Discover Oncology

Research

Prediction model for the early recurrence of stage IA–IIA non-small cell lung cancer based on hematological indexes and imaging features

Wei Zhao[1,2] · Yiyuan Sun[1] · Xin Liu[3] · Mingxiang Zhang[3] · Bohui Zhu[1]

Received: 2 November 2024 / Accepted: 25 April 2025
Published online: 07 May 2025

Abstract

Background Some patients with non-small cell lung cancer (NSCLC) experience early relapse within 2 years post-surgery. Screening patients who are prone to recurrence is crucial. This study aimed to determine factors influencing early recurrence within 2 years of surgery for stage IA–IIA NSCLC and to establish a prediction model.

Methods We retrospectively analyzed the hematological indices and imaging indicators of patients with stage IA–IIA NSCLC who underwent surgery at our hospital, and relevant clinical data were obtained through long-term follow-up from September 2019 to September 2020. Least absolute shrinkage and selection operator (LASSO) regression and univariate and multivariate Cox regression analyses were used to identify high-risk factors influencing postoperative recurrence, establish a predictive model, and construct a nomogram associated with recurrence-free survival.

Results Among 186 patients (90 male and 96 female), 29 (15.6%) experienced recurrence or metastasis during the follow-up period. Univariate analysis identified several significant factors, including tumor size, direct bilirubin, indirect bilirubin, albumin, globulin, serum creatinine, platelet-lymphocyte ratio, lymphocyte-monocyte ratio, prognostic nutrition index, albumin-alkaline phosphatase ratio, marginal lobulation, air bronchogram sign, pathological type of squamous cell carcinoma, tumor stage IB, and solid nodules. LASSO regression was used to further select variables and construct a multivariate Cox model showing globulin levels, air bronchogram signs, and solid nodules as independent prognostic factors for early recurrence within 2 years in patients with stage IA–IIA NSCLC. The Cox model stratified patients into high- and low-risk groups and was validated by Kaplan–Meier survival analysis, which demonstrated that high-risk patients had a significantly lower survival rate than low-risk patients, demonstrating the robust discriminative power of the predictive model.

Conclusion Globulin content, air bronchogram signs, and solid nodules were independent prognostic factors for early recurrence within 2 years in patients with stage IA–IIA NSCLC. The proposed model, developed based on the above factors and the albumin-alkaline phosphatase ratio, can effectively predict recurrence risk, potentially aiding clinicians in quantifying prognostic risk, making personalized survival assessments, and devising the most effective treatment plans.

Keywords Non-small cell lung cancer · Recurrence · Predictive model · Hematological indexes · Imaging features

Wei Zhao and Yiyuan Sun contributed equally to this work as first authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02514-2.

✉ Bohui Zhu, zxqh724@163.com | [1]Department of Oncology, Gongli Hospital of Shanghai Pudong New Area, No. 219, Miaopu Road, Pudong New Area, Shanghai 200135, China. [2]Department of Medical Oncology, The First Affiliated Hospital of University of Science and Technology of China (USTC), Hefei 230001, China. [3]The First Affiliated Hospital of University of Science and Technology of China (USTC), Hefei 230001, China.

Abbreviations

AAPR	Albumin-alkaline phosphatase ratio
AUC	Area under the curve
CI	Confidence Interval
CT	Computed tomography
DCA	Decision curve analysis
GGN	Ground-glass nodule
GGO	Ground-glass opacity
HR	Hazard ratio
LASSO	Least absolute shrinkage and selection operator
LMR	Lymphocyte-monocyte ratio
NLR	Neutrophil-lymphocyte ratio
NSCLC	Non-small cell lung cancer
OR	Odds ratio
PLR	Platelet-lymphocyte ratio
PNI	Prognostic nutrition index
RFS	Recurrence-free survival
ROC	Receiver operating characteristic
SCLC	Small cell lung cancer
SD	Standard deviation

1 Introduction

Lung carcinoma is a common malignancy and the primary contributor to cancer-associated fatalities. Despite significant advancements in treatment, the average five-year survival rate for patients with lung cancer remains alarmingly low at only 15% [1]. Lung cancer is classified into two main histological types: small-cell lung cancer (SCLC), which accounts for approximately 15% of lung cancer cases, and non-small cell lung cancer (NSCLC), which constitutes approximately 85% of lung cancer cases. NSCLC can be further subdivided into three primary subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [2]. Treatment may differ depending on the tumor type and stage. Surgery remains the primary treatment choice for early stage NSCLC, including stages I–IIA [3]. Unlike SCLC, the spread and metastasis of NSCLC occur relatively late. Complete surgical resection offers a cure for patients with early stage NSCLC. However, approximately 30–75% of these patients experience postoperative recurrence [4]. Some patients relapse within 2 years, with a worse prognosis [5]. Therefore, early screening of high-risk groups prone to early recurrence and timely intervention are imperative for improving the prognosis of these patients.

Traditional methods of evaluating cancer prognosis mainly depend on the tumor node metastasis stage and smoking status of the patient. For example, compared with SCLC, NSCLC generally has a higher five-year survival rate and lower early recurrence rate; compared with patients with lung cancer who do not smoke, patients with lung cancer who smoke usually have a lower five-year survival rate [6]. However, the prognosis of patients with cancer at the same stage is not identical. In addition to these traditional features, several routine clinical indicators, including serological and blood cytological indicators, the lymphocyte-monocyte ratio (LMR), platelet-lymphocyte ratio (PLR), neutrophil–lymphocyte ratio (NLR), albumin-alkaline phosphatase ratio (AAPR), and prognostic nutrition index (PNI), have been demonstrated to correlate with the prognosis of lung carcinoma [7–12]. Moreover, advancements in computed tomography (CT) have led to the identification of features that predict the prognosis of patients with tumors. Several high-risk CT features associated with lung cancer recurrence have been proposed [13]. However, the prognostic value of combining hematological indices with imaging features in lung cancer remains uncertain.

Predictive models are widely recognized as reliable tools for evaluating prognosis and aiding in clinical decision-making for various malignant tumors, including uterine squamous cell, prostate, pulmonary, and gastric carcinomas [14–17]. Therefore, this study aimed to evaluate the prognostic significance of hematological parameters and CT imaging features in predicting early recurrence of stage IA–IIA lung cancer within 2 years. Additionally, we sought to establish a prognostic model based on blood cytological parameters and high-risk CT imaging features. This model could aid in the identification of patients at a high risk of early recurrence, allowing for timely intervention to improve patient outcomes.

2 Materials and methods

2.1 Source of patient data

Between September 2019 and September 2020, we collected the clinical and pathological data of patients with stage IA–IIA NSCLC who underwent surgical treatment at the Department of Thoracic Surgery, First Affiliated Hospital, University of Science and Technology, China. This study was reviewed and approved by the institutional Ethics Committee.

2.2 Inclusion and exclusion criteria

The inclusion criteria were as follows: (1) no history of preoperative radical radiotherapy or chemotherapy, (2) radical resection performed, and (3) patients with stage IA–IIA NSCLC pathologically diagnosed post-surgery.

The exclusion criteria were as follows: (1) patients with other systemic tumors, (2) patients with multiple lung cancers and metastatic lung cancers, (3) patients without preoperative chest CT, and (4) patients with incomplete follow-up and medical records.

Between September 2019 and September 2020, 699 patients with stage IA–IIA NSCLC were diagnosed at the Department of Thoracic Surgery of our hospital. A total of 191 patients were lost to follow-up; 281 had missing chest CT imaging data before surgery, and 41 had incomplete preoperative serological and blood cytological examinations. Therefore, these patients were excluded from this study. A total of 186 patients (90 men and 96 women) met the inclusion criteria and were included in the study.

2.3 Collection of clinical and follow-up data

Clinical data, including patient name, sex, tumor stage, and surgical method, were collected from the electronic medical record system of our hospital. Each patient was assessed based on the classification outlined in the eighth edition of the American Joint Committee on Cancer Staging Compendium. All pathological diagnoses were performed by pathologists with several years of experience, and the recurrence of the disease was unknown. Patients were followed-up by telephone, and prognostic information was collected. The follow-up period was 24 months.

2.4 Definition of outcome

The main outcome index of this study was the diagnosis of recurrence within 24 months of thoracic surgery. Chest CT was performed every 3 months for 24 months after surgical resection. If recurrence was detected, the radiologist assisted in the joint assessment to determine the patient's outcome. The interval devoid of recurrence, herein referred to as the recurrence-free period, was operationally defined as the duration from the day after the surgical intervention to the point at which either recurrence manifested or until the final follow-up date. Early recurrence, delineated by the reappearance of neoplastic growth within the initial 24 months of postoperative intervention, is a distinctive hallmark [18].

2.5 CT features

All CT images were consistently evaluated by two chest radiologists who were aware that all patients had been diagnosed with lung cancer; however, the pathological outcomes and recurrence were unclear. The imaging features used for analysis included tumor location, lobulated border, spiculated margin, air bronchogram, bronchovascular bundle thickening, pleural retraction, nodule type, nodule shape, cavity formation, central low density, peritumoral interstitial thickening, and pleural contact (defined as a tumor-pleural contact length >1/4 tumor perimeter). According to the proportion of ground-glass opacity (GGO), nodules were divided into pure ground-glass nodule (GGN), partial GGN, and solid nodules. GGO is characterized by a hazy area that does not obscure the underlying vascular or bronchial structures.

The final results were determined by averaging the measurements from the two reviewers. If the discrepancy between the two measurements exceeded 5%, a third measurement was conducted by consensus and used as the final result. (Air Bronchogram: A radiographic feature indicating air-filled bronchi within a consolidation area. Solid Nodule: A lung nodule without ground-glass opacity, indicative of higher malignancy risk. GGO: hazy lung region that retains bronchial and vascular markings).

2.6 Predictive variables

The following variables were included as predictive variables and further analyzed to complete the model: age, sex, surgery type (lobectomy, partial or pneumonectomy), pathological type (adenocarcinoma, squamous cell carcinoma, carcinoma in situ, or other), stage (IA1, IA2, IA3, IB, or IIA), tumor size, γ-glutamyl transpeptidase, glutamic oxaloacetic transaminase, alkaline phosphatase, direct bilirubin, globulin, glutamic pyruvic transaminase, indirect bilirubin, serum creatinine, albumin, urea nitrogen, uric acid, blood glucose, carbon dioxide, monocyte count, neutrophil count, lymphocyte count, hemoglobin levels, platelet count, NLR, PLR, LMR, PNI, AAPR, tumor location (upper, middle, or lower lobes), lobulated border (yes or no), spiculated margin (yes or no), air bronchogram (yes or no), bronchovascular bundle thickening (yes or no), pleural retraction (yes or no), nodule type (solid nodule, partially GGN, or pure GGN), nodule shape (round, oval, or other), cavity (yes or no), central low density (yes or no), peritumor interstitial thickening (yes or no), pleural contact (yes or no). These 41 variables were incorporated into subsequent examinations as prospective predictors.

2.7 Selection of variables

To ascertain potential risk factors associated with early recurrence in individuals diagnosed with NSCLC, each predictive variable was integrated into a least absolute shrinkage and selection operator (LASSO) regression analysis within the designated training dataset. Through a cross-validated LASSO regression analysis, the magnitudes of the regression coefficients were penalized to mitigate the risk of overfitting by variables in the training dataset. The coefficients of the factors with low correlation were reduced to zero under punishment to select predictive variables with the greatest impact for further analysis.

2.8 Model development and verification

The model was developed using Cox proportional hazards regression analysis. The validity of the estimator was confirmed by testing and verifying the Cox proportional hazards hypothesis. The complete dataset was partitioned into training and testing datasets at a 60:40 ratio through random sampling. Subsequently, a Cox regression model was developed using the training dataset. To evaluate the efficacy of the model, a predictive algorithm was employed to compute the risk assessment of individual patients within the training cohort. Subsequently, individuals in the testing subset were categorized into high- and low-risk cohorts according to the median risk score generated using the Cox proportional hazards model. Kaplan–Meier analysis was used to draw survival curves, and working characteristic (receiver operating characteristic [ROC]) curves of the participants at 6, 12, and 18 months were drawn to evaluate the accuracy of prognosis. To visualize and quantify the impact of the variables on recurrence-free survival (RFS) rates at 6-month, 12-month, and 18-month, a nomogram was constructed and verified.

2.9 Statistical analysis

Continuous variables are presented as mean ± standard deviation (SD) for normally distributed data, whereas non-normally distributed data are presented as median with interquartile range. Categorical variables are presented as frequencies and percentages within each group. Dummy variables were created for four non-ordered categorical variables: pathological type, tumor location, tumor nature, and tumor shape. LASSO regression analysis was employed to identify predictors correlated with RFS. Cox regression was used to develop a prediction model that was visualized using a nomogram. Finally, column-line plots were examined using calibration curves. Survival curves were generated using Kaplan–Meier analysis.

Statistical analyses were performed using R software version 4.2.2, which can be accessed at https://www.r-project.org/. The following R packages were used: tidyverse and flextable for data description and trilinear table plotting, respectively; survivor, caret, glmnet, survminer, limma, and time ROC for LASSO regression screening of independent variables and Cox model construction and validation; and regplot and RMS for column line plotting and evaluation, respectively. Statistical significance was set at $P < 0.05$.

3 Results

3.1 Baseline characteristics

At the end of the observation period, 186 patients were monitored. The summarized data of these patients are as follows: mean age, 59.2 years (SD, 10.9 years); 90 men (48.4%) and 96 women (51.6%). In total, 29 patients experienced early recurrence within 2 years, with a recurrence rate of 15.6%. The total average recurrence time was 22.5 ± 4.7 months, and the average tumor size was 1.71 ± 0.91 cm. In terms of tumor staging, stages IA1, IA2, IA3, IB, and IIA were detected in 55 (29.6%), 86 (46.2%), 27 (14.5%), 15 (8.1%), and 3 (1.6%) cases, respectively (Table 1, Supplementary Table 1). The numbers

Table 1 CT features of 186 patients (N = 186)

CT features	Number
Tumor position	
UP	116 (62.4%)
Middle	15 (8.1%)
Lower	55 (29.6%)
Lobulated border	
No	112 (60.2%)
Yes	74 (39.8%)
Spiculated margin	
No	72 (38.7%)
Yes	114 (61.3%)
Air bronchogram	
No	141 (75.8%)
Yes	45 (24.2%)
Bronchovascular bundle thickening	
No	46 (24.7%)
Yes	140 (75.3%)
Pleural retraction	
No	109 (58.6%)
Yes	77 (41.4%)
Nodule type	
Solid	60 (32.3%)
Part-GGN	95 (51.1%)
Pure-GGN	31 (16.7%)
Nodule shape	
Circle	93 (50.0%)
Oval	13 (7.0%)
Other	80 (43.0%)
Cavity	
No	145 (78.0%)
Yes	41 (22.0%)
Center low density	
No	118 (63.4%)
Yes	68 (36.6%)
Peritumoral interstitial thicken	
No	114 (61.3%)
Yes	72 (38.7%)
Pleural contact	
No	163 (87.6%)
Yes	23 (12.4%)

of cases in the training and testing datasets were 113 and 73, respectively. The average ages of the training and testing datasets were 59.0 ± 11.6 and 59.5 ± 9.78 years, respectively. No substantial variance in age distribution was observed between the two cohorts ($P = 0.745$). The training cohort included 53 male (47.0%) and 60 female (53.0%) patients, whereas the testing cohort included 37 male (50.7%) and 36 female (49.3%) patients. Notably, there was no marked variance in the sex distribution between the two cohorts ($P = 0.698$). Among the 113 patients in the training set, 17 relapsed within 2 years, with a recurrence rate of 15%. Among the 73 patients in the testing set, 12 had early recurrence, with a recurrence rate of 16.4%. There was no significant difference in the recurrence rate between the two cohorts ($P = 0.961$). The average maximum tumor diameter of the patients in the training group was 1.70 ± 0.934 cm, and that of the patients in the testing group was 1.71 ± 0.877 cm. The mean tumor size was not significantly different between the two cohorts ($P = 0.948$), suggesting an equitable distribution of data among the groups (Table 2). CT imaging analysis demonstrated significant associations between lobulated borders and both albumin levels ($P = 0.042$) and the lymphocyte-monocyte ratio ($P = 0.017$), and between nodule type and the lymphocyte-monocyte ratio ($P = 0.018$) (Supplementary Table 2). Other CT features showed no significant correlation with hematological indices.

3.2 Univariate Cox regression

Univariate regression analysis was conducted to analyze the factors affecting early postoperative recurrence of stage IA–IIA NSCLC. The results showed that tumor size (hazard ratio [HR]: 2.072, 95% confidence interval [CI]: 1.376–3.120; $P < 0.001$), direct bilirubin (HR: 1.060, 95% CI: 1.033–1.088; $P < 0.001$), indirect bilirubin (HR: 1.098, 95% CI: 1.027–1.174; $P = 0.006$), albumin (HR: 0.935, 95% CI: 0.908–0.962; $P < 0.001$), globulin (HR: 0.909, 95% CI: 0.871–0.949; $P < 0.001$), serum creatinine (HR: 1.029, 95% CI: 1.013–1.045; $P < 0.001$), PLR (HR: 1.007, 95% CI: 1.000–1.013; $P = 0.037$), LMR (HR: 0.757, 95% CI: 0.576–0.996; $P = 0.046$), PNI (HR: 0.938, 95% CI: 0.913–0.965; $P < 0.001$), AAPR (HR: 0.009, 95% CI: 0.001–0.088; $P < 0.001$), lobulated border (HR: 5.619, 95% CI: 1.831–17.247; $P = 0.003$), presence of air bronchogram (HR: 2.620, 95% CI: 1.011–6.793; $P = 0.048$), pathological type (squamous cell carcinoma) (HR: 5.496, 95% CI: 1.927–15.671; $P = 0.001$), tumor stage (IB) (HR: 4.310, 95% CI: 1.515–12.265; $P = 0.006$), and nodule type (solid nodule) (HR: 4.357, 95% CI: 1.534–12.375) had a statistically significant effect on early recurrence within 2 years. The forest plot displays variables with univariate Cox regression *P*-values of <0.05 (Fig. 1, Supplementary Table 3).

3.3 Selection of variables and establishment of multi-factor Cox regression model

To identify independent risk factors for RFS, we included 15 variables with a significance level of <0.05 from the univariate Cox regression analysis in a multivariate Cox regression analysis. To mitigate the risk of overfitting and to streamline the model, we employed LASSO regression analysis, which penalizes the absolute values of the coefficients and effectively eliminates variable covariance. Combining the results of the LASSO and multicollinearity analyses (Supplementary Fig. 1a, b), nine variables exhibiting non-zero coefficient values and corresponding λ values and likelihood of bias were identified. These included tumor size, direct bilirubin, globulin, blood creatinine, PNI, AAPR, lobulated borders, air bronchograms, and nodule type (solid). After stepwise multi-factor Cox analysis of the nine factors, four factors comprising globulin content (HR: 0.921, 95% CI: 0.863–0.983; $P = 0.013$), AAPR (HR: 0.102, 95% CI: 0.006–1.721; $P = 0.113$), presence of air bronchograms (HR: 4.022, 95% CI: 1.382–11.704; $P = 0.011$), and nodule type (solid) (HR: 3.596, 95% CI: 1.223–10.575; $P = 0.020$) were selected to construct the risk model (Supplementary Table 4). Among these, globulin content, presence of an air bronchogram, and solid nodules were delineated as stand-alone prognostic factors for early recurrence in individuals diagnosed with stage IA–IIA NSCLC.

3.4 Model evaluation

To maximize the feasibility and robustness of the prediction model, patients were categorized into high- and low-risk cohorts according to the median prognostic score derived from the multifactorial Cox model established using the training dataset. Both the full and validation cohorts were classified using identical risk score thresholds, and the corresponding Kaplan–Meier survival curves were illustrated separately. The Kaplan–Meier survival curves indicated that in all three datasets, the training, testing, and total datasets, RFS exhibited a marked discrepancy between the high-risk and

Table 2 Baseline data in the training and testing datasets

Variables	Train (N = 113)	Test (N = 73)	*P*-value
Age (27–86 years)			
Mean (SD)	59.0 (11.6)	59.5 (9.78)	0.745
Sex (Male, Female)			
Male	53 (46.9%)	37 (50.7%)	0.698
Female	60 (52.2%)	36 (47.9%)	
Recurrence (No, Yes)			
No	96 (85.0%)	61 (83.6%)	0.961
Yes	17 (15.0%)	12 (16.4%)	
Recurrence time (1–24 months)			
Mean (SD)	22.6 (4.37)	22.2 (5.18)	0.599
Stage (IA1, IA2, IA3, IB, IIA)			
IA1	35 (31.0%)	20 (27.4%)	0.721
IA2	51 (45.1%)	35 (47.9%)	0.822
IA3	14 (12.4%)	13 (17.8%)	0.417
IB	12 (10.6%)	3 (4.1%)	0.188
IIA	1 (0.9%)	2 (2.7%)	0.701
Tumor size (0.4–5 cm)			
Mean (SD)	1.70 (0.934)	1.71 (0.877)	0.948
Adenocarcinoma (No, Yes)			
No	36 (31.9%)	14 (19.2%)	0.0827
Yes	77 (68.1%)	59 (80.8%)	
Squamous cell carcinoma (No, Yes)			
No	103 (91.2%)	68 (93.2%)	0.831
Yes	10 (8.8%)	5 (6.8%)	
Carcinoma in situ (No, Yes)			
No	90 (79.6%)	66 (90.4%)	0.081
Yes	23 (20.4%)	7 (9.6%)	
Pathological type (Other) (No, Yes)			
No	110 (97.3%)	71 (97.3%)	1
Yes	3 (2.7%)	2 (2.7%)	
Globulin (2.2–39.2 g/L)			
Mean (SD)	27.4 (7.37)	25.1 (8.84)	0.0692
Neutrophil–lymphocyte ratio (0.67–21.78)			
Mean (SD)	3.42 (3.94)	2.72 (2.10)	0.116
Platelet-lymphocyte ratio (25.19–373.91)			
Mean (SD)	132 (54.0)	134 (57.6)	0.831
Lymphocyte-monocyte ratio (0.52–8.80)			
Mean (SD)	4.01 (1.72)	3.86 (1.68)	0.556
Albumin-alkaline phosphatase ratio (0.02–1.10)			
Mean (SD)	0.566 (0.194)	0.526 (0.191)	0.17

low-risk cohorts, with statistical significance at $P < 0.001$, $P = 0.008$, and $P < 0.001$, respectively. These findings underscore the potential utility of the model in discerning populations at a higher risk of premature recurrence, as depicted in Fig. 2.

ROC curves play a pivotal role in assessing the precision of predictive models. The area under the curve (AUC) values acquired for the RFS prognostic model within the training cohort were 0.780, 0.845, and 0.856 at 6, 12, and 18 months, respectively. AUC values obtained for the RFS prognostic model in the testing set were 0.918, 0.883, and 0.882 for 6, 12, and 18 months, respectively. AUC values obtained for the RFS prognostic model in the total datasets were 0.848, 0.855, and 0.860 at 6, 12, and 18 months, respectively. ROC analysis demonstrated that the prognostic model exhibited high predictive accuracy across the training, testing, and total datasets (Fig. 3).

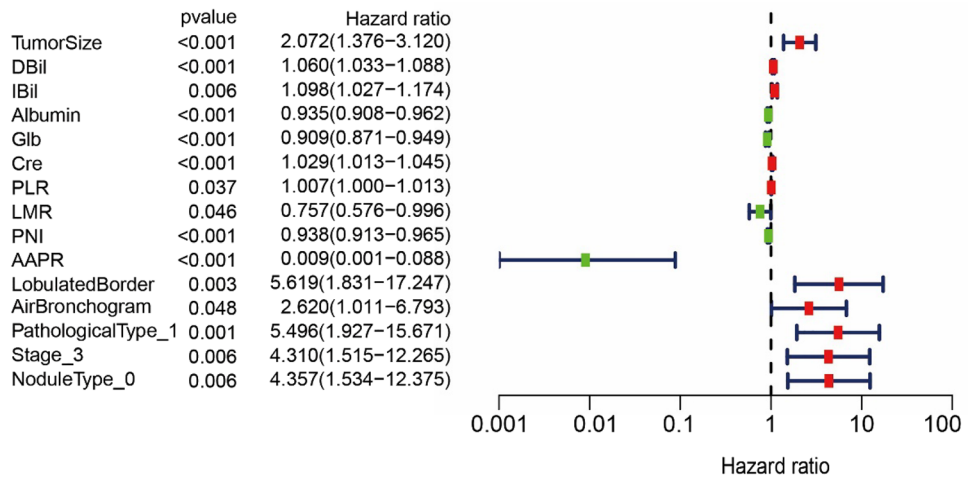

Fig. 1 Variables associated with early recurrence predicted via univariate Cox regression analysis. DBil: Direct bilirubin, Ibil: Indirect bilirubin, Glb: Globulin, Cre: creatinine, Pathological type 1: Squamous cell carcinoma, Stage 3: IB stage, Nodule_Type_0: Solid nodule

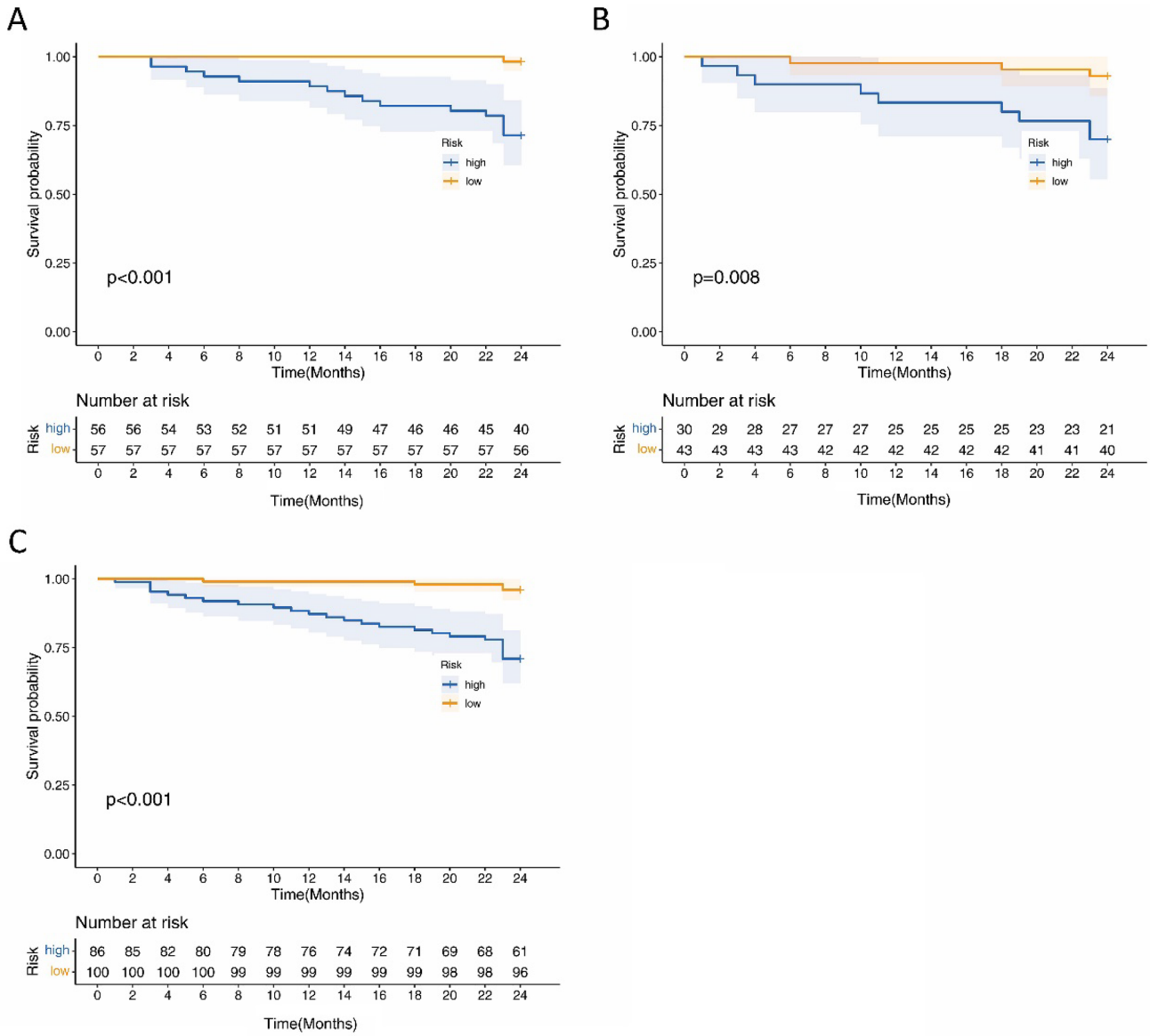

Fig. 2 Effect of the Cox model risk score on early postoperative recurrence. Kaplan–Meier analysis stratified patients into high- and low-risk groups based on calculated risk scores. **A** training dataset, **B** testing dataset, **C** total dataset

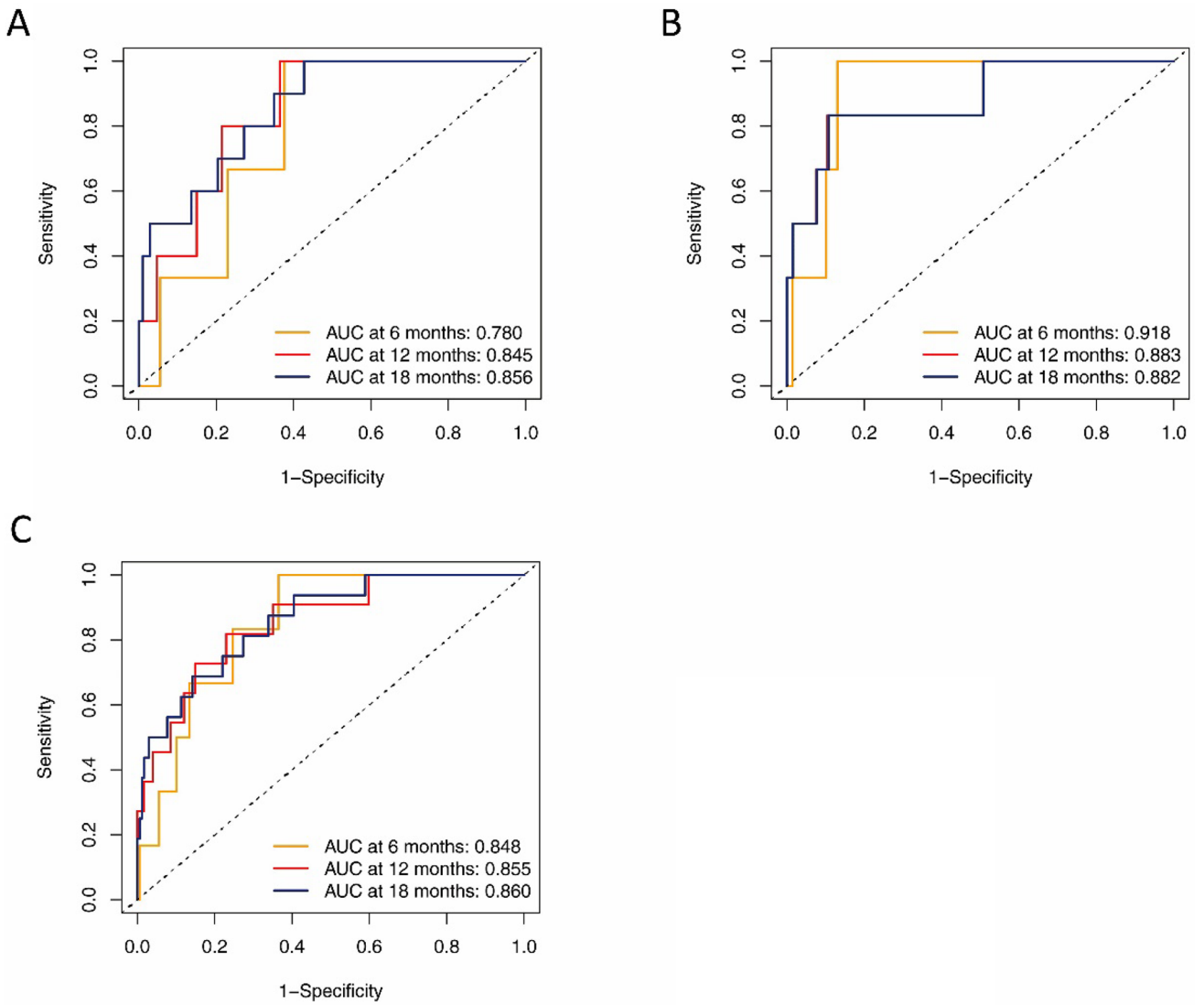

Fig. 3 ROC curve of RFS at 6, 12, and 18 months predicted using the multivariate Cox regression model. **A** training dataset, **B** testing dataset, **C** total dataset

3.5 Construction and verification of nomogram

We constructed a nomogram for RFS at 6, 12, and 18 months using a multifactorial Cox regression model (Fig. 4). Each factor was assigned a weighted number of points, and the total number of risk points per patient was used to predict RFS at 6, 12, and 18 months. To further assess the accuracy of the plotted column-line graphs, we assessed the nomogram using calibration curves and decision curve analysis (DCA). Calibration curves showed that the prediction lines were close to the diagonal, indicating the high accuracy of the column-line graph predictions (Supplementary Fig. 2). The net benefit of this model is represented on the y-axis in the DCA plots (Supplementary Fig. 3).

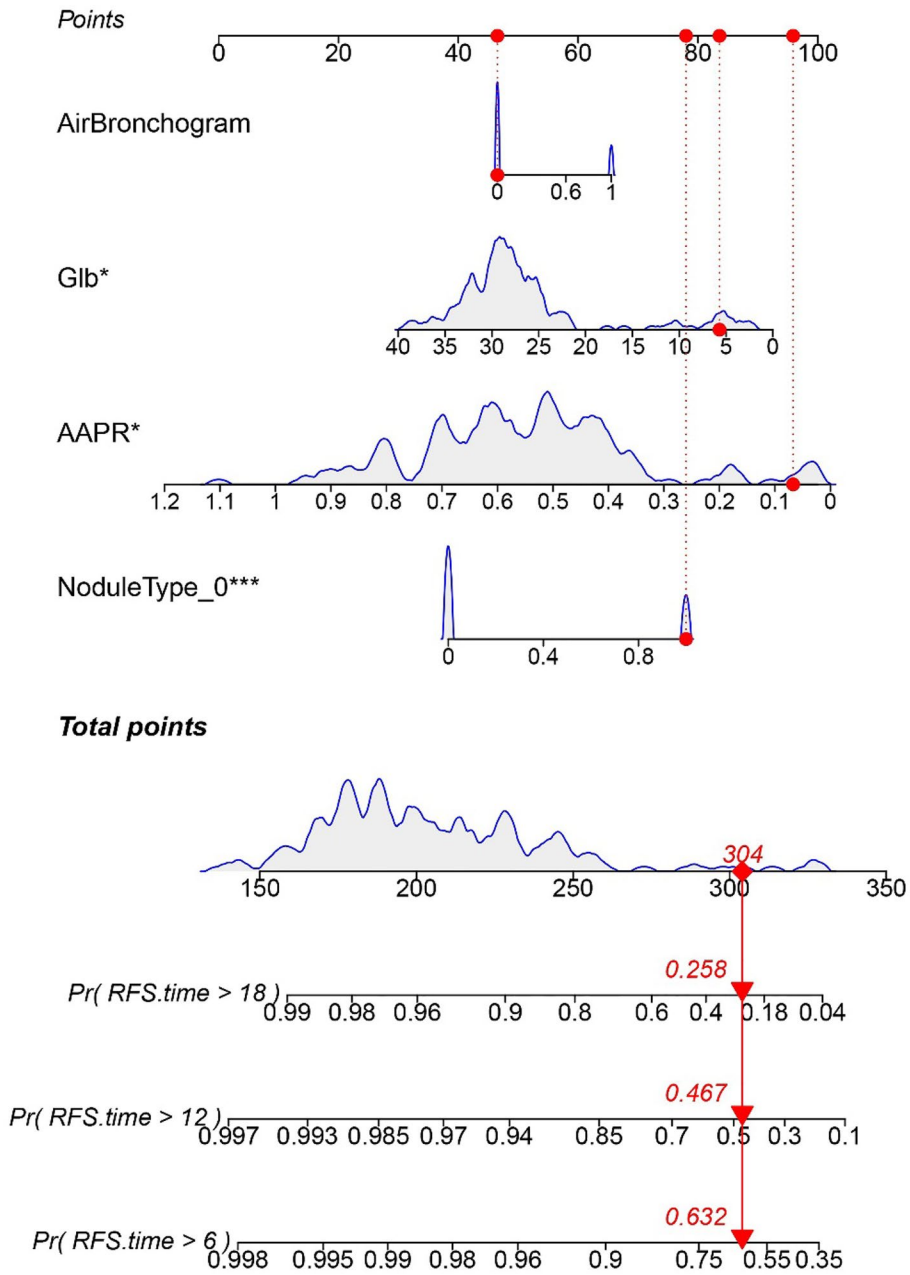

Fig. 4 The 6-month, 12-month, and 18-month RFS of patients with stage IA–IIA NSCLC was predicted using a nomogram. By drawing a vertical line upward from the scale corresponding to the four variables, the points for each variable were obtained, thereby obtaining the total number of points. Subsequently, the probability of RFS can be obtained by drawing a downward vertical line from the total points to the RFS axis of 6, 12, and 18 months

4 Discussion

Lung cancer is the predominant malignant neoplasm in global epidemiology, reigning supreme in terms of both prevalence and fatality rates [1]. Although CT is increasingly used, an increasing number of patients with early lung cancer have been identified and treated prior to the onset of clinical symptoms. Concurrently, owing to the use of medications, radiotherapy, chemotherapy, targeted therapy, and immunological examinations, the survival rate of patients with NSCLC has improved significantly. However, >20% of patients with early NSCLC relapse post-surgery and die [4]. Therefore, reducing the likelihood of postoperative recurrence in early stage NSCLC is crucial for the early prediction of patients at a high risk of recurrence and the selection of appropriate postoperative adjuvant therapy.

This study demonstrated that globulin level was an independent protective factor against early recurrence in patients with stage IA–IIA NSCLC within 2 years (HR: 0.921, 95% CI: 0.863–0.983; $P = 0.013$). Serum globulin is a mixture of proteins, including a large number of immunoglobulins, complements, and various glycoproteins with defense

functions. Immunoglobulins can be classified into antibodies and membrane immunoglobulins [19]. Antibodies are generally found in the serum and other body or exocrine fluids and can bind to specific antibodies. Membrane immunoglobulins, which exist on the human B cell membrane, are antigen receptors of B cells that recognize specific antibody molecules. The immune system plays a crucial role in the initiation and progression of lung cancer [20]. Immunoglobulins, which are vital components of the immune defense system of the body, mainly function to kill or dissolve target cells during monitoring and response to cancer cells [21]. When detecting serum immunoglobulin levels in patients with lung cancer, previous studies found that patients with NSCLC exhibited significantly higher levels of anti-CD25IgG ($Z = -7.48$, $P < 0.001$), indicating that the anti-CD25IgG antibody may serve as a valuable biomarker for the prognosis of lung cancer [22]. In this study, decreased globulin levels were associated with a higher risk of early recurrence within 2 years, possibly because these patients had a poor immune status and were more vulnerable to tumor cell escape. Globulins, especially immunoglobulins, play a critical role in immune surveillance by recognizing and eliminating tumor cells [23, 24]. Lower globulin levels suggest impaired immune responses, allowing tumor cells to evade detection and escape [25, 26]. This can lead to reduced tumor control and increased metastasis. Furthermore, decreased globulins may reflect a weakened inflammatory response, which supports tumor progression and recurrence [27]. Unlike the levels of tumor markers such as CEA or CYFRA 21-1, which reflect tumor burden, globulin levels provide insights into immune function and systemic inflammation, both of which are crucial for preventing recurrence [28, 29]. Low globulin levels indicate compromised immune surveillance because immunoglobulins play a key role in tumor recognition and immune response [23]. Reduced globulin levels are associated with impaired immune function and poor prognosis in various cancers [27]. Compared to general inflammatory markers such as CRP, which reflect broad inflammatory processes, globulin levels are more specifically associated with immune activity [30, 31]. However, while globulin levels should not replace imaging or conventional tumor markers in detecting recurrence, they can serve as complementary tools to assess recurrence risk. In summary, early detection and intervention in patients with low globulin levels are crucial for enhancing their probability of survival.

An air bronchogram indicates that a part of the air cavity in the original alveoli is replaced by inflammatory exudates, which combine the solid lung tissue with the aerated bronchus, often producing a transparent bronchial shadow [32]. Air bronchograms can be observed in various lung diseases, including pneumonia, pulmonary tuberculosis, lung cancer, and chronic obstructive pulmonary disease. Because an air bronchogram often occurs in pulmonary inflammatory diseases, previous studies have classified it as a benign lesion [33]. However, in 1996, a retrospective study of 132 patients with solitary pulmonary nodules showed that in 17 cases of benign lesions, only one case (5.9%) exhibited an air bronchogram, whereas among 115 cases of lung cancer, 33 (28.7%) showed air bronchograms ($P < 0.05$), indicating that air bronchograms are significantly associated with lung cancer [34]. A recent meta-study showed that patients with the disappearance of bronchial inflation sign are more likely to have local recurrence after radiotherapy than patients with benign lesions after conventional radiotherapy (odds ratio (OR): 7.15, 95% CI: approximately 2.08–24.59, $P < 0.01$), indicating that this sign is associated with the prognosis and development of tumors [35]. In this study, we selected the preoperative CT signs of the patients. These findings suggest that the presence of an air bronchogram is associated with the recurrence of early stage NSCLC. Moreover, it was identified as an autonomous risk factor for predicting early recurrence within 2 years in patients with stage IA–IIA NSCLC (HR: 4.022, 95% CI: 1.382–11.704; $P = 0.011$). Therefore, caution should be exercised when treating patients with these symptoms.

GGO refers to a gray-white blurred shadow on CT of the human chest, similar to ground glass, with a hazy cloud shape, and the vascular structure and bronchial texture can be seen. This sign often indicates an uneven composition within the nodule, such as fluid, cells, or pulmonary interstitial thickening of lung tissue. Short-term GGNs often indicate focal hyperemia, edema, or inflammation, whereas long-term partial GGNs can be precancerous lesions, lung cancer in situ, or early lung cancer [36]. The type of pulmonary nodules can be classified according to the GGO ratio, which can be divided into solid nodules, partial GGN, and pure GGN. Previous studies have reported that solid nodules have the lowest malignant probability, with a malignant proportion of only 7%, pure GGN has a malignancy rate of approximately 18%, and partial GGN has the highest malignant probability of approximately 64% [37]. However, studies on the GGO ratio and prognosis of lung cancer have reported different results. A recent study has shown that a tumor GGO of <17% is an independent risk factor for early recurrence within 5 years in patients with pathologically graded NSCLC (OR: 6.54, 95% CI: 1.47–29.15; $P = 0.014$) [38]. Numerous previous studies have demonstrated prognostic differences between GGO lung cancer and solid lung cancers: lung cancers of the malignant GGO type are usually inert tumors with slow growth and long cycles, with a reported mean doubling time of approximately 800 days, and malignant solid lung cancer nodules have relatively short growth cycles compared to GGNs, with changes usually occurring within 3–6 months and in a few patients, even metastases [39, 40]. This may be owing to the fact that normal lung tissue is largely absent from solid lung

cancer nodules compared to GGO lung cancer and that their composition is more complex than that of ground-glass lung cancer [41]. This study revealed that solid nodules independently increased the risk of early recurrence in patients with stage IA–IIA NSCLC (HR: 3.596, 95% CI: 1.223–10.575; $P = 0.020$). These results indicate that a decrease in the GGO ratio of lung cancer nodules may be a poor prognostic factor for NSCLC.

Correlations between specific CT imaging features and systemic biomarkers, such as associations of lobulated borders with albumin levels and the lymphocyte-monocyte ratio, and of the nodule type with the lymphocyte-monocyte ratio, highlight the potential of imaging biomarkers for tumor characterization. These findings are consistent with those of recent studies that emphasized the prognostic value of CT-derived radiomic features in cancer assessment [42]. The link between lobulated borders and low albumin levels may reflect aggressive tumor growth and nutritional depletion, whereas an association with a high lymphocyte-monocyte ratio suggests tumor-related inflammation and immune infiltration [43, 44]. However, the biological underpinnings of these radiomics features remain unclear. The lack of significant associations with other CT variables indicates a complex interplay between imaging features and systemic biomarkers, which is likely influenced by tumor heterogeneity and technical limitations [45, 46]. Future research combining advanced imaging modalities with gene expression, histopathology, and biomarker analysis is needed to elucidate these mechanisms and improve the predictive accuracy of non-invasive tumor assessments.

Identifying independent prognostic factors and developing accurate predictive models are essential for optimal treatment planning, counseling, and follow-up of patients with early stage NSCLC. Based on data from our hospital, hematological parameters and imaging features of patients with early stage NSCLC were analyzed, and globulin content, presence of air bronchograms, and solid nodule type were found to be independent prognostic factors for early recurrence in patients with stage IA–IIA NSCLC. A novel nomogram was developed to predict the RFS of patients with NSCLC at 6, 12, and 18 months, and the prediction model showed satisfactory performance. After internal verification, the model showed good discriminant ability. Overall, the results of this study demonstrate the effectiveness of the model, both internally and temporally. Using this model, line maps were employed to quantify and visualize model factors, enabling separate predictions of the 6-, 12-, and 18-month relapse-free survival of patients with early stage NSCLC.

However, the results of this study should be interpreted with caution. First, data collection was retrospective, and there may have been deviations that were difficult to eliminate. Second, although the results of this study have been internally verified, our research has the limitations of being a single-center study and the lack of independent external data for verification. Third, despite the inclusion of consecutive patients, the mean patient age was 59.5 years. The cohort included relatively few patients with lung cancer who were aged <60 years. Therefore, whether the prognostic model was applicable to patients in all age groups was difficult to determine. Fourth, CT imaging quality can vary between institutions owing to differences in equipment, protocols, and radiologist interpretation. This can lead to discrepancies in detecting subtle lesions or changes in disease status. We suggest that future studies incorporate multimodal imaging techniques or automated image analysis to enhance consistency. Finally, although a combination of hematological indexes and imaging features was modeled, the relationship between the blood index and imaging features is not clear and requires further study.

5 Conclusion

This study showed that globulin content, presence of air bronchograms, and solid nodule type were independent prognostic factors for early recurrence within 2 years in patients with stage IA–IIA NSCLC. A novel Cox regression model was constructed to predict the 6-month, 12-month, and 18-month RFS in patients with NSCLC, and an RFS diagram was drawn using the model. The model showed good discriminant ability through internal verification and can assist clinicians in quantifying the risk of recurrence in patients with NSCLC, allowing for efficient and convenient optimization of treatment plans and personalization of patient follow-ups. Large-scale, forward-looking, multicenter studies are necessary to conduct additional external validation and application of the prediction model.

Acknowledgements We would like to thank all individuals who helped us with our article.

Author contributions Wei Zhao and Bohui Zhu conducted the studies, participated in data collection, and drafted the manuscript. Xin Liu and Mingxiang Zhang were responsible for organizing and analyzing the imaging data. Bohui Zhu and Yiyuan Sun performed the statistical analysis and participated in the design. All the authors have read and approved the final version of the manuscript.

Funding No external funding was received for conducting this study.

Data availability All data generated or analyzed in this study are included in this article. Further inquiries can be directed to the corresponding authors.

Declarations

Ethics approval This study was conducted in accordance with the Declaration of Helsinki (revised in 2013). This study was approved by the Ethics Committee of the First Affiliated Hospital of University of Science and Technology of China (USTC). Written informed consent was obtained from all the participants.

Competing interests The authors declare no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Rice SR, Saboury B, Houshmand S, Salavati A, Kalbasi A, Goodman CR, Werner TJ, Vujaskovic Z, Simone CB, Alavi A. Quantification of global lung inflammation using volumetric 18F-FDG PET/CT parameters in locally advanced non-small-cell lung cancer patients treated with concurrent chemoradiotherapy: a comparison of photon and proton radiation therapy. Nucl Med Commun. 2019;40(6):618–25.
3. Markers D. Retracted: a retrospective study of effectiveness of thoracoscopic lobectomy and segmentectomy in patients with early-stage non-small-cell lung cancer. Dis Markers. 2023;2023:9815364.
4. Gardner LD, Loffredo PhDC, Langenberg P, George DMS, Deepak J, Harris CC, Amr S. Associations between history of chronic lung disease and non-small cell lung carcinoma in Maryland: variations by sex and race. Ann Epidemiol. 2018;28(8):543–8.
5. Matsumura Y, Yano M, Yoshida J, Koike T, Kameyama K, Shimamoto A, Nishio W, Yoshimoto K, Utsumi T, Shiina T, Watanabe A, Yamato Y, Watanabe T, Takahashi Y, Sonobe M, Kuroda H, Oda M, Inoue M, Tanahashi M, Adachi H, Saito M, Hayashi M, Otsuka H, Mizobuchi T, Moriya Y, Takahashi M, Nishikawa S, Suzuki H. Early and late recurrence after intentional limited resection for cT1aN0M0, non-small cell lung cancer: from a multi-institutional, retrospective analysis in Japan. Interact Cardiovasc Thorac Surg. 2016;23(3):444–9.
6. Morgensztern D, Ng SH, Gao F, Govindan R. Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey. J Thoracic Oncol. 2010;5(1):29–33.
7. Zhang W, Wang W, Wu J, Tian J, Yan W, Yuan Y, Yao Y, Shang A, Quan W. Immune cell-lipoprotein imbalance as a marker for early diagnosis of non-small cell lung cancer metastasis. Front Oncol. 2022;12:942964.
8. Portale G, Bartolotta P, Azzolina D, Gregori D, Fiscon V. Prognostic role of platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte, and lymphocyte-to-monocyte ratio in operated rectal cancer patients: systematic review and meta-analysis. Langenbecks Arch Surg. 2023;408(1):85.
9. Zhang CL, Jiang XC, Li Y, Pan X, Gao MQ, Chen Y, Pang B. Independent predictive value of blood inflammatory composite markers in ovarian cancer: recent clinical evidence and perspective focusing on NLR and PLR. J Ovarian Res. 2023;16(1):36.
10. Ren W, Wang H, Xiang T, Liu G. Prognostic role of preoperative onodera's prognostic nutritional index (OPNI) in gastrointestinal stromal tumors: a systematic review and meta-analysis. J Gastrointest Cancer. 2023;54(3):731–8.
11. Tsai MH, Chuang HC, Lin YT, Yang KL, Lu H, Huang TL, Tsai WL, Su YY, Fang FM. The prognostic value of preoperative albumin-to-alkaline phosphatase ratio on survival outcome for patients with locally advanced oral squamous cell carcinoma. Technol Cancer Res Treat. 2022;21:15330338221141254.
12. Liu Y, Hu C, Wu J, Zhou J, Wang W, Wang X, Guo J, Wang Q, Zhang X, Xie J, Xing Y, Ding X, Hu D. Relationship between lymphocyte to monocyte ratio and brain metastasis in non-smalll cell lung cancer patients. Am J Transl Res. 2022;14(6):3936–45.
13. Zhang R, Wei Y, Shi F, Ren J, Zhou Q, Li W, Chen B. The diagnostic and prognostic value of radiomics and deep learning technologies for patients with solid pulmonary nodules in chest CT images. BMC Cancer. 2022;22(1):1118.
14. Yin J, Yang Y, Ma K, Yang X, Lu T, Wang S, Shi Y, Zhan C, Zhu Y, Wang Q. Clinicopathological characteristics and prognosis of pulmonary pleomorphic carcinoma: a population-based retrospective study using SEER data. J Thorac Dis. 2018;10(7):4262–73.
15. Zhou H, Li X, Zhang Y, Jia Y, Hu T, Yang R, Huang KC, Chen ZL, Wang SS, Tang FX, Zhou J, Chen YL, Wu L, Han XB, Lin ZQ, Lu XM, Xing H, Qu PP, Cai HB, Song XJ, Tian XY, Zhang QH, Shen J, Liu D, Wang ZH, Xu HB, Wang CY, Xi L, Deng DR, Wang H, Lv WG, Shen K, Wang SX, Xie X, Cheng XD, Ma D, Li S. Establishing a nomogram for stage IA-IIB cervical cancer patients after complete resection. Asian Pacific J Cancer Prevent. 2015;16(9):3773–7.
16. He J, Albertsen PC, Moore D, Rotter D, Demissie K, Lu-Yao G. Validation of a contemporary five-tiered gleason grade grouping using population-based data. Eur Urol. 2017;71(5):760–3.
17. Han DS, Suh YS, Kong SH, Lee HJ, Choi Y, Aikou S, Sano T, Park BJ, Kim WH, Yang HK. Nomogram predicting long-term survival after d2 gastrectomy for gastric cancer. J Clin Oncol. 2012;30(31):3834–40.

18. Kiankhooy A, Taylor MD, Lapar DJ, Isbell JM, Lau CL, Kozower BD, Jones DR. Predictors of early recurrence for node-negative t1 to t2b non-small cell lung cancer. Ann Thorac Surg. 2014;98(4):1175–83.
19. Wang R, Shi G, Zhang H, Wang T, Ren W, Jiao Q. Globulin and albumin/globulin ratios as potential biomarkers for the diagnosis of acute and chronic peri-prosthetic joint infections: a retrospective study. Surg Infect. 2023;24(1):58–65.
20. Gunawardana NC, Durham SR. New approaches to allergen immunotherapy. Ann Allergy Asthma Immunol. 2018;121(3):293–305.
21. Pan B, Wang Z, Zhang X, Shen S, Ke X, Qiu J, Yao Y, Wu X, Wang X, Tang N. Targeted inhibition of RBPJ transcription complex alleviates the exhaustion of CD8(+) T cells in hepatocellular carcinoma. Commun Biol. 2023;6(1):123.
22. Ye L, Li X, Sun S, Guan S, Wang M, Guan X, Lee KH, Wei J, Liu B. A study of circulating anti-CD25 antibodies in non-small cell lung cancer. Clin Transl Oncol. 2013;15(8):633–7.
23. Schwartz-Albiez R. Naturally occurring antibodies directed against carbohydrate tumor antigens. Adv Exp Med Biol. 2012;750:27–43.
24. Alexandrescu DT, Wiernik PH. Serum globulins as marker of immune restoration after treatment with high-dose rituximab for chronic lymphocytic leukemia. Med Oncol. 2008;25(3):309–14.
25. Sun L, Kees T, Almeida AS, Liu B, He XY, Ng D, Han X, Spector DL, Mcneish IA, Gimotty P, Adams S, Egeblad M. Activating a collaborative innate-adaptive immune response to control metastasis. Cancer Cell. 2021;39(10):1361-1374.e9.
26. Peng L, Zhang Y, Wang Z. immune responses against disseminated tumor cells. Cancers. 2021;13(11):2515.
27. Niwa N, Matsumoto K, Ide H. The clinical implication of gamma globulin levels in patients with nonmuscle-invasive bladder cancer. Urol Oncol. 2019;37(4):291.e1-291.e7.
28. Kato J, Sumikawa Y, Hida T, Kamiya T, Horimoto K, Kamiya S, Sato S, Takahashi H, Sawada M, Yamashita T. Serum cytokeratin 19 fragment 21–1 is a useful tumor marker for the assessment of extramammary Pagt's disease. J Dermatol. 2017;44(6):666–70.
29. Gopal V, Nisha Y, Ganesan P, Kayal S, Bobby Z, Adithan S, Penumadu P, Ramakrishnalay VP, Bandlamudi BP, Bahttacharjee A, Dahagama S, Dubashi B. Role of CYFRA 21–1 and CEA as prognostic and predictive markers in locally advanced and metastatic gastric carcinoma. J Cancer Res Ther. 2024;20(5):1412–9.
30. Ramírez Velázquez C, Martínez Juárez G, Lozano Nuevo J, Olvera Medel A, Higuera Acuña L, García Quijano E, Guido Bayardo R, Castrejón VI. High-affinity C-reactive protein as inflammatory marker. Rev Alerg Mex. 2007;54(1):7–13.
31. Lubas W, Gutkowski K. C-reactive protein in cardiovascular diseases. Przegl Lek. 2006;63(7):562–6.
32. Volmonen K, Sederholm A, Rönty M, Paajanen J, Knuuttila A, Jartti A. Association of CT findings with invasive subtypes and the new grading system of lung adenocarcinoma. Clin Radiol. 2023;78(3):e251–9.
33. Karimian M, Azami M. Chest computed tomography scan findings of coronavirus disease 2019 (COVID-19) patients: a comprehensive systematic review and meta-analysis. Pol J Radiol. 2021;86:e31–49.
34. Kui M, Templeton PA, White CS, Cai ZL, Bai YX, Cai YQ. Evaluation of the air bronchogram sign on CT in solitary pulmonary lesions. J Comput Assist Tomogr. 1996;20(6):983–6.
35. Kim TH, Woo S, Halpenny DF, Kim YJ, Yoon SH, Suh CH. Can high-risk CT features suggest local recurrence after stereotactic body radiation therapy for lung cancer? A systematic review and meta-analysis. Eur J Radiol. 2020;127:108978.
36. Yin W, Wang W, Zou C, Li M, Chen H, Meng F, Dong G, Wang J, Yu Q, Sun M, Xu L, Lv Y, Wang X, Yin R. Predicting tumor mutation burden and EGFR mutation using clinical and radiomic features in patients with malignant pulmonary nodules. J Personal Med. 2022;13(1):16.
37. Sihoe ADL. Should sublobar resection be offered for screening-detected lung nodules? Transl Lung Cancer Res. 2021;10(5):2418–26.
38. Ko KH, Hsu HH, Huang TW, Gao HW, Cheng CY, Hsu YC, Chang WC, Chu CM, Chen JH, Lee SC. Predictive value of 18F-FDG PET and CT morphologic features for recurrence in pathological stage IA non-small cell lung cancer. Medicine. 2015;94(3):e434.
39. Kim EA, Johkoh T, Lee KS, Han J, Fujimoto K, Sadohara J, Yang PS, Kozuka T, Honda O, Kim S. Quantification of ground-glass opacity on high-resolution CT of small peripheral adenocarcinoma of the lung: pathologic and prognostic implications. AJR Am J Roentgenol. 2001;177(6):1417–22.
40. Uehara H, Tsutani Y, Okumura S, Nakayama H, Adachi S, Yoshimura M, Miyata Y, Okada M. Prognostic role of positron emission tomography and high-resolution computed tomography in clinical stage IA lung adenocarcinoma. Ann Thorac Surg. 2013;96(6):1958–65.
41. Suzuki K, Koike T, Asakawa T, Kusumoto M, Asamura H, Nagai K, Tada H, Mitsudomi T, Tsuboi M, Shibata T, Fukuda H, Kato H. 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 Thoracic Oncol. 2011;6(4):751–6.
42. Borhani AA, Zhang P, Diergaarde B, Darwiche S, Chuperlovska K, Wang SC, Schoen RE, Su GL. Role of tumor-specific and whole-body imaging biomarkers for prediction of recurrence in patients with stage III colorectal cancer. Abdom Radiol. 2024;50:1907–15.
43. Shi J, Han X, Wang J, Han G, Zhao M, Duan X, Mi L, Li N, Yin X, Shi H, Li C, Gao J, Xu J, Yin F. Matrine prevents the early development of hepatocellular carcinoma like lesions in rat liver. Exp Ther Med. 2019;18(4):2583–91.
44. Averyanova EV, Shkolnikova MN, Chugunova OV, Mazko ON. Effects of triterpenoids in fatty products on liver condition of laboratory animals with acute toxic hepatitis. Vopr Pitan. 2023;92(4):81–91.
45. Gu Y, Huang H, Tong Q, Cao M, Ming W, Zhang R, Zhu W, Wang Y, Sun X. Multi-view radiomics feature fusion reveals distinct immuno-oncological : characteristics and clinical prognoses in hepatocellular carcinoma. Cancers. 2023;15(8):2338.
46. Jiang L, You C, Xiao Y, Wang H, Su GH, Xia BQ, Zheng RC, Zhang DD, Jiang YZ, Gu YJ, Shao ZM. Radiogenomic analysis reveals tumor heterogeneity of triple-negative breast cancer. Cell Rep Med. 2022;3(7):100694.