

Original Article

Development and validation of a CT-based nomogram for predicting invasiveness in solitary pulmonary ground-glass nodules

Yongming Wu^{1,2*}, Yunlu Zhang^{3*}, Jue Li⁴, Wenpeng Song^{1,2}, Guowei Che^{1,2}, Long Tian^{1,2}

¹Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ²Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ³Health Management Center, General Practice Medical Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ⁴Laboratory of Precision Therapeutics, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu, Sichuan, China. *Equal contributors.

Received March 9, 2026; Accepted August 7, 2026; Epub September 15, 2026; Published September 30, 2026

Abstract: This study aimed to develop and validate a nomogram based on clinical and computed tomography (CT) features for the preoperative prediction of invasiveness in patients with solitary pulmonary ground-glass nodules (GGNs). We retrospectively analyzed 846 patients with solitary GGNs who underwent surgical resection at the Lung Cancer Center of West China Hospital between November 2022 and January 2024. All nodules were pathologically confirmed as atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), or invasive adenocarcinoma (IAC). Patients were classified into a non-invasive group (including AAH, AIS, and MIA) and an IAC group based on postoperative pathological results. Clinical data and CT features before surgery were collected. Patients were randomly assigned to a training cohort (n = 592) and a validation cohort (n = 254) in a ratio of 7:3. Least absolute shrinkage and selection operator (LASSO) regression was used for feature selection, followed by multivariate logistic regression to identify independent predictors and construct a nomogram model. The variance inflation factor (VIF) was used to evaluate multicollinearity among the candidate predictors. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), calibration curves, and decision curve analysis (DCA). Multivariate logistic regression analysis revealed that the maximum tumor diameter (OR = 39.82, 95% CI: 16.29-97.32), consolidation-to-tumor ratio ($0.5 < \text{CTR} < 1$, OR = 2.51, 95% CI: 1.36-4.63), vacuole sign (OR = 2.29, 95% CI: 1.29-4.07), pleura traction sign (OR = 3.03, 95% CI: 1.66-5.53), and the presence of solid component on the mediastinal window (OR = 3.27, 95% CI: 1.46-7.28) were independent risk factors for IAC. The nomogram constructed based on these variables demonstrated good discriminative ability, with AUC values of 0.930 and 0.913 in the training and validation cohorts, respectively. Calibration curves demonstrated good agreement between the predicted and observed probabilities, and DCA indicated favorable clinical utility. Consequently, this nomogram based on the maximum tumor diameter, CTR, vacuole sign, pleura traction sign, and the presence of solid component on the mediastinal window, provides an effective and noninvasive tool for preoperative prediction of the invasiveness of solitary GGNs.

Keywords: Lung adenocarcinoma, ground-glass nodule, invasiveness, nomogram

Introduction

The widespread adoption of low-dose computed tomography (LDCT) has markedly increased the detection of pulmonary nodules [1]. Based on density, pulmonary nodules can be classified as solid nodules and ground-glass nodules (GGNs) [2]. Persistent GGNs are predominantly associated with lung adenocarcinoma, and typically follow a stepwise pathological progres-

sion from atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and minimally invasive adenocarcinoma (MIA) to invasive adenocarcinoma (IAC) [3, 4]. Previous studies have shown that for patients with AIS and MIA who undergo radical surgery, regardless of whether lobectomy or segmentectomy is performed, the 10-year recurrence-free survival (RFS) rate remains 100% [5, 6]. In contrast, patients with IAC remain at risk of postopera-

tive recurrence and metastasis even at stage IA1 [7]. Chen et al. [8, 9] defined AIS, MIA and pure GGN as the window of surgical cure for lung adenocarcinoma, and proposed that for lesions within this curative window, surgical intervention can achieve excellent outcomes, and that limited resection without mediastinal lymph node dissection may be sufficient. Therefore, precise preoperative differentiation between non-invasive lesions and IAC is critical for determining both the optimal timing of surgery and the most appropriate surgical strategy.

Currently, the assessment of tumor invasiveness primarily relies on postoperative histopathological examination. Although intraoperative frozen section analysis is frequently used to guide surgical decision-making, the degree of invasion may be underestimated because of sampling limitations [10]. Previous studies have reported that several CT imaging features, such as the spiculation sign, CTR, and vascular convergence sign, are correlated with the invasive nature of lung adenocarcinoma. These preoperative CT characteristics may help distinguish IAC from AIS or MIA, and assist clinicians in determining the optimal surgical timing and extent of resection [11-13].

Several studies have attempted to identify distinct CT imaging characteristics to develop predictive models for assessing the risk of IAC. However, limitations in sample size and the restricted range of included features limit the generalizability of the models [13, 14]. Moreover, many previous studies have focused primarily on patients with multiple GGNs [15, 16]. Predictive models specifically designed for solitary GGNs remain scarce, and existing studies often suffer from insufficient sample sizes and limited variable selection [17]. To address these limitations, we integrated CT imaging features, clinical characteristics, and preoperative tumor markers to identify risk factors for IAC and developed a predictive model for preoperative invasiveness assessment in patients with solitary GGNs.

Patients and methods

This study was reviewed and approved by the Biomedical Ethics Committee of West China Hospital (Approval No. 2024-186) and the requirement for informed consent was waived

because of the retrospective nature of the study. All patient data were anonymized to ensure privacy and confidentiality.

Study population

A retrospective analysis was conducted on patients with surgically treated solitary pulmonary GGNs at the Lung Cancer Center of West China Hospital, from November 2022 to January 2024. The inclusion criteria were: (1) Availability of preoperative thin-section chest CT images acquired within two weeks before surgery, showing a solitary pulmonary GGN; (2) Age ≥ 18 years; (3) Nodules ≤ 3 cm on CT; (4) Postoperative pathological diagnosis of IAC, MIA, AIS, or AAH; and (5) No history of any antitumor therapy (e.g., chemotherapy, radiotherapy, targeted therapy, or immunotherapy) prior to surgery. The exclusion criteria were: (1) Poor-quality CT images that compromised accurate assessment; (2) Incomplete clinical or pathological data; and (3) Solid nodules without any ground-glass component.

Based on postoperative pathological findings, patients were classified into an IAC group and a NIAC group, which included MIA, AAH, and AIS. Patients were then randomly assigned to a training cohort (70%) and an internal validation cohort (30%) using a computer-generated random number sequence.

Data collection

Demographic and clinical characteristics: The following data were retrospectively retrieved from electronic medical records: demographic and clinical characteristics (age, sex, smoking history, personal or family history of malignancy) and preoperative serum tumor markers, including carcinoembryonic antigen (CEA), neuron-specific enolase (NSE), and cytokeratin fragment 19 (CYFRA 21-1).

CT image evaluation: Preoperative CT images were independently evaluated by two experienced thoracic radiologists (with 8 and 25 years of experience in thoracic imaging, respectively), who were blinded to the pathological results. The following imaging features were assessed: maximum tumor diameter, tumor location, consolidation-to-tumor ratio (CTR), spiculation sign, vacuole sign, lobulation sign, pleura traction sign, vascular convergence sign,

Nomogram for predicting invasiveness of solitary GGNs

air bronchogram sign, and the presence or absence of solid component on mediastinal window. The CTR was defined as the ratio of the maximum diameter of the solid component to the maximum diameter of the tumor, both of which were measured in the lung window. Any discrepancies between the two readers were resolved by consensus discussion with a senior third radiologist.

Statistical analysis

All statistical analyses were performed using R software (version 4.3.3). Continuous variables were tested for normality using the Shapiro-Wilk test. Normally distributed continuous variables were presented as mean $\pm$ standard deviation (SD) and were compared using Student's t-test; otherwise, they were expressed as median with interquartile range (Q1, Q3) and compared using the Mann-Whitney U test. Categorical variables were reported as frequencies (percentages) and analyzed via the chi-square test or Fisher's exact test. A two-sided *P* value <0.05 was considered statistically significant.

Feature selection and model construction

The dataset was first randomly divided into training and validation cohorts at a ratio of 7:3. In the training cohort, the least absolute shrinkage and selection operator (LASSO) regression method was applied to identify the most relevant predictive variables. The optimal penalty parameter (λ) in the LASSO regression was selected via 10-fold cross-validation. Variables with non-zero coefficients from the LASSO regression were retained. Prior to constructing the multivariable logistic regression model, multicollinearity among the candidate predictors selected by LASSO regression was assessed using the variance inflation factor (VIF). A VIF value greater than 5 was considered indicative of severe multicollinearity. The results of the multivariable analysis were used to construct the final prediction nomogram.

Model performance and validation

The model performance was evaluated in both the training and validation cohorts. The discrimination ability of the model was evaluated using the area under the receiver operating characteristic curve (AUC), the calibration was

evaluated using calibration curves and the Hosmer-Lemeshow test, and clinical utility was estimated via decision curve analysis (DCA).

Results

Patient characteristics

A total of 846 patients with lung adenocarcinoma were included in this study, including 13 patients with AAH, 9 patients with AIS, 598 patients with MIA and 226 patients with IAC. The overall cohort had a median age of 52 years (IQR: 41-59). Among them, 629 patients (74.3%) were female and 217 (25.7%) were male. All patients were randomly split into a training cohort ($n = 592$, 70%) and a validation cohort ($n = 254$, 30%) using a 7:3 ratio. The training cohort included 431 NIAC cases and 161 IAC cases, while the validation cohort included 189 NIAC cases and 65 IAC cases. The baseline clinical and imaging characteristics of the training and validation cohorts are shown in **Table 1**. No significant differences were observed between the two cohorts in baseline characteristics (all $P > 0.05$) ([Table S1](#)), indicating good comparability.

Feature selection

In the training cohort, all candidate variables were included in the LASSO regression analysis to identify the most representative IAC predictors. The coefficient path plot shows that, as the regularization parameter λ varies, the coefficients of the independent variables are gradually compressed to zero (**Figure 1A**). The cross-validation curve illustrates the mean squared error (MSE) across a range of λ values (**Figure 1B**). The dashed line on the left indicates λ at the minimum MSE (0.00786), and the dashed line on the right represents the λ positioned one standard error above this minimum (0.02633, λ_{1se}). To prioritize a parsimonious and clinically interpretable model, predictors were selected using the λ_{1se} criterion. At this threshold ($\lambda = 0.02633$), nine variables with non-zero coefficients were identified as potential predictors: age, tumor location, maximum tumor diameter, CTR, lobulation sign, vacuole sign, pleura traction sign, air bronchogram sign, and solid component on the mediastinal window (**Figure 1C**). The complete list of all candidate variables along with their corresponding LASSO coefficients is provided in [Table S2](#).

Nomogram for predicting invasiveness of solitary GGNs

Table 1. Comparison of clinical-radiological features between NIAC and IAC groups

Features	Training Cohort (n = 592)			Validation Cohort (n = 254)		
	NIAC n = 431	IAC n = 161	P	NIAC n = 189	IAC n = 65	P
Age(years), Median (Q1, Q3)	48 (39, 56)	57 (52, 66)	<0.001	49 (39, 57)	59 (53, 67)	<0.001
Sex, n (%)			0.048			0.126
Female	329 (76.3%)	110 (68.3%)		146 (77.2%)	44 (67.7%)	
Male	102 (23.7%)	51 (31.7%)		43 (22.8%)	21 (32.3%)	
Smoking history, n (%)			0.220			0.025
No	378 (87.7%)	135 (83.9%)		169 (89.4%)	51 (78.5%)	
Yes	53 (12.3%)	26 (16.1%)		20 (10.6%)	14 (21.5%)	
Family history of malignancy, n (%)			0.179			0.556
No	346 (80.3%)	137 (85.1%)		160 (84.7%)	53 (81.5%)	
Yes	85 (19.7%)	24 (14.9%)		29 (15.3%)	12 (18.5%)	
History of malignancy, n (%)			0.680			0.757
No	403 (93.5%)	149 (92.5%)		169 (89.4%)	59 (90.8%)	
Yes	28 (6.5%)	12 (7.5%)		20 (10.6%)	6 (9.2%)	
Location, n (%)			0.234			0.545
RM	28 (6.5%)	11 (6.8%)		15 (7.9%)	4 (6.2%)	
LL	65 (15.1%)	23 (14.3%)		26 (13.8%)	9 (13.8%)	
LU	124 (28.8%)	33 (20.5%)		56 (29.6%)	13 (20.0%)	
RL	78 (18.1%)	39 (24.2%)		37 (19.6%)	15 (23.1%)	
RU	136 (31.6%)	55 (34.2%)		55 (29.1%)	24 (36.9%)	
Maximum tumor diameter(cm), Median (Q1, Q3)	0.80 (0.70, 1.00)	1.50 (1.20, 2.10)	<0.001	0.90 (0.70, 1.00)	1.50 (1.10, 2.00)	<0.001
CYFRA211, n (%)			0.021			0.700
Abnormal	60 (13.9%)	35 (21.7%)		33 (17.5%)	10 (15.4%)	
Normal	371 (86.1%)	126 (78.3%)		156 (82.5%)	55 (84.6%)	
CEA, n (%)			0.454			0.271
Abnormal	5 (1.2%)	3 (1.9%)		2 (1.1%)	2 (3.1%)	
Normal	426 (98.8%)	158 (98.1%)		187 (98.9%)	63 (96.9%)	
NSE, n (%)			0.903			0.203
Abnormal	23 (5.3%)	9 (5.6%)		8 (4.2%)	6 (9.2%)	
Normal	408 (94.7%)	152 (94.4%)		181 (95.8%)	59 (90.8%)	
CTR, n (%)			<0.001			<0.001
≤0.5	343 (79.6%)	71 (44.1%)		156 (82.5%)	22 (33.8%)	
0.5<CTR<1	88 (20.4%)	90 (55.9%)		33 (17.5%)	43 (66.2%)	

Nomogram for predicting invasiveness of solitary GGNs

Lobulation sign, n (%)			<0.001		<0.001
Absent	402 (93.3%)	106 (65.8%)		173 (91.5%)	45 (69.2%)
Present	29 (6.7%)	55 (34.2%)		16 (8.5%)	20 (30.8%)
Spiculation sign, n (%)			<0.001		<0.001
Absent	328 (76.1%)	88 (54.7%)		156 (82.5%)	33 (50.8%)
Present	103 (23.9%)	73 (45.3%)		33 (17.5%)	32 (49.2%)
Vacuole Sign, n (%)			<0.001		<0.001
Absent	248 (57.5%)	53 (32.9%)		106 (56.1%)	21 (32.3%)
Present	183 (42.5%)	108 (67.1%)		83 (43.9%)	44 (67.7%)
Pleura traction sign, n (%)			<0.001		<0.001
Absent	334 (77.5%)	68 (42.2%)		141 (74.6%)	33 (50.8%)
Present	97 (22.5%)	93 (57.8%)		48 (25.4%)	32 (49.2%)
Vascular convergence sign, n (%)			0.177		>0.999
Absent	16 (3.7%)	2 (1.2%)		5 (2.6%)	1 (1.5%)
Present	415 (96.3%)	159 (98.8%)		184 (97.4%)	64 (98.5%)
Air Bronchogram Sign, n (%)			<0.001		<0.001
Absent	408 (94.7%)	105 (65.2%)		179 (94.7%)	42 (64.6%)
Present	23 (5.3%)	56 (34.8%)		10 (5.3%)	23 (35.4%)
Solid component on mediastinal window, n (%)			<0.001		<0.001
Absent	410 (95.1%)	89 (55.3%)		176 (93.1%)	29 (44.6%)
Present	21 (4.9%)	72 (44.7%)		13 (6.9%)	36 (55.4%)

NIAC, non-invasive adenocarcinoma; IAC, invasive adenocarcinoma; CTR, consolidation-to-tumor ratio; RM, right middle; RU, right upper; RL, right lower; LL, left lower; LU, left upper; CYFRA211, cytokeratin fragment 19; NSE, neuron-specific enolase; CEA, carcinoembryonic antigen.

Nomogram for predicting invasiveness of solitary GGNs

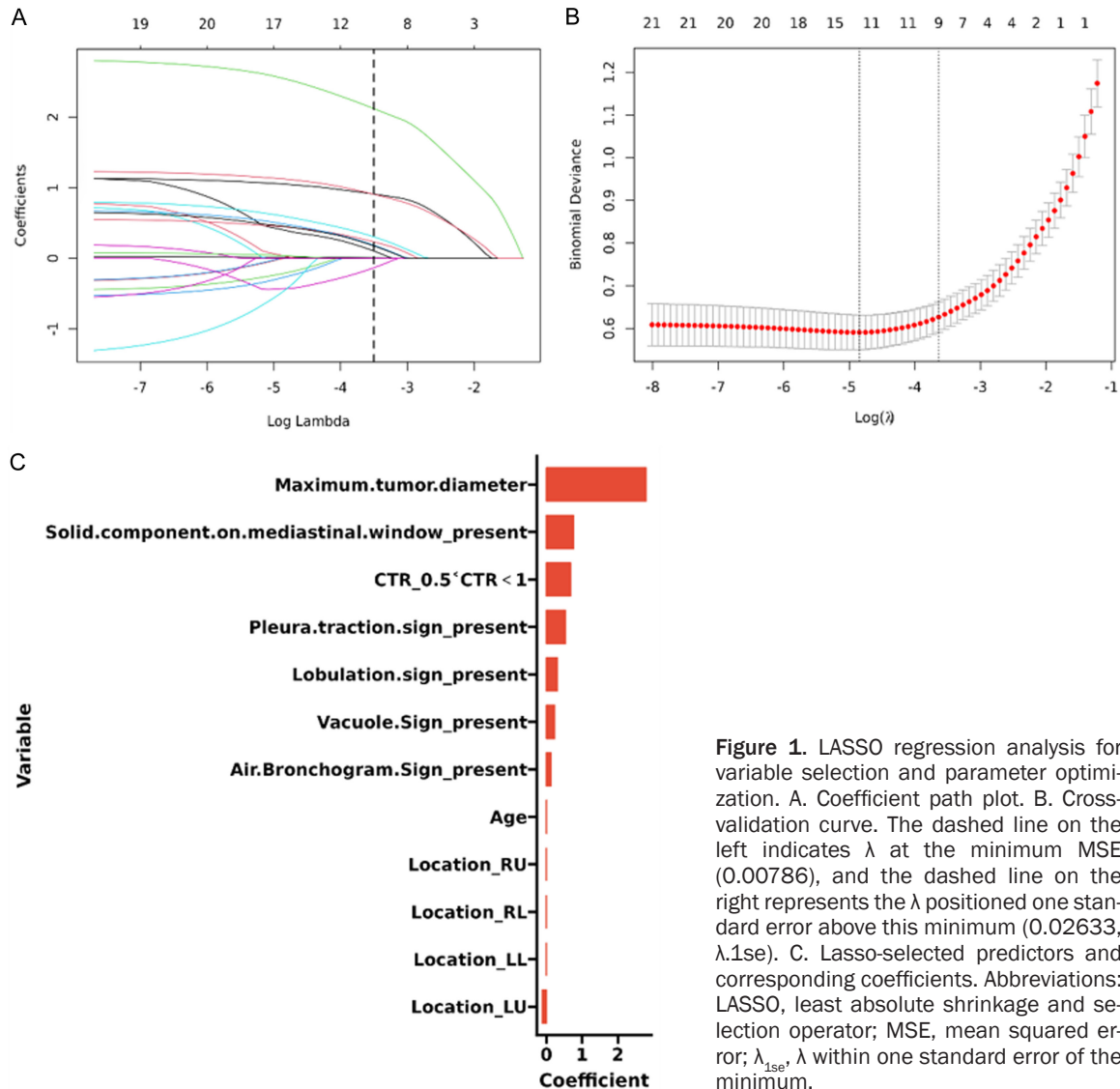


Figure 1. LASSO regression analysis for variable selection and parameter optimization. A. Coefficient path plot. B. Cross-validation curve. The dashed line on the left indicates λ at the minimum MSE (0.00786), and the dashed line on the right represents the λ positioned one standard error above this minimum (0.02633, λ_{1se}). C. Lasso-selected predictors and corresponding coefficients. Abbreviations: LASSO, least absolute shrinkage and selection operator; MSE, mean squared error; λ_{1se} , λ within one standard error of the minimum.

Prior to constructing the multivariable logistic regression model, multicollinearity was assessed for the candidate predictors selected by LASSO regression, using the VIF. All VIF values were below 2 (range: 1.016-1.872), indicating no severe multicollinearity (Table S3).

These nine variables screened by LASSO regression were subsequently entered into multivariate logistic regression analysis. The multivariate logistic regression analysis identified that the maximum tumor diameter (OR = 39.82, 95% CI: 16.29-97.32, $P < 0.001$), the presence of solid component on the mediastinal window (OR = 3.27, 95% CI: 1.46-7.28, $P = 0.004$), pleura traction sign (OR = 3.03, 95% CI: 1.66-5.53, $P < 0.001$), $0.5 < \text{CTR} < 1$ (OR = 2.51, 95% CI:

1.36-4.63, $P = 0.003$), and the vacuole sign (OR = 2.29, 95% CI: 1.29-4.07, $P = 0.005$) were independent predictors for IAC in solitary GGNs (Table 2).

Construction of the IAC risk prediction model

Based on the five independent predictors: maximum tumor diameter, CTR, vacuole sign, pleura traction sign, and solid component on the mediastinal window, a risk prediction model for IAC was developed, and a nomogram was constructed (Figure 2). The nomogram indicated that maximum tumor diameter contributed the most to the prediction of IAC, followed by CTR and the presence of solid component on the mediastinal window. Each variable was as-

Nomogram for predicting invasiveness of solitary GGNs

Table 2. Multivariable logistic regression of clinical-radiological features and invasiveness of solitary ground glass nodules

Features	OR	95% CI	P
Age	1.02	0.99, 1.04	0.131
Location			
RM	-	-	-
LL	0.91	0.25, 3.29	0.888
LU	0.55	0.16, 1.88	0.340
RL	1.83	0.53, 6.29	0.339
RU	1.29	0.39, 4.29	0.674
Maximum tumor diameter	39.82	16.29, 97.32	<0.001
0.5<CTR<1	2.51	1.36, 4.63	0.003
Lobulation sign	2.03	0.91, 4.53	0.083
Vacuole Sign	2.29	1.29, 4.07	0.005
Pleura traction sign	3.03	1.66, 5.53	<0.001
Air Bronchogram Sign	1.86	0.75, 4.60	0.182
Solid component on mediastinal window	3.27	1.46, 7.28	0.004

CTR, consolidation-to-tumor ratio; RM, right middle; RU, right upper; RL, right lower; LL, left lower; LU, left upper; OR, odds ratio; CI, confidence interval.

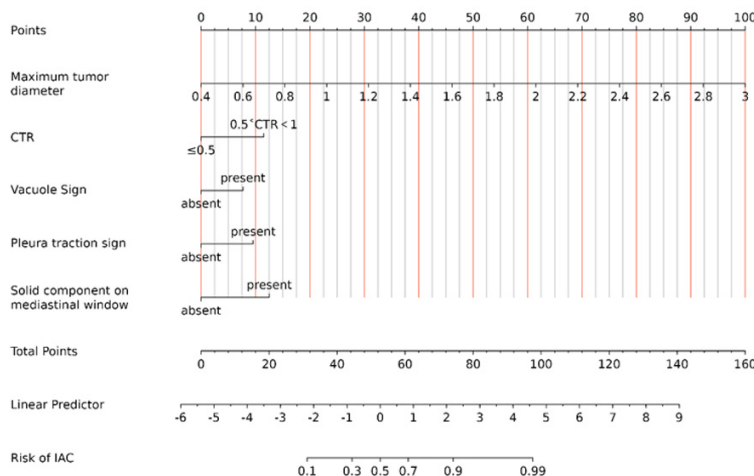


Figure 2. A nomogram predicting the risk of invasive adenocarcinoma in patients with solitary pulmonary ground-glass nodules.

signed a point score on the nomogram scale. By summing the points corresponding to all the variables, a total score is obtained. A vertical line drawn from the total points axis to the risk probability axis indicates the individual's predicted probability of IAC.

Validation of the IAC risk prediction model

The nomogram demonstrated excellent discriminative ability, with an AUC of 0.930 (95% CI: 0.909-0.952) in the training cohort and 0.913 (95% CI: 0.873-0.954) in the validation

cohort (**Figure 3**). To classify individuals into high- or low-risk groups, the optimal cutoff probability was determined by maximizing the Youden index (sensitivity + specificity - 1). The optimal cutoff was identified as 0.339. At this threshold, the nomogram achieved a sensitivity of 80.0% and a specificity of 88.4% in the validation cohort. Calibration curves demonstrated that the predicted probabilities of the nomogram closely aligned with the observed outcomes, as the 'logistic calibration' line approximated the ideal reference line in the training cohort (**Figure 4A**) and the validation cohort (**Figure 4B**). The Hosmer-Lemeshow test indicated good calibration in the training cohort ($\chi^2 = 10.207$, $P = 0.251$) and the validation cohort ($\chi^2 = 13.111$, $P = 0.108$), confirming no significant deviation between the predicted and observed probabilities.

The clinical utility of the model was further quantified using decision curve analysis (DCA), as shown in **Figure 5**. The net benefit of the model strategy (red curve) was compared against two default strategies: intervening on all patients (Treat All, grey curve) and intervening on no patients (Treat None, black line). In both the training (**Figure 5A**) and validation (**Figure 5B**) cohorts, the decision curve for our nomogram remained above the Treat All and Treat None reference lines at threshold probabilities between 0.1 and 0.8, demonstrating that applying the model yields a positive net clinical benefit.

Discussion

For patients with lung adenocarcinoma manifesting as GGNs, accurate preoperative distinguishing between IAC and non-invasive lesions

Nomogram for predicting invasiveness of solitary GGNs

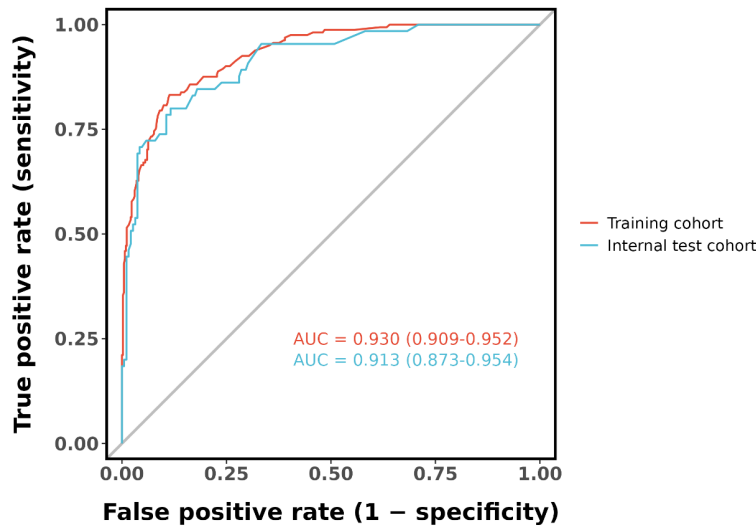


Figure 3. Receiver operating characteristic (ROC) curves of the nomogram in training cohort and validation cohort.

(AIS, MIA, and AAH) is critical for determining optimal surgical timing and strategy. We developed and validated a clinical-radiographic predictive model based on five independent predictors that included CTR, maximum tumor diameter, vacuole sign, pleura traction sign, and the presence of solid component on the mediastinal window to predict invasiveness in lung adenocarcinoma presenting as solitary GGNs. The model demonstrated excellent diagnostic performance, with AUCs of 0.930 (95% CI: 0.909-0.952) in the training cohort and 0.913 (95% CI: 0.873-0.954) in the validation cohort. Furthermore, the model exhibited good calibration and provided a net clinical benefit across a wide range of threshold probabilities. This model offers a practical, individualized risk assessment to aid clinicians in optimizing surgical strategies for patients with lung adenocarcinoma presenting as solitary GGNs.

Predictive models provide clinicians with an objective and quantifiable reference standard, thereby facilitating more rational and consistent clinical decisions. Models for predicting malignancy risk of pulmonary nodules, such as the Brock, Mayo and VA models [18-20], are already widely used in clinical practice. Nevertheless, there is currently no widely accepted model for predicting the invasiveness of lung adenocarcinoma. Several studies have previously developed predictive models for assessing the invasiveness of lung adenocarcinoma

[15-17, 21, 22]. Liu et al. [15] reported a nomogram incorporating CYFRA 21-1, density type, maximum diameter, and air bronchogram to non-invasively differentiate IAC from pre-invasive/microinvasive solitary lung adenocarcinoma lesions, achieving AUCs of 0.925 (95% CI: 0.897-0.953) in the training cohort and 0.895 (95% CI: 0.829-0.961) in the testing cohort. However, their study did not exclude solid nodules, which are known to have a higher likelihood of invasiveness compared with GGNs. Yang et al. [22] constructed a nomogram based on maximum diameter, CTR, maximum CT value, mean CT

value, spiculation and vascular convergence sign for predicting invasiveness of GGNs, which achieved AUCs of 0.910 (95% CI: 0.885-0.934) in the training group and 0.902 (95% CI: 0.859-0.944) in the validation group. Dong et al. [17] constructed a clinical-radiomics nomogram based on 473 GGNs for the preoperative differentiation between AIS/MIA and IAC. The model yielded AUCs of 0.845 (95% CI: 0.817-0.873) in the training set and 0.823 (95% CI: 0.795-0.851) in the test set. Compared to these models, our model, focused exclusively on patients with solitary GGNs, demonstrated a higher AUC, and was developed from a larger cohort of 846 patients. However, owing to the retrospective design of this study and differences in the predictive factors included in various models, a fair head-to-head comparison with previously established models could not be conducted. Moreover, most of these models have not undergone external validation. Future studies should further validate different models using external data and explore predictors that may better predict IAC.

Deep learning-based artificial intelligence (AI) demonstrated distinct advantages in extracting radiomic features for the assessment of pulmonary nodules [23, 24]. Predictive models developed using AI tools have demonstrated robust performance in assessing the invasiveness of lung adenocarcinoma [16, 25, 26]. Shi

Nomogram for predicting invasiveness of solitary GGNs

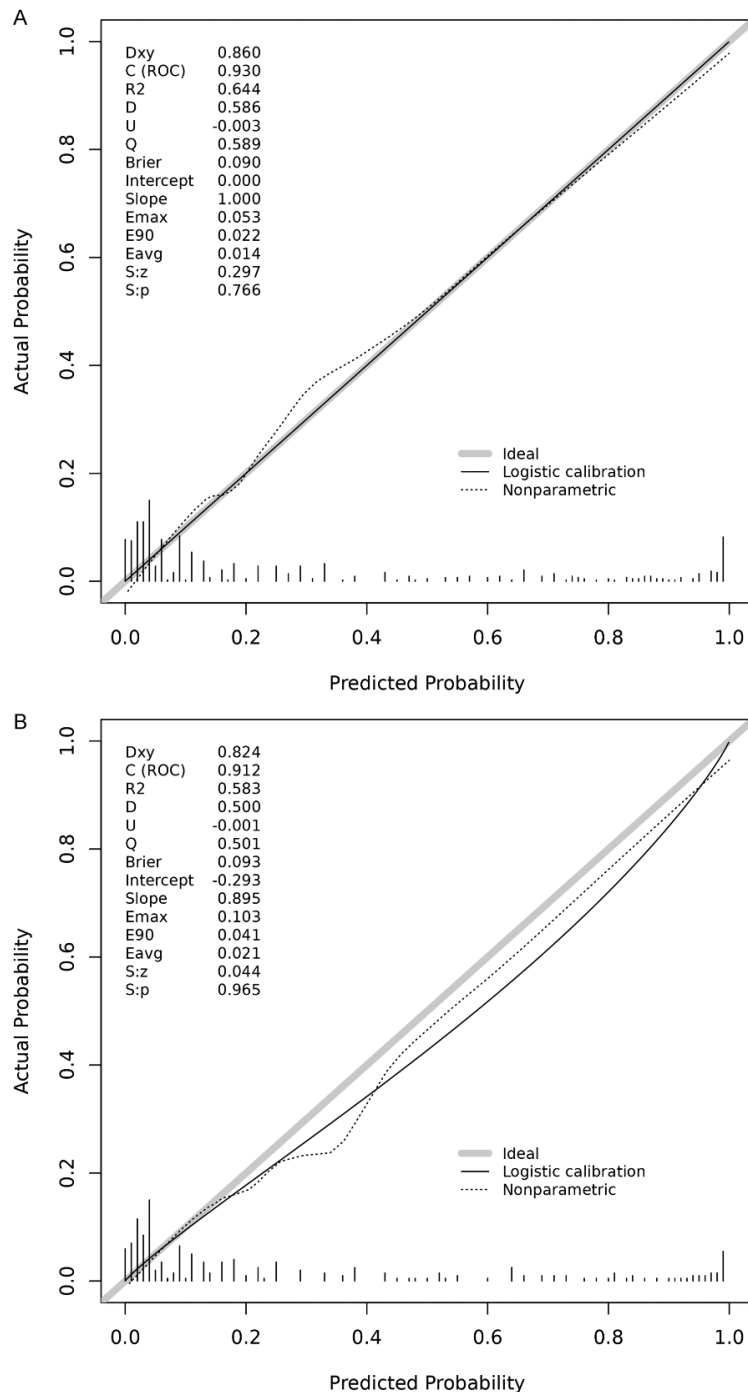


Figure 4. Calibration curves of the nomogram. Calibration curve of the nomogram in training cohort (A) and validation cohort (B).

et al. [16] developed an AI-driven clinical-radiomics model based on age, median tumor size, sphericity, gray level co-occurrence matrix features *lmc1*, gray level size zone matrix features *small area emphasis*, and gray level dependence matrix features *dependence en-*

tropy, which demonstrated excellent and generalizable performance for predicting invasiveness of stage T1 lung adenocarcinoma, achieving AUCs of 0.941 (95% CI: 0.927-0.955) in the training cohort, 0.934 (95% CI: 0.911-0.958) in the internal validation cohort, and 0.905 (95% CI: 0.878-0.931) in the external validation cohort. Yao et al. [25] developed a surgical resection image-based model (SuRImage), a deep learning system using smartphone photographs of surgical resection specimens for intraoperative diagnosis of stage IA lung adenocarcinoma, which achieved AUCs of 0.84 for IAC identification. Despite these encouraging results, these AI-based approaches often require specialized software, computational resources, or additional imaging techniques, which may limit their widespread adoption in routine clinical practice. Moreover, these studies did not incorporate morphological features such as lobulation and spiculation into their analyses. Our model incorporated only routinely available clinical and CT imaging features, maximum tumor diameter, CTR, vacuole sign, pleura traction sign, and solid component on the mediastinal window, that can be easily obtained in daily practice without additional imaging techniques or specialized software. Nevertheless, future studies should consider integrating AI tools to further improve predictive performance. For example, AI-based automated segmentation and feature extraction could be combined with our simplified clinical-radiographic model to achieve a more comprehensive and accurate prediction of GGN invasiveness, while maintaining clinical usability.

Nomogram for predicting invasiveness of solitary GGNs

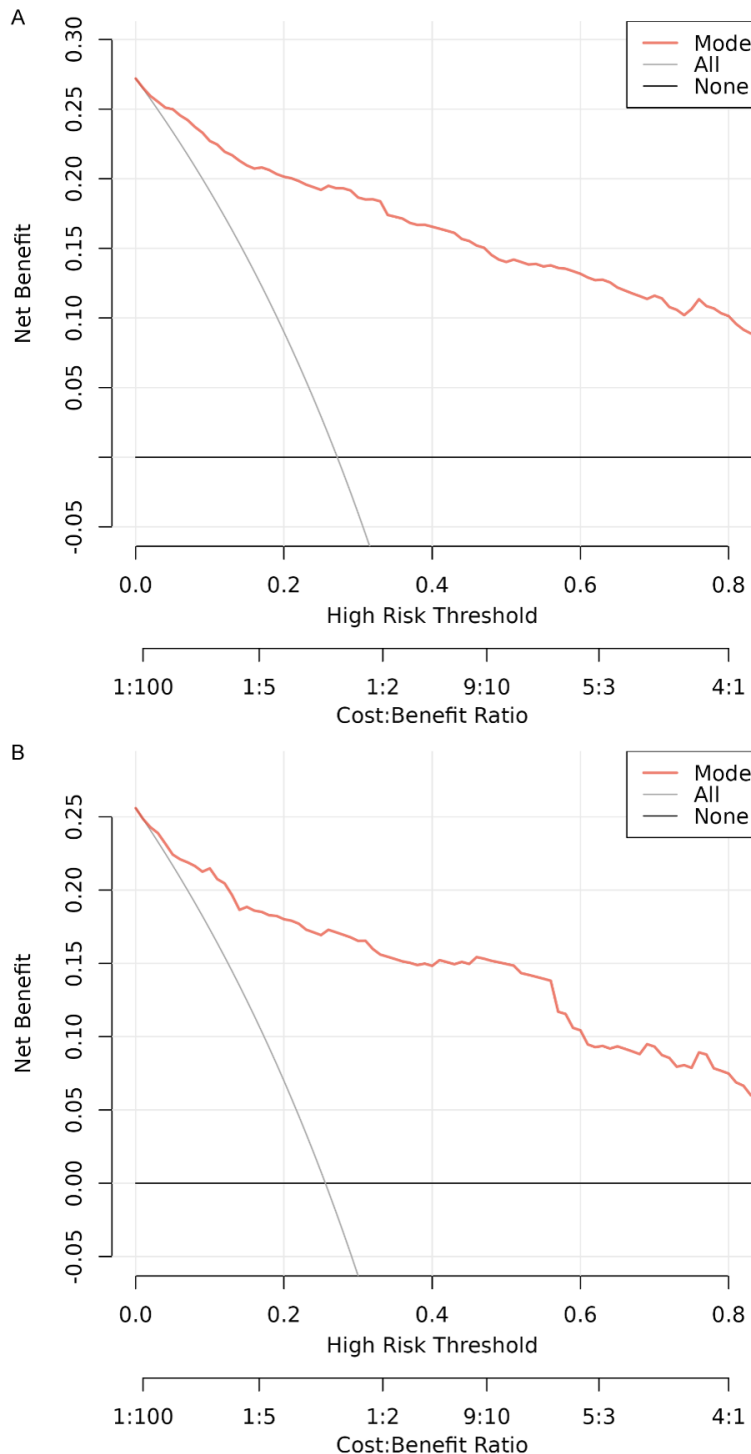


Figure 5. Decision curve analysis (DCA) of the nomogram. DCA of the nomogram in training cohort (A) and validation cohort (B).

We found that the maximum tumor diameter was the strongest predictor of IAC, which is consistent with previous studies [22, 27]. Nevertheless, the optimal cutoff value of maxi-

mum tumor diameter for distinguishing IAC remains controversial among different studies [28, 29]. Zhang et al. [29] reported that the optimal threshold for identifying IAC based on the maximum diameter of GGNs was 10 mm. Liu et al. [30] conducted a retrospective analysis of 190 cases of GGNs and reported that the median diameter of the IAC was 13 mm, whereas the median diameter of the NIAC group was 8 mm. In our study, the median diameter of the IAC was 15 mm, whereas the median diameter of the NIAC group remained at 8 mm. Previous studies have identified age as a risk factor for the invasiveness of lung adenocarcinoma [17]. In the present study, univariate analysis demonstrated that patients with IAC were significantly older than those with NIAC in both the training and validation cohorts. Nevertheless, age did not remain an independent predictor in multivariate regression analysis. To exclude the possibility of multicollinearity between age and maximum tumor diameter, we performed VIF analysis for all candidate variables prior to constructing the multivariable logistic regression model. All VIF values were below 2.0, indicating no significant collinearity between age and maximum tumor diameter. It is plausible that older patients may have undergone longer follow-up prior to surgery, leading to delayed intervention and thus larger tumor diameters at the time of resection. However, since our study did not collect

data on the interval between initial detection and resection, this hypothesis could not be tested directly. Therefore, future studies should incorporate follow-up data, such as the

interval between initial detection and resection and tumor volume doubling time, to further elucidate the temporal relationship among age, tumor growth, and the development of invasive features.

Consolidation-to-tumor ratio, defined as the ratio of the solid component diameter on the lung window to the total tumor diameter, is widely used in clinical practice to estimate tumor aggressiveness [31]. In lung adenocarcinoma manifesting as GGNs, the solid component typically corresponds to the invasive growth pattern [2]. Therefore, a higher CTR implies a larger invasive component and thus a higher risk of IAC. Our study further confirmed that GGNs with a $CTR > 0.5$ were associated with a higher risk of IAC, which is consistent with previous findings [32]. The Japan Clinical Oncology Group (JCOG 0201) study [11] revealed that patients with $CTR \leq 0.5$ have a favorable prognosis, with a 5-year survival rate of approximately 97%, and this criterion can be used to identify NIAC. Notably, Lai et al. [33] found that GGNs with solid component on both the lung window and the mediastinal window exhibited a significantly higher proportion of invasive pathological subtypes and a markedly worse prognosis compared to those with solid component visible only on the lung window. The presence of solid component on the mediastinal window was identified as an independent risk factor for IAC in our study. This association can be attributed to the underlying histopathological changes. In the pre-invasive stage of lung adenocarcinoma, tumor cells grow in a lepidic pattern along the alveolar walls, with most of the lung tissue structure remaining intact, which appears as a ground-glass opacity on CT. In the IAC group, as tumor cells accumulate extensively and invade deeper layers, alveolar walls collapse and fibrotic scar tissue proliferates. These proliferating tumor cells fill and replace the alveolar air spaces, leading to a significant increase in tissue density [30]. Concurrently, as the disease progresses, fibroblast activation, collagen deposition, and extracellular matrix remodeling further increase lesion density. With disease progression, GGNs may initially exhibit solid components on the lung window and subsequently become visible on the mediastinal window as well. Therefore, the presence of a solid component on the mediastinal window indicates a higher risk of IAC.

However, the specific mechanisms underlying these changes remain to be further investigated.

In our study, the incidence of the vacuole sign was notably high in the IAC group, reaching 67.3%, which is consistent with the findings reported by Liu et al. [30]. The vacuole sign is manifested as round or oval intranodular hypodense shadows on CT, measuring ≤ 5 mm in diameter, which may be solitary or multiple [34]. The pathological mechanisms underlying its formation may involve unclosed or dilated small bronchi, alveolar spaces that remain unsealed by cancer cells growing along the alveolar walls, or tumor cell invasion into the alveolar septa and interstitium, which leads to architectural destruction and coalescence of adjacent air spaces [35]. In addition, tumor necrosis or mucin secretion may also contribute to the formation of vacuole-like structures. Therefore, the presence of a vacuole sign is strongly associated with high tumor invasiveness. Zhang et al. [29] reported that the pleura traction sign was an independent risk factor for IAC, which is consistent with our results. This sign is generally considered to reflect fibrotic traction between the tumor and the pleura, which often occurs during tumor progression and stromal remodeling. Previous studies have reported that CT morphological features such as spiculation and lobulation are associated with IAC [17, 22]. However, these features were not found to be risk factors in our cohort. Further studies are warranted to elucidate the relationship between CT imaging characteristics and the invasiveness of lung adenocarcinoma.

This study has several limitations. First, the retrospective, single-center design may introduce selection bias. Future multicenter prospective studies are warranted to further validate and refine the model. Second, owing to the retrospective design of this study and differences in the predictive factors included across various models, a direct head-to-head performance comparison with previously established models could not be conducted. Future studies should consider prospective comparisons using shared validation datasets to better establish the relative performance and practical value of our model. Third, this study primarily evaluated the overall discriminative

ability of the nomogram. Although the optimal cutoff probability was determined by maximizing the Youden index, clinically applicable low-, intermediate-, and high-risk stratification criteria for GGN invasiveness were not formally established. This limits the clinical translational value of the model, and further work is required to define clinically meaningful risk thresholds. Fourth, the pathological subtype distribution of our cohort was imbalanced, with a limited number of pre-invasive lesions (13 AAH cases and 9 AIS cases) compared to MIA (n = 598) and IAC (n = 226). This disproportionate representation may introduce statistical bias and potentially reduce the robustness and generalizability of the predictive model. Future studies with a more balanced distribution of pathological subtypes are needed to validate our findings. Fifth, although we provided a pathophysiological explanation for the vacuole sign and mediastinal window solid component as independent risk factors, the mechanistic understanding remains limited. Future research is needed to explore the underlying mechanisms. Sixth, although common serum tumor markers were included, their predictive capability for invasiveness in early-stage lung adenocarcinoma remains limited. Emerging biomarkers or liquid biopsy techniques may offer additional predictive value in the future.

Conclusion

In summary, we developed and validated an efficient nomogram integrating five readily obtainable CT features including maximum tumor diameter, CTR, vacuole sign, pleura traction sign, and the presence of a solid component on the mediastinal window for the preoperative prediction of invasiveness in patients with solitary GGNs. As a simple and noninvasive tool, this nomogram may help clinicians better assess the invasive risk of GGNs and support individualized surgical decision-making.

Acknowledgements

This paper was supported by the Postdoctoral Fellowship Program of CPSF (GZC20251396), the Postdoctor Research Fund of West China Hospital, Sichuan University (2025HXBH092), and the Natural Science Foundation of Sichuan Province (2024NSFSC1928).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Long Tian, Department of Thoracic Surgery/Lung Cancer Center, West China Hospital, Sichuan University, No. 37 Guoxue Alley, Chengdu 610041, Sichuan, China. Tel: +86-28-85422494; Fax: +86-28-85422493; E-mail: tianlong@wchscu.edu.cn

References

- [1] National Lung Screening Trial Research Team; Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM and Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395-409.
- [2] Bankier AA, MacMahon H, Colby T, Gevenois PA, Goo JM, Leung ANC, Lynch DA, Schaefer-Prokop CM, Tomiyama N, Travis WD, Verschakelen JA, White CS and Naidich DP. Fleischner society: glossary of terms for thoracic imaging. *Radiology* 2024; 310: e232558.
- [3] Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, Dacic S, Jain D, Kerr KM, Lantuejoul S, Noguchi M, Papotti M, Rekhtman N, Scagliotti G, van Schil P, Sholl L, Yatabe Y, Yoshida A and Travis WD. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol* 2022; 17: 362-387.
- [4] Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, Garg K, Austin JH, Asamura H, Rusch VW, Hirsch FR, Scagliotti G, Mitsudomi T, Huber RM, Ishikawa Y, Jett J, Sanchez-Cespedes M, Sculier JP, Takahashi T, Tsuboi M, Vansteenkiste J, Wistuba I, Yang PC, Aberle D, Brambilla C, Flieder D, Franklin W, Gazdar A, Gould M, Hasleton P, Henderson D, Johnson B, Johnson D, Kerr K, Kuriyama K, Lee JS, Miller VA, Petersen I, Roggli V, Rosell R, Saijo N, Thunnissen E, Tsao M and Yankelewitz D. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 2011; 6: 244-285.
- [5] Li D, Deng C, Wang S, Li Y, Zhang Y and Chen H. Ten-year follow-up of lung cancer patients with resected adenocarcinoma in situ or minimally invasive adenocarcinoma: wedge resection is curative. *J Thorac Cardiovasc Surg* 2022; 164: 1614-1622, e1611.

Nomogram for predicting invasiveness of solitary GGNs

- [6] Yotsukura M, Asamura H, Motoi N, Kashima J, Yoshida Y, Nakagawa K, Shiraishi K, Kohno T, Yatabe Y and Watanabe SI. Long-term prognosis of patients with resected adenocarcinoma in situ and minimally invasive adenocarcinoma of the lung. *J Thorac Oncol* 2021; 16: 1312-1320.
- [7] Moon Y, Park JK, Lee KY and Kim ES. Prognosis after wedge resection in patients with 8(th) edition TNM stage IA1 and IA2 non-small cell lung cancer. *J Thorac Dis* 2019; 11: 2361-2372.
- [8] Fu F, Chen Z and Chen H. Treating lung cancer: defining surgical curative time window. *Cell Res* 2023; 33: 649-650.
- [9] Fu F, Zhang Y and Chen H. The surgical curative time window for lung adenocarcinoma: a comprehensive understanding. *Chin J Clin Thorac Cardiovasc Surg* 2023; 30: 1674-1677.
- [10] Wang H, Chen A, Wang K, Yang H, Wen W, Ren Q, Chen L, Xu X and Zhu Q. CT imaging features of lung ground-glass nodule patients with upgraded intraoperative frozen pathology. *Discov Oncol* 2024; 15: 29.
- [11] Asamura H, Hishida T, Suzuki K, Koike T, Nakamura K, Kusumoto M, Nagai K, Tada H, Mitsudomi T, Tsuboi M, Shibata T and Fukuda H; Japan Clinical Oncology Group Lung Cancer Surgical Study Group. Radiographically determined noninvasive adenocarcinoma of the lung: survival outcomes of Japan Clinical Oncology Group 0201. *J Thorac Cardiovasc Surg* 2013; 146: 24-30.
- [12] Lee SM, Park CM, Goo JM, Lee HJ, Wi JY and Kang CH. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology* 2013; 268: 265-273.
- [13] Yang Y, Wang W, Yang Y, Wang B, Hu H, Jiang Z, Cai D, Duan Y, Jiang J, Luo J, Zhao G, Huang Y and Ye L. Invasiveness assessment by CT quantitative and qualitative features of lung cancers manifesting ground-glass nodules in 555 patients: a retrospective cohort study. *Chin J Clin Thorac Cardiovasc Surg* 2024; 31: 51-58.
- [14] Zheng H, Li J, Wang S, Zhong F, Wang B, An W and Liao M. Evaluation on the invasion of lung cancer by diagnostic model nomogram based on the CT characteristics of pulmonary ground glass nodules. *Radiologic Practice* 2021; 36: 470-474.
- [15] Liu X, Wang D, Yang X, Yue X, Zheng C, Xia L and Kan X. A nomogram for preoperative prediction of invasiveness in solitary pulmonary adenocarcinoma: a multicenter study. *J Thorac Dis* 2025; 17: 8210-8223.
- [16] Shi W, Hu Y, Chang G, Yang Y, Qian H, Song Y, Wei Z, Gao L, Yi H, Wu S, Wang K, Huo H, Mao Y, Sun Y, Li M, Ai S, Zhao L, Li X and Zheng H. Development and validation of a nomogram for preoperative prediction of the invasiveness of stage T1 lung adenocarcinoma utilizing AI-driven radiomics. *BMC Cancer* 2026; 26: 199.
- [17] Dong N, Wei S, Zheng L, Huang D, Zhang G, Li Y, Zhang H, Wang A, Huang R, Zhao X and Liang P. Nomogram integrating clinical-radiological and radiomics features for differentiating invasive from non-invasive pulmonary adenocarcinomas presenting as ground-glass nodules. *Am J Cancer Res* 2025; 15: 797-810.
- [18] McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, Gingras M, Atkar-Khattra S, Berg CD, Evans K, Finley R, Yee J, English J, Nasute P, Goffin J, Puksa S, Stewart L, Tsai S, Johnston MR, Manos D, Nicholas G, Goss GD, Seely JM, Amjadi K, Tremblay A, Burrowes P, MacEachern P, Bhatia R, Tsao MS and Lam S. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013; 369: 910-919.
- [19] Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD and Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997; 157: 849-855.
- [20] Gould MK, Ananth L and Barnett PG; Veterans Affairs SNAP Cooperative Study Group. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest* 2007; 131: 383-388.
- [21] Zhang H, Miao L, Ma L, Sun X, Ouyang LN, Jing Y, Wang Y, Wang X, Wang P and Zhu L. Explainable machine learning for predicting invasiveness of pulmonary adenocarcinoma presenting as ground-glass nodules using CT images. *Acad Radiol* 2026; 33: 266-280.
- [22] Yang Y, Xu J, Wang W, Ma M, Huang Q, Zhou C, Zhao J, Duan Y, Luo J, Jiang J and Ye L. A nomogram based on the quantitative and qualitative features of CT imaging for the prediction of the invasiveness of ground glass nodules in lung adenocarcinoma. *BMC Cancer* 2024; 24: 438.
- [23] Long Y, Li Y, Zheng Y, Lin W, Qing H, Zhou P and Liu J. Quantitative measures of pure ground-glass nodules from an artificial intelligence software for predicting invasiveness of pulmonary adenocarcinoma on low-dose CT: a multicenter study. *Biomedicines* 2025; 13: 1600.
- [24] Ahn Y, Lee SM, Noh HN, Kim W, Choe J, Do KH and Seo JB. Use of a commercially available deep learning algorithm to measure the solid portions of lung cancer manifesting as subsolid lesions at CT: comparisons with radiologists and invasive component size at pathologic examination. *Radiology* 2021; 299: 202-210.

- [25] Yao L, Cai L, Weng M, Li Q, Liu F, Lu Y, Cui J, Lin H, Yao H, Xie D, Wu S, Huang L, Cai C, Lei Y, Xie R, Zhang Q, Li M, Zhan W, Li F, Zeng W, Zeng F, Zhong H, Liang Z, Dai J, Lin B, Zhang D, Zeng B, Wang G, Wing-Chi Chan L, Lanuti M, Qiao G, Lu C, Liu Z, Zhang Q, Zhang Y and Zhou H. Deep learning model for pathological invasiveness prediction using smartphone-based surgical resection images in clinical stage IA lung adenocarcinoma (SuRIImage): a prospective, multicentric, diagnostic study. *Lancet Digit Health* 2026; 8: 100965.
- [26] Nakamura T, Kudo Y, Matsubayashi J, Ichinose A, Park J, Shimada Y, Hagiwara M, Kakihana M, Ohira T, Nagao T, Masumoto J and Ikeda N. Artificial intelligence for accurate tumor size assessment and non-invasive adenocarcinoma prediction in small-sized lung cancer. *Eur J Surg Oncol* 2026; 52: 111734.
- [27] Qin L, Zhao L, Li XM, Jia QC, Liu ZY, Miao SD, Zhang X, Li YZ, Wang RT and Fu S. A CT-based radiomics nomogram incorporating adipose tissue to differentiate invasive adenocarcinomas among part-solid pulmonary nodules. *BMC Cancer* 2025; 25: 1471.
- [28] Moon Y, Sung SW, Lee KY, Sim SB and Park JK. Pure ground-glass opacity on chest computed tomography: predictive factors for invasive adenocarcinoma. *J Thorac Dis* 2016; 8: 1561-1570.
- [29] Zhang N, Liu JF, Wang YN and Yang L. A nomogram to predict invasiveness in lung adenocarcinoma presenting as ground glass nodule. *Transl Cancer Res* 2020; 9: 1660-1669.
- [30] Liu J, Yu X, Huang H, Chen S, Pan L, Feng Y, Chen K, Mao G and Ma H. High risk factors in images for infiltrating lung adenocarcinoma manifesting as peripheral ground-glass nodules. *Chin J Clin Thorac Cardiovasc Surg* 2024; 31: 85-91.
- [31] Suzuki K, Koike T, Asakawa T, Kusumoto M, Asamura H, Nagai K, Tada H, Mitsudomi T, Tsuboi M, Shibata T, Fukuda H and Kato H; Japan Lung Cancer Surgical Study Group (JCOG LCSSG). A prospective radiological study of thin-section computed tomography to predict pathological noninvasiveness in peripheral clinical IA lung cancer (Japan Clinical Oncology Group 0201). *J Thorac Oncol* 2011; 6: 751-756.
- [32] Ye T, Deng L, Xiang J, Zhang Y, Hu H, Sun Y, Li Y, Shen L, Wang S, Xie L and Chen H. Predictors of pathologic tumor invasion and prognosis for ground glass opacity featured lung adenocarcinoma. *Ann Thorac Surg* 2018; 106: 1682-1690.
- [33] Lai J, Li Q, Fu F, Zhang Y, Li Y, Liu Q and Chen H. Subsolid lung adenocarcinomas: radiological, clinical and pathological features and outcomes. *Semin Thorac Cardiovasc Surg* 2022; 34: 702-710.
- [34] Lin M, Zhao X, Huang H, Lin H and Li K. A nomogram for predicting lymphovascular invasion in lung adenocarcinoma: a retrospective study. *BMC Pulm Med* 2024; 24: 588.
- [35] Xiang W, Xing Y, Jiang S, Chen G, Mao H, Labh K, Jia X and Sun X. Morphological factors differentiating between early lung adenocarcinomas appearing as pure ground-glass nodules measuring ≤ 10 mm on thin-section computed tomography. *Cancer Imaging* 2014; 14: 33.

Nomogram for predicting invasiveness of solitary GGNs

Table S1. Clinical and radiological features of all patients with solitary pulmonary ground-glass nodules

Features	Cohort			P
	Overall n = 846	Training Cohort n = 592	Validation Cohort n = 254	
Age (years), Median (Q1, Q3)	52 (41, 59)	51 (41, 59)	52 (41, 59)	0.539
Sex, n (%)				0.843
Female	629 (74.3%)	439 (74.2%)	190 (74.8%)	
Male	217 (25.7%)	153 (25.8%)	64 (25.2%)	
Smoking history, n (%)				0.987
No	733 (86.6%)	513 (86.7%)	220 (86.6%)	
Yes	113 (13.4%)	79 (13.3%)	34 (13.4%)	
Family history of cancer, n (%)				0.428
No	696 (82.3%)	483 (81.6%)	213 (83.9%)	
Yes	150 (17.7%)	109 (18.4%)	41 (16.1%)	
History of cancer, n (%)				0.084
No	780 (92.2%)	552 (93.2%)	228 (89.8%)	
Yes	66 (7.8%)	40 (6.8%)	26 (10.2%)	
Location, n (%)				0.974
RM	58 (6.9%)	39 (6.6%)	19 (7.5%)	
LL	123 (14.5%)	88 (14.9%)	35 (13.8%)	
LU	226 (26.7%)	157 (26.5%)	69 (27.2%)	
RL	169 (20.0%)	117 (19.8%)	52 (20.5%)	
RU	270 (31.9%)	191 (32.3%)	79 (31.1%)	
Maximum tumor diameter(cm), Median (Q1, Q3)	1.00 (0.80, 1.20)	0.90 (0.75, 1.20)	1.00 (0.80, 1.30)	0.637
CYFRA211, n (%)				0.750
Abnormal	138 (16.3%)	95 (16.0%)	43 (16.9%)	
Normal	708 (83.7%)	497 (84.0%)	211 (83.1%)	
CEA, n (%)				0.759
Abnormal	12 (1.4%)	8 (1.4%)	4 (1.6%)	
Normal	834 (98.6%)	584 (98.6%)	250 (98.4%)	
NSE, n (%)				0.950
Abnormal	46 (5.4%)	32 (5.4%)	14 (5.5%)	
Normal	800 (94.6%)	560 (94.6%)	240 (94.5%)	
CTR, n (%)				0.966
≤0.5	592 (70.0%)	414 (69.9%)	178 (70.1%)	
>0.5	254 (30.0%)	178 (30.1%)	76 (29.9%)	
Lobulation sign, n (%)				0.995
Absent	726 (85.8%)	508 (85.8%)	218 (85.8%)	
Present	120 (14.2%)	84 (14.2%)	36 (14.2%)	
Spiculation sign, n (%)				0.221
Absent	605 (71.5%)	416 (70.3%)	189 (74.4%)	
Present	241 (28.5%)	176 (29.7%)	65 (25.6%)	
Vacuole Sign, n (%)				0.822
Absent	428 (50.6%)	301 (50.8%)	127 (50.0%)	
Present	418 (49.4%)	291 (49.2%)	127 (50.0%)	
Pleura traction sign, n (%)				0.864
Absent	576 (68.1%)	402 (67.9%)	174 (68.5%)	
Present	270 (31.9%)	190 (32.1%)	80 (31.5%)	
Vascular convergence sign, n (%)				0.586
Absent	24 (2.8%)	18 (3.0%)	6 (2.4%)	
Present	822 (97.2%)	574 (97.0%)	248 (97.6%)	

Nomogram for predicting invasiveness of solitary GGNs

Air Bronchogram Sign, n (%)				0.890
Absent	734 (86.8%)	513 (86.7%)	221 (87.0%)	
Present	112 (13.2%)	79 (13.3%)	33 (13.0%)	
Solid component on mediastinal window, n (%)				0.201
Absent	704 (83.2%)	499 (84.3%)	205 (80.7%)	
Present	142 (16.8%)	93 (15.7%)	49 (19.3%)	

NIAC, non-invasive adenocarcinoma; IAC, invasive adenocarcinoma; CTR, consolidation-to-tumor ratio; RM, right middle; RU, right upper; RL, right lower; LL, left lower; LU, left upper; CYFRA211, cytokeratin fragment 19; NSE, neuron-specific enolase; CEA, carcinoembryonic antigen.

Table S2. The coefficients of Lasso regression analysis (λ_{1se})

Variables	Coefficient
(Intercept)	-5.319818772
Age	0.007635497
Sex	0.000000000
Smoking history	0.000000000
Family history of cancer	0.000000000
History of cancer	0.000000000
LL	0.000000000
LU	-0.116112645
RL	0.000000000
RU	0.000000000
Maximum tumor diameter	2.789374934
CYFRA211	0.000000000
CEA	0.000000000
NSE	0.000000000
0.5<CTR<1	0.676368714
Lobulation sign	0.304247251
Spiculation sign	0.000000000
Vacuole Sign	0.233974832
Pleura traction sign	0.532778007
Vascular convergence sign	0.000000000
Air Bronchogram Sign	0.124118177
Solid component on mediastinal window	0.756640240

CTR, consolidation-to-tumor ratio; RM, right middle; RU, right upper; RL, right lower; LL, left lower; LU, left upper; CYFRA211, cytokeratin fragment 19; NSE, neuron-specific enolase; CEA, carcinoembryonic antigen.

Table S3. Variance inflation factor (VIF) analysis

Features	VIF
Age	1.255
Location	1.016
Maximum tumor diameter	1.872
CTR	1.261
Lobulation sign	1.264
Vacuole Sign	1.049
Pleura traction sign	1.170
Air Bronchogram Sign	1.423
Solid component on mediastinal window	1.627

VIF, variance inflation factor; CTR, consolidation-to-tumor ratio. All VIF values are below 5, indicating no severe multicollinearity.