

RESEARCH

Open Access



^{18}F -FDG PET/CT-based deep learning models and a clinical-metabolic nomogram for predicting high-grade patterns in lung adenocarcinoma

Yue Guo^{1†}, Xibin Jia^{2†}, Chuanxu Yang², Chao Fan², Hui Zhu¹, Xu Chen¹ and Fugeng Liu^{1*}

Abstract

Background To develop and validate deep learning (DL) and traditional clinical-metabolic (CM) models based on ^{18}F -FDG PET/CT images for noninvasively predicting high-grade patterns (HGPs) of invasive lung adenocarcinoma (LUAD).

Methods A total of 303 patients with invasive LUAD were enrolled in this retrospective study; these patients were randomly divided into training, validation and test sets at a ratio of 7:1:2. DL models were trained and optimized on PET, CT and PET/CT fusion images, respectively. CM model was built from clinical and PET/CT metabolic parameters via backwards stepwise logistic regression and visualized via a nomogram. The prediction performance of the models was evaluated mainly by the area under the curve (AUC). We also compared the AUCs of different models for the test set.

Results CM model was established upon clinical stage (OR: 7.30; 95% CI: 2.46–26.37), cytokeratin 19 fragment 21 - 1 (CYFRA 21-1, OR: 1.18; 95% CI: 0.96–1.57), mean standardized uptake value (SUV_{mean}, OR: 1.31; 95% CI: 1.17–1.49), total lesion glycolysis (TLG, OR: 0.994; 95% CI: 0.990–1.000) and size (OR: 1.37; 95% CI: 0.95–2.02). Both the DL and CM models exhibited good prediction efficacy in the three cohorts, with AUCs ranging from 0.817 to 0.977. For the test set, the highest AUC was yielded by the CT-DL model (0.895), followed by the PET/CT-DL model (0.882), CM model (0.879) and PET-DL model (0.817), but no significant difference was revealed between any two models.

Conclusions Deep learning and clinical-metabolic models based on the ^{18}F -FDG PET/CT model could effectively identify LUAD patients with HGP. These models could aid in treatment planning and precision medicine.

Clinical trial number Not applicable.

Keywords Positron emission tomography-computed tomography, Fluorodeoxyglucose F18, Lung adenocarcinoma, Deep learning, Histological subtypes

[†]Yue Guo and Xibin Jia contributed equally to this work.

*Correspondence:

Fugeng Liu
liufugengbjhmoh@163.com

¹Department of Nuclear Medicine, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, PR China

²Faculty of Information Technology, Beijing University of Technology, Beijing, China



Background

Lung adenocarcinoma (LUAD) is the most prevalent histological subtype of lung cancer [1]. LUAD has strong morphological heterogeneity, with invasive nonmucinous adenocarcinoma being divided into lepidic, acinar, papillary, micropapillary and solid types [2]. The micropapillary and solid subtypes are often associated with a worse prognosis, even when only a small amount are present [3]. In addition, non-traditional cribriform and complex glandular patterns also indicate poorer outcomes [4, 5]. Identifying high-grade patterns (HGPs) at the time of diagnosis is highly important for therapeutic strategies and precision medicine. In clinical practice, the confirmation of HGP is usually based on surgical samples. Patients who have unresectable LUAD or those who are intolerant to surgery rely on invasive percutaneous lung or endobronchial biopsies. However, the amount of biopsy tissue is so limited that HGPs are sometimes missed. Thus, a noninvasive and practical method for pretreatment prediction of HGP in LUAD patients is needed.

Deep learning (DL) technique has emerged as a novel and efficient tool in lung cancer screening, diagnosis and prognosis in recent years [6, 7]. Mostly using convolutional neural networks (CNNs), DL builds deep networks to continuously learn from medical image data and achieve both feature selection and model fitting. Some researchers have used DL techniques to predict the invasiveness risk, gene mutation status and prognosis of patients with lung cancer, the results of which are satisfactory [8–10].

¹⁸F-Fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT), as non-invasive multimodality imaging equipment, provides metabolic and anatomical information concurrently. The maximum standardized uptake value (SUVmax) is significantly associated with histological grade [11] and is higher in LUAD patients with micropapillary and solid contents [12]. Another study demonstrated that ¹⁸F-FDG PET/CT radiomics signatures could achieve promising prediction efficacy in identifying the presence of micropapillary and solid components in LUAD [13].

Recently, the integration of the DL technique in PET/CT has drawn the attention of several researchers, which were used to discriminate LUAD from lung squamous cell carcinoma [14] or predict lung cancer progression and survival [15], epidermal growth factor receptor mutations [16] and cell proliferation [17]. However, no PET/CT-based deep learning model for predicting HGP in LUAD patients has been reported. In this study, we aimed to develop and validate DL models based on ¹⁸F-FDG PET/CT images to predict HGP in LUAD patients. We also aimed to establish and validate a traditional predictive model based on clinical and PET/CT metabolic

parameters (CM model) and compare the predictive efficiency of the two types of models.

Methods

Patients

Consecutive individuals with lung nodules or masses who underwent diagnostic ¹⁸F-FDG PET/CT scanning between October 2017 and March 2022 at Beijing Hospital were retrospectively recruited first. The inclusion criteria were as follows: (1) histopathologically confirmed LUAD via surgical resection, and the subtypes were detailed in pathological reports; (2) histopathologically confirmed LUAD via biopsy specimens with HGP (the absence of HGP in did not indicate the entire tumor lacked HGP); and (3) all the clinical data needed in the study were available. The exclusion criteria were as follows: (1) patients who received any antineoplastic treatment or invasive procedure before PET/CT; (2) patients whose time interval between the PET/CT scan and pathological examination exceeded 4 weeks; and (3) patients whose image quality was not good enough or unsuitable for region of interest (ROI) delineation, such as severe respiratory artefacts or inflammation-like tumours. HGP, including micropapillary and solid subtypes, which are recognized by the WHO, as well as cribriform and complex glandular patterns, which are nontraditional but have poor prognoses similar to those of the abovementioned two histologic subtypes, were recorded from retrospective reviews of pathological reports. Finally, 303 patients were enrolled in the study and randomly divided into a training set ($n=208$), a validation set ($n=29$) and a test set ($n=66$) at a ratio of 7:1:2. The training set was used for model training, the validation set was used for model hyperparameter adjustment, and the test set was used for assessing model performance. The recruitment process is illustrated in the flowchart (Fig. 1), and the comparisons among the three sets are detailed in Table 1. Potential clinical risk factors, such as age, sex, smoking history, carcinoembryonic antigen (CEA), cytokeratin 19 fragment 21-1 (CYFRA 21-1) and clinical stage (according to the eighth edition of the International Association for the Study of Lung Cancer Staging System), were collected from the patients' medical records.

PET/CT image acquisition

Before intravenous injection of 5.18 MBq/kg ¹⁸F-FDG, each patient fasted for at least 4 h with a blood glucose level less than 11.1 mmol/L. Sixty minutes later, patients underwent PET/CT (Biograph mCT, Siemens Healthcare and Vereos digital PET/CT, Philips Medical Systems) scanning. Acquisition began with a spiral CT scan (120 kV, automatic mA or 100-mA tube current, and 3-mm layer thickness) without contrast medium from the skull base to the upper femur in the supine position,

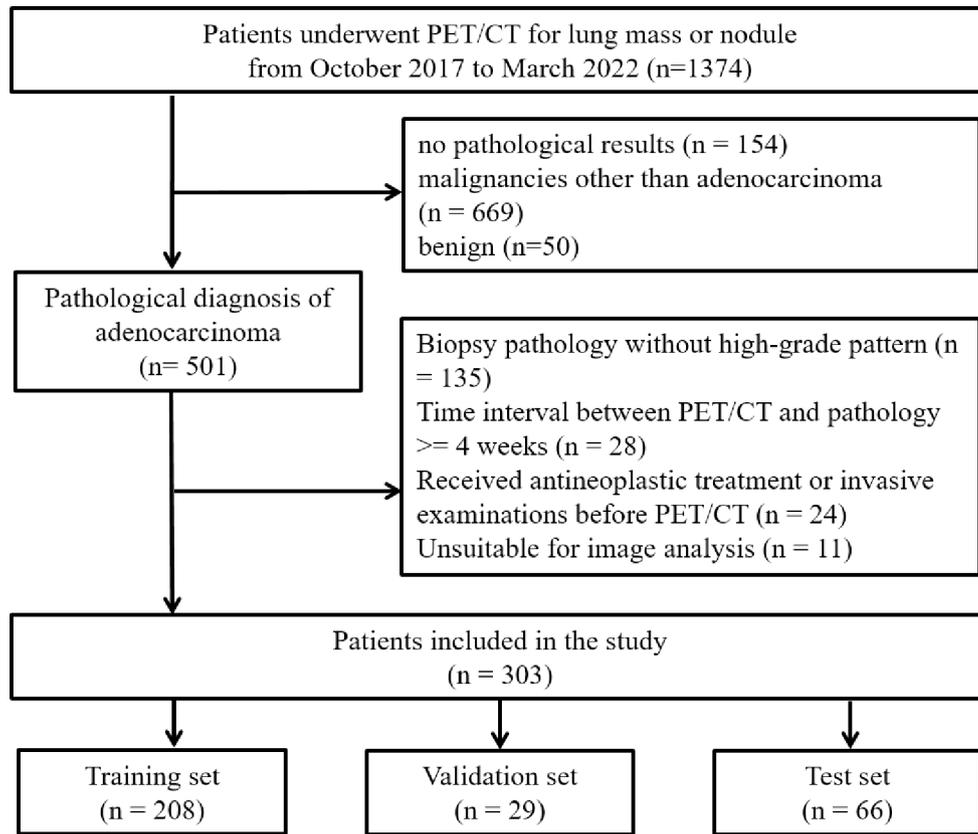


Fig. 1 Flowchart of patient selection

Table 1 Comparison of demographic and metabolic features between training set and validation or test set

	Training set (n=208)	Validation set (n=29)	Test set (n=66)	Total (n=303)	p1-value	p2-value
Sex					0.2763	0.0490
Male	94 (45.19%)	10 (34.48%)	39 (59.09%)	143 (47.19%)		
Female	114 (54.81%)	19 (65.52%)	27 (40.91%)	160 (52.81%)		
Age	66.0 (58.8;75.0)	64.0 (60.0;68.0)	66.0 (60.0;75.5)	66.0 (59.0;74.0)	0.0682	0.6666
HGP					0.6792	0.5975
Absence	80 (38.46%)	10 (34.48%)	23 (34.85%)	113 (37.29%)		
Presence	128 (61.54%)	19 (65.52%)	43 (65.15%)	190 (62.71%)		
Smoking Status					0.0828	0.0704
Never	139 (66.83%)	24 (82.76%)	36 (54.55%)	199 (65.68%)		
Ever/current	69 (33.17%)	5 (17.24%)	30 (45.45%)	104 (34.32%)		
Clinical Stage					0.6129	0.2184
I~II	125 (60.10%)	16 (55.17%)	34 (51.52%)	175 (57.76%)		
III~IV	83 (39.90%)	13 (44.83%)	32 (48.48%)	128 (42.24%)		
CEA	3.5 (2.1;10.0)	4.2 (1.7;10.0)	3.8 (2.0;8.8)	3.6 (2.0;9.9)	0.8782	0.8950
CYFRA21-1	2.7 (2.1;3.9)	2.6 (2.2;4.7)	3.4 (2.5;4.8)	2.8 (2.2;4.4)	0.7049	0.0197
SUVmax	9.0 (3.0;16.3)	8.6 (3.6;16.1)	13.3 (5.3;18.6)	9.5 (3.1;16.7)	0.9424	0.0735
SUVmean	5.3 (1.7;10.0)	5.0 (2.0;9.2)	7.8 (3.0;10.5)	5.5 (1.8;10.1)	0.9481	0.0831
MTV	2.9 (1.4;7.1)	4.2 (2.8;7.6)	6.0 (2.7;15.5)	3.6 (1.9;8.7)	0.0828	0.0001
TLG	11.9 (4.2;44.6)	22.0 (7.3;66.2)	35.1 (14.2;155.4)	16.1 (5.4;65.8)	0.2359	0.0002
Size (cm)	2.3 (1.5;3.3)	2.5 (2.2;3.6)	2.8 (2.1;4.1)	2.5 (1.7;3.5)	0.1639	0.0069

HGP high-grade patterns, CEA carcinoembryonic antigen, CYFRA21-1 Cytokeratin 19 Fragment, SUVmax maximum standardized uptake value, SUVmean mean standardized uptake value, MTV metabolic tumor volume, TLG total lesion glycolysis, p1-value for comparing the training set with the validation set, p2-value for comparing the training set with the test set

immediately followed by a PET scan (2 min per bed position for 5~7 beds) in three-dimensional list mode with the same coverage. The PET data were corrected for attenuation by CT images and reconstructed via a Gaussian filter with an ordered subset expectation-maximization algorithm (2 iterations, 20 subsets for mCT and 3 iterations, 8 subsets for Vereos).

PET/CT image interpretation and tumor delineation

A nuclear medicine physician (GY with 15 years of experience) completed the two tasks. When encountering uncertain situations, a senior physician (LFG with 30 years of experience) delineated the lesion. Both doctors were blinded to the participants' clinical and pathological information. By semiautomatically drawing the volume of interest (VOI) on reconstructed PET images, metabolic parameters, including the SUVmax, mean standardized uptake value (SUVmean), metabolic tumor volume (MTV) and total lesion glycolysis (TLG), were calculated using a threshold of 40% of the SUVmax. In cases where the delineation was not very accurate, minor manual adjustments were implemented. On the axial CT image, the size of the lesion was measured, which was represented by the longest diameter (in cm) on the lung window.

All the participants' PET/CT images were subsequently exported from the workstation in Digital Imaging and Communications in Medicine format and imported into the open-source software 3D Slicer (version 5.2.2, <http://www.slicer.org>). Lesions were manually segmented on both PET and CT images slice by slice. Each annotated tumor was labelled with the presence or absence of HGP. Finally, the segmentation image files in the Neuroimaging Informatics Technology Initiative format were exported.

Data preprocessing

In the training and validation sets, we first identify the slices where the lesion is located from the two-dimensional (2D) PET and CT images according to the manually annotated segmentation labels. For the sake of obtaining more training samples, we selected five slices containing the largest tumor area instead of one slice to represent each case, which was used as a model input. In this way, 1040 and 145 slices were obtained in the training and validation set, respectively. In view of the classification target and removing the impact of large surroundings, we cropped each slice image along the lesion and resized it to a fixed input dimension requested by the deep learning network. The selection and cropping procedure is shown in Fig. 2. In addition, data augmentation was performed, including random rotation, mirroring, flipping, color enhancement, and Gaussian blurring, to increase data diversity. In this way, five times more training data were obtained. In the test set, we used only one slice with the largest lesion size to represent each patient. Then, the same cropping and resizing processes were performed as in the training and validation sets.

Deep learning model construction

Considering the training sample size as well as the representation learning capability, the ResNet-18 residual network is finally adopted as the feature extractor. Additionally, to alleviate overfitting and increase the robustness of the DL model, the dropout technique was employed. The network architecture is shown in Fig. 3. The network consists of a 7×7 convolution layer, a 2×2 pooling layer, two ResNet Block1 layers, three ResNet Block2 layers and a fully connected layer. For the cropped input images, a 7×7 convolution kernel was employed for feature extraction, and downsampling was conducted because of the large receptive field of neighboring pixels. Then, the low-level feature maps of the transformed

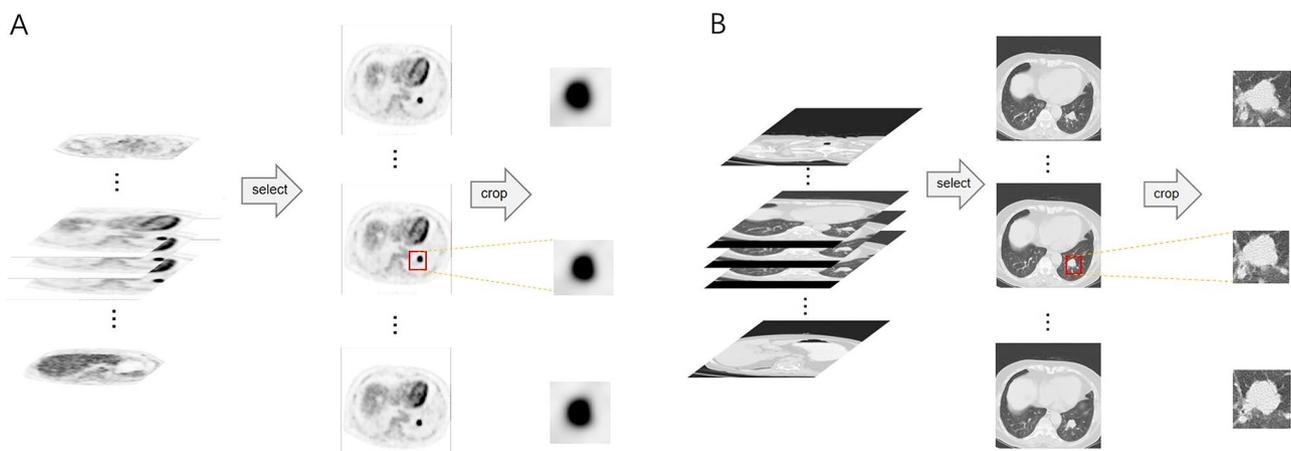


Fig. 2 Image preprocessing—selecting and cropping procedures. (A) PET images. (B) CT images.

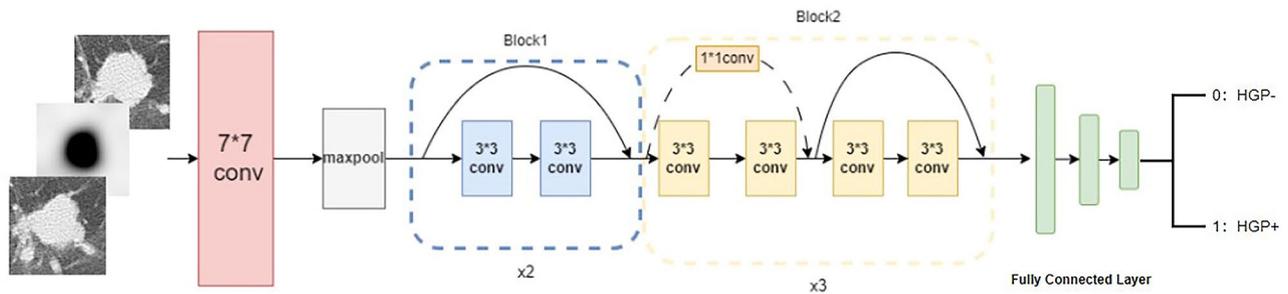


Fig. 3 Architecture of ResNet-18 network

image passed several ResNet blocks to obtain the category discriminant high-level features. Finally, the extracted features are fed into a fully connected layer for two-category classification. Compared with Block 1, Block 2 contains an additional 1×1 -convolution layer used to match the channel sizes of the input and output. Supposing that the input image is x and the extracted features are f , the above feature extraction processing of the deep learning model can be formulated as Eq. 1:

$$f = \text{block2}_{*3} (\text{block1}_{*2} (\text{maxpool} (\text{conv}_{7*7}(x)))) \quad (1)$$

The classification performed by the fully connected layer on the extracted convolutional features f is formulated in Eq. 2:

$$R_{cls} = \text{sigmoid}(fc(\text{reshape}(f))) \quad (2)$$

As shown in Eq. 2, the output features f of the convolutional network were first transformed to match the input dimension of the fully connected layer and then passed through the sigmoid function to obtain the final classification results.

ResNet-18 was optimized by the Adam optimizer with an initial learning rate of $1e-3$, which decays by 0.5 times every 10 epochs. Binary cross-entropy loss was used as the loss function, which can be represented as:

$$BCE(y, \hat{y}) = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y}))$$

In this function, y denotes the real label of each object (0 represents HGP negative, whereas 1 represents HGP positive), and \hat{y} represents the predicted probability of the DL model. Three DL models were trained on PET alone, CT alone and PET/CT fusion images, and the outputs were the PET-DL model, CT-DL model and PET/CT-DL model, respectively.

Establishment of clinical-metabolic nomogram prediction model

A nomogram prediction model was also built based on the training set, with the clinical and metabolic variables

selected via backwards stepwise logistic regression. The best model and included variables were determined by the Akaike information criterion (AIC) values. The CM model was subsequently visualized in the form of a nomogram.

Statistical analysis

R software (version 4.3.1, Vienna, Austria; URL <https://www.R-project.org/>) was used for statistical analysis. Continuous variables with a normal distribution are expressed as the means \pm standard deviations, whereas other variables are expressed as medians (interquartile ranges). Categorical variables are presented as numbers (percentages). Comparisons of continuous variables were conducted via Student's t-test or the Wilcoxon/Kruskal-Wallis rank sum test, and categorical variables were analysed via Pearson's chi-square test or Fisher's exact test. The "MASS" package was used to select optimal variables and establish a predictive model, with odds ratios (ORs) with 95% confidence intervals (CIs) being calculated. The nomogram was drawn via the "rms" package. For the performance assessment, we used mainly the receiver operating characteristic (ROC) curve and the area under the curve (AUC). Furthermore, we calculate the F1 score, accuracy, sensitivity and precision of different models to reflect their prediction ability more comprehensively. A Delong test was conducted to compare the AUCs of the DL models and CM model in the test set. Decision curve analysis (DCA) was performed to assess the clinical utility by quantifying the net benefits under different threshold probabilities for the test set. A two-sided p value less than 0.05 was interpreted as statistically significant.

Results

Patient characteristics

Among the 303 LUAD patients, 143 (47.2%) were males and 160 (52.8%) were females, with ages ranging from 27 to 88 years (median: 66.0 years); 199 (65.7%) were non-smokers, whereas 104 (34.3%) were former or current smokers. With respect to the clinical stage, 175 (57.8%) patients were classified as early stage (stage I or II), and 128 were classified as advanced stage (III or IV). In

addition, 190 (62.7%) and 113 (37.3%) patients had LUAD with or without HGP, respectively. Among the training, validation and test sets, the HGP positivity rate was not significantly different, with a p value of 0.8236. Table 2 summarizes the demographic and metabolic characteristics of the patients. The metabolic parameters showed no significant difference between the two PET/CT scanners in HGP positive patients, with all p values > 0.05 (Table S1).

Prediction performance of DL models

To obtain the best prediction efficacy, the dropout rate of the fully connected layer was adjusted based on the CT images of the test set. The highest AUC value in the test set was achieved when the dropout rate was set at 0.55. This algorithm was then used for training PET and PET/CT fusion images, and the performance metrics are displayed in Fig. 4. In the training set, the PET/CT model achieved the highest AUC (0.890), F1 score (0.776) and sensitivity (0.873), whereas the PET model yielded the highest accuracy (0.803) and precision (0.892). In the validation set, the highest AUC, F1 score, accuracy and sensitivity were generated by the PET/CT model, with corresponding values of 0.977, 0.894, 0.901 and 0.986, respectively. The PET model exhibited the best precision (0.913). In the test set, the CT model was at the top of the AUC and sensitivity lists, with values of 0.895 and 0.920, respectively, the PET model had the highest accuracy (0.789) and precision (0.837), while the PET/CT model contributed the highest F1 score (0.752). For further validating the robustness and reliability of the proposed DL

models, we conducted 5-fold cross-validation in training and test datasets. PET/CT-DL model showed the highest mean AUC (0.908), accuracy (0.850) and sensitivity (0.864) while PET-DL model demonstrated the highest F1 score (0.876) and specificity (0.934). The results were detailed in Supplementary Table S2. In addition, we applied Grad-CAM [18] in four patients in the test set to identify the key image features contributing to prediction, which was displayed in Fig. 5.

Establishment and performance assessment of the CM model

After backwards stepwise regression selection, five variables were ultimately included in the predictive model: clinical stage (OR: 7.30; 95% CI: 2.46–26.37), CYFRA 21-1 (OR: 1.18; 95% CI: 0.96–1.57), SUVmean (OR: 1.31; 95% CI: 1.17–1.49), TLG (OR: 0.994; 95% CI: 0.990–1.000) and size (OR: 1.37; 95% CI: 0.95–2.02). The AIC values of the initial and final models were 190.1 and 180.74, respectively. The details are shown in Table 3.

The CM model was visualized via a nomogram (Fig. 6). The performance metrics (Fig. 4) revealed good predictive ability, with AUCs of 0.893 (95% CI: 0.848~0.938), 0.916 (95% CI: 0.857~1.000) and 0.879 (95% CI: 0.797~0.961) in the training, validation and test sets, respectively. The F1 score, accuracy, sensitivity and precision of the CM model in the three cohorts are also displayed in Fig. 4.

Table 2 Demographic and metabolic characteristics of included LUAD patients

Variables	Training set (n=208)			Validation set (n=29)			Test set (n=66)		
	HGP (+) n=128	HGP (-) n=80	p-value	HGP (+) n=19	HGP (-) n=10	p-value	HGP (+) n=43	HGP (-) n=23	p-value
Age (years)	67.0 (60.0, 76.0)	64.0 (58.0, 73.2)	0.1811	62.4 ± 7.4	63.2 ± 10.6	0.8067	68.9 ± 12.4	63.4 ± 8.1	0.0360
Sex			0.0195			0.4137			0.0592
Male	66 (70.21%)	28 (29.79%)		8 (80.00%)	2 (20.00%)		29 (74.36%)	10 (25.64%)	
Female	62 (54.39%)	52 (45.61%)		11 (57.89%)	8 (42.11%)		14 (51.85%)	13 (48.15%)	
Smoking Status			0.0478			0.6328			
Never	79 (56.83%)	60 (43.17%)		15 (62.50%)	9 (37.50%)		19 (52.78%)	17 (47.22%)	0.0208
Ever/Current	49 (71.01%)	20 (28.99%)		4 (80.00%)	1 (20.00%)		24 (80.00%)	6 (20.00%)	
Clinical Stage			< 0.0001			0.0084			< 0.0001
I~II	50 (40.00%)	75 (60.00%)		7 (43.75%)	9 (56.25%)		14 (41.18%)	20 (58.82%)	
III~IV	78 (93.98%)	5 (6.02%)		12 (92.31%)	1 (7.69%)		29 (90.62%)	3 (9.38%)	
CEA	5.2 (2.5;14.3)	2.4 (1.5;3.8)	< 0.0001	6.5 (2.4;16.4)	2.8 (1.5;4.0)	0.1299	4.6 (2.9;12.1)	2.7 (1.6;3.7)	0.0022
CRFRA21-1	3.2 (2.2;4.6)	2.5 (2.0;2.9)	0.0001	3.7 (2.4;6.1)	2.3 (1.9;2.6)	0.0735	4.2 (2.8;5.8)	2.7 (2.0;3.2)	0.0010
SUVmax	14.0 (8.4;21.0)	2.5 (1.5;6.8)	< 0.0001	14.0 ± 7.7	3.9 ± 2.5	< 0.0001	14.7 (9.1;25.2)	3.1 (2.0;12.7)	< 0.0001
SUVmean	8.3 (4.8;12.3)	1.5 (0.9;4.0)	< 0.0001	7.8 (4.7;10.5)	1.8 (1.2;2.9)	0.0003	8.7 (5.4;14.8)	1.7 (1.1;7.4)	< 0.0001
MTV	3.4 (1.3;8.7)	2.7 (1.8;4.9)	0.2258	5.4 (2.8;8.1)	3.7 (3.0;6.8)	0.6357	7.1 (3.7;16.6)	4.1 (2.2;9.2)	0.0633
TLG	29.3 (9.6;72.2)	4.9 (2.1;12.3)	< 0.0001	44.4 (10.6;71.2)	5.0 (3.6;18.4)	0.0029	70.4 (28.2;194.3)	11.4 (3.0;37.3)	0.0001
Size (cm)	2.8 (1.9;3.6)	1.8 (1.2;2.5)	< 0.0001	2.8 ± 0.8	2.5 ± 1.2	0.4420	3.0 (2.3;4.5)	2.3 (1.6;2.8)	0.0049

HGP high-grade patterns, CEA carcinoembryonic antigen, CYFRA21-1 Cytokeratin 19 Fragment, SUVmax maximum standardized uptake value, SUVmean mean standardized uptake value, MTV metabolic tumor volume, TLG total lesion glycolysis

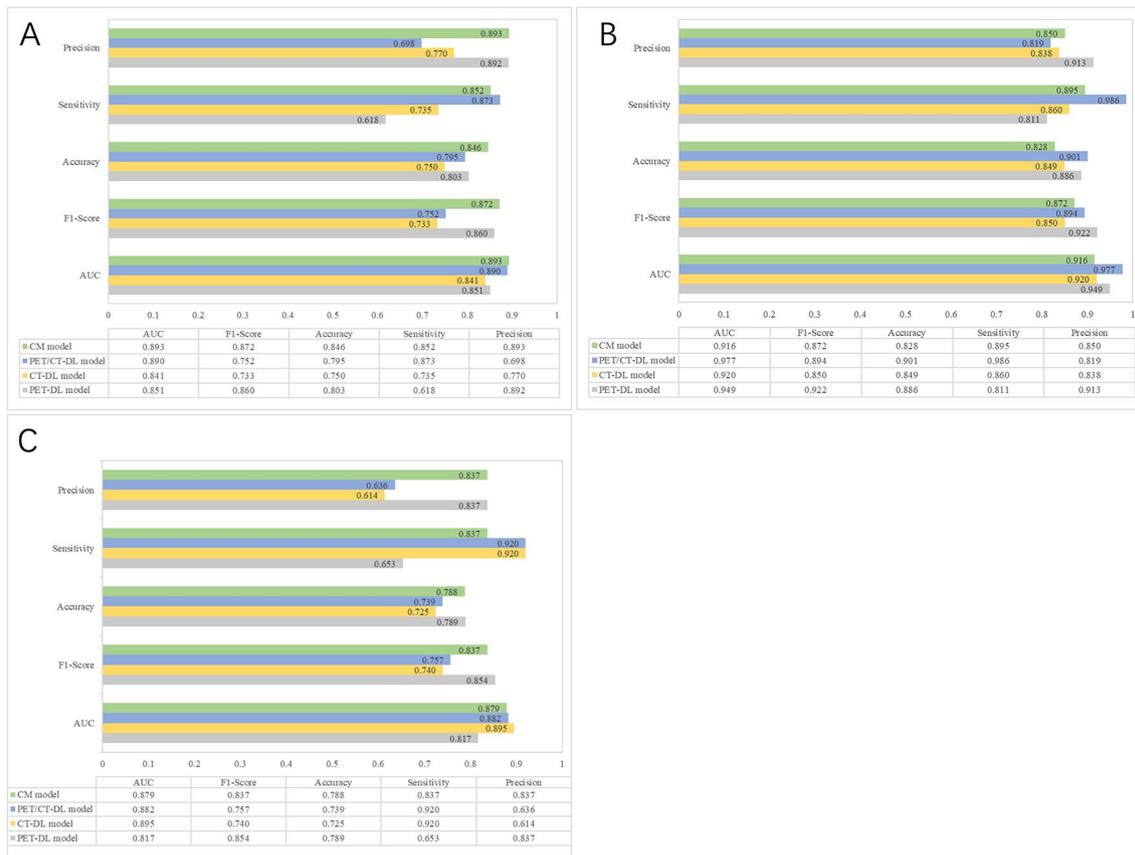


Fig. 4 Predictive performance of the DL and CM models in LUAD. Performance metrics of models for predicting HGP in the (A) training set, (B) validation set, and (C) test set

Comparisons of the predictive efficacy and clinical utility of DL models and CM model for test set

As shown in Fig. 7, the CT-DL model exhibited the best predictive performance in the test set, followed by the PET/CT-DL model, CM model and PET-DL model. However, no significant differences were revealed between any two models through Delong’s test (all $p > 0.05$). The details are displayed in Table 4.

According to the DCA (Fig. 8), all the DL models and CM model had good net benefits when the threshold probability was above 0.1 in the test set, demonstrating the potential clinical utility of the PET/CT-based DL and CM models.

Discussion

LUAD with HGP is often associated with poorer clinical outcomes and needs intensive surveillance strategies [19]. In this study, we established two types of models based on ¹⁸F-FDG PET/CT: one integrated traditional clinical and metabolic parameters, while the other included image-based deep learning models. All the models exhibited good ability to predict HGP, with AUCs of 0.817, 0.895, 0.882 and 0.897 for the PET-DL model, CT-DL model, PET/CT-DL model and CM nomogram model in the test

set, respectively. These results suggest that both the PET/CT-based traditional clinical-metabolic and deep learning predictive methods are reliable in the clinic.

HGP in lung adenocarcinoma tend to be more aggressive in terms of tumor behavior, leading to a high risk of lymph node metastasis and early locoregional recurrence [20]. Research published recently revealed a potential molecular mechanism [21]. The increased expression of FAM83A-AS1 (a type of long noncoding RNA) in HGP increases the degree of malignancy and glycolysis in LUAD via the miR-202-3p/hexokinase II (HK2) axis. This finding explains why LUAD with HGP is associated with increased FDG uptake, which has been reported in some studies [11, 22, 23]. In our study, all the ¹⁸F-FDG PET/CT metabolic parameters were greater in LUAD patients with HGP than in those without HGP, which was in line with the findings of previous studies. Moreover, some clinical characteristics are also related to HGP. Li et al. [24] retrospectively analysed 3100 invasive LUAD samples and reported that tumors with solid or micropapillary components presented higher CEA and CYFRA levels, larger tumor sizes and a greater proportion of males. Another study by Wang et al. [25] revealed that there were more males in the HGP-positive group, as well

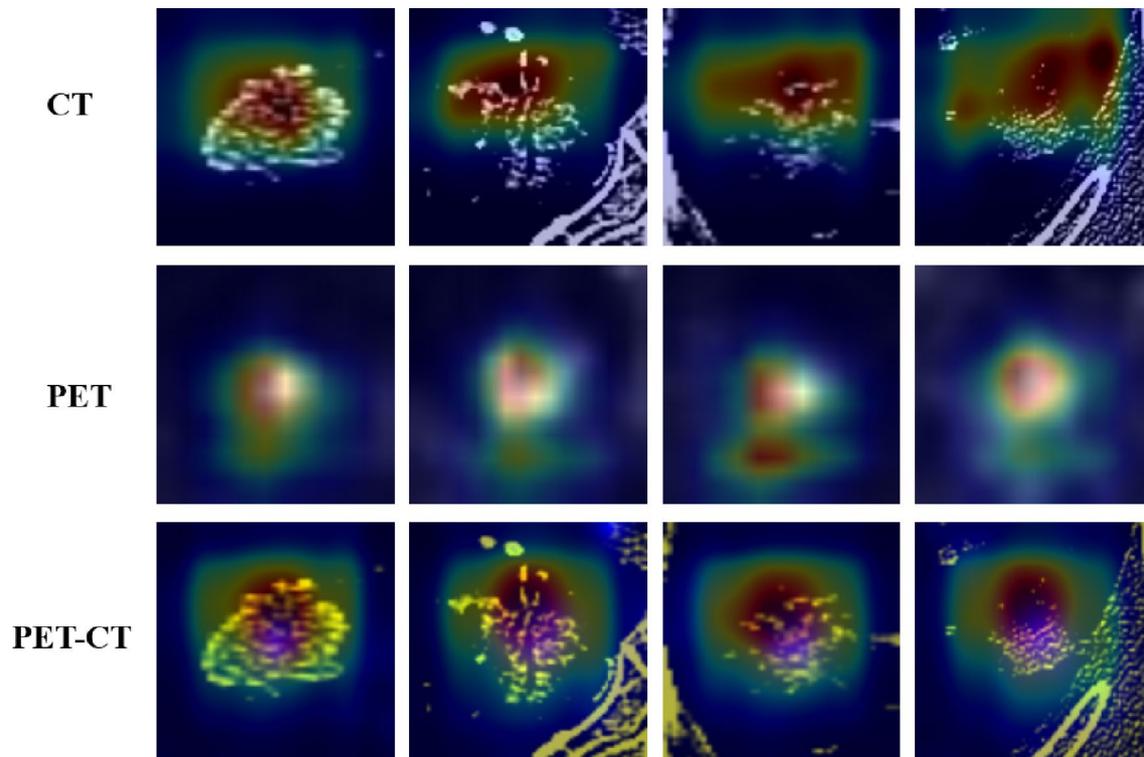


Fig. 5 Grad-CAM of ResNet-18 based DL models in four samples of test set

Table 3 Backward Stepwised regression for CM model

Variables(references)	Initial model (AIC = 190)		Ultimate model (AIC = 180.74)	
	OR (95%CI)	p-value	OR (95%CI)	p-value
Age	1.00 (0.96–1.04)	0.992		
Sex (female)	0.57(0.22–1.41)	0.223		
Smoking status (ever/current)	0.88(0.32–2.41)	0.807		
Clinical stage (III~IV)	6.33 (2.00–24.44)	0.003	7.30 (2.46–26.37)	< 0.001
CEA	1.00 (0.99–1.01)	0.603		
CYFRA211	1.20 (0.96–1.60)	0.190	1.18 (0.96–1.57)	0.200
SUVmax	1.02 (0.56–2.23)	0.951		
SUVmean	1.29 (0.36–3.50)	0.658	1.31 (1.17–1.49)	< 0.001
MTV	1.04 (0.93–1.15)	0.507		
TLG	0.99 (0.98–1.003)	0.103	0.99 (0.98–1.00)	0.04
Size	1.28 (0.81–2.05)	0.300	1.37 (0.95–2.02)	0.096

CM clinical-metabolic, AIC Akaike information criterion, OR odds ratios, CI confidence intervals, CEA carcinoembryonic antigen, CYFRA21-1 Cytokeratin 19 Fragment, SUVmax maximum standardized uptake value, SUVmean mean standardized uptake value, MTV metabolic tumor volume, TLG total lesion glycolysis

as larger lesions and many more smokers. Higher TNM stages are also more common in patients with micropapillary or solid patterns [13]. The abovementioned results were confirmed by our results. After backwards stepwise regression, the model combining clinical and metabolic features ultimately included five variables, namely, clinical stage, CYFRA21-1, SUVmean, TLG and tumor size, yielding excellent prediction performance.

Machine learning-based radiomics, which has the ability to capture first-order to higher-order features from medical images, has been employed for HGP prediction in LUAD patients in recent years. By extracting

radiomic features from PET/CT data, Zhou et al. constructed a radiomic-clinical combined model that could discriminate LUAD patients with HGP from those without HGP well, with AUCs ranging from 0.859~0.880 in the training and two test cohorts [13]. Another PET/CT-based radiomic model from Choi et al. revealed that, in comparison with the conventional predictor SUVmax, the radiomic model demonstrated better performance in identifying aggressive LUAD subtypes, achieving an accuracy of 83.5% and a high negative predictive value of 95.6% [26]. However, radiomic feature extraction and selection could be laborious tasks, which hinders their

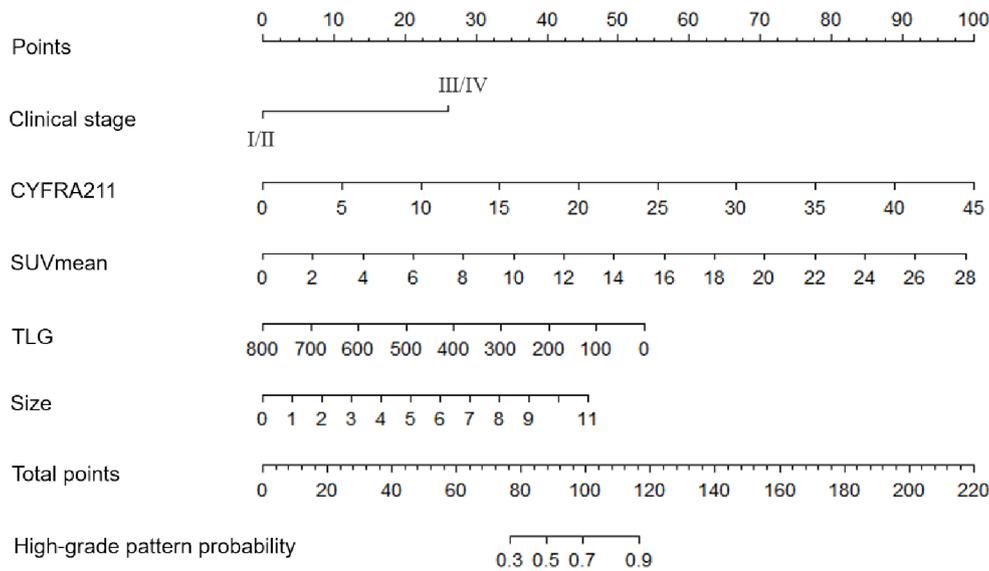


Fig. 6 Nomogram of the CM model for predicting high-grade patterns of lung adenocarcinoma

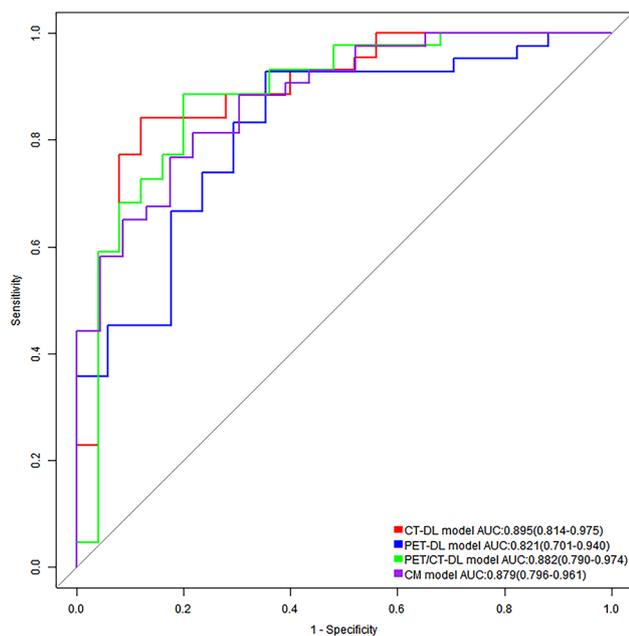


Fig. 7 ROC curves of DL and CM models for predicting high-grade patterns of lung adenocarcinoma in test set

Table 4 Performance comparisons of DL models and CM model for the test set

	Z value	p-value
CT-DL model VS PET-DL model	1.0039	0.318
CT-DL model VS PET/CT-DL model	0.56949	0.569
PET-DL model VS PET/CT-DL model	-0.79473	0.428
CT-DL model VS CM model	0.27056	0.787
PET-DL model VS CM model	-0.78296	0.435
PET/CT-DL model VS CM model	0.050187	0.960

CM clinical-metabolic, DL deep learning

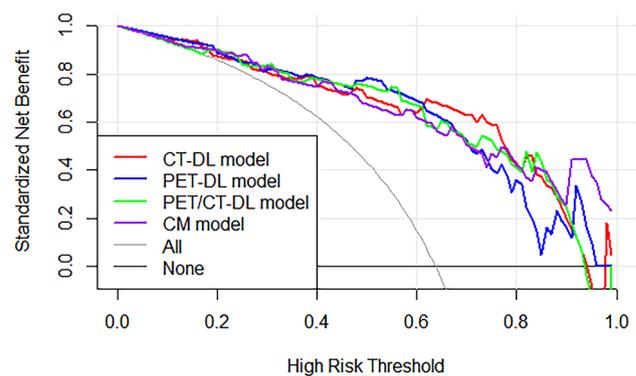


Fig. 8 DCA curves of DL and CM models for predicting high-grade patterns of lung adenocarcinoma in test set

feasibility in the clinic. Deep learning techniques based on convolutional neural networks can discover high-level semantic information and serve as alternatives to hand-crafted radiomic methods [27]. For HGP detection in pathology, deep learning methods have been verified to be of great assistance to pathologists, achieving a precision of 0.775 and a recall of 0.896 for the identification of micropapillary patterns in one study [28] and AUCs of greater than or equal to 0.85 for micropapillary, solid and cribriform pattern diagnosis in another study [1]. Nevertheless, few publications have reported the utilization of deep learning in the prediction of HGP in LUAD patients on the basis of PET/CT. Therefore, in the current study, we adopted parameter-compact ResNet as the primary framework to build prediction models. To avoid overfitting during training and enhance the robustness of the model for test set data, several operations were conducted, including data augmentation and the dropout technique. The satisfactory results implied that the

data-driven deep learning strategy could automatically learn the discriminant category features effectively from CT and PET images and facilitate the classification ability in distinguishing LUAD with HGP from those without HGP.

The comparison results in the test set revealed that the DL models were not significantly better than the CM model in the test cohort. Even in the training cohort, the AUCs of the DL models were lower than those of the CM model, leading us to doubt the advantages of the DL techniques. Nevertheless, it is worth noting that the DL models were established on the basis of images only, whereas the clinic-metabolic model included two clinical characteristics, stage and serum CRFRA21-1, which may contribute to the prediction efficacy. Previous publications have verified that combining classic clinical parameters with DL features could increase the predictive capacity of these methods. Like CT-based DL models for predicting lymph node metastasis in colorectal cancer, the model yielded the highest AUC after clinical factors were integrated into the DL model [29]. Shao et al. also developed several transfer learning-based PET/CT models for the prediction of EGFR mutations in LUAD and reported that the three-stream transfer learning model integrating PET, CT and clinical data was the most promising, with an AUC of 0.833 in the training set and 0.730 in the test set [30]. Although the performance preponderance of the DL models was not remarkable, DL approaches could still offer more confidence in HGP prediction because of their powerful image feature extraction and classification capability. In further studies, we aim to construct models that are more efficient by combining image-based data with clinical features.

There are several limitations of our research. First, as a data-driven computer technology, the sample size was not large enough for the deep learning technique. In addition to the DL-CT model, the DL-PET and DL-PET/CT models both demonstrated better predictive efficacy in the training set than in the test set. This overfitting phenomenon could impact the model generalization performance. Second, this was a single-center study, and the lack of external or prospective validation weakened the reliability of the model to some extent. Finally, the retrospective nature of this study may have led to selection bias. Large-scale and multicenter studies are needed to guarantee model efficiency. In subsequent study, the state-of-the-art CNN networks and radiomics features could be applied to increase the discrimination capacity of HGP in LUAD.

Conclusions

The deep learning and clinic-metabolic models based on ^{18}F -FDG PET/CT showed promising predictive performance in identifying high-grade patterns in lung

adenocarcinoma, especially the deep learning models, which verified their powerful automatic learning and feature recognition abilities. This noninvasive prediction method could further broaden the clinical utility of ^{18}F -FDG PET/CT in lung adenocarcinoma.

Abbreviations

2D	Two-dimensional
AIC	Akaike information criterion
AUC	Areas under the ROC curve
CEA	Carcinoembryonic antigen
CI	Confidence interval
CM	Clinical-metabolic
CYFRA21-1	Cytokeratin 19 fragment
DCA	Decision curve analysis
DL	Deep learning
CT	Computed tomography
FDG	Fluorodeoxyglucose
HGP	High-grade pattern
OR	Odds ratios
PET	Positron emission tomography
LUAD	Lung adenocarcinoma
MTV	Metabolic tumour volume
ROC	Receiver operating characteristic
ROI	Region of interest
SUVmax	Maximum standardized uptake value
SUVmean	Mean standardized uptake value
TLG	Total lesion glycolysis
VOI	Volume-of-interest

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12880-025-01684-3>.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

The study was designed by GY and LFG. Data curation and interpretation were performed by GY, ZH and CX. Data analysis was performed by JXB, YCX and FC. The first draft of the manuscript was written by GY and revised by JXB and LFG. All authors read and approved the final manuscript.

Funding

This study was supported by National High Level Hospital Clinical Research Funding (BJ-2021-186, BJ-2024-190).

Data availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Beijing Hospital (approval no. 2022BJYYEC-KY246-01) and complied with the Declaration of Helsinki. Written informed consent was waived by the Institutional Review Board.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 December 2024 / Accepted: 21 April 2025

Published online: 28 April 2025

References

- Zhao Y, He S, Zhao D, Ju M, Zhen C, Dong Y, et al. Deep learning-based diagnosis of histopathological patterns for invasive non-mucinous lung adenocarcinoma using semantic segmentation. *BMJ Open*. 2023;13(7):e069181.
- Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol*. 2022;17(3):362–87.
- Yanagawa N, Shiono S, Abiko M, Katahira M, Osakabe M, Ogata SY. The clinical impact of solid and micropapillary patterns in resected lung adenocarcinoma. *J Thorac Oncology: Official Publication Int Association Study Lung Cancer*. 2016;11(11):1976–83.
- Kuang M, Shen X, Yuan C, Hu H, Zhang Y, Pan Y, et al. Clinical significance of complex glandular patterns in lung adenocarcinoma. *Am J Clin Pathol*. 2018;150(1):65–73.
- Warth A, Muley T, Kossakowski C, Stenzinger A, Schirmacher P, Dienemann H, et al. Prognostic impact and clinicopathological correlations of the cribriform pattern in pulmonary adenocarcinoma. *J Thorac Oncol*. 2015;10(4):638–44.
- Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective. *Sem Cancer Biol*. 2023;89:30–7.
- de Margerie-Mellon C, Chassagnon G. Artificial intelligence: A critical review of applications for lung nodule and lung cancer. *Diagn Interv Imaging*. 2023;104(1):11–7.
- Huang H, Zheng D, Chen H, Wang Y, Chen C, Xu L, et al. Fusion of CT images and clinical variables based on deep learning for predicting invasiveness risk of stage I lung adenocarcinoma. *Med Phys*. 2022;49(10):6384–94.
- Wang S, Yu H, Gan Y, Wu Z, Li E, Li X, et al. Mining whole-lung information by artificial intelligence for predicting EGFR genotype and targeted therapy response in lung cancer: a multicohort study. *Lancet Digit Health*. 2022;4(5):e309–19.
- Kim H, Goo JM, Lee KH, Kim YT, Park CM. Preoperative CT-based deep learning model for predicting Disease-Free survival in patients with lung adenocarcinomas. *Radiology*. 2020;296(1):216–24.
- Sun X-Y, Chen T-X, Chang C, Teng H-H, Xie C, Ruan M-M, et al. SUVmax of 18FDG PET/CT predicts histological grade of lung adenocarcinoma. *Acad Radiol*. 2021;28(1):49–57.
- Chang C, Sun X, Zhao W, Wang R, Qian X, Lei B, et al. Minor components of micropapillary and solid subtypes in lung invasive adenocarcinoma (≤ 3 cm): PET/CT findings and correlations with lymph node metastasis. *Radiol Med*. 2019;125(3):257–64.
- Zhou L, Sun J, Long H, Zhou W, Xia R, Luo Y et al. Imaging phenotyping using 18F-FDG PET/CT radiomics to predict micropapillary and solid pattern in lung adenocarcinoma. *Insights into Imaging*. 2024;15(1).
- Zhao H, Su Y, Lyu Z, Tian L, Xu P, Lin L, et al. Non-invasively discriminating the pathological subtypes of Non-small cell lung Cancer with pretreatment (18) F-FDG PET/CT using deep learning. *Acad Radiol*. 2024;31(1):35–45.
- Huang B, Sollee J, Luo YH, Reddy A, Zhong Z, Wu J, et al. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT. *EBioMedicine*. 2022;82:104127.
- Xiao Z, Cai H, Wang Y, Cui R, Huo L, Lee EY, et al. Deep learning for predicting epidermal growth factor receptor mutations of non-small cell lung cancer on PET/CT images. *Quant Imaging Med Surg*. 2023;13(3):1286–99.
- Hu D, Li X, Lin C, Wu Y, Jiang H. Deep learning to predict the cell proliferation and prognosis of Non-Small cell lung Cancer based on FDG-PET/CT images. *Diagnostics (Basel, Switzerland)*. 2023;13:19.
- Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via Gradient-Based localization. *INT J COMPUT Vis*. 2020;128(2):336–59.
- Hou Y, Song W, Chen M, Zhang J, Luo Q, Um SW, et al. The presence of lepidic and micropapillary/solid pathological patterns as minor components has prognostic value in patients with intermediate-grade invasive lung adenocarcinoma. *Translational Lung cancer Res*. 2022;11(1):64–74.
- Cha MJ, Lee HY, Lee KS, Jeong JY, Han J, Shim YM, et al. Micropapillary and solid subtypes of invasive lung adenocarcinoma: clinical predictors of histopathology and outcome. *J Thorac Cardiovasc Surg*. 2014;147(3):921–e82.
- Xiong X, Zhang L, Zang B, Xu D, Chen C, Dong G et al. High-risk pathological subtype associated FAM83A-AS1 promotes malignancy and Glycolysis of lung adenocarcinoma via miR-202-3p/HK2 axis. *Oncol Rep*. 2023;49(5).
- Bu L, Tu N, Wang K, Zhou Y, Xie X, Han X, et al. Relationship between (18) F-FDG PET/CT Semi-Quantitative parameters and international association for the study of lung cancer, American thoracic society/European respiratory society classification in lung adenocarcinomas. *Korean J Radiol*. 2022;23(1):112–23.
- Önner H, Coşkun N, Erol M, Eren Karanis M. The role of Histogram-Based textural analysis of (18)F-FDG PET/CT in evaluating tumor heterogeneity and predicting the prognosis of invasive lung adenocarcinoma. *Mol Imaging Radionucl Ther*. 2022;31:33–41.
- Li Z, Wu W, Pan X, Li F, Zhu Q, He Z, et al. Serum tumor markers level and their predictive values for solid and micropapillary components in lung adenocarcinoma. *Cancer Med*. 2022;11(14):2855–64.
- Wang Z, Zhang N, Liu J, Liu J. Predicting micropapillary or solid pattern of lung adenocarcinoma with CT-based radiomics, conventional radiographic and clinical features. *Respir Res*. 2023;24(1).
- Choi W, Liu C-J, Alam SR, Oh JH, Vaghjiani R, Humm J, et al. Preoperative 18F-FDG PET/CT and CT radiomics for identifying aggressive histopathological subtypes in early stage lung adenocarcinoma. *Comput Struct Biotechnol J*. 2023;21:5601–8.
- Chen M, Copley SJ, Viola P, Lu H, Aboagye EO. Radiomics and artificial intelligence for precision medicine in lung cancer treatment. *Sem Cancer Biol*. 2023;93:97–113.
- Gao Y, Ding Y, Xiao W, Yao Z, Zhou X, Sui X, et al. A semi-supervised learning framework for micropapillary adenocarcinoma detection. *Int J Comput Assist Radiol Surg*. 2022;17(4):639–48.
- Li M, Gong J, Bao Y, Huang D, Peng J, Tong T. Special issue the advance of solid tumor research in China: prognosis prediction for stage II colorectal cancer by fusing computed tomography radiomics and deep-learning features of primary lesions and peripheral lymph nodes. *Int J Cancer*. 2023;152(1):31–41.
- Shao X, Ge X, Gao J, Niu R, Shi Y, Shao X, et al. Transfer learning-based PET/CT three-dimensional convolutional neural network fusion of image and clinical information for prediction of EGFR mutation in lung adenocarcinoma. *BMC Med Imaging*. 2024;24(1):54.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.