Am J Nucl Med Mol Imaging 2023;13(5):230-244
www.ajnmmi.us /ISSN:2160-8407/ajnmmi0152467

Original Article

Prediction of EGFR mutation status in lung adenocarcinoma based on ^{18}F-FDG PET/CT radiomic features

Jian-Ling Tan[1*], Liang Xia[2*], Su-Guang Sun[1], Hui Zeng[1], Di-Yu Lu[2#], Xiao-Jie Cheng[1#]

[1]Department of Nuclear Medicine, The Sixth Hospital of Wuhan, Affiliated Hospital of Jianghan University, Wuhan, Hubei, China; [2]Department of Nuclear Medicine, Wuhan Central Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. [*]Equal contributors. [#]Co-corresponding author.

Received July 27, 2023; Accepted October 16, 2023; Epub October 20, 2023; Published October 30, 2023

Abstract: The earlier identification of EGFR mutation status in lung adenocarcinoma patients is crucial for treatment decision-making. Radiomics, which involves high-throughput extraction of imaging features from medical images for quantitative analysis, can quantify tumor heterogeneity and assess tumor biology non-invasively. This field has gained attention from researchers in recent years. The aim of this study is to establish a model based on ^{18}F-FDG PET/CT radiomic features to predict the epidermal growth factor receptor (EGFR) mutation status of lung adenocarcinoma and evaluate its performance. 155 patients with lung adenocarcinoma who underwent ^{18}F-FDG PET/CT scans and EGFR gene detection before treatment were retrospectively analyzed. The LIFEx packages was used to perform 3D volume of interest (VOI) segmentation manually on DICOM images and extract 128 radiomic features. The Wilcoxon rank sum test and least absolute shrinkage and selection operator (LASSO) regression algorithm were applied to filter the radiomic features and establish models. The performance of the models was evaluated by the receiver operating characteristic (ROC) curve and the area under the curve (AUC). Among the models we have built, the radiomic model based on ^{18}F-FDG PET/CT has the best prediction performance for EGFR gene mutation status, with an AUC of 0.90 (95% CI 0.84~0.96) in the training set and 0.79 (95% CI 0.64~0.94) in the test set. In conclusion, we have established a radiomics model based on ^{18}F-FDG PET/CT, which has good predictive performance in identifying EGFR gene mutation status in lung adenocarcinoma patients.

Keywords: ^{18}F-FDG PET/CT, radiomics, lung adenocarcinoma, epidermal growth factor receptor (EGFR)

Introduction

Lung cancer is the second most common malignant tumor in the world as well as the leading cause of cancer-related deaths [1-3]. Non-small cell lung cancer (NSCLC) accounts for more than 85% of lung malignancies, with adenocarcinoma being the most common pathological type [4]. Advances in cancer genomics have demonstrated that NSCLC is driven by somatic mutations in key oncogenes [5], with epidermal growth factor receptor (EGFR) mutations being the most common genetic alteration in lung adenocarcinoma [6]. EGFR and its mediated signaling pathway regulate many physiological processes such as cell growth, proliferation and differentiation. Overexpression or mutation of EGFR plays an important role in the development, differentiation, and drug resistance of adenocarcinoma [7]. It is reported that over 50% of non-small cell lung cancer patients in Asia have EGFR mutations [8]. In recent years, molecular targeted therapy has developed rapidly, and several EGFR-tyrosine kinase inhibitors (TKIs) have been developed as small molecule targeted therapeutic agents for the treatment of NSCLC [9]. Compared to chemotherapy, EGFR tyrosine kinase inhibitor (TKI) therapy can effectively prolong the progression free survival (PFS) and overall survival (OS) of patients with lung adenocarcinoma with EGFR mutation [10-12]. Studies have shown that approximately 70%-80% of EGFR-mutated lung cancer patients respond significantly to EGFR-TKI treatment and achieve good clinical outcomes [13]. According to the National Comprehensive Cancer Network (NCCN) and Chinese Society of Clinical Oncology (CSCO) guidelines, EGFR TKIs have been approved as first-line standard therapy for driver mutation-

positive advanced lung adenocarcinoma [14]. However, in contrast, wild type patients experience limited benefits from EGFR-TKI therapy [13, 15]. Therefore, it is of great importance to identify whether a lung adenocarcinoma patient has EGFR mutation before TKIs targeted therapy. Currently, the most commonly used genetic testing method in clinical practice is biopsy. However, this method has some limitations. Firstly, it is invasive and some patients may not be able to undergo the procedure due to poor health conditions. Additionally, due to the location of the lesion, some patients may experience difficulties in obtaining a sufficient sample for testing [16, 17]. Due to the heterogeneity of tumors, the tissue obtained from a biopsy may not be sufficient to accurately determine the type of EGFR mutation [18]. Analysis of circulating cell-free tumor DNA (ctDNA) is another method for assessing EGFR mutation status. Although this method is relatively simple to operate, unfortunately, research has shown that ctDNA testing has a relatively high false negative rate and is costly [19]. Therefore, it is urgent to develop a simple and non-invasive method to predict the gene mutation status of lung adenocarcinoma patients.

Radiomics is a high-throughput method of extracting and analyzing a large number of medical imaging features, which can more comprehensively and objectively describe tumor heterogeneity [20], and make up for the shortcomings of traditional qualitative diagnosis. Currently, the radiomics method used to predict lung adenocarcinoma gene mutation is mainly based on the single-modality image features of CT or PET [21], which has some limitations. ^{18}F-FDG PET/CT is a non-invasive multimodal imaging approach that can provide both functional and metabolic information about the lesions, as well as anatomical information. It has been widely used in tumor diagnosis, staging, re-staging, efficacy monitoring and prognosis evaluation [22, 23]. Multimodal radiomics approaches can extract more meaningful radiomics features from different perspectives, thereby overcoming the limitations of unimodal models and providing a more comprehensive and reliable description of tumor [24-28]. Therefore, the purpose of this study is to build a radiomics model to predict the EGFR mutation status of lung adenocarcinoma based on ^{18}F-FDG PET/CT multimodal radiomics features and verify it.

Materials and methods

Patients

This study was performed in compliance with the Declaration of Helsinki and the relevant ethical guidelines. Ethics Committee of the Central Hospital of Wuhan approved this study. Due to the retrospective nature of the investigation and the use of anonymized patient data, the requirement for informed consent was waived. This study retrospectively analyzed data from patients diagnosed with lung adenocarcinoma in our hospital from June 2019 to May 2022. The inclusion criteria were: (1) Histologically confirmed lung adenocarcinoma; (2) Availability of complete ^{18}F-FDG PET/CT images reconstructed in digital imaging and communications in medicine (DICOM) format before treatment; (3) The EGFR mutation detection results are available and the time interval between the ^{18}F-FDG PET/CT examination and EGFR testing is no more than 4 weeks; (4) Single lesion (maximum diameter >1 cm); (5) No history of other malignant tumors. Exclusion criteria included: (1) No biopsy or lack of gene detection results; (2) Genetic testing has identified the presence of other mutations, such as the Kirsten ratsarcoma viral oncogene homolog (KRAS); (3) Poor image quality or with serious artifacts; (4) Pure ground-glass nodules without FDG metabolism; (5) Incomplete clinical data; (6) Metastatic tumor of the lung or other types of lung cancer. The patient selection flowchart is shown in **Figure 1**.

EGFR mutation detection

EGFR mutation detection was performed on tumor histological specimens obtained through surgical resection or biopsy. Real-time fluorescence PCR was used to detect mutations of exons of the EGFR gene, using the Roche Cobas DNA sample preparation and EGFR mutation detection kit, with the specific steps carried out in accordance with the kit instructions. PCR analysis was performed using Roche Cobas Z480. If any exon mutation was detected, the tumor was classified as EGFR mutant type (EGFR-MT). Otherwise, it was classified as EGFR wild type (EGFR-WT).

^{18}F-FDG PET/CT image acquisition

The ^{18}F-FDG was produced by GE Minitrace cyclotron, with radiochemical purity >95%.

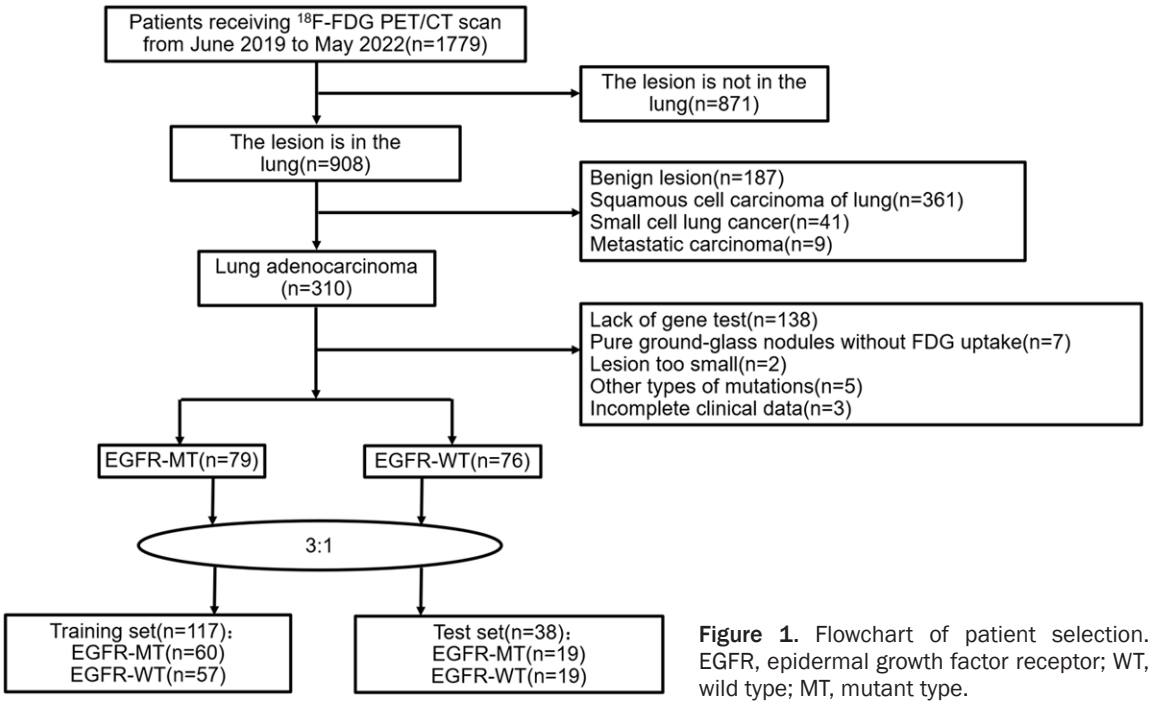

Figure 1. Flowchart of patient selection. EGFR, epidermal growth factor receptor; WT, wild type; MT, mutant type.

Patients were required to fast for more than 6 hours before the imaging agent was injected, and their fasting blood glucose levels were kept at ≤11.1 mmol/L. According to the patient's body mass, 3.7-5.5 MBq/kg of ^{18}F-FDG was injected through the back of the hand or elbow vein, and they rested for approximately 60 minutes in a quiet environment. After the bladder was emptied, the whole body PET/CT imaging was conducted using the combined imaging U510 and GE Discovery VCT PET CT scanners. The scan extended from the top of skull to the upper middle thigh. The CTScout image was used to locate the scanning position. First, CT scanning was conducted (tube voltage, 120 kV; tube current, 80 mA; slice thickness, 2.4 mm; 64 and 16 row CT), followed by PET scanning (2 minutes per bed position, with 6-8 beds collected). The original image was reconstructed using an iterative algorithm after attenuation correction, resulting in transverse, sagittal and coronal fusion images (Xeleris, conjoined image U510) from PET, CT and PET/CT.

Image analysis and feature extraction

In our study, three radiologists, who were blinded to all clinical and histologic data, retrospectively interpreted PET/CT studies. A nuclear medicine physician with 3 years of experience in pulmonary PET/CT performed 3D volume of interest (VOI) segmentation manually on DICOM images using LIFEx packages (version 7.2.0, http://www.lifexsoft.org). Two senior nuclear medicine physicians (8 and 10 years of experience in pulmonary PET/CT) reviewed all the VOIs and the radiologists negotiated to reach a consensus for controversial cases. The image analysis processing steps were as follows: (1) The VOI of lesion was manually delineated slice by slice on the CT scans from PET/CT. A CT-based VOI was labeled as "lesion_CT". If multiple lesions were present, the largest one was considered the labeled lesion assigned to the patient. (2) A PET-based VOI labeled "lesion_PT" was created by fusing voxels from PET scans and the "Lesion_CT" VOI. (3) A PET-based VOI labeled "lesion_PT_40%Peak" was created by selecting voxels with a threshold of 40% of the maximum Standardized Uptake Value (SUVmax) within the "lesion_PT".

In summary, we created three VOIs for each patient: "lesion_CT", "lesion_PT", "lesion_PT_40%Peak". The software program then automatically calculated and extracted 128 CT and PET radiomic features, including 25 basic features, 17 morphological features, 30 histogram features, 24 Gray-Level Co-occurrence Matrix (GLCM) features, 11 Gray-Level Run

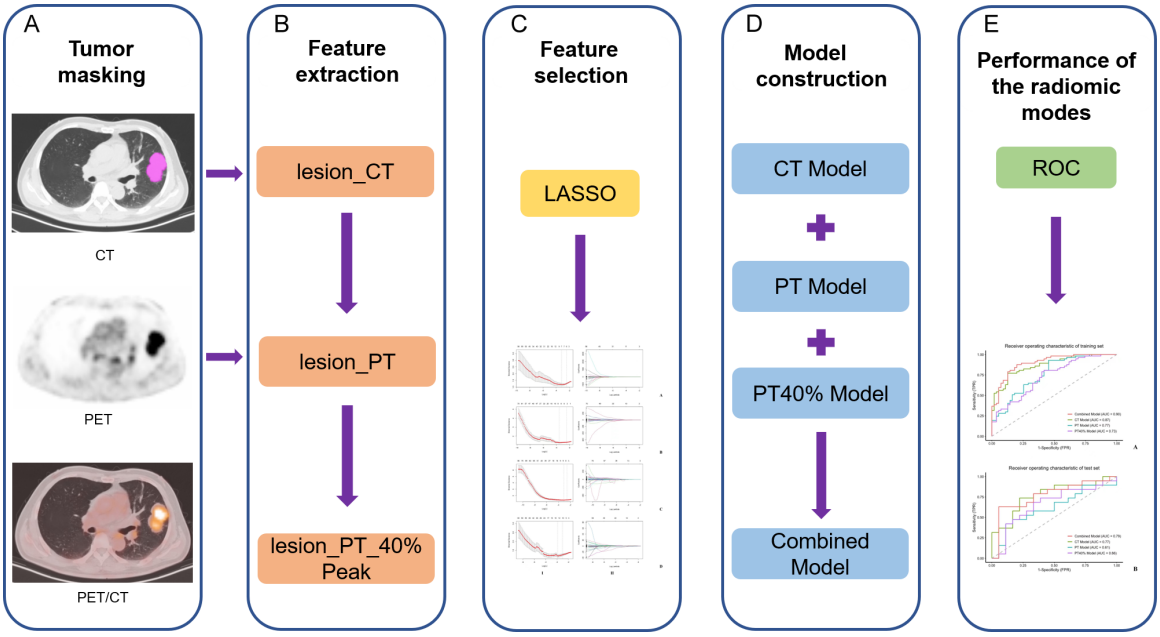

Figure 2. The workflow of our study. A. Tumor masking; B. Feature extraction; C. Feature selection; D. Model construction; E. Performance of the radiomic modes.

Length Matrix (GLRLM) features, 5 Neighborhood Grey-Level Different Matrix (NGLDM) features and 16 Grey-Level Zone Length Matrix (GLZLM) features. This resulted in a total of 384 (128 * 3) radiomic features, which are provided in the Supplementary Material.

Feature selection and modeling

Due to the relatively large number of radiomic features and small number of patients, the Wilcoxon rank sum test and least absolute shrinkage and selection operator (LASSO) regression algorithm were applied to avoid overfitting of the model. Three single-modality radiomic models and a dual-modality radiomic model were then trained to identify EGFR mutation status using radiomic features derived from the three types of VOIs ("lesion_CT", "lesion_PT", "lesion_PT_40%Peak"). Logistic regression was used to establish these models, and their performance was evaluated by the receiver operating characteristic (ROC) curve and the area under the curve (AUC) in the training and test set.

Statistical analysis

We performed statistical analysis using IBM SPSS Statistics version 26.0 and R (version 4.0.5, http://www.r-project.org). Continuous variables were tested using the independent samples t-test, and categorical variables were tested using the chi-square test or Fisher exact probability method. The "glmnet" package was used to perform LASSO logistic regression analysis and binary logistic regression models. The "ggplot" package was used to create ROC curves and calculate and compare AUCs, while the "ggpubr" package was used to compare RadScores. A two-sided *p*-value <0.05 was accepted as indicative of statistical significance. Our workflow is shown in **Figure 2**.

Results

Clinical characteristics of patients

A total of 155 lung adenocarcinoma patients were enrolled in our study, as shown in **Table 1**. There were 79 cases (51.0%) with EGFR mutant and 76 cases (49.0%) with wild type. All patients were randomly assigned to a training set (117 cases) and a test set (38 cases) in a 3:1 ratio. In the training set, there were statistically significant differences in the gender (P=0.005) and maximum diameter (P=0.049) between the EGFR mutation and wild type. In the test set, there was a statistically significant difference in gender (P=0.038) between the EGFR mutation and wild type. However, the

Table 1. Clinical characteristics of all patients included in the study

Factors	Training Set (n=117)		*P* value	Test Set (n=38)		*P* value
	EGFR-MT	EGFR-WT		EGFR-MT	EGFR-WT	
Age	67.83±9.04	64.73±11.04	0.099	64.05±10.64	68.53±9.01	0.170
Gender						
Male	28	41	0.005*	9	16	0.038*
Female	32	16		10	3	
Smoking history						
Yes	28	24	0.620	11	8	0.517
No	32	33		8	11	
TNM Staging						
I~II	24	16	0.174	9	4	0.170
III~IV	36	41		10	15	
Maximum diameter						
<3 cm	33	21	0.049*	9	6	0.329
≥3 cm	27	36		10	13	
Tumor location						
LUL	23	20	0.548	7	5	0.674
LLL	6	11		3	5	
RUL	19	15		6	8	
RLL	12	11		3	1	

Notes: LUL, left upper lung; LLL, left lower lung; RUL, right upper lung; RLL, right lower lung; TNM Staging Based on American Joint Committee on Cancer (AJCC) 8th edition; *Only statistically significant (P<0.05) results are reported for analysis.

other clinical characteristics (age, smoking history, TNM staging, and tumor location) showed no statistically significant differences in either the training or test sets (all P>0.05).

Feature selection

The Wilcoxon rank sum test and least absolute shrinkage and selection operator (LASSO) regression algorithm were applied to select the optimal subset of radiomic features (**Figure 3**). As previously mentioned, three single-modality radiomic models (CT Model, PT Model and PT40% Model) and a dual-modality radiomic model (Combined Model) were trained to identify EGFR mutation status with 117 patients included in the training set. Based on "lesion_CT", the 7 features were Spherical Disproportion, Intensity Histogram Skewness, Intensity Histogram Minimum Grey Level, Intensity Histogram Interquartile Range, GLCM_Inverse Difference Moment, GLRLM_Long Runs Emphasis and GLSZM_Small Zone High Grey Level Emphasis. The 7 features derived from "lesion_PT" were Skewness, Maximum Grey Level, Area Under Curve Csh, Total Lesion Glycolysis, GLCM_Difference Variance, NGTDM_Coarseness, GLSZM_Normalised Zone Size Non Uniformity. The 9 features then drawn from "lesion_PT_40%Peak" were Spherical Disproportion, Sphericity, Maximum Grey Level, Area Under Curve Csh, Total Lesion Glycolysis, Minimum Histogram Gradient Grey Level, GLRLM_Short Run Low Grey Level Emphasis, NGTDM_Contrast, GLSZM_Small Zone High Grey Level Emphasis. Eventually, 12 features developed on these three VOIs were selected for the Combined Model, which were Spherical Disproportion (CT), Intensity Histogram Skewness (CT), Intensity Histogram Minimum Grey Level (CT), Intensity Histogram Interquartile Range (CT), GLCM_Inverse Difference Moment (CT), GLRLM_Long Runs Emphasis (CT), Skewness (PT), Maximum Grey Level (PT), GLCM_Difference Variance (PT), Maximum Grey Level (PT40%), GLRLM_Short Run Low Grey Level Emphasis (PT40%), NGTDM_Contrast (PT40%).

In addition, a radiomics signature score (RadScore) was calculated for each patient within each model: CT Model RadScore: -8.3841695 + 6.8798866 * SphericalDisproportion + 0.3512650 * IntensityHistogramSkewness + 0.0909245 * IntensityHistogramInterquartile-

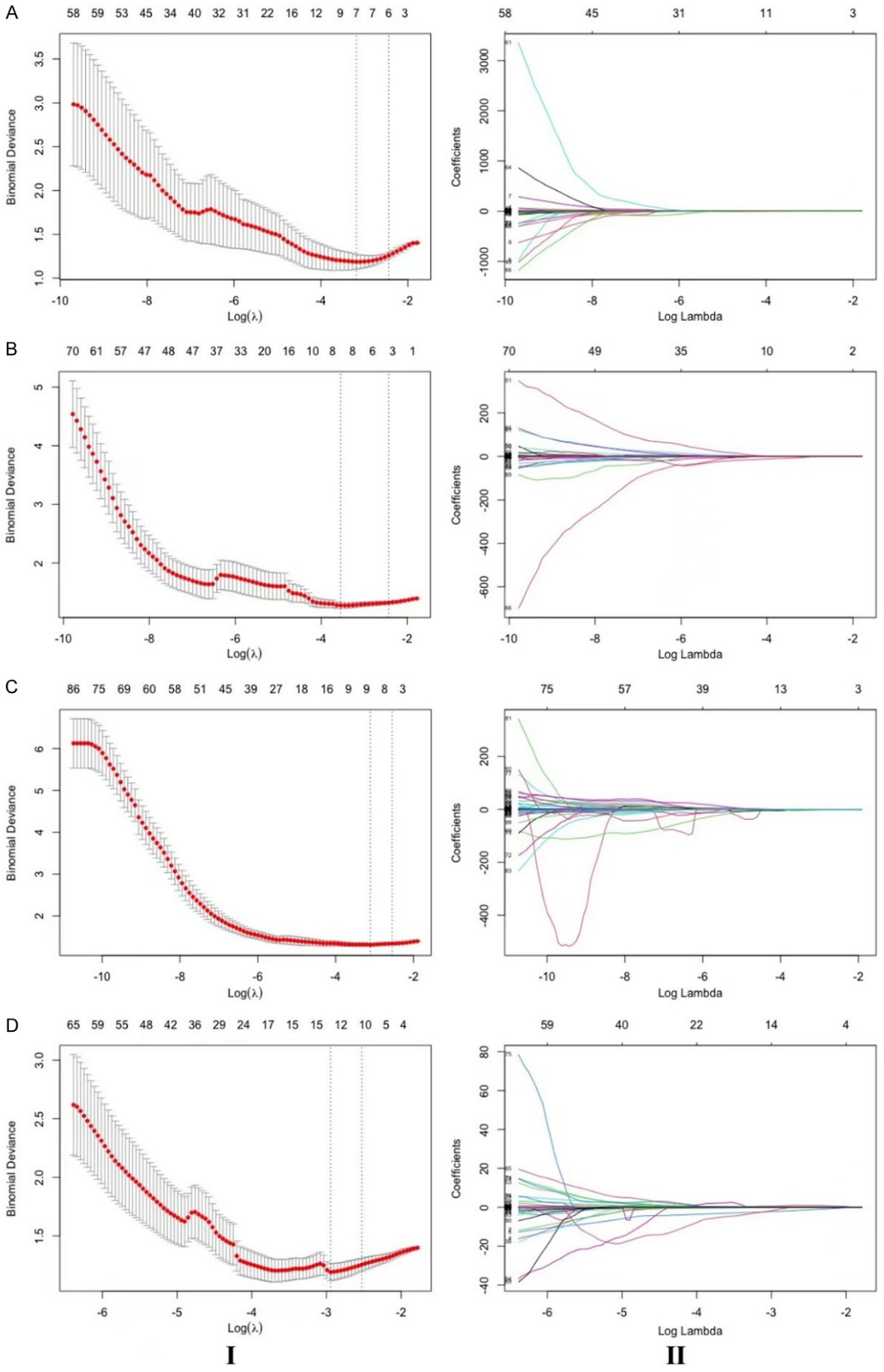
A
58 59 53 45 34 40 32 31 22 16 12 9 7 7 6 3
Binomial Deviance
Log(λ)
58 45 31 11 3
Coefficients
Log Lambda
B
70 61 57 47 48 47 37 33 20 16 10 8 8 6 3 1
Binomial Deviance
Log(λ)
70 49 35 10 2
Coefficients
Log Lambda
C
86 75 69 60 58 51 45 39 27 18 16 9 9 8 3
Binomial Deviance
Log(λ)
75 57 39 13 3
Coefficients
Log Lambda
D
65 59 55 48 42 36 29 24 17 15 15 12 10 5 4
Binomial Deviance
Log(λ)
59 40 22 14 4
Coefficients
Log Lambda
I
II

Figure 3. The Wilcoxon rank sum test and least absolute shrinkage and selection operator (LASSO) regression algorithm were applied to select the optimal subset of radiomic features. I, Optimal feature selection according to AUC value. II, LASSO coefficient profiles of the radiomic features. A. CT Model; B. PT Model; C. PT40% Model; D. Combined Model.

Range + 0.0619512 * IntensityHistogramInterquartileRange + 9.3825251 * GLCM_InverseDifferenceMoment - 3.0919950 * GLRLM_LongRunsEmphasis - 0.0004478 GLSZM_SmallZoneHighGreyLevelEmphasis; PT Model RadScore: -0.465306 + 0.411846 * Skewness + 0.242730 * MaximumGreyLevel - 0.726558 * AreaUnderCurveCsh - 0.001263 * TotalLesionGlycolysis - 0.051487 * GLCM_DifferenceVariance + 29.863704 * NGTDM_Coarseness - 1.973408 * GLSZM_NormalisedZoneSizeNonUniformity; PT40% Model RadScore: 3.043534 - 1.219807 * SphericalDisproportion - 1.508150 * Sphericity + 0.220315 * MaximumGreyLevel - 0.434433 * AreaUnderCurveCsh - -0.001865 * TotalLesionGlycolysis - -0.010617 * MinimumHistogramGradientGreyLevel + 11.485984 * GLRLM_ShortRunLowGreyLevelEmphasis - -1.593508 * NGTDM_Contrast - -0.001262 * GLSZM_SmallZoneHighGreyLevelEmphasis; Combined Model RadScore: -9.12922 + 6.13339 * CT_SphericalDisproportion + 0.67808 * CT_IntensityHistogramSkewness + 0.09336 * CT_IntensityHistogramMinimumGreyLevel + 0.06373 * CT_IntensityHistogramInterquartileRange + 11.47291 * CT_GLCM_InverseDifferenceMoment - 3.82125 * CT_GLRLM_LongRunsEmphasis + 0.42365 * PT_Skewness + 0.16443 * PT_MaximumGreyLevel - 0.03230 * PT_GLCM_DifferenceVariance - 0.06691 * PT40%_MaximumGreyLevel + 7.62081 * PT40%_GLRLM_ShortRunLowGreyLevelEmphasis - 0.91629 * PT40%_NGTDM_Contrast.

The median and the interquartile range for the selected radiomic features in each model and the calculated RadScore are shown in **Table 2**. For each model, the selected features and RadScore were significantly different between the EGFR mutant and wild type groups in the training set (P<0.05). Specifically, lesions with EGFR mutant had higher RadScore than those with EGFR wild type in both the training and test sets. While RadScore with CT Model, PT Model, PT40% Model and Combined Model in training set were 1.471 vs -1.06, 0.693 vs -0.48, 0.31 vs -0.066 and 1.669 vs -1.444, respectively, those in test set were 1.424 vs -2.028, 0.465 vs -0.124, 0.209 vs -0.601 and 1.978 vs -2.296, respectively. The RadScore for each patient in the two sets is shown as the bargraphs in **Figure 4** with the four models.

Performance of the radiomic models

ROCs of the models in both the training and test sets are displayed in **Figure 5**. All the models had good predictive performance in the training set, and the AUCs with CT Model, PT Model, PT40% Model and Combined Model were 0.87 (95% CI 0.80~0.94), 0.77 (95% CI 0.68~0.86), 0.73 (95% CI 0.64~0.82) and 0.90 (95% CI 0.84~0.96), respectively. In the test set, the AUCs with CT Model, PT Model, PT40% Model and Combined Model were 0.77 (95% CI 0.62~0.92), 0.61 (95% CI 0.53~0.80), 0.66 (95% CI 0.50~0.84) and 0.79 (95% CI 0.64~0.94), respectively. There was a statistically significant difference in AUCs between the Combined Model and PT Model (P=0.0011), and between the Combined Model and the PT40% Model (P<0.0001). However, there was no statistical difference in AUCs between the Combined Model and the CT Model (P=0.112). The predictive abilities of the four models, including AUC, sensitivity, specificity and accuracy, are listed in **Table 3**.

Discussion

In this study, we have established a predictive model based on 12 radiomic features derived from pre-therapy ^{18}F-FDG PET/CT images of patients to predict EGFR mutation status between EGFR mutant and wild type, which showed good predictive performance in the training set (AUC=0.90) and test set (AUC=0.79). Among the 12 selected features in our study, Spherical Disproportion is a radiomics feature used to describe the morphological characteristics of tumors. It refers to the measurement of the difference between the long and short axes of the tumor, indicating how closely it approximates a spherical shape. The larger the value of Spherical Disproportion, the more irregular the tumor shape. We found that

Table 2. The median and the interquartile range for the selected radiomic features in each model and the calculated RadScore

Characteristic	Training set (n=117)		p	Training set (n=38)		p
	EGFR mutant (n=60)	EGFR wild type (n=57)		EGFR mutant (n=19)	EGFR wild type (n=19)	
CT Model (7 features)						
SphericalDisproportion	1.525 (1.369~1.76)	1.378 (1.239~1.472)	0.0028	1.537 (1.366~1.763)	1.414 (1.331~1.498)	0.0142
IntensityHistogramSkewness	-0.404 (-1.452~0.2)	-1.143 (-3.015~-0.557)	0.0027	-0.732 (-1.102~-0.264)	-2.248 (-3.316~-0.787)	0.0286
IntensityHistogramMinimumGreyLevel	13 (7~19)	10 (2.5~15.5)	0.0359	16 (12.5~19.5)	10.5 (3~14.75)	0.0283
IntensityHistogramInterquartileRange	29 (18~43)	20 (3~37.5)	0.0101	31 (12.5~36.5)	5 (3~31.5)	0.0259
GLCM_InverseDifferenceMoment	0.173 (0.104~0.261)	0.283 (0.141~0.43)	0.0065	0.221 (0.11~0.335)	0.38 (0.281~0.449)	0.0363
GLRLM_LongRunsEmphasis	1.201 (1.099~1.415)	1.501 (1.159~1.913)	0.0034	1.373 (1.11~1.692)	1.768 (1.387~2.096)	0.0423
GLSZM_SmallZoneHighGreyLevelEmphasis	4203.223 (2853.801~4872.091)	4709.551 (3961.609~5904.823)	0.0145	4537.519 (4008.971~4796.108)	5052.106 (4319.99~5675.182)	0.0112
RadScore	1.471 (0.409~2.254)	-1.06 (-2.514~-0.203)	<0.0001	1.424 (-0.268~2.902)	-2.028 (-3.678~-0.303)	0.0041
PT Model (7 features)						
Skewness	1.047 (0.711~1.553)	0.9 (0.499~1.306)	0.0069	1.377 (0.915~1.573)	0.827 (0.416~1.182)	0.0919
MaximumGreyLevel	9.653 (3.928~13.669)	15.667 (9.115~20.127)	0.0011	9.79 (4.164~16.824)	15.981 (13.603~22.9)	0.0049
AreaUnderCurveCsh	0.968 (0.592~1.258)	1.437 (0.963~2.213)	0.0004	0.854 (0.574~1.438)	1.714 (1.12~2.358)	0.0107
TotalLesionGlycolysis	44.791 (9.57~159.482)	121.389 (28.732~318.699)	0.0278	60.342 (10.618~199.473)	279.194 (46.494~1089.488)	0.0726
GLCM_DifferenceVariance	13.09 (1.331~24.829)	33.618 (11.422~65.837)	0.0023	21.218 (2.116~37.129)	38.898 (25.511~56.917)	0.0387
NGTDM_Coarseness	0.018 (0.008~0.028)	0.012 (0.005~0.02)	0.0431	0.01 (0.005~0.017)	0.005 (0.003~0.011)	0.0346
GLSZM_NormalisedZoneSizeNonUniformity	0.323 (0.208~0.392)	0.421 (0.292~0.525)	0.0033	0.32 (0.154~0.417)	0.445 (0.364~0.49)	0.0171
RadScore	0.693 (-0.039~1.146)	-0.48 (-1.694~0.408)	<0.0001	0.465 (-0.608~0.899)	-0.124 (-1.232~0.46)	0.0221
PT40% Model (9 features)						
SphericalDisproportion	1.475 (1.409~1.648)	1.609 (1.401~1.818)	0.0124	1.519 (1.342~1.629)	1.825 (1.497~2.482)	0.0172
Sphericity	0.678 (0.607~0.71)	0.622 (0.55~0.714)	0.0124	0.658 (0.614~0.745)	0.548 (0.403~0.668)	0.0172
MaximumGreyLevel	9.653 (3.928~13.669)	15.667 (9.115~20.127)	0.0009	9.79 (4.164~16.824)	15.981 (13.603~22.9)	0.0061
AreaUnderCurveCsh	1.663 (0.802~2.513)	2.396 (1.723~3.639)	0.0026	1.956 (0.749~2.809)	2.866 (2.049~4.628)	0.0309
TotalLesionGlycolysis	24.014 (5.831~74.908)	61.755 (20.289~187.743)	0.0134	29.783 (9.937~120.825)	119.146 (30.201~627.434)	0.0568
MinimumHistogramGradientGreyLevel	15 (5~23)	22 (10~32.5)	0.0034	19 (8.5~28.5)	30 (18.75~37.75)	0.0614
GLRLM_ShortRunLowGreyLevelEmphasis	0.004 (0.001~0.019)	0.002 (0.001~0.007)	0.0339	0.003 (0.001~0.013)	0.001 (0.001~0.002)	0.0098
NGTDM_Contrast	0.283 (0.066~0.504)	0.458 (0.206~0.847)	0.0135	0.476 (0.189~0.639)	0.447 (0.313~0.679)	0.0532
GLSZM_SmallZoneHighGreyLevelEmphasis	153.222 (2.51~499.008)	462.913 (191.73~924.533)	0.0016	366.565 (38.853~555.567)	660.679 (422.653~877.655)	0.0255
RadScore	0.31 (0.047~1.045)	-0.066 (-1.214~0.31)	<0.0001	0.209 (-0.5~0.452)	-0.601 (-1.181~0.031)	0.0192
Combined Model (12 features)						
CT_SphericalDisproportion	1.525 (1.369~1.76)	1.378 (1.239~1.472)	0.0283	1.537 (1.366~1.763)	1.414 (1.331~1.498)	0.0142
CT_IntensityHistogramSkewness	-0.404 (-1.452~0.2)	-1.143 (-3.015~-0.557)	0.0284	-0.732 (-1.102~-0.264)	-2.248 (-3.316~-0.787)	0.0286
CT_IntensityHistogramMinimumGreyLevel	13 (7~19)	10 (2.5~15.5)	0.0358	16 (12.5~19.5)	10.5 (3~14.75)	0.0283
CT_IntensityHistogramInterquartileRange	29 (18~43)	20 (3~37.5)	0.0101	31 (12.5~36.5)	5 (3~31.5)	0.0259
CT_GLCM_InverseDifferenceMoment	0.173 (0.104~0.261)	0.283 (0.141~0.43)	0.0007	0.221 (0.11~0.335)	0.38 (0.281~0.449)	0.0363
CT_GLRLM_LongRunsEmphasis	1.201 (1.099~1.415)	1.501 (1.159~1.913)	0.0003	1.373 (1.11~1.692)	1.768 (1.387~2.096)	0.0423
PT_Skewness	1.047 (0.711~1.553)	0.9 (0.499~1.306)	0.0639	1.377 (0.915~1.573)	0.827 (0.416~1.182)	0.0919

PT_MaximumGreyLevel	9.653 (3.928~13.669)	15.667 (9.115~20.127)	0.0011	9.79 (4.164~16.824)	15.981 (13.603~22.9)	0.0049
PT_GLCM_DifferenceVariance	13.09 (1.331~24.829)	33.618 (11.422~65.837)	0.0023	21.218 (2.116~37.129)	38.898 (25.511~56.917)	0.0387
PT40%_MaximumGreyLevel	9.653 (3.928~13.669)	15.667 (9.115~20.127)	0.0009	9.79 (4.164~16.824)	15.981 (13.603~22.9)	0.0061
PT40%_GLRLM_ShortRunLowGreyLevelEmphasis	0.004 (0.001~0.019)	0.002 (0.001~0.007)	0.0339	0.003 (0.001~0.013)	0.001 (0.001~0.002)	0.0098
PT40%_NGTDM_Contrast	0.283 (0.066~0.504)	0.458 (0.206~0.847)	0.0135	0.476 (0.189~0.639)	0.447 (0.313~0.679)	0.0532
RadScore	1.669 (0.324~2.805)	-1.444 (-3.467~-0.424)	<0.0001	1.978 (-0.186~3.826)	-2.296 (-3.218~0.402)	0.0041

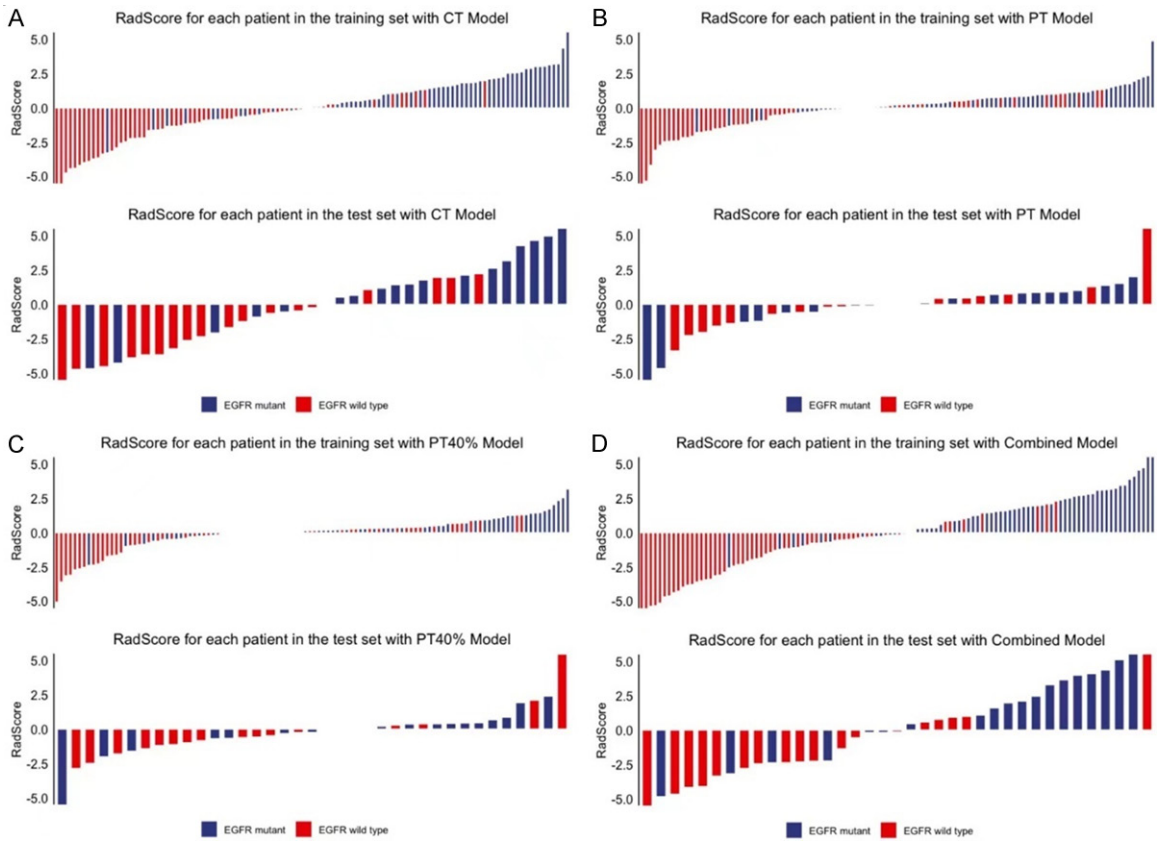

Figure 4. The RadScore for each patient in the two sets with the four models.

compared with the wild type, the morphology of EGFR mutant lung adenocarcinoma seems to be more irregular, which is similar to the results of Li et al. [9]. Intensity Histogram Minimum Grey Level represents the minimum gray value in VOI. Our research suggests that the value of wild type is lower than that of mutant type, indicating that the former may contain more cavities. Previous studies have shown that the CT signs of EGFR mutant and wild type are different. When examining the correlation between the EGFR gene mutation status and the CT signs of lung adenocarcinoma lesions, Zhou et al. found that the EGFR mutation rate in lung adenocarcinoma patients with cavities was 35.29% [29], indicating that the cavity signs are associated with EGFR wild type and consistent with our research results. Similarly, Hasegawa et al. suggested that the presence of cavitation in primary lung adenocarcinoma is associated with non-EGFR mutations and may indicate an adverse prognosis [30]. Maximum Grey Level reflects the distribution characteristics of tumor pixel gray level [31, 32], thus revealing the spatial distribution of ^{18}F-FDG in tumor, as well as is less susceptible to noise than SUVmax [17]. Previous studies have shown that EGFR mutations may activate pertinent intracellular signaling pathways, which can increase tumor glycolysis and lead to intense ^{18}F-FDG uptake in PET scans [33]. Our study shows that the value of wild type in the training and testing sets is higher than that of mutant type. It is consistent with the widely accepted view that lung adenocarcinomas without an EGFR mutation are more invasive than EGFR-mutated tumors [9, 34]. All other features are associated with image uniformity and tumor heterogeneity. In this study, EGFR mutant type was found to be more heterogeneous, which is in line with prior research [5, 9, 14, 17, 32, 35]. In addition, each patient's RadScore was calculated. In the training and testing sets, the RadScore of EGFR mutant patients was higher than that of EGFR wild type (both $P<0.001$).

Radiomics texture analysis can impart different meanings relying on the imaging modality.

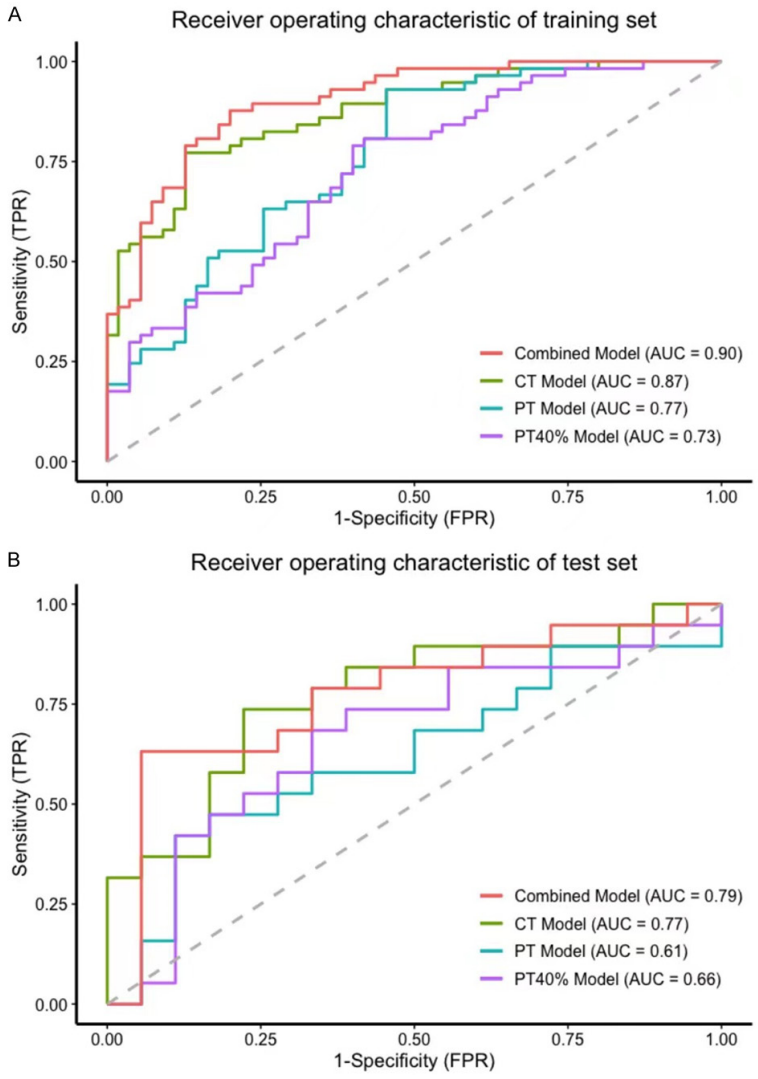

Figure 5. ROCs of the four models in both the training and test sets.

There have been several studies in the past that have found a correlation between CT radiomics and EGFR mutation. Tu et al. suggested that CT radiological features may be predictive factors for identifying EGFR mutations [36]. In our study, the radiomics model based solely on CT imaging had an AUC of 0.77 in the test set, which is consistent with previous research findings [37-39]. Unlike traditional CT imaging, PET imaging can reflect the glucose metabolism characteristics of tumors [18, 35, 40]. SUVmax is currently the most researched conventional metabolic parameter in PET imaging. However, there has been a long-standing debate among scholars regarding the relationship between SUVmax and EGFR mutations in lung adenocarcinoma in recent years [41-45]. It is well known that tumors are heterogeneous, and that simple SUV measurements do not take into account the spatial relationships between image voxels, which may provide more insight into the biology of these mutations and help to assess the degree of tumor heterogeneity [40]. Conventional semi-quantitative indicators generated from PET do not have sufficient predictive value for clinical practice. Unlike SUVmax, radiomics can extract numerous features through high throughput and use sophisticated mathematical models to quantify the spatial relationship between image voxels, providing a better representation of tumors and allowing for the prediction of EGFR mutation status. In this study, the radiomics model based solely on PET imaging had an AUC of 0.61 in the test set. Due to the limited spatial resolution of PET imaging, it may be difficult to determine the tumor boundary on PET images. Previous studies have not conducted in-depth exploration of this issue. Most previous researchers have selected a threshold of 40% to 42% SUVmax to determine the boundary of the tumor on PET images. To explore whether this method can effectively improve the predictive performance of radiomics, we selected 40% SUVmax as the threshold and established a 40% PET unimodal model. The results showed that the AUC of this model on the test set was 0.66, and the difference between it and the unthreshold PET model was not statistically significant. This may be because both models are based on PET images, so they may overlap in some features, leading to similar predictive performance. Additionally, setting a threshold may exclude some tumor tissue and reduce the features extracted by radiomics. Therefore, we believe that using 40% SUVmax as a threshold to determine the tumor boundary on PET imag-

Table 3. The predictive performance of each model in identifying EGFR mutations

Model	Training set				Test set			
	AUC (95% CI)	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC (95% CI)	Sensitivity (%)	Specificity (%)	Accuracy (%)
CT	0.87 (0.80~0.94)	87.27	71.93	79.46	0.77 (0.62~0.92)	68.42	77.78	72.97
PT	0.77 (0.68~0.86)	52.63	74.55	63.39	0.61 (0.53~0.80)	47.37	77.78	62.16
PT40%	0.73 (0.64~0.82)	38.69	87.27	62.5	0.66 (0.50~0.84)	21.06	88.89	54.05
Combined	0.90 (0.84~0.96)	73.69	87.27	80.36	0.79 (0.64~0.94)	68.42	83.33	75.68

Notes: AUC, area under the ROC curve; CI, confidence interval.

es may not significantly improve the predictive performance of radiomics models, although further confirmation in larger samples is needed. It is generally accepted that radiomic features derived from PET and CT are complementary. CT-based radiomic analysis reflects the pattern of tissue density distribution, whereas radiomic texture analysis based on PET images is related to the variability of the metabolic phenotype [9]. Thus, a radiomic signature based on the combination of PET and CT radiomic features could reflect the heterogeneity of tumors from different angles, thus enhancing its ability to predict EGFR mutational status. Therefore, we compared the PET/CT combined model with the unimodal models. The results showed that the PET/CT radiomics model had a significantly higher AUC compared to the PET-base model (P=0.0011). However, there was no significant difference between the PET/CT model and the CT model (P=0.112). We believe that one possible reason may be attributed to the limited sample size and selection bias. Additionally, although we excluded pure ground-glass nodules without FDG metabolism, some lesions with less solid components may have lower FDG metabolism, resulting in fewer PET radiomics features extracted, which led to a less significant improvement in the AUC of the combined model. However, this does not mean that PET radiomics features have no contribution in the combined model. Compared with the single-modality CT model, the combined model improved the accuracy of the training and test sets to 80.36% and 75.68%, respectively, and the specificity to 87.27% and 83.33%. Therefore, we believe that the predictive performance of the PET/CT double-modality radiomics model has the best predictive performance.

Our study has some limitations that should be acknowledged. Firstly, our sample size is relatively small, and we obtained all data from a single center. It has been noted that the frequency of EGFR mutations may have a correlation with race, with higher rates of mutations found in Asian populations compared to other ethnicities [5]. Given that all patients in this study were of Asian descent, it is imperative that we conduct large-scale, multi-center studies to enhance the generalization ability of our model across diverse races and regions. Additionally, we only collected the imaging data of patients. In future investigations, we plan to integrate clinical information from patients, which may further enhance the stability and predictive performance of our model. Lastly, it is a retrospective study, and prospective validation may offer additional evidence for further clinical applications.

In summary, we have constructed a radiomics model based on ^{18}F-FDG PET/CT for predicting EGFR gene mutation status in lung adenocarcinoma patients. Our model has shown good performance. By providing a straightforward and non-invasive screening method, this study offers valuable support for clinicians identifying candidates for molecular targeted therapy, especially for those patients who are unable to undergo biopsy.

Acknowledgements

This work was fully sponsored by Scientific Research Projects From Wuhan Municipal Health Commission (WX21C26).

Disclosure of conflict of interest

None.

Address correspondence to: Di-Yu Lu, Department of Nuclear Medicine, Wuhan Central Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. Tel: +86-18164105189; E-mail: diyulu2013@163.com; Xiao-Jie Cheng, Department of Nuclear Medicine, The Sixth Hospital of Wuhan, Affiliated Hospital of Jianghan University, Wuhan, Hubei, China. Tel: +86-13627125820; E-mail: chengxiaojie2008@163.com

References

[1] Shi Y, Au JS, Thongprasert S, Srinivasan S, Tsai CM, Khoa MT, Heeroma K, Itoh Y, Cornelio G and Yang PC. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol 2014; 9: 154-162.

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[3] Lin C, Hu F, Chu H, Ren P, Ma S, Wang J, Bai J, Han X and Ma S. The role of EGFR-TKIs as adjuvant therapy in EGFR mutation-positive early-stage NSCLC: a meta-analysis. Thorac Cancer 2021; 12: 1084-1095.

[4] Hsu WH, Yang JC, Mok TS and Loong HH. Overview of current systemic management of EGFR-mutant NSCLC. Ann Oncol 2018; 29: i3-i9.

[5] Yip SS, Kim J, Coroller TP, Parmar C, Velazquez ER, Huynh E, Mak RH and Aerts HJ. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. J Nucl Med 2017; 58: 569-576.

[6] Jia Y, Yun CH, Park E, Ercan D, Manuia M, Juarez J, Xu C, Rhee K, Chen T, Zhang H, Palakurthi S, Jang J, Lelais G, DiDonato M, Bursulaya B, Michellys PY, Epple R, Marsilje TH, McNeill M, Lu W, Harris J, Bender S, Wong KK, Jänne PA and Eck MJ. Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. Nature 2016; 534: 129-132.

[7] Dong H, Yin H, Zhao C, Cao J, Xu W and Zhang Y. Design, synthesis and biological evaluation of novel osimertinib-based HDAC and EGFR dual inhibitors. Molecules 2019; 24: 2407.

[8] Zhang YL, Yuan JQ, Wang KF, Fu XH, Han XR, Threapleton D, Yang ZY, Mao C and Tang JL. The prevalence of EGFR mutation in patients with non-small cell lung cancer: a systematic review and meta-analysis. Oncotarget 2016; 7: 78985-78993.

[9] Li X, Yin G, Zhang Y, Dai D, Liu J, Chen P, Zhu L, Ma W and Xu W. Predictive power of a radiomic signature based on (18)F-FDG PET/CT images for EGFR mutational status in NSCLC. Front Oncol 2019; 9: 1062.

[10] Duan H, Lu J, Lu T, Gao J, Zhang J, Xu Y, Wang M, Wu H, Liang Z and Liu T. Comparison of EGFR mutation status between plasma and tumor tissue in non-small cell lung cancer using the Scorpion ARMS method and the possible prognostic significance of plasma EGFR mutation status. Int J Clin Exp Pathol 2015; 8: 13136-13145.

[11] Han JY, Park K, Kim SW, Lee DH, Kim HY, Kim HT, Ahn MJ, Yun T, Ahn JS, Suh C, Lee JS, Yoon SJ, Han JH, Lee JW, Jo SJ and Lee JS. First-SIGNAL: first-line single-agent iressa versus gemcitabine and cisplatin trial in never-smokers with adenocarcinoma of the lung. J Clin Oncol 2012; 30: 1122-1128.

[12] Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Munoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombaron P, Bernabe R, Bearz A, Artal A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M and Paz-Ares L; Spanish Lung Cancer Group in collaboration with Groupe Francais de Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13: 239-246.

[13] Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA and Fukuoka M. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361: 947-957.

[14] Jiang M, Yang P, Li J, Peng W, Pu X, Chen B, Li J, Wang J and Wu L. Computed tomography-based radiomics quantification predicts epidermal growth factor receptor mutation status and efficacy of first-line targeted therapy in lung adenocarcinoma. Front Oncol 2022; 12: 985284.

[15] Miller VA, Riely GJ, Zakowski MF, Li AR, Patel JD, Heelan RT, Kris MG, Sandler AB, Carbone DP, Tsao A, Herbst RS, Heller G, Ladanyi M, Pao W and Johnson DH. Molecular characteristics of bronchioloalveolar carcinoma and adenocarcinoma, bronchioloalveolar carcinoma subtype, predict response to erlotinib. J Clin Oncol 2008; 26: 1472-1478.

[16] Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, Liu Y, Gevaert O, Wang K, Zhu Y, Zhou H, Liu Z and Tian J. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur Respir J 2019; 53: 1800986.

[17] Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, Wang Y, Dai M and Han J. Value of pretherapy (18)F-FDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2020; 47: 1137-1146.
[18] Liu Q, Sun D, Li N, Kim J, Feng D, Huang G, Wang L and Song S. Predicting EGFR mutation subtypes in lung adenocarcinoma using (18) F-FDG PET/CT radiomic features. Transl Lung Cancer Res 2020; 9: 549-562.
[19] Moding EJ, Diehn M and Wakelee HA. Circulating tumor DNA testing in advanced non-small cell lung cancer. Lung Cancer 2018; 119: 42-47.
[20] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A and Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012; 48: 441-446.
[21] Lv Z, Fan J, Xu J, Wu F, Huang Q, Guo M, Liao T, Liu S, Lan X, Liao S, Geng W and Jin Y. Value of (18)F-FDG PET/CT for predicting EGFR mutations and positive ALK expression in patients with non-small cell lung cancer: a retrospective analysis of 849 Chinese patients. Eur J Nucl Med Mol Imaging 2018; 45: 735-750.
[22] Ren H, Xu W, You J, Song X, Huang H, Zhao N, Ren X and Zhang X. Analysis of the role of PET/CT SUVmax in prognosis and its correlation with clinicopathological characteristics in resectable lung squamous cell carcinoma. Zhongguo Fei Ai Za Zhi 2016; 19: 192-199.
[23] Minamimoto R, Jamali M, Gevaert O, Echegaray S, Khuong A, Hoang CD, Shrager JB, Plevritis SK, Rubin DL, Leung AN, Napel S and Quon A. Prediction of EGFR and KRAS mutation in non-small cell lung cancer using quantitative (18)F FDG-PET/CT metrics. Oncotarget 2017; 8: 52792-52801.
[24] Li J, Zhang B, Ge S, Deng S, Hu C and Sang S. Prognostic value of (18)F-FDG PET/CT radiomic model based on primary tumor in patients with non-small cell lung cancer: a large single-center cohort study. Front Oncol 2022; 12: 1047905.
[25] Kim G, Kim J, Cha H, Park WY, Ahn JS, Ahn MJ, Park K, Park YJ, Choi JY, Lee KH, Lee SH and Moon SH. Metabolic radiogenomics in lung cancer: associations between FDG PET image features and oncogenic signaling pathway alterations. Sci Rep 2020; 10: 13231.
[26] Desseroit MC, Visvikis D, Tixier F, Majdoub M, Perdrisot R, Guillevin R, Cheze Le Rest C and Hatt M. Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I-III. Eur J Nucl Med Mol Imaging 2016; 43: 1477-1485.
[27] Putora PM, Szentesi K, Glatzer M, Rodriguez R, Müller J, Baty F and Früh M. SUVmax and tumour location in PET-CT predict oncogene status in lung cancer. Oncol Res Treat 2016; 39: 681-686.
[28] Wang D, Zhang M, Gao X and Yu L. Prognostic value of baseline 18F-FDG PET/CT functional parameters in patients with advanced lung adenocarcinoma stratified by EGFR mutation status. PLoS One 2016; 11: e0158307.
[29] Zhou G, Xu S, Liu X, Ge J, He Q, Cao W, Ding J and Kai X. Relationship between the image characteristics of artificial intelligence and EGFR gene mutation in lung adenocarcinoma. Front Genet 2022; 13: 1090180.
[30] Hasegawa M, Sakai F, Ishikawa R, Kimura F, Ishida H and Kobayashi K. CT features of epidermal growth factor receptor-mutated adenocarcinoma of the lung: comparison with non-mutated adenocarcinoma. J Thorac Oncol 2016; 11: 819-826.
[31] Chicklore S, Goh V, Siddique M, Roy A, Marsden PK and Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging 2013; 40: 133-140.
[32] Li S, Li Y, Zhao M, Wang P and Xin J. Combination of (18)F-fluorodeoxyglucose PET/CT radiomics and clinical features for predicting epidermal growth factor receptor mutations in lung adenocarcinoma. Korean J Radiol 2022; 23: 921-930.
[33] Cho A, Hur J, Moon YW, Hong SR, Suh YJ, Kim YJ, Im DJ, Hong YJ, Lee HJ, Kim YJ, Shim HS, Lee JS, Kim JH and Choi BW. Correlation between EGFR gene mutation, cytologic tumor markers, 18F-FDG uptake in non-small cell lung cancer. BMC Cancer 2016; 16: 224.
[34] Ko KH, Hsu HH, Huang TW, Gao HW, Shen DH, Chang WC, Hsu YC, Chang TH, Chu CM, Ho CL and Chang H. Value of ^{18}F-FDG uptake on PET/CT and CEA level to predict epidermal growth factor receptor mutations in pulmonary adenocarcinoma. Eur J Nucl Med Mol Imaging 2014; 41: 1889-1897.
[35] Zhang M, Bao Y, Rui W, Shangguan C, Liu J, Xu J, Lin X, Zhang M, Huang X, Zhou Y, Qu Q, Meng H, Qian D and Li B. Performance of (18)F-FDG PET/CT radiomics for predicting EGFR mutation status in patients with non-small cell lung cancer. Front Oncol 2020; 10: 568857.
[36] Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, Li Q, Zhang D, Liu S and Li Z. Radiomics signature: a potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. Lung Cancer 2019; 132: 28-35.
[37] Chen Q, Li Y, Cheng Q, Van Valkenburgh J, Sun X, Zheng C, Zhang R and Yuan R. EGFR muta-

tion status and subtypes predicted by CT-based 3D radiomic features in lung adenocarcinoma. Onco Targets Ther 2022; 15: 597-608.
[38] Ozkan E, West A, Dedelow JA, Chu BF, Zhao W, Yildiz VO, Otterson GA, Shilo K, Ghosh S, King M, White RD and Erdal BS. CT gray-level texture analysis as a quantitative imaging biomarker of epidermal growth factor receptor mutation status in adenocarcinoma of the lung. AJR Am J Roentgenol 2015; 205: 1016-1025.
[39] Li M, Zhang L, Tang W, Jin YJ, Qi LL and Wu N. Identification of epidermal growth factor receptor mutations in pulmonary adenocarcinoma using dual-energy spectral computed tomography. Eur Radiol 2019; 29: 2989-2997.
[40] Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, D'Amico TA, DeCamp MM, Dilling TJ, Dobelbower M, Doebele RC, Govindan R, Gubens MA, Hennon M, Horn L, Komaki R, Lackner RP, Lanuti M, Leal TA, Leisch LJ, Lilenbaum R, Lin J, Loo BW Jr, Martins R, Otterson GA, Reckamp K, Riely GJ, Schild SE, Shapiro TA, Stevenson J, Swanson SJ, Tauer K, Yang SC, Gregory K and Hughes M. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2017; 15: 504-535.
[41] Huang CT, Yen RF, Cheng MF, Hsu YC, Wei PF, Tsai YJ, Tsai MF, Shih JY, Yang CH and Yang PC. Correlation of F-18 fluorodeoxyglucose-positron emission tomography maximal standardized uptake value and EGFR mutations in advanced lung adenocarcinoma. Med Oncol 2010; 27: 9-15.
[42] Mak RH, Digumarthy SR, Muzikansky A, Engelman JA, Shepard JA, Choi NC and Sequist LV. Role of 18F-fluorodeoxyglucose positron emission tomography in predicting epidermal growth factor receptor mutations in non-small cell lung cancer. Oncologist 2011; 16: 319-326.
[43] Na II, Byun BH, Kim KM, Cheon GJ, Choe du H, Koh JS, Lee DY, Ryoo BY, Baek H, Lim SM, Yang SH, Kim CH and Lee JC. 18F-FDG uptake and EGFR mutations in patients with non-small cell lung cancer: a single-institution retrospective analysis. Lung Cancer 2010; 67: 76-80.
[44] Putora PM, Früh M and Müller J. FDG-PET SUVmax values do not correlate with epidermal growth factor receptor mutation status in lung adenocarcinoma. Respirology 2013; 18: 734-735.
[45] Chung HW, Lee KY, Kim HJ, Kim WS and So Y. FDG PET/CT metabolic tumor volume and total lesion glycolysis predict prognosis in patients with advanced lung adenocarcinoma. J Cancer Res Clin Oncol 2014; 140: 89-98.

Supplementary Material

Feature list	(128 features)
MORPHOLOGICAL 17	MORPHOLOGICAL_Volume MORPHOLOGICAL_ApproximateVolume MORPHOLOGICAL_voxelsCounting MORPHOLOGICAL_SurfaceArea MORPHOLOGICAL_SurfaceToVolumeRatio MORPHOLOGICAL_Compacity MORPHOLOGICAL_Compactness1 MORPHOLOGICAL_Compactness2 MORPHOLOGICAL_SphericalDisproportion MORPHOLOGICAL_Sphericity MORPHOLOGICAL_Asphericity MORPHOLOGICAL_MaxValueCoordinates MORPHOLOGICAL_CenterOfMass MORPHOLOGICAL_WeightedCenterOfMass MORPHOLOGICAL_CentreOfMassShift MORPHOLOGICAL_Maximum3DDiameter MORPHOLOGICAL_IntegratedIntensity
BASIC 25	BASED_Mean BASED_Variance BASED_Skewness BASED_Kurtosis BASED_Median BASED_MinimumGreyLevel BASED_10thPercentile BASED_25thPercentile BASED_50thPercentile BASED_75thPercentile BASED_90thPercentile BASED_StandardDeviation BASED_MaximumGreyLevel BASED_InterquartileRange BASED_Range BASED_MeanAbsoluteDeviation BASED_RobustMeanAbsoluteDeviation BASED_MedianAbsoluteDeviation BASED_CoefficientOfVariation BASED_QuartileCoefficientOfDispersion BASED_AreaUnderCurveCsh BASED_Energy BASED_RootMeanSquare BASED_TotalLesionGlycolysis (only for PET) BASED_TotalCalciumScore (only for CT)

HISTOGRAM 30	HISTOGRAM_IntensityHistogramMean HISTOGRAM_IntensityHistogramVariance HISTOGRAM_IntensityHistogramSkewness HISTOGRAM_IntensityHistogramKurtosis HISTOGRAM_IntensityHistogramMedian HISTOGRAM_IntensityHistogramMinimumGreyLevel HISTOGRAM_IntensityHistogram10thPercentile HISTOGRAM_IntensityHistogram25thPercentile HISTOGRAM_IntensityHistogram50thPercentile HISTOGRAM_IntensityHistogram75thPercentile HISTOGRAM_IntensityHistogram90thPercentile HISTOGRAM_IntensityHistogramStandardDeviation HISTOGRAM_IntensityHistogramMaximumGreyLevel HISTOGRAM_IntensityHistogramMode HISTOGRAM_IntensityHistogramInterquartileRange HISTOGRAM_IntensityHistogramRange HISTOGRAM_IntensityHistogramMeanAbsoluteDeviation HISTOGRAM_IntensityHistogramRobustMeanAbsoluteDeviation HISTOGRAM_IntensityHistogramMedianAbsoluteDeviation HISTOGRAM_IntensityHistogramCoefficientOfVariation HISTOGRAM_IntensityHistogramQuartileCoefficientOfDispersion HISTOGRAM_IntensityHistogramEntropyLog10 HISTOGRAM_IntensityHistogramEntropyLog2 HISTOGRAM_AreaUnderCurveCsh HISTOGRAM_Uniformity HISTOGRAM_RootMeanSquare HISTOGRAM_MaximumHistogramGradient HISTOGRAM_MaximumHistogramGradientGreyLevel HISTOGRAM_MinimumHistogramGradient HISTOGRAM_MinimumHistogramGradientGreyLevel
Grey Level Co-occurrence Matrix (GLCM) 24	GLCM_JointMaximum GLCM_JointAverage GLCM_JointVariance GLCM_JointEntropyLog2 GLCM_JointEntropyLog10 GLCM_DifferenceAverage GLCM_DifferenceVariance GLCM_DifferenceEntropy GLCM_SumAverage GLCM_SumVariance GLCM_SumEntropy GLCM_AngularSecondMoment GLCM_Contrast GLCM_Dissimilarity GLCM_InverseDifference GLCM_NormalisedInverseDifference GLCM_InverseDifferenceMoment GLCM_NormalisedInverseDifferenceMoment GLCM_InverseVariance GLCM_Correlation GLCM_Autocorrelation GLCM_ClusterTendency GLCM_ClusterShade GLCM_ClusterProminence

Grey-Level Run Length Matrix (GLRLM) 11	GLRLM_ShortRunsEmphasis GLRLM_LongRunsEmphasis GLRLM_LowGreyLevelRunEmphasis GLRLM_HighGreyLevelRunEmphasis GLRLM_ShortRunLowGreyLevelEmphasis GLRLM_ShortRunHighGreyLevelEmphasis GLRLM_LongRunLowGreyLevelEmphasis GLRLM_LongRunHighGreyLevelEmphasis GLRLM_GreyLevelNonUniformity GLRLM_RunLengthNonUniformity GLRLM_RunPercentage
Neighborhood Grey-Level Different Matrix (NGLDM) 5	NGTDM_Coarseness NGTDM_Contrast NGTDM_Busyness NGTDM_Complexity NGTDM_Strength
Grey-Level Zone Length Matrix (GLZLM) 16	GLSZM_SmallZoneEmphasis GLSZM_LargeZoneEmphasis GLSZM_LowGrayLevelZoneEmphasis GLSZM_HighGrayLevelZoneEmphasis GLSZM_SmallZoneLowGreyLevelEmphasis GLSZM_SmallZoneHighGreyLevelEmphasis GLSZM_LargeZoneLowGreyLevelEmphasis GLSZM_LargeZoneHighGreyLevelEmphasis GLSZM_GreyLevelNonUniformity GLSZM_NormalisedGreyLevelNonUniformity GLSZM_ZoneSizeNonUniformity GLSZM_NormalisedZoneSizeNonUniformity GLSZM_ZonePercentage GLSZM_GreyLevelVariance GLSZM_ZoneSizeVariance GLSZM_ZoneSizeEntropy