



## OPEN ACCESS

## EDITED BY

Natsuo Tomita,  
Nagoya City University, Japan

## REVIEWED BY

Shreya Shukla,  
Tata Memorial Hospital, India  
Zhenbin Qiu,  
Guangdong Provincial People's Hospital Lung  
Cancer Institute, China

## \*CORRESPONDENCE

Wen Gao

✉ gaowen@jsph.org.cn

Mei Yuan

✉ yuameijiangsu@163.com

<sup>†</sup>These authors have contributed  
equally to this work

RECEIVED 01 March 2025

ACCEPTED 19 May 2025

PUBLISHED 18 June 2025

## CITATION

Wang F, Li C, Li S, Zhang T, Yu T, Zhang W,  
He J, Yuan M and Gao W (2025) CT-based  
radiomics integrated model for brain  
metastases in stage III/IV ALK-positive lung  
adenocarcinoma patients.  
*Front. Oncol.* 15:1585930.  
doi: 10.3389/fonc.2025.1585930

## COPYRIGHT

© 2025 Wang, Li, Li, Zhang, Yu, Zhang, He,  
Yuan and Gao. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# CT-based radiomics integrated model for brain metastases in stage III/IV ALK-positive lung adenocarcinoma patients

Fen Wang<sup>1†</sup>, Caiyun Li<sup>2†</sup>, Shuke Li<sup>3†</sup>, Teng Zhang<sup>4</sup>, Tongfu Yu<sup>4</sup>,  
Wei Zhang<sup>4</sup>, Jing He<sup>3</sup>, Mei Yuan<sup>4\*</sup> and Wen Gao<sup>3\*</sup>

<sup>1</sup>Department of Medical Imaging, The Affiliated Hua'an No.1 People's Hospital of Nanjing Medical University, Huaian, Jiangsu, China, <sup>2</sup>Department of Radiology, First Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China, <sup>3</sup>Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, <sup>4</sup>Department of Radiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

**Purpose:** This study aims to develop and validate a computed tomography (CT)-based radiomics nomogram for predicting brain metastases in lung adenocarcinoma with anaplastic lymphoma kinase positive (ALK+).

**Methods:** Of 117 patients were retrospectively reviewed, among them, 34 patients from another hospital. Patients were randomly allocated into training (70%) and validation (30%) cohorts. We integrated the radiomics score (Rad\_score) with independent clinic-radiological variables to build the nomogram model. The DeLong test and Decision curve analysis (DCA) were utilized to evaluate performance of three models. Cox regression analysis was used to identify statistically significant factors for progression-free survival (PFS) in ALK-positive lung adenocarcinoma, with model discrimination evaluated by the concordance index (C-index). The patients were divided into low-risk and high-risk groups. Finally, the Log-rank test was used to ascertain significant differences between the two risk groups in the nomogram models.

**Results:** From Stage III/IV lung cancer cases, we extracted 1834 radiomics features, identifying two features can serve as standalone indicators of BM. The AUC of radiomics model was 0.905 and 0.880 in the validation and external test cohort, respectively. The AUC of nomogram model was 0.940 in the validation cohort and 0.896 in the external test cohort, respectively. The statistical difference merely exists between nomogram and clinical model ( $P=0.009$ ,  $P=0.012$ ) in validation and external test cohorts, respectively. The multivariate Cox regression analysis showed that lymphadenopathy (Hazard ratio (HR) = 5.41, 95% confidence interval (CI): 1.38–21.16,  $P = 0.015$ ) and rad\_score (HR = 25.67, 95% CI: 5.41–121.94,  $P < 0.001$ ) were independent predictive factors for PFS. The Concordance index (C-Index) for training cohort (C-Index(95%CI): 0.887 (0.826–0.956); testing cohort: 0.798 (0.676–0.938), and the external cohort with 0.927 (0.857–0.996). Patients in the low-risk group showed a significantly better PFS compared to those in the high-risk group in the training cohort and validation cohort ( $P$  all  $< 0.010$ , respectively), whereas the results were not consistent in the external test cohort ( $P=0.130$ ).

**Conclusion:** CT-derived radiomic signatures show promise as a tool for predicting BM within 2 years after detection of primary lung adenocarcinoma detection with ALK+. Combining these radiomic signatures with clinical features can enhance risk stratification for these patients.

#### KEYWORDS

lung adenocarcinoma, anaplastic lymphoma kinase, radiomics, brain metastasis, computed tomography

## Introduction

When lung adenocarcinoma progresses, it has been observed that approximately 30–43% of patients develop brain metastases (BM) (1–3). The prognosis for patients with brain-metastatic non-small cell lung cancer remains poor, with median overall survival (OS) typically around 17 months for those with lung adenocarcinoma (4). Among the various oncogenic drivers in patients with non-small cell lung cancer (NSCLC), rearrangements in the anaplastic lymphoma kinase (ALK) gene are considered to be potent drivers, presented in approximately 5% of cases, second only to epidermal growth factor receptor (EGFR) mutations (5, 6). Moreover, the research have indicated that patients with ALK rearrangements are particularly susceptible to developing brain metastases, with an incidence as high as 66% in ALK-positive (ALK+) patients, whereas the overall incidence in all lung adenocarcinoma cases is lower (7). The advent of the ALK inhibitor, crizotinib, significantly improved the treatment of ALK+ advanced lung adenocarcinoma, demonstrating better outcomes than chemotherapy (8). Brain metastases are generally considered to be the final stage of advanced disease (staging III/IV) and deemed as an ominous sign of disease progression and death. In a systematic review including 21 studies, the median OS for ALK-positive NSCLC patients with baseline brain metastases was 23 months (9). A real-world study separately analyzed data from patients with and without baseline brain metastases, showing that the median OS for ALK-positive NSCLC patients with baseline brain metastases was 27.1 months after first-line ALK tyrosine kinase inhibitor (TKI) treatment, while it was 36.9 months for patients without brain metastases (10). However, the effectiveness of crizotinib in controlling brain metastases is grim, as it struggles to penetrate the blood-brain barrier. In response, second- and third-generation ALK inhibitors have emerged (11), showing varying levels of control over intracranial metastases. The median PFS could be 24.0 months. Consequently, the brain remains the most common site of progression in patients with or without baseline BM. BM status may significantly influence prognosis and therapeutic response, necessitating the development of accurate prediction models. Developing reliable predictive and prognostic indicators for brain metastases (BM) is imperative, not only pretreatment but also during longitudinal surveillance. Therefore, developing a prognostic marker for predicting the development of TKI resistance in ALK+ patients would be very significant. This would allow for earlier identification

of patients who may require alternative treatment options and could potentially improve patient outcomes.

Ongoing research efforts are focused on utilizing radiomics signatures to extract high-dimensional data from clinical images. This data is then analyzed using data mining techniques to explore the potential biological behavior of tumors and make preoperative diagnoses for assessing therapeutic efficacy (12–14). The driver gene status of lung cancer, such as the echinoderm microtubule-associated protein-like 4 (EML4)-ALK fusion, highlights the heterogeneity of tumors at the molecular level. Radiomics has shown promise in predicting lymph node metastases across various types of tumors (15–17). The nomogram was depicted with the combination of clinic-radiological and radiomics features can aid in prediction. However, few studies have investigated the application of this integrated nomogram model in predicting brain metastases specifically in stage III/IV lung adenocarcinoma. Therefore, the aim of the aforementioned study was to develop and validate predictive radiomics models and nomogram models that can serve as reliable auxiliary tools for predicting BM in patients with ALK-positive lung adenocarcinoma, thereby providing valuable insights to improved patient management and treatment decision-making.

## Materials and methods

### Study cohort

This retrospective study at a dual medical institution strictly adhered to the principles outlined in the Declaration of Helsinki. This study was reviewed and approved by the ethics committee of the first affiliated hospital of Nanjing medical university (Permit Number: 2023-SRFA-337). The written informed consent was waived because of the retrospective and anonymous nature of the data analysis.

The study encompassed patients with lung adenocarcinoma confirmed to harbor ALK-positive mutations through pathological analysis from June 2016 to August 2023. The inclusion and exclusion of participants were elucidated in ([Supplementary Data 1.1](#)).

In our hospital study, before exclusion, a total of 117 participants were diagnosed with lung adenocarcinoma with ALK-positive, and finally, 83 patients were ultimately confirmed

as stage III/IV NSCLC through medical imaging, who were categorized according to the 7th Edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual [2010] being included in the study (18). These participants were then categorized into two groups based on the presence or absence of brain metastasis (BM). The presence of BM (BM+) group consisted of 26 patients who were diagnosed with brain metastasis within a minimum follow-up period of 2 years. The absence of BM (BM-) group, on the other hand, consisted of 57 patients who did not develop brain metastasis during the follow-up period of at least 2 years. Patients were randomly assigned to the training set ( $n=58$ ), with 37 participants from the BM+ group and 21 from the BM- group. The remaining 25 participants were assigned to the validation cohort, with 5 participants from the BM+ group and 20 participants from the BM- group.

In addition to the internal study cohort, we also included an external test cohort from another hospital. This cohort consisted of 34 cases, with 16 cases of ALK+ lung adenocarcinoma with brain metastasis and 18 cases of ALK+ lung adenocarcinoma without brain metastasis. The same methods used in the internal study were applied to review and analyze the patients in the external test cohort. The entire workflow is visualized in Figure 1.

## Clinical observation

In this study, we used various imaging techniques, routinely used in clinical practice for diagnosing BM (16), specially contrast-enhanced T1-weighted imaging (T1-CE), T2 fluid-attenuated inversion recovery (T2-FLAIR), T2-weighted imaging (T2WI), and diffusion-weighted imaging (DWI). All magnetic resonance (MR) scans were performed using a 3.0 T MR scanner (Siemens Medical Solutions) with an eight-channel head and neck coil following standard brain imaging procedures. To ensure the accuracy of the diagnosis, all MR scans were reviewed and analyzed by experienced radiologists who were board-certified with a minimum of 5 years of experience in interpreting brain imaging (Authors #2, #4). The dual reporting system involving two radiologists helped minimize the risk of diagnostic error and ensured reliable confirmation of brain metastasis.

Beyond MR scans, the study also included routine laboratory tests and other imaging modalities, such as chest CT enhancement scan with a full abdominal CT enhancement scan, to monitor disease progression and evaluate metastasis in other areas of the body. The follow-up interval for these assessments was typically 4–6 weeks.

To capture all relevant clinical information, we also collected data from our hospital's Electronic Medical Records System (EMRS) and a collaborating hospital, encompassing age ( $<56$  year,  $\geq 56$  year), Gender (male, female), smoking history (no, yes), clinical stage (III/IV), and distant metastasis status. Data collection protocols were harmonized between institutions to ensure consistency.

## Image acquisition and evaluation

Pre-treatment thoracic CT images were acquired with a Siemens Somatom Sensation CT scanner (Siemens Healthineers, Erlangen, Germany). The scan encompassed the area from the chest inlet to the lower level of the costophrenic angle, bilaterally including the axillary regions. The detailed CT parameters are exhibited in (Supplementary Data 1.2). Two board-certified radiologists (author #1 with 5 years' experience and author #2 with 3 years' experience in thoracic imaging), independently interpreted the CT images blinded to patients' pathological diagnoses. A senior radiologist (author #5 with 30 years of experience) adjudicated any assessments discrepancies between the initial readers. The morphological assessment included comprehensive evaluation of the tumor characteristics and surrounding features, such as maximal axial diameter, location, pleural indentation, pleural effusion, lymphadenopathy, and carcinomatous lymphangitis (CL). The final maximal axial diameter measurements represent the mean values from both primary readers. For quantitative analysis, regions of interest (ROIs) within the nodule or mass were manually outlined along the tumor margins.

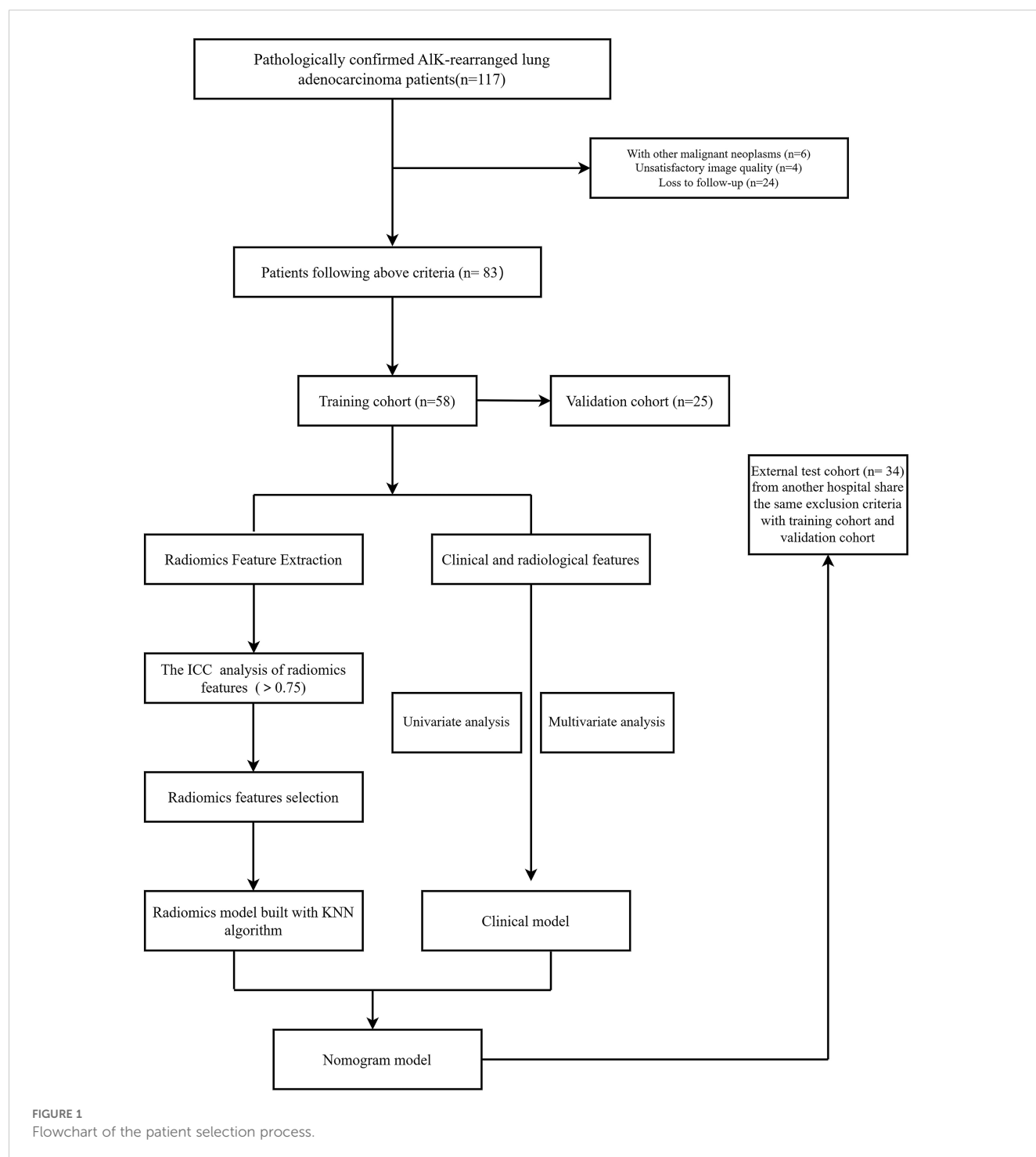
Authors #1, #4, and #2 subsequently evaluated a separate cohort of stage III/IV ALK+ lung adenocarcinoma patients from another hospital (Figure 2).

## Interobserver and interobserver reproducibility evaluation

To assess the inter-observer reproducibility of the 1843 radiomics features, two authors #2 and #1 independently performed the tumor segmentation procedure for all cases approximately 3 months later using the same annotation tool to minimize recall bias. The inter-observer and intra-observer agreement of the 1834 extracted features were evaluated using intraclass correlation coefficients (ICCs) (19). The intraclass correlation coefficient (ICC) refers to the reliability coefficient obtained by using the same scale to repeatedly measure the same group of respondents at certain time intervals. In practice, the level of reliability is often evaluated by calculating the intraclass correlation coefficient. The formula is as follows:  $ICC = \frac{MSa - MSe}{MSa + (n - 1)MS}$ , Where  $MSa$  represents the mean square between groups (study subjects),  $MSe$  represents the mean square within groups (error), and  $n$  represents the number of repeated measurements. An ICC value greater than 0.75 suggests good reproducibility of the feature extraction process (19, 20) for subsequent analysis.

## Image segmentation, radiomics feature extraction and analysis

The volumes of interest (VOIs) encompassing the complete tumor information. Semi-automatically contoured on thin-section CT images (1.0mm/1.5mm slice thickness) using an in-house software called MultiLabel (Shanghai Key Laboratory of Magnetic



Resonance, China). In cases where the border of the lesion was unclear, author #1 manually adjusted the precise edge of the VOIs. To ensure accurate segmentation, author #2 reviewed all the VOIs. Subsequently, the entire cohort of CT images with the VOIs segmentation information was converted to the NII format for further radiomics analysis. Radiomic features were extracted from each VOI using the pyradiomics package (<http://www.radiomics.io/pyradiomics.html>). We extracted the 105

original features from each VOI, including first order, Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM). Different image filters were also implemented subsequently, including exponential(n=91), Gradient(n=91), local binary patterns(lbp).3D(n=273), log.sigma (n=273), logarithm(n=91), square(n=91), squareroot (n=91),

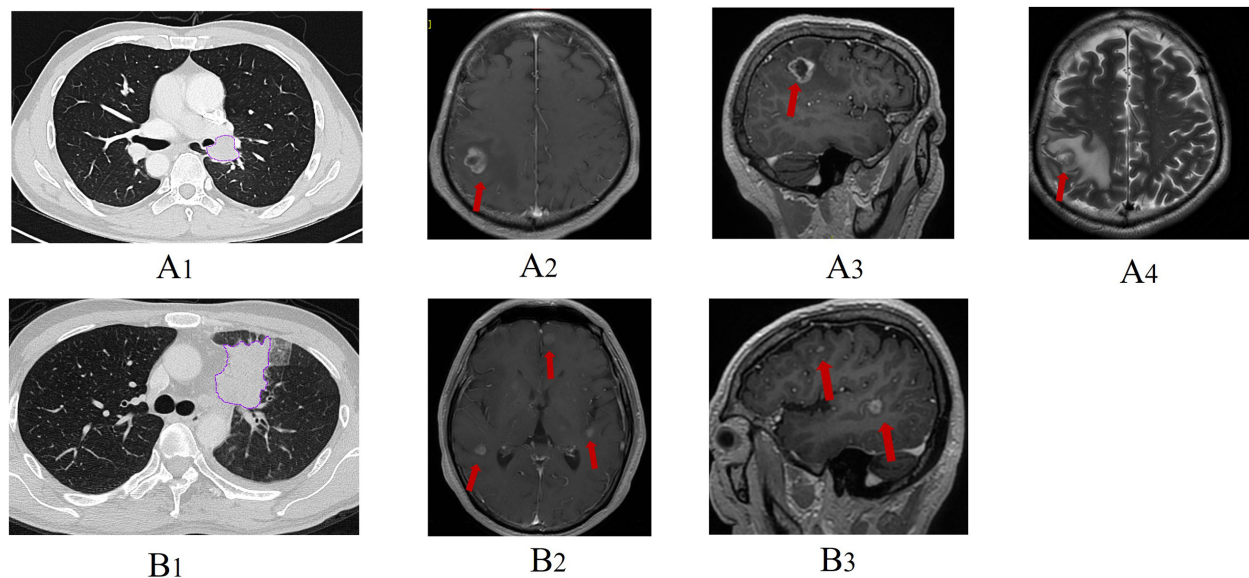


FIGURE 2

Representative clinical cases and related thoracic CT/brain MRI images. One patient had a pre-treatment thoracic CT image (A1) and brain MRI image (A2–A4) which indicated brain metastasis at baseline examination. Another patient (B1) was free from brain metastasis at baseline evaluation but developed brain metastasis after chemotherapy (B2, B3, arrow).

wavelet( $n=728$ ). The extracted radiomic features were normalized to a standard unit and zero-centered using the following equation:  $\vec{x}_n^{\text{normalized}} = \frac{\vec{x}_n - \bar{x}_n}{\sqrt{x_{1n}^2 + x_{2n}^2 + \dots + x_{mn}^2}}$  E1, where  $\vec{x}_n$  is the value of feature N, and  $\bar{x}_n$  is the average value of all features.

In the dimensionality reduction and feature selection process in the training cohort, the specific steps are as follows: 1) The Synthetic Minority Oversampling Technique (SMOTE) (21) was employed to address the issue of imbalanced data distribution for the minority class by interpolating existing minority class samples. 2) Pearson correlation coefficients (PCC) (22) were calculated for the extracted feature. The features with interclass correlation scores greater than 0.9 exhibited high correlation with the target variable and were preserved, while removing one feature from a pair of features with high correlation to avoid redundancy. 3) The least absolute shrinkage and selection operator (LASSO) (23) was applied to further select features by imposing a penalty on the number of nonzero coefficients to select a subset of features that are most informative for classification. 4) Several machine learning algorithms were compared during the data training process to identify an excellent classifier for the prediction model in patients with brain metastases (BM+), including logistic regression(LR) (24), support vector machine (SVM) (25), NaiveBayes(NB) (26), multilayer perceptron (MLP) (27), extreme gradient boosting (XGBoost) (28), and k-nearest neighbor algorithm (KNN) (29), Random forest(RF) (30), Extra trees(ET) (31), Light Gradient Boosting Decision Machine(LGB) (32), a gradient-boosting machine (GBM) (33), Adaptive Boosting(ADA) (34). In this step, 5-fold cross-validation was employed to train the classifier to evaluate the model's stability. The performance of different algorithms was evaluated, and the best-performing classifier was selected for the prediction model.

## Clinical utilization

The nomogram model (35) was developed based on the selected clinical parameters and Radiomics-score (Rad\_score) (36). The calibration curve and decision curve analysis (DCA) (37) were then visualized to evaluate the clinical utility of the nomogram model.

Additionally, Cox regression analyses were performed. Variables demonstrating statistical significance ( $p < 0.05$ ) in the univariate analysis were incorporated into the multivariate analysis. Concordance index (C-index) was used to assess the prognostic capability of the nomogram model in three different cohorts.

## PFS analysis

After regular and complete imaging follow-up, the progression-free survival (PFS) of patients were achieved, which is defined as the time to from the date to the first occurrence of disease progression, death or the last visit. According to the cutoff of the rad\_score, participants were categorized into a high-risk group and a low-risk group with the X-tile method (38). Based on the predetermined cutoff value of the rad\_score, patients were stratified into high-risk (Rad\_score above the cutoff) and low-risk (Rad\_score below the cutoff) groups. Kaplan-Meier survival analysis was conducted in all three cohorts to assess PFS differences between the groups.

## Statistical analysis

Statistical analysis was carried out using several software, such as SPSS 25.0, R software (version 4.1.0; <https://www.r-project.org>),



R software (version 3.3.4; <https://www.r-project.org>, the “carnet” and “ggplot” packages) and Python software (version 3.7.0; <http://www.python.org>; scikitplot, sklearn, matplotlib.pyplot, lightgbm, xgboost, sklearn.neighbors, sklearn.svm, numpy, and shap packages). The characteristics being compared were analyzed using different statistical tests depending on the type of variables. The student t-test was used for continuous parameters, while Chi-square or Fisher’s exact tests, and Mann-Whitney U test were used for categorical variables. The ultimate parameters were determined by univariate and multivariate logistic regression (39, 40). And the corresponding results were described as an odds ratio (OR) with a 95% confidence interval (CI) (41). The performance of the radiomics features to predict the presence of BM in stage III-IV lung cancer was assessed using the receiver operating characteristic curve (ROC) (42) and the area under the curve (AUC). ROC curve is a graphical tool used to evaluate the performance of a binary classification model. It illustrates the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) by plotting these rates at various threshold settings. The corresponding various metrics, including accuracy, sensitivity, specificity and F1 score were used to evaluate the performance of the models (28). A *P*-value less than 0.05 was considered statistically significant to determine the presence of a significant difference. Survival differences were analyzed using the Kaplan–Meier method, and the log-rank test was applied to evaluate statistical significance.

## Results

### Baseline characteristics of patients

In our hospital, we finally included 83 patients, of which 26 patients had BM (brain metastasis). These patients were randomly divided into the training cohort (58 patients; 21BM+, 37 BM-) and the validation cohort (25 patients, 5BM+,20BM-). Additionally, we also enrolled 34 patients for the external test cohort. Among these patients, 16 had BM, while 18 did not. The clinic-radiological features were listed in Table 1.

### Establishment of the clinical model

The comparative analysis of clinico-radiological features between BM + and BM- group is summarized in Table 1. While baseline demographic variables (age, gender, smoking history) and tumor characteristics (diameter, pleural effusion, pleural indentation, location) showed no significant differences between the BM- and BM+ groups (*P*>0.05). However, significant differences were observed in the variables of lymphadenopathy (*P*=0.003) and carcinomatous lymphangitis (*P*=0.010). To further identify predictors of BM, a multivariable logistic regression was performed, and a clinical model was established (Table 2). The final clinical model identified two independent predictors of brain metastasis: lymphadenopathy (odds ratio (OR): 5.133, 95% confidence interval (CI) 1.218–21.630, *P*=0.026) and

carcinomatous lymphangitis (OR:9.545, 95% CI 1.021–89.223, *P*=0.048). The clinical model can effectively predict the presence of brain metastasis based on the identified clinico-radiological predictors, with an area under the curve (AUC) of 0.746, specificity of 33.3%, sensitivity of 97.3%, and accuracy of 74.1% in the training cohort, with an AUC of 0.680, specificity of 40.0%, sensitivity of 75.0%, and accuracy of 68.0% in the validation cohort. In the external test cohort, the AUC of the clinical model was 0.661 (95% CI 0.487–0.836), with an accuracy of 58.8%.

### Feature selection and radiomics signature construction

The intra-observer ICCs and inter-observer ICCs calculated based on extracted 1834 features ranged from 0.768 to 1.000 and 0.751–0.999, respectively. These ICC values indicate a high level of agreement and consistency among the observers in evaluating the features.

To reduce the dimensionality of the feature space and mitigate the risk of bias and potential overfitting, we employed several methods, including SMOTE and PCC for initial feature selection, followed by LASSO and 5-fold cross-validation (Supplementary Figure 1). Firstly, SMOTE and PCC help identify the features with non-zero coefficients. Then, LASSO, combined with 5-fold cross-validation further refined the selection by zeroing out non-informative features, ensuring the most predictive ones remained while avoiding overfitting. The top features ranked by LASSO for each CT image contributed to compile radiomics signatures, termed Rad\_score. Similarly, a radiomics signature was built using a K-Nearest Neighbors (KNN) classifier from pre-treatment thoracic CT images. The KNN algorithm selected the most predictive features, again using a penalty determined by 5-fold cross-validation. Notably, the KNN algorithm utilized a nearest neighbor value of 7, from pre-treatment thoracic CT images. These features, along with their corresponding regression coefficients, were combined to calculate the Rad\_score and build the radiomics signature model.

Remarkably, only 2 out of the initial 1834 radiomics features were selected for the final model.

The selected features were chosen for the final model, which weighted by their coefficients are visualized in Supplementary Figure 2. The performance of each classifier for both training and validation cohorts is detailed in (Supplementary Table S1).

### Radiomics signature model performance

In the training cohort, the predictive power of the model was analyzed with an AUC of 0.938 (95% CI (0.882–0.993)), specificity of 94.6%, a sensitivity of 66.7%, and accuracy of 84.5%. The model also exhibited reasonable performance in the validation cohort, with an AUC of 0.905(95% CI (0.801–1.000)), specificity of 85.0%, sensitivity of 80.0%, and accuracy of 84.0%. Note that the diagnostic performance in the external test cohort was lower than that of the validation cohort, exhibiting an AUC of 0.880(95% CI (0.767–0.993)), the same goes for accuracy (79.4%) (Table 3).

TABLE 1 Patient characteristics and clinico-radiological features in training and validation sets.

Characteristics	Training cohort (n=58)		<i>P</i> - value	Validation cohort (n=25)		<i>P</i> - value
	BM+(n=21)	BM-(n=37)		BM+(n=5)	BM-(n=20)	
age (≥56yr)	9 (43%)	20 (54%)	0.412	2 (40%)	10 (40%)	0.689
gender (male/female)	9/12 (43%/57%)	20/17 (54%/50%)	0.412	2 (40%)	7 (28%)	0.835
smoking history (yes)	3 (14%)	11 (30%)	0.187	3 (60%)	11 (44%)	0.840
diameter (> 3cm)	12 (57%)	15 (41%)	0.223	1 (20%)	2 (24%)	0.504
pleural effusion (+)	9 (43%)	8 (22%)	0.088	1 (20%)	6 (24%)	1.000
pleural indentation (+)	13 (62%)	19 (51%)	0.437	4 (80%)	11 (44%)	0.615
carcinomatous lymphangitis (+)	7 (33%)	1 (3%)	0.002	2 (40%)	5 (20%)	0.597
lymphadenopathy (+)	18 (86%)	16 (43%)	0.002	2 (40%)	12 (48%)	0.623
location						
RUL	6 (28%)	4 (11%)		1 (20%)	4 (16%)	
RML	2 (10%)	3 (8%)		0 (0)	2 (8%)	
RLL	2 (10%)	9 (24%)	0.349	2 (40%)	4 (16%)	0.034
LUL	8 (38%)	13 (35%)		2 (40%)	6 (24%)	
LLL	3 (14%)	8 (22%)		0 (0)	4 (16%)	

Data are numbers of patients and parentheses indicate the percentage; BM, brain metastasis; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; yr, year; +, positive.

TABLE 2 Univariate and multivariate analyses of clinical model.

Characteristics	Univariate analysis	<i>P</i> <sup>1</sup> -value	Multivariate analysis	<i>P</i> -value
	OR (95%CI)		OR (95%CI)	
age (years)	0.638 (0.217-1.876)	0.414		
gender	1.569 (0.533-4.616)	0.414		
diameter	1.958 (0.661-5.789)	0.226		
location	0.764 (0.511-1.142)	0.189		
smoking history	0.394 (0.096-1.615)	0.196		
pleural effusion	2.719 (0.847-8.725)	0.093		
lymphadenopathy	7.875 (1.972-31.444)	0.003	5.133 (1.218-21.630)	0.026
carcinomatous lymphangitis	18.000 (2.026-159.926)	0.010	9.545 (1.021-89.223)	0.048
pleural indentation	1.539 (0.517-4.585)	0.438		

*P*<sup>1</sup>-value was derived from the univariate logistic regression analyses between each of the variables; OR, odds ratio; CI, confidence interval.

### Combination of clinico-radiological features and radiomics score

Candidate clinico-radiological factors were integrated with Rad\_score into an assessment tool to predict the two groups. We used univariate and multivariate logistic regression (LR) analyses to select predictive variables and assess for interaction. To determine the independently significant factors of predicted model, variables at

univariate analysis with statistically significant were candidates for stepwise multivariate regression analysis. The nomogram model was formulated based on the regression coefficient of multivariate logistic regression analysis. In the nomogram model, we identified radiological variables lymphadenopathy and carcinomatous lymphangitis and Rad\_score as independent predictors for predictive of pre-treatment BM based on logistic regression (*P* < 0.05) (Figure 3). The AUC of the nomogram model was 0.961 (95% CI (0.920-1.000)) in the training

TABLE 3 The diagnostic performance of different types of models for predicting BM in stage III/IV lung adenocarcinoma.

Models	Datasets	ACC	AUC	95%CI	SEN	SPE	PPV	NPV
Clinical model	Training cohort	0.74.1	0.746	0.625-0.868	33.3% (7/21)	97.3% (36/37)	87.5% (7/8)	72.0% (36/50)
	Validation cohort	0.680	0.680	0.472-0.888	40.0% (2/5)	75.0% (15/20)	28.6% (2/7)	83.3% (15/18)
	External test cohort	0.588	0.661	0.487-0.836	81.3% (13/16)	38.9% (7/18)	54.2% (13/14)	70.0% (7/10)
Radiomics model	Training cohort	0.845	0.938	0.882-0.993	66.7% (14/21)	94.6% (35/37)	87.5% (14/15)	83.3% (35/42)
	Validation cohort	0.840	0.905	0.801-1.000	80.0% (4/5)	85.0% (17/20)	57.1% (4/7)	94.4% (17/18)
	External test cohort	0.794	0.880	0.767-0.993	68.7% (11/16)	88.9% (16/18)	84.6% (11/13)	76.2% (16/21)
Nomogram model	Training cohort	0.897	0.961	0.920-1.000	76.2% (16/21)	97.3% (36/37)	94.1% (16/17)	87.8% (36/41)
	Validation cohort	0.880	0.940	0.846-1.000	80.0% (4/5)	90.0% (18/20)	66.7% (4/6)	94.7% (18/19)
	External test cohort	0.794	0.896	0.792-1.000	81.3% (13/16)	77.8% (14/18)	76.5% (13/17)	82.4% (14/17)

SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; ACC, accuracy; CI, confidence interval.

cohort, 0.940(95% CI (0.846-1.000)) in the validation cohort, and The AUC of 0. 896(95% CI (0.792-1.000)) in the external test cohort (accuracy 79.4%) (Table 3). Figures 4A–C display the ROC curve for the clinical, radiomics and nomogram models across different datasets. The prediction performance of these models in the training, validation and external test cohort is displayed in Supplementary Figure 3.

Moreover, the Hosmer-Lemeshow test indicated the goodness of fit of logistic regression models ( $P>0.05$ ). The calibration curve was generated to illustrate the reliability and accuracy of the model’s predictions across various predicted probabilities. Well-calibrated models, as shown in Supplementary Figure 4, exhibited a close alignment between predicted and observed probabilities, indicating robust generalization and predictive capabilities. When we compared these models, we found that the nomogram model was superior to clinical model in three cohorts ( $P=0.001$  for training,  $P=0.009$  for validation,  $P=0.012$  for external test cohorts), while the predicted performance of nomogram model was comparable to radiomics model in three cohorts (Figures 5A–C). The radiomics model surpassed the clinical model in the training ( $P=0.007$ ) and external cohorts ( $P=0.033$ ), as shown in Figures 5A, C. Regarding clinical utility in validation and external test cohort, DCA curve showed that the nomogram model is comparable to the radiomics model, and both outperformed the clinical model (Figures 6A, B). The multivariate Cox regression analysis showed that lymphadenopathy (HR = 5.41, 95% CI: 1.38-21.16,  $P = 0.015$ ) and rad\_score (HR = 25.67, 95% CI: 5.41–121.94,  $P< 0.001$ ) were independent predictive factors for PFS (Table 4). Concordance index (C-index) for all the nomogram models in each cohort: training cohort (C-Index(95%CI):0.887 (0.826-0.956); testing cohort(C-Index(95%CI):0.798 (0.676-0.938). The C-index achieved 0.927 in the external cohort with 95%CI (0.857-0.996)).

### PFS analysis

The PFS analysis comprised 106 patients distributed across three cohorts: training cohort (n=54), validation cohort (n=22), and external test cohort (n=30). The rad\_score threshold (0.20) was computed by X-tile for dividing patients into high- and low-risk groups (Figure 7A). Patients in the low-risk group showed a significantly better PFS compared to those in the high-risk group in the training cohort and validation cohort ( $P$  all  $< 0.010$ , respectively) (Figures 7B, C), whereas the results were not consistent in the external test cohort ( $P=0.130$ ) (Figure 7D), which maybe contributed to that the follow-up duration of enrolled patients varied significantly.

### Discussion

Our research has developed a radiomics model to predict brain metastasis (BM) in stage III/IV ALK-positive lung adenocarcinoma patients after detecting the primary tumor. The model has demonstrated promising results. Furthermore, we have identified that the CT-based radiomics signature, transformed into quantitative Rad\_score, can serve as an independent predictor in the nomogram model. When incorporated into the nomogram, the Rad\_score significantly improves the model’s performance compared to a clinical model based solely on clinical variables, as evidence in the training, validation, and external test cohorts.

In our analysis of 117 patients, clinical characteristics such as age, gender, and smoking history were not significantly associated with brain metastasis (BM), consistent with finding from a previous study (35). However, unlike the findings of Amol Mujoomdar et al. (43), we



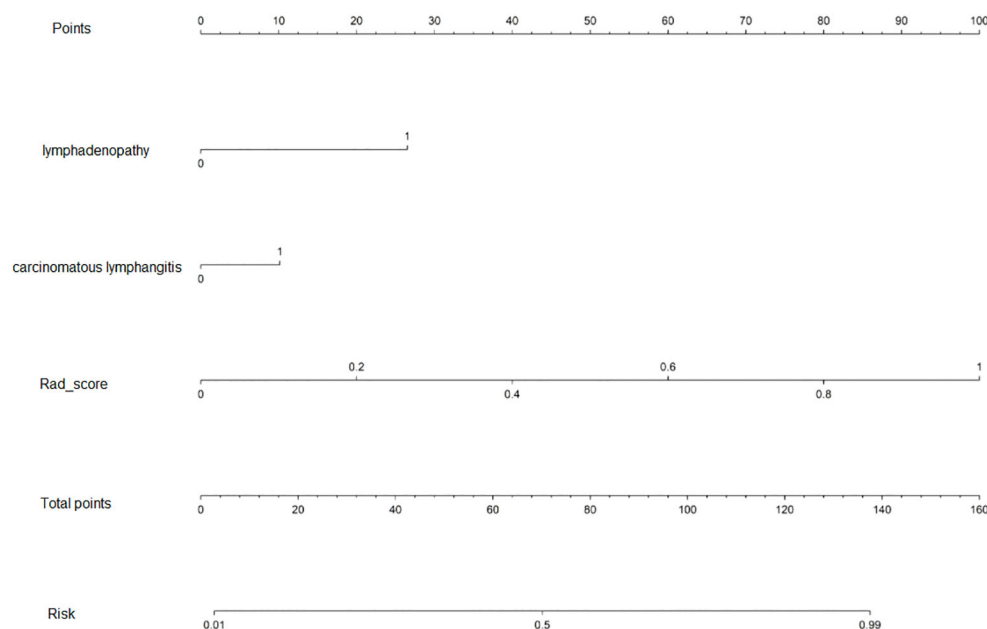


FIGURE 3

The nomogram is based on a nomogram model to predict the ALK-positive in stage III/IV lung adenocarcinoma patients. The probability of the stage III/IV ALK-positive lung adenocarcinoma, label patient value at each axis, draw a straight line perpendicular to the point axis, and sum the points for all variables in the left column.

observed that no significant difference in the diameter of the tumor between the BM and the non-BM group. This difference may be attributed to the inclusion of multiple subtypes of non-small cell lung cancer (NSCLC) in Amol Mujoomdar's study, whereas our focus was specifically on lung adenocarcinoma (40). Furthermore, methodological differences between two study may have contribute to the results (40), suggesting the need for further investigation (44).

Additionally, our study revealed no significant correlation between BM and pericardial effusion, pleural indentation, slightly contracting the findings of Lv et al. (43). However, we did observe a notable association between lymphadenopathy and an increased BM risk. Patients with BM often exhibited common carcinomatous lymphangitis and lymphadenopathy compared to those without

BM. This observation may be explained by an increased occurrence of metastasis in lymphadenopathy and carcinomatous lymphangitis, leading to the shedding of tumor cells that could potentially migrate to the brain with lymphatic vessels. However, the low sensitivity of clinical models in different datasets requires attention because that small sample sizes and imbalanced data can result in this phenomenon due to insufficient data diversity, which also echoed in previous literature (40). So, we endeavored to address this limitation through DCA curve analysis to make the performance comparison more reliable. DCA curve showed that the nomogram model is comparable to the radiomics model, and both outperformed the clinical model. The multivariate Cox regression analysis revealed that lymphadenopathy and rad\_score were independent predictive

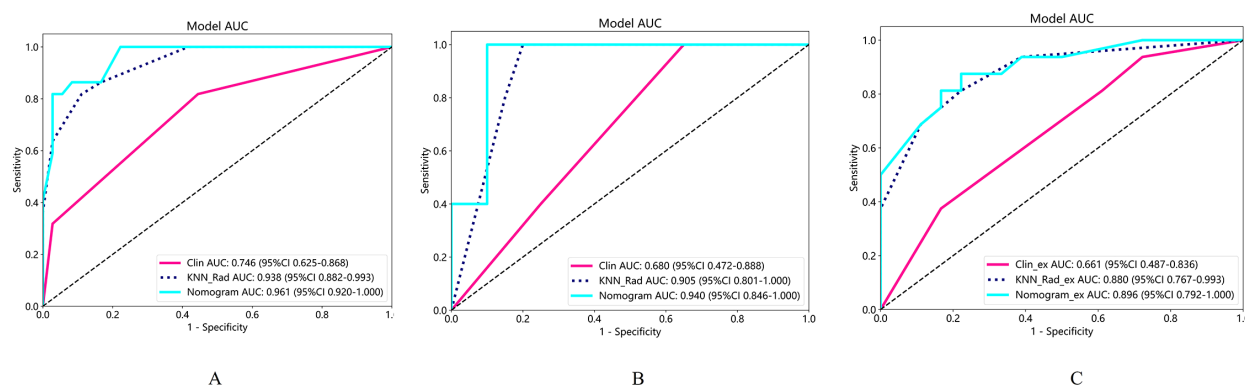


FIGURE 4

The ROC curve with AUCs of training set (A), validation set (B) and external test set (C) were exhibited.

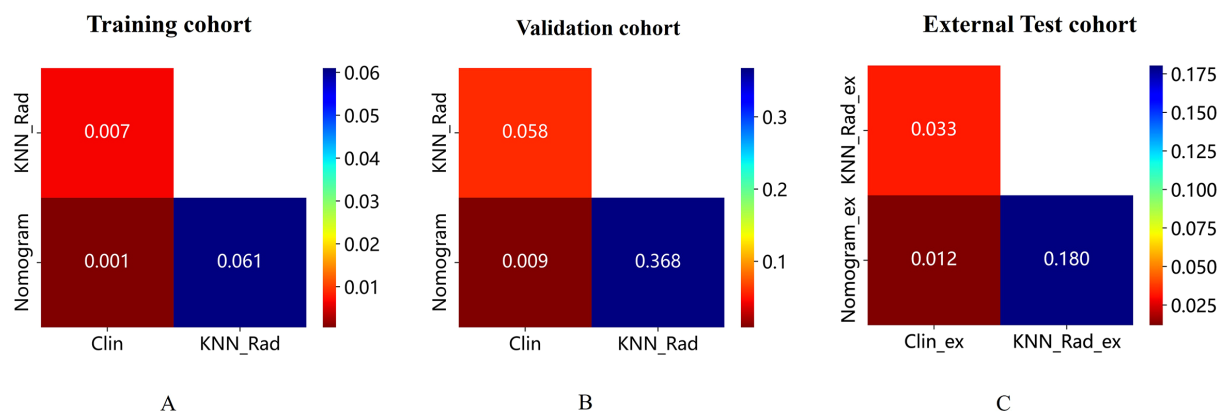


FIGURE 5

(A-C) shows DeLong test for the AUC of the three models in the training, validation, and external test cohorts, respectively. Radiomics showed superior predictive performance than clinical models in training and external tests, respectively; When combined with clinical and radiological features, the AUC of nomogram model was superior to clinical model in the three cohort, while comparable to radiomics model in three cohorts. AUC, the area under the curve; ROC, receiver operating characteristic curve.

factors for PFS, significantly increasing the risk of disease progression. The impact of rad\_score was stronger, suggesting it may be a more critical prognostic indicator. In both the training and validation cohorts, the low-risk group had significantly better progression-free survival (PFS) than the high-risk group ( $P$  all  $<0.010$ , respectively). This consistency supports the robustness of the risk stratification in predicting PFS in these cohorts. While the survival difference between the low- and high-risk groups was not statistically significant in the external test cohort, it might be because the follow up of the enrolled patients of external cohort varied significantly.

Given the limitations of the clinical model, there is a need for more accurate prediction model. Radiomics, an emerging field based on the hypothesis that radiological images reflect tumor biological characteristics and holds promise in this regard (14). Based on our current knowledge, there have been no previous explorations on the predictive role of CT-based radiomics analysis for brain metastases in

patients with ALK-positive lung adenocarcinoma. However, several obstacles need to be overcome. One of the significant challenges is the high dependency of these engineered features on imaging acquisition algorithms. To address this issue, we conducted image normalization. The resulting radiomics signature demonstrated favorable prediction performance in an external test cohort, with an area under the curve (AUC) of 0.880, 95%CI, 0.767-0.993. Among the selected radiomic features, 3D radiomics feature (lbp\_3D\_glszm\_GrayLevelVariance) was the top ranked feature for predicting brain metastasis of tumors. As tumors invasiveness increases, its shape becomes irregular and surface areas expands, marking 3D feature particularly informative and therefore had better diagnostic performance (45). In comparison, Xu et al. (46) attempted to develop a radiomic signature for predicting pre-treatment brain metastasis (BM) in stage III/IV ALK-positive NSCLC patients. They found that only one radiomic feature (W\_GLCM\_LH\_Correlation) was an independent predictor, with

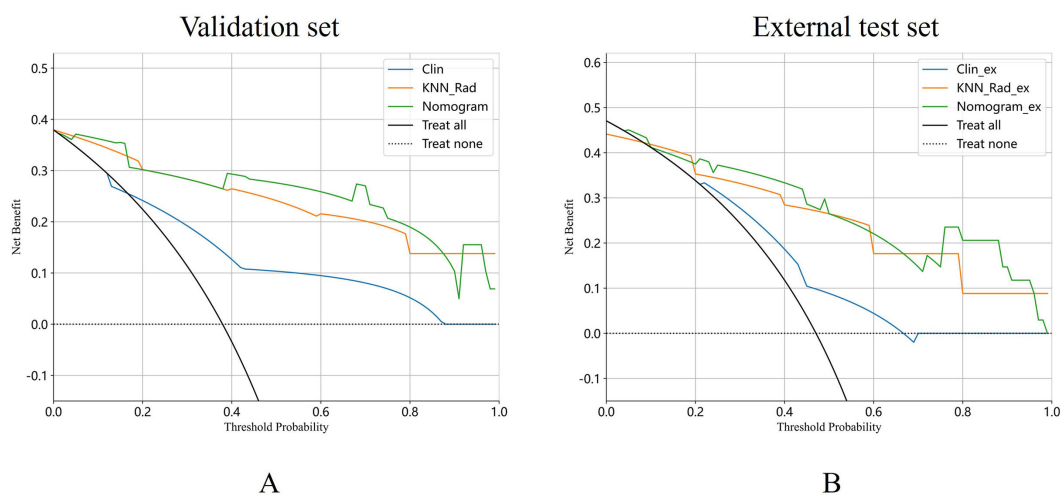


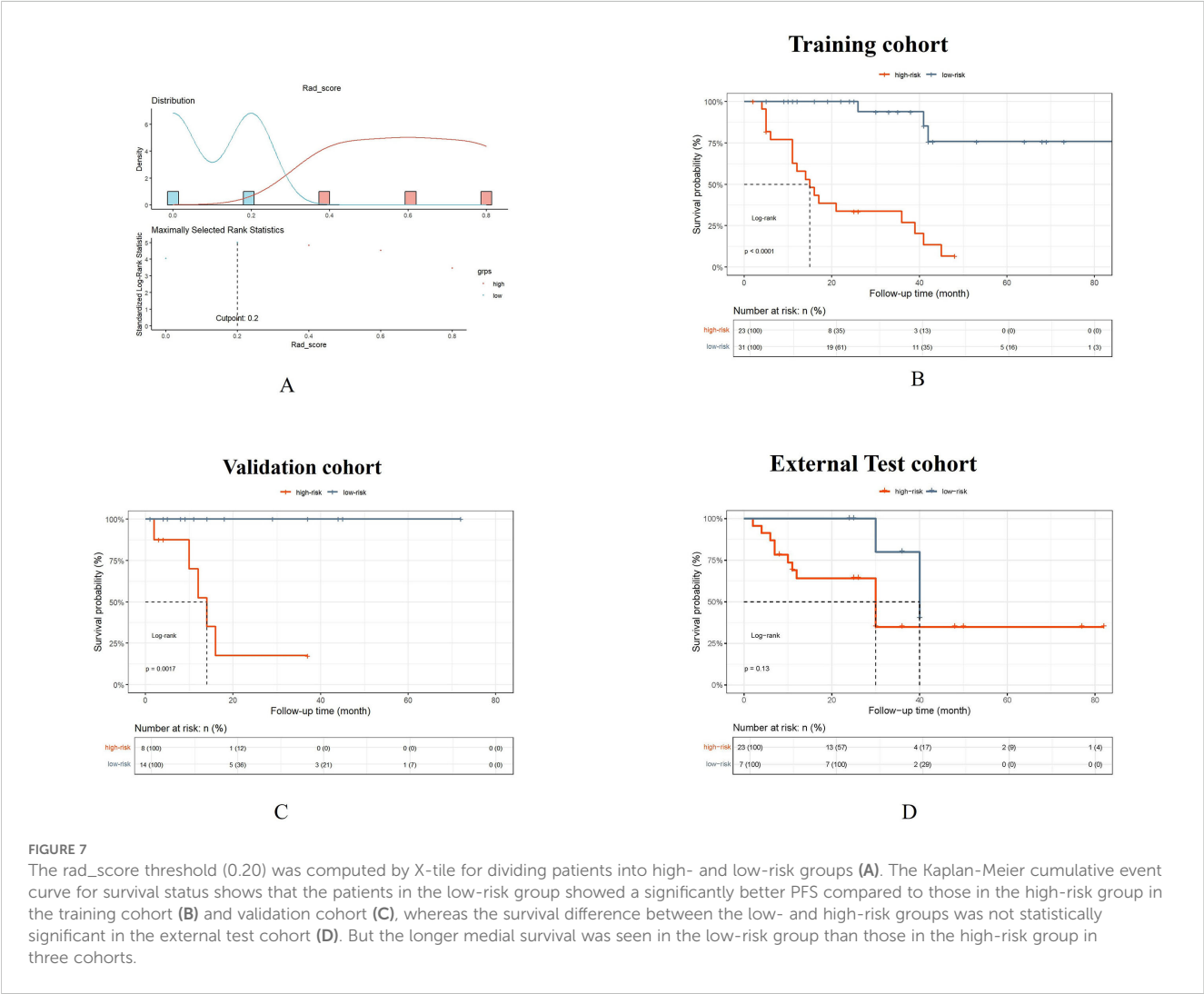
FIGURE 6

DCA curves of the clinical model, radiomics and nomogram model were exhibited in validation cohort (A) and external test cohort (B).

TABLE 4 Univariable and multivariable Cox regression analysis for risk factors associated with PFS in lung adenocarcinoma with ALK-positive.

Characteristics	Univariate Analysis			Multivariate Analysis		
	Hazard.Ratio	95% CI	P-value	Hazard.Ratio	95%CI	P-value
age	0.78	0.34-1.87	0.581			
gender	1.49	0.63-3.55	0.364			
location	0.96	0.69-1.32	0.787			
pleural effusion	2.55	1.05-6.21	0.039	1.45	0.48-1.4	0.510
lymphadenopathy	5.71	1.67-19.53	0.005	5.41	1.38-21.16	<b>0.015</b>
diameter	1.14	0.93-1.4	0.199			
carcinomatous lymphangitis	5.44	2.09-14.14	0.001	0.57	0.12-2.6	0.465
pleural indentation	1.32	0.54-3.18	0.543			
smoking history	0.85	0.25-2.93	0.800			
Radscore	23.66	6.33-88.39	<0.001	25.67	5.41-12.94	<b>&lt;0.001</b>

CI, confidence interval.  
The bold of P-value indicates statistically significant difference.



an AUC of 0.687 in the training cohort and 0.642 in the validation cohort. This feature also showed moderate performance in predicting BM during follow-up, with AUCs of 0.682 for stage III and 0.653 for stage IV. However, these AUC values were lower than those achieved by the radiomics model in the external test cohort of our study (AUC=0.803). It is worth noting that in previous research (35), there were 30 patients with pre-treatment BM out of a total of 77 patients. Dividing patients without BM at baseline into different stages subsequently reduced the sample size, which could have mitigated the statistical power compared to the initial cohort. Additionally, their research lacked external validation. It is reported that they tried to build a radiomic signature to predict pre-treatment BM for ALK-positive NSCLC patients and found that only 5 radiomic features were independent predictors (AUC=0.828), but they only performed internal test, therefore the independent model assessment could not be committed to avoid overfitting. In addition, because some patients did not undergo enhanced CT in their study, they used plain CT images to extract the radiomic features, which may affect the segmentation of the tumor.

When tumors are enriched in gene mutations, there are changes in downstream signaling pathways that eventually lead to comprehensive cellular responses and trigger processes such as cell proliferation and inflammation (47–49). The ALK gene, especially, is involved in cell proliferation signaling. The most common way of mutating in the ALK gene in NSCLC is the formation of the EML4-ALK fusion gene (50). Research by Martinengo et al. (51) shows that the hypoxia response is specifically enriched in a large series of human ALK-positive lymphoma and NSCLC cases, and they provide evidence that ALK specifically regulates the expression of HIF1 $\alpha$  and HIF2 $\alpha$  under hypoxic conditions in both anaplastic large-cell lymphoma and NSCLC, and that both HIF1 $\alpha$  and HIF2 $\alpha$  are essential for NSCLC growth and metastasis. Wang et al. (52) previously demonstrated that radiomics analysis was able to detect subtle changes in the metastatic parenchyma regardless of whether morphological metastasis was visible.

After comparison of several machine learning methods, as mentioned in other literature as usual (53, 54), we choose the KNN algorithm to be the best classifier to build the radiomics model. KNN, a widely used classification algorithm (55), is non-linear space partitioning approach may explain its superior performance compared to traditional SVM and RF classifiers. Our study performed by comparing different ML models showed that the KNN model with AUCs values of 0.938 and 0.905 in training and validation cohorts, respectively. In contrast, LR and MLP exhibited lower yet comparable performance, but have the risk of overfitting. Compared with the KNN model, these classifiers, including Light GBM, SWM, Randomforest, XGboost, Extra trees, GradientBoosting and AdaBoost all displayed the lower diagnostic performance in validation cohort. When selecting the optimal predictor, it is essential to qualitatively consider all these factors to achieve the best balance for the particular requirements (52). The study confirmed that the radiomics and machine learning analysis with KNN classifier can be method to predict early

metastasis in stage III/IV lung cancer patients with specific driver gene mutation (ALK-positive). Here, we need to describe the selected features used to build the model derived from the local binary patterns, a local-level feature extraction technique, which allows identification of a small, but biologically important, tumor niche area (a small number of pixels) within an otherwise homogeneous, larger tumor region (56). Even though the local binary patterns features have not been frequently used in radiomics study, the value of local binary patterns features for the analysis of small volume VOIs are also confirmed (57). In advanced-stage patients with large volume of tumors, local binary patterns features helped identify small but biologically important tumor niche within the homogeneous, tumor region.

There are several limitations to this study that should be acknowledged. Firstly, the small sample size and limited external validation due to low incidence of lung adenocarcinoma with ALK-positive and the high proportion of loss to follow-up need to expand the sample size in further investigations to enhance the reliability and generalizability of the study findings, which align with the precious study (58). Secondly, the regional characteristics of our sample population may introduce selection bias, limiting the generalizability of our findings. The remaining two radiomics features for model building in our study could raise the risk of overfitting, however, we have regularized the feature selection process to minimize overfitting. We also could find out the rare radiomics features for predicted model in the previous study (59). The Alvarez-Jimenez has preliminarily studied the cross-scale associations that may exist between digital pathology and CT imaging which can be used to identify relevant radiomic and histopathology features to accurately distinguish lung adenocarcinomas from squamous cell carcinomas (60). We also acknowledge the importance of exploring correlations between radiomics features and histopathological data for providing deeper insights into the underlying biological mechanisms of brain metastases. Future research should address these limitations to provide a more comprehensive understanding of the topic.

## Conclusion

In conclusion, our preliminary study demonstrated that radiomics signature derived from pretreatment CT can serve as a non-invasive biomarker for predicting BM in patients with stage III/IV ALK-positive lung adenocarcinoma, backed by an independent external test dataset. Furthermore, an individualized model that combines the rad-score and clinico-radiological data may aid in the management of patients with advanced ALK-positive lung adenocarcinoma.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving humans were approved by the ethics committee of the first affiliated hospital of Nanjing medical university. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the written informed consent was waived because of the retrospective and anonymous nature of the data analysis. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because the written informed consent was waived because of the retrospective and anonymous nature of the data analysis.

## Author contributions

FW: Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. CL: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. SL: Formal analysis, Methodology, Resources, Software, Visualization, Writing – review & editing. TZ: Conceptualization, Investigation, Methodology, Resources, Writing – review & editing. TY: Conceptualization, Investigation, Methodology, Resources, Writing – review & editing. WZ: Conceptualization, Investigation, Methodology, Resources, Writing – review & editing. JH: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. MY: Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. WG: Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Safiri S, Sohrabi MR, Carson-Chahhoud K, Bettampadi D, Taghizadieh A, Almasi-Hashiani A, et al. Burden of tracheal, bronchus, and lung cancer and its attributable risk factors in 204 countries and territories, 1990 to 2019. *J Thorac Oncol.* (2021) 16:945–59. doi: 10.1016/j.jtho.2021.03.030
3. Sacks P, Rahman M. Epidemiology of brain metastases. *Neurosurg Clin N Am.* (2020) 31:481–8. doi: 10.1016/j.nec.2020.06.001
4. Pan K, Concannon K, Li J, Zhang J, Heymach JV, Le X. Emerging therapeutics and evolving assessment criteria for intracranial metastases in patients with oncogene-driven non-small-cell lung cancer. *Nat Rev Clin Oncol.* (2023) 20:716–32. doi: 10.1038/s41571-023-00808-4
5. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature.* (2007) 448:561–6. doi: 10.1038/nature05945
6. Takamori S, Toyokawa G, Shimokawa M, Kinoshita F, Kozuma Y, Matsubara T, et al. Radiological features of brain metastases from non-small cell lung cancer harboring EGFR mutation. *Anticancer Res.* (2018) 38:3731–4. doi: 10.21873/anticancer.12653
7. Yousefi M, Bahrami T, Salmaninejad A, Nosrati R, Ghaffari P, Ghaffari SH. Lung cancer-associated brain metastasis: Molecular mechanisms and therapeutic options. *Cell Oncol (Dordr).* (2017) 40:419–41. doi: 10.1007/s13402-017-0345-5
8. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med.* (2014) 371:2167–77. doi: 10.1056/NEJMoa1408440
9. Petrelli F, Lazzari C, Ardito R, Borgonovo K, Bulotta A, Conti B, et al. Efficacy of ALK inhibitors on NSCLC brain metastases: A systematic review and pooled analysis of 21 studies. *PLoS One.* (2018) 13:e0201425. doi: 10.1371/journal.pone.0201425
10. Davies J, Martinec M, Coudert M, Delmar P, Crane G. Real-world anaplastic lymphoma kinase (ALK) rearrangement testing patterns, treatment sequences, and survival of ALK inhibitor-treated patients. *Curr Med Res Opin.* (2019) 35:535–42. doi: 10.1080/03007995.2018.1533458
11. Yu Y, Zhu F, Zhang W, Lu S. Comparison of efficacy and safety of brigatinib in first-line treatments for patients with anaplastic lymphoma kinase-positive non-small-cell lung cancer: A systematic review and indirect treatment comparison. *J Clin Med.* (2022) 11:2963. doi: 10.3390/jcm11112963

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The work was supported by the National Natural Science Foundation of China (Grant No. 82141121); The Specialized Diseases Clinical Research Fund of Jiangsu Province Hospital (Grant No. DL202412); The Joint Open Research Fund of Nanjing University of Posts and Telecommunications & Jiangsu Province Hospital; The Medical Engineering Translational Fund of Jiangsu Province Hospital (RG202401).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1585930/full#supplementary-material>



12. Pécoux CL, Sun A, Slotman BJ, De Ruyscher D, Belderbos J, Gore EM. Prophylactic cranial irradiation for patients with lung cancer. *Lancet Oncol.* (2016) 17: e277–93. doi: 10.1016/S1470-2045(16)30065-1
13. Patchell RA, Tibbs PA, Walsh JW, Dempsey RJ, Maruyama Y, Kryscio RJ, et al. A randomized trial of surgery in the treatment of single metastases to the brain. *N Engl J Med.* (1990) 322:494–500. doi: 10.1056/NEJM199002223220802
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
15. Wu J, Tha KK, Xing L, Li R. Radiomics and radiogenomics for precision radiotherapy. *J Radiat Res.* (2018) 59:25–31. doi: 10.1093/jrr/rxx102
16. Oh JE, Kim MJ, Lee J, Hur BY, Kim B, Kim DY, et al. Magnetic resonance-based texture analysis differentiating KRAS mutation status in rectal cancer. *Cancer Res Treat.* (2020) 52:51–9. doi: 10.4143/crt.2019.050
17. Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics approach to prediction of occult mediastinal lymph node metastasis of lung adenocarcinoma. *AJR Am J Roentgenol.* (2018) 211:109–13. doi: 10.2214/AJR.17.19074
18. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J Clin.* (2017) 67:93–9. doi: 10.3322/caac.21388
19. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012
20. Zhou T, Tu W, Dong P, Duan S, Zhou X, Ma Y, et al. CT-based radiomic nomogram for the prediction of chronic obstructive pulmonary disease in patients with lung cancer. *Acad Radiol.* (2023) 30:2894–903. doi: 10.1016/j.acra.2023.03.021
21. Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, et al. FeAture Explorer (FAE): A tool for developing and comparing radiomics models. *PLoS One.* (2020) 15: e0237587. doi: 10.1371/journal.pone.0237587
22. Yu R, Cai L, Gong Y, Sun X, Li K, Cao Q, et al. MRI-based machine learning radiomics for preoperative assessment of human epidermal growth factor receptor 2 status in urothelial bladder carcinoma. *J Magn Reson Imaging.* (2024) 60:2694–704. doi: 10.1002/jmri.29342
23. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med.* (2007) 26:5512–28. doi: 10.1002/sim.3148
24. Paul A, Mukherjee DP, Das P, Gangopadhyay A, Chintha AR, Kundu S. Improved random forest for classification. *IEEE Trans Image Process.* (2018) 27:4012–24. doi: 10.1109/TIP.2018.2834830
25. Huang S, Cai N, Pacheco PP, Narrantes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics.* (2018) 15:41–51. doi: 10.21873/cgp.20063
26. Sugahara S, Ueno M. Exact learning augmented naive bayes classifier. *Entropy (Basel).* (2021) 23:1703. doi: 10.3390/e23121703
27. Lorencin I, Andelić N, Španjol J, Car Z. Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis. *Artif Intell Med.* (2020) 102:101746. doi: 10.1016/j.artmed.2019.101746
28. Alie MS, Negesse Y, Kindie K, Merawi DS. Machine learning algorithms for predicting COVID-19 mortality in Ethiopia. *BMC Public Health.* (2024) 24:1728. doi: 10.1186/s12889-024-19196-0
29. Hosseini MP, Hosseini A, Ahi K. A review on machine learning for EEG signal processing in bioengineering. *IEEE Rev BioMed Eng.* (2021) 14:204–18. doi: 10.1109/RBME.2020.2969915
30. Chen Z, He N, Huang Y, Qin WT, Liu X, Li L. Integration of A deep learning classifier with A random forest approach for predicting malonylation sites. *Genomics Proteomics Bioinf.* (2018) 16:451–9. doi: 10.1016/j.gpb.2018.08.004
31. Ghiassi MM, Zendehboudi S. Application of decision tree-based ensemble learning in the classification of breast cancer. *Comput Biol Med.* (2021) 128:104089. doi: 10.1016/j.combiomed.2020.104089
32. Yan J, Xu Y, Cheng Q, Jiang S, Wang Q, Xiao Y, et al. LightGBM: accelerated genomically designed crop breeding through ensemble learning. *Genome Biol.* (2021) 22:271. doi: 10.1186/s13059-021-02492-y
33. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot.* (2013) 7:21. doi: 10.3389/fnbot.2013.00021
34. Zheng Z, Yang Y. Adaptive boosting for domain adaptation: toward robust predictions in scene segmentation. *IEEE Trans Image Process.* (2022) 31:5371–82. doi: 10.1109/TIP.2022.3195642
35. Wang H, Chen YZ, Li WH, Han Y, Li Q, Ye Z. Pretreatment thoracic CT radiomic features to predict brain metastases in patients with ALK-rearranged non-small cell lung cancer. *Front Genet.* (2022) 13:772090. doi: 10.3389/fgene.2022.772090
36. Shi Z, Huang X, Cheng Z, Xu Z, Lin H, Liu C, et al. MRI-based quantification of intratumoral heterogeneity for predicting treatment response to neoadjuvant chemotherapy in breast cancer. *Radiology.* (2023) 308:e222830. doi: 10.1148/radiol.222830
37. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J.* (2021) 21:1643–8. doi: 10.1016/j.spinee.2021.02.024
38. Li Y, Zhao J, Li R, Yao X, Dong X, Zhao Y, et al. Predicting prognosis in patients with stage IA lung adenocarcinoma with a micropapillary component using a nomogram based on computed tomography radiomics and clinicopathologic factors: a retrospective analysis. *Transl Lung Cancer Res.* (2024) 13:2585–602. doi: 10.21037/tlcr-24-544
39. Zheng Z, Liu C, Shen Y, Xia L, Xiao L, Sun Y, et al. Serum albumin levels as a potential marker for the predictive and prognostic factor in sudden sensorineural hearing loss: A prospective cohort study. *Front Neurol.* (2021) 12:747561. doi: 10.3389/fneur.2021.747561
40. Li H, Chen XL, Liu H, Liu YS, Li ZL, Pang MH, et al. MRI-based multiregional radiomics for preoperative prediction of tumor deposit and prognosis in resectable rectal cancer: a bicenter study. *Eur Radiol.* (2023) 33:7561–72. doi: 10.1007/s00330-023-09723-9
41. Wu YJ, Liu YC, Liao CY, Tang EK, Wu FZ. A comparative study to evaluate CT based semantic and radiomic features in preoperative diagnosis of invasive pulmonary adenocarcinomas manifesting as subsolid nodules. *Sci Rep.* (2021) 11:66. doi: 10.1038/s41598-020-79690-4
42. Wang Q, Guo A. An efficient variance estimator of AUC and its applications to binary classification. *Stat Med.* (2020) 39:4281–300. doi: 10.1002/sim.8725
43. Mujoondar A, Austin JH, Malhotra R, Powell CA, Pearson GD, Shiao MC, et al. Clinical predictors of metastatic disease to the brain from non-small cell lung carcinoma: primary tumor size, cell type, and lymph node metastases. *Radiology.* (2007) 242:882–8. doi: 10.1148/radiol.2423051707
44. Lv Z, Fan J, Xu J, Wu F, Huang Q, Guo M, et al. Value of 18F-FDG PET/CT for predicting EGFR mutations and positive ALK expression in patients with non-small cell lung cancer: a retrospective analysis of 849 Chinese patients. *Eur J Nucl Med Mol Imaging.* (2018) 45:735–50. doi: 10.1007/s00259-017-3885-z
45. Wan Q, Zhou J, Xia X, Hu J, Wang P, Peng Y, et al. Diagnostic performance of 2D and 3D T2WI-based radiomics features with machine learning algorithms to distinguish solid solitary pulmonary lesion. *Front Oncol.* (2021) 11:683587. doi: 10.3389/fonc.2021.683587
46. Xu X, Huang L, Chen J, Wen J, Liu D, Cao J, et al. Application of radiomics signature captured from pretreatment thoracic CT to predict brain metastases in stage III/IV ALK-positive non-small cell lung cancer patients. *J Thorac Dis.* (2019) 11:4516–28. doi: 10.21037/jtd.2019.11.01
47. Diederichs S, Bulk E, Steffen B, Ji P, Tickenbrock L, Lang K, et al. S100 family members and trypsinogens are predictors of distant metastasis and survival in early-stage non-small cell lung cancer. *Cancer Res.* (2004) 64:5564–9. doi: 10.1158/0008-5472.CAN-04-2004
48. Koay EJ, Lee Y, Cristini V, Lowengrub JS, Kang Y, Lucas FAS, et al. A visually apparent and quantifiable CT imaging feature identifies biophysical subtypes of pancreatic ductal adenocarcinoma. *Clin Cancer Res.* (2018) 24:5883–94. doi: 10.1158/1078-0432.CCR-17-3668
49. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife.* (2017) 6:e23421. doi: 10.7554/eLife.23421
50. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. *N Engl J Med.* (2017) 377:829–38. doi: 10.1056/NEJMoa1704795
51. Martinengo C, Poggio T, Menotti M, Scalzo MS, Mastini C, Ambrogio C, et al. ALK-dependent control of hypoxia-inducible factors mediates tumor growth and metastasis. *Cancer Res.* (2014) 74:6094–106. doi: 10.1158/0008-5472.CAN-14-0268
52. Wang G, Wang B, Wang Z, Li W, Xiu J, Liu Z, et al. Radiomics signature of brain metastasis: prediction of EGFR mutation status. *Eur Radiol.* (2021) 31:4538–47. doi: 10.1007/s00330-020-07614-x
53. Lissos CS, Lissos CG, Mezger MF, Wolf D, Schmidt SA, Thaiss WM, et al. Deep neural networks and machine learning radiomics modelling for prediction of relapse in mantle cell lymphoma. *Cancers (Basel).* (2022) 14:2008. doi: 10.3390/cancers14082008
54. Granata V, Fusco R, De Muzio F, Cutolo C, Setola SV, Dell'Aversana F, et al. Radiomics and machine learning analysis based on magnetic resonance imaging in the assessment of liver mucinous colorectal metastases. *Radiol Med.* (2022) 127:763–72. doi: 10.1007/s11547-022-01501-9
55. Dondi F, Gatta R, Albano D, Bellini P, Camoni L, Treglia G, et al. Role of radiomics features and machine learning for the histological classification of stage I and stage II NSCLC at [<sup>18</sup>F]FDG PET/CT: A comparison between two PET/CT scanners. *J Clin Med.* (2022) 12:255. doi: 10.3390/jcm12010255
56. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. *AJNR Am J Neuroradiol.* (2018) 39:208–16. doi: 10.3174/ajnr.A5391
57. Peeken JC, Shouman MA, Kroenke M, Rauscher I, Maurer T, Gschwend JE, et al. A CT-based radiomics model to detect prostate cancer lymph node metastases in PSMA radioguided surgery patients. *Eur J Nucl Med Mol Imaging.* (2020) 47:2968–77. doi: 10.1007/s00259-020-04864-1
58. Deng F, Liu Z, Fang W, Niu L, Chu X, Cheng Q, et al. MRI radiomics for brain metastasis sub-pathology classification from non-small cell lung cancer: a machine learning, multicenter study. *Phys Eng Sci Med.* (2023) 46:1309–20. doi: 10.1007/s13246-023-01300-0
59. Xie PY, Zeng ZM, Li ZH, Niu KX, Xia T, Ma DC, et al. MRI-based radiomics for stratifying recurrence risk of early-onset rectal cancer: a multicenter study. *ESMO Open.* (2024) 9:103735. doi: 10.1016/j.esmoop.2024.103735
60. Alvarez-Jimenez C, Sandino AA, Prasanna P, Gupta A, Viswanath SE, Romero E. Identifying cross-scale associations between radiomic and pathomic signatures of non-small cell lung cancer subtypes: preliminary results. *Cancers (Basel).* (2020) 12:3663. doi: 10.3390/cancers12123663