



Computed-tomography-based radiomic nomogram for predicting the risk of indeterminate small (5–20 mm) solid pulmonary nodules

Chun-Ran Zhang*

Qi Wang*

Hui Feng

Yu-Zhi Cui

Xiao-Bo Yu

Gao-Feng Shi

*Contributed equally to this work.

PURPOSE

This study aims to develop a diagnostic model that combines computed tomography (CT) images and radiomic features to differentiate indeterminate small (5–20 mm) solid pulmonary nodules (SSPNs).

METHODS

This study retrospectively enrolled 413 patients who had had SSPNs surgically removed and histologically confirmed between 2017 and 2019. The SSPNs included solid malignant pulmonary nodules (n = 210) and benign pulmonary nodules (n = 203). The least absolute shrinkage and selection operator was used for radiomic feature selection, and random forest algorithms were used for radiomic model construction. The clinical model and nomogram were established using univariate and multivariable logistic regression analyses combined with clinical symptoms, subjective CT findings, and radiomic features. The area under the curve (AUC) of the receiver operating characteristic curve was used to evaluate the performance of the models.

RESULTS

The AUC for the clinical model was 0.77 in the training cohort [n = 289; 95% confidence interval (CI): 0.71–0.82; P = 0.001] and 0.75 in the validation cohort (n = 124; 95% CI: 0.66–0.83; P = 0.016). The AUCs for the nomogram were 0.92 (95% CI: 0.89–0.95; P < 0.001) and 0.85 (95% CI: 0.78–0.91; P < 0.001), respectively. The radiomic score (Rad-score), sex, pleural indentation, and age were the independent predictors that were used to build the nomogram.

CONCLUSION

The radiomic nomogram derived from clinical features, subjective CT signs, and the Rad-score can potentially identify the risk of indeterminate SSPNs and aid in the patient's preoperative diagnosis.

KEYWORDS

Computed tomography, solid pulmonary nodules, lung cancer, nomogram, radiomics

From the Department of Radiology (C-R.Z., Q.W., H.F., Y-Z.C., G-F.S. ✉ gaofengs62@sina.com), Hebei Medical University Fourth Affiliated Hospital and Hebei Provincial Tumor Hospital, Shijiazhuang, China; Siemens Healthineers Ltd. (X-B.Y.), Beijing, China.

Received 30 April 2021; revision requested 02 June 2021; last revision received 30 March 2022; accepted 30 April 2022.



Epub: 27.01.2023

Publication date: 29.03.2023

DOI: 10.4274/dir.2022.22395

With the wide application of computed tomography (CT) in pulmonary nodule screening and the improvement of public health awareness, an increasing number of small solid pulmonary nodules (SSPNs) are being detected.^{1,2} According to the Fleischner guidelines, the nodules are categorized into solid and subsolid nodules depending on their density. Solid nodules are dense enough to mask the traveling blood vessels and bronchial shadows, while subsolid nodules refer to the ground glass density containing the lesion density that is not enough to mask the traveling blood vessels and bronchial shadows. Nodules with different densities have different malignant probabilities. A survey found that the malignant probability of small nodules (SNs) in patients undergoing surgical resection ranged from 5% to 70%.^{3,4} The classic definition of indeterminate solitary pulmonary nodules (which may be malignant) is pulmonary nodules that do not meet the quintessential benign radiological criteria.⁵ Currently, the incidence of lung carcinoma both in China and globally remains high.^{6,7} In the early stage, lung carcinoma is often found in the form of pulmonary

nodules. Assessing the nature of the nodules is an essential step in making SN management decisions. One study showed that solid nodules smaller than 5 mm had a malignancy rate of 0.4%, solid nodules larger than 20 mm had a malignancy rate of 31%,⁸ and solid nodules between 5 and 20 mm were difficult to diagnose.⁹

A computer-aided diagnosis is increasingly used in the diagnosis of pulmonary nodules. Some people build clinical models to diagnose hamartoma and adenocarcinoma, and some use radiomics to distinguish between benign and malignant pulmonary nodules.^{10,11} The latest research techniques, such as the use of radiomic features and the attached vessel tortuosity, are also used to distinguish between lung adenocarcinoma and granuloma. We developed a nomogram to predict the risk probability of indeterminate solid lung nodules, which combined clinical data, subjective CT signs, and radiomics, and included nodules ranging from 5 to 20 mm.

Methods

Patient selection

Our retrospective study was approved (approval no: 2020KY082) by the Hospital Institutional Review Committee, and the requirement for informed consent was waived. Two radiologists (with 3 and 10 years of working experience, respectively) who were blinded to the final pathological diagnosis and clinical data, independently reviewed the CT images of all patients with SSPNs between January 2017 and December 2019. Interpretation discrepancy, if any, was resolved by consensus. The admission criteria for patients were as follows: (1) SSPNs 5–20 mm in diameter; (2) aged older than 15 years; (3) no history of malignant tumors in the past 5 years; (4) confirmation by surgical or CT-guided biopsy pathology; (5) no radiotherapy or chemotherapy before the exam-

ination; (6) preoperative chest thin-layer CT image (≤ 1.25 mm); and (7) interval between chest CT scan and surgery of less than one month. The exclusion criteria were as follows: (1) SPNs (non-solid and part-solid); (2) impacts of diffuse pulmonary disease on imaging diagnosis; (3) lesions accompanied by a cavity; and (4) pathologically confirmed metastatic carcinoma.

In total, 413 participants were enrolled in this study (58.0 ± 10.5 years old), including 199 (48.18%) women and 214 (51.82%) men. The prevalence of malignant SSPNs was 50.85%. The most malignant SSPN was lung adenocarcinoma with 196 (93.33%) cases, while the others included 8 (3.81%) squamous cell carcinomas, 5 (2.38%) adenosquamous carcinomas, and 1 (0.48%) neuroendocrine carcinoma. Figure 1 shows a flowchart for nodule screening.

Chest CT scanning technology and image analysis

A thoracic spiral CT was performed from the apex pulmonis to the costophrenic angle using the second-generation Somatom

Definition Flash CT scanner of Siemens or a GE Revolution Spiral system (GE Healthcare). The enhanced scan was performed using a high-pressure syringe, injecting non-ionic iodine (iohexol; 350 mg/mL; injection amount, 1.5–2 mL/kg; injection rate, 3 mL/s) intravenously through the elbow. The arterial phase scan was performed 25 s after injection of the contrast agent. The acquisition parameters were set as follows: tube voltage, 120 kV; tube current, 80–300 mAs; pitch, 0.2; and scanning matrix, 512×512 . The scanning layer thickness was 5.0 mm, and the reconstruction layer thickness of the standard algorithm was 1.0–1.25 mm. The mediastinum window was set [width, 350 Hounsfield unit (HU); level, 40 HU], and the lung window was also set (width, 1.200 HU; level, –600 HU). The picture archiving and communication system was used to store the images and export them in the Digital Imaging and Communications in Medicine format.

Two radiologists analyzed the characteristics of each SSPN while blinded to the patients' pathological results, including their age, sex, smoking history, clinical symptoms,

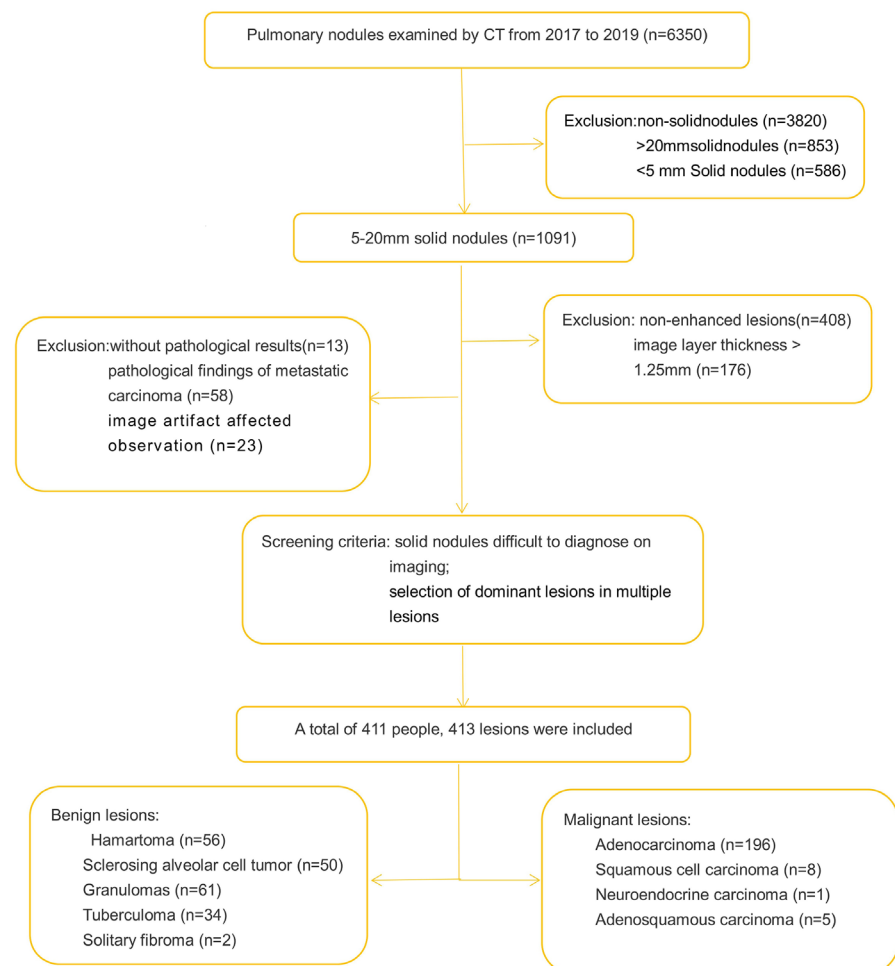


Figure 1. Nodule screening flow chart. CT, computed tomography.

Main points

- Radiomics has great advantages in assessing the risk of indeterminate small solid pulmonary nodules (SSPNs).
- The radiomic nomogram derived from clinical features, subjective computed tomography signs, and the radiomic score (Rad-score) is superior to the clinical model in evaluating the risk of indeterminate SSPNs.
- The Rad-score, sex, age, and pleural indentation are independent predictors in assessing the risk of indeterminate SSPNs.

respiratory disease history, family history of lung cancer, and extrathoracic malignancy history (>5 years ago). The image-based features included the size, location (whether located in the upper lobe), density, shape, margin (regular or irregular), marginal spiculation, pleural indentation, pulmonary nodules significantly enhanced [yes (>15 HU) or no (\leq 15 HU)], and emphysema. Nodule size was defined as the mean of the longest diameter and vertical diameter of the largest plane of the nodule on the axial CT image. The regular margin was determined to have circular or elliptical smooth shapes. The marginal spiculation sign was a straight-line shadow from the margins of the lesion to the surroundings, which were radial and unbranched. Pleural indentation, also known as pleural traction, was due to the formation of fibrosis in the tumor, which pulls the visceral pleura toward the tumor. The enhancement index was measured as the difference between the enhancement scan and plain scan at the CT central level of nodules. If the diameter of the nodule was less than 1 cm, the two radiologists measured the CT value independently three times and took the average value.

Nodule segmentation and feature extraction

Dedicated radiomic software (Radiomics, version 1.0.3, Siemens, Forchheim, Germany) was used for semiautomatic lesion segmentation. The accuracy of the segmentation was confirmed layer by layer from the axial, coronal, and sagittal positions (Figure 2). Pleural indentation, the long cords around the lesion, and the blood vessels and trachea at the edge of the lesion were not delineated, and any non-conforming layers were manually erased or filled.

Feature extraction was performed for each lung nodule, and 1,691 features were computed, including first-order statistical features, shape-based features, and texture features based on the gray-level co-occurrence matrix and gray-level size zone matrix. The interobserver reproducibility of segmentation was evaluated using 50 randomly selected cases that were re-segmented by the same radiologist following the same delineation principle one week later.

Data Analysis

Radiomic feature selection and radiomic feature model construction

An intraclass correlation coefficient (ICC) threshold of 0.8 was used to evaluate the

reproducibility of the radiomic features. The ICC of a radiomic feature of >0.8 was considered stable and selected for subsequent analysis. Next, the least absolute shrinkage and selection operator (LASSO) method was used to further select the features using the five-fold cross-validation method. The prediction model was constructed using the random forest (RF) method, which is a categorization method that involves multiple nodes of a decision tree and bootstrap resampling technology. Each tree uses a stochastically selected input variable or each node variable combination to form a forest.^{12,13} The RF has the potential to overcome the overfitting problem with high accuracy and a robust anti-interference ability. Each nodule's characteristic radiomic score (Rad-score) was calculated using the constructed model. The radiomic model was developed using the Python Scikit-learn package (Python 3.6; Scikit-learn version 0.24.0; <http://scikit-learn.org/>).

Construction and evaluation of the nomogram

The Rad-scores, clinical features, and subjective CT signs were analyzed in the training cohort, and the nomogram was established with univariate and multivariable logistic regressions. The backward stepwise multivariable logistic regression was utilized to obtain similar results with fewer variables. The area under the curve (AUC) of each prediction model was estimated using bootstrapping (1,000 times) based on the prediction score.

Clinical value of the radiological nomogram

A decision curve analysis (DCA) was performed to evaluate the nomogram's clinical efficacy, which was calculated as a series benefit of the model under a series of threshold probabilities. Figure 3 shows the flow chart of the data process.

Statistical analysis

In this study, continuous variables conforming to a normal distribution were expressed as the mean \pm standard deviation; otherwise, they were expressed as the median (the first quantile; the third quantile). Categorical variables were presented as frequencies with percentages. The Wilcoxon rank-sum test was used for continuous variables, and χ^2 or Fisher's exact test was used for categorical variables. The performance of the models in the training and validation cohorts was quantified by the receiver operating characteristic (ROC) analysis of the AUC, accuracy, sensitivity, specificity, positive predictive value, and negative predic-

tive value. A 95% confidence interval (CI) for each feature was also recorded. The calibration curve was used to evaluate the consistency between the observed outcomes and predicted results, and the Hosmer-Lemeshow test was performed to evaluate the fit. The AUC comparison was examined using the DeLong test. We evaluated the performance of the model with the Nagelkerke R-squared and omnibus tests. Statistical analyses in this study were performed using R software (version 4.0.3; R Foundation for Statistical Computing, Vienna, Austria, <http://www.Rproject.org>) and SPSS 24.0 software (SPSS Inc., Chicago, IL, USA). The R package information is shown in Appendix 1. The reported statistical significance levels were both two-sided with $P < 0.050$ but $P < 0.100$ in the univariate logistic regression analysis.

Results

Clinical feature analysis and clinical model establishment

The patients were randomized in a 7:3 ratio into the training and validation cohorts. The patient's clinical characteristics are described in Table 1. In the training cohort, potential predictors were determined using a univariate logistic regression analysis and incorporated into a multivariable logistic regression analysis. Sex [odds ratio (OR): 0.46; 95% CI: 0.27–0.81; $P = 0.007$], morphology (OR: 2.07; 95% CI: 1.23–3.49; $P = 0.006$), age (OR: 1.04; 95% CI: 1.01–1.07; $P = 0.004$), pleural indentation (OR: 3.16; 95% CI: 1.86–5.39; $P < 0.001$), emphysema (OR: 2.63; 95% CI: 1.07–6.48; $P = 0.036$), and significant lung nodule enhancement (OR: 2.11; 95% CI: 1.25–3.57; $P = 0.005$) were malignancy-independent predictors, and a clinical model was constructed using these six independent predictors (Table 2). The AUCs for the clinical model in the training and validation cohorts were 0.77 (95% CI: 0.71–0.82; $P < 0.001$) and 0.75 (95% CI: 0.66–0.83; $P < 0.001$), respectively.

Establishment and verification of the radiomic model

In total, 1,691 radiological features were abstracted from each patient's lung image. The radiological features of 29 non-zero coefficients were selected using LASSO regression in the training cohort and included in the Rad-score calculation. Figure 4 shows the distribution of the Rad-scores of benign and malignant SSPNs in the training and validation cohorts. No significant differences were

found in the Rad-scores between the benign ($P = 0.448$) and validation ($P = 0.055$) cohorts, and malignant nodule groups in the training but the Rad-scores of the malignant nodule

group [0.72 (0.62; 0.78) vs. 0.64 (0.57; 0.74); $P < 0.001$] and benign nodule group [0.34 (0.23; 0.51) vs. 0.38 (0.25; 0.53); $P < 0.001$] in the training and validation cohorts were significantly different.

Malignant pulmonary nodules typically have higher Rad-scores than benign pulmonary nodules. The AUC values of the radiomic model in the training and validation groups were 0.90 (95% CI: 0.86–0.93) and 0.83 (95% CI: 0.76–0.91), respectively, which showed good performance in predicting SSPNs.

Establishment and verification of the nomogram

The multivariable logistic regression analysis showed that the Rad-score, sex (OR: 0.23; 95% CI: 0.10–0.49; $P < 0.001$), age (OR: 1.05; 95% CI: 1.02–1.09; $P = 0.004$), and pleural indentation (OR: 2.42; 95% CI: 1.20–4.87; $P = 0.013$) were independent predictors. By integrating these four independent factors, a combined model was constructed and presented in the form of a nomogram (Figure 5). The AUC of the nomogram was 0.92 (95% CI: 0.89–0.95); $P < 0.001$) in the training cohort and 0.85 (95% CI: 0.78–0.91); $P < 0.001$) in the validation cohort. The calibration curve showed that the predicted results of the nomogram were in good agreement with the actual results in the training and verification queues. The Hosmer–Lemeshow test yielded a non-significant statistic ($P = 0.584$ and $P = 0.716$ for the training and validation cohorts, respectively), which suggests a good fit for probability.

Model performance comparison

Table 3 shows the clinical model and nomogram's diagnostic performance, and Figures 6 and 7 show the ROC curves of these models. The cut-off values for the clinical model and nomogram AUCs were 0.49 and 0.56, respectively. Using the omnibus test with the model coefficients, the nomogram was significantly improved compared to the clinical model ($P < 0.001$). Based on the DeLong test, statistically significant differences were found between the clinical model and nomogram in predicting the risk of SSPNs (Figure 8). The DCA curve (Figure 9) shows that the nomogram increases the net benefit more than the clinical model in distinguishing indeterminate SSPNs.

Discussion

In the present retrospective study, we developed and verified a nomogram combining routine clinical features, subjective

Table 1. Clinical characteristics and subjective CT findings of the SSPN patients

	Test	Train	P value
	n = 124	n = 289	
Sex			0.485
Female	63 (50.81%)	136 (47.06%)	
Male	61 (49.19%)	153 (52.94%)	
Age	57.50 (52.00; 64.00)	60.00 (52.00; 66.00)	0.157
Location			0.230
Upper	65 (52.42%)	133 (46.02%)	
Middle and lower	59 (47.58%)	156 (53.98%)	
Average diameter	1.25 (1.04; 1.55)	1.25 (0.98; 1.53)	0.280
Density			0.316
Uniform	62 (50.00%)	129 (44.64%)	
Non-uniform	62 (50.00%)	160 (55.36%)	
Morphology			0.194
Regular	67 (54.03%)	136 (47.06%)	
Irregular	57 (45.97%)	153 (52.94%)	
Boundary			0.119
Clear	94 (75.81%)	197 (68.17%)	
Unclear	30 (24.19%)	92 (31.83%)	
Spicule sign			0.990
No	76 (61.29%)	177 (61.25%)	
Yes	48 (38.71%)	112 (38.75%)	
Pleural indentation			0.370
No	64 (51.61%)	163 (56.40%)	
Yes	60 (48.39%)	126 (43.60%)	
Emphysema			0.844
No	109 (87.90%)	256 (88.58%)	
Yes	15 (12.10%)	33 (11.42%)	
Significant lung nodule enhancement			0.459
No	68 (54.84%)	147 (50.87%)	
Yes	56 (45.16%)	142 (49.13%)	
Clinical symptoms			0.049
No	101 (81.45%)	209 (72.32%)	
Yes	23 (18.55%)	80 (27.68%)	
Smoking history			0.882
No	81 (65.32%)	191 (66.09%)	
Yes	43 (34.68%)	98 (33.91%)	
Respiratory disease history			1.000
No	121 (97.58%)	282 (97.58%)	
Yes	3 (2.42%)	7 (2.42%)	
Lung cancer history			0.229
No	108 (87.10%)	263 (91.00%)	
Yes	16 (12.90%)	26 (9.00%)	
Extrathoracic malignancy history			0.850
No	117 (94.35%)	274 (94.81%)	
Yes	7 (5.65%)	15 (5.19%)	

CT, computed tomography; SSPN, small solid pulmonary nodule.

findings from CT images, and radiomic features to distinguish indeterminate SSPNs. Our results suggest that the performance of a nomogram is superior to the clinical model in predicting the risk of indeterminate SSPNs.

The DCA curve demonstrates the clinical usefulness of a nomogram.

Radiomics is a rapidly developing field that can extract thousands of quantitative image features from images and describe

the biological characteristics and heterogeneity of lesions by analyzing these features. Radiomics identifies information from conventional images that are not visible to the naked eye or limited by the size or shape of the lesion.¹⁴⁻¹⁶ In recent years, using a radiomic texture feature analysis for lung nodule assessments has received increased attention. Most studies have focused on general small pulmonary nodules (including solid and subsolid nodules) or used radiomics and clinical features to identify the nature of pulmonary nodules,¹⁷⁻¹⁹ for which CT signs have not been included. We constructed an integrated model combining clinical features, subjective CT signs, and the Rad-scores for indeterminate SSPNs (5–20 mm). SNs require stricter research standards and are more challenging to diagnose through imaging.

Our study revealed that sex, age, shape, pleural indentation, emphysema, and enhancement are independent predictors of indeterminate SSPNs. This finding is consistent with the findings of previous studies.²⁰⁻²⁴ However, we also found that in most previous studies, the position of the upper lobe and smoking history were independent predictors of the malignancy of pulmonary nodules. This finding differs from those in previous studies. First, the discrepancy may be caused by regional differences. The incidence of tuberculosis in China is very high and mainly affects the upper lobe. Additionally, according to our previous statistics on the risk factors for lung cancer screening in the Hebei province, smoking history is not an independent predictor. This may be due to air pollution, which has caused the incidence of nodules in non-smokers to increase significantly. Additionally, malignant pulmonary nodules may be mostly adenocarcinoma, which has a higher incidence in women than in men, and smoking is rare in women.

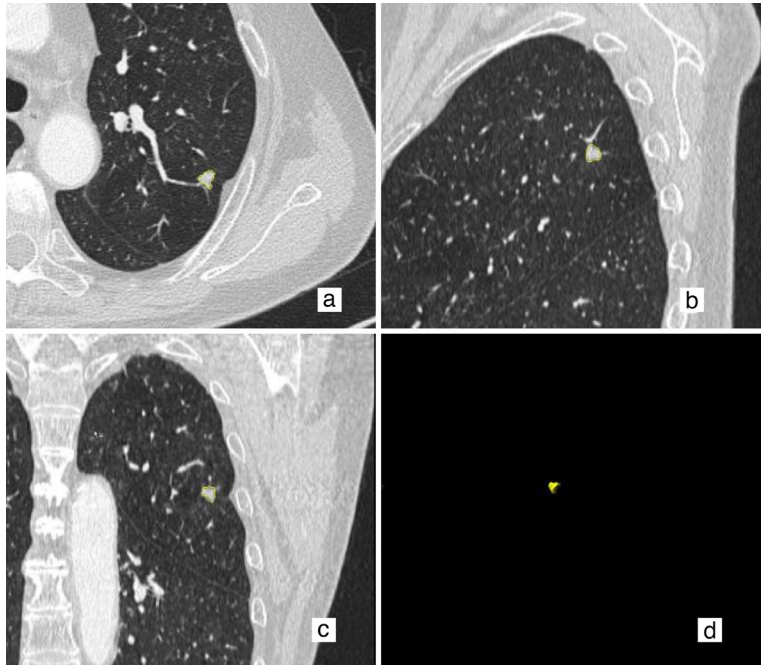


Figure 2. (a, b, and c) are used to determine the region of interest boundary layer by layer from the axial, sagittal, and coronal positions, respectively. (d) Is a three-dimensional diagram of the nodule segmentation.

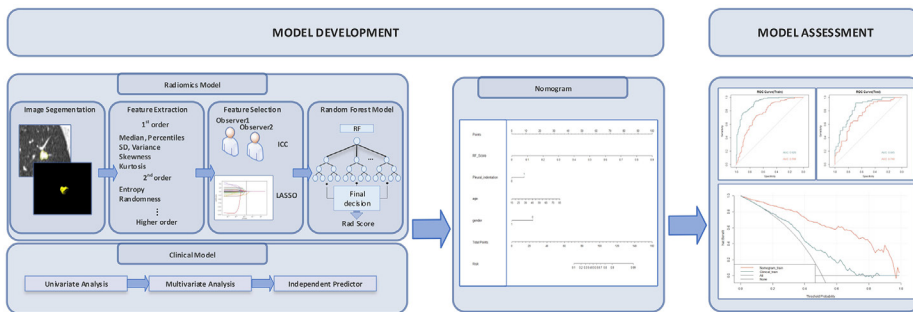


Figure 3. Flow chart of data processing. RF, random forest; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; B, regression coefficient.

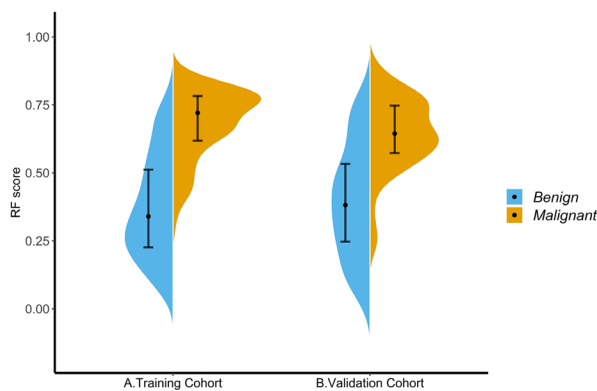


Figure 4. Distribution of the radiomic scores between benign and malignant nodules in the training and validation cohorts, with the ordinate representing the ability to differentiate between benign and malignant nodules.

Our model covers various pathological types, such as adenocarcinoma, squamous cell carcinoma, small cell carcinoma, inflammation, tuberculosis, hamartoma, and other pathological types. We extracted 1,691 features from each nodule, and the most critical and reproducible features were selected to construct the prediction model. However, our results were not significantly different from those of other studies. Because the SSPNs included in our study were difficult to diagnose on imaging, we excluded 20–30 mm nodules and subsolid nodules. Larger-diameter nodules and subsolid nodules are more likely to become

Table 2. Clinical model screening results

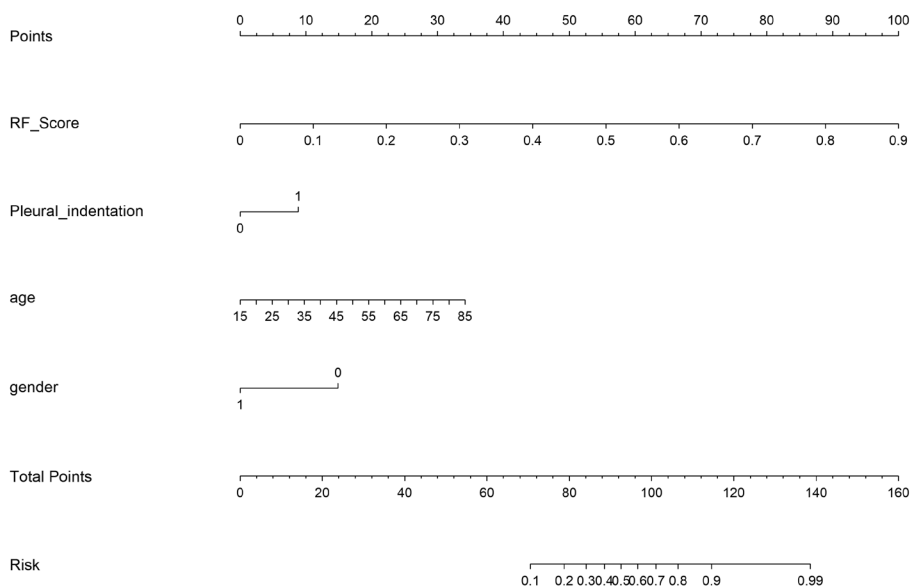
Variables	Variable category		Univariate logistic regression			Multivariable logistic regression		
	Reference category	Other category	B	OR (95% CI)	P value	B	OR (95% CI)	P value
Sex	Female	Male	-0.55	0.58 (0.36-0.92)	0.021	-0.77	0.46 (0.270-0.81)	0.007
Age	-	-	0.04	1.04 (1.02-1.07)	<0.001	0.04	1.04 (1.01-1.07)	0.004
Morphology	Regular	Irregular	0.97	2.65 (1.65-4.28)	<0.001	0.73	2.07 (1.23-3.49)	0.006
Pleural indentation	No	Yes	1.16	3.17 (1.96-5.2)	<0.001	1.15	3.16 (1.86-5.39)	<0.001
Emphysema	No	Yes	0.9	2.45 (1.15-5.58)	0.021	0.97	2.63 (1.07-6.48)	0.036
Significant lung nodule enhancement	No	Yes	0.83	2.29 (1.44-3.69)	0.001	0.75	2.11 (1.25-3.57)	0.005
Density	Uniform	Non-uniform	0.83	2.29 (1.43-3.7)	0.001	-	-	-
Boundary	Clear	Unclear	0.59	1.81 (1.10-3.02)	0.02	-	-	-
Spicule sign	No	Yes	0.53	1.70 (1.06-2.75)	0.029	-	-	-
Location	Middle and lower	Upper	0.41	1.51 (0.95-2.41)	0.083	-	-	-
Average diameter	-	-	-0.03	0.97 (0.63-1.45)	0.865	-	-	-
Clinical symptoms	No	Yes	0.23	1.26 (0.75-2.12)	0.384	-	-	-
Smoking history	No	Yes	-0.18	0.84 (0.51-1.37)	0.479	-	-	-
Lung cancer history	No	Yes	-0.04	0.96 (0.43-2.17)	0.926	-	-	-
Respiratory disease history	No	Yes	0.26	1.30 (0.28-6.68)	0.737	-	-	-
Extrathoracic malignancy history	No	Yes	1.03	2.79 (0.93-10.26)	0.074	-	-	-

OR, odds ratio; CI, confidence interval; B, regression coefficient.

Table 3. Performance comparison between different models

		AUC	Sensitivity	Specificity	Accuracy	PPV	NPV	Cut-off	R ²	P value
Training cohort	Clinical model	0.77	0.70	0.73	0.72	0.72	0.71	0.49	0.28	<0.001
	Nomogram	0.92	0.82	0.82	0.82	0.83	0.81	0.56	0.61	<0.001
Validation cohort	Clinical model	0.75	0.66	0.70	0.68	0.68	0.68	0.49	0.24	<0.001
	Nomogram	0.85	0.74	0.81	0.77	0.76	0.79	0.56	0.46	<0.001

AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value. R², Nagelkerke R-square.

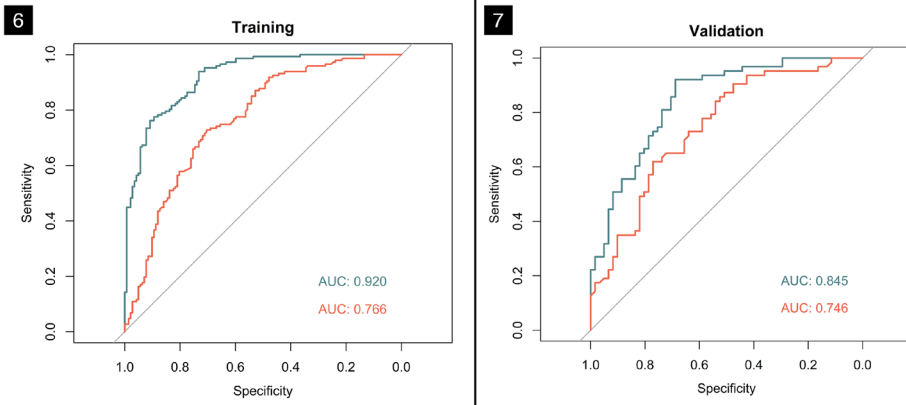


malignant. Models containing these factors may show better efficacy.^{25,26}

Our research has several limitations. First, this is a retrospective study, but our results are further verified in the internal cohort. Second, we only included nodules with pathological results from surgery or biopsy, indicating selection bias. Additionally, our model was established based on clinical, radiomic, and image features of the pulmonary nodules. In addition, some other articles have applied other information, such as the nodules with attached vessel tortuosity and topological skeleton of the nodules.^{27,28} In future studies, our model may be further improved if these latest research results are combined.

In conclusion, in this retrospective study, we constructed a nomogram combining clinical features, subjective CT imaging find-

Figure 5. Nomogram for predicting the risk of indeterminate small solid pulmonary nodules.



Figures 6, 7. Receiver operating characteristic curves of the two prediction models in the training and verification queues when predicting uncertain small solid pulmonary nodules.

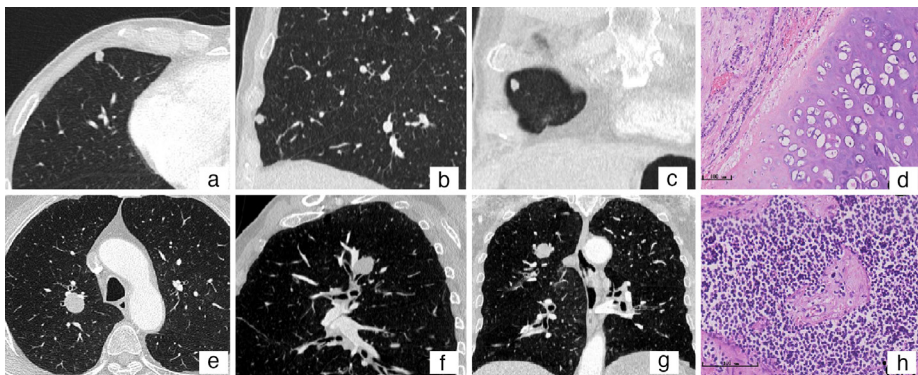


Figure 8. From left to right: axial computed tomography images, predicted results of the nomogram, and pathological diagnosis images of uncertain pulmonary nodules (hematoxylin and eosin staining, $\times 400$). First row: a 74-year-old man presented with a small nodule (approximately 7.5 mm in diameter) in the middle lobe of his right lung (a-c), with a total nomogram score of 56.3, a 0.01 probability of malignancy, and a pathologically proven hepatocellular tumor (d). Second row: a 69-year-old man presented with a nodule (approximately 18.5 mm in diameter) in the superior lobe of his right lung (e-g), which had a total nomogram score of 107.5 and a probability of malignancy of 0.81. Pathology confirmed a small cell carcinoma (h).

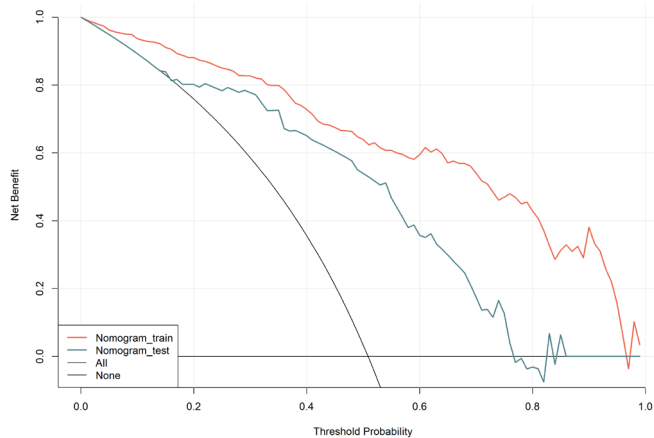


Figure 9. Decision curve analysis of the nomogram. The x-axis represents the threshold probability, and the y-axis represents the net-income benefit. The red line represents the clinical benefits of small solid pulmonary nodules (SSPNs) predicted by the nomogram in the training group. The blue line indicates the clinical benefits of SSPNs predicted by the nomogram in the validation group. The black slash represents the assumption that all nodules are malignant. The black horizontal line represents the assumption that all nodules are benign. The decision curve shows that the nomogram increases the net benefits in forecasting uncertain SSPNs.

ings, and radiomic features to differentiate indeterminate SSPNs. This nomogram is non-invasive and repeatable and has high predictive accuracy to help with the preoperative diagnosis of patients.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Hassannezhad R, Vahed N. Prediction of the risk of malignancy among detected lung nodules in the National Lung Screening Trial. *J Am Coll Radiology*. 2018;15(11):1529-1535. [\[CrossRef\]](#)
- Leleu O, Basille D, Auquier M, et al. Lung cancer screening by low-dose CT scan: baseline results of a french prospective study. *Clin Lung Cancer*. 2020;21(2):145-152. [\[CrossRef\]](#)
- McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first-screening CT. *N Eng J Med*. 2013;369(10):910-919. [\[CrossRef\]](#)
- White CS, Dharaiya E, Campbell E, Boroczky L. The vancouver lung cancer risk prediction model: assessment by using a subset of the National Lung Screening Trial cohort. *Radiology*. 2017;283(1):264-272. [\[CrossRef\]](#)
- Erasmus JJ, McAdams HP, Connolly JE. Solitary pulmonary nodules: part II. Evaluation of the indeterminate nodule. *Radiographics*. 2000;20:59-66. [\[CrossRef\]](#)
- Chen WQ, Li H, Sun KX, et al. Report of cancer incidence and mortality in China, 2014. *Zhonghua Zhong Liu Za Zhi*. 2018;40(1):5-13. [\[CrossRef\]](#)
- Qu RY, Zhou BS. Analysis of the distribution and trend of lung cancer mortality in China between 2004 and 2010 [in Chinese]. *Chin J Health Stat*. 2014;31:932-5. [\[CrossRef\]](#)
- Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. *Radiology*. 2005;235(1):259-265. [\[CrossRef\]](#)
- Liu X, Liang M, Wang Y, et al. The outcome differences of CT screening for lung cancer pre and post following an algorithm in Zhuhai, China. *Lung Cancer*. 2011;73(2):230-236. [\[CrossRef\]](#)
- Liu A, Wang Z, Yang Y, et al. Preoperative diagnosis of malignant pulmonary nodules in lung cancer screening with a radiomics nomogram. *Cancer Commun (Lond)*. 2020;40(1):16-24. [\[CrossRef\]](#)
- Chen X, Feng B, Chen Y, et al. A CT-based radiomics nomogram for prediction of lung adenocarcinomas and granulomatous lesions in patient with solitary sub-centimeter solid nodules. *Cancer Imaging*. 2020;20(1):45. [\[CrossRef\]](#)

12. Mei X, Wang R, Yang W, et al. Predicting malignancy of pulmonary ground-glass nodules and their invasiveness by random forest. *J Thorac Dis.* 2018;10(1):458-463. [\[CrossRef\]](#)
13. Li XX, Li B, Tian LF, Zhang L. Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. *IET Image Processing.* 2018;12(7):1253-1264. [\[CrossRef\]](#)
14. Yang M, Ren Y, She Y, et al. Imaging phenotype using radiomics to predict dry pleural dissemination in non-small cell lung cancer. *Ann Transl Med.* 2019;7(12):259. [\[CrossRef\]](#)
15. Xu X, Huang L, Chen J, et al. Application of radiomics signature captured from pretreatment thoracic CT to predict brain metastases in stage III/IV ALK-positive non-small cell lung cancer patients. *J Thorac Dis.* 2019;11:4516-4528. [\[CrossRef\]](#)
16. Shen Y, Xu F, Zhu W, Hu H, Chen T, Li Q. Multiclassifier fusion based on radiomics features for the prediction of benign and malignant primary pulmonary solid nodules. *Ann Transl Med.* 2020;8(5):171. [\[CrossRef\]](#)
17. Ruilong Z, Daohai X, Li G, Xiaohong W, Chunjie W, Lei T. Diagnostic value of 18F-FDG-PET/CT for the evaluation of solitary pulmonary nodules: a systematic review and meta-analysis. *Nucl Med Commun.* 2017;38(1):67-75. [\[CrossRef\]](#)
18. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med.* 1997;157(8):849-855. [\[CrossRef\]](#)
19. Li Y, Chen KZ, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. *Clin Lung Cancer.* 2011;12(5):313-319. [\[CrossRef\]](#)
20. Swensen SJ, Viggiano RW, Midthun DE, et al. Lung nodule enhancement at CT: multicenter study. *Radiology.* 2000;214(1):73-80. [\[CrossRef\]](#)
21. Soardi GA, Perandini S, Larici AR, et al. Multicentre external validation of the BIMC model for solid solitary pulmonary nodule malignancy prediction. *Eur Radiol.* 2017;27(5):1929-1933. [\[CrossRef\]](#)
22. Ruilong Z, Daohai X, Li G, Xiaohong W, Chunjie W, Lei T. Diagnostic value of 18F-FDG-PET/CT for the evaluation of solitary pulmonary nodules: a systematic review and meta-analysis. *Nucl Med Commun.* 2017;38(1):67-75. [\[CrossRef\]](#)
23. Xiang Y, Sun Y, Liu Y, et al. Development and validation of a predictive model for the diagnosis of solid solitary pulmonary nodules using data mining methods. *J Thorac Dis.* 2019;11(3):950-958. [\[CrossRef\]](#)
24. She Y, Zhao L, Dai C, et al. Development and validation of a nomogram to estimate the pretest probability of cancer in Chinese patients with solid solitary pulmonary nodules: a multi-institutional study. *J Surg Oncol.* 2017;116(6):756-762. [\[CrossRef\]](#)
25. Chu ZG, Zhang Y, Li WJ, Li Q, Zheng YN, Lv FJ. Primary solid lung cancerous nodules with different sizes: computed tomography features and their variations. *BMC Cancer.* 2019;19(1):1060. [\[CrossRef\]](#)
26. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Eng J Med.* 2013;369(10):910-919. [\[CrossRef\]](#)
27. Mahdi MBT, Orooji M, Teimouri M, Shahabifar R. Segmentation of the pulmonary nodule and the attached vessels in the CT scan of the chest using morphological features and topological skeleton of the nodule. *IET Image Processing.* 2020;14(87):1520-1528. [\[CrossRef\]](#)
28. Tavakoli MB, Orooji M, Teimouri M, Shahabifar R. Distinguishing adenocarcinomas from granulomas in the CT scan of the chest: performance degradation evaluation in the automatic segmentation framework. *BMC Res Notes.* 2021;14(1):87. [\[CrossRef\]](#)

Appendix 1. Packages used in this study			
Statistical analysis	Packages	Version	Web
Feature selection: LASSO	"glmnet" in R	4.1-1	https://CRAN.R-project.org/package=glmnet
Radiomic model establishment: random forest	"randomForestSRC" in R	2.11.0	https://CRAN.R-project.org/package=randomForestSRC
Descriptive table	"compareGroups" in R	4.5.1	https://CRAN.R-project.org/package=compareGroups
Clinical model and nomogram establishment: logistic regression and nomogram	"rms" in R	6.0-1	https://CRAN.R-project.org/package=rms
Model performance evaluation: DeLong test and 95% CI for ROC curve analysis	"pROC" in R	1.18.0	https://CRAN.R-project.org/package=pROC
Model performance evaluation: decision curve analysis	"rmda" in R	1.6	https://CRAN.R-project.org/package=rmda
Model performance evaluation: Hosmer–Lemeshow Goodness of Fit test	"ResourceSelection" in R	0.3-5	https://CRAN.R-project.org/package=ResourceSelection

CI, confidence interval; ROC, receiver operating characteristic curve.