Machine learning algorithms for predicting malignancy grades of lung adenocarcinoma and guiding treatments: CT radiomics-based comparisons

Jun Zhu[1#^], Jiayu Tao[2#], Fengfeng Zhang[2#], Jie Yao[1], Ke Chen[1], Yuxuan Wang[1], Xiaochen Lu[1], Bin Ni[1], Maoshan Zhu[3]

[1]Department of Thoracic Surgery, the First Affiliated Hospital of Soochow University, Suzhou, China; [2]Department of Oncology, the First Affiliated Hospital of Soochow University, Suzhou, China; [3]Department of Thoracic Surgery, Lianyungang Affiliated Hospital of Nanjing University of Chinese Medicine, Lianyungang, China

Contributions: (I) Conception and design: M Zhu, B Ni, J Zhu; (II) Administrative support: M Zhu, B Ni; (III) Provision of study materials or patients: J Zhu; (IV) Collection and assembly of data: J Zhu, J Tao, F Zhang, J Yao, K Chen, Y Wang, X Lu; (V) Data analysis and interpretation: J Zhu, J Tao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Maoshan Zhu, MMed. Department of Thoracic Surgery, Lianyungang Affiliated Hospital of Nanjing University of Chinese Medicine, 160 Chaoyang Middle Road, Haizhou District, Lianyungang 222004, China. Email: zhumaoshan110@163.com; Bin Ni, MD. Department of Thoracic Surgery, the First Affiliated Hospital of Soochow University, 188 Shizi Street, Gusu District, Suzhou 215006, China. Email: nb_fyywk@163.com.

Background: Lung adenocarcinoma (LUAD) is the most frequently diagnosed subtype of non-small cell lung cancer (NSCLC). Notably, prognosis can vary significantly among LUAD patients with different tumor subtypes. The advent of radiomics and machine learning (ML) technologies enables the development of non-invasive pathology predictive models. We attempted to develop computed tomography (CT) radiomics-based diagnostic models, enhanced by ML, to predict LUAD malignancy grade and guide surgical strategies.

Methods: In this retrospective analysis, a total of 168 surgical patients with histology-confirmed LUAD were divided into low-risk group (n=93) and intermediate-to-high-risk group (n=75) based on postoperative pathology. The region of interest (ROI) was delineated on the preoperative CT images for all patients, followed by the extraction of radiomic features. Patients were randomly allocated to a training set (n=117) and a testing set (n=51) in a 7:3 ratio. Within the training set, clinical-radiological model (CM) and radiomics model (RM) were developed utilizing patients' clinical characteristics, radiological semantic features, and radiomic features, along with the calculation of Rad scores. After the Rad scores were combined with independent risk factors among clinical-radiological features, logistic regression (LR), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), K-nearest neighbors (KNN), and naïve Bayes model (NBM) were employed to create different comprehensive models (COMs). The optimal model was identified based on the receiver operating characteristic (ROC) curves and the DeLong test. Finally, Shapley additive explanations (SHAP) were utilized to visualize the predictive processes of the models.

Results: Among the 168 patients enrolled, there were 50 males (29.76%) aged 56 (49.25, 67.00) years and 118 females (70.24%) aged 56.5 (42.00, 64.00) years; Diameter (P<0.001), and consolidation-to-tumor ratio (CTR) ≥0.5 (P=0.002) were identified as independent risk factors for the malignancy degree of LUAD during CM creation. The CM had an area under the ROC curve (AUC) of 0.909 [95% confidence interval (CI): 0.856–0.962] in the training set and 0.920 (95% CI: 0.846–0.994) in the testing set. The RM, comprising seven radiomic features, achieved an AUC of 0.961 (95% CI: 0.926–0.996) in the training set and 0.957

^ ORCID: 0009-0009-1276-1927.

(95% CI: 0.905–1.000) in the testing set. Among models created using various ML algorithms, the XGBoost model was identified as the optimal model. SHAP visualization revealed the model prediction processes and the values of different features.

Conclusions: We constructed and validated a robust, integrative model leveraging ML and CT radiomics, which amalgamates radiomics, clinical, and radiological attributes to precisely identify LUADs with elevated postoperative pathological grades. This enables doctors to formulate different surgical plans according to the pathology of the patients' tumors before the operation.

Keywords: Radiomics; lung nodules; prediction model; lung cancer

Introduction

Lung cancer represents the most prevalent malignancy and the leading cause of cancer-related mortality in China (1). The tumor-node-metastasis (TNM) staging system is the most widely used classification of lung cancer (2), with TNM stages being significantly associated with prognosis. As the predominant histological subtype of lung cancer, lung adenocarcinomas (LUADs) have completely different prognoses even if the diseases are

within the same TNM stage, possibly due to their distinct histological subtypes and grades. LUADs are histologically divided into seven different types: adenocarcinoma in situ (AIS), microinvasive adenocarcinoma (MIA), lepidic-predominant adenocarcinoma (LCA), acinar-predominant adenocarcinoma (ACA), papillary-predominant adenocarcinoma (PPA), micropapillary-predominant adenocarcinoma (MPA), and solid-predominant adenocarcinoma (SPA). Among these subtypes, AIS and MIA are categorized as non-invasive adenocarcinomas, whereas the remaining five are collectively termed invasive adenocarcinomas (IA). Additionally, certain non-conventional complex glandular patterns, such as cribriform and fused-gland architectures, have been identified as indicative of a poorer prognosis. To delineate the grading of IA and facilitate prognostic guidance, the International Association for the Study of Lung Cancer (IASLC) pathology committee has developed a novel grading model (3-5): (I) well-differentiated adenocarcinomas: lepidic predominant tumors, with no or less than 20% high-grade patterns (solid, micropapillary, or complex gland), typically having an excellent prognosis (6); (II) moderately-differentiated adenocarcinomas: acinar or papillary predominant tumors, with no or less than 20% of high-grade patterns, typically having moderate malignancy; and (III) poorly-differentiated adenocarcinomas: any tumor with 20% or more of high-grade patterns, typically having the poorest prognosis and usually necessitating adjuvant post-operative therapy (7,8).

The IASLC histological class has long been validated as a prognostic factor for adenocarcinomas (9). Tailored surgical treatments are needed for different classes of LUAD to maximize survival benefits. For IA ≤1 cm, lobectomy or segmentectomy offers superior survival benefits over wedge

Highlight box

Key findings
- This study developed a model for predicting the malignancy grade of lung adenocarcinoma (LUAD), with XGBoost emerging as the superior machine learning (ML) algorithm, offering a foundation and research direction for surgical interventions.

What is known and what is new?
- LUAD prognosis ties to histological type and grade, with surgery scope dependent on grade. Sublobectomy is favored for peripheral small non-small cell lung cancer, yet studies on optimal surgical approaches for various pathological subtypes are limited. Radiomics aids in diagnosing lung nodules, with limited research guiding surgical approaches for LUAD.
- Clinical models identify independent risk factors, but their performance is moderate; radiomics models outperformed the clinical models. The logistic regression model surpasses the single model, with XGBoost being the top-performing comprehensive model in ML.

What is the implication, and what should change now?
- Radiomics and ML aid clinical surgery decisions, guide model optimization, and support surgical method selection.
- It is essential to enhance and standardize image data processing and delineation and improve model generalization.

resection. Furthermore, in patients with LCA or ACA, wedge resection yielded relapse-free survival (RFS) and overall survival (OS) rates comparable to those achieved with anatomical lung resections; in contrast, in patients with PPA, MPA, or SPA, wedge resection was associated with a lower OS compared to those undergoing anatomical lung resections (9,10). Altorki *et al.* (11) in a phase 3 trial found that sublobectomy, including wedge resection and segmentectomy, was non-inferior to lobectomy for disease-free survival (DFS) in peripheral non-small cell lung cancer (NSCLC) patients with a tumor size of 2 cm or less and pathologically confirmed node-negative disease in the hilar and mediastinal lymph nodes. Wedge resection, a less invasive sublobar surgery, better preserves lung function, particularly in elderly patients with peripheral LUAD. Chiang *et al.* (12) confirmed better perioperative outcomes for wedge resection patients, with shorter operative time, length of hospital stay, and post-operative chest tube duration and less blood loss. McGuire *et al.* (13) found no significant differences in recurrence, mortality, or DFS between wedge resection and lobectomy in 423 patients with stage IA tumors sized 2 cm or smaller.

Precise preoperative prediction of tumor pathology equips surgeons to select optimal surgical approaches tailored to individual patients. Currently, two primary modalities for obtaining pathological diagnosis of pulmonary lesions in clinical practice are percutaneous lung biopsy (14,15) and bronchial biopsy (16). Although these invasive procedures represent effective diagnostic strategies, they are associated with several drawbacks. Patients often experience discomfort, and the diagnostic yield is suboptimal. Additionally, financial burden, potential needle-tract metastases, and technical challenges pose significant barriers. Specifically, percutaneous biopsy of pulmonary nodules measuring less than 1 cm or located centrally is technically demanding and carries a higher risk of complications. Intraoperative frozen section (IFS) has emerged as a valuable intraoperative tool, enabling rapid discrimination between benign and malignant nodules, it is still difficult to distinguish lung cancer subtypes during surgery. Yeh *et al.* (17) reviewed frozen sections and permanent section slides from 361 resected stage I LUAD ≤3 cm in size for predominant histological subtype and the presence or absence of lepidic, acinar, papillary, micropapillary, and solid patterns and found that determining the malignancy grades of IFS was quite difficult even for experienced pulmonary pathologists; notably, MIA was often over-assessed as IA, which could result in elevated

malignancy class and potential need for reoperation (18). Therefore, we tried to develop a computed tomography (CT) radiomics model (RM) to predict the pathological subtypes of $T_1N_0M_0$ LUAD, which may inform the option for wedge resection with RFS and OS akin to more extensive resections.

High-resolution CT enables visualization of lung nodule heterogeneity at the cellular level (19), and radiomics employs computer algorithms to mine large amounts of usable image data, thus revealing the underlying pathophysiological patterns, analyzing and quantifying the radiological data of lesions, and detecting features indiscernible to the human eye. Recently, machine learning (ML) algorithms, enhanced by deep learning (DL) (20-23), have bolstered radiomics with high repeatability and accuracy, enabling personalized and precise cancer diagnostics and treatments. Consequently, this study utilized CT radiomics and ML algorithms to establish a reliable non-invasive clinical prediction model for pulmonary nodules measuring ≤3 cm. The model was designed to classify the pathological subtypes of LUAD before patients underwent treatment. Specifically, patients were stratified into a low-risk (LR) group and an intermediate-to-high-risk (IHR) group. Multivariate statistical analyses were employed to identify the optimal model. The investigation further delved into how this model could be effectively used to guide subsequent surgical treatment plans. Successful implementation of the proposed model is projected to minimize surgical complications, hasten postoperative recovery, maintain pulmonary function, and ultimately enhance patient prognosis. We present this article in accordance with the TRIPOD reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-2025-310/rc).

Methods

Patient enrolment and grouping

This study fully complied with the Declaration of Helsinki and its subsequent amendments. This study was approved by the Ethics Committee of the First Affiliated Hospital of Soochow University (No. 2024668). As the analyzed data were anonymous and did not encroach on patient privacy, the need for obtaining signed informed consent for this retrospective analysis was waived. Eligible lung cancer patients were enrolled from the Department of Thoracic Surgery at the First Affiliated Hospital of Soochow

University between 1 September 2023 and 1 September 2024. The inclusion criteria were as follows: (I) having undergone high-resolution CT scan within one week before surgery; (II) with nodule diameter (including maximum transverse diameter and maximum longitudinal diameter) ≤3 cm; (III) having complete clinical baseline data; and (IV) having complete postoperative pathology data, which confirmed the diagnosis of LUAD. The exclusion criteria were as follows: (I) with a prior history of malignancy; (II) with pathologically confirmed mucinous adenocarcinoma; (III) having received anti-inflammatory treatment within one month before surgery; (IV) with incomplete or low-quality CT data; and (V) with incomplete or missing clinical data. Ultimately, following a strict screening process, 168 LUAD patients were included in the final analysis out of a total of 202 patients.

Patients were divided into LR and IHR based on postoperative pathology. Patients were classified as LR group (n=93) if postoperative pathology revealed AIS, MIA, or LCA, without or with ≤20% high-grade patterns (solid, micropapillary, or complex gland). Conversely, those with ACA, PPA, or with >20% papillary, micropapillary, solid, or complex glandular components were classified into the IHR group (n=75). The dataset was randomly allocated into training and testing sets in a ratio of 7:3.

Acquisition of CT data

The CT data in Digital Imaging and Communications in Medicine (DICOM) format were sourced from the in-hospital database. Patients were scanned using TOSHIBA 256 iCT and Definition AS128 CT scanners (Toshiba, Tokyo, Japan), with the following scanning parameters: tube voltage, 120 kV; tube current, 110–240 mA; rotation time, 0.5 s; slide thickness during scanning, 5 mm; and reconstructed slide thickness, 1 mm. Patients were asked to hold their breath during CT scans after inspiration. The scanning area extended from the thoracic inlet to a point 5 cm below the costophrenic angle.

Acquisition of clinical-radiological data

The clinical information and radiological data were obtained from the hospital database. The clinical information included gender, age, smoking history, respiratory conditions, and postoperative pathology. Radiological semantic features (24,25) were assessed by two radiologists with ≥3 years of pulmonary nodule reading experience,

who independently reviewed CT data without knowledge of clinical details or final pathology. Another radiologist with >10 years of imaging experience arbitrated any discrepancies. The radiological semantic features assessed included the following: (I) maximum nodule diameter (mm); (II) the consolidation-to-tumor ratio (CTR), coded as 0 if CTR below 50% or 1 if CTR greater than 50%; (III) lobulation, coded as 0 if lobulation absent and 1 if present; (IV) nodule shape regularity, coded as 0 if irregular and 1 if regular; (V) spicule sign, coded as 0 if absent and 1 if present; (VI) marginal blurring, coded as 0 if blurred and 1 if clear; (VII) pleural indentation sign, coded as 0 if absent and 1 if present; (VIII) vascular convergence sign, coded as 0 if absent and 1 if present; (IX) vacuole sign, coded as 0 if absent and 1 if present; (X) air bronchogram sign, coded as 0 if absent and 1 if present; and (XI) relative position of nodules to the hilum, classified as peripheral or central.

Delineation of CT images

The delineation of CT images was performed using the open-source software 3D Slicer (version 5.6.2, https://download.slicer.org/) according to the following principles: (I) manual delineation in 3D Slicer; (II) including nodule-bound vessels and abnormal bronchi, while excluding unrelated ones; (III) delineating the spicule sign and pleural indentation sign-related traction lines; and (IV) incorporating any intra-nodular cavities.

Creation of clinical model and its performance assessment

Independent predictors from clinical and radiological semantic features were selected using univariate and multivariate LR analyses on the clinical baseline features of the training set to create a clinical-radiological model (CM). Model performance was assessed using receiver operating characteristic (ROC) curves and areas under the ROC curves (AUCs).

Extraction and screening of radiomic features

Image features were extracted using Slicer Radiomics, a Python package in 3D Slicer, with a resampling parameter of 1 mm × 1 mm × 1 mm. This package can extract a total of 851 features including shape-based features, first-order statistics, second-order texture features, and wavelet transforms. Shape-based features are fundamental extrinsic descriptors of region of interest (ROI), including aspects

such as sphericity and compactness, which define the shape, size, and surface characteristics. First-order statistics summarize ROI intensity and variations, for example, means and medians, but ignore spatial relationships. In contrast, the second-order features capture voxel interrelationships. Typically, these features are extracted from various matrices, including the Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GLRLM), and Gray-Level Zone of Dependence Matrix (GLZDM), which encapsulate detailed spatial information within the ROI. Among the 851 extracted features, the Max-Relevance and Min-Redundancy (mRMR) algorithm was employed within the training set to identify the top 10 features that exhibited the strongest correlation with pathological subtypes while minimizing inter-feature redundancy. Subsequently, the least absolute shrinkage and selection operator (LASSO) was utilized for further feature refinement, with the aim of selecting those features associated with the lowest λ value for inclusion in the final model construction.

Development and validation of RM and multi-ML model

The imaging model was built using optimal features and their coefficients from LASSO, and radiomics scores (Rad score) were yielded for all patients. The comprehensive models (COMs) were constructed using the Rad scores and clinical-radiological features. To determine the optimal COM, a suite of ML algorithms was employed, including LR, decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), K-nearest neighbors (KNN), and naïve Bayes model (NBM). The performance of these models was evaluated using the ROC curves, AUC values, F1 scores, sensitivity, and specificity. The DeLong test was utilized to assess AUC differences. The net benefits were assessed via decision curve analysis (DCA). Shapley additive explanations (SHAP) was utilized to visualize and analyze the prediction processes of the best COM.

Statistical analysis

Statistical analysis was performed using packages in the R Studio software (version 4.2.1). The compare groups package was used for baseline description and difference analysis. The normally distributed variables were presented as mean ± standard deviation (SD), and the non-normally distributed variables as median (Q1, Q3). Qualitative data are presented as cases and percentages. The glm package

was used for multivariate LR analysis for identifying independent predictors for CMs. The mRMRe package was applied for feature selection via mRMR, whereas glmnet was used for LASSO regression to build the RM. The Rad scores, together with clinical-radiological features, were used for the creation of COMs. The rpart package was used for constructing the DT model, randomForest for the RF model, xgboost for the XGBoost model, e1071 for the NBM and SVM, and kknn for KNN. After the models were established, the ROCR package was used for ROC curve plotting and AUC calculation, with pROC conducting DeLong tests to compare model AUCs. The nricens package was used to compute the net reclassification index (NRI) to assess the incremental benefit of the COMs over single models. Finally, the "dcurves" package was applied for multi-model DCA to evaluate the predictive performance of the COMs.

Results

Baseline data and CM

A total of 168 LUAD patients entered the final analysis, including 50 males (29.76%) aged 56 (49.25, 67.00) years and 118 females (70.24%) aged 56.5 (42.00, 64.00) years, with 93 cases in the LR group and 75 cases in the IHR group. The training set consisted of 117 patients, which included 66 cases of LR and 51 cases of IHR. The clinical data (including age, gender, smoking history, and pulmonary diseases) and radiological features (including maximum nodule diameter, CTR, peripheral or central location, lobulation, shape, spicule sign, marginal blurring, pleural indentation sign, vascular convergence sign, vacuole sign, and air-bronchogram sign) are shown in *Table 1*.

Univariate analysis indicated that air bronchogram sign, vacuole sign, vascular convergence sign, pleural indentation sign, marginal blurring, irregularity, lobulation sign, CTR ratio ≥0.5, tumor diameter, age, and gender were significantly associated with the malignancy grade of LUAD (all P<0.05). Subsequent multivariate analysis, using the backward approach of Akaike information criterion (AIC) for model selection, identified tumor diameter (P<0.001), and CTR ≥0.5 (P=0.002) as independent risk factors for the malignancy grade of LUAD, with the AIC value of model being 93.64 (*Table 2*). The CM of the established clinical model was calculated using the following formula, with the AUC of the training set and testing set being 0.909 [95% confidence interval (CI): 0.856–0.962] and 0.920 (95% CI:

Table 1 Clinical and CT radiological characteristics of patients with different degrees of malignancy

Characteristics	All (N=117)	LR (N=66)	IHR (N=51)	P value
Gender				0.052
Male	36 (30.77)	15 (22.73)	21 (41.18)	
Female	81 (69.23)	51 (77.27)	30 (58.82)	
Age, years	55.00 (43.00; 65.00)	50.50 (39.00; 59.75)	59.00 (49.50; 67.00)	0.005
Smoking				>0.99
No	110 (94.02)	62 (93.94)	48 (94.12)	
Yes	7 (5.98)	4 (6.06)	3 (5.88)	
Pulmonary disease				0.03
No	113 (96.58)	66 (100.00)	47 (92.16)	
Yes	4 (3.42)	0 (0.00)	4 (7.84)	
Diameter, mm	10.00 (8.00; 16.00)	8.50 (7.25; 10.00)	16.00 (12.50; 21.00)	<0.001
CTR				<0.001
<50%	71 (60.68)	52 (78.79)	19 (37.25)	
≥50%	46 (39.32)	14 (21.21)	32 (62.75)	
Location				>0.99
Central	4 (3.42)	2 (3.03)	2 (3.92)	
Peripheral	113 (96.58)	64 (96.97)	49 (96.08)	
Lobulation				<0.001
No	52 (44.44)	43 (65.15)	9 (17.65)	
Yes	65 (55.56)	23 (34.85)	42 (82.35)	
Irregularity				<0.001
No	48 (41.03)	40 (60.61)	8 (15.69)	
Yes	69 (58.97)	26 (39.39)	43 (84.31)	
Spicule sign				0.18
No	10 (8.55)	8 (12.12)	2 (3.92)	
Yes	107 (91.45)	58 (87.88)	49 (96.08)	
Marginal burring				0.04
No	41 (35.04)	29 (43.94)	12 (23.53)	
Yes	76 (64.96)	37 (56.06)	39 (76.47)	
Pleural indentation sign				<0.001
No	64 (54.70)	46 (69.70)	18 (35.29)	
Yes	53 (45.30)	20 (30.30)	33 (64.71)	
Vascular convergence sign				0.001
No	31 (26.50)	26 (39.39)	5 (9.80)	
Yes	86 (73.50)	40 (60.61)	46 (90.20)	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	All (N=117)	LR (N=66)	IHR (N=51)	P value
Vacuole sign				<0.001
No	55 (47.01)	45 (68.18)	10 (19.61)	
Yes	62 (52.99)	21 (31.82)	41 (80.39)	
Air bronchogram sign				<0.001
No	83 (70.94)	59 (89.39)	24 (47.06)	
Yes	34 (29.06)	7 (10.61)	27 (52.94)	

Categorical variables are presented as n (%) and continuous variables are presented as median (Q1; Q3). CT, computed tomography; CTR, consolidation-to-tumor ratio; IHR, intermediate-to-high-risk group; LR, low-risk group.

Table 2 Univariate and multivariate logistic regression table for clinical-radiological risk factors

Characteristics	Univariable logistic regression			Multivariable logistic regression		
	β	OR (95% CI)	P value	β	OR (95% CI)	P value
Gender	−0.867	0.42 (0.186–0.93)	0.03			
Age	0.035	1.035 (1.008–1.066)	0.01			
Diameter	0.378	1.46 (1.295–1.693)	<0.001	0.388	1.474 (1.294–1.732)	<0.001
CTR ≥0.5	1.833	6.256 (2.816–14.58)	<0.001	1.773	5.887 (2.031–18.85)	0.002
Lobulation sign	2.166	8.725 (3.751–22.08)	<0.001			
Irregularity	2.113	8.269 (3.497–21.57)	<0.001			
Marginal blurring	0.935	2.547 (1.154–5.878)	0.02			
Pleural indentation sign	1.439	4.217 (1.965–9.364)	<0.001			
Vascular convergence sign	1.788	5.98 (2.253–18.99)	0.001			
Vacuole sign	2.173	8.786 (3.825–21.74)	<0.001			
Air bronchogram sign	2.249	9.482 (3.812–26.39)	<0.001			

CI, confidence interval; CTR, consolidation-to-tumor ratio; OR, odds ratio.

0.846–0.994), respectively (*Figure 1*).

$$CM = -5.723 + 0.3882 \times Diameter + 1.7728 \times CTR \qquad [1]$$

The calibration curve and Brier score of CM represent great accuracy with a Brier score of 11.9% (95% CI: 8.1–15.7%) in training set and with a score of 11.4% (95% CI: 5.9–17.0%) in testing set (*Figure 2*).

Development and assessment of the RM

Based on the mRMR algorithm, a subset of 10 radiomic features demonstrating the highest correlation with pathological subtypes and the lowest degree of inter-feature correlation were identified from a pool of 851 radiomic features. Subsequently, LASSO regression with 10-fold cross-validation was applied to further refine this subset to seven features and RM was built (see *Table 3* for feature selection and *Figures 3-5* for feature screening and model validation). The AUCs for the training and testing sets were 0.961 (95% CI: 0.926–0.996) and 0.957 (95% CI: 0.905–1.000) (see *Figure 6*), respectively. The DeLong test revealed that RM outperformed CM in the training set (P=0.04); however, there is no significant difference in AUC between RM and CM in the testing set (P=0.15). The calibration curves and Brier score of RM show great accuracy with a brier score of 6.6% (95% CI: 3.3–9.8%)

in training set and 9.1% (95% CI: 3.6–14.6%) in testing set (see *Figure* 7). Based on the coefficients derived from LASSO, the calculation formula for Rad score and RM is as follows:

$$Rad\ score = -9.3144607379 + 0.000381896 \times Maximum.1$$
$$- 2.6627831219 \times LargeDependenceLowGrayLevelEmphasis.8$$
$$+ 0.0175793888 \times Maximum2DDiameterRow \quad [2]$$
$$+ 0.7730640440 \times RunEntropy.8 + 0.0281735245 \times JointAverage.7$$
$$- 2.6943239782 \times Sphericity + 0.5648530023 \times JointEntropy.8$$
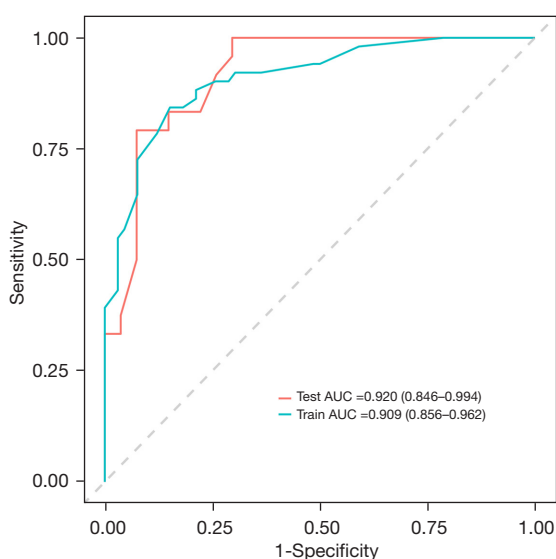
Figure 1 ROC curve of the training set and testing set of the CM. Values in brackets indicate 95% CI. AUC, area under the curve; CI, confidence interval; CM, clinical-radiological model; ROC, receiver operating characteristic.

$$RM = 0.1086 + 1.2910 \times Rad\ score \quad [3]$$

Development and validation of multiple ML COMs

The Rad score, along with two risk factors among clinical-radiological features, was used to create and validate COMs via algorithms including LR, DT, KNN, NBM, RF, SVM, and XGBoost. *Figures 8,9* present the performance metrics (including sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, F1 score, and performance radar chart) and AUC curves for these models in the training and testing sets. The AUC values are listed in *Table 4*, and the results of DeLong test for AUC comparison are presented in *Table 5*.

For the LR model, the NRI was used to compare the performance difference between LR and single CM and RM models. The findings indicated that the LR model enhanced the predictive capability of the CM model in training set, with the integrated discrimination improvement (IDI) being 0.207 (95% CI: 0.129–0.284) (P<0.001) for the training set and 0.099 (95% CI: −0.017–0.2146) (P=0.09) for the testing set. The LR model showed no significant change in performance compared to RM in the training set [IDI: 0.004 (95% CI: −0.011–0.020), P=0.59] but demonstrated significant improvement in the testing set [IDI: 0.101 (95% CI: 0.010–0.192), P=0.03].

Comparisons of all COMs models in the training set showed that RF, and KNN exhibited high AUCs, significantly outperforming DT, LR, SVM, XGBoost and NBM (all P<0.05). However, RF and KNN had low AUCs
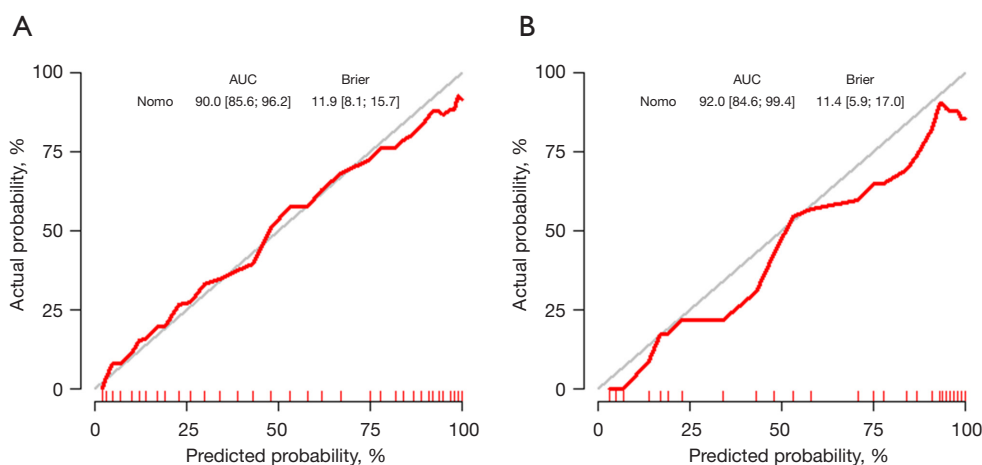

Figure 2 The calibration curve and Brier score of CM. (A) Calibration curve of training set; (B) calibration curve of testing set. Values in square brackets indicate 95% CI. AUC, area under the curve; CI, confidence interval; CM, clinical-radiological model.

Table 3 The 10 best features after mRMR and their meanings

Features	Meanings
Maximum2DDiameterRow	The maximum 2D diameter of the target region in the image in the row direction
Sphericity	Used to measure the similarity between the shape of the ROI and a sphere
Maximum	The maximum value of a certain second-order texture feature statistic
LargeDependenceLowGrayLevelEmphasis	The emphasis degree of pixel pairs with large dependence and low gray levels in the gray-level co-occurrence matrix
RunEntropy	Used to measure the disorder degree of the gray-level run length distribution in the image
MCC	Measure the linear correlation of gray values in the image
JointAverage	Used to measure the average situation of the joint distribution of two or more image features
Skewness	It describes the skewness of the image gray-level value distribution, reflecting the asymmetry of the data distribution
SmallDependenceEmphasis	It represents the emphasis degree of pixel pairs with small dependence in the gray-level co-occurrence matrix
JointEntropy.8	Used to measure the uncertainty of the joint distribution of two or more image features

mRMR, Max-Relevance and Min-Redundancy; ROI, region of interest.
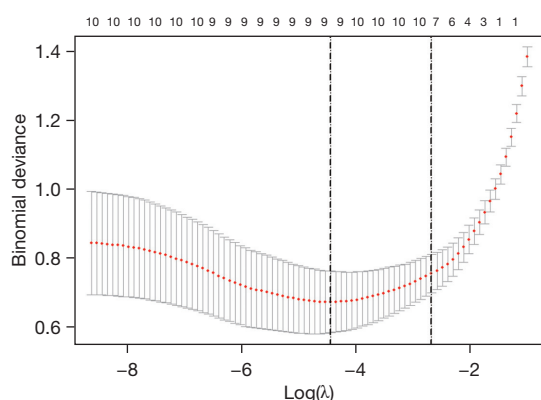
Figure 3 LASSO image feature screening of RM. LASSO, least absolute shrinkage and selection operator; RM, radiomics model.

in the testing set, with RF significantly underperforming XGBoost (P=0.04) and SVM (P=0.02) and showing no significant difference from other models (all P>0.05). Meanwhile, DT shows the lowest AUC in both training set and the testing set. Excluding RF, KNN and DT, the AUC of the XGBoost is the highest both in the training set and the testing set. Moreover, in both training set and testing set, XGBoost has the best performance in terms of sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, and F1 score among all the COMs. Hence, XGBoost was chosen as the optimal COM.

The performance of XGBoost model peaked at iteration 19, with feature importance values for gain, cover, and frequency detailed in *Table 6*. Rad score was found to be the predominant influencing factor.

To better explain the performance of the XGBoost model, we computed the SHAP values for the collective and individual features, thereby ascertaining their respective contributions to the models. A visual representation of these values is provided in *Figure 10*, which indicates the importance of three key features—diameter, CTR, and Rad score—within the models. Each dot in the SHAP beeswarm plot represents the eigenvalue of each case, with dark purple indicating low values and yellow indicating high values. SHAP values at the specific dot show the feature's impact on prediction probability. We also found that all three features had positive impacts on the prediction of LUAD malignancy grade. The SHAP partial dependence plot (PDP) was used to visualize the eigenvalues in each case and their impact on predictions, which confirmed that all three features were positive predictors, with higher eigenvalues correlating with predictions of more malignant subtypes.

Discussion

Summary of key findings

In the present study, we leveraged clinical and radiological semantic feature data to create a CM for predicting the
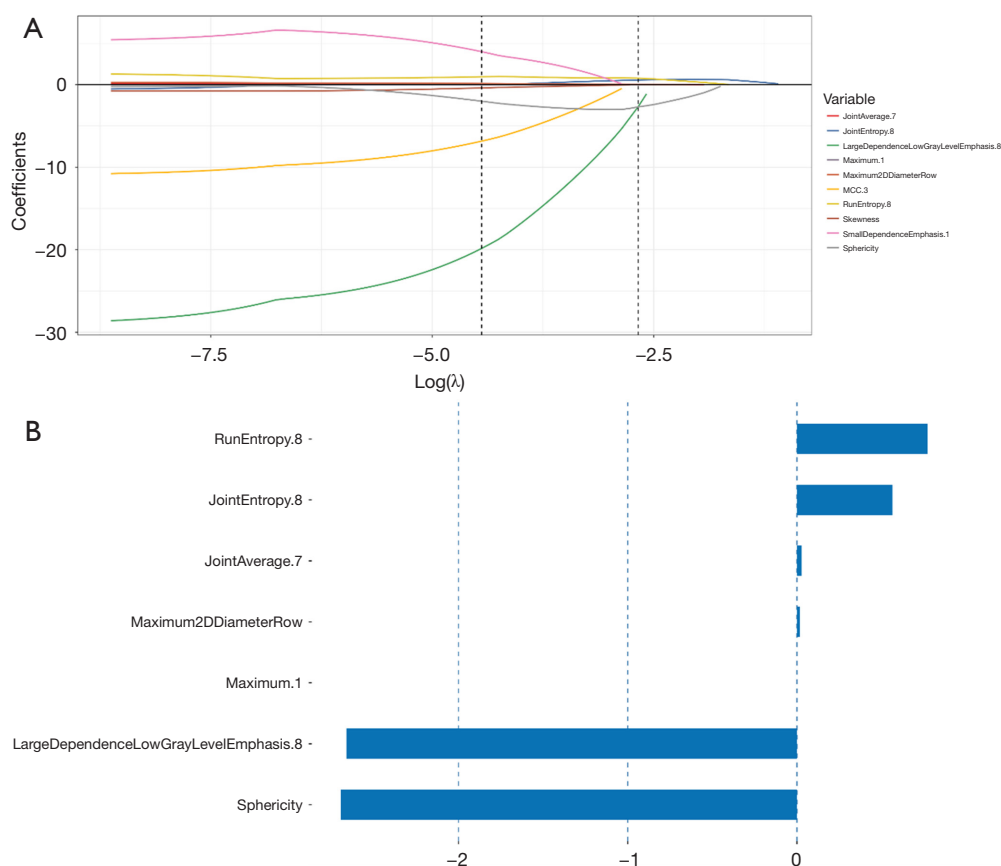
Figure 4 The process of feature selection and the corresponding feature coefficients in RM. (A) Coefficient paths of variables in a LASSO model as a function of log (λ). (B) Radiomics feature coefficient plot. LASSO, least absolute shrinkage and selection operator; RM, radiomics model.

malignancy grade of LUAD. The analysis revealed diameter (P<0.001), and CTR ≥0.5 (P=0.002) as independent risk factors. However, the CM exhibited moderate performance, with an AUC of 0.909 (95% CI: 0.856–0.962) for the training set and 0.920 (95% CI: 0.846–0.994) for the testing set. The calibration curve and Brier score of CM represent great accuracy. Subsequently, we used radiomics and ML techniques to integrate radiological and clinical features into more robust predictive models for LUAD subtyping. These COMs created using multiple ML methods demonstrated improved AUCs when compared with single CM or RM. XGBoost was selected as the optimal prediction model, which had an AUC of 0.966 (95% CI: 0.933–0.999) in the training set and 0.975 (95% CI: 0.943–1.000) in the testing set. In the testing set, XGBoost had performance metrics as follows: sensitivity 1.00, specificity 0.78, positive predictive value 0.80, negative predictive value 1.00, precision 0.80,

recall 1.00, and F1 score 0.89. SHAP was utilized to visualize the impacts of different features on predictions. Our study confirmed that ML-based COMs integrating radiomics and clinical-radiological features could reliably predict the feasibility of wedge resection for an LUAD and guiding surgical decision-making, with XGBoost demonstrating optimal performance.

Comparison to existing literature

The IASLC subtype is a proven independent predictor of NSCLC. Zhang *et al.* demonstrated that LCA patients had the best prognosis, followed by ACA and PPA patients, and those with SPA or MPA had the poorest outcomes (all P<0.01) (26). Zombori *et al.* (27) also observed a favorable prognosis for LCA, yet no significant difference in OS or DFS was noted between tumors with lepidic component
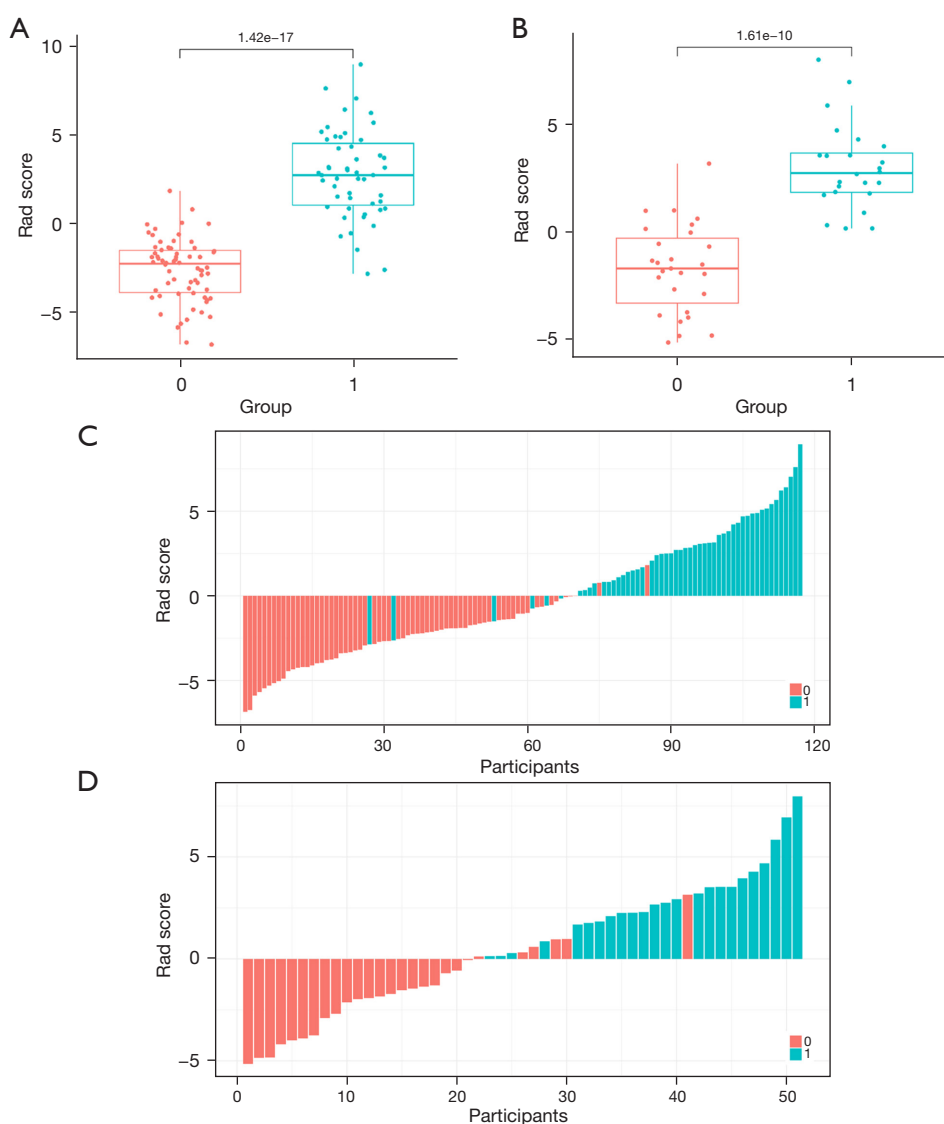
Figure 5 Distribution of Rad score of RM. (A,B) Rad score values between Group 0 (LR) and Group 1 (IHR) in training set and testing set; (C,D) Rad score waterfall plot in training and testing sets. IHR, intermediate-to-high-risk group; LR, low-risk group; RM, radiomics model.

as the second principal pattern and those without lepidic pattern. Xu *et al.* (28) revealed significantly reduced long-term survival in patients with over 5% micropapillary pattern compared to those with 5% or less.

Radical lobectomy remains the gold standard treatment for resectable NSCLC. However, the Japanese JCOG0802 study (29) showed that the decrement in pulmonary function following lobectomy might compromise OS due to the onset of complications: although patients who underwent segmentectomy experienced a higher rate of

local recurrence compared to those who received lobectomy (10.5% *vs.* 5.4%; P=0.0018), the proportion of deaths attributed to causes other than NSCLC was lower in the segmentectomy group than in the lobectomy group (47% *vs.* 63%). Thus, sublobar resection might be a preferred procedure for small peripheral NSCLC. However, few studies have investigated which LUAD subtypes within peripheral lung cancers with different patterns may be more suitably treated with wedge resection. Nitadori *et al.* (30) linked a higher proportion of high-risk components (e.g.,

MPA and SPA) to a greater recurrence risk in wedge-resected lung cancer patients. Song *et al.* (10) demonstrated equivalent prognoses between wedge resection and anatomic excision in patients with LCA or PPA.

Radiomics, a widely used quantitative tool, serves as a reliable clinical diagnostic aid, particularly for lung nodules. By extracting lesion information (including data related to or complementary to pathology, hematology, and genomics) from radiological images, it can reveal cellular-level tumor heterogeneity and inform treatment (31). Prior research has

predominantly concentrated on the application of radiomics for the prediction of benign versus malignant pulmonary nodules (32), tumor staging, tumor genotyping (33), as well as clinical outcomes and prognosis. Additionally, a limited number of studies have integrated ML or artificial intelligence (AI) methodologies. However, there is a paucity of research that employs these techniques to inform surgical strategies for LUAD. In our present study, we developed models integrating CT radiomics and clinical-radiological features to guide surgery, enhanced by ML and finally found that the XGBoost algorithm significantly improved model performance, which aligned with findings from prior research (34,35). Among the 10 radiomic features screened in our current study, one was from the First Order Features category, two from the Shape Features (3D) domain, and seven from the Second-Order GLCM Features and Gray Level Dependence Matrix (GLDM) Features. The specific definitions of these features are listed in *Table 3*. It was found the clinical-radiomic LR model outperformed individual clinical and radiomic models, and the XGBoost model, which was created using ML algorithm, further enhanced LR performance with an NRI, making it a reliable tool for surgical method selection.
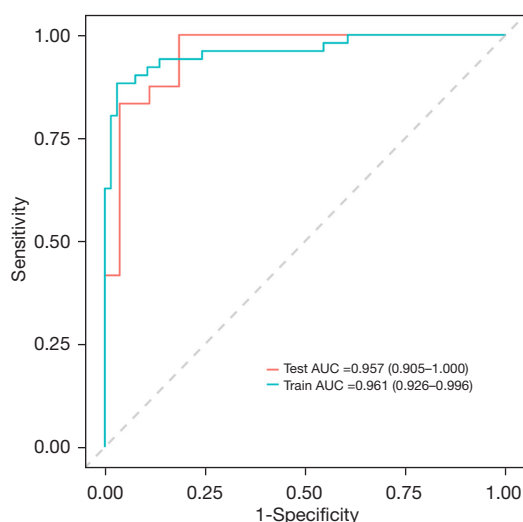

Figure 6 ROC curves of RM for the training and testing sets. Values in brackets indicate 95% CI. AUC, area under the curve; RM, radiomic model; ROC, receiver operating characteristic.

Future clinical implications, limitations and research directions

Specifically, all patients with suspected LUAD detected by CT can have their imaging data included in the model to calculate the malignancy degree. For patients with LUAD
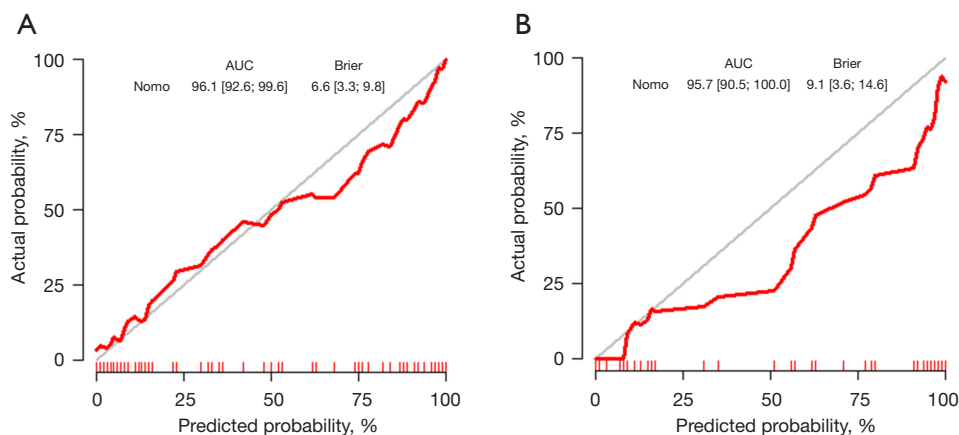

Figure 7 Calibration curves and Brier score of RM. (A,B) Calibration curves of RM in training and testing sets. Values in square brackets indicate 95% CI. AUC, area under curve; CI, confidence interval; RM, radiomic model.
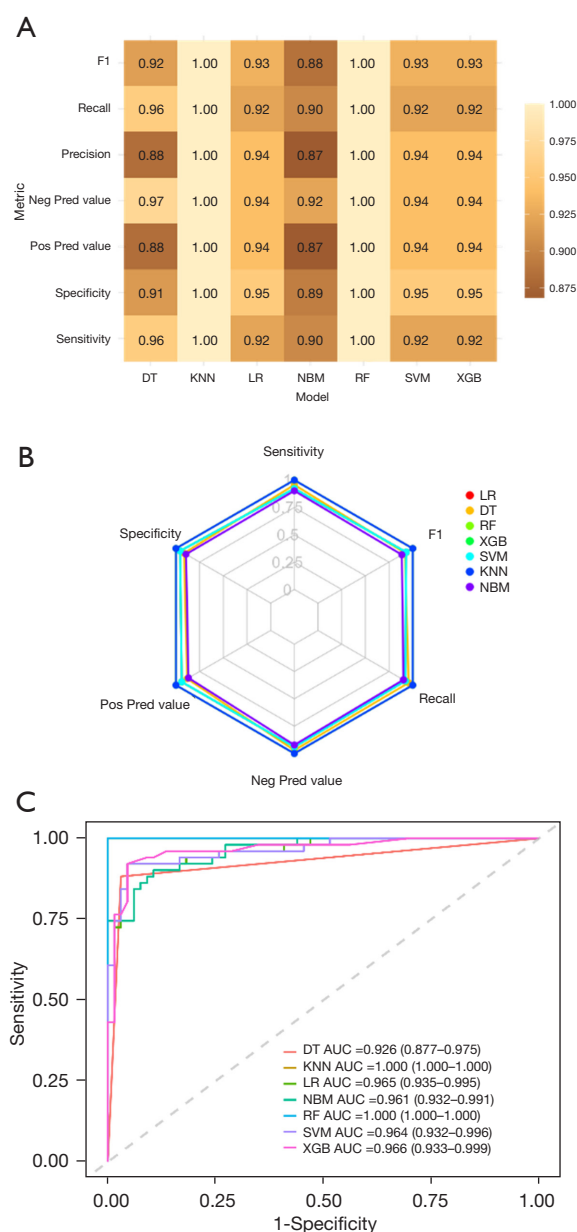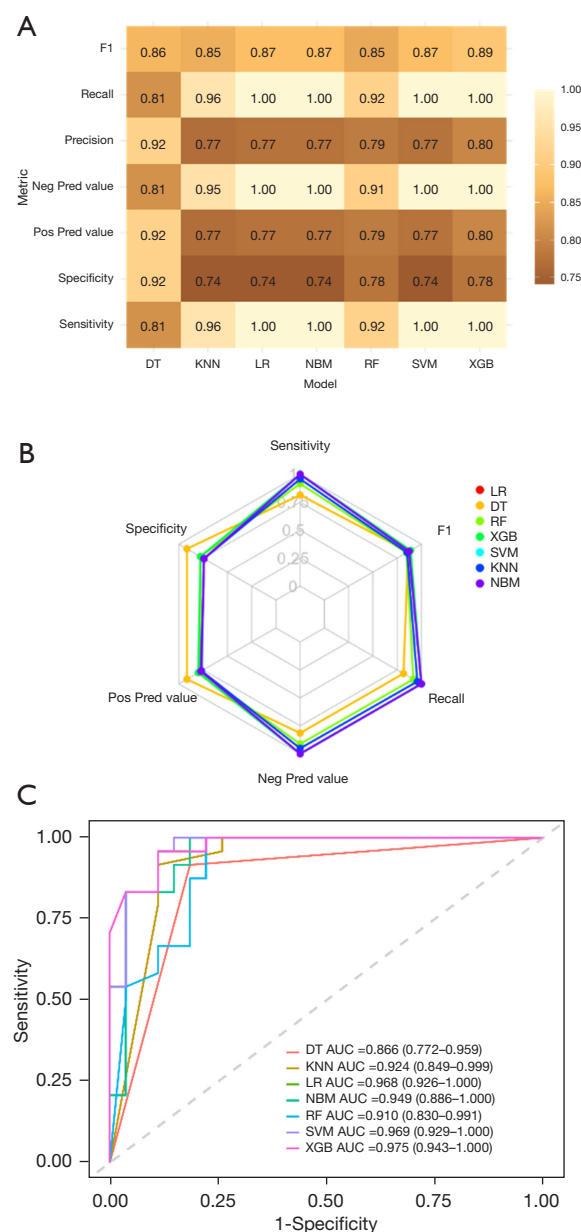
Figure 8 Comparison of the performances of various ML algorithms in training set. (A) Heatmaps of the performance of various COMs in training set. (B) The radar chart visualization of prediction performance. (C) ROC curves for different models in training set. Values in brackets indicate 95% CI. AUC, area under the curve; CI, confidence interval; COMs, comprehensive models; DT, decision tree; KNN, K-nearest neighbors; LR, logistic regression; ML, machine learning; NBM, naïve Bayes model; RF, random forest; ROC, receiver operating characteristic; SVM, support vector machine; XGB, extreme gradient boosting.

Figure 9 Comparison of the performances of various ML algorithms in testing set. (A) Heatmaps of the performance of various COMs in testing set. (B) The radar chart visualization of prediction performance. (C) ROC curves for different models in testing set. Values in brackets indicate 95% CI. AUC, area under the curve; CI, confidence interval; COMs, comprehensive models; DT, decision tree; KNN, K-nearest neighbors; LR, logistic regression; ML, machine learning; NBM, naïve Bayes model; RF, random forest; ROC, receiver operating characteristic; SVM, support vector machine; XGB, extreme gradient boosting.

Table 4 AUC for different models in the training set and the testing set

Model	AUC	95% CI
Training set		
CM	0.909	0.856–0.962
RM	0.961	0.926–0.996
LR	0.965	0.935–0.995
DT	0.926	0.877–0.975
RF	1.000	1.000–1.000
XGB	0.966	0.933–0.999
SVM	0.964	0.932–0.996
KNN	1.000	1.000–1.000
NBM	0.961	0.932–0.991
Testing set		
CM	0.920	0.846–0.994
RM	0.957	0.905–1.000
LR	0.968	0.926–1.000
DT	0.866	0.772–0.959
RF	0.910	0.830–0.991
XGB	0.975	0.943–1.000
SVM	0.969	0.929–1.000
KNN	0.924	0.849–0.999
NBM	0.949	0.886–1.000

AUC, area under the curve; CI, confidence interval; CM, comprehensive model; DT, decision tree; KNN, K-nearest neighbors; LR, logistic regression; NBM, naïve Bayes model; RF, random forest; RM, radiomic model; SVM, support vector machine; XGB, extreme gradient boosting.

predicted to have low malignancy (LR) by the XGBoost model, such as those with postoperative pathology showing AIS, MIA, or LCA with ≤20% high-grade patterns, pulmonary wedge resection can be performed. This is because for such patients, wedge resection can achieve prognoses comparable to more extensive resections and better perioperative outcomes, while minimizing lung function loss. For patients predicted to have IHR by the XGBoost model, such as those with ACA, PPA, or >20% high-grade patterns (solid, micropapillary, or complex

gland), segmentectomy/lobectomy plus lymph node dissection should be performed to obtain better survival benefits. In this way, by using the model's prediction results, doctors can develop more precise and personalized surgical plans for LUAD patients with different malignancy degrees, thereby improving the prognosis of patients.

Our study had several limitations: first, its retrospective design would have introduced selection bias. Second, it was a single-center study with a limited sample size. Although the training set and testing set were used in this study and 10-fold cross-validation was applied, multi-center external validation for model robustness is warranted. Third, ROIs were manually delineated, which compromised repeatability and accuracy. Huang *et al.* (36) analyzed pre-treatment FDG-PET/CT scan data using ML to predict lung cancer progression and OS. The results showed that the accuracy and sensitivity of the CT automatic segmentation model were significantly higher than those of the manual segmentation model. The sensitivity of the PET manual segmentation model was significantly higher than that of the automatic segmentation model, while there was no significant difference in the performance of the PET CT ensemble model between manual and automatic segmentations. We are also currently exploring automatic segmentation and DL-based alternatives for nodule delineation to assess reproducibility, and the results of this comparison will be reported in future studies. Fourth, although the COMs were created using clinical data, radiological semantic features, and radiomic features, it was based on a simple information fusion method, necessitating the development and use of better feature fusion techniques. Finally, the feasibility of pulmonary wedge resection can be influenced by a variety of factors beyond just pathological subtype; the distance of the nodule from the visceral pleura, the pulmonary lobe where the nodule is located, and the physical performance of the patient are all concerned before developing a surgical plan.

Conclusions

Focusing on the common condition in clinical settings, we employed CT radiomics and ML to mine information on lesions on preoperative chest CT images, aiming to predict optimal candidates for lung wedge resection. The COM created using the XGBoost ML algorithm was found to have strong diagnostic performance and

Table 5 P value of DeLong test of AUC for different models in the training set and the testing set

Model	DT	LR	RF	XGB	SVM	KNN	NBM
Training set							
DT	>0.99	0.18	<0.01	0.03	0.02	<0.01	0.04
LR	0.18	>0.99	0.03	0.98	0.96	0.03	0.86
RF	<0.01	0.03	>0.99	0.04	0.03	>0.99	0.01
XGB	0.03	0.98	0.04	>0.99	0.90	0.04	0.64
SVM	0.02	0.96	0.03	0.90	>0.99	0.03	0.79
KNN	<0.01	0.03	>0.99	0.04	0.03	>0.99	0.01
NBM	0.04	0.86	0.01	0.64	0.79	0.01	>0.99
Testing set							
DT	>0.99	0.06	0.12	0.01	0.01	0.15	0.06
LR	0.06	>0.99	0.22	0.80	0.96	0.32	0.81
RF	0.12	0.22	>0.99	0.04	0.02	0.59	0.25
XGB	0.01	0.80	0.04	>0.99	0.59	0.07	0.55
SVM	0.01	0.96	0.02	0.59	>0.99	0.06	0.76
KNN	0.15	0.32	0.59	0.07	0.06	>0.99	0.36
NBM	0.06	0.81	0.25	0.55	0.76	0.36	>0.99

AUC, area under the curve; DT, decision tree; KNN, K-nearest neighbors; LR, logistic regression; NBM, naïve Bayes model; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting.

Table 6 The influence values of different features in the XGBoost model

Feature	Gain	Cover	Frequency
Rad score	0.5921357	0.4176745	0.4730159
Diameter	0.2724704	0.2997670	0.3269841
CTR ≥0.5	0.1353940	0.2825584	0.2000000

Gain: the degree of contribution of features to the improvement of model accuracy in the process of tree construction. Cover: the number or range of samples covered by the features. Frequency: the frequency with which features are used when constructing trees. CTR, consolidation-to-tumor ratio; XGBoost, extreme gradient boosting.

certain clinical application potential. Wedge resection in patients with LUAD deemed low-grade malignant by the XGBoost model could achieve comparable prognoses and better perioperative outcomes, along with minimal lung function loss, when compared with anatomical resections. Furthermore, patients with high-grade malignancy predicted by the XGBoost model should receive segmentectomy/lobectomy plus lymph node dissection for better survival benefits. Despite limitations, the XGBoost model offers a basis for auxiliary diagnosis. The integration of ML and radiomics is viable, offering new directions for diagnostic model development.
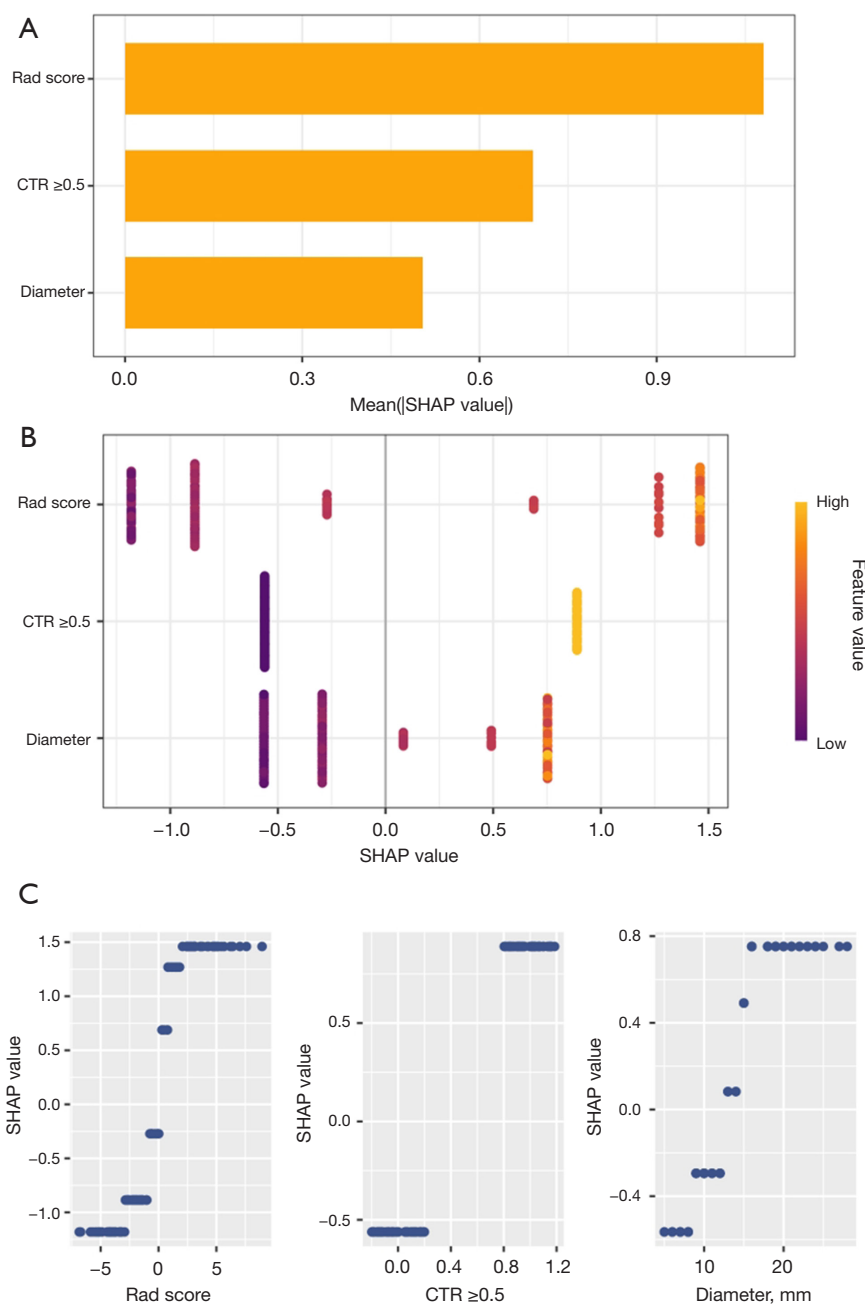
Figure 10 SHAP visualization of the prediction process of the model. (A) The mean SHAP values of each feature. (B) The SHAP bee swarm plot illustrates the positive or negative impacts of each feature on the prediction probability by means of purple and yellow colors. (C) SHAP partial dependence plot of the features in XGBoost model. CTR, consolidation-to-tumor ratio; SHAP, Shapley additive explanation; XGBoost, extreme gradient boosting.

References

1. Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin Med J (Engl) 2022;135:584-90.

2. Gao F, Li M, Zhang Z, et al. Morphological classification of pre-invasive lesions and early-stage lung adenocarcinoma based on CT images. Eur Radiol 2019;29:5423-30.

3. Moreira AL, Ocampo PSS, Xia Y, et al. A Grading System for Invasive Pulmonary Adenocarcinoma: A Proposal From the International Association for the Study of Lung Cancer Pathology Committee. J Thorac Oncol 2020;15:1599-610.

4. Willner J, Narula N, Moreira AL. Updates on lung adenocarcinoma: invasive size, grading and STAS. Histopathology 2024;84:6-17.

5. Fujikawa R, Muraoka Y, Kashima J, et al. Clinicopathologic and Genotypic Features of Lung Adenocarcinoma Characterized by the International Association for the Study of Lung Cancer Grading System. J Thorac Oncol 2022;17:700-7.

6. Okubo Y, Kashima J, Teishikata T, et al. Prognostic Impact of the Histologic Lepidic Component in Pathologic Stage IA Adenocarcinoma. J Thorac Oncol 2022;17:67-75.

7. Jeon HW, Kim YD, Sim SB, et al. Significant difference in recurrence according to the proportion of high grade patterns in stage IA lung adenocarcinoma. Thorac Cancer 2021;12:1952-8.

8. Mikubo M, Tamagawa S, Kondo Y, et al. Micropapillary and solid components as high-grade patterns in IASLC grading system of lung adenocarcinoma: Clinical implications and management. Lung Cancer 2024;187:107445.

9. Tan KS, Reiner A, Emoto K, et al. Novel Insights Into the International Association for the Study of Lung Cancer Grading System for Lung Adenocarcinoma. Mod Pathol 2024;37:100520.

10. Song W, Hou Y, Zhang J, et al. Comparison of outcomes following lobectomy, segmentectomy, and wedge resection based on pathological subtyping in patients with pN0 invasive lung adenocarcinoma ≤1 cm. Cancer Med 2022;11:4784-95.

11. Altorki N, Wang X, Kozono D, et al. Lobar or Sublobar Resection for Peripheral Stage IA Non-Small-Cell Lung Cancer. N Engl J Med 2023;388:489-98.

12. Chiang XH, Lu TP, Hsieh MS, et al. Thoracoscopic Wedge Resection Versus Segmentectomy for cT1N0 Lung Adenocarcinoma. Ann Surg Oncol 2021;28:8398-411.

13. McGuire AL, Hopman WM, Petsikas D, et al. Outcomes: wedge resection versus lobectomy for non-small cell lung cancer at the Cancer Centre of Southeastern Ontario 1998-2009. Can J Surg 2013;56:E165-70.

14. Kodama H, Takaki H, Taniguchi J, et al. Efficacy of Percutaneous Direct Puncture Biopsy of Malignant Lung Tumors Contacting to the Pleura. In Vivo 2023;37:2237-43.

15. Bourgouin PP, Rodriguez KJ, Fintelmann FJ. Image-Guided Percutaneous Lung Needle Biopsy: How we do it. Tech Vasc Interv Radiol 2021;24:100770.

16. Kramer T, Annema JT. Advanced bronchoscopic techniques for the diagnosis and treatment of peripheral lung cancer. Lung Cancer 2021;161:152-62.

17. Yeh YC, Nitadori J, Kadota K, et al. Using frozen section to identify histological patterns in stage I lung adenocarcinoma of ≤ 3 cm: accuracy and interobserver agreement. Histopathology 2015;66:922-38.

18. Fu Z, Shen X, Deng C, et al. Prediction of the pathological subtypes by intraoperative frozen section for patients with cT1N0M0 invasive lung adenocarcinoma (ECTOP-1015): a prospective multicenter study. Int J Surg 2024;110:5444-51.

19. Choi ER, Lee HY, Jeong JY, et al. Quantitative image variables reflect the intratumoral pathologic heterogeneity of lung adenocarcinoma. Oncotarget 2016;7:67302-13.

20. Choi RY, Coyner AS, Kalpathy-Cramer J, et al. Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol 2020;9:14.

21. Sultan AS, Elgharib MA, Tavares T, et al. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. J Oral Pathol Med 2020;49:849-56.

22. Adams SJ, Mikhael P, Wohlwend J, et al. Artificial Intelligence and Machine Learning in Lung Cancer Screening. Thorac Surg Clin 2023;33:401-9.

23. Sandino CM, Cole EK, Alkan C, et al. Upstream Machine Learning in Radiology. Radiol Clin North Am 2021;59:967-85.

24. Liu Z, Yang L, Liang J, et al. Radiomic features add incremental benefit to conventional radiological feature-based differential diagnosis of lung nodules. Eur Radiol 2024. [Epub ahead of print]. doi: 10.1007/s00330-024-11221-5.

25. Wu G, Jochems A, Refaee T, et al. Structural and functional radiomics for lung cancer. Eur J Nucl Med Mol Imaging 2021;48:3961-74.

26. Zhang H, Sun FH, Chen ZC, Wang Q. Validation of prognostic value of pathological staging in pathological stage I lung adenocarcinoma. Zhonghua Wai Ke Za Zhi 2022;60:580-6.

27. Zombori T, Nyári T, Tiszlavicz L, et al. The more the

28. micropapillary pattern in stage I lung adenocarcinoma, the worse the prognosis-a retrospective study on digitalized slides. Virchows Arch 2018;472:949-58.

28. Xu L, Zhou H, Wang G, et al. The prognostic influence of histological subtypes of micropapillary tumors on patients with lung adenocarcinoma ≤ 2 cm. Front Oncol 2022;12:954317.

29. Saji H, Okada M, Tsuboi M, et al. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. Lancet 2022;399:1607-17.

30. Nitadori J, Bograd AJ, Kadota K, et al. Impact of micropapillary histologic subtype in selecting limited resection vs lobectomy for lung adenocarcinoma of 2cm or smaller. J Natl Cancer Inst 2013;105:1212-20.

31. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J Nucl Med 2020;61:488-95.

32. Zhang Y, Feng W, Wu Z, et al. Deep-Learning Model of ResNet Combined with CBAM for Malignant-Benign Pulmonary Nodules Classification on Computed Tomography Images. Medicina (Kaunas) 2023;59:1088.

33. Kirienko M, Sollini M, Corbetta M, et al. Radiomics and gene expression profile to characterise the disease and predict outcome in patients with lung cancer. Eur J Nucl Med Mol Imaging 2021;48:3643-55.

34. Le NQK, Kha QH, Nguyen VH, et al. Machine Learning-Based Radiomics Signatures for EGFR and KRAS Mutations Prediction in Non-Small-Cell Lung Cancer. Int J Mol Sci 2021;22:9254.

35. Liu Z, Luo C, Chen X, et al. Noninvasive prediction of perineural invasion in intrahepatic cholangiocarcinoma by clinicoradiological features and computed tomography radiomics based on interpretable machine learning: a multicenter cohort study. Int J Surg 2024;110:1039-51.

36. Huang B, Sollee J, Luo YH, et al. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT. EBioMedicine 2022;82:104127.