Am J Cancer Res 2025;15(10):4200-4215
www.ajcr.us /ISSN:2156-6976/ajcr0167670

Original Article

Imaging features and clinical markers for predicting postoperative recurrence in early-stage lung cancer

Shuhong Guan[1*], Ying Ding[2*], Long Zhang[2], Junkang Huangfu[2], Tianyu Chen[2], Jun Zhou[1]

[1]*Department of Respiratory and Critical Care Medicine, Changzhou First People's Hospital, Changzhou 212003, Jiangsu, China;* [2]*Suzhou Medical College of Soochow University, Suzhou 215000, Jiangsu, China.* [*]*Equal contributors.*

Received July 23, 2025; Accepted September 29, 2025; Epub October 15, 2025; Published October 30, 2025

Abstract: This study aimed to evaluate the predictive value of imaging features and clinical markers for postoperative recurrence in patients with early-stage lung cancer and to establish a nomogram and a neural network model for prediction. A total of 439 patients with early-stage lung cancer who underwent surgical treatment at Changzhou First People's Hospital between January 2020 and January 2023 were retrospectively enrolled. Clinical characteristics, preoperative imaging findings, postoperative pathology, laboratory test results, and recurrence status were collected. By April 1, 2025, 85 of the 439 patients had relapsed, accounting for 19.36% of the cohort. Univariate analysis revealed significant differences between the recurrence and non-recurrence groups in terms of age, tumor density, CYFRA21-1, CA19-9, and CA125 (all $P<0.05$). Multivariate logistic regression analysis identified solid tumor density (OR=2.132), CYFRA21-1≥3.3 ng/mL (OR=2.307), CA19-9≥37 U/mL (OR=2.901), and CA125≥35 U/mL (OR=5.974) as independent risk factors for postoperative recurrence. In addition, increasing age was associated with higher recurrence risk (OR=1.121) (all $P<0.05$). The nomogram model based on these predictors demonstrated an area under the receiver operating characteristic curve (AUC) of 0.804 in the training set and 0.760 in the validation set, both exceeding 0.7, indicating good predictive performance. The neural network model yielded AUC values of 0.882 in the training set and 0.734 in the validation set, also showing favorable performance. DeLong test revealed a significant difference in AUC between the two models in the training set (Z=-3.514, $P<0.001$), but no significant difference in the validation set (Z=0.374, $P=0.709$). External validation showed that the nomogram achieved a sensitivity of 74.36%, specificity of 73.84%, and accuracy of 73.93%, while the neural network model achieved a sensitivity of 79.49%, specificity of 68.60%, and accuracy of 70.62%. In conclusion, this study developed a nomogram and a neural network model incorporating imaging features and clinical markers to predict postoperative recurrence in early lung cancer. These models may serve as valuable tools to identify high-risk patients and guide individualized clinical management.

Keywords: Early lung cancer, imaging, clinical indicators, influencing factors, prediction model

Introduction

With advances in imaging techniques, the detection rate of early-stage lung cancer (stages I and II) has significantly increased [1-3]. Surgery is the primary treatment approach; however, postoperative recurrence remains a major factor influencing long-term survival [4, 5]. Recurrence risk varies among individuals and is influenced by multiple clinical and pathological factors. Despite the availability of diverse treatment options, including adjuvant/neoadjuvant chemotherapy, targeted therapy, and immunotherapy, the necessity of postoperative adjuvant therapy in early-stage disease remains controversial [6-8]. Current guidelines recommend adjuvant chemotherapy for stage IB patients with high-risk features such as poor differentiation, vascular invasion, or intrapulmonary dissemination, while stage-IA patients are advised to undergo regular follow-up [9-11]. However, some stage-IA patients still experience poor outcomes, indicating that TNM staging alone is insufficient to guide treatment and follow-up. This approach neglects other prognostic factors, potentially delaying intervention in high-risk patients.

High-resolution imaging techniques (e.g., low-dose spiral CT) can detect subtle pulmonary

https://doi.org/10.62347/YERO8788

abnormalities during follow-up, serving as early indicators of recurrence [12-15]. However, relying solely on imaging manifestations is sometimes insufficient for early postoperative recurrence prediction, as some non-specific lesions (such as inflammation or fibrosis) may also present similar characteristics on imaging, leading to misdiagnosis or missed detection. Clinical markers also hold promise for assessing recurrence risk. Blood-based tumor markers and genetic alterations can provide supportive evidence [16-18], but their accuracy is limited by biological variability, and a single marker rarely captures the full complexity of lung cancer pathophysiology [19-22]. Therefore, integrating imaging features with clinical markers is critical to improving the accuracy and reliability of recurrence risk assessment.

In recent years, the rapid development of big data and artificial intelligence technology has enabled predictive modeling based on multidimensional indicators, offering new approaches for the diagnosis and prognosis of early-stage lung cancer [23-25]. Building on this, this study aimed to investigate imaging features and clinical markers in early-stage lung cancer, analyze their combined predictive value, and develop two predictive models - a nomogram and a neural network. These models may provide novel strategies for accurate recurrence risk assessment, ultimately enhancing prognosis and quality of life in patients with early-stage lung cancer.

Materials and methods

Patient selection

A total of 439 patients with early-stage lung cancer who underwent surgical treatment in Changzhou First People's Hospital between January 2020 and January 2023 were retrospectively enrolled.

Inclusion criteria: (1) Patients who underwent radical resection with postoperative pathological confirmation of stage IA, IB or IIA NSCLC; (2) Postoperative follow-up and adjuvant therapy performed in our hospital; (3) Aged between 18 and 80 years; (4) Complete clinical data, including preoperative imaging, postoperative pathological reports, and laboratory tests.

Exclusion criteria: (1) Presence of other malignant tumors; (2) Severe cardiopulmonary dysfunction or other major systemic diseases; (3) Severe intraoperative complications (e.g., massive bleeding, infection); (4) Inability to obtain recurrence information through telephone follow-up.

This study was approved by the Ethics Committee of Changzhou First People's Hospital. A total of 511 patients were initially screened. According to the exclusion criteria, 15 patients with other malignancies, 23 patients with severe cardiopulmonary dysfunction or other systemic diseases, and 7 patients with severe intraoperative or postoperative complications (such as massive hemorrhage and infection) were excluded. Additionally, 27 patients were lost to follow-up, as recurrence status could not be confirmed by telephone. Finally, 439 patients with early-stage lung cancer were included in this study. The screening process is illustrated in **Figure 1**.

Data extraction

Clinical information was collected from the electronic medical record system of our hospital, including demographic characteristics, preoperative imaging, postoperative pathology, laboratory indicators, and recurrence status. (1) General characteristics: sex, age, body mass index (BMI), smoking history, alcohol consumption history, chronic disease (diabetes or hypertension), type of surgery (lobectomy or sublobectomy), and adjuvant therapy (chemotherapy, targeted therapy, or immunotherapy). (2) Imaging features: tumor location (peripheral vs. central), tumor density (solid, part-solid, or pure ground-glass), and margin characteristics (pleural indentation, lobulation, spiculation, cavity sign). (3) Pathological findings: pathological type (adenocarcinoma vs. other) and tumor stage (IA, IB or IIA). (4) Laboratory indicators: neuron-specific enolase (NSE), cytokeratin fragments 21-1 (CYFRA21-1), squamous cell carcinoma antigen (SCC), carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA19-9), and carbohydrate antigen 125 (CA125). Each marker was categorized as normal or abnormal according to the cut-off values (NSE: 16.3 ng/mL, CYFRA21-1: 3.3 ng/mL, SCC: 1.5 ng/mL, CEA: 5 ng/mL, CA19-9: 37 U/mL, and CA125: 35 U/mL). (5) Postoperative recurrence: defined as the primary outcome. Recurrence status was determined by telephone follow-up until April 1, 2025. Patients with confirmed recur-

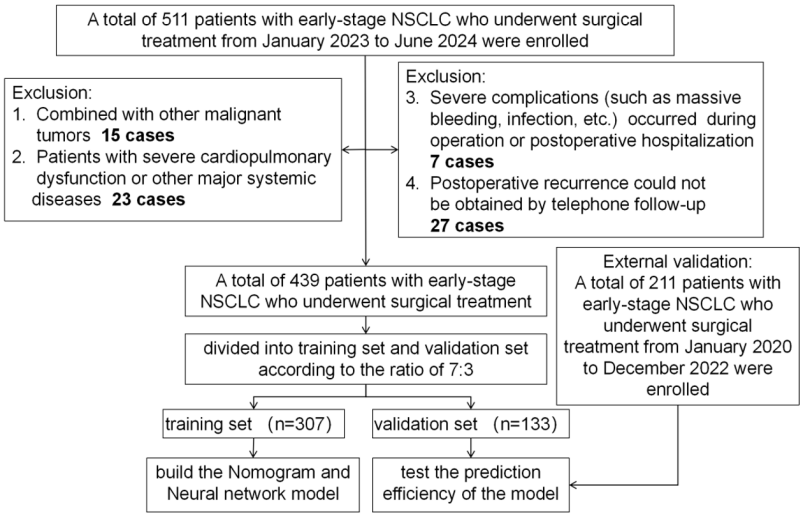

Figure 1. Study flowchart.

rence by this date were classified into the recurrence group, and those without recurrence into the non-recurrence group.

Model construction

(1) Nomogram model: patients were randomly divided into a training set and a validation set in a 7:3 ratio. The training set was used for model development, and the validation set for performance evaluation. First, clinical data were compared between the recurrence and non-recurrence groups, and variables with significant differences were identified. Then, these variables were included as independent variables, with recurrence as the dependent variable, and logistic regression analysis was performed to determine predictors of postoperative recurrence. Finally, the results of logistic regression were visualized to construct the nomogram, and model performance was assessed using the receiver operating characteristic (ROC) curve and calibration curve.

(2) Neural network model: the subjects were randomly divided into a training set and a validation set in the same 7:3 ration. The neural network consisted of an input layer, a hidden layer, and an output layer. Significant variables identified by univariate analysis were included in the input layer, with each node representing one variable. Recurrence was defined as the output variable, and the hidden layer performed automatic feature extraction through a multilayer perceptron. The model was implemented in R4.5.1 software using the *mlbench* and *neuralnet* packages. Detailed parameters were as follows: the input layer consists of 5 nodes, corresponding to variables from the multivariate regression analysis; the hidden layer contained 10 nodes; the hidden layer used the ReLU (Rectified Linear Unit) activation function, and the output layer used the Sigmoid activation function to yield recurrence probability. To reduce overfitting, L2 regularization was applied with a penalty parameter of 0.01. The model was trained using the Adam optimizer with a learning rate of 0.001 for 1000 epochs, with a batch size of 32.

Statistical analysis

SPSS 26.0 and R 4.5.1 software were used for statistical analysis. Continuous variables with normal distribution were expressed as mean ± standard deviation (SD), and between-group comparison was conducted using independent sample t test; those with non-normal distribution were expressed as median (interquartile range), and the between-group comparison was conducted using the rank-sum test. Categorical variables were expressed as frequency (percentage) and compared using the chi-square test. The DeLong test was applied to compare differences in the areas under the ROC curves (AUCs). A two-tailed $P<0.05$ was considered statistically significant.

Results

Clinical characteristics of the study population

A total of 439 patients with early-stage lung cancer were included, including 267 males (60.82%) and 172 females (39.18%), with an average age of (56.29±7.19) years. Detailed baseline characteristics are summarized in **Table 1**.

Postoperative recurrence

All 439 patients underwent surgical treatment. Representative imaging before surgery, after

Table 1. Clinical data of study subjects

Index	Circumstance
Gender	
Male	267 (60.82)
Female	172 (39.18)
Age (years)	56.29±7.19
BMI (kg/m^2)	24.16±1.23
Smoking	
Yes	211 (48.06)
No	228 (51.94)
Alcohol consumption	
Yes	237 (53.99)
No	202 (46.01)
Chronic disease	
With	134 (30.52)
Without	305 (69.48)
Type of surgery	
Lobectomy	301 (68.56)
Sublobectomy	138 (31.44)
Adjuvant therapy	
With	102 (23.23)
Without	337 (76.77)
Tumor type	
Peripheral	279 (63.55)
Central	160 (36.45)
Solid density	
Yes	228 (51.94)
No	211 (48.06)
Pleural indentation	
With	146 (33.26)
Without	293 (66.74)
Lobulation	
With	289 (65.83)
Without	150 (34.17)
Spiculation	
With	198 (45.10)
Without	241 (54.90)
Cavity sign	
With	76 (17.31)
Without	363 (82.69)
Pathological classification	
Adenocarcinoma	362 (82.46)
other	77 (17.54)
Tumor stage	
IA and IB	351 (79.95)
IIA	88 (20.05)
NSE (ng/mL)	
≥16.3	208 (47.38)
<16.3	231 (52.62)
CYFRA21-1 (ng/mL)	
≥3.3	170 (38.72)
<3.3	269 (61.28)
SCC (ng/mL)	
≥1.5	82 (18.68)
<1.5	357 (81.32)
CEA (ng/mL)	
≥5	251 (57.18)
<5	188 (42.82)
CA19-9 (U/mL)	
≥37	152 (34.62)
<37	287 (65.38)
CA125 (U/mL)	
≥35	60 (13.67)
<35	379 (86.33)

surgery, and at 1 month postoperatively are shown in **Figure 2**. Among the 439 patients, 85 patients (19.36%) experienced recurrence. The cohort was randomly divided into a training set (n=307) and a validation set (n=133) at a 7:3 ratio. Group comparison, univariate and multivariate Logistic regression analysis were performed in the training set to construct the predictive model for postoperative recurrence. Model performance was further verified in the validation dataset. As shown in **Figure 3**, recurrence occurred in 52 of 307 patients (16.94%) in the training set and in 33 of 133 patients (24.81%) in the validation set.

Comparison of clinical data between recurrence and non-recurrence groups

Clinical variables were compared between the recurrence and non-recurrence groups. As shown in **Table 2**, significant differences were observed in age, tumor density, CYFRA21-1, CA19-9, and CA125 levels (all $P<0.05$), whereas other indicators were comparable between groups (all $P>0.05$).

Univariate and multivariate logistic regression analysis

Univariate logistic regression was performed in the training set. Variables that differed between the two groups (age, tumor density, CYFRA21-1, CA19-9 and CA125; see **Table 3** for variable assignments) were included as independent variables, with postoperative

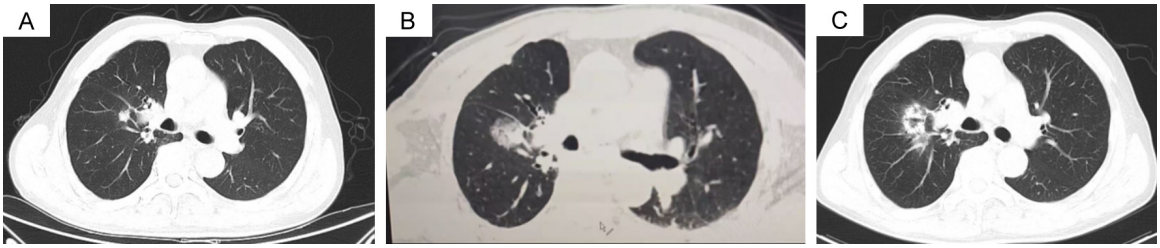

Figure 2. Typical imaging data of a patient. (A) Before surgery, (B) after surgery, (C) 1 month after surgery.

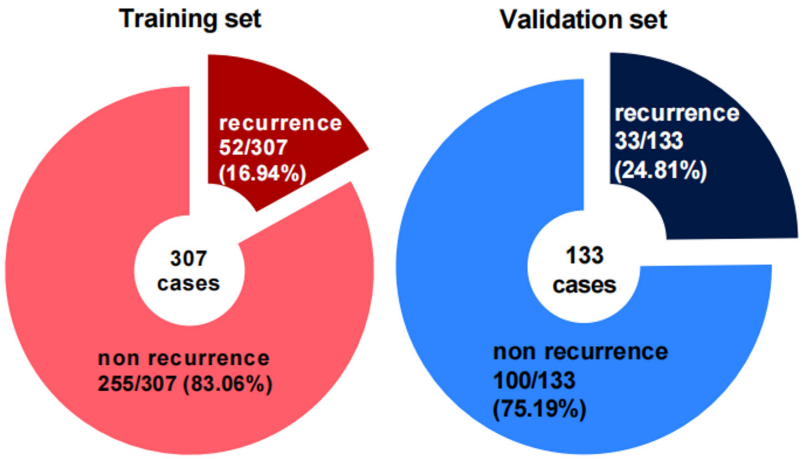

Figure 3. Proportion of subjects with recurrence in the two data sets.

recurrence as the dependent variable. As shown in **Table 4**, all five variables were significantly associated with postoperative recurrence (all P<0.05).

Multivariate logistic regression analysis was then conducted using the variables identified above. As presented in **Table 5**, older age (OR=1.121), solid tumor density (OR=2.132), CYFRA21-1≥3.3 ng/mL (OR=2.307), CA19-9≥37 U/mL (OR=2.901), and CA125≥35 U/mL (OR=5.974) were independent risk factors for postoperative recurrence (all *P*<0.05).

Construction of the nomogram

Significant predictors identified in the multivariate analysis were incorporated into a logistic regression model, and the results were visualized as a nomogram (**Figure 4**). The nomogram assigns each predictor a weighted score according to its contribution to recurrence risk. By summing the scores of individual predictors for a given patient, a total score is obtained, which can be projected onto the risk scale to estimate the probability of postoperative recurrence.

Validation of the nomogram model

The ROC curve was used to test the discrimination of the nomogram model. As shown in **Figure 5A**, **5B**, the AUC was 0.804 (95% CI=0.736-0.871) in the training set and 0.760 (95% CI=0.673-0.847) in the validation set, both exceeding 0.7, indicating good discrimination. Calibration curves (**Figure 6A**, **6B**) demonstrated that predicted recurrence risk closely matched the observed outcomes in both datasets.

Comparison of predictive efficacy between the model and individual factors

The ROC curves of the nomogram and individual predictors are presented in **Figure 7**. DeLong's test showed that the AUC of the nomogram was significantly higher than that of age, tumor density, CYFRA21-1, CA19-9, and CA125 (Z=3.862, 4.297, 4.795, 4.831 and 5.348, respectively; all *P*<0.05). The AUCs and the cut-off values of each factor are summarized in **Table 6** (cut-off values of binary variables correspond to their respective risk thresholds).

Neural network model

The neural network architecture is shown in **Figure 8**, with five input nodes corresponding to the variables identified in the multivariate regression analysis. Importance analysis

Table 2. Differences in clinical data between two groups

Index	Recurrence group (n=85)	Non-recurrence group (n=354)	χ^2/t	*P*
Gender			0.668	0.414
Male	55 (20.60)	212 (79.40)		
Female	30 (17.44)	142 (82.56)		
Age (years)	59.82±6.40	55.45±7.11	5.190	0.000
BMI (kg/m^2)	24.01±1.19	24.20±1.23	1.297	0.197
Smoking			2.984	0.084
Yes	48 (22.75)	163 (77.25)		
No	37 (16.23)	191 (83.77)		
Alcohol consumption			0.569	0.451
Yes	49 (20.68)	188 (79.32)		
No	36 (17.82)	166 (82.18)		
Chronic disease			1.758	0.185
With	31 (23.13)	103 (76.87)		
Without	54 (17.70)	251 (82.30)		
Type of surgery			0.501	0.479
Lobectomy	61 (20.27)	240 (79.73)		
Sublobectomy	24 (17.39)	114 (82.61)		
Adjuvant therapy			3.726	0.054
With	13 (12.75)	89 (87.25)		
Without	72 (21.36)	265 (78.64)		
Tumor type			0.247	0.619
Peripheral	56 (20.07)	223 (79.93)		
Central	29 (18.13)	131 (81.88)		
Solid density			16.602	0.000
Yes	61 (26.75)	167 (73.25)		
No	24 (11.37)	187 (88.63)		
Pleural indentation			0.319	0.572
With	33 (22.60)	113 (77.40)		
Without	52 (17.75)	241 (82.25)		
Lobulation			0.071	0.790
With	57 (19.72)	232 (80.28)		
Without	28 (18.67)	122 (81.33)		
Spiculation			1.281	0.258
With	43 (21.72)	155 (78.28)		
Without	42 (17.43)	199 (82.57)		
Cavity sign			1.871	0.171
With	19 (25.00)	57 (75.00)		
Without	66 (18.18)	297 (81.82)		
Pathological classification			0.964	0.326
Adenocarcinoma	67 (18.51)	295 (81.49)		
other	18 (23.38)	59 (76.62)		
Tumor stage			3.235	0.072
IA and IB	62 (17.66)	289 (82.34)		
IIA	23 (26.14)	65 (73.86)		
NSE (ng/mL)			3.494	0.062
≥16.3	48 (23.08)	160 (76.92)		
<16.3	37 (16.02)	194 (83.98)		

CYFRA21-1 (ng/mL)			8.979	0.003
≥3.3	45 (26.47)	125 (73.53)		
<3.3	40 (14.87)	229 (85.13)		
SCC (ng/mL)			3.601	0.058
≥1.5	22 (26.83)	60 (73.17)		
<1.5	63 (17.65)	294 (82.35)		
CEA (ng/mL)			3.441	0.064
≥5	41 (16.33)	210 (83.67)		
<5	44 (23.40)	144 (76.60)		
CA19-9 (U/mL)			11.868	0.001
≥37	43 (28.29)	109 (71.71)		
<37	42 (14.63)	245 (85.37)		
CA125 (U/mL)			33.186	0.000
≥35	28 (46.67)	32 (53.33)		
<35	57 (15.04)	322 (84.96)		

Table 3. Variable assignment

Variable	Variable assignment
Dependent variable	
Recurrence situation	1= Recurrence, 0= Non-recurrence
Independent variable	
Age	Enter actual value
Tumor density	1= "solid", 0= "partial solid/pure ground glass density"
CYFRA21-1	1= "≥3.3 ng/mL", 0= "<3.3 ng/mL"
CA19-9	1= "≥37 U/mL", 0= "<37 U/mL"
CA125	1= "≥35 U/mL", 0= "<35 U/mL"

Table 4. Univariate analysis

Variable	β	SE	Wald χ^2	P	OR	95% CI
Age	0.088	0.022	15.324	0.000	1.091	1.045-1.140
Tumor density is solid	0.824	0.321	6.587	0.010	2.280	1.215-4.278
CYFRA21-1≥3.3 ng/mL	0.855	0.309	7.680	0.006	2.352	1.284-4.306
CA19-9≥37 U/mL	0.801	0.309	6.736	0.009	2.228	1.217-4.079
CA125≥35 U/mL	1.742	0.376	21.416	0.000	5.707	2.729-11.934

Table 5. Multivariate logistic regression analysis

Variable	β	SE	Wald χ^2	P	OR	95% CI
Age	0.115	0.026	18.699	0.000	1.121	1.065-1.181
Tumor density is solid	0.757	0.358	4.468	0.035	2.132	1.057-4.300
CYFRA21-1≥3.3 ng/mL	0.836	0.346	5.824	0.016	2.307	1.170-4.550
CA19-9≥37 U/mL	1.065	0.355	8.986	0.003	2.901	1.446-5.820
CA125≥35 U/mL	1.787	0.429	17.344	0.000	5.974	2.576-13.854
Constant	-9.767	1.705	32.819	-	-	-

(**Figure 9**) indicated that the most influential predictors of recurrence were age, tumor density, CA125, CYFRA21-1, and CA19-9. As shown in **Figure 10**, the neural network demonstrated

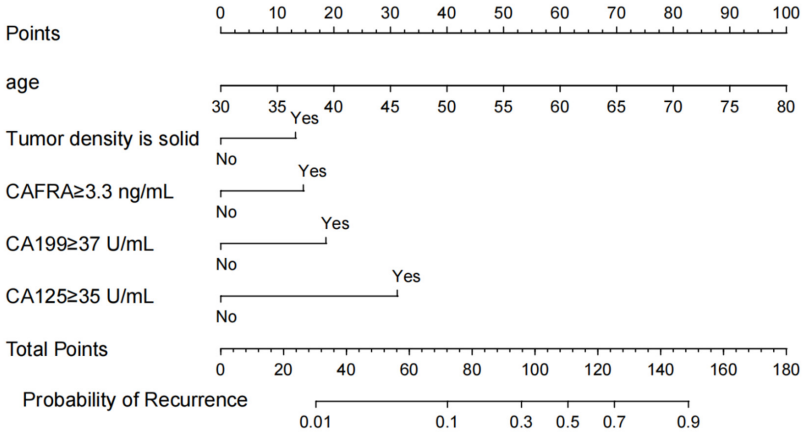

Figure 4. Nomogram model.

good predictive accuracy for both recurrent and non-recurrent cases. ROC analysis (**Figure 11**) showed an AUC of 0.882 (95% CI=0.838-0.926) in the training set and 0.734 (95% CI=0.632-0.836) in the validation set. DeLong's test indicated a significant difference between the neural network and nomogram in the training set (Z=-3.514, P<0.001), but no significant difference in the validation set (Z=0.374, P=0.709).

External model verification

According to the same inclusion and selection criteria, an independent cohort of 211 patients with early-stage lung cancer who underwent surgery between February 2023 and July 2024 was used for external validation. This cohort included 139 males and 72 females, with a mean age of (57.68±7.18) years. As of April 1, 2025, 39 patients (16.67%) had experienced recurrence. Using a cut-off value of 0.264, patients were stratified into high- and low-risk groups. For predicting postoperative recurrence, the nomogram achieved a sensitivity of 74.36% (29/39), specificity of 73.84% (127/172), and accuracy of 73.93% (156/211). The neural network model achieved a sensitivity of 79.49% (31/39), specificity of 68.60% (118/172), and accuracy of 70.62% (149/211). Detailed data are shown in **Table 7**.

Discussion

Lung cancer is a highly prevalent malignant tumor worldwide. In 2022, approximately 2.5 million new cases were reported, accounting for 12.4% of all cancers, with 1.8 million deaths, representing 18.7% of cancer-related mortality [26]. Non-small cell lung cancer (NSCLC), the most common subtype, is primarily managed through surgery, chemotherapy, and radiotherapy, which can alleviate symptoms and prolong survival. However, the 5-year survival rate remains below 20% [27-29]. In recent years, targeted therapy and immunotherapy have brought new hope, but the treatment of advanced lung cancer still faces challenges such as distant metastasis, low chemotherapy sensitivity, and drug resistance [30-33]. Therefore, early detection, timely intervention, and effective control are crucial for improving prognosis and survival rates.

In the present study, among 439 patients with early-stage lung cancer, 85 cases experienced postoperative recurrence (19.36%), consistent with 15.6% reported by Zhao et al. [34]. Multivariate analysis identified solid tumors (OR=2.132), CYFRA21-1≥3.3 ng/mL (OR=2.307), CA19-9≥37 U/mL (OR=2.901), and CA125≥35 U/mL (OR=5.974) as risk factors for postoperative recurrence in NSCLC patients. In addition, recurrence risk increased with the age (OR=1.121). Advanced age is associated with impaired immune function, poor postoperative recovery, higher complication rates, and a microenvironment more conducive to tumor recurrence, thus increasing recurrence risk [35]. Solid tumors are characterized by high cellular density and strong angiogenic potential, which facilitate dissemination via the blood or lymphatic system, increasing the risk of recurrence [35]. Increasing evidence has shown that ground-glass components is associated with a better prognosis. For instance, Hattori et al. [36, 37], in an analysis of 671 patients with stage IA NSCLC from the JCOG0201 trial, demonstrated that across all IA substages, patients with ground-glass opacity (GGO) components had higher 5-year survival rate compared with those with purely solid nodules, regardless of the proportion of solid components within mixed nodules. Similarly,

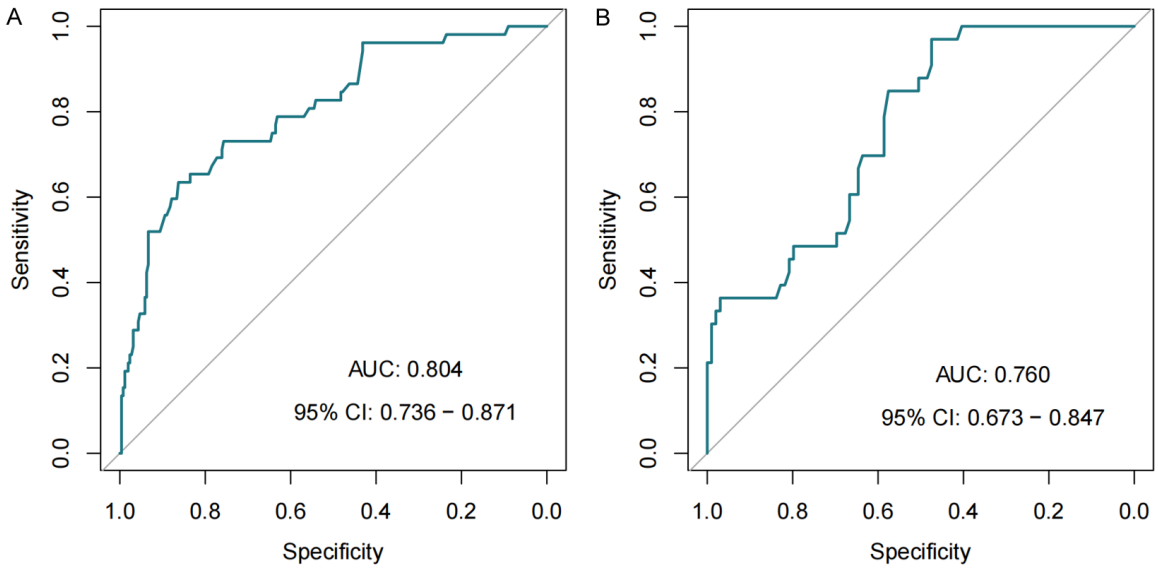

Figure 5. ROC curves for the nomogram in predicting recurrence in training set (A) and validation set (B).

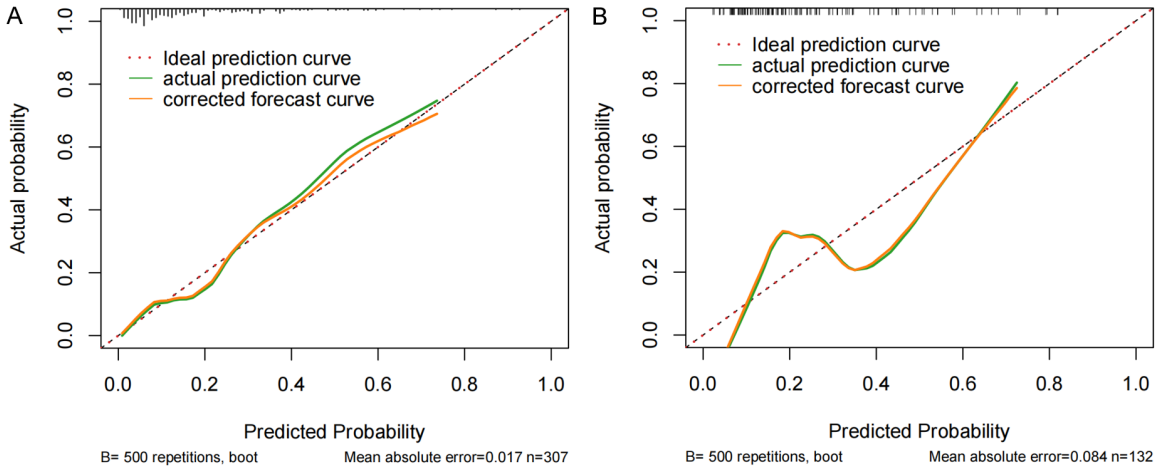

Figure 6. Calibration curves for the nomogram in training set (A) and validation set (B).

other researchers have reported a 5-year cumulative recurrence rate of only 7.5% in patients with lung adenocarcinoma harboring ground-glass components, markedly lower than the 24.5% recurrence observed in patients without GGO [38].

CYFRA21-1, a fragment of cytokeratin 19 mainly released by tumor cells into the bloodstream, is an important diagnostic biomarker for lung cancer and is closely related to postoperative recurrence and prognosis [39-41]. This study found that the risk of postoperative recurrence in early-stage lung cancer patients increases with the elevation of CYFRA21-1 levels, similar to the results of Zhang et al. [42]. Normally, CYFRA21-1 is distributed in lung epithelial cells but not released into circulation; however, during carcinogenesis it is released into the blood. Elevated levels may indicate aggressive tumor biology, higher tumor burden, and greater cellular heterogeneity, thereby increasing postoperative recurrence risk [43-45].

CA19-9 and CA125 are two widely used tumor markers used in the diagnosis, treatment monitoring, and prognosis assessment of various malignant tumors. Initially, CA19-9 was mainly used for the detection of pancreatic cancer and cholangiocarcinoma, while CA125 was com-

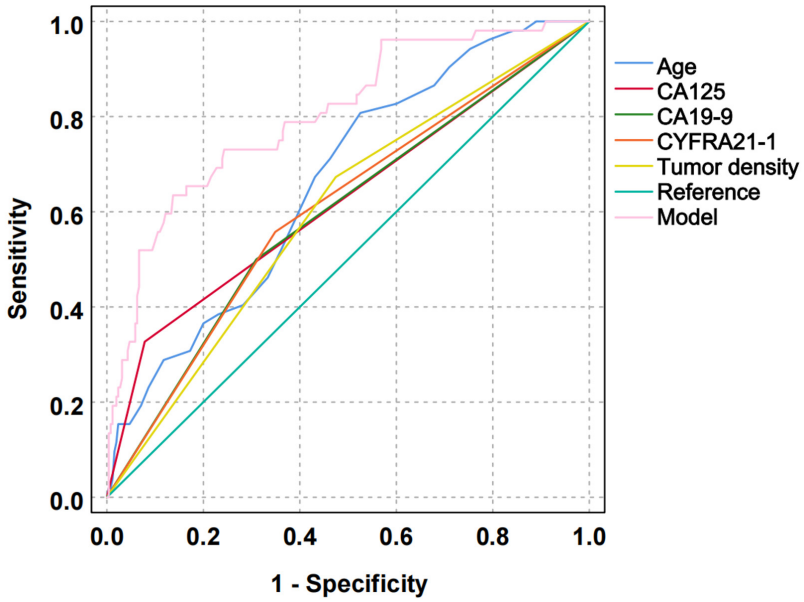

Figure 7. Comparison of predictive efficacy.

Table 6. Comparison of predictive efficacy

Indicators	AUC	Cut off value	95% CI	Z	P
Model	0.804	0.264	0.736-0.871	-	-
Age	0.663	54	0.607-0.716	3.862	0.000
Tumor density is solid	0.599	1	0.542-0.655	4.297	0.000
CYFRA21-1	0.604	1	0.547-0.659	4.795	0.000
CA19-9	0.594	1	0.538-0.650	4.831	0.000
CA125	0.624	1	0.567-0.679	5.348	0.000

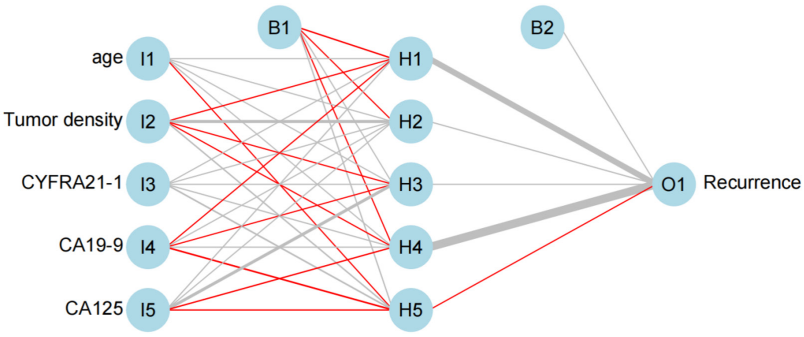

Figure 8. Neural network diagram.

monly used for the monitoring of ovarian cancer [46-48]. In recent years, the clinical significance of these two markers in lung cancer patients has also been confirmed [49]. CA125 may affect intercellular signal transduction through its interaction with cell surface receptors, thereby promoting the invasion and metastasis of tumor cells [49]. Our study suggest that elevated CA19-9 and CA125 can also predict postoperative recurrence risk in early-stage lung cancer patients. Sun et al. [49] demonstrated significant differences in CA19-9 and CA125 levels between lung cancer patients before and 6 months after surgery, with both markers associated with prognosis.

Currently, in clinical practice, postoperative risk assessment for lung cancer patients is largely based on TNM stage and histopathological type. However, compared with the 7th edition, the 8th edition of TNM staging system reclassified stage IB disease into stage IB (3 cm <T2a≤4 cm) and stage IIA (4 cm <T2b≤5 cm). For these substages, especially stage IB, the role of postoperative adjuvant therapy remains controversial [50]. Although recurrence rates are generally higher in patients with more advanced stages, our study found no significant difference in recurrence among different stages. This may be related to some patients receiving adjuvant therapy. According to the NCCN guidelines, surgical resection is recommended for stage I NSCLC, while stage II patients are recommended for surgery followed by adjuvant therapy. In addition, adjuvant therapy may be considered for high-risk stage IB patients [51]. The 2020 JCOG0802 trial included 1106 patients and demonstrated comparable 5-year recurrence-free survival and overall survival between lobectomy and sublobar resection [52].

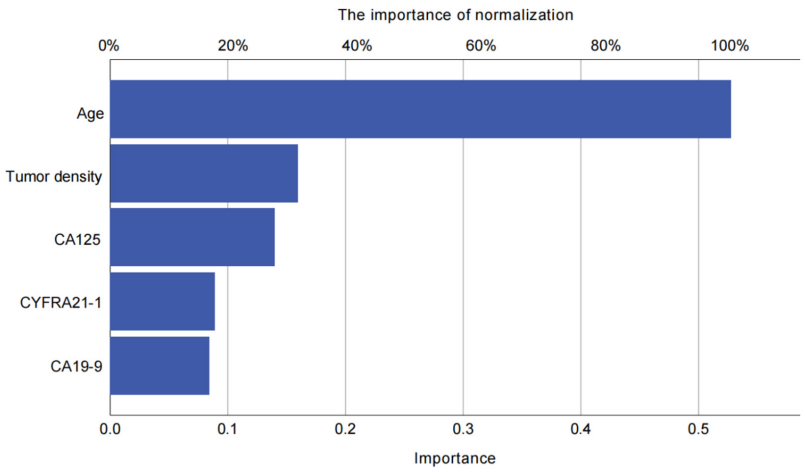

Figure 9. Importance ranking of independent variables.

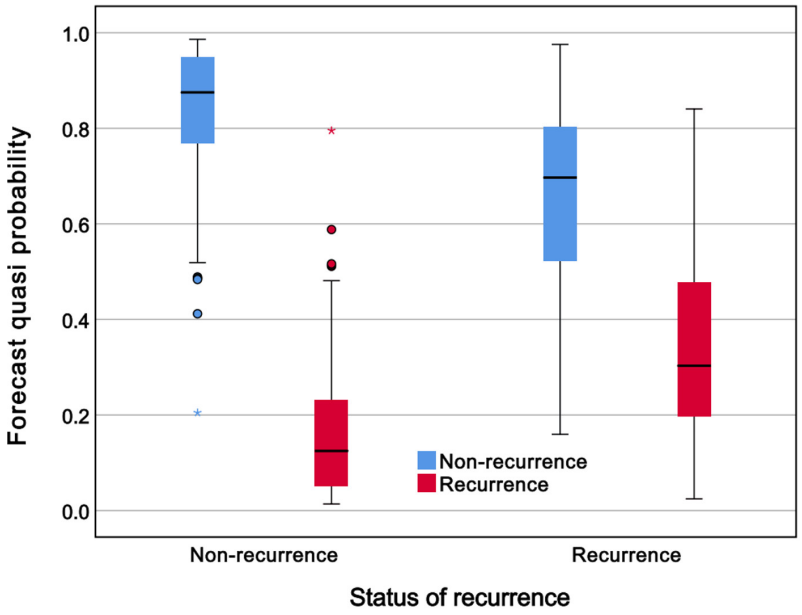

Figure 10. Predicted quasi-probability plots.

Similarly, the 2023 CALGB140503 trial found no difference in the 5-year disease-free survival between lobectomy and sublobar resection when the tumor diameter was ≤2 cm [53]. In line with these findings, our study observed no significant difference in postoperative recurrence among patients of different surgical types. This may reflect real-world clinical practice, in which treatment plans are tailored according to individual stage and condition, with adjuvant therapy administered when necessary. However, the absence of differences may also be attributable to the limited sample size in our cohort. Additionally, traditional imaging features such as spiculation and pleural indentation were not significantly associated with recurrence in the univariate analysis. These features may have greater prognostic relevance in advanced-stage lung cancer, and further studies are warranted to clarify their role in early-stage disease.

Disease prediction models are used to predict health conditions, disease risks, or disease progression. Among them, logistic regression and neural networks are two of the most commonly used models [54-61]. In this study, we developed both models to predict postoperative recurrence in early-stage lung cancer patients. The logistic regression model achieved an AUC of 0.804 in the training set and 0.760 in the validation set, showing good discrimination and prediction accuracy. The neural network model achieved an AUC of 0.882 in the training set and 0.734 in the validation, also demonstrating good predictive performance. These results indicate that both models have good accuracy and reliability in predicting recurrence risk of early-stage lung cancer after surgery.

Limitations and prospects

This study has several limitations. First, the sample size was relatively limited; although the models performed well in internal and external validation, their generalizability requires further assessment in larger multicenter cohorts. Second, the follow-up periods of the study subjects were not fully consistent, which may affect

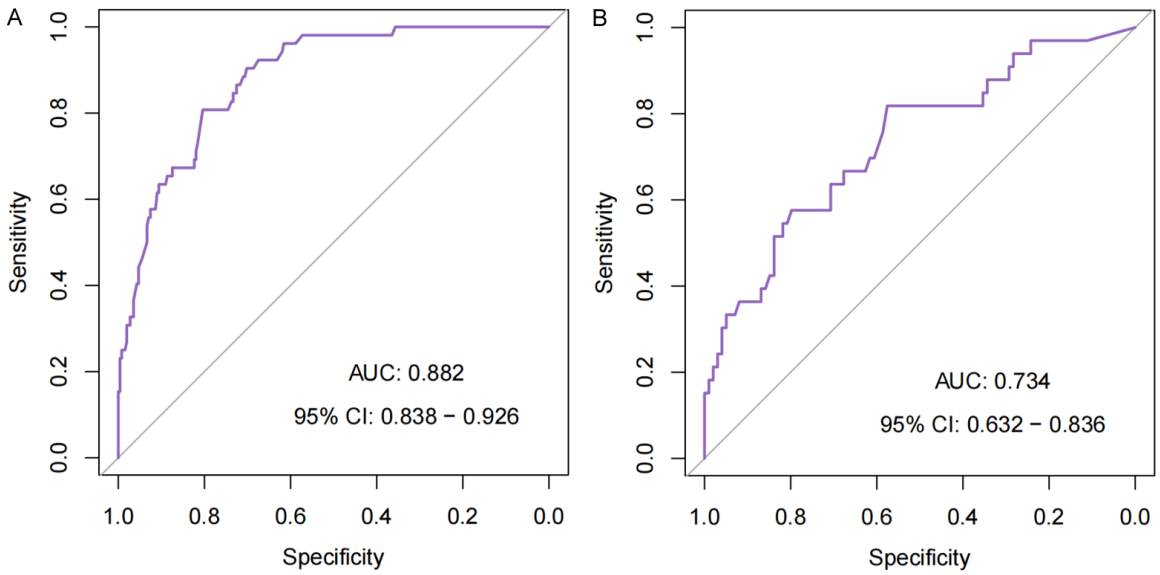

Figure 11. ROC curves of the neural network model in predicting recurrence risk in training set (A) and validation set (B).

Table 7. Validation of the model

Model	Forecast situation	Actual recurrence situation		Total
		Recurrence	Non-recurrence	
Nomogram model	Recurrence	29	45	74
	Non-recurrence	10	127	137
Total		39	172	211
Neural network model	Recurrence	31	54	85
	Non-recurrence	8	118	126
Total		39	172	211

the reliability of recurrence assessment. Future prospective studies with uniform follow-up periods are needed to address this issue. Third, the models primarily relied on static preoperative and postoperative data and did not incorporate dynamic monitoring indicators during follow-up. Future research could incorporate postoperative dynamic monitoring data to further enhance the predictive performance of the model.

Conclusions

The nomogram model and neural network model constructed in this study both showed high predictive efficiency and effectively identified patients at elevated recurrence risk. This provides strong support for clinicians in developing individualized treatment plans. For patients with a higher risk of recurrence, more intensive follow-ups (e.g., shortening the regular 6-month interval to 3 months) and timely adjuvant treatments - determined by a multidisciplinary team based on a comprehensive assessment of the patient's specific condition - may reduce the risk of recurrence. Conversely, for low-risk patients, follow-up frequency can be appropriately reduced to alleviate the psychological and economic burdens on the patients.

Acknowledgements

This work was supported by Changzhou Social Development Science and Technology Support Project (CE20235057).

Disclosure of conflict of interest

None.

Address correspondence to: Jun Zhou, Department of Respiratory and Critical Care Medicine, Changzhou First People's Hospital, No. 185, Qianqian Street, Changzhou 212003, Jiangsu, China. Tel: +86-0519-68870008; E-mail: yunxu-1969@163.com

References

[1] Zheng X, Li R, Fan L, Ge Y, Li W and Feng F. Prognostic predictors of radical resection of stage I-IIIB non-small cell lung cancer: the role of preoperative CT texture features, conventional imaging features, and clinical features in a retrospectively analyzed. BMC Pulm Med 2023; 23: 122.

[2] Zeng D, Tang X, Nong F, Yi J, Yao Y and Luo S. Construction and validation of a recurrent risk nomogram model for non-small cell lung cancer within 1 year after radical resection. J Oncol 2022; 2022: 8967162.

[3] Tian C, Liu G, Xu Y, Xia G, Zhang T, Huang J, Liu F and Li B. Beneficial effects of postoperative radiotherapy for IIIA-N2 non-small cell lung cancer after radical resection analysed using the propensity score-matching method. Oncol Lett 2023; 25: 205.

[4] Mei W, Yao W, Song Z, Jiao W, Zhu L, Huang Q, An C, Shi J, Yu G, Sun P, Zhang Y, Shen J, Xu C, Yang H, Wang Q and Zhu Z. Development and validation of prognostic nomogram for T(1-3) N(0)M(0) non-small cell lung cancer after curative resection. BMC Cancer 2023; 23: 715.

[5] Akcam TI, Tekneci AK, Ergin TM, Memmedov R, Ergonul AG, Ozdil A, Turhan K, Cakan A and Cagirici U. Factors influencing postoperative recurrence of early-stage non-small cell lung cancer. Acta Chir Belg 2024; 124: 121-130.

[6] Grant C, Hagopian G and Nagasaka M. Neoadjuvant therapy in non-small cell lung cancer. Crit Rev Oncol Hematol 2023; 190: 104080.

[7] Tan AC and Tan DSW. Targeted therapies for lung cancer patients with oncogenic driver molecular alterations. J Clin Oncol 2022; 40: 611-625.

[8] Lahiri A, Maji A, Potdar PD, Singh N, Parikh P, Bisht B, Mukherjee A and Paul MK. Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer 2023; 22: 40.

[9] Zhou K, Zhao Y, Liang L, Cao J, Lin H, Peng Z and Mei J. Adjuvant chemotherapy may improve long-term outcomes in stage IB non-small cell lung cancer patients with previous malignancies: a propensity score-matched analysis. Front Oncol 2022; 12: 938195.

[10] Chen Z, Wu X, Fang T, Ge Z, Liu J, Wu Q, Zhou L, Shen J and Zhou C. Prognostic impact of tumor spread through air spaces for T2aN0 stage IB non-small cell lung cancer. Cancer Med 2023; 12: 15246-15255.

[11] Shimomura M, Miyagawa-Hayashino A, Omatsu I, Asai Y, Ishihara S, Okada S, Konishi E, Teramukai S and Inoue M. Spread through air spaces is a powerful prognostic predictor in patients with completely resected pathological stage I lung adenocarcinoma. Lung Cancer 2022; 174: 165-171.

[12] Hage Chehade A, Abdallah N, Marion JM, Oueidat M and Chauvet P. Lung and colon cancer classification using medical imaging: a feature engineering approach. Phys Eng Sci Med 2022; 45: 729-746.

[13] Heidt CM, Bohn JR, Stollmayer R, von Stackelberg O, Rheinheimer S, Bozorgmehr F, Senghas K, Schlamp K, Weinheimer O, Giesel FL, Kauczor HU, Heussel CP and Heussel G. Delta-radiomics features of ADC maps as early predictors of treatment response in lung cancer. Insights Imaging 2024; 15: 218.

[14] Prosper AE, Kammer MN, Maldonado F, Aberle DR and Hsu W. Expanding role of advanced image analysis in ct-detected indeterminate pulmonary nodules and early lung cancer characterization. Radiology 2023; 309: e222904.

[15] Guo ZJ, Zhang Q, Liu Y, Bai ZM, Qiang L, Li HY, Zhang YH, Chi HW, Men M, Xu Q, Liang F, Feng WQ, Lu JY, Liu HT and Zeng YH. Preliminary investigation and imaging analysis of early lung cancer screening among petroleum workers in north China. Cancer Invest 2021; 39: 321-332.

[16] Tsukui T, Wolters PJ and Sheppard D. Alveolar fibroblast lineage orchestrates lung inflammation and fibrosis. Nature 2024; 631: 627-634.

[17] Davies MPA, Sato T, Ashoor H, Hou L, Liloglou T, Yang R and Field JK. Plasma protein biomarkers for early prediction of lung cancer. EBioMedicine 2023; 93: 104686.

[18] Li Y, Yu ND, Ye XL, Jiang MC and Chen XQ. Construction of lung cancer serum markers based on relieff feature selection. Comput Methods Biomech Biomed Engin 2024; 27: 1215-1223.

[19] Gasparri R, Noberini R, Cuomo A, Yadav A, Tricarico D, Salvetto C, Maisonneuve P, Caminiti V, Sedda G, Sabalic A, Bonaldi T and Spaggiari L. Serum proteomics profiling identifies a preliminary signature for the diagnosis of early-stage lung cancer. Proteomics Clin Appl 2023; 17: e2200093.

[20] Duffy MJ. Circulating tumor DNA (ctDNA) as a biomarker for lung cancer: early detection, monitoring and therapy prediction. Tumour Biol 2024; 46: S283-S295.

[21] Wadowska K, Bil-Lula I, Trembecki L and Sliwinska-Mosson M. Genetic markers in lung cancer diagnosis: a review. Int J Mol Sci 2020; 21: 4569.

[22] Wang H, Li Y, Han J, Lin Q, Zhao L, Li Q, Zhao J, Li H, Wang Y and Hu C. A machine learning-based PET/CT model for automatic diagnosis of early-stage lung cancer. Front Oncol 2023; 13: 1192908.

[23] Chen H, Ma Y, Xu J, Wang W, Lu H, Quan C, Yang F, Lu Y, Wu H and Qiu M. Circulating microbiome DNA as biomarkers for early diagnosis and recurrence of lung cancer. Cell Rep Med 2024; 5: 101499.
[24] Sun T, Liu J, Yuan H, Li X and Yan H. Construction of a risk prediction model for lung infection after chemotherapy in lung cancer patients based on the machine learning algorithm. Front Oncol 2024; 14: 1403392.
[25] Yang Y, Xu L, Sun L, Zhang P and Farid SS. Machine learning application in personalised lung cancer recurrence and survivability prediction. Comput Struct Biotechnol J 2022; 20: 1811-1820.
[26] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-263.
[27] Li Y, Yan B and He S. Advances and challenges in the treatment of lung cancer. Biomed Pharmacother 2023; 169: 115891.
[28] Lam DC, Liam CK, Andarini S, Park S, Tan DSW, Singh N, Jang SH, Vardhanabhuti V, Ramos AB, Nakayama T, Nhung NV, Ashizawa K, Chang YC, Tscheikuna J, Van CC, Chan WY, Lai YH and Yang PC. Lung cancer screening in Asia: an expert consensus report. J Thorac Oncol 2023; 18: 1303-1322.
[29] Berg CD, Schiller JH, Boffetta P, Cai J, Connolly C, Kerpel-Fronius A, Kitts AB, Lam DCL, Mohan A, Myers R, Suri T, Tammemagi MC, Yang D and Lam S; International Association for the Study of Lung Cancer (IASLC) Early Detection and Screening Committee. Air pollution and lung cancer: a review by international association for the study of lung cancer early detection and screening committee. J Thorac Oncol 2023; 18: 1277-1289.
[30] Li S, Wang A, Wu Y, He S, Shuai W, Zhao M, Zhu Y, Hu X, Luo Y and Wang G. Targeted therapy for non-small-cell lung cancer: new insights into regulated cell death combined with immunotherapy. Immunol Rev 2024; 321: 300-334.
[31] Presley CJ, Dalal N, Davenport AP, Gounden A, Ramchandran K and Tonorezos E. Survivorship in advanced lung cancer: understanding a new landscape and opportunities. Am Soc Clin Oncol Educ Book 2024; 44: e433298.
[32] Passaro A, Wang J, Wang Y, Lee SH, Melosky B, Shih JY, Wang J, Azuma K, Juan-Vidal O, Cobo M, Felip E, Girard N, Cortot AB, Califano R, Cappuzzo F, Owen S, Popat S, Tan JL, Salinas J, Tomasini P, Gentzler RD, William WN Jr, Reckamp KL, Takahashi T, Ganguly S, Kowalski DM, Bearz A, MacKean M, Barala P, Bourla AB, Girvin A, Greger J, Millington D, Withelder M, Xie J, Sun T, Shah S, Diorio B, Knoblauch RE, Bauml JM, Campelo RG and Cho BC; MARIPOSA-2 Investigators. Amivantamab plus chemotherapy with and without lazertinib in EGFR-mutant advanced NSCLC after disease progression on osimertinib: primary results from the phase III MARIPOSA-2 study. Ann Oncol 2024; 35: 77-90.
[33] Bessede A, Peyraud F, Besse B, Cousin S, Cabart M, Chomy F, Rey C, Lara O, Odin O, Nafia I, Vanhersecke L, Barlesi F, Guegan JP and Italiano A. TROP2 is associated with primary resistance to immune checkpoint inhibition in patients with advanced non-small cell lung cancer. Clin Cancer Res 2024; 30: 779-785.
[34] Zhao W, Sun Y, Liu X, Zhang M and Zhu B. Prediction model for the early recurrence of stage IA-IIA non-small cell lung cancer based on hematological indexes and imaging features. Discov Oncol 2025; 16: 684.
[35] Long B, Xiong Z, Liu S, Cheng Y, Li M and Liao W. Clinic information, pathological, and imaging characteristics in 2 058 surgical patients with lung cancer from a single center. Zhong Nan Da Xue Xue Bao Yi Xue Ban 2024; 49: 247-255.
[36] Hattori A, Suzuki K, Takamochi K, Wakabayashi M, Aokage K, Saji H and Watanabe SI; Japan Clinical Oncology Group Lung Cancer Surgical Study Group. Prognostic impact of a ground-glass opacity component in clinical stage IA non-small cell lung cancer. J Thorac Cardiovasc Surg 2021; 161: 1469-1480.
[37] Hattori A, Matsunaga T, Fukui M, Takamochi K and Suzuki K. Prognostic influence of a ground-glass opacity component in hypermetabolic lung adenocarcinoma. Eur J Cardiothorac Surg 2022; 61: 249-256.
[38] Xu SJ, Chen RQ, Tu JH, You CX, Chen C, Zhang ZF, Divisi D, Migliore M, Bongiolatti S, Durand M, Sato M, Kuroda H, Yang CF, Yu SB and Chen SC. Effects of a ground-glass opacity component on the recurrence and survival of pathological stage IA3 lung adenocarcinoma: a multi-institutional retrospective study. Transl Lung Cancer Res 2023; 12: 1078-1092.
[39] Fu Y, Li D, Zhu Y, Yan S, Wang X, Lian Z, Xie Q and He Z. Application value of CYFRA21-1 combined with NSE, CEA, and SCC-Ag in lung cancer. Clin Lab 2024; 70.
[40] Li J, Chen Y, Wang X, Wang C and Xiao M. The value of combined detection of CEA, CYFRA21-1, SCC-Ag, and pro-GRP in the differential diagnosis of lung cancer. Transl Cancer Res 2021; 10: 1900-1906.
[41] Huang Z, Peng K, Hong Z, Zhang P and Kang M. Nomogram for predicting recurrence and metastasis of stage IA lung adenocarcinoma

treated by videoassisted thoracoscopic lobectomy. Asian J Surg 2022; 45: 2691-2699.
[42] Zhang Z, Xie S, Cai W, Hong ZN, Yang C, Lin Y, Zhu J, Lin Z, Christoph DC, Bohnenberger H, Kepka L, Brueckl WM, Van Houtte P, Kang M and Lin J. A nomogram to predict the recurrence-free survival and analyze the utility of chemotherapy in stage IB non-small cell lung cancer. Transl Lung Cancer Res 2022; 11: 75-86.
[43] Yuan J, Sun Y, Wang K, Wang Z, Li D, Fan M, Bu X, Chen J, Wu Z, Geng H, Wu J, Xu Y, Chen M and Ren H. Development and validation of reassigned CEA, CYFRA21-1 and NSE-based models for lung cancer diagnosis and prognosis prediction. BMC Cancer 2022; 22: 686.
[44] Tang J, Ge QM, Huang R, Shu HY, Su T, Wu JL, Pan YC, Liang RB, Zhang LJ, Shao Y and Yu Y. Clinical significance of CYFRA21-1, AFP, CA-153, CEA, and CA-199 in the diagnosis of lung cancer ocular metastasis in hypertension population. Front Cardiovasc Med 2021; 8: 670594.
[45] Du Y, Wen Y and Huang J. Analysis of variation of serum CEA, SCC, CYFRA21-1 in patients with lung cancer and their diagnostic value with EBUS-TBNA. J Med Biochem 2024; 43: 363-371.
[46] Luo G, Jin K, Deng S, Cheng H, Fan Z, Gong Y, Qian Y, Huang Q, Ni Q, Liu C and Yu X. Roles of CA19-9 in pancreatic cancer: biomarker, predictor and promoter. Biochim Biophys Acta Rev Cancer 2021; 1875: 188409.
[47] Boyd LNC, Ali M, Kam L, Puik JR, Rodrigues SMF, Zwart ES, Daams F, Zonderhuis BM, Meijer LL, Le Large TYS, Giovannetti E, van Laarhoven HWM and Kazemier G. The diagnostic value of the CA19-9 and bilirubin ratio in patients with pancreatic cancer, distal bile duct cancer and benign periampullary diseases, a novel approach. Cancers (Basel) 2022; 14: 344.
[48] Zhang M, Cheng S, Jin Y, Zhao Y and Wang Y. Roles of CA125 in diagnosis, prediction, and oncogenesis of ovarian cancer. Biochim Biophys Acta Rev Cancer 2021; 1875: 188503.
[49] Sun A. Clinical role of serum tumor markers SCC, NSE, CA 125, CA 19-9, and CYFRA 21-1 in patients with lung cancer. Lab Med 2023; 54: 638-645.
[50] Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS and Wistuba I; WHO Panel. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol 2015; 10: 1243-1260.
[51] Riely GJ, Wood DE, Ettinger DS, Aisner DL, Akerley W, Bauman JR, Bharat A, Bruno DS, Chang JY, Chirieac LR, DeCamp M, Desai AP, Dilling TJ, Dowell J, Durm GA, Gettinger S, Grotz TE, Gubens MA, Juloori A, Lackner RP, Lanuti M, Lin J, Loo BW, Lovly CM, Maldonado F, Massarelli E, Morgensztern D, Mullikin TC, Ng T, Owen D, Owen DH, Patel SP, Patil T, Polanco PM, Riess J, Shapiro TA, Singh AP, Stevenson J, Tam A, Tanvetyanon T, Yanagawa J, Yang SC, Yau E, Gregory KM and Hang L. Non-small cell lung cancer, version 4.2024, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2024; 22: 249-274.
[52] Saji H, Okada M, Tsuboi M, Nakajima R, Suzuki K, Aokage K, Aoki T, Okami J, Yoshino I, Ito H, Okumura N, Yamaguchi M, Ikeda N, Wakabayashi M, Nakamura K, Fukuda H, Nakamura S, Mitsudomi T, Watanabe SI and Asamura H; West Japan Oncology Group and Japan Clinical Oncology Group. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. Lancet 2022; 399: 1607-1617.
[53] Altorki N, Wang X, Kozono D, Watt C, Landrenau R, Wigle D, Port J, Jones DR, Conti M, Ashrafi AS, Liberman M, Yasufuku K, Yang S, Mitchell JD, Pass H, Keenan R, Bauer T, Miller D, Kohman LJ, Stinchcombe TE and Vokes E. Lobar or sublobar resection for peripheral stage IA non-small-cell lung cancer. N Engl J Med 2023; 388: 489-498.
[54] Choi W, Nadeem S, Alam SR, Deasy JO, Tannenbaum A and Lu W. Reproducible and interpretable spiculation quantification for lung cancer screening. Comput Methods Programs Biomed 2021; 200: 105839.
[55] Park DJ, Park MW, Lee H, Kim YJ, Kim Y and Park YH. Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci Rep 2021; 11: 7567.
[56] Kuwana M, Avouac J, Hoffmann-Vold AM, Smith V, Toenges G, Alves M and Distler O. Development of a multivariable prediction model for progression of systemic sclerosis-associated interstitial lung disease. RMD Open 2024; 10: e004240.
[57] Tong C, Miao Q, Zheng J and Wu J. A novel nomogram for predicting the decision to delayed extubation after thoracoscopic lung cancer surgery. Ann Med 2023; 55: 800-807.
[58] Liu C, Zhao W, Xie J, Lin H, Hu X, Li C, Shang Y, Wang Y, Jiang Y, Ding M, Peng M, Xu T, Hu A, Huang Y, Gao Y, Liu X, Liu J and Ma F. Development and validation of a radiomics-

based nomogram for predicting a major pathological response to neoadjuvant immunochemotherapy for patients with potentially resectable non-small cell lung cancer. Front Immunol 2023; 14: 1115291.

[59] Li J, Pan B, Huang Q, Zhan C, Lin T, Qiu Y, Zhang H, Xie X, Lin X, Liu M, Wang L and Zhou C. A nomogram for predicting cancer-specific survival in young patients with advanced lung cancer based on competing risk model. Clin Respir J 2024; 18: e13800.

[60] Paez R, Kammer MN, Balar A, Lakhani DA, Knight M, Rowe D, Xiao D, Heideman BE, Antic SL, Chen H, Chen SC, Peikert T, Sandler KL, Landman BA, Deppen SA, Grogan EL and Maldonado F. Longitudinal lung cancer prediction convolutional neural network model improves the classification of indeterminate pulmonary nodules. Sci Rep 2023; 13: 6157.

[61] Luo Y, Yuan H, Pei Q, Chen Y, Xian J, Du R and Ye T. Artificial neural network-based diagnostic models for lung cancer combining conventional indicators with tumor markers. Exp Biol Med (Maywood) 2023; 248: 829-838.