

# Development and Validation of a Preoperative CT-Based Nomogram to Differentiate Invasive from Non-Invasive Pulmonary Adenocarcinoma in Solitary Pulmonary Nodules

Xin Song<sup>1,2</sup>, Qingtao Zhao<sup>1</sup>, Hua Zhang<sup>1</sup>, Wenfei Xue<sup>1</sup>, Zhifei Xin<sup>1</sup>, Jianhua Xie<sup>1,2</sup>, Xiaopeng Zhang<sup>1</sup>

<sup>1</sup>Department of Thoracic Surgery, Hebei General Hospital, Shijiazhuang, People's Republic of China; <sup>2</sup>North China University of Science and Technology, Tangshan, People's Republic of China

Correspondence: Xiaopeng Zhang, Hebei General Hospital, No. 348, Heping Western Road, Xinhua District, Shijiazhuang, 050000, People's Republic of China, Tel +8613722865878, Email 13722865878@163.com

**Purpose:** This study aimed to develop and validate a preoperative CT-based nomogram combined with clinical and radiological features for distinguishing invasive from non-invasive pulmonary adenocarcinoma.

**Patients and Methods:** A total of 167 patients with solitary pulmonary nodules and pathologically confirmed adenocarcinoma treated between January 2020 and December 2020 at Hebei General Hospital were retrospectively assessed. To evaluate the probability of invasive pulmonary adenocarcinoma, we developed three models, the multivariate logistic regression model, the stepwise logistic regression model, and the cross-validation model. The Akaike information criterion (AIC) was used to compare the relative strength of different models, and the area under the curve (AUC) was used to quantify the predictive accuracy. The best performing model was presented as a nomogram, calibrated and evaluated for clinical utility.

**Results:** The stepwise logistic regression model revealed highest and mean attenuations of non-enhanced CT images, and lobulation and vacuole presence were predictive factors of invasive pulmonary adenocarcinoma. This model (AIC = 67.528) with the lowest AIC value compared with that of the multivariate logistic regression model (AIC = 69.301) or the cross-validation model (AIC = 81.216) was identified as the best model, and its AUC value (0.9967; 95% CI, 0.9887–1) was higher than those of the other two models. The calibration curve showed optimal agreement in invasive pulmonary adenocarcinoma probability as predicted by the nomogram and the actual value.

**Conclusion:** We developed and validated a nomogram that could estimate the preoperative probability of invasive pulmonary adenocarcinoma in patients with solitary pulmonary nodules, which may be useful in clinical decision-making associated with personalized surgical intervention and therapeutic regimen selection.

**Keywords:** solitary pulmonary nodule, nomogram, invasive pulmonary adenocarcinoma, lung cancer

## Introduction

With the development of imaging techniques, including high-resolution Computed Tomography (CT) and low-dose CT, as well as increased health awareness, more and more early pulmonary nodules are detected. A solitary pulmonary nodule (SPN) is defined as a single, well-defined radiographic opacity that measures up to 30 mm in diameter and is completely surrounded by the lung parenchyma without atelectasis, mediastinal lymphadenopathy, or pleural effusion.<sup>1</sup> Lung adenocarcinoma is the commonest histopathological type of peripheral carcinoma, and its incidence has been increasing in recent years.<sup>2–7</sup> According to the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinoma, lung adenocarcinoma can be divided into four groups: (1) preinvasive lesions, including atypical adenomatous hyperplasia (AAH) and

adenocarcinoma in situ (AIS); (2) minimally invasive adenocarcinoma (MIA); (3) invasive pulmonary adenocarcinoma (IPA); (4) Variants of invasive adenocarcinoma.<sup>8</sup>

Benign pulmonary lesions have more typical imaging signs, and radiologists and thoracic surgeons could easily distinguish them. Patients are generally expected to be cured only by internal conservative treatment. Imaging manifestations of advanced and rapidly progressing lung lesions are also typical. Patients are not eligible for surgery, and the progression of lesions can be controlled by chemotherapy, radiotherapy, targeted therapy, immunotherapy and other medical tools. However, early and intermediate pulmonary cancers are not easily distinguishable by high-resolution CT, and there are surgical opportunities for both grades. For adenocarcinoma with the highest incidence, early-stage lesions include AAH, AIS and MIA, and sub-lobectomy is feasible to achieve cure; for intermediate-stage lesions, invasive adenocarcinoma, lobectomy is feasible. At present, the main means for determining the surgical method is to obtain frozen sections of the diseased tissue during the operation. An effective model to differentiate between early and middle lung lesions should (1) have minimal economic burden for patients to obtain the optimal treatment effect, (2) decrease the time of unnecessary follow-up of patients with effective diagnosis and treatment plan, and (3) provide intervention measures for lesions at the appropriate time to avoid further tumor progression and losing the opportunity to operation.

Nomograms are typically used as an easy applied math prognostic model that quantifies the hazard of a clinical event like cancer. Since pneumonic glandular cancer has the highest incidence rate, this study aimed to develop a predictive model that differentiates invasive from non-invasive pulmonary adenocarcinoma in cases with solitary pulmonary nodules, and to establish a nomogram that combines CT and clinical features for the pretreatment identification of IPA risk in patients with SPNs.

## Materials and Methods

### Patient Selection

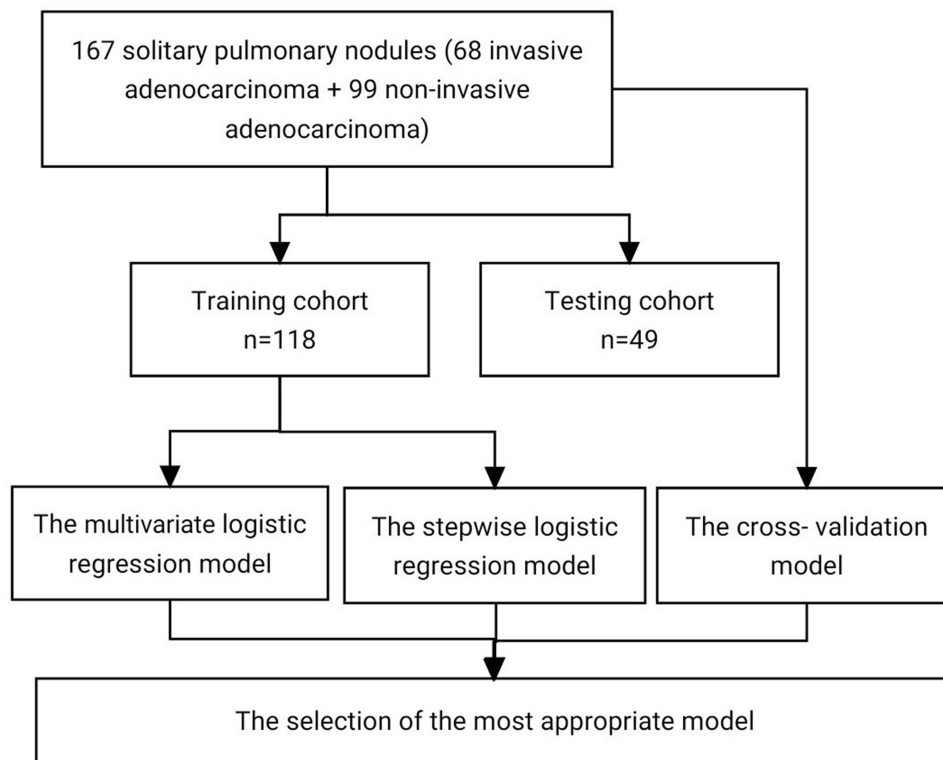
This retrospective study was approved by the ethics committee of Hebei General Hospital. All patients provided signed informed consent. The clinical data and CT images of 252 patients administered complete surgical resection and pathologically confirmed with AAH, AIS, MIA, and IPA in Hebei General Hospital (Shijiazhuang, Hebei, China) between January 2020 and December 2020 were reviewed. The flowchart summarizes the study design of each approach and shows in [Figure 1](#). Inclusion criteria were: (1) SPNs on lung window images (level and width of  $-700$  and  $1000$  Hounsfield unit [HU], respectively); (2) nodules of  $5-30$  mm in longest diameter on axial images; (3) aged over 18 years; (4) no lung cancer or other malignant tumor in the past five years; (5) asymptomatic at diagnosis; (6) no preoperative treatment; (7) all lesions diagnosed as primary pulmonary adenocarcinoma. Exclusion criteria were: (1) nodules larger than  $3$  cm in diameter; (2) complications with other pulmonary lesions; (3) a history of lung disease; (4) metastatic tumors; (5) disease histopathologically confirmed to comprise other pathological types such as squamous cell carcinoma and small cell lung cancer. Based on these inclusion and exclusion criteria, 85 patients were excluded. Finally, 167 patients ( $58$  males ( $60 \pm 11$  years) and  $109$  females ( $58 \pm 11$  years)) were analyzed.

### Variables

In this study, clinical variables such as gender, age, body mass index (BMI), smoking history, history of cancer, and family history of cancer were collected, as well as radiological chest CT data, including SPN diameter (maximum transverse size), highest attenuation of non-enhanced CT images, mean attenuation of non-enhanced CT images, margin (clear or blurred), spiculation, lobulation, pleural indentation, vacuole, vascular convergence and SPN location (upper or non-upper lobes). Both clinical and pathological data were assessed by 2 senior radiologists blinded to the current study. Laboratory characteristics, including carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCC), neuron specific enolase (NSE) and cytokeratin 19 fragment (CYFRA21-1), were also recorded.

### Histopathological Analysis

All pathological specimens were formalin fixed, stained with hematoxylin-eosin and evaluated by two experienced pathologists. Histopathological evaluations were conducted by examining hematoxylin-eosin-stained slides by light



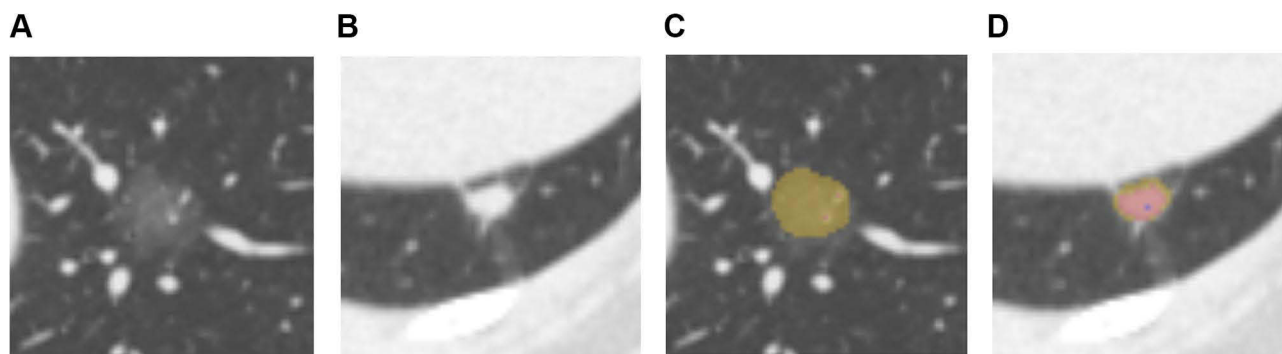
**Figure 1** Flowchart for building the prediction model.

microscopy. All specimens were classified according to the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinomas.<sup>8,9</sup> AAH is a localized, small (usually 0.5 cm or less) proliferation of mildly to moderately atypical type II pneumocytes and/or Clara cells lining the alveolar wall and sometimes, respiratory bronchioles. AIS is a localized small ( $\leq 3$  cm) adenocarcinoma with growth restricted to neoplastic cells along preexisting alveolar structures (lepidic growth), lacking stromal, vascular, or pleural invasion. MIA is a small, solitary adenocarcinoma ( $\leq 3$  cm), with a predominantly lepidic pattern and  $\leq 5$  mm invasion in greatest dimension for any one focus.<sup>10</sup>

## CT Examination and Imaging Analysis

All chest CT examinations included the entire thorax, and were performed in the supine position. Single scans were obtained during deep inspiration and breath-hold. A SIEMENS SOMATOM Definition Flash system (Siemens Healthineers, Erlangen, Germany) was used for CT, with the following parameters: 120 kV tube voltage; 110 mA tube current; 80 mm (128 $\times$ 0.625 mm) collimation width; 1 mm slice thickness; 1.2 pitch; 280 ms rotation time. The scans were analyzed in lung window setting (1000 HU window width and  $-700$  HU window level) and the mediastinal window (350 HU window width and 40 HU window level).

All SPNs were automatically detected using uAI platform (United Imaging Healthcare, Shanghai, China), an artificial intelligence software. All regions of interest (ROIs) were manually drawn on the CT images on each section as an area that was as large as possible from all sections to cover the target nodule. The representative segmentation results were shown in Figure 2. Radiographic features were assessed independently by two experienced radiologists blinded to pathological results. Discrepancies between observers were resolved by consensus. The radiographic characteristics analyzed for each solitary pulmonary nodule included: (1) SPN diameter; (2) highest attenuation of non-enhanced CT images; (3) mean attenuation of non-enhanced CT images; (4) margin (clear or blurred); (5) spiculation (present or absent); (6) lobulation (present or absent); (7) pleural indentation (present or absent); (8) vacuole (present or absent); (9) vascular convergence; (10) SPN location (upper or non-upper lobes). SPN diameter was defined as the average of the



**Figure 2** Representative segmentation results of SPN. (A and C) A 53-year-old male with SPN. (B and D) A 51-year-old female with SPN. (A and B) Original CT image of SPN. (C and D) Manual segmentation result of SPN.

major and vertical short diameters measured on the largest section in lung window.<sup>3,11</sup> Attenuations were measured manually for the lesion region. A clear margin was defined as sharply demarcated, with a round or oval smooth shape.<sup>3</sup> Spiculation was defined as strands extending from the nodule margin into the lung parenchyma without reaching the pleural surface.<sup>9</sup> Lobulation was defined as a part of the nodule surface (except for the portions in contact with the pleura) exhibiting a concave or straightened configuration.<sup>3</sup> A pleural indentation was defined as a trumpet-shaped attenuation extending toward the pleura from the nodule.<sup>9</sup> A vacuole was defined as a small spot of round or ovoid air attenuation occurring within a solitary pulmonary nodule.<sup>12</sup> Vascular convergence was defined with vessels crowding the internal part of the nodule or abnormally angled toward the nodule compared with the normal lung parenchyma.<sup>9</sup>

## Collection of Tumor Markers

The tumor markers assessed included CEA, SCC, NSE and CYFRA21-1. Serum carcinoembryonic antigen is an established tumor marker in non-small cell lung cancer (NSCLC), although modest elevation of CEA is often seen without malignancy.<sup>13</sup> CEA also plays a significant role as a regulator of tumor cell proliferation and differentiation. Overexpression of CEA contributes to the progression of many epithelial cancers and to immune dysfunction.<sup>14</sup> SCC is a specific marker of squamous cell carcinoma and an independent prognostic factor of cervical squamous cell carcinoma. NSE represents a highly specific marker of neuronal and peripheral neuroendocrine cells, and is elevated in the serum of patients with neuroblastoma, melanoma, and seminoma.<sup>15</sup> Cyfra 21-1 is mainly expressed in epithelial-derived cells and can be used for staging and prognostic analysis of nasopharyngeal carcinoma.<sup>16</sup>

## Statistical Analysis

In this study, three statistical methods were used to establish prediction models, and the final screening variables were determined by comparing the area under the curves (AUCs) for the three methods. Nominal categorical variables were compared by Pearson's Chi square test or Fisher's exact test. The distribution of ordinal categorical variables was examined by the Shapiro–Wilk test, and Student's *t*-test was performed to compare the training and validation cohorts or the invasive and non-invasive pulmonary adenocarcinoma groups for normally distributed features; otherwise, the Mann–Whitney *U*-test was used. In multivariate logistic regression model, the patients were divided into the training (n=118) and validation (n=49) cohorts at a ratio of 7:3. In the training cohort, univariate logistic regression analysis was used to determine associations of SPN invasion and clinical characteristics with CT imaging variables. In stepwise logistic regression model, the univariate logistic regression analysis described in multivariate logistic regression model was used to evaluate the clinicopathological factors in the training cohort. Factors with a *P*-value <0.05 were considered in stepwise multivariate logistic regression analysis. Meanwhile, the Akaike information criterion (AIC) was used to determine a stopping rule, and only factors with a *P*-value below 0.05 were incorporated into the nomogram. The factors were chosen to construct the stepwise logistic regression model. In cross-validation model, all patients were selected for construction by using a 10-fold cross-validation method. The AIC was applied to compare the relative strength of different models and identify the most

appropriate model. The AUC was used to quantify the predictive accuracy of the three models. The predictive accuracies of the first two models were assessed by the AUCs for the validation cohort. The predictive accuracy of the last model was assessed by the AUC for the overall cohort. Then, the nomogram of the most appropriate model was established by comparing the three models. In addition, calibration of the nomogram was performed by a visual calibration curve with 200 bootstrap resamples. The R software (version 4.1.0; <http://www.R-project.org>) was used to perform all statistical analyses. The reported statistical significance was set at a *P*-value <0.05. The “glmnet” package was used to conduct the least absolute shrinkage and selection operator (LASSO) model. The “rms” package was used to build nomograms and generate calibration curves. The “pROC” package was used to plot receiver operating characteristic (ROC) curves.

## Results

The characteristics of the 167 included participants with SPNs are described in Table 1. Overall, 58 (34.7%) male and 109 (65.3%) female participants with a mean age of 58.44±10.95 years were assessed in the current study. The prevalence of IPA was 40.7%. The data of 118 patients were utilized to develop a predictive model, while those of the remaining 49 were used for validation. There were no significant differences between the two cohorts in terms of patient characteristics, image-based features, and tumor biomarkers (*P*-values ranging between 0.0564 and 0.9619). There was also no significant difference in the prevalence of IPA between the two cohorts (*P*=0.1614). The rates of IPA in the training and validation cohorts were 37.3% and 49.0%, respectively. The clinical, radiological, and laboratory characteristics of patients in the training cohort are summarized in Table 2. Univariable analysis revealed significant differences between the invasive and non-invasive pulmonary adenocarcinoma cases in terms of age, diameter, highest attenuation of non-enhanced CT images, mean attenuation of non-enhanced CT images, spiculation, lobulation, pleural indentation, vacuole, and vascular convergence in the training cohort. Independent predictors of IPA were determined by multivariate logistic regression analysis, including highest attenuation of non-enhanced CT images (*P*=0.0013), mean attenuation of non-enhanced CT images (*P*=0.0459), and lobulation (*P*=0.0100) (Table 3). Then, the multivariate logistic regression model was reflected by the following formula: probability of IPA in the SPN:  $P = e^x / (1 + e^x)$ , where  $x = 1.1344 + (0.0106 \times \text{highest attenuation of non-enhanced CT images}) + (0.0037 \times \text{mean attenuation of non-enhanced CT images}) + (2.1363 \times \text{lobulation})$ .

Univariate logistic regression analysis demonstrated significant differences between invasive and non-invasive pulmonary adenocarcinoma in terms of age, diameter, highest attenuation of non-enhanced CT images, mean attenuation of non-enhanced CT images, spiculation, lobulation, pleural indentation, vacuole, and vascular convergence in the training cohort. In addition, highest attenuation of non-enhanced CT images, mean attenuation of non-enhanced CT images, lobulation, and vacuole remained viable predictive indicators in the stepwise logistic regression model as shown in Table 4. Then, the stepwise logistic regression model was reflected by the following formula: probability of IPA in the SPN:  $P = e^x / (1 + e^x)$ , where  $x = 0.9902 + (0.0099 \times \text{highest attenuation of non-enhanced CT images}) + (0.0046 \times \text{mean attenuation of non-enhanced CT images}) + (2.3892 \times \text{lobulation}) + (1.4948 \times \text{vacuole})$ .

After 10-fold cross-validation with the LASSO, the most valuable five features with nonzero coefficients were selected to establish the cross-validation model, including age, diameter, highest attenuation of non-enhanced CT images, mean attenuation of non-enhanced CT images, and lobulation (Figure 3 and 4). Then, the cross-validation model was reflected by the following formula: probability of IPA in the SPN:  $P = e^x / (1 + e^x)$ , where  $x = 0.3096 + (0.0027 \times \text{age}) + (0.0270 \times \text{diameter}) + (0.0049 \times \text{highest attenuation of non-enhanced CT images}) + (0.0024 \times \text{mean attenuation of non-enhanced CT images}) + (0.6675 \times \text{lobulation})$ .

The stepwise logistic regression model (AIC = 67.528) with the lowest AIC value compared with the multivariate logistic regression model (AIC = 69.301) or the cross-validation model (AIC = 81.216) was identified as the best model. On the other hand, as shown in Table 5, the AUC of the stepwise logistic regression model (AUC = 0.9967; 95% CI: 0.9887–1) was higher than the multivariate logistic regression model (AUC = 0.9933; 95% CI: 0.9805–1) or the cross-validation model (AUC = 0.9731; 95% CI: 0.9437–1). Therefore, the stepwise logistic regression model showed the best discrimination. The ROC curves of the three models were presented in Figure 5. In this case, four momentous risk factors that predicted IPAs were selected to construct a nomogram based on the stepwise logistic regression model (Figure 6). In the nomogram, the impact of highest attenuation of non-enhanced CT images on the risk of IPA was the highest (100

**Table 1** Clinical and Imaging Characteristics of All Patients with Solitary Pulmonary Nodules

Characteristics	Overall Cohort (n=167)	Training Cohort (n=118)	Testing Cohort (n=49)	p-value
Clinical factors				
Gender				0.2814
Male	58(34.7)	44(37.3)	14(28.6)	
Female	109(65.3)	74(62.7)	35(71.4)	
Age (year)	58.44±10.95	58.35±10.82	58.67±11.36	0.6487
BMI (kg/m <sup>2</sup> )	24.45±3.14	24.57±3.18	24.16±3.07	0.4417
Smoking history				0.5961
Yes	35(21)	26(22)	9(18.4)	
No	132(79)	92(78)	40(81.6)	
History of cancer				0.9619
Yes	7(4.2)	5(4.2)	2(4.1)	
No	160(95.8)	113(95.8)	47(95.9)	
Family history of cancer				0.7193
Yes	17(10.2)	13(11)	4(8.2)	
No	150(89.8)	105(89)	45(91.8)	
CEA (ng/mL)	6.63±28.53	5.87±27.17	8.47±31.79	0.0615
SCC (ng/mL)	1.46±0.89	1.51±0.94	1.32±0.76	0.2219
NSE (ng/mL)	12.42±2.75	12.65±2.99	11.87±1.97	0.1461
CYFRA21-I (ng/mL)	2.27±1.07	2.26±1.13	2.31±0.93	0.3921
Radiologic factors				
Nodule diameter (mm)	13.99±6.99	13.14±6.28	16.04±8.18	0.0564
Highest attenuation	-138.66±231.15	-155.02±231.93	-99.29±226.76	0.0961
Mean attenuation	-419.62±265.98	-431.09±261.38	-392±277.55	0.4199
Margin				0.1008
Clear	69(41.3)	44(37.3)	25(51.0)	
Blurred	98(58.7)	74(62.7)	24(49.0)	
Spiculation				0.2332
Yes	44(26.3)	28(23.7)	16(32.7)	
No	123(73.7)	90(76.3)	33(67.3)	
Lobulation				0.4037
Yes	47(28.1)	31(26.3)	16(32.7)	
No	120(71.9)	87(73.7)	33(67.3)	

(Continued)



**Table 1** (Continued).

Characteristics	Overall Cohort (n=167)	Training Cohort (n=118)	Testing Cohort (n=49)	p-value
Pleural indentation				0.0680
Yes	64(38.3)	40(33.9)	24(49.0)	
No	103(61.7)	78(66.1)	25(51.0)	
Vacuole				0.9516
Yes	38(22.8)	27(22.9)	11(22.4)	
No	129(77.2)	91(77.1)	38(77.6)	
Vascular convergence				0.7463
Yes	102(61.1)	73(61.9)	29(59.2)	
No	65(38.9)	45(38.1)	20(40.8)	
Lesion location				0.8887
Upper lobe	107(64.1)	76(64.4)	31(63.3)	
Non-upper lobe	60(35.9)	42(35.6)	18(36.7)	
Pathology type				0.1614
Invasive adenocarcinoma	68(40.7)	44(37.3)	24(49.0)	
Non-invasive adenocarcinoma	99(59.3)	74(62.7)	25(51.0)	

**Abbreviations:** BMI, body mass index; CEA, carcinoembryonic antigen; SCC, squamous cell carcinoma antigen; NSE, neuron specific enolase; CYFRA21-1, cytokeratin 19 fragment.

points), followed by mean attenuation of non-enhanced CT images (25 points), lobulation (18 points) and vacuole (6 points). The following is an example scenario to calculate the total points and predict the possibility of a lesion being an IPA: For a patient with SPN, highest and mean attenuations of non-enhanced CT images are 200 HU (90 points) and -100 HU (20 points) respectively. At the same time, there was not lobulation sign (0 point) or vacuole sign (0 point) on his or her CT images. In total, this patient had approximately 110 points on the nomogram. Theoretically, the possibility of this SPN being an IPA is over 90%.

A calibration curve validating the model's performance is shown in [Figure 7](#), indicating that the probability of IPA in SPNs predicted by the nomogram agreed well with the actual value.

## Discussion

This study used data from Hebei General Hospital to develop and validate a novel predictive model to assess the risk of IPA in patients who radially presented with SPNs. Patients with high scores need to achieve cure by lobectomy, while those with low scores need sub-lobectomy. Therefore, the model established to distinguish invasive from non-invasive pulmonary adenocarcinoma in SPNs could improve the management and prognosis of patients with lung adenocarcinoma. In the present study, four characteristics, including highest attenuation of non-enhanced CT images, mean attenuation of non-enhanced CT images, lobulation and vacuole, were significant predictors of IPA with SPNs. Importantly, we developed a nomogram to estimate the preoperative patient-specific "risk" of IPA with excellent predictive accuracy.

In lung cancer screening, radiologists usually stratify lung nodules by explaining features such as location, density, diameter, volume, and margins.<sup>17</sup> Due to the spread of malignant cells in the lung interstitium, malignant nodules are more likely to have irregular edges and lobulated, sharp or prickly signs.<sup>18</sup> Lung adenocarcinoma is the most common

**Table 2** Univariate Logistic Regression Analysis of the Associations of Clinical and Radiological Factors with Pathological Status in the Training Cohort (n=118)

Characteristics	Invasive Adenocarcinoma (n=44)	Non-Invasive Adenocarcinoma (n=74)	P-value	OR	95% CI
Clinical factors					
Gender			0.3084	1.4896	0.6899–3.2162
Male	19	25			
Female	25	49			
Age (year)	62.89±9.02	55.65±10.96	0.0008	1.0759	1.0332–1.1264
BMI (kg/m <sup>2</sup> )	23.96±2.89	24.93±3.3	0.1110	0.9025	0.7909–1.0196
Smoking history			0.5495	1.3111	0.5306–3.1733
Yes	11	15			
No	33	59			
History of cancer			0.8980	1.1270	0.1441–7.0649
Yes	2	3			
No	42	71			
Family history of cancer			0.9261	1.0577	0.3015–3.3990
Yes	5	8			
No	39	66			
CEA (ng/mL)	12.07±44.07	2.19±1.57	0.0625	1.2675	1.0398–1.6631
SCC (ng/mL)	1.45±0.86	1.55±0.99	0.5710	0.8873	0.5713–1.3240
NSE (ng/mL)	12.37±2.20	12.82±3.38	0.4320	0.9472	0.8179–1.0765
CYFRA21-1 (ng/mL)	2.51±1.27	2.11±1.02	0.0714	1.3784	0.9865–2.0067
Radiologic factors					
Nodule diameter (mm)	16.89±6.41	10.92±5.06	<0.0001	1.1942	1.1100–1.2991
Highest attenuation	49.16±118.95	-276.42±194.46	<0.0001	1.0144	1.0096–1.0208
Mean attenuation	-184.50±237.23	-577.72±133.81	<0.0001	1.0089	1.0062–1.0125
Margin			0.30826	1.4896	0.6899–3.2162
Clear	19	25			
Blurred	25	49			
Spiculation			<0.0001	11.3333	4.2984–34.1366
Yes	22	6			
No	22	68			
Lobulation			<0.0001	9.0357	3.6508–24.4433

(Continued)



**Table 2** (Continued).

Characteristics	Invasive Adenocarcinoma (n=44)	Non-Invasive Adenocarcinoma (n=74)	P-value	OR	95% CI
Yes	23	8			
No	21	66			
Pleural indentation			0.0004	4.3500	1.9578–9.9875
Yes	24	16			
No	20	58			
Vacuole			0.0282	2.6724	1.1162–6.5412
Yes	15	12			
No	29	62			
Vascular convergence			0.0002	5.5794	2.3128–15.1115
Yes	37	36			
No	7	38			
Lesion location			0.3530	0.6933	0.3193–1.5091
Upper lobe	26	50			
Non-upper lobe	18	24			

**Abbreviations:** BMI, body mass index; CEA, carcinoembryonic antigen; SCC, squamous cell carcinoma antigen; NSE, neuron specific enolase; CYFRA21-1, cytokeratin 19 fragment; OR, odds ratio; CI, confidence interval.

histological type of peripheral lung cancer. Once discovered, lung nodules must be evaluated to determine the possibility of malignant transformation and determine treatment recommendations.<sup>2–7</sup> The study by Kenji Suzuki et al showed that there was no statistically significant difference in intraoperative and postoperative complications between segmentectomy and lobectomy.<sup>19</sup> If patients with surgical indications can predict the pathological type of pulmonary nodules before

**Table 3** Multivariate Logistic Regression Analysis in the Training Cohort (n=49)

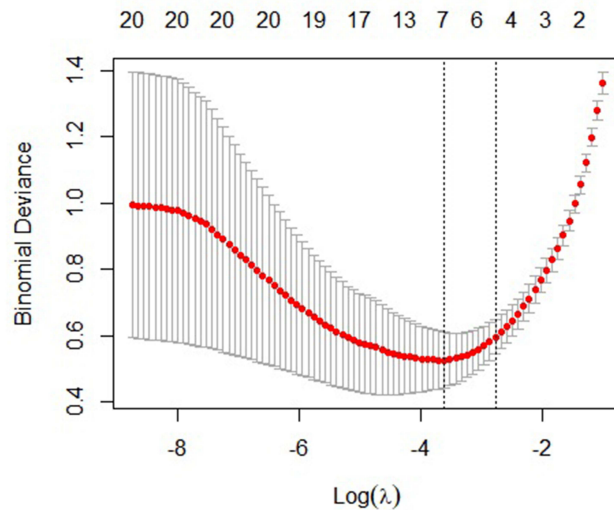
Variable	Regression Coefficient	P-value	OR	95% CI
Highest attenuation	0.0106	0.0013	1.0106	1.0048–1.0180
Mean attenuation	0.0037	0.0459	1.0037	1.0002–1.0076
Lobulation	2.1363	0.0100	8.4680	1.8733–52.8273

**Abbreviations:** OR, odds ratio; CI, confidence interval.

**Table 4** Stepwise Multivariate Logistic Regression Analysis

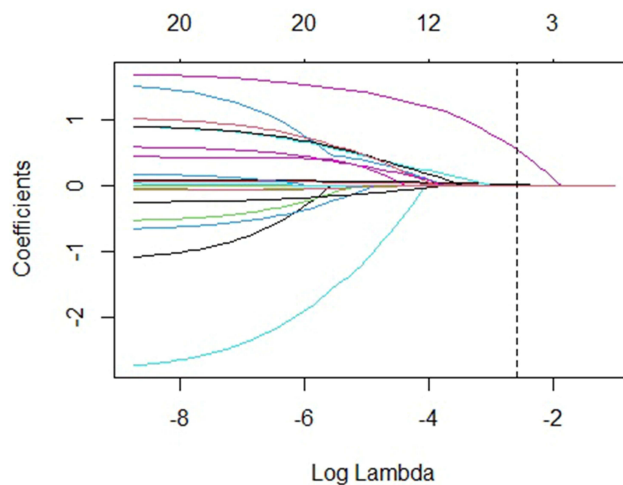
Variable	Regression Coefficient	P-value	OR	95% CI
Highest attenuation	0.0099	0.0046	1.0099	1.0036–1.0177
Mean attenuation	0.0046	0.0196	1.0046	1.0010–1.0089
Lobulation	2.3892	0.0069	10.9052	2.2043–76.7978
Vacuole	1.4948	0.0599	4.4585	0.9869–23.2938

**Abbreviations:** OR, odds ratio; CI, confidence interval.



**Figure 3** Feature selection in LASSO logistic regression. Tuning parameter  $\lambda$  selection by 10-fold cross-validation with minimum criteria. Binomial deviance (y-axis) was plotted versus  $\log(\lambda)$  (x-axis). The dotted vertical lines were drawn at the optimal value of  $\lambda$ , where the model provides the best fit to the data. The optimal value of  $\lambda$  was 0.062, and the corresponding value of  $\log(\lambda)$  was  $-2.778$ .

**Abbreviation:** LASSO, least absolute shrinkage and selection operator.



**Figure 4** Coefficient profiles determined through LASSO logistic regression analysis of features. At  $\log(\lambda)=-3$ , a dotted vertical line was drawn, including 14 optimal features with non-zero coefficients.

**Abbreviation:** LASSO, least absolute shrinkage and selection operator.

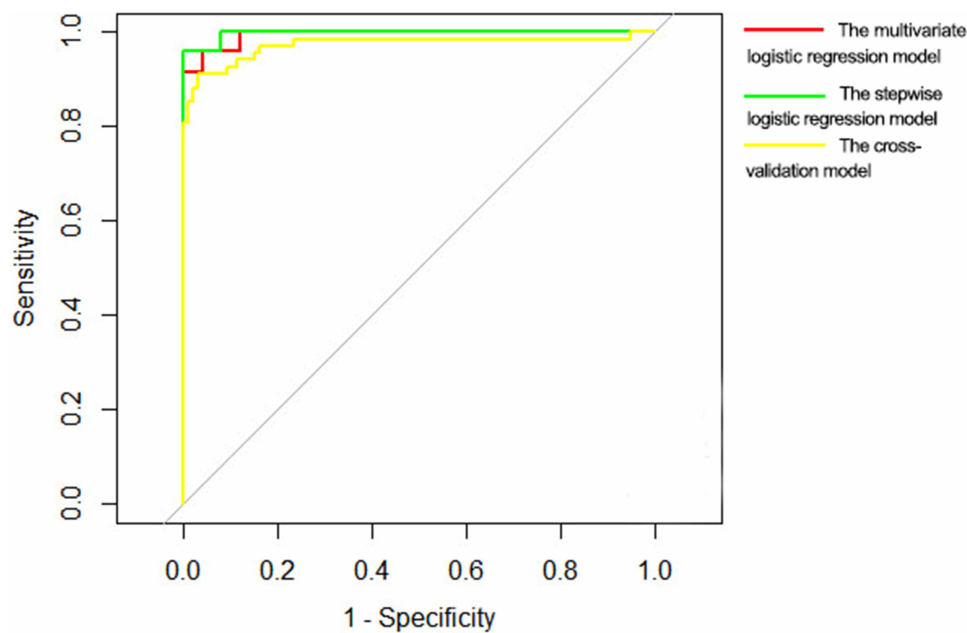
surgery, then the surgical procedure can be determined before surgery, and intraoperative freezing pathology can be avoided during the operation, thereby reducing the operation time.

On the other hand, this study showed that IPA is more largely related to attenuations, lobulation and vacuole on CT, while less affected by clinical characteristics and tumor markers. This model can help thoracic surgeons choose more

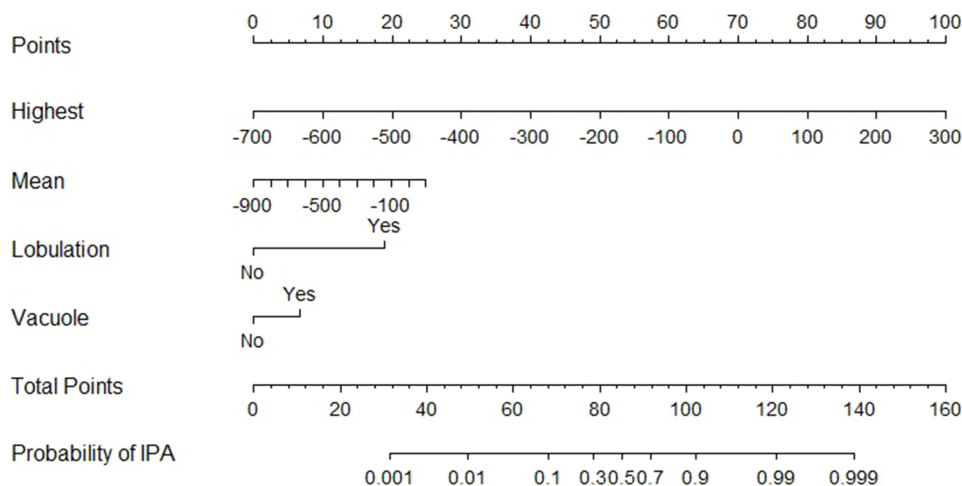
**Table 5** Comparison and Validation of Different Models

Model	AUC	Sensitivity	Specificity	95% CI	Cut-Off Value
Multivariate logistic regression model	0.9933	0.9583	0.9600	0.9805–1	-0.4565
Stepwise logistic regression model	0.9967	0.9583	1.0000	0.9887–1	0.8185
Cross-validation model	0.9731	0.9118	0.9697	0.9437–1	0.0952

**Abbreviations:** AUC, area under the curve; CI, confidence interval.

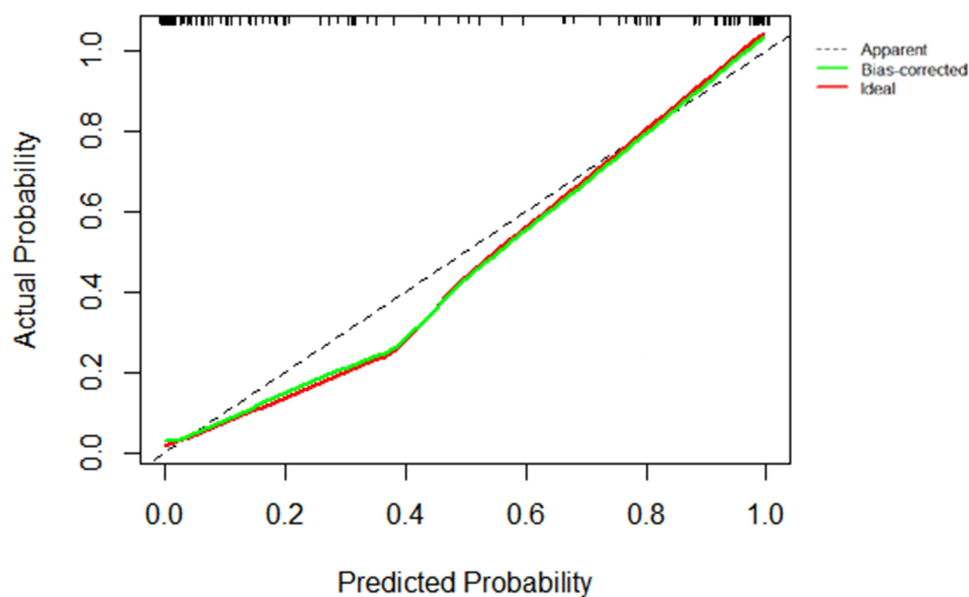


**Figure 5** Area under the curve (AUC) comparison among different models. The red line represents the multivariate logistic regression model. The green line represents the stepwise logistic regression model. The yellow line represents the cross-validation model.



**Figure 6** Nomogram for predicting IPA. To obtain the nomogram-predicted probability, the patient’s values are located at each axis, and a vertical line is drawn to the “Points” axis to determine the number of points attributed to each variable, obtain the total number of points for all variables, and derive the sum on the “Total Points” line to assess the individual probability of IPA.

reasonable management methods, such as determining the patient’s CT follow-up interval, selecting appropriate surgical method, and developing affordable treatments. According to the National Comprehensive Cancer Network guidelines, high-risk patients should continue undergoing CT scanning in stable condition until they no longer need the final treatment, eg, surgery, chemotherapy and radiotherapy.<sup>20</sup> For patients with stage I–II NSCLC, medical history collection, chest CT and physical examination should be conducted every 6 months for 2 to 3 years after surgery, and once a year thereafter. However, a Phase III randomized controlled clinical trial (IFCT-0302) used a different approach. In the first two years, patients were followed up every six months, then yearly until the fifth year. During each follow-up examination, patients underwent clinical examination, chest and abdominal CT scanning and bronchoscopy.<sup>21</sup> Therefore, accurate prediction of invasive adenocarcinoma in SPNs could help effectively develop optimal follow-up



**Figure 7** Calibration curve reflecting the relationship between predicted and actual probabilities. The x-axis shows the nomogram-predicted probability and the y-axis depicts the actual probability of malignancy. Perfect prediction would correspond to the 45° dashed line. Red line, entire cohort; green line, bias corrected by bootstrapping (B = 200 repetitions), indicating observed nomogram performance.

plans for patients. The nomogram developed in this study can reasonably help clinicians strengthen the hierarchical management of patients.

The present results indicate that attenuation, lobulation and vacuole of the SPN can well predict the severity of lung adenocarcinoma, which reflects tumor transformation from heterogeneous to homogeneous type. As the disease progresses, malignant tumor cells are more likely to diffuse into the lung interstitium and produce irregular lobulation. At the same time, there are vacuoles with more tumor necrotic and mucinous annular cavity structures in the lesion.

Compared with the published studies concentrating on predicting risk factors IPA, our study not only screened variables by using a variety of statistical methods, but we also discussed the effect of predicting the malignant degree of lung cancer on surgical approach.<sup>6,7</sup> There are following advantages on this study: Firstly, the predicting model can develop optimal follow-up plans for patients and reduce the effect of radiation to some extent. Secondly, it can make the surgical plan in advance.

This study had several limitations. First, it was a single-center retrospective study, and only lesions that underwent surgery were chosen, indicating selection bias. The accuracy of the developed predictive model should be confirmed in multicenter prospective studies, which could enable the use of the current nomogram as a clinical tool for predicting IPA. Secondly, the imaging data extracted in this study are all manually measured, which is labor-intensive and time-consuming; if an automated method for extracting imaging data can be found, a lot of manpower and time costs can be reduced. Thirdly, the sample was small in the growth nomogram study and needed external validation; more importantly, more follow-up samples are needed to develop a 5-year growth predictive nomogram for SPNs.

Factors such as smoking history, malignancy history and family history were not included. This may be related to the low positive rate in the selected patients. Previous evidence suggests that the likelihood of malignancy increases with nodule diameter.<sup>22</sup> This study did not include such factors due to increasing health awareness of the public in recent years with fewer patients with lung cancer developing to the next stage. Most tumor markers had less clinical significance in this study, which in part reflects the early stage of lesions in most patients.

## Conclusion

In conclusion, we have established and validated a novel and easy-to-use nomogram for predicting the risk of invasive adenocarcinoma in patients with SPNs. The model can accurately and effectively identify invasive adenocarcinomas and help thoracic surgeons to accurately develop individualized treatment programs for patients.

## Abbreviations

CT, Computed Tomography; SPN, solitary pulmonary nodule; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IPA, invasive pulmonary adenocarcinoma; HU, Hounsfield unit; BMI, body mass index; CEA, carcinoembryonic antigen; SCC, squamous cell carcinoma antigen; NSE, neuron specific enolase; CYFRA21-1, cytokeratin 19 fragment; ROI, regions of interest; NSCLC, non-small cell lung cancer; AUC, area under the curve; AIC, the Akaike information criterion; LASSO, the least absolute shrinkage and selection operator; ROC, receiver operating characteristic; CI, confidence interval; OR, odds ratio.

## Acknowledgments

This work was supported by governmental funding for specialist capacity building and specialist leader training projects of department of finance and health commission of hebei province (No. [2018]34).

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. 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. doi:10.1002/jso.24704
2. Wu YJ, Liu YC, Liao CY, Tang EK, Wu FZ. A comparative study to evaluate CT-based semantic and radiomic features in preoperative diagnosis of invasive pulmonary adenocarcinomas manifesting as subsolid nodules. *Sci Rep*. 2021;11(1):66. doi:10.1038/s41598-020-79690-4
3. 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. doi:10.1186/s40644-020-00320-3
4. Tang EK, Chen CS, Wu CC, et al. Natural history of persistent pulmonary subsolid nodules: long-term observation of different interval growth. *Heart Lung Circ*. 2019;28(11):1747–1754. doi:10.1016/j.hlc.2018.08.015
5. Hsu HT, Tang EK, Wu MT, et al. Modified lung-RADS improves performance of screening LDCT in a population with high prevalence of non-smoking-related lung cancer. *Acad Radiol*. 2018;25(10):1240–1251. doi:10.1016/j.acra.2018.01.012
6. Wu FZ, Chen PA, Wu CC, et al. Semiquantitative visual assessment of sub-solid pulmonary nodules  $\leq 3$  cm in differentiation of lung adenocarcinoma spectrum. *Sci Rep*. 2017;7(1):15790. doi:10.1038/s41598-017-16042-9
7. Wu FZ, Huang YL, Wu CC, et al. Assessment of selection criteria for low-dose lung screening CT among Asian ethnic groups in Taiwan: from mass screening to specific risk-based screening for non-smoker lung cancer. *Clin Lung Cancer*. 2016;17(5):e45–e56. doi:10.1016/j.clcc.2016.03.004
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: executive summary. *Proc Am Thorac Soc*. 2011;8(5):381–385. doi:10.1513/pats.201107-042ST
9. Wang L, Shen W, Xi Y, Liu S, Zheng D, Jin C. Nomogram for predicting the risk of invasive pulmonary adenocarcinoma for pure ground-glass nodules. *Ann Thorac Surg*. 2018;105(4):1058–1064. doi:10.1016/j.athoracsur.2017.11.012
10. Huang L, Lin W, Xie D, et al. Development and validation of a preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule: a machine learning approach, multicenter, diagnostic study. *Eur Radiol*. 2021;32(3):1983–1996.
11. Feng B, Chen X, Chen Y, et al. Differentiating minimally invasive and invasive adenocarcinomas in patients with solitary sub-solid pulmonary nodules with a radiomics nomogram. *Clin Radiol*. 2019;74(7):570e571–570 e511. doi:10.1016/j.crad.2019.03.018
12. Qiu T, Ru X, Yin K, Yu J, Song Y, Wu J. Two nomograms based on CT features to predict tumor invasiveness of pulmonary adenocarcinoma and growth in pure GGN: a retrospective analysis. *Jpn J Radiol*. 2020;38(8):761–770. doi:10.1007/s11604-020-00957-x
13. Okamura K, Takayama K, Izumi M, Harada T, Furuyama K, Nakanishi Y. Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer. *Lung Cancer*. 2013;80(1):45–49. doi:10.1016/j.lungcan.2013.01.002
14. Primrose JN, Perera R, Gray A, et al. Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: the FACS randomized clinical trial. *JAMA*. 2014;311(3):263–270. doi:10.1001/jama.2013.285718
15. Yang Q, Zhang P, Wu R, Lu K, Zhou H. Identifying the best marker combination in CEA, CA125, CY211, NSE, and SCC for lung cancer screening by combining ROC curve and logistic regression analyses: Is it feasible? *Dis Markers*. 2018;2018:2082840. doi:10.1155/2018/2082840
16. Song X, Wang S, Wang Z, Cao WJ, Chen FJN. Serum CYFRA21-1 as an effective tumor biomarker for patients with nasopharyngeal carcinoma. *Neoplasma*. 2015;62(1):124–129. doi:10.4149/neo\_2015\_016
17. Liu J, Xu H, Qing H, et al. Comparison of radiomic models based on low-dose and standard-dose CT for prediction of adenocarcinomas and benign lesions in solid pulmonary nodules. *Front Oncol*. 2020;10:634298. doi:10.3389/fonc.2020.634298
18. Feng B, Chen X, Chen Y, et al. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. *Eur Radiol*. 2020;30(12):6497–6507. doi:10.1007/s00330-020-07024-z
19. Suzuki K, Saji H, Aokage K, et al. Comparison of pulmonary segmentectomy and lobectomy: safety results of a randomized trial. *J Thorac Cardiovasc Surg*. 2019;158(3):895–907. doi:10.1016/j.jtcvs.2019.03.090
20. Duan XQ, Wang XL, Zhang LF, et al. Establishment and validation of a prediction model for the probability of malignancy in solid solitary pulmonary nodules in northwest China. *J Surg Oncol*. 2021;123(4):1134–1143. doi:10.1002/jso.26356
21. Liu G, Xu Z, Zhang Y, et al. Machine-learning-derived nomogram based on 3D radiomic features and clinical factors predicts progression-free survival in lung adenocarcinoma. *Front Oncol*. 2021;11:692329. doi:10.3389/fonc.2021.692329

22. Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with bayesian analysis. Part II. Application. *Radiology*. 1993;186(2):415–422. doi:10.1148/radiology.186.2.8421744

Cancer Management and Research

Dovepress

### Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>