# Development of a diagnostic model for malignant solitary pulmonary nodules based on radiomics features

**Wei Zhao[1#], Chenxi Zou[1#], Chunsun Li[1], Jie Li[2], Zirui Wang[1], Liang'an Chen[1]**

[1]Department of Respiratory and Critical Care Medicine, General Hospital of the People's Liberation Army, Beijing, China; [2]Department of Pathology, General Hospital of the People's Liberation Army, Beijing, China

*Contributions:* (I) Conception and design: W Zhao; (II) Administrative support: L Chen; (III) Provision of study materials or patients: W Zhao, C Zou, J Li; (IV) Collection and assembly of data: W Zhao, C Zou; (V) Data analysis and interpretation: W Zhao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

*Correspondence to:* Liang'an Chen. Department of Respiratory and Critical Care Medicine, General Hospital of the People's Liberation Army, 28 Fuxing Road, Beijing 100853, China. Email: chenliangan301@163.com.

**Background:** This study proposed a precise diagnostic model for malignant solitary pulmonary nodules (SPNs). This model can be used to identify objective and quantifiable image features and guide the clinical treatment strategy adopted for SPNs. This model will help clinicians optimize management strategies for SPN.

**Methods:** In this retrospective study, the clinical data of 455 patients of SPN with defined pathological diagnosis between September 2016 and August 2019 were collected and analyzed. The data included pathological diagnosis, preoperative computed tomography (CT) diagnosis, gender, age, smoking history, family history of tumor, previous history, and contact history data. The quantitative image features and radiomic information of the SPNs were provided using computer-aided detection (CAD) "digital lung" software. The Chi-squared test was used to assess the accuracy between CAD and conventional CT in the diagnosis of SPNs. The diagnostic model for benign or malignant SPNs was developed using a multivariate logistic regression analysis that comprises 6 radiomic factors (irregularity, average diameter, COPD910, proportion of emphysema, proportion of fat, and average density of related blood vessels). The area under the receiver operating characteristic curve was used to evaluate the performance of the model in determining SPN risk of malignancy.

**Results:** There was a statistical difference in the accuracy of CAD and conventional CT in diagnosing SPNs. According to the golden standard pathological diagnosis, the diagnostic accuracy of CAD (81%) was higher than that of conventional CT (63.7%) (P<0.05). Six variables (i.e., irregularity, the mean diameter, COPD910, the proportion of emphysema, the proportion of fat, and the vascular density) were identified using multivariable logistic regression to establish the diagnostic model for distinguish benign or malignant SPNs. The area under the receiver operating characteristic (ROC) curve (AUC) of the diagnostic model was 0.876 (95% CI: 0.8445–0.9076), and its sensitivity and specificity were 81.25% and 82.56% respectively.

**Conclusions:** The proposed diagnostic model, which comprises 6 radiomic factors, is accurate and effective at diagnosing benign or malignant SPNs.

**Keywords:** Solitary pulmonary nodule (SPN); malignant; computer-aided detection (CAD); radiomics; diagnostic model

## Introduction

Lung cancer is the leading cause of cancer-related death worldwide (1), and with its high morbidity and mortality rate, it accounts for 23% of all deaths from malignancies. Despite improvements in the diagnosis and treatment of lung cancer (2,3), the cure rate is only approximately 10% (4). As the symptoms do not always distinctly appear in early stage lung cancer, the optimal time for treatment is often missed (5). A previous study indicated that the overall 5-year survival rate of patients with stage IA of lung cancer is approximately 92% (6). Thus, early diagnosis and therapy for tumors is crucial if the survival rate of patients is to be improved. Lung cancer mainly presents as solitary pulmonary nodules (SPNs) in the early stage. Thus, the accurate diagnosis of SPNs is important to detect lung cancer early.

The clinical methods for evaluating SPN status are based on image characteristics and clinical information, and depend on the judgment of a radiologist. This traditional evaluation method is time consuming and inaccurate. Some classical diagnostic models for tumors based on clinical data, tumor markers, and image parameters, including size, density, growth rate, calcification, and morphology, have been proposed (7). Most lung cancer lesions present as tiny nodules in the early stage, and show no classical characteristics. In recent years, machine-learning relative methods, such as support vector machines (8) and deep learning technology (9), have been used to optimize classical imaging features to improve diagnostic accuracy. Due to the lack of a strong correlation between the imaging characteristics and SPN status, clinical guidelines have not yet integrated computer-based tools to differentiate between benign and malignant lesions. Thus, screening out more objective and quantifiable imaging omics parameters and combining digital transforming technology is essential for SPN prediction.

Computer-aided detection (CAD) technology can detect pulmonary nodules more quicky and more efficiently than imaging professionals (10,11). CAD significantly improves pulmonary nodule detection and has great accuracy (12-14). Target lesions can be identified, separated, and measured using the CAD "digital lung" software, which quickly calculates and quantifies imaging omics parameter data to make a diagnostic judgment. A few classical diagnostic models for SPN have been established based on the image features extracted by professional radiologists (7,15). We propose a novel clinical diagnostic model based on quantified image features using CAD combined with clinical data. This model could guide the clinical strategy adopted for the treatment of SPNs. This model diagnoses whether SPNs are benign or malignant.

We present the following article in accordance with the TRIPOD reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-462/rc).

## Methods

### Participants and ethics

This retrospective study was approved by the Ethics Committee of the General Hospital of the People's Liberation Army (PLAGH) (No. S2016-019-01), and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Individual consent for this retrospective analysis was waived.

A total of 455 patients diagnosed with SPNs between September 2016 and August 2019 were enrolled in this study. To be eligible for inclusion in the study, patients had to meet the following inclusion criteria: (I) have received CT scans within 2 weeks before surgery; (II) have single or multiple primary SPNs <3 cm in diameter with the nodules confirmed to be of the same histological subtype; (III) have had the lesions surgically resected and have pathological confirmation of atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), invasive adenocarcinoma (IAC), or benign lesions; (IV) have complete medical records, including past medical history, smoking history and contact history records; and (V) have no lymph nodes or distant metastases (stage IA).

### CT scan acquisition and CAD data processing

Computed tomography (CT) scanning was performed by 2 radiologists on a helical multidetector scanner (Brilliance iCT 64-slice, Philips Inc., Netherlands) on patients in the supine position. The scanning parameters for all patients were as follows: 120 kV tube voltage, 512×512 image matrix, 5.0 mm thickness, 80 mm collimation, and 0.758 pitch. The images were output in the DICOM format. Quantifiable data in the radiographs in the DICOM format were extracted using a Medical Imaging Diagnostic Workstation based on CAD "digital lung" software (Food and Drug Administration (FDA) approved application number K143586; National Medical Products

**Table 1** CT parameters used for screening for SPNs

| Classification | Parameters |
| --- | --- |
| Size | Volume |
| | Area |
| | SA/V |
| | MaxLumenDiameter |
| | MeanLumenDi |
| | MASS |
| Density | Mean Intensity |
| | Ground Glass |
| | COPD950P |
| | COPD910P |
| | Emphysema% |
| | AgatstonCal |
| | FatRatio |
| | VolumeCal |
| | CavityRatio |
| Irregularity | Irregularity |
| Associated microvessel | VesselTortuosity |
| | VesselNumber |
| | VesselVolume |
| Fractal dimension | FractalDiameter |
| Lung intensity | LungIntensity |
| Pleural ratio | Pleuralratio |

CT, computerized tomography; SPNs, solitary pulmonary nodules; SA/V, surface area/volume.

Administration (NMPA) approved application number 20170024). Data processing was conducted according to the standard operation protocol. The comprehensive feature parameters of the target regions in the lesion were quantified using CAD. The features comprised SPN size, density, morphological, environment, lung tissue density, and fractal dimension. The detailed CT parameters are listed in *Table 1*.

### Statistical analysis

A total of 455 patients were divided into benign group and malignant group. There were 176 cases in the benign group, including 113 cases of benign nodules and 63 cases of AAH. The malignant group included 279 cases, including 52 cases of AIS, 57 cases of MIA, and 170 cases of IA.

A statistical description based on the patients' clinical features, CT imaging features, and radiomics information was conducted using SPSS 24.0, R 3.5.0, and SAS 9.3 software. The specific analytical methods for the different parameters were as follows: normal distribution parameters are presented as mean ± standard deviation, and the differential analysis was achieved through variance; non-normal distribution parameters are presented as median and interquartile range, and were analyzed using a non-parametric test; the qualitative data are presented as percentages using the chi-square test or Fisher's exact test.

In this study, the diagnostic model was built using logistic regression. A univariate binary logistic regression analysis was first used to screen the statistically significant univariable, which were then used in the multivariate binary logistic regression analysis. The strength of the major-related factors was evaluated using the P values, odds ratios (ORs), and 95% confidence intervals (CIs). The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the predictive performance of the model in determining SPN status (benign or malignant). Calibration curves were plotted to assess the model with the Hosmer-Lemeshow test.

### Results

#### Patient characteristics

A total of 455 patients who met the inclusion criteria were recruited and included in the training set. The detailed clinical characteristics of all the patients are listed in *Table 2*. Of the 455 patients, 198 (43.5%) were male and 257 (56.5%) were female. The patients were allocated into 5 groups based on their pathological diagnosis. In the 5 groups of SPNs, there were 63 (13.8%) AAH, 52 (11.4%) AIS, 57 (12.5%) MIA, 170 (37.4%) IA, and 113 (24.8%) BL cases.

#### Comparing CAD "digital lung" auxiliary diagnosis and conventional CT diagnosis

In this study, a data set of 342 (AAH, AIS, MIA, or IA) patients was used to assess the diagnostic accuracy between the CAD "digital lung" and conventional CT before surgery. According to the golden standard pathological diagnosis, the accuracy assessment of the conventional CT is displayed in Table S1; it was 63.7% (n=218). The error rate

Page 4 of 7

Zhao et al. Pulmonary nodules diagnosed by CT quantitative feature model

**Table 2** Characteristics of enrolled patients

| Clinical characteristics | N (%) |
|---|---|
| Pathological diagnosis | |
| AAH | 63 (13.8) |
| AIS | 52 (11.4) |
| MIA | 57 (12.5) |
| IA | 170 (37.4) |
| BL | 113 (24.8) |
| Gender | |
| Male | 198 (43.5) |
| Female | 257 (56.5) |
| Family history | |
| Yes | 352 (77.4) |
| No | 103 (22.6) |
| Smoking history (year) | |
| No | 357 (78.5) |
| Quitting smoking >5 | 9 (2.0) |
| Yes | 89 (19.6) |
| Medical history | |
| No | 255 (56.0) |
| History of malignancy | 35 (7.7) |
| Chronic or benign of lung | 26 (5.7) |
| Benign lung outside | 139 (30.5) |
| Contact history | |
| No | 453 (99.6) |
| Yes | 2 (0.4) |
| Multiple primary tumors | |
| No | 356 (78.2) |
| Yes | 99 (21.8) |
| Left or right | |
| Left | 146 (32.1) |
| Right | 309 (67.9) |
| Lab | |
| Upper left | 85 (18.7) |
| Low left | 59 (13.0) |
| Upper right | 167 (36.7) |
| Right middle lobe | 39 (8.6) |
| Low right | 105 (23.1) |

AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IA, 1A; BL, benign lesion.

was 12.0% (n=41), including 2.6% (n=9) under-diagnosed and 9.4% (n=32) over-diagnosed cases. An indefinite event accounted for 24.3% (n=83) of the cases. For CAD, the diagnostic accuracy and indefinite event comprised 81.0% (n=277) and 15.5% (n=53) of the cases, respectively. The error rate was 12.0% (n=12), including 0.6% (n=2) under-diagnosed and 2.9% (n=10) over-diagnosed cases. The indefinite events accounted for 24.3% (n=83) of the cases.

The diagnostic accuracy was assessed in relation to the infiltration degree and different maximum diameter sizes using the 2 strategies before surgery (Tables S2-S5). There was no significant difference (P=0.33) in the degree of infiltration (Table S6). However, the CAD "digital lung" was more accurate than the conventional CT at making different maximum diameter size assessments (Table S7) (P=0.00).

### Correlation analysis between the quantitative images features and malignancy of SPNs

The training patients were divided into 2 groups, benign (BL or AAH) and malignant (AIS, MIA, or IA). A total of 31 clinical characters were identified and statistically compared to analyze their correlation with SPN malignancy; 19 clinical characters, including gender, age, family history, medical history, whether the primary tumor was malignant, volume, surface area, average diameter, the longest diameter, MASS, irregular, COPD950, COPD910, emphysema proportion, irregularity, related blood count, blood vessels, the average density, fat percentage, and fractal dimension, were statistically different (P<0.05; Table S4). The other 12 factors (i.e., smoking history, right and left lung, pulmonary lobe location, special contact history, average density, ground glass proportion, calcification score, calcification volume, related vascular tortuosity, related vascular volume, cavity proportion, and pleural proportion, were not significantly correlated (P>0.05; data not shown).

### Multivariate binary logistic regression analysis

The 19 screened characteristics were evaluated using a regression analysis. The selected 6 characteristics (*Table 3*) were examined in relation to whether the SPN was malignant or not [P<0.05; that is, X8 (irregularity), X10 (average diameter), X13 (COPD910), X14 (proportion of emphysema), X17 (proportion of fat), and X19 (average density of related blood vessels)].

**Table 3** Predictors of malignant SPNs

| Variables | β | S.E | z value | Pr (>|z|) | OR | 95% CI |
|---|---|---|---|---|---|---|
| (Intercept) | −5.778 | 0.806 | −7.167 | 7.67E−13 | 0.003 | 0.001–0.015 |
| X8 | 0.402 | 0.061 | 6.545 | 5.95E−11 | 1.495 | 1.326–1.687 |
| X10 | 1.099 | 0.204 | 5.392 | 6.96E−08 | 3.001 | 2.013–4.474 |
| X13 | 0.076 | 0.009 | 8.055 | 7.95E−16 | 1.079 | 1.059–1.099 |
| X14 | −0.396 | 0.082 | −4.818 | 1.45E−06 | 0.673 | 0.573–0.791 |
| X17 | 0.043 | 0.022 | 1.928 | 0.053912 | 1.043 | 0.999–1.090 |
| X19 | −0.002 | 0.001 | −2.973 | 0.002944 | 0.998 | 0.997–0.999 |

SPNs, solitary pulmonary nodules; X8, irregularity; X10, average diameter; X13, COPD910; X14, proportion of emphysema; X17, proportion of fat; X19, average density of related blood vessels.
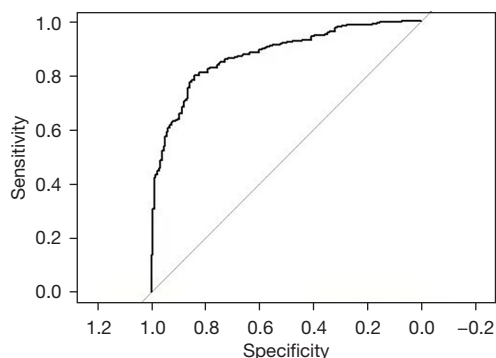


**Figure 1** ROC curve for the diagnostic model. ROC, receiver operator characteristic.

### The establishment of a prediction model for SPN

Six clinical features were used to develop the predictive model. The predictive model was established based on the coefficients and ORs of the variables. The following formula was used for which the value of P represents the probability of a malignant SPN:

$$P = \frac{\exp\left(-5.778 + 0.402 \times (X8) + 1.099 \times (X10) + 0.076 \times (X13) - 0.396 \times (X14) + 0.043 \times (X17) - 0.002 \times (X19)\right)}{1 + \exp\left(-5.778 + 0.402 \times (X8) + 1.099 \times (X10) + 0.076 \times (X13) - 0.396 \times (X14) + 0.043 \times (X17) - 0.002 \times (X19)\right)} \quad [1]$$

where X8 refers to MASS, X10 to mean diameter, X13 to COPD910P, X14 to emphysema% and X17 to proportion of fat.

### The evaluation of the diagnostic model

The ROC for the model was presented and the AUC was 0.876 (95% CI: 0.8445–0.9076), which suggests that the research model performed well in distinguishing between benign and malignant SPNs (*Figure 1*). The Hosmer-Lemeshow test was used to evaluate the calibration. There was no statistically significant difference between the model-predicted and actual-observed values ($\chi^2$=7.1314, P=0.5225), indicating its excellent predictive ability. A nomogram was simultaneously established based on the multivariable logistic analysis for which the derivation cohort was used as a quantitative tool (*Figure 2*).

### Discussion

A clinical diagnostic model for predicting benign or malignant SPNs was developed successfully. This model could be used to guide the treatment of SPNs or other solid tumors. The diagnostic accuracy between CAD "digital lung" and conventional CT before surgery was assessed using a data set comprising 342 (AAH, AIS, MIA, or IA) patients. The diagnostic accuracy of CAD was higher than that of conventional CT in the pathological subgroups. The under-diagnosed and over-diagnosed events were found to be fewer using CAD. Consistent with the findings of our previous studies (10,12), CAD was faster and more efficient at accurately diagnosing SPNs than professional radiologists.

The CAD "digital lung" was used to identify and isolate lesions and segment the quantitative parameters for visual features. The univariate binary logistic regression analysis showed that radiomic features, such as MASS, SA/V, COPD950, COPD910, the proportion of emphysema, irregularity, the number of related vessels, vascular density, the proportion of fat and fractal dimension, were significantly correlated to SPN malignancy. We also verified that some classic parameters (i.e., gender, age, family history of malignancy, previous medical history, multiple primary, volume, surface area, average diameter, and
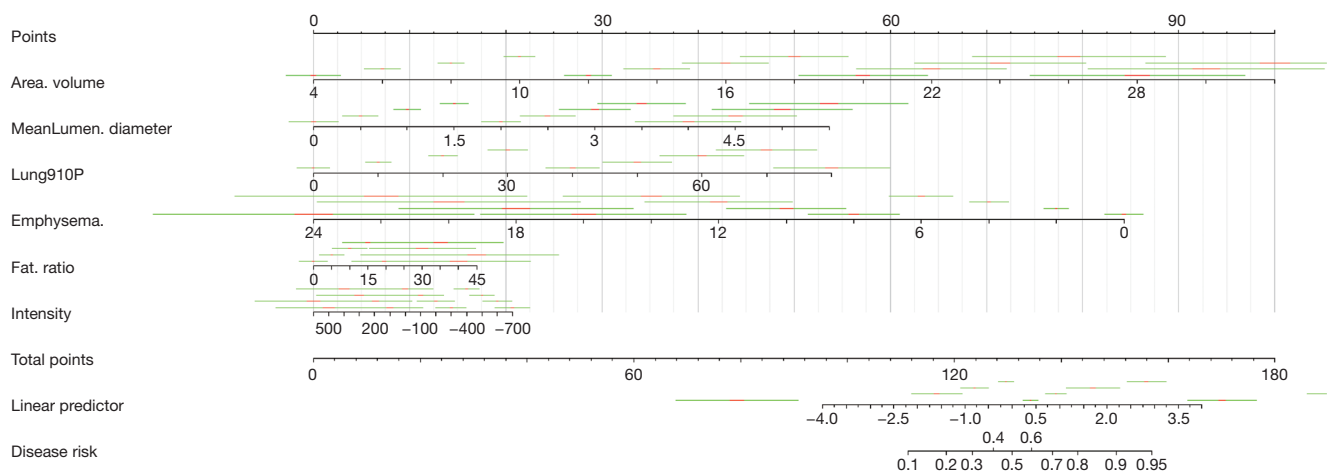
Page 6 of 7

Zhao et al. Pulmonary nodules diagnosed by CT quantitative feature model

**Figure 2** Nomogram for predicting benign or malignant SPNs. SPNs, solitary pulmonary nodules.

maximum diameter) can be used to determine malignancy. These feature parameters combined with other signatures (i.e., size, density, morphology, and the surrounding environment) were analyzed using multivariate logistic regression analysis to assess the nature of SPNs. Similarly, SA/V, COPD910, the proportion of fat, mean diameter, and vascular density were found to be risk factors for malignancy. However, the proportion of emphysema was a protective factor. The results indicate that the multivariate logistic regression analysis was more comprehensive than univariate binary logistic regression analysis for SPN analyses. Finally, 6 variable predictors, including MASS, mean diameters, COPD910, the proportion of emphysema, and the proportion of fat, were used in a multivariate binary logistic regression analysis to derive a prediction model for malignancy. Our ROC analyses illustrate that the mode (AUC of 0.876) had a powerful predictive performance. The Hosmer-Lemeshow test showed that the difference between the model-predicted and actual-observed values was not statistically significant ($\chi^2$=7.1314, P=0.5225), indicating its excellent predictive ability.

Classical models, such as the Mayo (7) and VA models (15) were used to diagnose the probability of SPN malignancy. In these models, multivariate statistical methods are based on a large amount of radiographic data and clinical information. The Mayo model was established using data for 3 clinical characteristics (i.e., age, cigarette-smoking history, and history of cancer) and 3 radiological features (i.e., spiculation, upper lobe location, and diameter) from 629 patients identified by CT; 65% of the patients had benign lesions while 35% had malignant lesions. The

AUC of the model was 0.876. Thus, this model was more accurate in the diagnosis of benign lesions, but it showed a low diagnostic rate for early or low-grade malignant lung cancer. Conversely, the model in this study showed that 38.6% of the patients had benign lesions (AHH or BL), and 61.4% had malignant lesions (AIS, MIA, or IA). The malignant cases in our model were all in the pIA stage of lung cancer; thus, the model was effective at detecting and diagnosing SPN. The Veterans Affairs (VA) model (15) was based on the following 4 factors from 532 patients: nodule diameter, smoking history, smoking cessation time, and age. Malignant cases were confirmed through pathology while benign cases with stable lesions were confirmed in the 2-year follow-up period. The AUC was 0.79. Conversely, our model was developed based on quantitative image features. Unlike classical models that need observers to judge image features, the model imports data from CAD into its formula. The model has the advantage of eliminating subjective bias for image feature acquisition.

There were some limitations to this research. First, the sample size should be expanded for model verification and to compare the diagnostic ability with that of the classical model. Second, non-adenocarcinoma malignant tumors should be included to avoid the unilateralism of this model.

In conclusion, the proposed predictive model, which comprises 6 radiomics factors, is accurate and effective at diagnosing benign or malignant SPNs.

## Acknowledgments

and development program (No. 2019YFC0120504).

## Footnote

*Reporting Checklist: The authors have completed the TRIPOD reporting checklist.* Available at https://atm.amegroups.com/article/view/10.21037/atm-22-462/rc

*Data Sharing Statement:* Available at https://atm.amegroups.com/article/view/10.21037/atm-22-462/dss

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-462/coif). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was approved by the Ethics Committee of the General Hospital of the People's Liberation Army (PLAGH) (No. S2016-019-01), and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Individual consent for this retrospective analysis was waived.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.
2. Brahmer JR, Govindan R, Anders RA, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of non-small cell lung cancer (NSCLC). J Immunother Cancer 2018;6:75.
3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature 2018;553:446-54.
4. Detterbeck FC, Boffa DJ, Kim AW, et al. The Eighth Edition Lung Cancer Stage Classification. Chest 2017;151:193-203.
5. National Lung Screening Trial Research Team; Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395-409.
6. Sozzi G, Boeri M. Potential biomarkers for lung cancer screening. Transl Lung Cancer Res 2014;3:139-48.
7. Swensen SJ, Silverstein MD, Ilstrup DM, et al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. Arch Intern Med 1997;157:849-55.
8. Hawkins S, Wang H, Liu Y, et al. Predicting Malignant Nodules from Screening CT Scans. J Thorac Oncol 2016;11:2120-8.
9. Yang W, Sun W, Li Q, et al. Diagnostic Accuracy of CT-Guided Transthoracic Needle Biopsy for Solitary Pulmonary Nodules. PLoS One 2015;10:e0131373.
10. Das M, Mühlenbruch G, Mahnken AH, et al. Small pulmonary nodules: effect of two computer-aided detection systems on radiologist performance. Radiology 2006;241:564-71.
11. Ley S, Ley-Zaporozhan J. Novelties in imaging in pulmonary fibrosis and nodules. A narrative review. Pulmonology 2020;26:39-44.
12. Thawani R, McLane M, Beig N, et al. Radiomics and radiogenomics in lung cancer: A review for the clinician. Lung Cancer 2018;115:34-41.
13. Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res 2017;6:86-91.
14. Christe A, Leidolt L, Huber A, et al. Lung cancer screening with CT: evaluation of radiologists and different computer assisted detection software (CAD) as first and second readers for lung nodule detection at different dose levels. Eur J Radiol 2013;82:e873-8.
15. Gould MK, Ananth L, Barnett PG, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest 2007;131:383-8.

(English Language Editor: L. Huleatt)

**Table S1** The diagnostic accuracy assessment of the conventional CT and CAD

| Examination | Indefinite | Accuracy | Diagnostic error | | $\chi^2$ | P | Total |
| | | | Under-diagnosis | Over-diagnosis | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Conventional CT | 83 (24.3) | 218 (63.7) | 9 (2.6) | 32 (9.4) | 29.63 | 0.000 | 342 |
| CAD | 53 (15.5) | 277 (81.0) | 2 (0.6) | 10 (2.9) | | | |

CT, computerized tomography; CAD, computer-aided diagnosis.

**Table S2** The diagnostic accuracy of conventional chest CT with different infiltration

| SPN | Indefinite | Accuracy | Diagnostic error | | Total | Fisher | P |
| | | | Under-diagnosis | Over-diagnosis | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AAH | 32 (50.8) | 9 (14.3) | 9 (14.3) | 13 (20.6) | 63 | 178.85 | 0.000 |
| AIS | 25 (48.1) | 16 (30.8) | 0 (0.0) | 11 (21.2) | 52 | | |
| MIA | 18 (31.6) | 35 (61.4) | 0 (0.0) | 4 (7.0) | 57 | | |
| IA | 8 (4.7) | 158 (92.9) | 0 (0.0) | 4 (2.4) | 170 | | |
| Total | 83 (24.3) | 218 (63.7) | 9 (2.6) | 32 (9.4) | 342 | | |

CT, computerized tomography; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma.

**Table S3** The diagnostic accuracy of CAD with different infiltration

| SPN | Indefinite | Accuracy | Diagnostic error | | Total | Fisher | P |
| | | | Under-diagnosis | Over-diagnosis | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AAH | 30 (47.6) | 28 (44.4) | 2 (3.2) | 3 (4.8) | 63 | 103.73 | 0.000 |
| AIS | 15 (28.8) | 33 (63.5) | 0 (0.0) | 4 (7.7) | 52 | | |
| MIA | 6 (10.5) | 49 (86.0) | 0 (0.0) | 2 (3.5) | 57 | | |
| IA | 2 (1.2) | 167 (98.2) | 0 (0.0) | 1 (0.6) | 170 | | |
| Total | 53 (15.5) | 277 (81.0) | 2 (0.6) | 10 (2.9) | 342 | | |

CAD, computer-aided diagnosis; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma.

**Table S4** The diagnostic accuracy of conventional chest CT with different maximum diameter

| Maximum diameter | Indefinite, n (%) | Accuracy, n (%) | Diagnostic error | Total | $\chi^2$ | P |
| --- | --- | --- | --- | --- | --- | --- |
| d ≤10 mm | 46 (41.1) | 44 (39.3) | 22 (19.6) | 112 | 56.562 | 0.000 |
| 10 mm< d ≤20 mm | 27 (22.9) | 76 (64.4) | 15 (12.7) | 118 | | |
| d >20 mm | 10 (8.9) | 98 (87.5) | 4 (3.6) | 112 | | |
| Total | 83 | 218 | 41 | 342 | | |

CT, computerized tomography.

**Table S5** The diagnostic accuracy of CAD with different maximum diameter

| Maximum diameter | Indefinite | Accurate | Diagnostic error | Total | Fisher | P |
|---|---|---|---|---|---|---|
| d ≤10 mm | 33 (29.5) | 73 (65.2) | 6 (5.4) | 112 | 51.524 | 0.000 |
| 10 mm< d ≤20 mm | 20 (16.9) | 94 (79.7) | 4 (3.4) | 118 | | |
| d >20 mm | 0 | 110 (98.2) | 2 (1.8) | 112 | | |
| Total | 53 | 277 | 12 | | | |

CAD, computer-aided diagnosis.

**Table S6** The diagnostic accuracy assessment of the conventional CT and CAD with different infiltration

| SPN | conventional CT (%) | CAD (%) | Fisher | P |
|---|---|---|---|---|
| AAH | 14.3 | 44.4 | 9.09 | 0.33 |
| AIS | 30.8 | 63.5 | | |
| MIA | 61.4 | 86.0 | | |
| IA | 92.9 | 98.2 | | |

CT, computerized tomography; CAD, computer-aided diagnosis; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma.

**Table S7** The diagnostic accuracy assessment of the conventional CT and CAD with different diameter

| Maximum diameter | Conventional CT | | | CAD | | | $\chi^2$ | P |
|---|---|---|---|---|---|---|---|---|
| | Indefinite | Accurate | Diagnostic error | Indefinite | Accurate | Diagnostic error | | |
| d ≤10 mm | 46 (41.1) | 44 (39.3) | 22 (19.6) | 33 (29.5) | 73 (65.2) | 6 (5.4) | 18.47 | 0.000 |
| 10 mm< d ≤20 mm | 27 (22.9) | 76 (64.4) | 15 (12.7) | 20 (16.9) | 94 (79.7) | 4 (3.4) | 9.317 | 0.009 |
| d >20 mm | 10 (8.9) | 98 (87.5) | 4 (3.6) | 0 (0.0) | 110 (98.2) | 2 (1.8) | 12.457 | 0.001 |

CT, computerized tomography; CAD, computer-aided diagnosis.