



# Dual-layer spectral detector computed tomography multiparameter machine learning model for prediction of invasive lung adenocarcinoma

Jiayu Wan<sup>1,2,3#</sup>, Xue Lin<sup>1,2,3#</sup>, Zhaokai Wang<sup>4#</sup>, Peng Sun<sup>5</sup>, Shen Gui<sup>5</sup>, Tianhe Ye<sup>1,2,3</sup>, Qianqian Fan<sup>1,2,3</sup>, Weiwei Liu<sup>1,2,3</sup>, Feng Pan<sup>1,2,3</sup>, Bo Yang<sup>1,2,3</sup>, Xiaotong Geng<sup>1,2,3</sup>, Zhen Quan<sup>1,2,3</sup>, Lian Yang<sup>1,2,3</sup>

<sup>1</sup>Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; <sup>2</sup>Hubei Provincial Clinical Research Center for Precision Radiology & Interventional Medicine, Wuhan, China; <sup>3</sup>Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China; <sup>4</sup>Department of Thoracic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; <sup>5</sup>MSC Clinical & Technical Solutions, Philips Healthcare, Wuhan, China

**Contributions:** (I) Conception and design: J Wan, Z Quan, L Yang; (II) Administrative support: Z Quan, L Yang; (III) Provision of study materials or patients: J Wan, X Lin, Z Wang; (IV) Collection and assembly of data: T Ye, Q Fan, W Liu, B Yang, X Geng; (V) Data analysis and interpretation: J Wan, P Sun, S Gui; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Lian Yang, MD, PhD; Zhen Quan, MD, PhD. Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Jiefang Avenue #1277, Wuhan 430022, China; Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China; Hubei Provincial Clinical Research Center for Precision Radiology & Interventional Medicine, Wuhan, China. Email: yanglian@hust.edu.cn; 2022xh0063@hust.edu.cn.

**Background:** Lung adenocarcinoma (LUAD) is the leading cause of cancer-related deaths. High-resolution computed tomography (HRCT) has improved the detection of ground glass nodules (GGNs), which are early indicators of lung cancer. Accurate assessment of GGN invasiveness is crucial for determining the appropriate surgical approach. Dual-layer spectral detector computed tomography (DLCT) offers advanced imaging capabilities, including electron density and iodine density, which enhance the evaluation of GGN invasiveness. This study aims to develop a machine learning (ML) model that integrates DLCT parameters and clinical features to predict the invasiveness of GGNs in LUAD, aiding in surgical decision-making and prognosis improvement.

**Methods:** The retrospective study encompassed 272 patients who were diagnosed with LUAD, comprising 154 cases of invasive adenocarcinomas (IA) and 118 cases of pre-invasive minimally invasive adenocarcinoma (MIA) which were then randomly allocated into a training set and a test set. Six ML models were developed based on five DLCT parameters (conventional, iodine density, virtual noncontrast, electron density, and effective atomic number). Subsequently, a nomogram was constructed using multi-factor logistic regression, incorporating radiomic characteristics and clinicopathological risk factors.

**Results:** The ML model based on conventional plus electron density performed better than the models with other DLCT parameters, with the area under the curves (AUCs) of 0.945 and 0.964 in the training and test sets, respectively. The clinical model and radiomics score (Rad-score) were combined in the logistic regression to construct a joint model, of which the AUCs were 0.974 in the training sets and 0.949 in the test sets. The ML model effectively differentiated between IA and pre-invasive MIA, and further classified patients into high and medium risk categories for invasion using waterfall plots.

**Conclusions:** The ML model based on DLCT parameters helps predict the invasiveness of GGNs and classifies the GGNs into different risk grades.

**Keywords:** Dual-layer spectral detector computed tomography (DLCT); radiomics; machine learning (ML); ground-glass nodules (GGNs); adenocarcinoma

Submitted Sep 10, 2024. Accepted for publication Jan 09, 2025. Published online Feb 27, 2025.

doi: 10.21037/tlcr-24-822

View this article at: <https://dx.doi.org/10.21037/tlcr-24-822>

## Introduction

According to cancer statistics in 2022, the incidence of lung cancer has become the leading cause of death in cancer cases, and its predominant histological type is adenocarcinoma (1,2). The utilization of high-resolution computed tomography (HRCT) has resulted in an enhanced detection rate of ground glass nodules (GGNs), which are ultimately histologically diagnosed as early-stage lung cancer (3,4). The GGNs can be classified into two types: pure ground glass nodules (pGGNs) without solid components and mixed ground glass nodules (mGGNs) with solid components, with the nodules' malignancy degree increasing with the proportion of solid components (5,6). The presence of a GGN on HRCT typically suggests the presence of lung adenocarcinoma (LUAD) or its precursors, including invasive adenocarcinomas (IA), minimally invasive adenocarcinoma (MIA), adenocarcinoma *in situ* (AIS), and atypical adenomatous hyperplasia (AAH) (7). The degree of invasion exhibited by mGGNs is higher than that observed

in pGGNs in LUAD with GGN (8). AIS, AAH and MIA are considered as preinvasive-MIA. Preinvasive-MIA and IA show significant differences in terms of surgical approaches, postoperative treatments and prognoses. To optimize the preservation of functional lung parenchyma, limited wedge resection or segmentectomy is commonly employed for preinvasive-MIA, while the standard recommendation for IA is lobectomy, which effectively reduces the risk of tumor recurrence (9,10). Therefore, early and accurate identification of the invasiveness of GGNs is crucial for surgical decision-making, enabling the selection of an appropriate surgical approach and improving prognosis. The predictive value of GGNs morphology on conventional computed tomography (CT) and multiple quantitative CT features in determining the invasiveness of GGNs has been extensively investigated by many researchers (11,12). However, its efficacy is constrained in GGNs with scant or negligible solid components (13).

Recently, the utilization of dual-layer spectral detector computed tomography (DLCT), as a promising non-invasive imaging modality, has significantly enhanced the capabilities of conventional CT in clinical practices, offering more parameters such as virtual monoenergetic imaging (VMI), iodine density (ID), effective atomic number (Zeff), electron density (ED) and virtual non-contrast (VNC) (14). Among them, ED imaging enables the generation of electron density-based images that are not achievable with conventional CT, thereby providing a novel avenue for clinicians to enhance the diagnostic process of diseases. Previous studies have indicated the efficacy of DLCT parameters in the diagnosis of invasiveness in GGNs (15,16).

In the field of artificial intelligence (AI), radiomics is used to characterize lung lesions by extracting a variety of predefined high-throughput features using computer software, followed by statistical methods to identify the most relevant ones. Ultimately, machine learning (ML) techniques are utilized to establish diagnostic and predictive models. Although the potential of radiomics in determining GGN invasiveness based on DLCT has been demonstrated by several studies, these studies focused less on further classification of invasive pulmonary nodules and enrolled only a small number of patients (17,18). Accordingly, the prediction of the invasiveness of GGNs needs further

### Highlight box

#### Key findings

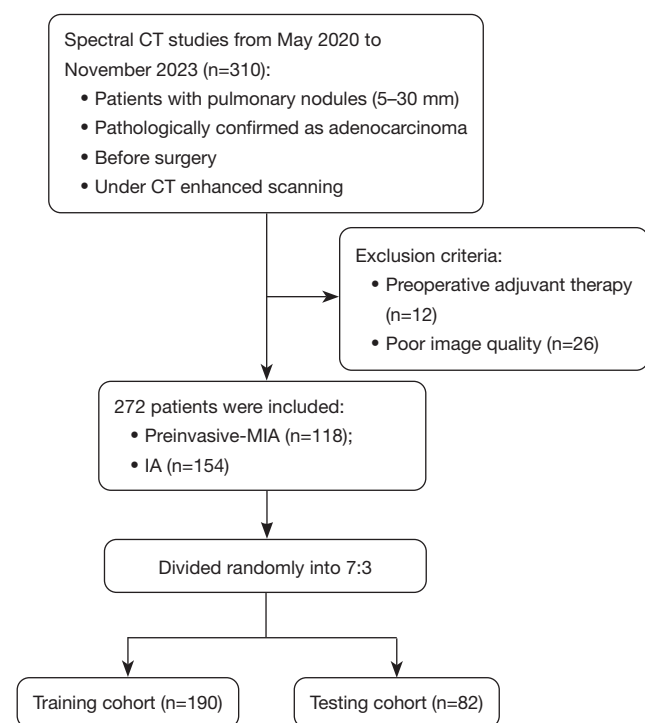
- We developed a machine learning model based on dual-layer spectral detector computed tomography (DLCT) parameters to predict the invasiveness of ground-glass nodules (GGNs) and further classify the GGNs into pre-invasive minimally invasive adenocarcinoma, medium, and high risk of invasive adenocarcinomas.

#### What is known and what is new?

- The current research efforts primarily focus on predicting the invasiveness of lung adenocarcinoma, with only a limited number of multi-class studies conducted to further elucidate the precise extent of invasion.
- We combined the traditional clinical features and DLCT based machine learning model to establish a predictive model of invasiveness of GGNs for the first time.

#### What is the implication, and what should change now?

- There are limitations in this study, including being a retrospective, single-center design with potential validation bias. For the next phase of work, we propose the establishment of a large, diverse, multicenter cohort to enhance, standardize, and validate the existing prevalent clinical risk prediction models.



**Figure 1** The flowchart of patient selection. CT, computed tomography; IA, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma.

exploration. Moreover, no previous precedents have established models for utilizing ML analysis based on DLCT to evaluate GGNs in LUAD. Therefore, the objective of this study is to develop an ML prediction model that assesses the invasiveness of GGNs in lung adenocarcinoma using DLCT parameters and incorporates clinical features, while also constructing visual charts to represent this predictive model. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tclr.amegroups.com/article/view/10.21037/tclr-24-822/rc>).

## Methods

### Study populations

The study was approved by the Institutional Review Board of Wuhan Union Hospital (No. 2024S048), and individual consent for this retrospective analysis was waived. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Between May 2020 and November 2023, a retrospective analysis was conducted on 310 patients who had undergone enhanced DLCT. The

inclusion criteria for this study were as follows: (I) patients who undergone complete surgical procedures and received pathological confirmation of AAH, AIS, MIA, or IA, with a time interval between the surgery and the most recent spectral scan of less than one month. (II) The diameter of pulmonary nodules measured on thin-layer lung window CT images ranged from 5–30 mm (7,19,20);

The exclusion criteria were as follows: (I) patients who underwent chemotherapy or radiotherapy prior to DLCT examination; (II) spectral image quality was poor, and there were artifacts in the lung nodule region. Based on these criteria, 38 out of 310 patients were excluded from our study (Figure 1).

Finally, the inclusion criteria were met by a total of 272 patients, who were then randomly assigned in a ratio of 7:3 to either the training set (n=190) or the test set (n=82). The research methodology is shown in Figure 1, providing a comprehensive overview.

### Spectral CT examination

The CT scans were performed using a dual-layer spectral CT system (IQon, Philips Healthcare, Best, The Netherlands) in accordance with established protocols. The acquisition parameters were set as follows: 120 kV, with automatic regulation of the tube current; a 512×512 matrix; collimation of 64×0.625 mm<sup>2</sup>; and reconstructed slice thickness and interval of 1.5 mm/1.5 mm. An intravenous injection of contrast medium (iohexol, 320 mg/mL) was administered at a rate of 2–2.5 mL/s. The administration of contrast medium (iohexol, 320 mg/mL) was performed intravenously at a flow rate of 2–2.5 mL/s. Reconstruction of conventional images employed the iDose4 algorithm developed by Philips Healthcare, whereas specialized spectral reconstruction algorithms were utilized for reconstructing spectral-based images (SBIs).

### Multimodal image preprocessing and DLCT images segmentation

In conventional CT, the density range of GGN is typically concentrated between –750 and –350 HU (21). The direct normalization of the images would lead to a decrease in the density resolution of GGN lesion areas. Therefore, the DICOM images underwent initial density normalization, during which the window levels were adjusted to 1,600 and –600 HU (22). The grayscale values of the GGN were normalized using a 0–1 normalization technique

subsequently, in order to mitigate the influence of contrast and brightness. Then, the automatic detection of pulmonary nodules was performed using a pre-trained 3D region-based convolutional neural network (3D-RCNN) network, followed by the application of the pre-trained 3D sphere representation-based center-points matching detection network (SCPM-Net) convolutional neural network for target nodule segmentation (23).

### ***Feature extraction and model establishment***

The findings of various studies have shown that venous phase imaging accurately depicts tumor microcirculation, enhancing its clinical applicability (24,25). The PyRadiomics package in Python was utilized to extract radiological features from each modality data (ID, Zeff Zeff, ED, and VNC) during the venous phase of the spectrum (26). The extracted features included both morphological and first-order characteristics, offering detailed information as described in PyRadiomics. (<http://pyradiomics.readthedocs.io/en/latest/>). The machine learning least absolute shrinkage and selection operator (LASSO) regression method was utilized for performing optimal feature selection, and cross-validation was applied to determine the optimal model parameters.

Subsequently, a radiomics model based on the optimal feature subset identified from the training dataset was developed using logic-based scoring. The characteristic of an ML algorithm is its extraordinary performance, and its performance in predicting results on large datasets is much higher than that of traditional regression (26,27). We developed a light gradient boosting machine (LightGBM). For the entire sample, a random selection was made to allocate 70% as the training set and the remaining 30% as the testing set. The training process of machine learning ML based models incorporates adjustments to mitigate overfitting. On this basis, we established radiomics models based on different modalities of radiomics (conventional, ID, VNC, ED, and Zeff). Additionally, a logistic regression analysis model incorporating clinical data was established. Subsequently, the radiomics features were integrated with the clinical data to construct a composite model. To facilitate visualization and provide a personalized tool for predicting the invasiveness of GGNs, we developed a nomogram. We calculated the area under the curve (AUC) as well as the corresponding sensitivity and specificity, and overall accuracy of the ML algorithm. An overview of the research methodology is presented in *Figure 2*.

### ***Statistical analysis***

The baseline characteristics were compared utilizing the Wilcoxon's rank sum test for continuous parameters, while the  $\chi^2$  test or Fisher's exact test was employed for categorical variables. Correlations between features were assessed using Spearman correlation. To assess the performance of the model on both training and test sets, we utilized the AUC and the precision-recall curve, along with calculating the odds ratios (ORs) for each selected variable. When comparing the performance of ML algorithms, an AUC close to 1 indicates better classification model performance. Risk factors were identified through optimization algorithms and a nomogram was subsequently created. The nomogram was subsequently assessed for discriminative ability and calibration using consistency statistics and calibration curves, respectively. All tests were double tailed, and a P value of <0.05 was considered statistically significant. The statistical analyses were performed using the R 4.0.4 and SPSS 16.0 software packages (R Core Team, Vienna, Austria; IBM Corp., Armonk, NY, USA). Additionally, decision curve analysis (DCA) was conducted to assess the clinical utility of the model by estimating net benefits across various threshold probabilities.

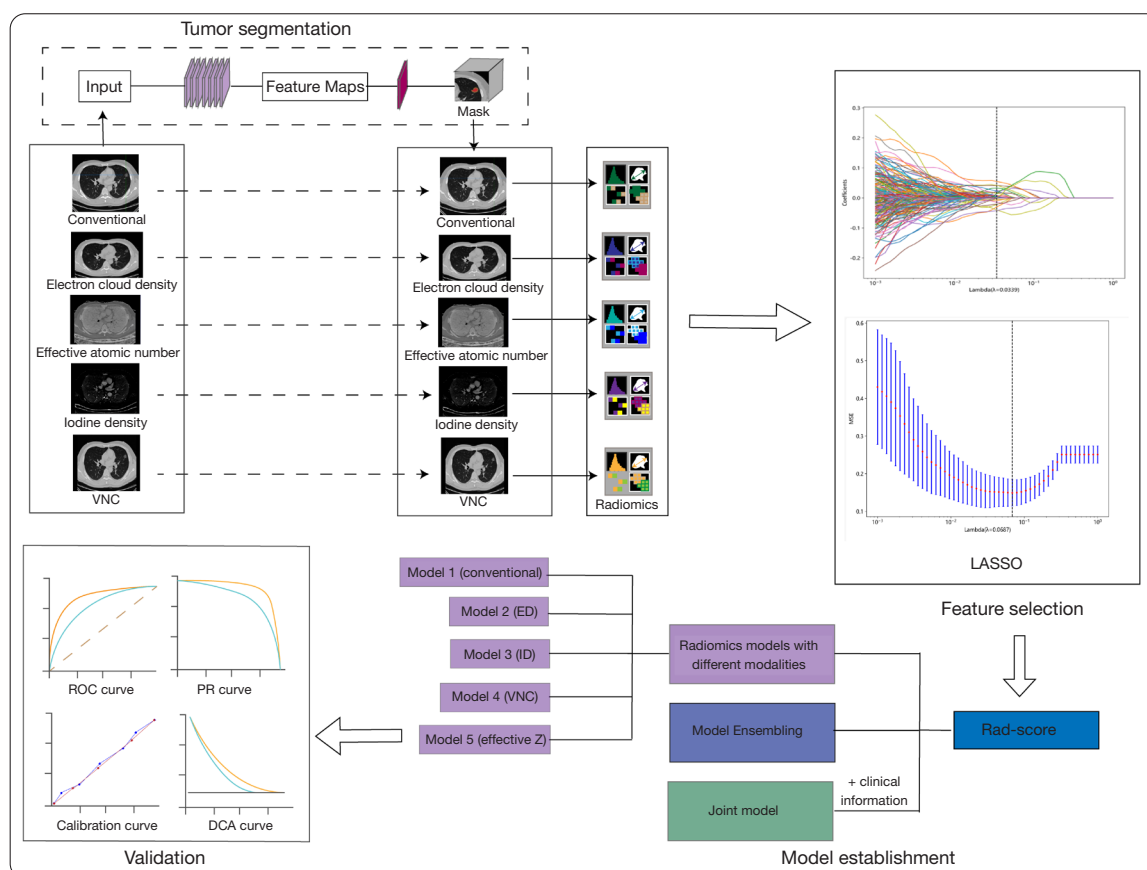
## **Results**

### ***Patient characteristics***

The HRCT detected a total of 272 GGNs. Of these, 122 were identified as pGGNs and 150 as mGGNs. In the training set, there were 85 cases of noninvasive and 105 cases of invasive GGNs included, while the test set consisted of 33 noninvasive and 49 invasive GGNs. Clinical features of the 272 patients have been succinctly outlined in *Table 1*.

### ***Univariate and multivariate logistic regression analyses of lung adenocarcinoma***

The results of univariate and multivariate logistic regression analyses of clinical features predicting the invasiveness of GGNs are shown in *Table 2*. The results of univariate logistic regression analysis revealed that age [OR = 0.947; 95% confidence interval (CI): 0.927–0.967], sex (OR = 0.449; 95% CI: 0.268–0.753), smoking (OR = 0.418; 95% CI: 0.187–0.935), alcohol history (OR = 0.233; 95% CI: 0.065–0.827) were significant factors of GGNs invasiveness. Using these clinical features, a clinical model was established, yielding



**Figure 2** The workflow of radiomics. DCA, decision curve analysis; ED, electron density; ID, iodine density; LASSO, least absolute shrinkage and selection operator; MSE, mean squared error; PR, precision recall; Rad-score, radiomics score; ROC, receiver operating characteristic; VNC, virtual non-contrast.

an AUC of 0.667 (95% CI: 0.591–0.743) in the training set and 0.706 (95% CI: 0.587–0.825) in the test set, respectively (Figure 3). The multivariate logistic regression analysis showed that the age (OR =0.951; 95% CI: 0.918–0.985) was an independent predictor of lung adenocarcinoma invasiveness. A calibration curve was used for calibration of the logistic regression analyses of the invasiveness of GGNs in lung adenocarcinoma in the clinical model (Figure S1).

### Performance comparison among different models

According to the LASSO penalized logistic regression analysis, finally 13–30 features with non-zero coefficients were selected to establish an ML model based on different DLCT parameters (Figure 4 and Figure S2). The weight of each feature based on Conventional plus ED model depicted was in Figure 4 and models based on other DLCT

parameters are shown in Figure S3. After establishing the ML models, the most important features derived from the LightGBM model based on conventional plus ED are shown in Figure 4 and other ML models are shown in Figure S4.

The conventional plus ED model outperformed ML models based on other DLCT parameters, achieving AUCs of 0.945 (95% CI: 0.917, 0.973) and 0.964 (95% CI: 0.931, 0.997) in the training and test sets, respectively (Figure 3). The AUCs for the ML model based on other DLCT parameters are depicted in Table 3 and Figure S5. Subsequently, the logistic regression model incorporated age variables that exhibited significant differences between groups, along with the radiomics score (Rad-score), resulting in the development of a composite model. The area under the receiver operating characteristic curve (AUC) for this composite model was 0.949 (95% CI: 0.922, 0.977)



**Table 1** Clinical characteristics of patients

Characteristics	Preinvasive-MIA (n=118)	IA (n=154)	P value
Age (years)	51.66 [41.75, 61.00]	58.34 [51.75, 67.00]	<0.001
Sex (female =yes, male =no)			0.003
Yes	95 (80.5)	98 (63.6)	
No	23 (19.5)	56 (36.4)	
Smoking history			0.07
Yes	8 (6.8)	21 (13.6)	
No	110 (93.2)	133 (86.4)	
Alcoholic history			0.001
Yes	2 (1.7)	13 (8.5)	
No	116 (98.3)	141 (91.5)	
Family history			0.45
Yes	5 (4.3)	4 (2.6)	
No	113 (95.7)	150 (97.4)	
CEA (<5 ng/mL)	1.70 [1.07, 2.00]	2.83 [1.30, 3.40]	0.15
SCCA (ng/mL)	0.84 [0.60, 1.10]	0.74 [0.50, 0.90]	0.37
NSE-1 (<16.3 ng/mL)	13.44 [10.68, 15.97]	13.38 [10.54, 14.78]	0.94
BMI (kg/m <sup>2</sup> )	23.31 [20.37, 25.56]	23.31 [20.76, 23.98]	0.99

Data are displayed as median [IQR] or number (%). P value is derived from the *t*-test (two-tailed distribution, equal variance assumption) between training set and test set. *P*<0.05 was considered statistically significant. BMI, body mass index; CEA, carcinoembryonic antigen; IA, invasive adenocarcinoma; IQR, interquartile range; MIA, minimally invasive adenocarcinoma; NSE, neuron-specific enolase; SCCA, squamous cell carcinoma antigen.

and 0.974 (95% CI: 0.947, 1.000) in the training and test sets, respectively (*Figure 3*).

### Clinical use

The DCA reveals that the LightGBM model based on conventional plus ED provides a net benefit to patients, when compared to both the treat-all and treat-none models. Notably, within a threshold probability range of 0 to 0.85, the ML model demonstrates significant advantages (*Figure S6*). The waterfall plots demonstrate the efficacy of the ML model in accurately discriminating between IA and pre-invasive MIA, as well as effectively stratifying IA into high and medium risk categories based on invasiveness (*Figure S7*).

### Development and evaluation of nomogram

The invasiveness of GGNs was evaluated using logistic

regression analysis with backward stepwise selection, which identified the ML model, age, gender, smoking history, and alcohol history as independent risk factors. These variables were incorporated into the development of the nomogram. The nomogram demonstrates superior clinical applicability in comparison to the intricate logistic regression formula, as it provides a more streamlined and lucid approach. According to the scores of each independent variable, a vertical projection is performed onto the axis representing the highest score points, corresponding to a specific score. Subsequently, scores are assigned to all five independent risk factor variables. The sum of each index score is projected onto the IA risk axis through its corresponding total points below, enabling prediction of GGN invasiveness in patients after DLCT (*Figure 5*). The higher the overall score, the greater the risk of IA. Therefore, the nomogram facilitates personalized prediction of GGN invasiveness based on patients' medical condition. The nomogram was calibrated utilizing a calibration curve (*Figure S8*). Calibration curve

**Table 2** Results of univariate and multivariate logistic regression analysis of clinical characteristics

Characteristics	Total, n	Univariate analysis		Multivariate analysis	
		Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value
Age	272	0.947 (0.927, 0.967)	<0.001	0.951 (0.918, 0.985)	<0.001
Sex (female = yes, male = no)	272				
Yes	193	Reference		Reference	
No	79	0.449 (0.268, 0.753)	0.002	0.507 (0.176, 1.465)	0.21
Smoking history	272				
Yes	277	Reference		Reference	
No	32	0.418 (0.187, 0.935)	0.03	1.312 (0.186, 9.246)	0.78
Alcoholic history	272				
Yes	292	Reference		Reference	
No	17	0.233 (0.065, 0.827)	0.02	0.497 (0.053, 4.671)	0.54
Family history	272				
Yes	297	Reference		–	–
No	12	1.657 (0.514, 5.340)	0.39	–	–
CEA	130	1.011 (0.899, 1.280)	0.56	–	–
SCCA	102	0.820 (0.524, 1.284)	0.38	–	–
NSE-1 (<16.3 ng/mL)	156	1.012 (0.952, 1.075)	0.70	–	–
BMI	272	0.994 (0.941, 1.050)	0.81	–	–

P<0.05 was considered statistically significant. BMI, body mass index; CEA, carcinoembryonic antigen; CI, confidence interval; NSE, neuron-specific enolase; SCCA, squamous cell carcinoma antigen.

results show a minor offset between the ideal and the actual curves.

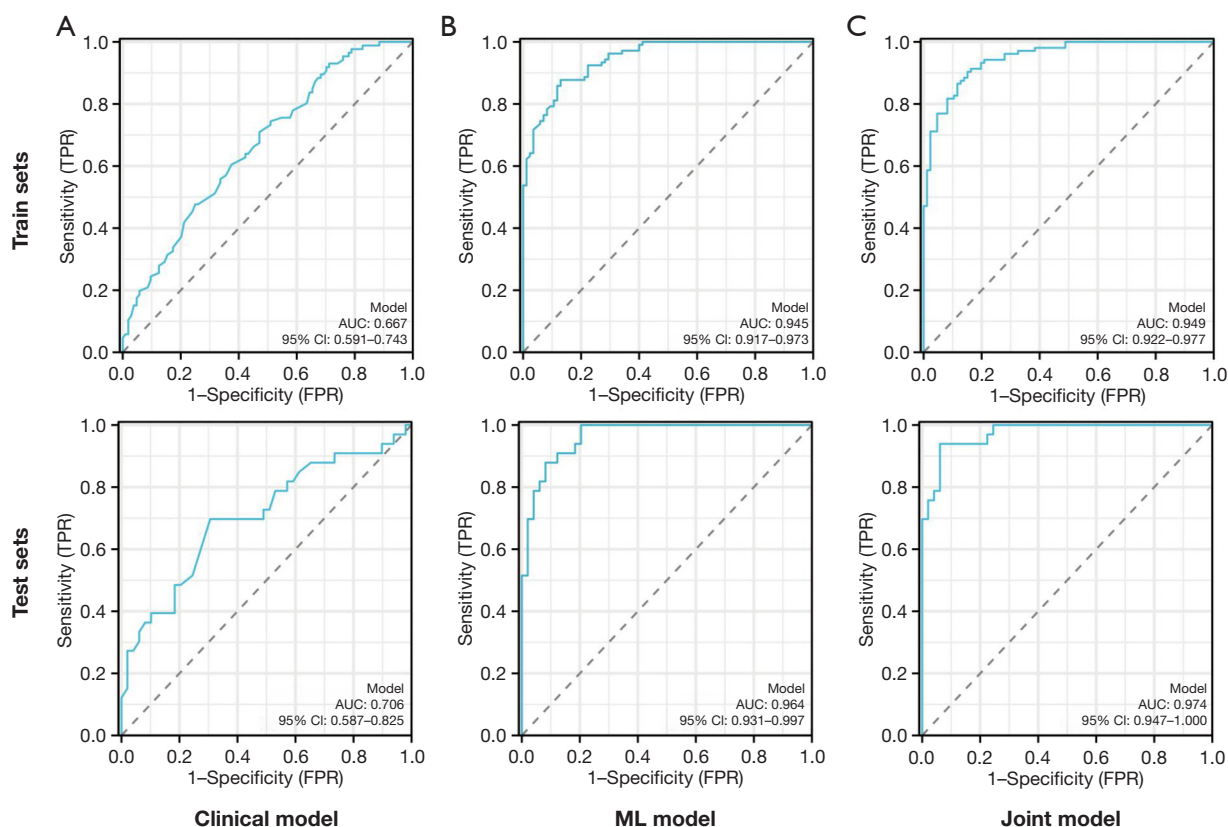
The heat maps illustrate the correlation between the Rad-scores of the machine learning prediction model and the clinical characteristics of LUAD in patients with GGNs (*Figure 6*). Patients exhibit higher Rad-scores in the IA group than in the pre-invasive MIA group.

## Discussion

The clinical presentation of early-stage LUAD often involves the presence of GGNs, which deviate from the typical manifestations and pose challenges in distinguishing between different subtypes of adenocarcinoma. Therefore, the application of radiomics is crucial in the early detection and prognosis assessment of patients. The current research efforts primarily focus on predicting the invasiveness of lung adenocarcinoma, with only a limited number of multi-class studies conducted to further elucidate the precise extent of

invasion (28–31). Therefore, in this study, we aim to predict the invasiveness of GGNs using an ML model based on DLCT parameters and further classify the GGNs into pre-invasive MIA, medium, and high risk of IA. We assessed the primary imaging characteristics, quantified parameters, and extracted 13 radiomics features from multimodal DLCT images derived from the LightGBM model based on conventional plus ED. Subsequently, the development of a clinical prognostic model for invasiveness was conducted, incorporating gender, age, smoking and drinking history. Subsequently, the model was validated using backward stepwise selection. Finally, a convenient nomogram incorporating with clinical prognostic model and Rad-score was established to ensure the prediction performance of the final model is more stable and reliable.

Despite numerous studies, establishing consistent and systematic criteria for assessing the invasiveness of GGNs using conventional CT signs remains challenging due to their atypical nature in early lung cancer and the



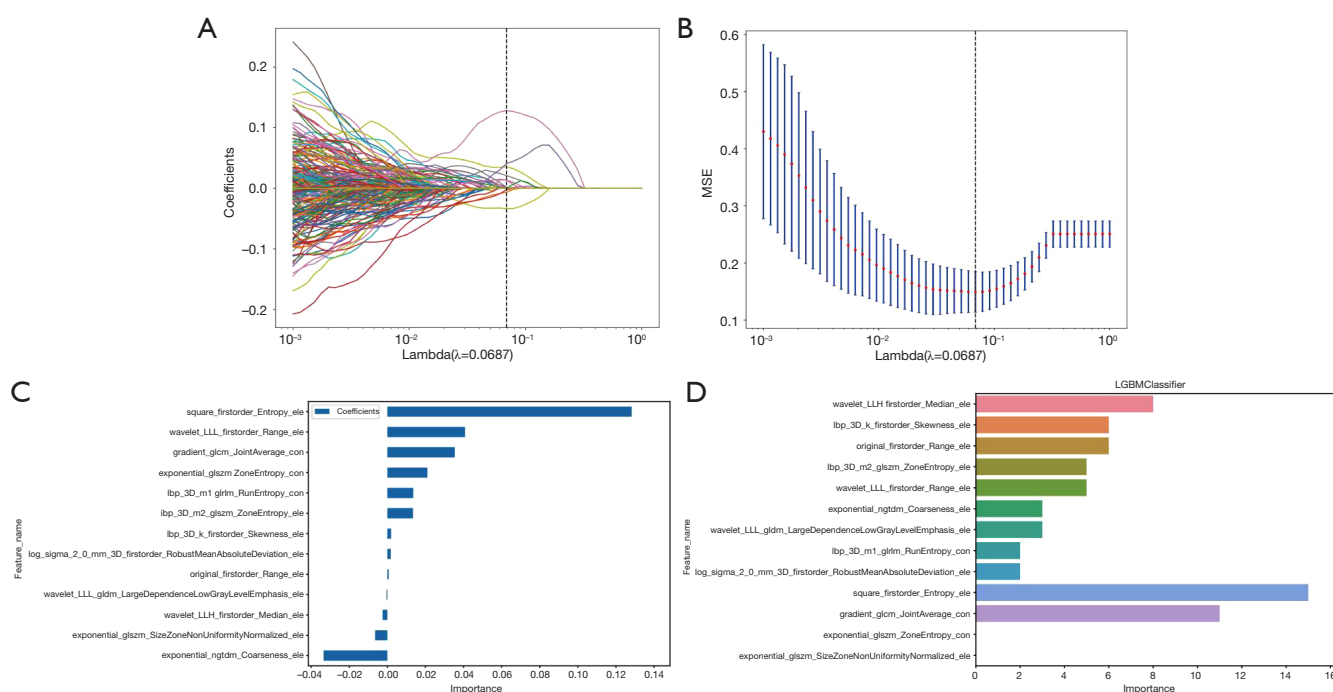
**Figure 3** Comparison of ROC curves for the clinical model, (A) ML model based on the conventional plus ED (B) and joint model combined the ML and clinical model (C) in the train and test sets. ROC, receiver operating characteristic; ML, machine learning; ED, electron density; PR, false positive rate; TPR, true positive rate.

extensive diagnostic expertise required. Previous studies have primarily focused on investigating the imaging characteristics and radiomics based on conventional CT. Recently, an increasing number of researchers have been utilizing DLCT for the analysis of GGNs invasiveness. Son *et al.* (32) revealed that monoenergetic CT values at higher energy levels, particularly at 140 keV, exhibited superior accuracy in diagnosing IA compared to monoenergetic CT values at lower energy levels. Yu *et al.* (15) further used the single-source dual-energy technique to quantitatively assess iodine concentration, standardized iodine concentration, water content, and monochromatic CT value in both plain and venous phases. They observed that the primary subtypes of pGGNs were represented by AIS and MIA. The quantitative parameters obtained from DLCT in these two phases, along with the maximum diameter of the lesion, can provide valuable insights for predicting non-invasive adenocarcinomas characterized by pGGNs. Zhang *et al.* (33) found that DLCT is better than conventional CT scans

at visualizing GGNs, leading to improved detection rates of mGGNs in 150 patients. Therefore, it is important to consider the potential presence of invasive adenocarcinoma when encountering mGGNs with higher mean CT value or ED value.

The comprehensive consideration of various risk factors is crucial for effectively predicting the probability of medium and high risk of IA. Several studies have attempted to integrate clinical-radiomics features for constructing radiomics models, which have demonstrated their effectiveness in achieving more accurate classification (28,34,35). Song *et al.* (36) conducted a study involving 346 patients with pGGNs and demonstrated that their proposed hybrid ensemble model exhibited high accuracy in predicting various invasiveness subtypes of pGGNs. Furthermore, the study revealed that a considerable number of misclassified cases initially diagnosed as MIA were predicted to be IA, indicating the challenging nature of distinguishing between these two levels. Importantly, the





**Figure 4** Radiomics feature selection using LASSO logistic regression and LightGBM model based on conventional plus ED. (A) LASSO logistic regression of radiomics features and the AUC versus the regularization parameter lambda. (B) The corresponding weights of radiomics features selected by LASSO logistic regression. (C) The top important features derived from the LightGBM model. At the Lambda value corresponding to the dashed line, the model constructed using the features preserved after dimensionality reduction performs the best. AUC, area under the curve; ED, electron density; LASSO, least absolute shrinkage and selection operator; MSE, mean squared error.

**Table 3** The AUC included in the LightGBM model based on different DLCT parameters

DLCT parameters	AUC	
	Training	Test
v_Conventional	0.938	0.939
v_EffectiveZ	0.958	0.900
v_ElectronDensity	0.950	0.943
v_IodineDensity	0.927	0.881
v_VNC	0.964	0.930
v_Conventional plus v_ElectronDensity	0.945	0.963

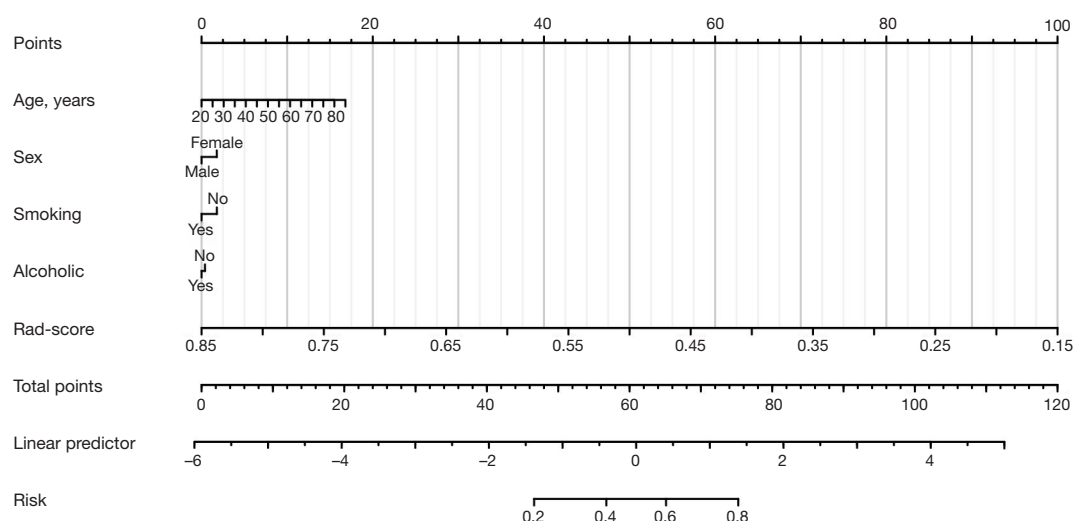
AUC, area under the curve; DLCT, detector computed tomography; LightGBM, light gradient boosting machine; VNC, virtual non-contrast.

hybrid ensemble model displayed no misclassification when differentiating IAC from PIA, highlighting its potential clinical utility.

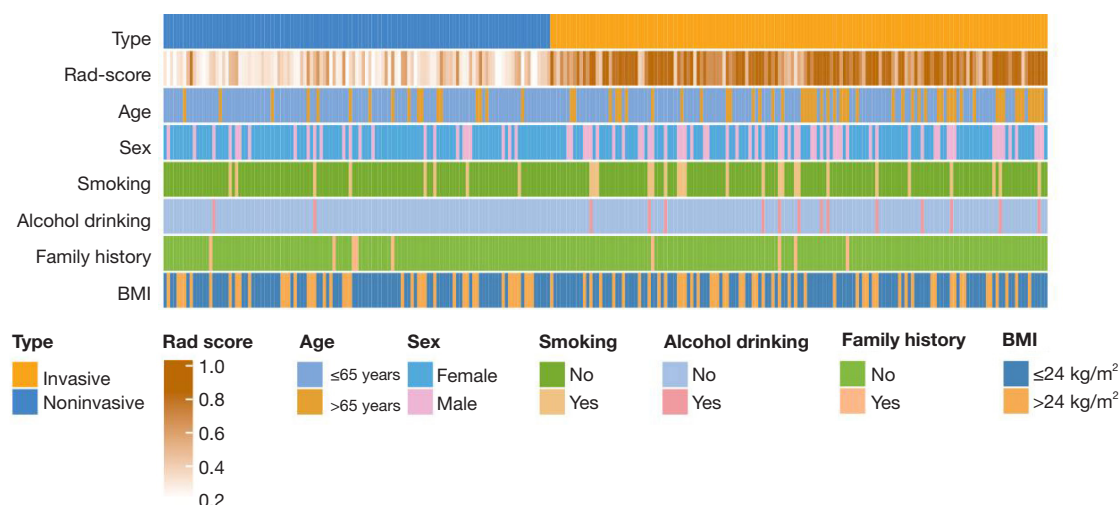
Currently, only a small number of studies have revealed the potential of radiomics based on spectral CT in

predicting the invasiveness of lung adenocarcinoma. Zheng *et al.* (18) investigated a cohort of 92 patients diagnosed with lung adenocarcinoma, comprising 30 cases of stage IA and 62 cases of preinvasive-MIA. The DLCT-based radiomics model exhibited exceptional performance, achieving an AUC of 0.957 in the training set and 0.865 in the test set. The radiomics features of their study were extracted from VMI, encompassing images acquired at 50 and 150 keV. In our previous study, we incorporated 63 DLCT multimodal radiomics features of pGGNs and ultimately identified 5 significant multimodal features. Moreover, we integrated age and gender with the Rad-score to establish a comprehensive joint model. Remarkably, the joint model exhibited superior diagnostic performance compared to the radiomics model. In the current study, we have further combined more clinical and DLCT radiomics features to construct a joint model, and the AUC performed better than the previous study (17).

In this study, we combined the traditional clinical features and DLCT based ML model to establish a predictive model of invasiveness of GGNs for the first



**Figure 5** Nomogram used for predicting invasiveness of GGNs in lung adenocarcinoma. Logistic regression algorithm was used to establish nomogram. The final score is calculated as the sum of the individual scores of each of the five variables included in the nomogram. Rad-score, radiomics score; GGNs, ground glass nodules.



**Figure 6** The heat map shows lung adenocarcinoma information of patients in different risk groups in the prospective validation cohort. BMI, body mass index; Rad-score, radiomics score.

time. Comparisons of ML models based on various DLCT parameters revealed that the LightGBM model based on conventional plus ED, exhibited the best performance in predicting GGNs invasiveness. The ML model outperformed the clinical model in accurately distinguishing IA from pre-MIA. Importantly, our study demonstrates that the LightGBM model exhibits an effective discriminatory ability in distinguishing between high and medium risk of IA. Consequently, the utilization of ML models enables

a more precise assessment of tumor invasiveness prior to surgical intervention. The calibration curve showed a strong agreement between the predicted probabilities generated by the ML model and those from both the training and test datasets, as well as with the actual probabilities. This study developed a visualized nomogram as a prediction model to assess the risk of IA. This nomogram enables clinical medical staff to calculate the probability of GGNs invasion, facilitating identification and intervention of associated risk

factors and enhancing clinical decision-making capabilities.

The present study, however, is subject to several limitations that should be acknowledged. Firstly, it should be noted that this was a retrospective study conducted at a single center, focusing exclusively on surgically resected GGNs. Consequently, our study design suffers from the presence of validation bias. The generalizability and reliability of the results might have limitations; thus, it is imperative to conduct future prospective studies with larger sample sizes in order to authenticate these findings. Although a multicentre DLCT study to further validate our findings presents challenges related to equipment variability, data standardization, logistical coordination, regulatory compliance, and financial constraints exist, with careful planning, collaboration, and adequate resources, it has the potential to strengthen research validity and contribute to advancements in pulmonary imaging and patient care. Secondly, we did not assess the differences in DLCT quantitative parameters [such as Zeff, normalized iodine concentration (NIC), spectrum curve slope] among the GGNs, future studies will assess more details of DLCT quantitative parameters to enhance the prediction ability of ML model. Thirdly, the invasiveness of GGNs was classified using a two-stage approach, initially distinguishing pre-invasive from IA, and subsequently analyzing the invasiveness of GGNs to predict medium and high-risk IA. Future studies will aim to simplify the model by employing an ML algorithm for one-step classification of the invasiveness of GGNs. For the next phase of work, we propose the establishment of a large, diverse, multicenter cohort to enhance, standardize, and validate the existing prevalent clinical risk prediction models.

## Conclusions

In conclusion, the ML model based on DLCT parameters has been proven to be a valuable tool for assessing the risk of GGNs invasion when considered collectively. With the ongoing advancements in artificial intelligence technology, a quantitative nomogram prediction model relying on DLCT radiomics features effectively differentiates between non-invasive and invasive GGNs lesions, thereby exhibiting significant potential for clinical applications.

## Acknowledgments

None.

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://tldr.amegroups.com/article/view/10.21037/tldr-24-822/rc>

*Data Sharing Statement:* Available at <https://tldr.amegroups.com/article/view/10.21037/tldr-24-822/dss>

*Peer Review File:* Available at <https://tldr.amegroups.com/article/view/10.21037/tldr-24-822/prf>

*Funding:* This study was supported by grants from the Major Program of Special Project for Technology Innovation of Hubei Province (No. 2023BCB014), the National Natural Science Foundation of China (Nos. 82172034, 82272083 and 82472058) and the Fundamental Research Funds for the Central Universities (No. 20242422).

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://tldr.amegroups.com/article/view/10.21037/tldr-24-822/coif>). P.S. and S.G. are employees of Philips Healthcare. The other authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by Ethics Committee of Wuhan Union Hospital (No. 2024S048) and individual consent for this retrospective analysis was waived.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin* 2024;74:12-49.

2. Asamura H, Nishimura KK, Giroux DJ, et al. IASLC Lung Cancer Staging Project: The New Database to Inform Revisions in the Ninth Edition of the TNM Classification of Lung Cancer. *J Thorac Oncol* 2023;18:564-75.
3. Ni Y, Yang Y, Zheng D, et al. The Invasiveness Classification of Ground-Glass Nodules Using 3D Attention Network and HRCT. *J Digit Imaging* 2020;33:1144-54.
4. Chen X, Yao B, Li J, et al. Feasibility of Using High-Resolution Computed Tomography Features for Invasiveness Differentiation of Malignant Nodules Manifesting as Ground-Glass Nodules. *Can Respir J* 2022;2022:2671772.
5. Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008;246:697-722.
6. Chen Z, Long Y, Zhang Y, et al. Detection efficacy of analog [18F]FDG PET/CT, digital [18F]FDG, and [13N]NH<sub>3</sub> PET/CT: a prospective, comparative study of patients with lung adenocarcinoma featuring ground glass nodules. *Eur Radiol* 2023;33:2118-27.
7. Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *J Thorac Oncol* 2022;17:362-87.
8. Yu F, Peng M, Bai J, et al. Comprehensive characterization of genomic and radiologic features reveals distinct driver patterns of RTK/RAS pathway in ground-glass opacity pulmonary nodules. *Int J Cancer* 2022;151:2020-30.
9. Jiang Y, Che S, Ma S, et al. Radiomic signature based on CT imaging to distinguish invasive adenocarcinoma from minimally invasive adenocarcinoma in pure ground-glass nodules with pleural contact. *Cancer Imaging* 2021;21:1.
10. Sun Y, Li C, Jin L, et al. Radiomics for lung adenocarcinoma manifesting as pure ground-glass nodules: invasive prediction. *Eur Radiol* 2020;30:3650-9.
11. Wu F, Tian SP, Jin X, et al. CT and histopathologic characteristics of lung adenocarcinoma with pure ground-glass nodules 10 mm or less in diameter. *Eur Radiol* 2017;27:4037-43.
12. Han L, Zhang P, Wang Y, et al. CT quantitative parameters to predict the invasiveness of lung pure ground-glass nodules (pGGNs). *Clin Radiol* 2018;73:504.e1-7.
13. Yu Y, Fu Y, Chen X, et al. Dual-layer spectral detector CT: predicting the invasiveness of pure ground-glass adenocarcinoma. *Clin Radiol* 2022;77:e458-65.
14. Li M, Fan Y, You H, et al. Dual-Energy CT Deep Learning Radiomics to Predict Macrotrabecular-Massive Hepatocellular Carcinoma. *Radiology* 2023;308:e230255.
15. Yu Y, Cheng JJ, Li JY, et al. Determining the invasiveness of pure ground-glass nodules using dual-energy spectral computed tomography. *Transl Lung Cancer Res* 2020;9:484-95.
16. Yang Y, Li K, Sun D, et al. Invasive Pulmonary Adenocarcinomas Versus Preinvasive Lesions Appearing as Pure Ground-Glass Nodules: Differentiation Using Enhanced Dual-Source Dual-Energy CT. *AJR Am J Roentgenol* 2019;213:W114-22.
17. Wang Y, Chen H, Chen Y, et al. A semiautomated radiomics model based on multimodal dual-layer spectral CT for preoperative discrimination of the invasiveness of pulmonary ground-glass nodules. *J Thorac Dis* 2023;15:2505-16.
18. Zheng Y, Han X, Jia X, et al. Dual-energy CT-based radiomics for predicting invasiveness of lung adenocarcinoma appearing as ground-glass nodules. *Front Oncol* 2023;13:1208758.
19. Travis WD, Brambilla E, Noguchi M, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 2011;6:244-85.
20. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* 2017;284:228-43.
21. Cohen JG, Goo JM, Yoo RE, et al. Software performance in segmenting ground-glass and solid components of subsolid nodules in pulmonary adenocarcinomas. *Eur Radiol* 2016;26:4465-74.
22. Shur JD, Doran SJ, Kumar S, et al. Radiomics in Oncology: A Practical Guide. *Radiographics* 2021;41:1717-32.
23. Luo X, Song T, Wang G, et al. SCPM-Net: An anchor-free 3D lung nodule detection network using sphere representation and center points matching. *Med Image Anal* 2022;75:102287.
24. Zegadło A, Żabicka M, Różyk A, et al. A New Outlook on the Ability to Accumulate an Iodine Contrast Agent in Solid Lung Tumors Based on Virtual Monochromatic Images in Dual Energy Computed Tomography (DECT): Analysis in Two Phases of Contrast Enhancement. *J Clin Med* 2021;10:1870.
25. Zhang Z, Zou H, Yuan A, et al. A Single Enhanced Dual-Energy CT Scan May Distinguish Lung Squamous Cell Carcinoma From Adenocarcinoma During the Venous

- phase. *Acad Radiol* 2020;27:624-9.
26. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017;77:e104-7.
  27. Deo RC. Machine Learning in Medicine. *Circulation* 2015;132:1920-30.
  28. Meng F, Guo Y, Li M, et al. Radiomics nomogram: A noninvasive tool for preoperative evaluation of the invasiveness of pulmonary adenocarcinomas manifesting as ground-glass nodules. *Transl Oncol* 2021;14:100936.
  29. Zheng H, Zhang H, Wang S, et al. Invasive Prediction of Ground Glass Nodule Based on Clinical Characteristics and Radiomics Feature. *Front Genet* 2021;12:783391.
  30. Shi L, Shi W, Peng X, et al. Development and Validation a Nomogram Incorporating CT Radiomics Signatures and Radiological Features for Differentiating Invasive Adenocarcinoma From Adenocarcinoma In Situ and Minimally Invasive Adenocarcinoma Presenting as Ground-Glass Nodules Measuring 5-10mm in Diameter. *Front Oncol* 2021;11:618677.
  31. Ren H, Xiao Z, Ling C, et al. Development of a novel nomogram-based model incorporating 3D radiomic signatures and lung CT radiological features for differentiating invasive adenocarcinoma from adenocarcinoma in situ and minimally invasive adenocarcinoma. *Quant Imaging Med Surg* 2023;13:237-48.
  32. Son JY, Lee HY, Kim JH, et al. Quantitative CT analysis of pulmonary ground-glass opacity nodules for distinguishing invasive adenocarcinoma from non-invasive or minimally invasive adenocarcinoma: the added value of using iodine mapping. *Eur Radiol* 2016;26:43-54.
  33. Zhang Z, Yin F, Kang S, et al. Dual-layer spectral detector CT (SDCT) can improve the detection of mixed ground-glass lung nodules. *J Cancer Res Clin Oncol* 2023;149:5901-6.
  34. Song L, Xing T, Zhu Z, et al. Hybrid Clinical-Radiomics Model for Precisely Predicting the Invasiveness of Lung Adenocarcinoma Manifesting as Pure Ground-Glass Nodule. *Acad Radiol* 2021;28:e267-77.
  35. Wu YJ, Liu YC, Liao CY, et al. 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.
  36. Song F, Song L, Xing T, et al. A Multi-Classification Model for Predicting the Invasiveness of Lung Adenocarcinoma Presenting as Pure Ground-Glass Nodules. *Front Oncol* 2022;12:800811.

**Cite this article as:** Wan J, Lin X, Wang Z, Sun P, Gui S, Ye T, Fan Q, Liu W, Pan F, Yang B, Geng X, Quan Z, Yang L. Dual-layer spectral detector computed tomography multiparameter machine learning model for prediction of invasive lung adenocarcinoma. *Transl Lung Cancer Res* 2025;14(2):385-397. doi: 10.21037/tlcr-24-822