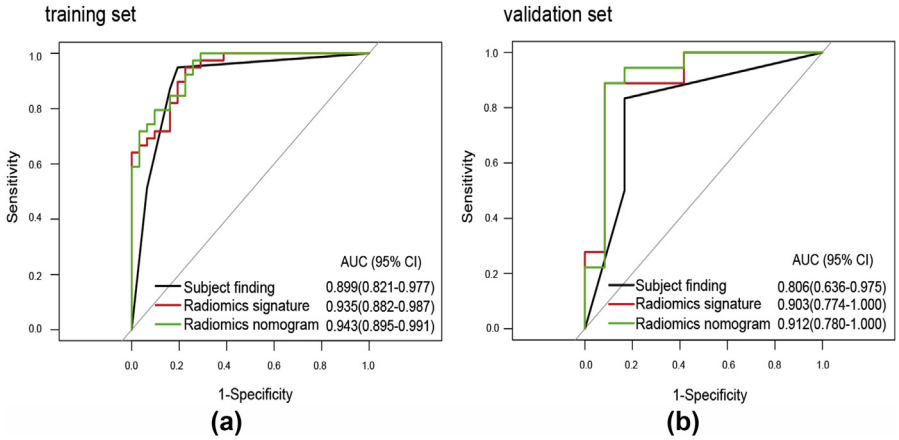


**Figure 7** ROC curves of the radiomics signature (red line), the subjective findings model (black line), and the combined model (green line). The solid dots represented the optimal cut-off values for the discrimination. The AUC and 95% CI for the models are shown in the lower right corner of the figure.

**Table 4**
Diagnostic performance of the subjective findings model, the radiomics signature, and the radiomics nomogram.

| Model | Cut-off | Cross-validated AUC (95% CI) | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|---|
| Subjective findings | 0.500 | 0.876 (0.798–0.953) | 0.772 (44/57) | 0.832 (36/43) | 0.800 (80/100) |
| Radiomics signature | 0.650 | 0.921 (0.867–0.975) | 0.770 (44/57) | 0.880 (38/43) | 0.820 (82/100) |
| Radiomics nomogram | 0.640 | 0.930 (0.879–0.982) | 0.840 (48/57) | 0.880 (38/43) | 0.860 (86/100) |

Numbers in the parentheses were used to calculate percentages.
CI, confidence interval.

in the validation set. Prospective and/or external validation studies were warranted. Second, current study was limited to lung adenocarcinoma nodules; however, adenocarcinoma is the most common histological type in the sub-solid nodules that reflected the actual clinical practice. Third, CT examinations were performed on different CT systems and imaging parameters. Nevertheless, the section thicknesses of reconstructed images were all $\leq$3 mm.

In conclusion, the current nomogram has the potential to accurately distinguish between MIA from IAC among sub-solid pulmonary nodules. This may assist clinicians in selecting the optimal surgical recommendations and improving patient outcomes.

## Conflict of interest

The authors have declared that no conflict of interest exists.

## Acknowledgements

This study was funded by a grant from the Guangxi Natural Science Foundation (no. 2016GXNSFBA380160).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crad.2019.03.018.

## References

1. Aberle DR, Adams AM, Berg CD, *et al*. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;**365**(5):395–409 https://doi.org/10.1056/NEJMoa1102873.
2. Hansell DM, Bankier AA, Macmahon H, *et al*. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008;**246**(3):697–722 https://doi.org/10.1148/radiol.2462070712.
3. Yip R, Yankelevitz DF, Hu M, *et al*. Lung cancer deaths in the national lung screening trial attributed to nonsolid nodules. *Radiology* 2016;**281**(2):589–96 https://doi.org/10.1148/radiol.2016152333.
4. Henschke CI, Yip R, Smith JP, *et al*. CT Screening for lung cancer: part-solid nodules in baseline and annual repeat rounds. *AJR Am J Roentgenol* 2015;**277**(2):1–9 https://doi.org/10.2214/AJR.16.16043.
5. Zhang Y, Shen Y, Qiang JW, *et al*. HRCT features distinguishing pre-invasive from invasive pulmonary adenocarcinomas appearing as ground-glass nodules. *Eur Radiol* 2016;**26**(9):1–8 https://doi.org/10.1007/s00330-015-4131-3.
6. Zhao H, Marshall HM, Yang IA, *et al*. Screen-detected sub-solid pulmonary nodules: long-term follow-up and application of the PanCan lung cancer risk prediction model. *Br J Radiol* 2016;**89**(1060):20160016 https://doi.org/10.1259/bjr.20160016.
7. Kim HY, Shim YM, Lee KS, *et al*. Persistent pulmonary nodular ground-glass opacity at thin-section CT: histopathologic comparisons. *Radiology* 2007;**245**(1):267–75 https://doi.org/10.1148/radiol.2451061682.
8. Travis WD, Brambilla E, Noguchi M, *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**(2):244–85 https://doi.org/10.1097/JTO.0b013e318206a221.
9. Travis WD, Brambilla E, Nicholson AG, *et al*. The 2015 World Health Organization Classification of Lung Tumours: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* 2015;**10**(9):1243–60 https://doi.org/10.1097/JTO.0000000000000630.
10. Montserrat N, González A, Méndez E, *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**(5):653–64 https://doi.org/10.1038/modpathol.2010.232.
11. Van Schil PE, Asamura H, Rusch VW, *et al*. Surgical implications of the new IASLC/ATS/ERS adenocarcinoma classification. *Eur Resp J* 2012;**39**(2):478–86 https://doi.org/10.1183/09031936.00027511.
12. Tsutani Y, Miyata Y, Nakayama H, *et al*. Appropriate sublobar resection choice for ground glass opacity-dominant clinical stage IA lung adenocarcinoma: wedge resection or segmentectomy. *Chest* 2014;**145**(1):66–71 https://doi.org/10.1378/chest.13-1094.
13. Zhang J, Wu J, Tan Q, *et al*. 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**(9):1196–202 https://doi.org/10.1097/JTO.0b013e31829f09a7.
14. Lee HJ, Lee CH, Jeong YJ, *et al*. IASLC/ATS/ERS international multidisciplinary classification of lung adenocarcinoma: novel concepts and radiologic implications. *J Thorac Imaging* 2012;**27**(6):340–53 https://doi.org/10.1097/RTI.0b013e3182688d62.
15. Kumar V, Gu Y, Basu S, *et al*. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012;**30**(9):1234–48.
16. Liang C, Huang Y, Lan H, *et al*. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I–II and stage III–IV colorectal cancer. *Oncotarget* 2016;**7**(21):31401–12 https://doi.org/10.18632/oncotarget.8919.
17. Lee SM, Park CM, Goo JM, *et al*. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology* 2013;**268**(1):265–73 https://doi.org/10.1148/radiol.13120949.
18. Chae HD, Park CM, Park SJ, *et al*. Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. *Radiology* 2014;**273**(1):285–93 https://doi.org/10.1148/radiol.14132187.
19. Huang YQ, Liang CH, He L, *et al*. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* 2016 Jun 20;**34**(18):2157–64. https://doi.org/10.1200/JCO.2015.65.9128.
20. Leijenaar RT, Carvalho S, Hoebers FJ, *et al*. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol* 2015;**54**(9):1423–9 https://doi.org/10.3109/0284186X.2015.1061214.
21. Aerts HJ, Velazquez ER, Leijenaar RT, *et al*. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;**5**:4006 http://doi.org//10.1038/ncomms5006.

22. Chen CH, Chang CK, Tu CY, *et al*. Radiomic features analysis in computed tomography images of lung nodule classification. *Plos One* 2018;**13**(2):e0192002 http://doi.org//10.1371/journal.pone.0192002.
23. Wu W, Parmar C, Grossmann P, *et al*. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol* 2016;**6**:71 http://doi.org//10.3389/fonc.2016.00071.
24. Huang Y, Liu Z, He L, *et al*. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* 2016;**281**(3):947 http://doi.org//10.1148/radiol.2016152234.
25. Jin C, Cao J, Cai Y, *et al*. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules. *J Thorac Cardiovasc Surg* 2016;**153**(2):462–9 https://doi.org/10.1016/j.jtcvs.2016.10.019.
26. Eguchi T, Yoshizawa A, Kawakami S, *et al*. Tumour size and computed tomography attenuation of pulmonary pure ground-glass nodules are useful for predicting pathological invasiveness. *Plos One* 2014;**9**(5):e97867 https://doi.org/10.1371/journal.pone.0097867.
27. Cohen JG, Reymond E, Lederlin M, *et al*. Differentiating pre- and minimally invasive from invasive adenocarcinoma using CT-features in persistent pulmonary part-solid nodules in Caucasian patients. *Eur J Radiol* 2015;**84**(4):738–44 https://doi.org/10.1016/j.ejrad.2014.12.031.
28. Penn A, Ma M, Chou BB, *et al*. Inter-reader variability when applying the 2013 Fleischner guidelines for potential solitary subsolid lung nodules. *Acta Radiol* 2014;**56**(10):1180 https://doi.org/10.1177/0284185114551975.
29. Van Riel SJ, Sánchez CI, Bankier AA, *et al*. Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. *Radiology* 2015;**277**(3):863–71 https://doi.org/10.1148/radiol.2015142700.
30. Mcwilliams A, Tammemagi MC, Mayo JR, *et al*. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;**369**(10):910–9 https://doi.org/10.1056/NEJMoa1214726.
31. Shen C, Liu Z, Guan M, *et al*. 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. *Translat Oncol* 2017;**10**(6):886–94 https://doi.org/10.1016/j.tranon.2017.08.007.
32. Lambin P, Leijenaar RTH, Deist TM, *et al*. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;**14**(12):749 https://doi.org/10.1038/nrclinonc.2017.141.