Construction of a risk prediction model for isolated pulmonary nodules 5–15 mm in diameter

Siting Xie[1,2#], Xingguang Luo[3#], Yuxin Guo[1#], Xiulian Huang[1], Jinyu Long[4], Ying Chen[5], Ping Lin[5], Jinhe Xu[1], Shangwen Xu[6], Chunlei Zhao[6], Baoquan Lin[7], Chunxia Su[8], Nagarashee Seetharamu[9], Duilio Divisi[10], Mingliang Jin[5], Zongyang Yu[11]

[1]Fuzong Clinical Medical College, Fujian Medical University, Fuzhou, China; [2]Respiratory Department of Xiamen Hongai Hospital, Xiamen, China; [3]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA; [4]College of Rehabilitation Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, China; [5]Department of Respiratory and Critical Care Medicine, The 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China, Fuzhou, China; [6]Department of Medical Imaging, Fuzhou General Hospital of Fujian Medical University, Dongfang Hospital of Xiamen University, The 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China, Fuzhou, China; [7]Department of cardiothoracic surgery, Fuzhou General Hospital of Fujian Medical University, Dongfang Hospital of Xiamen University, The 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China, Fuzhou, China; [8]Department of Medical Oncology, Shanghai Pulmonary Hospital and Thoracic Cancer Institute, Tongji University School of Medicine, Shanghai, China; [9]Division of Medical Oncology and Hematology, Northwell Health Cancer Institute, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Lake Success, NY, USA; [10]Department of Life, Health and Environmental Sciences, Thoracic Surgery Unit, University of L'Aquila, L'Aquila, Italy; [11]Department of Pulmonary and Critical Care Medicine, Fuzhou General Hospital of Fujian Medical University, Dongfang Hospital of Xiamen University, The 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China, Fuzhou, China

Contributions: (I) Conception and design: Z Yu, M Jin; (II) Administrative support: Z Yu; (III) Provision of study materials or patients: Z Yu, M Jin; (IV) Collection and assembly of data: X Huang, J Long, Y Chen, P Lin, J Xu, S Xu, C Zhao, B Lin, C Su; (V) Data analysis and interpretation: S Xie, X Luo, Y Guo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Zongyang Yu, MD. Department of Pulmonary and Critical Care Medicine, Fuzhou General Hospital of Fujian Medical University, Dongfang Hospital of Xiamen University, The 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China, 156 Xierhuan North Road, Fuzhou 350025, China. Email: yuzy527@sina.com; Mingliang Jin, MD. Department of Respiratory and Critical Care Medicine, The 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China, 156 Xierhuan North Road, Fuzhou 350025, China. Email: roupingjml@163.com.

Background: Based on current technology, the accuracy of detecting malignancy in solitary pulmonary nodules (SPNs) is limited. This study aimed to establish a malignant risk prediction model for SPNs 5–15 mm in diameter.

Methods: We collected clinical characteristics and imaging features from 317 patients with SPNs 5–15 mm in diameter from the 900th Hospital of the Joint Logistic Support Force as a training cohort and 100 patients with SPNs 5–15 mm in diameter as a validation cohort. Univariate logistic regression analysis, least absolute shrinkage and selection operator (LASSO), and binary logistic regression analysis were used to screen for the independent influencing factors of benign and malignant SPN and to establish a prediction model for benign and malignant SPN with a diameter of 5–15 mm. The model in this study was compared with the Mayo model, Veterans Affairs (VA) model, Brock model, and Peking University People's Hospital (PKUPH) model. Finally, the clinical application value of this model was assessed.

Results: Univariate logistic regression analysis showed that smoking history, nodule diameter, nodule location, nodule density, margin, calcification, lobulation sign, spiculation sign, and vascular cluster sign were statistically significant factors. The results of LASSO and binary logistic regression analysis showed that smoking history, nodule diameter, nodule density, margin, lobulation sign, and vascular cluster sign were independent influencing factors of SPNs. The prediction model was successfully constructed and demonstrated a good predictive performance, with an area under the curve (AUC) value of 0.814 [95% confidence interval (CI): 0.768–0.861; P<0.001] in the training cohort and 0.864 (95% CI: 0.794–0.934;

P<0.001) in the validation cohort. This model was shown to be highly accurate in predicting malignant SPNs and thus has a high clinical application value. Compared with previously described prediction models, including the Mayo model, VA model, Brock model, and PKUPH model, the proposed model demonstrated a significantly superior predictive ability.

Conclusions: The prediction model developed in this study can be used as an early screening method for SPNs 5–15 mm in diameter.

Keywords: Lung cancer; isolated pulmonary nodule; prediction model; independent influencing factors

Introduction

Lung cancer is associated with the highest cancer-related morbidity and mortality rate globally. Most patients with lung cancer are diagnosed at advanced stages, which despite significant scientific advances, is still associated with a 5-year survival rate (5YSR) of only 20% (1). Overall survival is stage-dependent, with 5YSR for stages I and II lung cancers being considerably better than for stages III or IV lung cancers. Specifically, 5YSR is 82% for stage IA and only 2.9% for stage IVB (2). Therefore, screening, early detection, and timely intervention can result in significantly improved outcomes and decrease risk of death amongst individuals with high-risk for developing lung cancer. According to the National Lung Screening Trial (NLST), low-dose computed tomography (LDCT) can reduce lung cancer mortality by 20% (3). The Dutch-Belgian Lung Cancer Screening Trial (NELSON) reported a 24–33% reduction in cumulative lung cancer mortality after 10 years in the screening group (4). Therefore, LDCT is recommended for individuals at high risk of lung cancer as a screening test by various organizations including U.S. Preventive Services Task Force (USPSTF), National Comprehensive Care Network (NCCN), and China National Lung Cancer Screening guidelines.

With the widespread use of LDCT, the detection rate of pulmonary nodules has increased, but the high false-positive rate has precluded wide-spread uptake of LDCT in several communities. In the NLST study, 96.4% of the detected pulmonary nodules were false positives (4), suggesting that the majority of solitary pulmonary nodules (SPNs) are benign. To better differentiate between benign and malignant SPNs, various malignant prediction models for SPN have been established. Internationally recognized models include the Mayo model (5), Veterans Affairs (VA) model (6), Brock model (7), Pan-Can model (8), and Peking University People's Hospital (PKUPH) model (9). However, these lung cancer prediction models exhibit variable predictive accuracies and differences in real-world applicability due to

---

Highlight box

Key findings
- This study aimed to establish a malignant risk prediction model for solitary pulmonary nodules (SPNs) 5–15 mm in size.

What is known and what is new?
- To better differentiate between benign and malignant SPNs, various malignant prediction models for SPNs have been established. However, these lung cancer prediction models exhibit variable predictive accuracies and differences in clinical applicability due to variations in populations, sample sizes, and influencing factors in the studies that led to these models.
- We used univariate logistic regression analysis, least absolute shrinkage and selection operator (LASSO), and binary logistic regression analysis to screen for the independent influencing factors of benign and malignant SPNs. Univariate logistic regression analysis showed that smoking history, nodule diameter, nodule location, nodule density, boundary, calcification, lobulated sign, spiculation sign, and vascular cluster sign were statistically significant factors. The results of LASSO and binary logistic regression analysis showed that smoking history, nodule diameter, nodule density, boundary, lobulated sign, and vascular cluster sign were independent influencing factors of SPNs. Moreover, the predictive value of this model was significantly better than that of the Mayo model, Veterans Affairs (VA) model, Brock model, and Peking University People's Hospital (PKUPH) model.

What is the implication, and what should change now?
- This prediction model, serving as an early screening method for SPNs with a diameter of 5–15 mm, demonstrates superior prediction efficacy and greater clinical applicability when compared to the Mayo model, VA model, Brock model, and PKUPH model.

variations in populations, sample sizes, and risk factors used in these prediction models.

Previous studies have indicated a significant correlation between the diameter of the SPN and the degree of malignancy, with larger diameters associated with higher degrees of malignancy (10,11). The 2017 Fleischner guideline eliminated the need for routine follow-up for SPNs <6 mm (12). As per the Chinese lung nodule guidelines (13), nonsolid SPNs with a diameter of 5–15 mm and exhibiting no obvious malignant computed tomography (CT) signs are categorized as intermediate-risk nodules. Solid SPNs with a diameter of 8–15 mm and/or with malignant CT signs are defined as high-risk nodules. The guidelines for triaging SPNs with diameters ranging from 5 to 15 mm are unclear and identifying malignant potential of these nodules is crucial for determining the need for further clinical intervention.

To enhance the accuracy of clinicians' diagnosis of benign and malignant SPNs with diameters ranging from 5 to 15 mm, we constructed a malignancy probability prediction model based on clinical and CT image characteristics. We present this article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-785/rc).

Methods

We enrolled 417 patients with SPNs measuring 5–15 mm in diameter who underwent surgical treatment at the 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China. The included pulmonary nodules were randomly assigned to a training set (n=317) and a validation set (n=100) in a 3:1 ratio. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Institutional Ethics Committee of the 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China (approval No. 2021-026) and informed consent was taken from all the patients. The inclusion criteria were as follows: (I) age over 18 years; (II) pulmonary nodules on LDCT; (III) surgical treatment with clear pathological results; (IV) time between CT examination and operation of 1 month or less; (V) nodules between 5 and 15 mm in diameter; (VI) complete clinical and imaging data available; and (VII) signing of informed consent form before surgery. Meanwhile, the exclusion criteria were as follows: (I) pulmonary nodules treated before surgery; including chemotherapy, immunotherapy, targeted, or

surgical regimens; (II) pathologically confirmed metastatic pulmonary nodules; and (III) presence of metastases to other sites. We collected data from medical records, including clinical characteristics, imaging features, and pathological findings.

Statistical analysis

The data were analyzed using the software SPSS 26.0 (IBM Corp., Armonk, NY, USA) and R v. 4.2.1 (The R Foundation for Statistical Computing, Vienna, Austria). The t-test was applied for measurement data conforming to a normal distribution, expressed as mean ± standard deviation (SD), whereas the Mann-Whitney test was used for nonnormally distributed data, expressed as the median and interquartile range. The chi-square test was used for counting data, expressed as frequency number and percentage [n (%)]. The clinical data and CT imaging characteristics of patients in the training group were analyzed through univariate logistic regression analysis to identify statistically significant influencing factors for differentiating between the benign and malignant groups SPNs with a diameter of 5–15 mm. Least absolute shrinkage and selection operator (LASSO) analysis was used to select the optimal predictors associated with lung cancer risk from variables with P<0.1 in the above-mentioned univariate logistic regression analysis. The final prediction model was constructed using binary logistic regression analysis. Receiver operating characteristic (ROC) curves were drawn, and the area under the ROC curve (AUC), truncation value, Youden index, sensitivity, specificity, positive predictive value, negative predictive value, accuracy, model forest plot, and nomogram were plotted.

Verification group data were used to plot ROC curves for the proposed model, Mayo model, PKUPH model, VA model, and Brock model, and AUC values were calculated (P<0.05 was considered a statistically significant difference). The predictive efficacy of these five models for benign and malignant SPN with a diameter of 5–15 mm was analyzed and compared. The calibration of the models was examined using the calibration curve. The clinical value of the prediction models was evaluated via decision curve analysis (DCA).

Results

Patient characteristics

The study included 417 patients with SPN who visited

3142

Xie et al. Lung cancer in high-risk nodules

the 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China between January 2015 and January 2023. In the training cohort, patients were categorized into a benign nodule group (80 males and 81 females; mean age 52 years; range, 44–61 years) and a malignant nodule group (75 males and 81 females; mean age 55 years; range, 47–64 years). Similarly, patients in the testing cohort were separated into a benign nodule group (25 males and 26 females; mean age 53 years; range, 44–65 years) and a malignant nodule group (17 males and 32 females; mean age 55 years; range, 49–63 years). *Table 1* displays the clinical characteristics and imaging features of the participants. *Table 2* displays the assignment of categorical variables associated with malignant SPNs.

In this study, the final pathological results of postoperative specimens were used as the gold standard to judge nodule properties. The pathological results for the 417 patients included in the study are shown in *Figure 1*.

Table 1 Clinical and computed tomography imaging characteristics of patients in the training and validation cohorts

Variables	Training cohort (N=317)			Validation cohort (N=100)		
	Benign (n=161)	Malignant (n=156)	P value	Benign (n=51)	Malignant (n=49)	P value
Age (years)	52.00 (44.00, 61.00)	55.00 (47.00, 64.00)	0.045	53.00 (44.00, 65.00)	55.00 (49.00, 63.00)	0.77
Longest diameter (mm)	9.00 (7.00, 12.00)	11.00 (9.00, 12.00)	0.002	10.00 (8.00, 11.00)	11.00 (9.00, 13.00)	0.12
Shortest diameter (mm)	8.00 (6.00, 10.00)	8.00 (7.00, 10.00)	0.04	8.00 (6.00, 9.00)	9.00 (7.00, 10.00)	0.03
Gender			0.77			0.15
Male	80 (49.69)	75 (48.08)		25 (49.02)	17 (34.69)	
Female	81 (50.31)	81 (51.92)		26 (50.98)	32 (65.31)	
Smoking history			0.03			0.42
Yes	15 (9.32)	28 (17.95)		8 (15.69)	5 (10.20)	
No	146 (90.68)	128 (82.05)		43 (84.31)	44 (89.80)	
Alcohol history			0.11			0.50
Yes	8 (4.97)	15 (9.62)		5 (9.80)	3 (6.12)	
No	153 (95.03)	141 (90.38)		46 (90.20)	46 (93.88)	
Personal cancer history			0.07			0.68
Yes	5 (3.11)	12 (7.69)		4 (7.84)	5 (10.20)	
No	156 (96.89)	144 (92.31)		47 (92.16)	44 (89.80)	
Family cancer history			0.77			0.98
Yes	11 (6.83)	12 (7.69)		1 (1.96)	1 (2.04)	
No	150 (93.17)	144 (92.31)		50 (98.04)	48 (97.96)	
Hypertension			0.91			0.93
Yes	24 (14.91)	24 (15.38)		9 (17.65)	9 (18.37)	
No	137 (85.09)	132 (84.62)		42 (82.35)	40 (81.63)	
Respiratory disease			0.98			-
Yes	2 (1.24)	2 (1.28)		0 (0.00)	0 (0.00)	
No	159 (98.76)	154 (98.72)		51 (100.00)	49 (100.00)	

Table 1 (*continued*)

Table 1 (*continued*)

Variables	Training cohort (N=317)			Validation cohort (N=100)		
	Benign (n=161)	Malignant (n=156)	P value	Benign (n=51)	Malignant (n=49)	P value
Location			0.08			0.06
Right upper lobe	48 (29.81)	59 (37.82)		22 (43.14)	20 (40.81)	
Right middle lobe	26 (16.15)	12 (7.69)		4 (7.84)	8 (16.33)	
Right lower lobe	35 (21.74)	27 (17.31)		5 (9.80)	11 (22.45)	
Left upper lobe	33 (20.50)	32 (20.51)		13 (25.49)	9 (18.37)	
Left lower lobe	19 (11.80)	26 (16.67)		7 (13.73)	1 (2.04)	
Density			<0.001			<0.001
Solid	85 (52.79)	42 (26.92)		29 (56.86)	9 (18.37)	
Partial solid	23 (14.29)	66 (42.31)		9 (17.65)	23 (46.94)	
Ground glass	53 (32.92)	48 (30.77)		13 (25.49)	17 (34.69)	
Boundary			<0.001			<0.001
Clear	96 (59.63)	44 (28.21)		38 (74.51)	9 (18.37)	
Blurred	65 (40.37)	112 (71.79)		13 (25.49)	40 (81.63)	
Calcification			0.02			0.045
Yes	8 (4.97)	1 (0.64)		2 (3.92)	0 (0.00)	
No	153 (95.03)	155 (99.36)		49 (96.08)	49 (100.00)	
Lobulated sign			<0.001			0.008
Yes	39 (24.22)	80 (51.28)		12 (23.53)	24 (48.98)	
No	122 (75.78)	76 (48.72)		39 (76.47)	25 (51.02)	
Spiculation sign			<0.001			0.003
Yes	43 (26.71)	73 (46.79)		36 (70.59)	20 (40.82)	
No	118 (73.29)	83 (53.21)		15 (29.41)	29 (59.18)	
Vacuole sign			0.78			0.46
Yes	15 (9.32)	16 (10.26)		4 (7.84)	6 (12.24)	
No	146 (90.68)	140 (89.74)		47 (92.16)	43 (87.76)	
Pleural indentation sign			0.17			0.03
Yes	29 (18.01)	38 (24.36)		5 (9.80)	13 (26.53)	
No	132 (81.99)	118 (75.64)		46 (90.20)	36 (73.47)	
Air bronchogram			0.86			0.07
Yes	9 (5.59)	8 (5.13)		2 (3.92)	7 (14.29)	
No	152 (94.41)	148 (94.87)		49 (96.08)	42 (85.71)	
Vascular cluster sign			<0.001			0.009
Yes	4 (2.48)	29 (18.59)		3 (5.88)	12 (24.49)	
No	157 (97.52)	127 (81.41)		48 (94.12)	37 (75.51)	

Data are presented as median (interquartile range) or n (%).

Table 2 Assignment of categorical variables associated with malignant solitary pulmonary nodules

Categorical variable	Assignment condition
Smoking history	Yes equals 1; no equals 0
Personal cancer history	Yes equals 1; no equals 0
Density	Solid equals 0; ground glass equals 1; partial solid equals 2
Boundary	Clear equals 1; blur equals 0
Calcification	Yes equals 1; no equals 0
Lobulated sign	Yes equals 1; no equals 0
Vascular cluster sign	Yes equals 1; no equals 0

Construction of the prediction model

Multivariate logistic regression analysis of malignant SPNs

We used univariate logistic regression analysis to examine the clinical characteristics and imaging features of patients in the training cohort, identifying nine statistically significant influencing factors for differentiating between benign and malignant SPNs 5–15 mm in diameter (*Figure 2*). LASSO regression analysis, which was based on variables with P<0.1 in the univariate logistic regression analysis, identified a total of nine factors with nonzero coefficients in the training cohort (*Figure 3A,3B*). Binary logistic regression analysis was
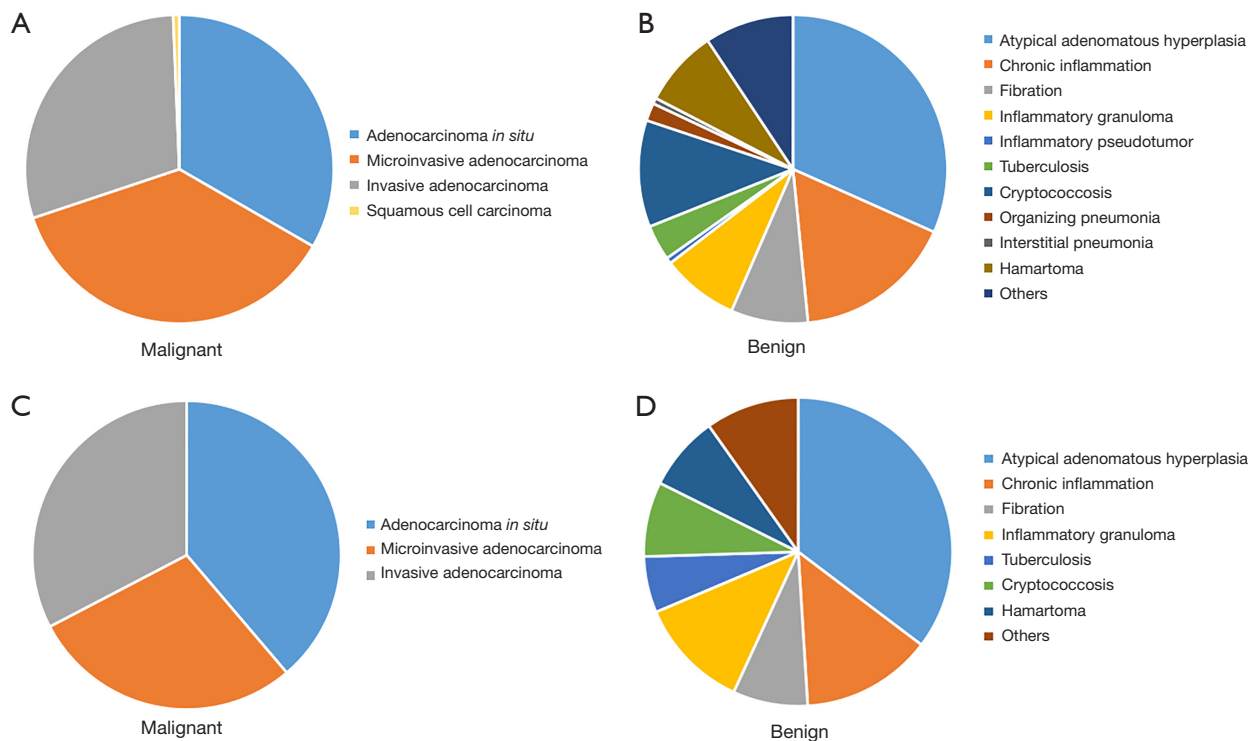
Figure 1 Pathological types of patients. (A) Malignant nodules in the training group. (B) Benign nodules in the training group. The category "Others" includes immunoglobin 4-related disease, lymphangioleiomyomatosis, Epstein-Barr virus-associated smooth muscle tumors, pulmonary leiomyoma, inflammatory myofibroblastic tumor, bronchiolar adenoma, spindle cell tumor, pulmonary sclerosing hemangioma, and silicotic nodules. (C) Malignant nodules in the validation group. (D) Benign nodules in the validation group. The category "Others" includes bronchiolar adenoma, pulmonary sclerosing hemangioma, lymph node, and silicotic nodules.

Characteristics	OR (95%CI)		P
Age	1.018(0.999,1.038)		0.06
Gender			
Male	0.938(0.603,1.456)		0.77
Smoking history			
Yes	2.129(1.089,4.163)		0.03
Alcohol history			
Yes	2.035(0.837,4.945)		0.12
Hypertension			
Yes	1.038(0.562,1.918)		0.91
Respiratory disease			
Yes	1.032(0.144,7.422)		0.98
Personal cancer history			
Yes	2.6(0.894,7.562)		0.08
Family cancer history			
Yes	1.136(0.486,2.657)		0.77
Longest diameter	1.132(1.045,1.226)		0.002
Shortest diameter	1.087(0.997,1.186)		0.059
Location			
Right middle lobe	0.375(0.172,0.822)		0.01
Right lower lobe	0.628(0.334,1.179)		0.15
Left upper lobe	0.789(0.425,1.463)		0.45
Left lower lobe	1.113(0.551,2.250)		0.77
Density			
Partial solid	5.807(3.182,10.598)		0
Ground glass	1.833(1.071,3.138)		0.03
Boundary			
Clear	0.266(0.166,0.426)		0
Calcification			
Yes	0.123(0.015,0.998)		0.050
Lobulated sign			
Yes	3.293(2.041,5.311)		0
Spiculation sign			
Yes	2.414(1.509,3.861)		0
Vacuole sign			
Yes	1.112(0.530,2.335)		0.78
Pleural indentation sign			
Yes	1.466(0.851,2.524)		0.17
Air bronchogram			
Yes	0.913(0.343,2.430)		0.86
Vascular cluster sign			
Yes	8.963(3.070,26.161)		0

Figure 2 Univariate logistic regression analysis and forest plot of solitary pulmonary nodules in the training cohort. The solid line is the invalid line. The risk factors appear on the right side of the invalid line is, and the protective factors appear on the left side. OR, odds ratio; CI, confidence interval.

then employed to screen six independent influencing factors of SPN, including smoking history, nodule diameter, nodule density, boundary, lobulated sign, and vascular cluster sign (*Figure 4*).
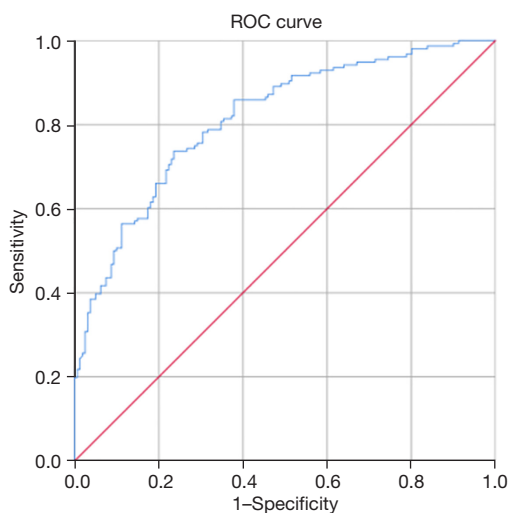
Construction of the prediction model

Based on the six independent influencing factors of SPNs, we constructed a prediction model with good performance, which demonstrated an AUC value of 0.814 [95% confidence interval (CI): 0.768–0.861; P<0.001] in the training cohort (*Figure 5*) and 0.864 (95% CI: 0.794–0.934;

P<0.001) in the validation cohort (*Figure 6*).

Assessment and comparison of the models' predictive performance

The calibration curves and the ideal curve of the predicted model in the training cohort (*Figure 7A*) and validation cohort (*Figure 7B*) showed a high degree of coincidence, indicating accurate predictions. For the Mayo model, PKUPH model, VA model, and Brock model, we compared the different parameters of various prediction models

3146

Xie et al. Lung cancer in high-risk nodules
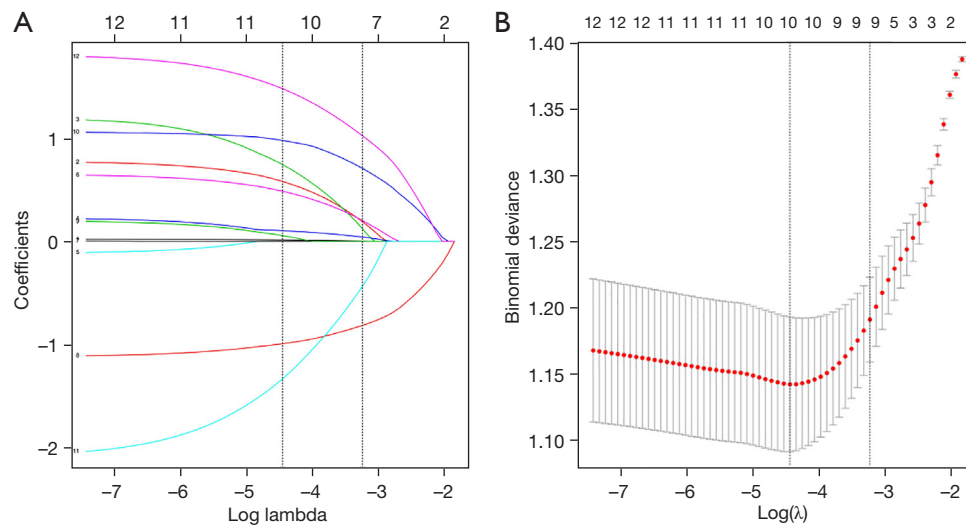
Figure 3 Selection of characteristics from LASSO regression. (A) Curves for the coefficient paths of the 12 features in the training cohort. The vertical black line defines the optimal value of λ for the 10-fold cross-validation in *Figure 2B*, with nine features with nonzero coefficients obtained with λ equal to 0.039. (B) Selection of the regularization parameter λ in the LASSO regression via 10-fold cross-validation based on a standard error of the minimum mean square error of distance. LASSO, least absolute shrinkage and selection operator.

Characteristics	β	OR (95%CI)	P
Age	0.018	1.018 (0.994,1.042)	0.14
Smoking history			
Yes	0.859	2.408 (1.094,5.298)	0.03
Personal cancer history			
Yes	0.148	3.088 (0.856,11.143)	0.09
Longest diameter	1.128	1.160 (1.044,1.289)	0.006
Density			
Partial solid	1.275	3.580 (1.758,7.291)	0
Ground glass	1.585	4.879 (2.336,10.188)	0
Boundary			
Clear	-0.892	0.410 (0.236,0.712)	0.002
Calcification			
Yes	-2.072	0.126 (0.011,1.498)	0.10
Lobulated sign			
Yes	1.138	3.12 (1.741,5.589)	0
Vascular cluster sign			
Yes	1.607	4.990 (1.517,16.414)	0.008
Constant	-3.608	0.027	0

Figure 4 Binary logistic regression analysis results and forest plot of benign and malignant nodules in the training cohort. The solid line is the invalid line. The risk factors appear on the right side of the invalid line is, and the protective factors appear on the left side. OR, odds ratio; CI, confidence interval.

(*Table 3*). The AUCs were found to be 0.610 (95% CI: 0.498–0.721), 0.705 (95% CI: 0.603–0.807), 0.502 (95% CI: 0.383–0.612), and 0.677 (95% CI: 0.570–0.785), respectively. The predictive value of the proposed model was significantly higher than that of the Mayo model, VA model, Brock model, and PKUPH model (*Figure 8*).
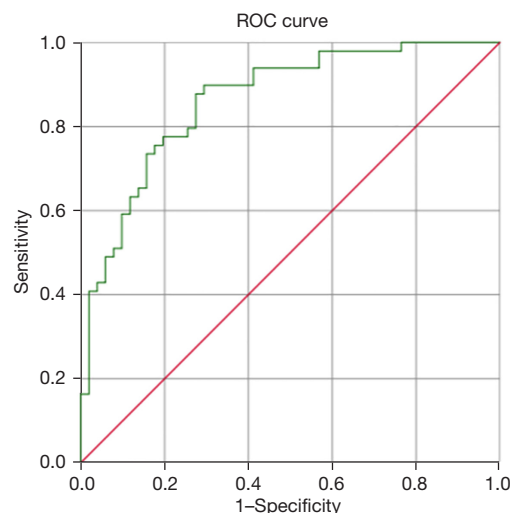
Analysis of the model's clinical utility

DCA for the predictive nomogram was performed (*Figure 9*),
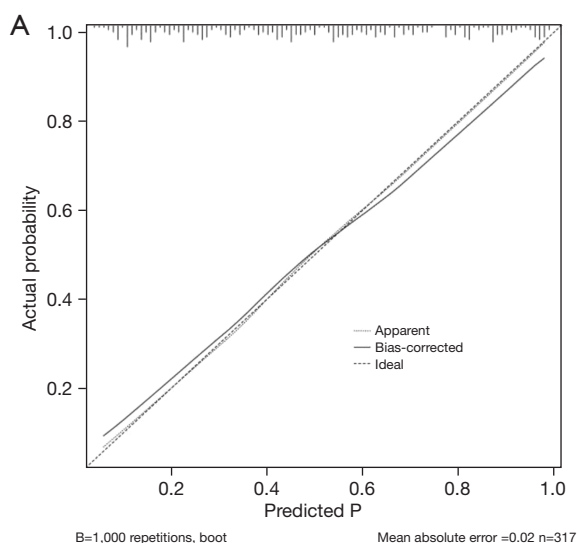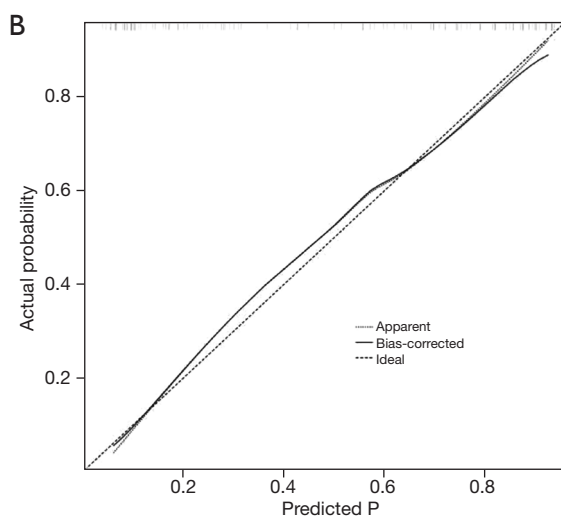
Figure 5 ROC curve of the prediction model for malignant solitary pulmonary nodules in the training cohort. ROC, receiver operating characteristic.

Figure 6 ROC curve of the prediction model for malignant solitary pulmonary nodules in the validation cohort. ROC, receiver operating characteristic.

Figure 7 Calibration curves of the predictive model for malignant single pulmonary nodules in the (A) training cohort and (B) validation cohort. The solid line represents the prediction efficacy of the model, and the closer the solid line and diagonal dashed line are, the better the prediction effect.

which suggested that the model has high clinical application value.

Discussion

The early screening and diagnosis of lung cancer are crucial steps in the reduction of the mortality associated

with this condition. Currently, LDCT is the preferred method for the early detection of SPNs, with its usage becoming increasingly widespread (13-16). Clinically, 18F-fluorodeoxyglucose positron emission tomography/CT (^{18}F-FDG-PET/CT) is also a useful tool for the qualitative detection of benign and malignant pulmonary nodules. Different from the principle of CT imaging,

Table 3 Comparison of different parameters of various prediction models

Prediction model	Sample size	Malignant rate	AUC	Limitations
The model	417	49.64%	0.814	Further expansion of the sample is necessary
Mayo model	629	23%	0.61	Definite pathological results were lacking
VA model	375	54%	0.502	VA model including older adult male smokers without considering imaging characterization
Brock model	1,871	1.4%	0.677	Brock model is suitable for predicting malignant pulmonary nodules during physical examination
PKUPH model	371	53.1%	0.705	CT selection time was uncertain during model building

AUC, area under the curve; VA, Veterans Affairs; PKUPH, Peking University People's Hospital; CT, computed tomography.
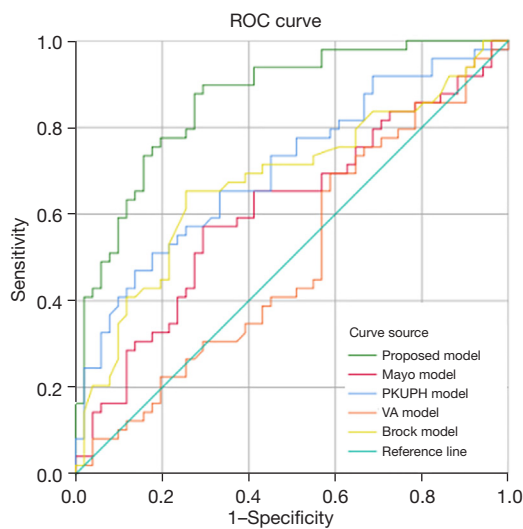
Figure 8 Comparison of verification results of the proposed model, Mayo model, PKUPH model, VA model, and Brock model according to the ROC curve. ROC, receiver operating characteristic; PKUPH, Peking University People's Hospital; VA, Veterans Affairs.

PET/CT comprehensively evaluates the benign and malignant nodules mainly by using on drugs to produce different values of the maximum standardized uptake value (SUVmax). However, international guidelines [National Comprehensive Cancer Network (NCCN), American College of Chest Physicians (ACCP), Fleischner, Asian Consensus] recommend LDCT mainly as a screening method for benign and malignant pulmonary nodules, and PET/CT is not routinely recommended. However, the high false-positive rate of LDCT has spurred the development of predictive models to efficiently and accurately discern benign and malignant SPNs. Various models, including the

Mayo model, Pan-Can model, Brock model, VA model, and PKUPH model, have been established, each with a unique applicability to certain populations and varying prediction parameters, which limit their more general application.

This study found that the VA model had a poor predictive effect, indicated by an AUC value of only 0.502. This may be attributed to the model including older adult male smokers without considering imaging characterization (6). The Mayo model yielded an AUC value of only 0.61, which can potentially be explained by the lack of clear pathological results for some patients during model establishment (5). The Brock model yielded an AUC value of 0.677, indicating average predictive performance. This result may be related to the small number of smoking patients included in this study (training cohort: 43/317, 13.6%). The PKUPH model, with an AUC value of 0.705, might be more suitable for assessing malignant risk in solid nodules, especially given that nearly two-thirds of SPNs selected for this study were subsolid nodules, with only two nodules exhibiting calcification (9).

We constructed a prediction model for malignant SPNs 5–15 mm in diameter to be used in early lung cancer screening. A total of 417 patients with SPNs 5–15 mm in diameter who underwent surgical treatment with a pathological diagnosis at the 900th Hospital of the Joint Logistic Support Force over an 8-year period were included in this study. Eight clinical characteristics and 12 imaging features were collected, and six independent influencing factors related to benign and malignant SPNs 5–15 mm in diameter were identified. The constructed prediction model demonstrated an AUC of 0.814 (95% CI: 0.768–0.861) for the training cohort (*Figure 4*) and 0.864 (95% CI: 0.794–0.934) for the validation cohort, exhibiting high sensitivity, specificity, and accuracy. This model outperformed the Mayo model, PKUPH model, VA model, and Brock model
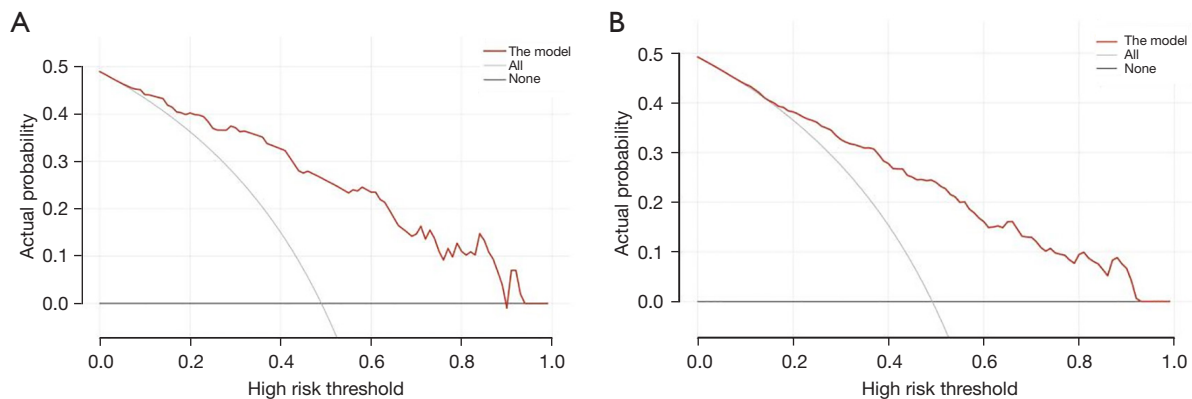
Figure 9 Decision curve analysis for the model's prediction efficacy of lung cancer in the validation cohort. The horizontal axes represent the threshold probabilities, the vertical axes represent the net benefit rates, the red line on the left represents the benefit curves of the training group, and the red line on the right represents the benefit curves of validation group. When the threshold probability is in the range of 0.01–0.95, the model can provide benefit to the classification of pulmonary nodules.

in predicting SPN malignancy.

This study observed a higher prevalence of subsolid nodules than that of solid nodules, with subsolid nodules exhibiting a significantly higher malignancy rate. Gender differences were not statistically significant and were not independent influencing factors for SPNs, aligning with the findings of several studies (6,9,17). Smoking emerged as an independent risk factor for lung cancer, which is consistent with the existing literature (18). Additionally, factors including density, size, border, lobulation, and vascular cluster sign were identified as independent risk factors for lung cancer, which is also in line with a previous study (19). Although the constructed prediction model performed well, it had certain limitations. First, as our study adopted a retrospective design, some laboratory results were unavailable and thus not included in the information collection. Second, the sample size in this study was small whereas individual differences in pulmonary nodules were large, potentially reducing the diagnostic accuracy of the model. Further expansion of the sample for verification is necessary to validate the model.

In conclusion, the prediction model for benign and malignant SPNs 5–15 mm in diameter that incorporated six independent factors including smoking history, nodule diameter, nodule density, boundary, lobulated sign, and vascular cluster sign, demonstrated low leakage and misdiagnosis rates with good accuracy. Compared with the Mayo model, VA model, Brock model, and PKUPH model, this model exhibited high predictive value as an early screening method for SPNs 5–15 mm in diameter,

providing a foundation for the diagnosis of benign and malignant SPNs of this size.

Conclusions

The proposed prediction model, built to serve as an early screening method for SPNs with a diameter of 5–15 mm, demonstrated superior prediction efficacy and greater clinical applicability as compared to the Mayo model, VA model, Brock model, and PKUPH model.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the

TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-785/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-785/dss

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-785/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-785/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. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Institutional Ethics Committee of the 900th Hospital of the Joint Logistic Support Force, People's Liberation Army of China (approval No. 2021-026) and informed consent was taken from all the patients.

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. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. Cancer Manag Res 2019;11:943-53.
3. Detterbeck FC, Chansky K, Groome P, et al. The IASLC Lung Cancer Staging Project: Methodology and Validation Used in the Development of Proposals for Revision of the Stage Classification of NSCLC in the Forthcoming

4. 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.
5. 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.
6. 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.
7. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 2013;369:910-9.
8. Tammemagi MC, Schmidt H, Martel S, et al. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. Lancet Oncol 2017;18:1523-31.
9. 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;12:313-9.
10. Bai C, Choi CM, Chu CM, et al. Evaluation of Pulmonary Nodules: Clinical Practice Consensus Guidelines for Asia. Chest 2016;150:877-93.
11. Ye Y, Sun Y, Hu J, et al. A clinical-radiological predictive model for solitary pulmonary nodules and the relationship between radiological features and pathological subtype. Clin Radiol 2024;79:e432-9.
12. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology 2017;284:228-43.
13. Zhou Q, Fan Y, Wang Y, et al. China National Guideline of Classification, Diagnosis and Treatment for Lung Nodules (2016 Version). Chinese Journal of Lung Cancer 2016:19:793-8.
14. Zhou Q, Fan Y, Wang Y, et al. China National Lung Cancer Screening Guideline with Low-dose Computed Tomography (2018 version). Chinese Journal of Lung Cancer 2018:21:67-75.
15. Zhou Q, Fan Y, Wang Y, et al. China National Lung Cancer Screening Guideline with Low-dose Computed Tomography(2023 Version). Chinese Journal of Lung

(Eighth) Edition of the TNM Classification of Lung Cancer. J Thorac Oncol 2016;11:1433-46.

Cancer 2023:26:1-9.

16. Zhou Q, Fan Y, Wu N, et al. Demonstration program of population-based lung cancer screening in China: Rationale and study design. Thorac Cancer 2014;5:197-203.

17. Liu Z, Ran H, Yu X, et al. Immunocyte count combined with CT features for distinguishing pulmonary tuberculoma from malignancy among non-calcified solitary pulmonary solid nodules. J Thorac Dis 2023;15:386-98.

18. Ozlü T, Bülbül Y. Smoking and lung cancer. Tuberk Toraks 2005;53:200-9.

19. Xia C, Liu M, Li X, et al. Prediction Model for Lung Cancer in High-Risk Nodules Being Considered for Resection: Development and Validation in a Chinese Population. Front Oncol 2021;11:700179.