

# Enhancing the prediction of the invasiveness of pulmonary adenocarcinomas presenting as pure ground-glass nodules: Integrating intratumor heterogeneity score with clinical-radiological features via machine learning in a multicenter study

DIGITAL HEALTH  
Volume 10: 1–12  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076241289181  
journals.sagepub.com/home/dhj



Hong Zheng<sup>1</sup>, Wei Chen<sup>2</sup>, Wanyin Qi<sup>3</sup>, Haibo Liu<sup>4</sup> and Zhichao Zuo<sup>5</sup> 

## Abstract

**Objective:** The invasiveness of lung adenocarcinoma significantly impacts clinical decision-making. However, assessing this invasiveness preoperatively, especially when it manifests as pure ground-glass nodules (pGGN) on CT scans, poses challenges. This study aims to quantify intratumor heterogeneity (ITH) and determine whether the ITH score can enhance the accuracy of invasiveness predictions.

**Methods:** A total of 524 patients with lung adenocarcinomas presenting as pGGN were enrolled in the study, with 177 (33.78%) receiving a pathologic diagnosis of invasiveness. Four diagnostic approaches were developed to predict the invasiveness of lung adenocarcinoma presenting as pGGN: (1) conventional lesion size, (2) ITH score, (3) clinical-radiological features (ClinRad), and (4) integration of the ITH score with ClinRad. ClinRad alone or in combination with the ITH score served as the input for 11 machine learning approaches. The trained models were evaluated in an independent validation cohort, and the area under the curve (AUC) was calculated to assess classification performance.

**Results:** The conventional lesion size showed the lowest performance, with an AUC of 0.826 (95% confidence interval [CI]: 0.758–0.894), while the ITH score outperformed it with an AUC of 0.846 (95% CI: 0.787–0.905). The CatBoost model performed best when the ITH score and ClinRad were both used as input features, leading to the development of an ITH-ClinRad-guided CatBoost classifier. CatBoost also excelled with ClinRad alone, resulting in a ClinRad-guided CatBoost classifier with an AUC of 0.830 (95% CI: 0.764–0.896), surpassed by the ITH-ClinRad-guided CatBoost classifier with an AUC of 0.871 (95% CI: 0.818–0.924).

**Conclusion:** The ITH-ClinRad-guided CatBoost classifier emerges as a promising tool with significant potential to revolutionize the management of lung adenocarcinomas presenting as pGGNs.

<sup>1</sup>Department of Radiology, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University/Hunan Cancer Hospital, Changsha, PR China

<sup>2</sup>Department of Radiology, The Second People's Hospital of Hunan Province, Brain Hospital of Hunan Province, Changsha, PR China

<sup>3</sup>Department of Radiology, The Affiliated Hospital of Southwest Medical University, Luzhou, PR China

<sup>4</sup>Department of Radiology, Xiangtan Central Hospital, Xiangtan, PR China

<sup>5</sup>School of Mathematics and Computational Science, Xiangtan University, Xiangtan, PR China

## Corresponding authors:

Wei Chen, Department of Radiology, The Second People's Hospital of Hunan Province, Brain Hospital of Hunan Province, Changsha 410007, Hunan Province, P. R. China.

Email: weichen.work@hotmail.com

Zhichao Zuo, School of Mathematics and Computational Science, Xiangtan University, Xiangtan 411105, Hunan Province, P. R. China.

Email: zuo\_z@smail.xtu.edu.cn



## Keywords

Intratumor heterogeneity, invasiveness, pulmonary adenocarcinoma, pure ground-glass nodule, machine learning

Submission date: 3 July 2024; Acceptance date: 17 September 2024

## Introduction

With advancements in lung cancer screening, lung adenocarcinomas can be detected earlier using computed tomography (CT). Ground glass nodules (GGNs) are focal nodular areas that exhibit increased lung attenuation, pulmonary vessels, and bronchial structures within the nodule, as defined by the Fleischner Society. GGNs are classified based on the presence of solid components into mixed GGNs and pure GGNs (pGGNs). pGGNs are characterized by the absence of solid components<sup>1–3</sup> and demonstrate a more indolent growth pattern and slow progression.<sup>4,5</sup> Even in cases where pGGNs are diagnosed as malignant after surgery, pathology is more likely to reveal atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), or minimally invasive adenocarcinoma (MIA), rather than invasive adenocarcinoma (IAC).<sup>6,7</sup> However, while many pGGNs follow an indolent course, a significant number are diagnosed pathologically as IAC, with incidence rates ranging from 35.4% to 44.2%.<sup>8–10</sup> AIS and MIA exhibit 10-year disease-free survival rates of 100% after complete resection,<sup>11</sup> whereas IAC has a 5-year survival rate of 89% after complete resection.<sup>12</sup> Consequently, while segmentectomy or wedge resection is often adequate for AAH/AIS or MIA, IAC is typically treated with lobectomy.<sup>13,14</sup> Preoperative assessment of the invasiveness of lung adenocarcinoma on CT significantly affects clinical decision-making and ensures appropriate management.<sup>15</sup> However, this assessment can be challenging owing to overlapping CT features among pGGNs, potentially leading to overdiagnosis and overtreatment.

Previous studies have shown that the invasiveness of pGGNs is primarily associated with the lesion size rather than the clinical-radiological features (ClinRad), such as age, margin, bubble lucency, or pleural indentation.<sup>16–19</sup> Consequently, management of pGGNs is based on the lesion size, following the guidelines of the 8th edition of TNM staging, in which lesion size is defined as the maximum diameter in the axial plane on CT.<sup>20,21</sup> Although lesion size is a convenient and straightforward method for assessing invasiveness, its utility is limited due to insufficient objectivity and inability to fully reflect biological properties. Lesion size overlooks intratumor heterogeneity (ITH) within the lesion, thus constraining diagnostic efficacy. ITH arises from various factors such as angiogenesis, metabolism, and proliferation, as well as diverse cell populations and their

uneven distribution within different intratumoral subregions.<sup>22</sup> Currently, multi-scale information captured through the simultaneous integration of local and global pixel data on CT images that considers pixel characteristics and their spatial distribution can quantify ITH on CT images.<sup>23,24</sup> The ITH score has been reported to be useful for prognosis assessment<sup>23</sup> and distinguishing histological subtypes.<sup>25</sup> Therefore, it is crucial to determine whether ITH score can aid in the assessment of the invasiveness of pGGNs and compare its efficacy with that of traditional diagnostic approaches, and whether it can be integrated into a new diagnostic biomarker.

Machine learning algorithms primarily achieve the mapping function from the input to the output by adjusting their parameters based on the given dataset, aiming to make the output as close as possible to the expected objective function. This approach has been widely applied in lung cancer diagnosis, treatment, and prognosis.<sup>26</sup> Given the lack of reliable, accurate, and efficient predictive models or diagnostic biomarkers, this study employed 11 machine learning algorithms to evaluate the invasiveness of pGGNs, with ClinRad and ITH scores as the input features. The primary objective of this study was to identify a new diagnostic biomarker capable of accurately predicting the invasiveness of adenocarcinoma on chest CT, thereby providing scientific evidence for clinical decision-making.

## Methods

### Study population

Ethical approval for this multicenter population-registry cancer study was obtained from the institutional ethics committees of the participating institutions (project approval number: No. 2021-07-009, No. KY2020147). Given the retrospective nature of the study, the requirement for written informed consent was waived. The study was conducted from January 2021 to June 2022 and included individuals with a pathological diagnosis of lung adenocarcinoma, which was categorized as AIS, MIA, or IAC, according to the IASLC/ATS/ERS classification.<sup>13</sup>

Inclusion Criteria: (1) lung lesions that manifested as pGGNs on preoperative CT scans with a thin slice thickness of  $\leq 1.5$  mm. (2) The pGGN lesion size had to exceed 5 mm

and be less than 30 mm in diameter on CT imaging. (3) Patients with no history of biopsy, chemotherapy, or radiotherapy prior to CT evaluation were included. Exclusion Criteria: (1) Patients with CT scans of poor image quality that significantly affected the observation and measurement of pGGNs; (2) Patients with a time gap of more than 14 days between the CT scan and surgery; (3) Patients with concurrent severe chronic conditions, such as other malignancies or significant comorbidities that could affect their overall health status and surgical outcomes.

Ultimately, 524 patients with persistent pGGNs were enrolled in the study. For cases with multiple lesions, representative lesions with a definitive pathological diagnosis were selected for inclusion. The procedural flow of this study is illustrated in Supplemental Figure 1. Patients were randomly assigned to either the training cohort (70%,  $n=366$ ) or the validation cohort (30%,  $n=158$ ). The training cohort was utilized to develop and optimize the predictive model, while the validation cohort was employed to independently assess the model's performance and generalizability.

### CT examination

Preoperative chest CT was conducted at the participating hospitals using a multi-detector row CT system (Revolution CT [GE Healthcare], MX16 CT [Philips Healthcare], uCT550 or uCT760 [Shanghai United Imaging Healthcare]). The imaging procedure involved breath-hold helical scanning in the supine position, which covered the area from the lung apex to below the costophrenic angle. The volumetric CT scan parameters included a tube voltage of 120 kV, tube current-time product ranging from 180 to 280 mA, matrix size of  $512 \times 512$ , and image preprocessing. Reconstruction of the CT image was performed using a standard algorithm with the layer thickness/interval set at 0.55 to 0.625 mm.

### Clinrad features

The medical records, including clinical characteristics (age and sex), pathological findings, and preoperative chest CT scans, were comprehensively reviewed. An experienced thoracic radiologist (H.L.) with over 10 years of experience assessed all lung lesions on the CT images and was blinded to most study information except for the presence of the lesion. pGGNs were assessed using a picture archiving and communication system with a lung window setting, and Hounsfield units (HU) of 1600 and window level of  $-600$  HU. The CT value was measured three times using a region-of-interest (ROI) cursor to calculate the mean CT value and measure portions of the pGGN away from the air space and blood vessels. The maximum diameter in the axial plane determined the lesion size on CT images, following the guidelines of the 8th edition of the TNM

staging system.<sup>20,21</sup> Additionally, CT morphological features, such as tumor location, margins, lobulation, spiculation, vacuole sign, vascular convergence sign, and pleural indentation, were documented.

### ITH score calculation

ITH score acquisition involved three steps, in accordance with previous literature<sup>23–25</sup> nodule region segmentation, subregion clustering within the pGGN, and ITH score calculation.

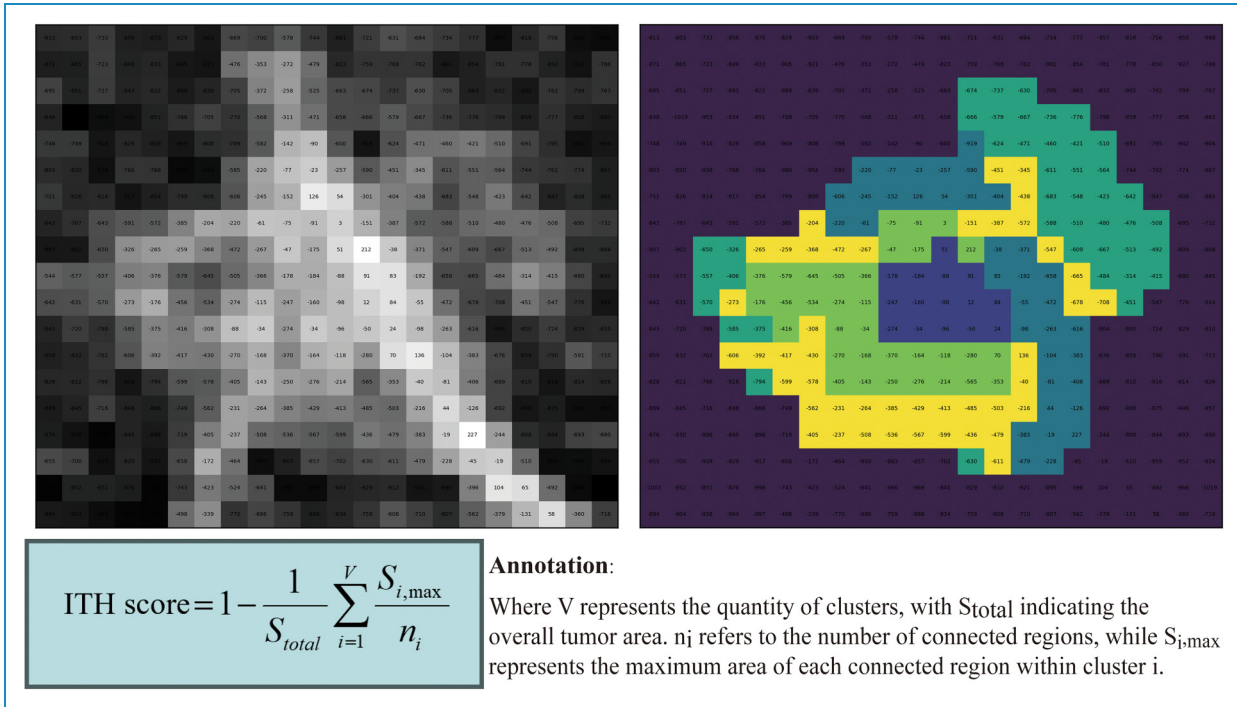
First, we used ITK-SNAP software version 3.6.0 ([www.itksnap.org](http://www.itksnap.org)) to segment the nodules in all CT images. A cardiothoracic radiologist (Z.Z.) manually delineated the ROI along the tumor border across each two-dimensional lesion section. The delineation was then reviewed by another radiologist (H.L.) with over 10 years of experience in the field, and adjustments were made, if required.

In the second step, we integrated both the local and global radiomic features. This involved a meticulous analysis of the variability in radiomic attributes within the tumor, considering both the localized and overall characteristics. We computed consistent radiomic properties for each pixel within the nodule region to capture intricate localized details. Subsequently, we employ clustering techniques based on the inherent local characteristics of the pixels to generate an informative visual map. The distribution patterns of clusters on the label map indicate the level of heterogeneity within the nodule, with pixels sharing the same cluster labels exhibiting similarities in intensity and neighboring textures.

In the third and final step, we set the optimal number of clusters to five based on previous research to precisely segment the intratumoral subregions.<sup>23,25</sup> The ITH score was used as an evaluative metric to quantify the diversity level within the label maps (Fig. 1).

### ITH-ClinRad-guided machine learning

We used 11 supervised machine-learning algorithms to integrate the ITH score with ClinRad and develop an optimal model. These algorithms, including logistic regression, gaussian naive bayes (Gaussian NB), k-nearest neighbor (KNN), support vector machine (SVM), multi-layer perceptron (MLP), classification and regression trees (CART), random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost), were imported from the scikit-learn machine learning package (version 1.4.2) in Python (version 3.12.1).<sup>27</sup> Subsequently, the models were developed using the training cohort, and a 10-fold cross-validation hyperparameter adjustment area under the curve (AUC) was used to optimize the model parameters. By employing 10-fold cross-validation on the training



**Figure 1.** Acquisition and calculation process of the ITH score.

cohort to select the optimal parameters for each model, we effectively mitigated the risk of overfitting. Comprehensive parameters are detailed in Supplemental Material File 1.

The predictive performance of all machine learning classifiers was evaluated using the AUC. The trained models were tested on an independent validation cohort, and the AUC of their classification performance was calculated. The model with the highest AUC value in the validation cohort was selected as the final model.<sup>28</sup> To further evaluate the selected model, confusion matrix-driven metrics, including accuracy, sensitivity (SEN), specificity (SPE), positive prediction value (PPV), and negative prediction value (NPV), were used, along with the corresponding 95% confidence intervals (95% CI).

### Statistical analysis

We used descriptive statistics for the demographic characteristics of the study participants. The normality of continuous variables was assessed using the Kolmogorov–Smirnov test. Continuous variables with a normal distribution were described using means  $\pm$  standard deviations and compared using the independent samples t-test. Non-normally distributed data were reported as medians (interquartile ranges) and analyzed using the Mann–Whitney U test. Categorical variables were evaluated using the chi-squared test, and statistical significance was set at  $p < 0.05$ .

## Results

### Demographic characteristics

The demographic characteristics of the study participants, all of whom were confirmed to have either IAC or AIC/MIA based on postoperative pathology, are summarized in Table 1. The training and validation cohorts did not differ significantly. A total of 524 participants were included in the final analysis; 143 (27.2%) were men, and the median age was 55 years (interquartile range: 47–64 years). Among them, 177 (33.7%) were diagnosed with IAC. pGGN with corresponding pathologic diagnosis are provided in Supplemental Material File 2.

### Machine learning models for identifying IAC

The general strategy for identifying IAC in pulmonary adenocarcinomas presenting as pGGNs using machine learning models is shown in Figure 2. We employed 11 supervised machine learning algorithms to merge the ITH score with ClinRad and developed an optimal approach. The trained models included logistic regression, Gaussian NB, KNN, SVM, MLP, CART, RF, GBM, XGBoost, LightGBM, and CatBoost, and each was cross-validated with 10-fold validation to determine the optimal model parameters in the training cohort. Model performance was assessed using the AUC in the validation cohort, with CatBoost achieving the highest AUC of 0.871 (95% CI:

**Table 1.** Baseline characteristics of the training, and validation cohort.

Variable	Training cohort (N=366)	Validation cohort (N=158)	Total (N=524)	p-Value
Location				0.383
RUL	118 (32.24%)	60 (37.97%)	178 (33.97%)	
RLL	33 (9.02%)	11 (6.96%)	44 (8.40%)	
RML	52 (14.21%)	27 (17.09%)	79 (15.08%)	
LUL	114 (31.15%)	38 (24.05%)	152 (29.01%)	
LLL	49 (13.39%)	22 (13.92%)	71 (13.55%)	
Margin				0.884
Ill defined	307 (83.88%)	131 (82.91%)	438 (83.59%)	
Well defined	59 (16.12%)	27 (17.09%)	86 (16.41%)	
Shape				0.124
Irregular	262 (71.58%)	124 (78.48%)	386 (73.66%)	
Regular	104 (28.42%)	34 (21.52%)	138 (26.34%)	
Lobulation				0.350
Absence	233 (63.66%)	108 (68.35%)	341 (65.08%)	
Presence	133 (36.34%)	50 (31.65%)	183 (34.92%)	
Spiculation				0.177
Absence	265 (72.40%)	124 (78.48%)	389 (74.24%)	
Presence	101 (27.60%)	34 (21.52%)	135 (25.76%)	
Vascular convergence sign				0.056
Absence	85 (23.22%)	50 (31.65%)	135 (25.76%)	
Presence	281 (76.78%)	108 (68.35%)	389 (74.24%)	
Vacuole sign				0.371
Absence	317 (86.61%)	142 (89.87%)	459 (87.60%)	
Presence	49 (13.39%)	16 (10.13%)	65 (12.40%)	
Pleural indentation				0.271
Absence	223 (60.93%)	105 (66.46%)	328 (62.60%)	
Presence	143 (39.07%)	53 (33.54%)	196 (37.40%)	

(continued)

Table 1. Continued.

Variable	Training cohort (N=366)	Validation cohort (N=158)	Total (N=524)	p-Value
Sex				0.895
Male	101 (27.60%)	42 (26.58%)	143 (27.29%)	
Female	265 (72.40%)	116 (73.42%)	381 (72.71%)	
Age	54.00 [47.00;64.00]	56.00 [49.00;64.00]	55.00 [47.00;64.00]	0.691
Lesion size(mm)	13.04 [9.85;17.03]	12.91 [9.43;16.76]	13.04 [9.85;17.00]	0.734
Mean CT value (Hu)	-605.60 [-666.14; -535.89]	-625.31 [-685.58; -558.91]	-609.48 [-670.50; -537.75]	0.158
ITH score	0.54 [0.39;0.66]	0.53 [0.37;0.69]	0.53 [0.38;0.67]	0.891
Pathologic diagnosis				0.074
AIS/MIA	233 (63.66%)	114 (72.15%)	347 (66.22%)	
IAC	133 (36.34%)	44 (27.85%)	177 (33.78%)	

Abbreviation: LUL, left upper lobe; LLL, left lower lobe; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; IAC, invasive lung adenocarcinoma; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; ITH, intratumor heterogeneity.

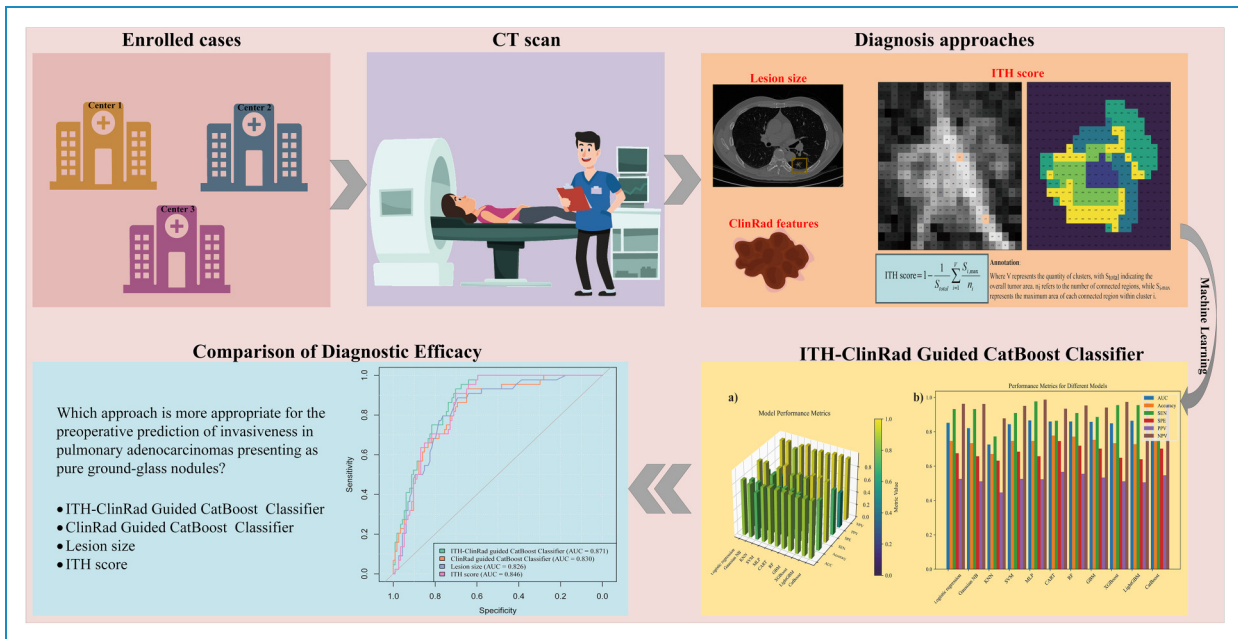
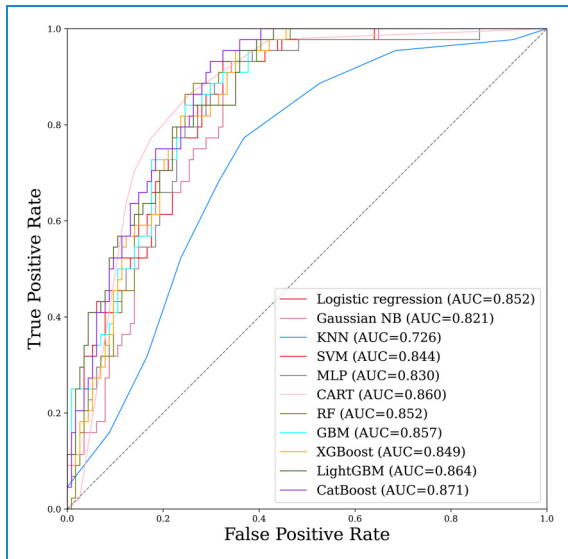


Figure 2. Study framework of enrolled cases, including CT scan, diagnosis approaches, the ITH-clinRad guided catBoost classifier, and comparison of diagnosis efficacy.

0.818–0.924), as illustrated in the receiver operating characteristic (ROC) curve in Figure 3. Detailed performance metrics, such as AUC, ACC, SEN, SPE, PPV, and NPV, are outlined in Table 2 and Figure 4, confirming the suitability of the CatBoost model for classification modeling.

Consequently, we developed an ITH-ClinRad-guided CatBoost classifier for IAC identification.

According to further investigation of the ClinRad-guided machine learning for IAC prediction, the input variables consisted solely of ClinRad and the ITH score was



**Figure 3.** ROC curve analysis of 11 machine learning algorithms with the merge of the ITH score with clinRad.

excluded. CatBoost still outperformed the other models with an AUC of 0.830 (95% CI: 0.764–0.896), as shown in Supplemental Fig. 2. Accordingly, we established a ClinRad-guided CatBoost classifier for IAC classification.

### Model interpretation

The SHapley Additive exPlanations (SHAP) method was used to visualize the selected variables in the ITH-ClinRad-guided CatBoost classifier for predicting IAC in pulmonary adenocarcinomas presenting as pGGNs and to explain their impact.<sup>29</sup> The results are presented as beeswarm plots, heatmaps, and force plots.

In Figure 5(a), the beeswarm plot illustrates the 13 most crucial features of the model. Each feature's importance line shows how all patients contributed to the results, with various colored dots indicating high- and low-risk values. The ITH score holds the highest importance in the model, followed by lesion size and mean CT value. The colored dots indicate that a higher ITH score, lesion size, and mean CT value correlate with a greater risk of IAC.

Figure 5(b) presents a heatmap showing the attributes of the SHAP values using different colored dots. Patients are arranged based on hierarchical clustering according to the similarity of explanations. The output of the ITH-ClinRad-guided CatBoost classifier is displayed above the heatmap matrix, with the gray dashed line denoting the baseline value. The bar chart on the right side of the figure highlights the global importance, reaffirming that the ITH score had the greatest significance, followed by lesion size and mean CT value.

In Figure 5(c), the force plot demonstrates the prediction explanation for a specific instance. Positive and negative

impacts are shown in red and blue, respectively. The length of the arrow signifies the magnitude of the feature's impact on the output. It is evident that the highest ITH score (0.800), followed by lesion size (20.12 mm) and mean CT value (−517.87 HU), positively influenced IAC risk in pulmonary adenocarcinomas presenting as pGGNs.

In real-world clinical decision-making, when applying the ITH-ClinRad-guided CatBoost classifier to predict the risk of IAC in pulmonary adenocarcinomas presenting as pGGNs, the classifier integrates ClinRad features such as lesion size and mean CT value, along with the ITH score. As illustrated in Figure 5, for a patient with a persistent pGGN, if the ITH-ClinRad-guided CatBoost classifier predicts a high risk of IAC, the recommended surgical approach would be lobectomy. Conversely, if the classifier predicts a low risk of IAC, the patient would typically undergo segmentectomy or wedge resection, which is often sufficient.

### Various diagnostic models for IAC prediction

The ITH-ClinRad-guided CatBoost classifier, ClinRad-guided CatBoost classifier, ITH score, and lesion size were utilized to predict IAC in pulmonary adenocarcinomas presenting as pGGNs. Among these models, conventional lesion size demonstrated the lowest performance with an AUC of 0.826 (95% CI: 0.758–0.894), which is lower than the AUC of 0.846 (95% CI: 0.787–0.905) for the ITH score. Figure 6 illustrates that the ITH-ClinRad-guided CatBoost classifier exhibited the highest performance, followed by the ITH score, ClinRad-guided CatBoost classifier, and lesion size. Detailed performance metrics, including AUC, ACC, SEN, SPE, PPV, and NPV, are presented in Table 3. Subgroup analyses were conducted to further explore the ability of the ITH-ClinRad-guided CatBoost classifier to differentiate between AIS from MIA/IAC. The results are provided in Supplemental Material File 3.

### Correlation between ITH score and clinRad

The relationship between ITH score and ClinRad is depicted in a chord diagram (Figure 7(a)) and matrix plot (Figure 7(b)), where dark green signifies a positive correlation and khaki indicates a negative correlation. The intensity of the color or thickness of the line reflects the strength of the correlation. Specifically, the ITH score was positively correlated with various radiographic features associated with the degree of malignancy, such as the mean CT value, lesion size, male sex, pleural indentation, vacuole sign, spiculation, and lobulation.

### Discussion

In the current study, we conducted an in-depth exploration by employing 11 supervised machine learning algorithms to

**Table 2.** Diagnostic efficacy across Various machine learning models.

Models	AUC	Accuracy	SEN	SPE	PPV	NPV
Logistic regression	0.852 (0.794-0.910)	0.747 (0.744-0.749)	0.932 (0.857-1.000)	0.675 (0.589-0.761)	0.526 (0.415-0.636)	0.963 (0.921-1.004)
Gaussian NB	0.821 (0.756-0.885)	0.734 (0.732-0.737)	0.932 (0.857-1.000)	0.658 (0.571-0.745)	0.512 (0.403-0.622)	0.962 (0.919-1.004)
KNN	0.726 (0.644-0.808)	0.671 (0.668-0.674)	0.773 (0.649-0.897)	0.632 (0.543-0.720)	0.447 (0.336-0.559)	0.878 (0.807-0.949)
SVM	0.844 (0.783-0.904)	0.747 (0.744-0.749)	0.909 (0.824-0.994)	0.684 (0.599-0.770)	0.526 (0.414-0.639)	0.951 (0.905-0.998)
MLP	0.830 (0.811-0.920)	0.747 (0.744-0.749)	0.977 (0.933-1.000)	0.658 (0.571-0.745)	0.524 (0.416-0.632)	0.987 (0.961-1.012)
CART	0.860 (0.802-0.918)	0.778 (0.776-0.781)	0.864 (0.762-0.965)	0.746 (0.666-0.826)	0.567 (0.449-0.686)	0.934 (0.883-0.985)
RF	0.852 (0.804-0.916)	0.772 (0.770-0.774)	0.909 (0.824-0.994)	0.719 (0.637-0.802)	0.556 (0.441-0.670)	0.953 (0.909-0.998)
GBM	0.857 (0.801-0.914)	0.753 (0.751-0.755)	0.886 (0.793-0.980)	0.702 (0.618-0.786)	0.534 (0.420-0.649)	0.941 (0.891-0.991)
XGBoost	0.849 (0.791-0.907)	0.734 (0.732-0.737)	0.955 (0.893-1.000)	0.649 (0.562-0.737)	0.512 (0.404-0.620)	0.974 (0.938-1.010)
LightGBM	0.864 (0.809-0.920)	0.728 (0.725-0.730)	0.955 (0.893-1.000)	0.640 (0.552-0.728)	0.506 (0.398-0.614)	0.973 (0.937-1.010)
CatBoost	0.871 (0.818-0.924)	0.766 (0.764-0.768)	0.932 (0.857-1.000)	0.702 (0.618-0.786)	0.547 (0.434-0.659)	0.964 (0.924-1.004)

*Abbreviation:* Gaussian NB, Gaussian Naive Bayes; KNN, K-nearest neighbor; SVM, supportvector machine; MLP, multi-layer perceptron; CART, classification and regression trees; RF, random forest; GBM, gradient boosting machine; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; CatBoost, categorical boosting.

AUC, area under the curve; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value.



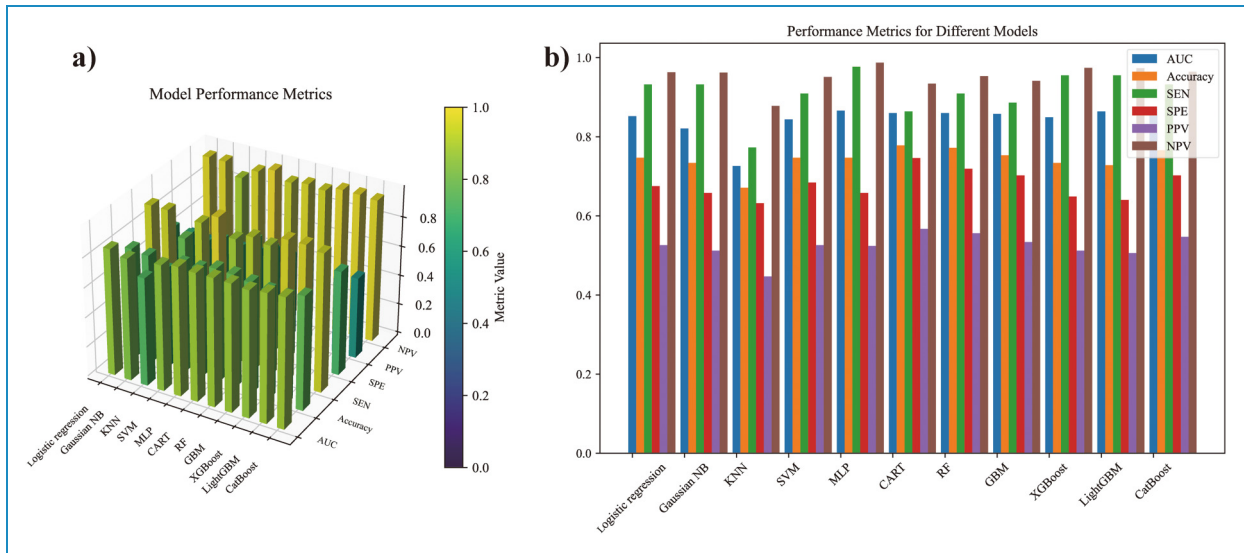


Figure 4. AUC, ACC, SEN, SPE, PPV, and NPV of 11 machine learning algorithms.

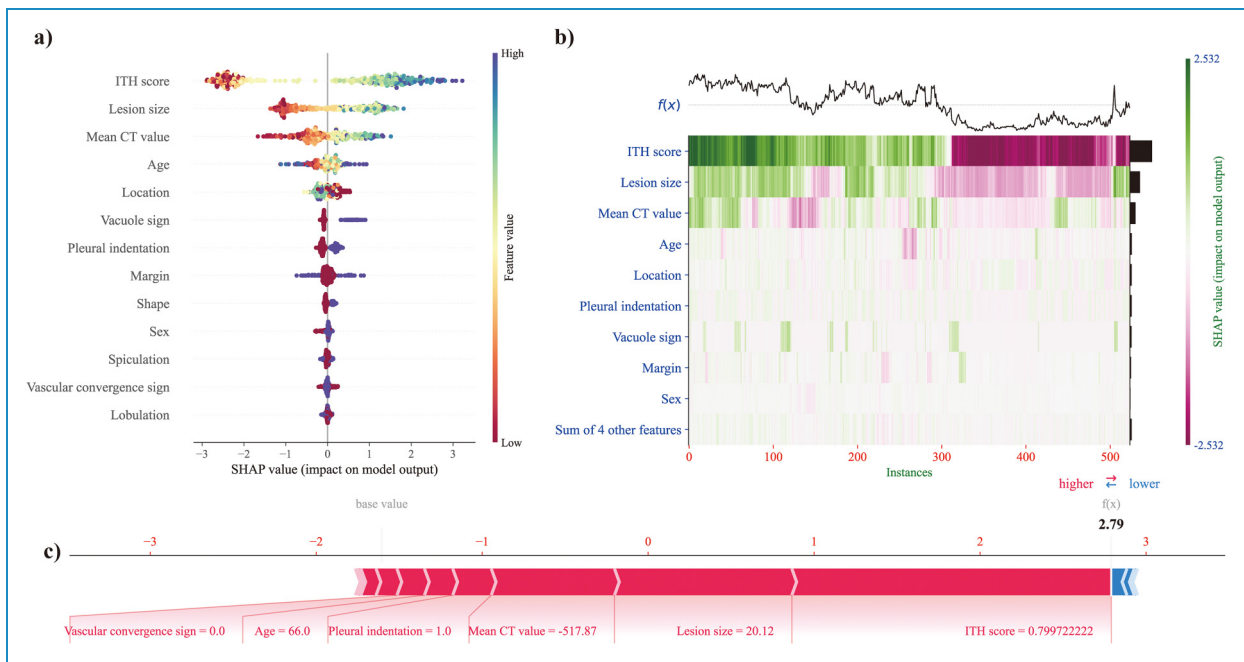
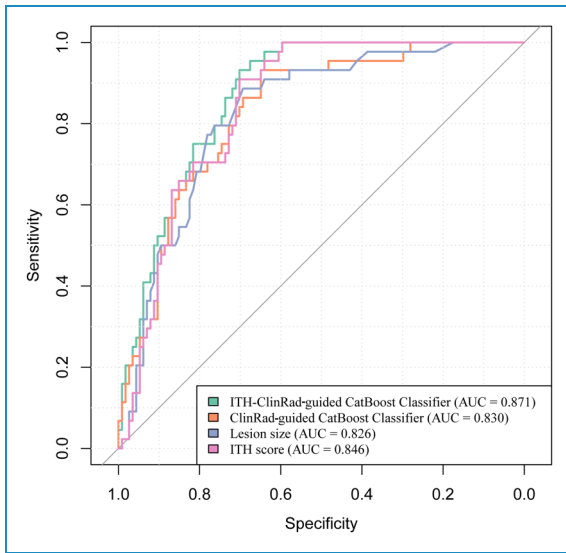


Figure 5. SHAP for model interpretation, including beeswarm plot (a), heatmap (b), and force plot (c).

merge the ITH score with ClinRad for the preoperative differentiation of IAC from AIS/MIA in pulmonary adenocarcinomas presenting as pGGNs. Among the models, the CatBoost algorithm showed the best performance; therefore, we built an ITH-ClinRad-guided CatBoost classifier for IAC identification, which outperformed the ITH score, ClinRad-guided CatBoost classifier, and lesion size.

Prior investigations have revealed that the degree of invasiveness of pGGNs is predominantly associated with the lesion size rather than ClinRad parameters such as

age, margin characteristics, bubble lucency, or pleural indentation.<sup>16–19</sup> To enhance the objectivity and fully capture intertumoral heterogeneity, we developed an ITH score that integrates local and global radiomic information, considering both pixel characteristics and their spatial distributions. This represents a significant advancement in the comprehensive capture of multiscale heterogeneity information.<sup>23–25</sup> In the current study, the ITH score achieved an AUC of 0.846 (95% CI: 0.787–0.905), outperforming lesion size, which had an AUC of 0.826 (95% CI:



**Figure 6.** ROC curve across the ITH-clinRad-guided catBoost classifier, ITH score, clinRad-guided catBoost classifier, and lesion size.

0.758–0.894). Previous studies have used ITH score for prognostic assessment<sup>23</sup> and differentiation of histological subtypes.<sup>25</sup> Zhang et al. further demonstrated that the ITH score surpassed conventional radiomic analysis in distinguishing histological subtypes of clinical T1 stage lung adenocarcinoma presenting as pGGNs.<sup>25</sup> The ITH score, computed by using clustering label maps, assists in reducing variations among images of diverse tumors while maintaining intrinsic heterogeneity within the same tumor. Additionally, the application of unsupervised machine-learning techniques to generate the ITH score offers quantitative data derived from intuitive cluster patterns, effectively addressing potential biases stemming from the intricate reduction of feature dimensionality in radiomic analysis.

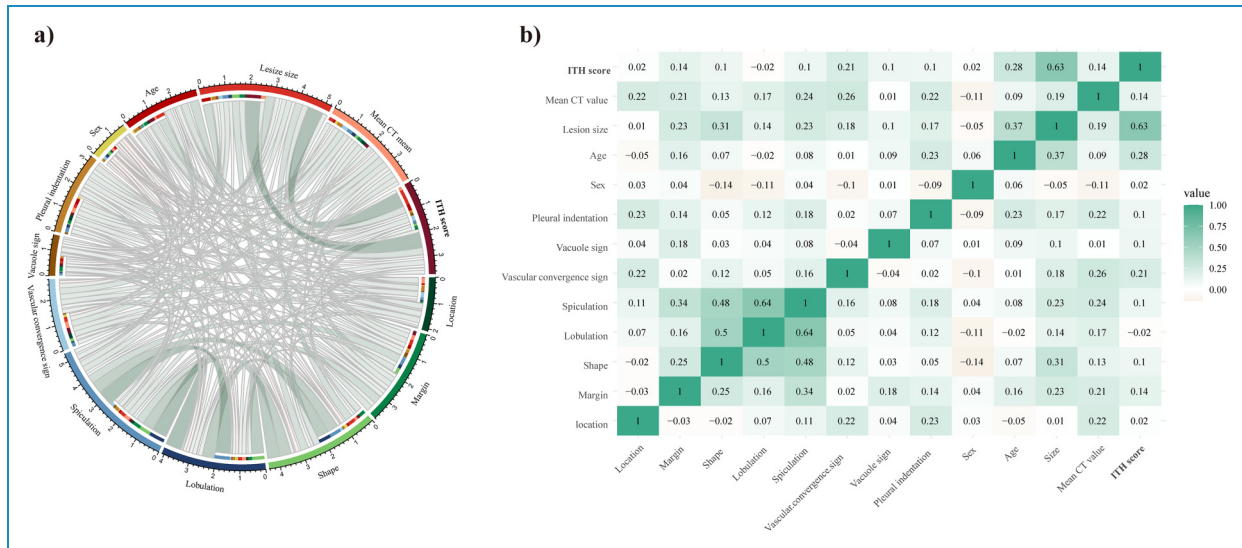
The ITH-ClinRad-guided CatBoost classifier integrates the ITH score and ClinRad for IAC identification. CatBoost combines categorical and boosting techniques, analyzes categorical features by calculating category frequencies, and incorporates hyperparameters to generate new numerical features. It utilizes oblivious trees as the base learner with fewer parameters, support for categorical variables, and use of combination categories to exploit feature relationships, thereby significantly enhancing feature dimensionality. By implementing sorted boosting, CatBoost reduces noise in the training process, prevents gradient bias and prediction shift issues, reduces overfitting, and improves the algorithm accuracy and generalization capabilities.<sup>30,31</sup>

This study investigated the potential of further leveraging the ITH score to enhance the prediction of the invasiveness of adenocarcinomas presenting with pGGNs. When the input variables solely consisted of ClinRad, excluding the ITH score, the ClinRad-guided CatBoost classifier achieved an

**Table 3.** Diagnostic efficacy across Various diagnosis approaches.

Models	AUC	Accuracy	SEN	SPE	PPV	NPV
ClinRad-guided CatBoost Classifier	0.830 (0.764–0.896)	0.728 (0.725–0.730)	0.932 (0.857–1.000)	0.649 (0.562–0.737)	0.506 (0.397–0.615)	0.961 (0.918–1.004)
Lesion size	0.826 (0.758–0.894)	0.747 (0.744–0.749)	0.886 (0.793–0.980)	0.693 (0.608–0.778)	0.527 (0.413–0.641)	0.940 (0.890–0.991)
ITH score	0.846 (0.787–0.905)	0.759 (0.757–0.762)	0.909 (0.824–0.994)	0.702 (0.618–0.786)	0.541 (0.427–0.654)	0.952 (0.907–0.998)
ITH-ClinRad-guided CatBoost Classifier	0.871 (0.818–0.924)	0.766 (0.764–0.768)	0.932 (0.857–1.000)	0.702 (0.618–0.786)	0.547 (0.434–0.659)	0.964 (0.924–1.004)

**Abbreviation:** AUC, area under the curve; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; ClinRad, clinical-radiological features; CatBoost, categorical boosting; ITH, intratumor heterogeneity.



**Figure 7.** Correlation between the ITH score and clinRad, including a chord diagram (a) and matrix plot (b).

AUC of 0.830 (95% CI: 0.764–0.896), which was lower than the AUC of 0.871 (95% CI: 0.818–0.924) obtained by the ITH-ClinRad-guided CatBoost classifier. Furthermore, to deepen our understanding of the relationship between the ITH score and ClinRad, we utilized correlation chord diagrams and matrix plots and found that an increased ITH score was associated with various radiographic features of malignancy levels.

This study has several limitations. First, the retrospective nature of the study may have introduced potential selection bias because the analysis was limited to patients with accessible postoperative pathological findings. Second, the calculation of the ITH score can be enhanced further by considering the entire tumor volume rather than relying solely on the most extensive CT slice. Finally, the findings indicate a correlation between the ITH score and invasiveness of lung adenocarcinoma presenting as pGGNs. Exploring the broad clinical applications of the ITH score, including its significance in selecting surgical plans, predicting patient outcomes, and potentially generalizing it across different cancer types could provide valuable insights.

In conclusion, this study focused on the preoperative prediction of the invasiveness of pulmonary adenocarcinomas presenting as pGGNs with promising results. The ITH-ClinRad-guided CatBoost classifier, ClinRad-guided CatBoost classifier, ITH score, and lesion size exhibited satisfactory performances in predicting IAC in pulmonary adenocarcinomas with pGGNs. Notably, the ITH-ClinRad-guided CatBoost classifier, which combines the ITH score and ClinRad using the CatBoost algorithm, may potentially aid physicians in formulating comprehensive treatment strategies.

**Acknowledgments:** We would like to extend our sincere appreciation to Editage for their meticulous language editing

services, which have significantly enhanced the quality and readability of this manuscript.

**Contributorship:** (I) Conception and design: Hong Zheng, Zhichao Zuo; (II) Administrative support: Wei Chen, Zhichao Zuo; (III) Provision of study materials or patients: Hong Zheng, Wanyin Qi, Haibo Liu; (IV) Collection and assembly of data: Wanyin Qi, Haibo Liu; (V) Data analysis and interpretation: Hong Zheng, Zhichao Zuo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors

**Consent to participate:** Given the retrospective nature of the study, the requirement for written informed consent was waived.

**Consent for publication:** Not applicable.

**Data availability:** The data for this study are available by contacting the corresponding author upon reasonable request.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval:** Ethical approval for this multicenter population-registry cancer study was obtained from the institutional ethics committees of the participating institutions (project approval number: No. 2021-07-009, No. KY2020147).

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Scientific Research Project of Hunan Provincial Health Commission, Changsha, China (Grant Number: B202309017964).

**Guarantor:** Wei Chen(weichen.work@hotmail.com)

**ORCID iD:** Zhichao Zuo  <https://orcid.org/0009-0004-4824-6138>

**Supplemental material:** Supplemental material for this article is available online.

## References

- National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395–409.
- Mazzone PJ and Lam L. Evaluating the patient with a pulmonary nodule: a review. *JAMA* 2022; 327: 264–273.
- Kobayashi Y, Ambrogio C and Mitsudomi T. Ground-glass nodules of the lung in never-smokers and smokers: clinical and genetic insights. *Transl Lung Cancer Res* 2018; 7: 487–497.
- Kobayashi Y, Mitsudomi T, Sakao Y, et al. Genetic features of pulmonary adenocarcinoma presenting with ground-glass nodules: the differences between nodules with and without growth. *Ann Oncol* 2015; 26:156–161.
- Lin MW, Su KY, Su TJ, et al. Clinicopathological and genomic comparisons between different histologic components in combined small cell lung cancer and non-small cell lung cancer. *Lung Cancer* 2018; 125: 282–290.
- Lee GD, Park CH, Park HS, et al. Lung adenocarcinoma invasiveness risk in pure ground-glass opacity lung nodules smaller than 2 cm. *Thorac Cardiovasc Surg* 2019; 67: 321–328.
- Suzuki K, Koike T, Asakawa T, et al. A prospective radiological study of thin-section computed tomography to predict pathological noninvasiveness in peripheral clinical IA lung cancer (Japan clinical oncology group 0201). *J Thorac Oncol* 2011; 6: 751–756.
- Cohen JG and Ferretti GR. Pure ground-glass nodules: are they really indolent? *J Thorac Dis* 2017; 9: 2839–2842.
- Qi L, Xue K, Li C, et al. Analysis of CT morphologic features and attenuation for differentiating among transient lesions, atypical adenomatous hyperplasia, adenocarcinoma in situ, minimally invasive and invasive adenocarcinoma presenting as pure ground-glass nodules. *Sci Rep* 2019; 9: 14586.
- Jin X, Zhao SH, Gao J, et al. CT Characteristics and pathological implications of early stage (T1N0M0) lung adenocarcinoma with pure ground-glass opacity. *Eur Radiol* 2015; 25: 2532–2540.
- Yotsukura M, Asamura H, Motoi N, et al. Long-Term prognosis of patients with resected adenocarcinoma in situ and minimally invasive adenocarcinoma of the lung. *J Thorac Oncol* 2021; 16: 1312–1320.
- Watanabe Y, Hattori A, Nojiri S, et al. Clinical impact of a small component of ground-glass opacity in solid-dominant clinical stage IA non-small cell lung cancer. *J Thorac Cardiovasc Surg* 2022; 163: 791–801. e4.
- Van Schil PE, Asamura H, Rusch VW, et al. Surgical implications of the new IASLC/ATS/ERS adenocarcinoma classification. *Eur Respir J* 2012; 39: 478–486.
- Ito H, Nakayama H, Murakami S, et al. Does the histologic predominance of pathological stage IA lung adenocarcinoma influence the extent of resection? *Gen Thorac Cardiovasc Surg* 2017; 65: 512–518.
- Adams SJ, Stone E, Baldwin DR, et al. Lung cancer screening. *Lancet* 2023; 401: 390–408.
- Fu F, Zhang Y, Wang S, et al. Computed tomography density is not associated with pathological tumor invasion for pure ground-glass nodules. *J Thorac Cardiovasc Surg* 2021; 162: 451–459. e3.
- Nelson DB, Godoy MCB, Benveniste MF, et al. Clinicoradiographic predictors of aggressive biology in lung cancer with ground glass components. *Ann Thorac Surg* 2018; 106: 235–241.
- Zuo Z, Wang P, Zeng W, et al. Measuring pure ground-glass nodules on computed tomography: assessing agreement between a commercially available deep learning algorithm and radiologists' readings. *Acta Radiol* 2023; 64: 1422–1430.
- Antonoff MB. The search for reliable markers of ground glass opacity prognosis: the truth remains largely unknown. *J Thorac Cardiovasc Surg* 2018; 156: 814–815.
- 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: 228–243.
- Martin MD, Kanne JP, Broderick LS, et al. Lung-RADS: pushing the limits. *Radiographics* 2017; 37:1975–1993.
- de Sousa VML and Carvalho L. Heterogeneity in lung cancer. *Pathobiology* 2018; 85: 96–107.
- Li J, Qiu Z, Zhang C, et al. ITHscore: comprehensive quantification of intra-tumor heterogeneity in NSCLC by multi-scale radiomic features. *Eur Radiol* 2023; 33:893–903.
- van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017; 77: e104–e107.
- Zhang J, Sha J, Liu W, et al. Quantification of Intratumoral Heterogeneity: Distinguishing Histological Subtypes in Clinical T1 Stage Lung Adenocarcinoma Presenting as Pure Ground-Glass Nodules on Computed Tomography. *Acad Radiol* 2024.
- Li Y, Wu X, Yang P, et al. Machine learning for lung cancer diagnosis, treatment, and prognosis. *Genomics Proteomics Bioinformatics* 2022; 20: 850–866.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011; 12:2825–2830.
- Obuchowski NA and Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol* 2018; 63:07TR01.
- Lundberg SM and Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp.4765–4774.
- Prokhorenkova L, Gusev G, Vorobev A, et al. Catboost: unbiased boosting with categorical features. In: Proceedings of the 31st International Conference on Neural Information Processing Systems., 2018, pp.6639–6649.
- Ye G, Wu G, Li K, et al. Development and validation of a deep learning radiomics model to predict high-risk pathologic pulmonary nodules using preoperative computed tomography. *Acad Radiol* 2024; 31:1686–1697.