Performance of ^{18}F-FDG PET/CT Radiomics for Predicting EGFR Mutation Status in Patients With Non-Small Cell Lung Cancer

Min Zhang[1†], Yiming Bao[2†], Weiwei Rui[3], Chengfang Shangguan[4], Jiajun Liu[1], Jianwei Xu[2], Xiaozhu Lin[1], Miao Zhang[1], Xinyun Huang[1], Yilei Zhou[1], Qian Qu[1], Hongping Meng[1], Dahong Qian[1,2*] and Biao Li[1*]

[1] Department of Nuclear Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China, [3] Department of Pathology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [4] Department of Oncology, Rujin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Objective: To assess the performance of pretreatment ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) radiomics features for predicting EGFR mutation status in patients with non-small cell lung cancer (NSCLC).

Patients and Methods: We enrolled total 173 patients with histologically proven NSCLC who underwent preoperative ^{18}F-FDG PET/CT. Tumor tissues of all patients were tested for EGFR mutation status. A PET/CT radiomics prediction model was established through multi-step feature selection. The predictive performances of radiomics model, clinical features and conventional PET-derived semi-quantitative parameters were compared using receiver operating curves (ROCs) analysis.

Results: Four CT and two PET radiomics features were finally selected to build the PET/CT radiomics model. Compared with area under the ROC curve (AUC) equal to 0.664, 0.683 and 0.662 for clinical features, maximum standardized uptake values (SUV$_{max}$) and total lesion glycolysis (TLG), the PET/CT radiomics model showed better performance to discriminate between EGFR positive and negative mutations with the AUC of 0.769 and the accuracy of 67.06% after 10-fold cross-validation. The combined model, based on the PET/CT radiomics and clinical feature (gender) further improved the AUC to 0.827 and the accuracy to 75.29%. Only one PET radiomics feature demonstrated significant but low predictive ability (AUC = 0.661) for differentiating 19 Del from 21 L858R mutation subtypes.

Conclusions: EGFR mutations status in patients with NSCLC could be well predicted by the combined model based on ^{18}F-FDG PET/CT radiomics and clinical feature, providing an alternative useful method for the selection of targeted therapy.

Keywords: positron emission tomography/computed tomography, radiomics, lung cancer, epidermal growth factor receptor, ^{18}F-fluorodeoxyglucose

INTRODUCTION

Lung cancer is the leading cause of cancer-related death in the world (1). Non-small cell lung cancer (NSCLC) accounts for approximately 80% to 85% of all lung cancers (2). Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) has become a first-line drug in the treatment of NSCLC. Because the efficacy of TKI therapy is closely related to EGFR mutation status, identification of mutation status before the administration of TKI is crucial in achieving the best curative effect. Furthermore, exon 19 deletion (19 del) and exon 21 L858R point mutation (21 L858R),the most common mutation subtypes of EGFR (3), demonstrate different clinical outcomes in patients with NSCLC after TKI treatment (4, 5). *Current molecular testing for identifying EGFR mutation status is mainly based on tumor tissue from biopsies and surgical resection* (6). However, focal tissue testing may sometimes be limited by invasive procedures or tissue samples that are not readily available (7), causing patients to lose potential opportunities for EGFR-TKI treatment.

Medical imaging can reflect tumor gene-driven phenotype (8). ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET/CT, as a noninvasive molecular imaging tool, has been widely used in the evaluation of glucose metabolic phenotype of tumor (6). Previous data has suggested that several genes associated with glucose metabolism, including GLUT1 (9), GPI, G6PD, PKM2, and GAPDH (10), are down-regulated in EGFR-mutated lung cancer. Therefore, numerous studies have explored the relationship between ^{18}F-FDG PET/CT images and EGFR mutation status. Some studies suggested that there was significantly lower maximum standardized uptake values (SUV_{max}) of NSCLCs with EGFR mutations than those with wild type (11–14), but other studies reported non-significant (15) or opposite results (16). These confusing findings may be related to intra-tumoral heterogeneity of EGFR mutation (17) that the PET-derived semi-quantitative parameters cannot well reflect.

Radiomics data obtained using mathematical algorithms can quantitatively describe the spatial relationship between voxels, and become an important tool to study tumor heterogeneity *in vivo* (18). To date, most studies using radiomics for the prediction of EGFR mutation status in NSCLC are based on chest CT images (19, 20), whereas few studies about the relationship between PET or PET/CT radiomics features and EGFR mutation status in lung cancer (21–23) are conducted.

In the present study, both PET and CT radiomics features that significantly discriminated EGFR mutation status were extracted and selected for establishing a robust predictive model. Then we compared the predictive performances of the radiomics model, clinical features, and conventional PET-derived semi-quantitative parameters. Moreover, we tried to investigate the possibility of PET/CT radiomics features for distinguishing the 19 del from the 21 L858R mutation which both are two main mutation subtypes of EGFR.

MATERIALS AND METHODS

Subjects

A total of 173 patients (115 men, 58 women; mean [± SD] age 60.9 ± 10.9 years [range, 27–86 years]) with histologically proven NSCLC, who had undergone pre-treatment ^{18}F-FDG PET/CT between January 2017 and March 2018, were included in this study. This retrospective study was approved by the Ethics Committee of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine.

^{18}F-FDG PET/CT Imaging

Patients were required to fast for at least 6 h before ^{18}F-FDG PET/CT scan using GE Discovery VCT64 system, and their serum glucose levels were maintained to < 7.8 mmol/L. Whole-body imaging was performed approximately 60 min after the intravenous administration of 5.55 MBq of ^{18}F-FDG per kilogram of body weight. Emission images were acquired for 3 min per bed position using 128 × 128 matrix size, 28 subsets, 2 iterations and full-width half-maximum post-filtering. CT images were acquired using 140 kV tube voltage, 220 mA tube current, and 3.75 mm section thickness. PET images were reconstructed based on an ordered-subset expectation maximization algorithm with photon attenuation correction from CT data.

EGFR Mutation Status Analysis

Tissue samples from lung tumors were obtained through biopsy or surgical resection followed by 10% formalin fixation, paraffin embedding, and sectioning. After extracting DNA from sample sections, the nucleotide sequence encoding the kinase domain (exons 18-21) of EGFR was tested using an amplification refractory mutation system polymerase chain reaction (24) or target sequencing method based on polymerase chain reaction (25) using the X10 system (Illumina, San Diego, CA, USA).

PET/CT Image Feature Extraction, Selection and Model Establishment

All segmentation was performed by experienced nuclear medicine physicians blinded to the mutation data using an open-source ITK-SNAP software (version 3.6, https://www.itksnap.org) to manually outline the contour of the volume of interest on CT images, and automatically delineated on PET images using a fixed SUV_{max} threshold of 2.5 as previously reported (21). The extraction and selection of radiomics features were performed according to the following steps (**Figure 1**):

1. Before extraction of radiomics features, filters including Laplacian of Gaussian, wavelet, square, square root, logarithm and exponential (26) (**Supplemental Table 1**), were applied to the original PET/CT images to highlight image features for more efficient feature extraction.
2. Based on the original and filtered PET/CT images mentioned above, several types of well-designed image features were calculated using pyradiomics python package (27). These features are designed in compliance with the Image Biomarker Standardization Initiative (28) including First-order statistics, Shape, Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Dependence Matrix (GLDM), Gray Level Run Length Matrix (GLRLM) and Neighboring Gray Tone Difference
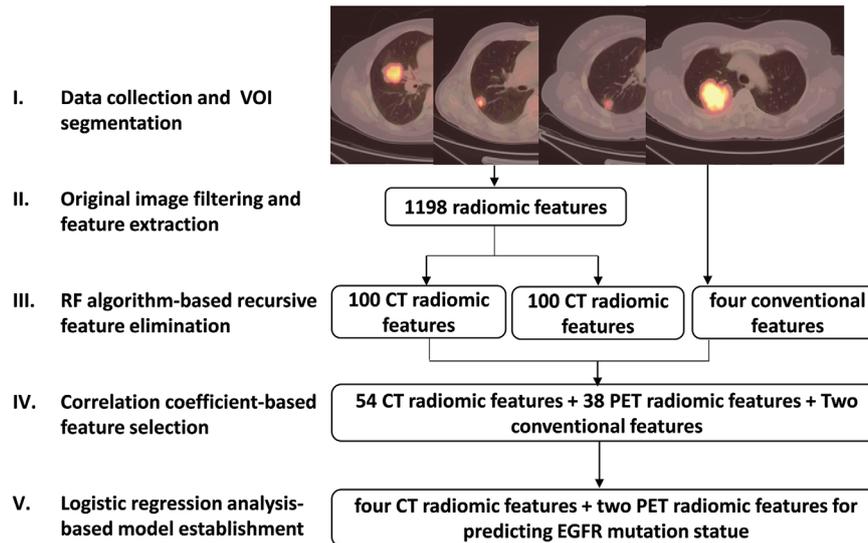
FIGURE 1 | Schematic diagram of image feature extraction and selection steps.

Matrix (NGTDM) (**Supplemental Table 2**). A total of 1198 PET and CT radiomics features were then extracted.

3. A recursive feature elimination (29) method based on random forest (RF) algorithm was developed to delete features with minimum weight coefficient. Compared to other regularization based embedded methods like Lasso and Ridge, this random forest-based wrapper feature selection method is more convenient and more intuitive for researchers to find out the most relevant features corresponding to the predication target. Among 1198 radiomics features, that with the lowest correlation with EGFR mutation status was removed during current random forest model training iteration, and the most suitable feature sub-package was reserved for next iteration. Finally, 100 CT and 100 PET radiomics features were retained (**Supplemental Table 3**).

4. The Spearman correlation coefficient (r) was used to assess the correlation between 100 PET/CT radiomics features and four conventional PET-derived semi-quantitative parameters including SUV_{max}, mean SUV (SUV_{mean}), metabolic tumor volume (MTV) and total lesion glycolysis (TLG) (illustrated in **Supplemental Figure 1**). In all pair features with r > 0.85 that were highly correlated and likely to provide redundance rather than complementary information about the mutation status, the one with the lower area under the curve (AUC) by receiver operating characteristic (ROC) analysis for predicting EGFR mutation status was excluded. As a result, 54 CT and 38 PET radiomics features as well as SUV_{max} and TLG were included.

5. The univariate and multivariate (**Supplemental Tables 4, 6**) logistic regression (LR) was ultimately used to screen out the CT and PET radiomics and clinical features that can be significant to establish a robust prediction model for differentiating EGFR mutation status, and the PET/CT

radiomics prediction score for EGFR mutation probability of each patient was calculated based on this model.

Statistical Analysis

Data were analyzed using SPSS version 19.0 (IBM Corporation, Armonk, NY, USA). Spearman correlation analysis was performed to remove redundant radiomics features. Continuous data were compared using the independent samples t test. The χ^2 test was used to compare categorical data such as patient sex. Univariate and multivariate logistic regression was used to screen out final significant variables. 10-fold cross-validation of prediction model based on selected features using machine learning algorithm of RF, support vector machine (SVM) or traditional statistics of LR were performed to test the generalization ability of the models. ROC curves were analyzed to evaluate the performance of PET/CT radiomics model for predicting EGFR mutation status. Statistical significance was set at $p < 0.05$.

RESULTS

Patient Characteristics

As shown in **Table 1**, 173 patients with NSCLC were enrolled in the present study, among whom 71 (41%) tested positively for an EGFR mutation (EGFR+) and 102 (59%) were EGFR-negative (EGFR-). Female patients demonstrated a significantly higher EGFR mutation rate (64% [37/58]) than male patients (30% [34/115]). There was no statistical difference in age between patients with or without EGFR mutations. 39% (68/173) and 61% (105/173) of patients were stage I/II and stage III/IV, respectively. Seventy-one percent (122/173) of the pathological types of NSCLCs were adenocarcinoma. Among the 71 patients who

TABLE 1 | Patient Characteristics.

	EGFR+	EGFR-	p Value
No. of patients	71	102	
Sex			
Male	34	81	<0.001
Female	37	21	
Age (y)			
Mean ± SD	60.06 ± 10.93	61.54 ± 10.89	NS
Range	27 ~ 86	32 ~ 83	
Clinical Stage			
I	18	28	NS
II	8	14	
III	19	23	
IV	26	37	
Histology			
Adenocarcinoma	60	62	0.004
Squamous cell carcinoma	8	31	
Large cell neuroendocrine carcinoma	0	4	
NSCLC-NOS	3	5	
EGFR mutation subtype			
18 G719S	3	/	
19 Del	29	/	
20 T790M	1	/	
21 L858R	38	/	

NOS, not otherwise specified; NS, not significant.

were EGFR+, 38 (54%) harbored the 21 L858R mutation, 29 (41%) had the 19 del mutation, 3 (4%) had the 18 G719A substitution mutation, and 1 (1%) had the 20 T790M substitution mutation. In all clinical features, only gender was an independent and significant variable for differentiating EGFR mutation status after multivariate logistic regression analysis (**Supplemental Table 4**).

Characteristic of Selected PET/CT Radiomics Features

Eventually, four CT and two PET radiomics features were selected to build the radiomics model based on the 173 patients, including ct_original_glszm_High Gray Level Zone Emphasis (GLSZM_HGLZE), ct_wavelet_HLL_glszm_Gray Level Non-Uniformity Normalized (GLSZM_GLNN), ct_wavelet_HLL_glszm_Zone Entropy (GLSZM_ZE), ct_exponential_gldm_Dependence Variance (GLDM_DV), pet_wavelet_LHH_firstorder_Skewness (First-order_Skewness (LHH)), pet_wavelet_LLL_firstorder_Skewness (First-order_Skewness (LLL)). The definitions of these selected radiomics features were shown in **Supplemental Table 5**. The PET/CT radiomics model prediction score for EGFR mutation probability of each patient was calculated using the following formula:

$$\text{PET/CT radiomics model prediction score} = -6.142 - 2.736$$
$$\times \text{GLSZM\_HGLZE} + 5.815 \times \text{GLSZM\_GLNUN} + 5.173$$
$$\times \text{GLSZM\_ZE} + 7.737 \times \text{GLDM\_DV} - 1.734 \times \text{First-order}$$
$$\_\text{Skewness} \quad (\text{LHH}) - 6.142 \times \text{First-order\_Skewness (LLL)}.$$

The median and the interquartile range for selected PET/CT radiomics features and conventional PET parameters (SUV_{max}

and TLG) was shown in **Table 2**. There was significant difference of every individual radiomics feature, SUV_{max} and TLG between the EGFR+ and EGFR- groups. Meanwhile, the tumors with EGFR+ had higher radiomics model score than those with EGFR- (0.722 vs. 0.170, $p < 0.001$). The PET/CT radiomics model prediction score for each patient was displayed in **Figure 2**.

Performance of the PET/CT Radiomics Model

The performance of PET/CT radiomics model was evaluated and compared with conventional PET-derived semi-quantitative parameters and clinical features for distinguishing EGFR+ from EGFR-. Both CT (AUC=0.792) and PET alone (AUC=0.738) radiomics model had better predictive performance than SUV_{max} (AUC=0.683), TLG (AUC=0.662) and gender (AUC=0.664). The AUC of PET/CT radiomics model further reached 0.868 with sensitivity of 92.8%, specificity of 66.3% and accuracy of 77.1%. Gender was only significant clinical predictor of EGFR mutation status (AUC=0.664), and used in the combined model in our study, whereas other clinical characteristics were excluded from the diagnostic model after multivariate regression analysis. The combined model, based on the PET/CT radiomics features and gender showed a comparable AUC (0.866) to PET/CT radiomics model. The sensitivity, specificity, and accuracy of different models and individual parameter in the training set were shown in **Table 3**. Subsequently, 10-fold cross-validation of the diagnostic model based on selected features using machine learning algorithm of SVM (**Table 3**), RF or traditional statistics of LR (**Supplemental Table 7**) were performed to further test the generalization ability of the models. The AUCs of PET radiomics, CT radiomics, PET/CT radiomics and combined models based on SVM were respectively 0.750, 0.754, 0.769 and 0.827.

In addition, we tried to investigate the possibility of radiomics features for discriminating two main mutation subtypes (**Table 4**). As previous reported, there was no difference of SUV_{max} or TLG between the 19 del and the 21 L858R mutation group. In all radiomics features, only one PET radiomics feature (pet_logarithm_glcm_Difference Variance, GLCM_DV) was significantly predictive (AUC=0.661) for differentiating these two mutation subtypes. However, it had low accuracy (43.1%) for the prediction of EGFR mutation subtypes.

DISCUSSION

EGFR-TKI is an important treatment for patients with NSCLC. When treated with TKI, patients with EGFR mutations experience significantly longer survival than those with wild-type EGFR. As such, identification of EGFR mutation status is crucial for TKI treatment to be effective; however, the molecular test for EGFR mutation status sometimes cannot be performed when a tumor sample is not available.

TABLE 2 | Characteristic of selected PET/CT radiomic features and conventional PET parameters.

Characteristic	EGFR- (N=102)	EGFR+ (N=71)	p Value
Conventional PET parameters			
SUV$_{max}$	11.500 (7.070-16.950)	6.900 (4.895-10.890)	<0.001
TLG	143.181 (25.241-358.192)	33.120 (8.854-168.031)	0.018
CT Radiomic features			
GLSZM_HGLZE	0.523 (0.353-0.659)	0.314 (0.240-0.445)	<0.001
GLDM_DV	0.390 (0.248-0.501)	0.530 (0.446-0.725)	<0.001
GLSZM_GLNUN	0.286 (0.218-0.379)	0.374 (0.283-0.483)	0.001
GLSZM_ZE	0.737 (0.610-0.849)	0.631 (0.479-0.725)	<0.001
PET Radiomic features			
First-order_Skewness (LHH)	0.561 (0.392-0.764)	0.374 (0.125-0.815)	0.019
First-order_Skewness (LLL)	1.008 (0.653-1.615)	0.773 (0.537-0.982)	<0.001
PET/CT Radiomic Score	0.170 (0.051-0.359)	0.722 (0.388-0.893)	<0.001

Data were expressed as median (interquartile range).
GLSZM, Gray Level Size Zone Matrix; GLDM, Gray Level Dependence Matrix; HGLZE, High Gray Level Zone Emphasis; DV, dependence variance; GLNUN, Gray Level Non Uniformity Normalized; ZE, zone entropy; LHH and LLL are two subtypes of wavelet filters.

Although a significant correlation between the tumor glucose metabolism level captured on PET images and EGFR mutation status has been found in multiple previous studies (11–14), namely lower SUV$_{max}$ in NSCLCs with EGFR mutation than those with wild type EGFR, conventional PET-derived semi-quantitative parameters didn't show enough satisfactory predictive ability to be applied in clinical practice. Consistent with previous studies, SUV$_{max}$ as a single pixel value only showed moderate AUC for distinguishing mutant EGFR from wild type in our study, whereas total lesion glycolysis (TLG) as a volumetric measurement of tumor glucose metabolism showed no higher predictive performance either. Therefore, our present study established a model based on ^{18}FDG PET/CT radiomics to improve the predictive performance for EGFR mutation status in patients with NSCLC.

In our study, four CT radiomics features and two PET features were selected to establish the predictive model with significantly higher AUC than that of SUV$_{max}$ and TLG. Among

FIGURE 2 | Distribution of PET/CT radiomic model prediction score of all patients. The tumors with EGFR+ had significantly higher score than those with EGFR- (p < 0.001).

these selected radiomics features, GLSZM_HGLZE from CT images measures the distribution of the higher gray-level values with a higher value indicating larger high-density areas proportion in tumor, which suggested that the tumors with EGFR+ had lower density than the EGFR- group in our study. In agreement with our finding, more ground-glass opacity and less solid components were observed in lung cancers with EGFR mutation (30) with lower mean CT values when compared to those with wild-type EGFR (31). The remaining 5 radiomics features, including three CT features (GLDM_DV, GLSZM_GLNN, GLSZM_ZE) and two PET features (First-order_Skewness (LHH), First-order_Skewness (LLL)), are all related to image uniformity and heterogeneity. In our study, the EGFR+ group was more heterogeneous on both PET and CT images than the EGFR- group. Our findings were similar to previous studies (21–23). They found that those image texture feature measuring the variability of gray-level intensity or the asymmetry of the distribution of gray-level values were significant predictive of EGFR mutation status. In summary, the NSCLCs with EGFR mutation had lower glucose metabolism and density, with more heterogeneity on both PET and CT images than those with wild-type EGFR. Owing to the bi-modal image features, PET/CT radiomics model in recent studies (0.79 in Zhang J's study (23); 0.80 in Li X's study (22); 0.77 in our study) has showed higher AUC than those generated by PET (0.67 in Yip, SS's study (21)) or CT (0.69 in Rios Velazquez, E's study (19); 0.56-0.75 in Sacconi, B's study (31)) radiomics features alone for predicting EGFR mutation status. However, compared with larger sample size in CT radiomics research, the current sample size in PET/CT radiomics-related studies is generally limited, and thus the generalization ability of PET/CT radiomics-based model remains to be further tested.

Clinical features in patients with NSCLC are also non-negligible variables in the evaluation of EGFR mutations, which are more likely to occur in Asians, adenocarcinomas, females, and nonsmokers (32). In our study, gender was only significant clinical predictor of EGFR mutation status. Smoking history was not included in our study due to the complexity of its definition, including the length of history, whether to quit or repeat smoking, etc. This complexity of smoking history made the simple
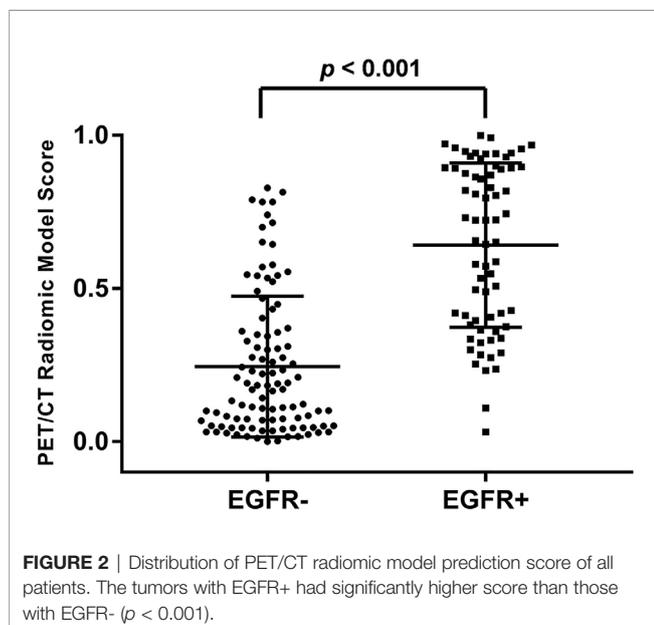
TABLE 3 | Predictive performance of EGFR mutation status using different models compared with conventional PET parameters and clinical feature.

Model/Parameters	Training set				10-fold cross validation using SVM algorithm			
	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)
Combined Model	0.866	82.60%	81.20%	81.80%	0.827	73.74%	76.07%	75.29%
PET/CT Radiomics Model	0.868	92.80%	66.30%	77.10%	0.769	67.11%	67.03%	67.06%
CT Radiomics Model	0.792	58.00%	87.10%	75.30%	0.754	64.22%	69.87%	67.65%
PET Radiomics Model	0.738	55.10%	82.20%	71.20%	0.750	60.29%	69.69%	67.06%
Gender	0.664	53.60%	79.20%	68.80%	/	/	/	/
SUV_{max}	0.683	84.10%	49.50%	63.50%	/	/	/	/
TLG	0.662	66.70%	64.40%	65.30%	/	/	/	/

SVM, support vector machine.

TABLE 4 | Predictive performance of EGFR mutation subtypes using PET/CT radiomic features compared with conventional PET parameters.

Parameters/Feature	21 L858R mutation(N =38)	19 Del mutation(N=29)	p Value	AUC	Sensitivity(%)	Specificity(%)	Accuracy(%)
SUV_{max}	7.5 (5.355-11.650)	6.695 (4.450-10.450)	0.134	/	/	/	/
TLG	37.98 (12.703-180.620)	26.014 (4.529-164.770)	0.408	/	/	/	/
GLCM_DV	1134.093 (801.011-1667.094)	808.42 (433.669-1353.409)	0.016	0.661	75.70%	57.10%	43.10%

Data were expressed as median (interquartile range).
GLCM, Gray Level Co-occurrence Matrix; DV, dependence variance.

classification of yes or no meaningless. Gender as only clinical characteristic was selected in combine model of our study. The addition of clinical characteristics to PET/CT radiomics model, to varying degrees, increase the diagnostic performance of diagnostic model in previous studies (22, 23) and our study, which finally reached 82.6% of diagnostic accuracy in Li X's study (22), 80.0% in Zhang J's study (23), and 75.3% in our study for predicting EGFR mutation status. It suggested that the combined model might be an alternative indicator of EGFR mutations when tissue samples are not available.

The 19 del and 21 L858R mutations are the main two EGFR mutation subtypes. Although both mutation subtypes are sensitive to EGFR-TKI treatment, it is now being recognized that patients with the 19 del mutation experience better clinical outcomes compared to those with the 21 L858R mutation (33, 34). Similar to a previous study investigating a large cohort of Chinese patients (14), we found that SUV_{max} or TLG had no ability to classify the 19 del and the 21 L858R mutation. We tried to investigate the possibility of PET/CT radiomics features for distinguishing these two subtypes. Only GLCM_DV from PET images, which measures the heterogeneity of different intensity-level matrix, showed significant but unsatisfactory predictive performance in our study (AUC=0.661). Liu Q, et al. recent study (35) established a predictive model for EGFR mutation subtypes using machine learning algorithm, which seemed to have better classification performance (AUC=0.77 and 0.92 for respectively predicting exons 19 del and 21 L858R mutations) than ours. However, the number of exons 19 del and 21 L858R mutations was small in Liu Q's study (only 44 and 31 cases respectively), especially when divided as the train and test cohorts, so the generalization ability of the predictive model was not clear.

The limited number of cases was one of the main factors restricting our research to obtain more reliable conclusions. Larger-scale data based on multi-center may be a solution in our next research. However, multi-center study may bring another important issue that affects the generalization ability of the model, that is, the variable PET imaging protocols among multi-centers including image acquisition and reconstruction conditions will be unable to ensure the uniformity and comparability of extracted radiomics features, thus affecting the sensitivity and specificity of radiomics model. Papp L, et al. suggested that larger matrix size/smaller voxel size, point-spread function reconstruction algorithms, and narrow Gaussian post-filtering helped minimize feature variations (36). The variability of PET radiomics is also feature-dependent. GLCM and shape features are the least sensitive to PET imaging system variations (36, 37). Although the single center study maintains the image acquisition and reconstruction methods consistent in all enrolled patients, thus avoiding the influence of the above-described factors as much as possible, the standardization of large databases from multi-centers will remain an unavoidable key step in further research.

In conclusion, EGFR mutations status in patients with NSCLC could be well predicted by the model based on ^{18}F-FDG PET/CT radiomics and clinical features, providing an alternative useful method for the selection of TKI therapy.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ruijin Hospital Ethics Committee Shanghai Jiao Tong University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.568857/full#supplementary-material

REFERENCES

1. Rosell R, Karachaliou N. Lung cancer in 2014: optimizing lung cancer treatment approaches. *Nat Rev Clin Oncol* (2015) 12(2):75–6. doi: 10.1038/nrclinonc.2014.225

2. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. *Adv Exp Med Biol* (2016) 893:1–19. doi: 10.1007/978-3-319-24223-1_1

3. Wu YL, Zhou C, Hu CP, Feng J, Lu S, Huang Y, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* (2014) 15(2):213–22. doi: 10.1016/S1470-2045(13)70604-1

4. Riely GJ, Pao W, Pham D, Li AR, Rizvi N, Venkatraman ES, et al. Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. *Clin Cancer Res* (2006) 12(3 Pt 1):839–44. doi: 10.1158/1078-0432.CCR-05-1846

5. Yang W, Gao Y, Li X, Zhang J, Liu T, Feng X, et al. Postoperative survival of EGFR-TKI-targeted therapy in non-small cell lung cancer patients with EGFR 19 or 21 mutations: a retrospective study. *World J Surg Oncol* (2017) 15(1):197. doi: 10.1186/s12957-017-1251-z

6. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, et al. Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2017) 15(4):504–35. doi: 10.6004/jnccn.2017.0050

7. Suda K, Murakami I, Sakai K, Tomizawa K, Mizuuchi H, Sato K, et al. Heterogeneity in resistance mechanisms causes shorter duration of epidermal growth factor receptor kinase inhibitor treatment in lung cancer. *Lung Cancer* (2016) 91:36–40. doi: 10.1016/j.lungcan.2015.11.016

8. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552

9. Chen L, Zhou Y, Tang X, Yang C, Tian Y, Xie R, et al. EGFR mutation decreases FDG uptake in nonsmall cell lung cancer via the NOX4/ROS/GLUT1 axis. *Int J Oncol* (2019) 54(1):370–80. doi: 10.3892/ijo.2018.4626

10. Takamochi K, Mogushi K, Kawaji H, Imashimizu K, Fukui M, Oh S, et al. Correlation of EGFR or KRAS mutation status with 18F-FDG uptake on PET-CT scan in lung adenocarcinoma. *PloS One* (2017) 12(4):e0175622. doi: 10.1371/journal.pone.0175622

11. Lee EY, Khong PL, Lee VH, Qian W, Yu X, Wong MP. Metabolic phenotype of stage IV lung adenocarcinoma: relationship with epidermal growth factor receptor mutation. *Clin Nucl Med* (2015) 40(3):e190–5. doi: 10.1097/RLU.0000000000000684

12. Cho A, Hur J, Moon YW, Hong SR, Suh YJ, Kim YJ, et al. Correlation between EGFR gene mutation, cytologic tumor markers, 18F-FDG uptake in non-small cell lung cancer. *BMC Cancer* (2016) 16:224. doi: 10.1186/s12885-016-2251-z

13. Yoshida T, Tanaka H, Kuroda H, Shimizu J, Horio Y, Sakao Y, et al. Standardized uptake value on (18)F-FDG-PET/CT is a predictor of EGFR T790M mutation status in patients with acquired resistance to EGFR-TKIs. *Lung Cancer* (2016) 100:14–9. doi: 10.1016/j.lungcan.2016.07.022

14. Lv Z, Fan J, Xu J, Wu F, Huang Q, Guo M, et al. 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(5):735–50. doi: 10.1007/s00259-017-3885-z

15. Lee SM, Bae SK, Jung SJ, Kim CK. FDG uptake in non-small cell lung cancer is not an independent predictor of EGFR or KRAS mutation status: a retrospective analysis of 206 patients. *Clin Nucl Med* (2015) 40(12):950–8. doi: 10.1097/RLU.0000000000000975

16. Ko KH, Hsu HH, Huang TW, Gao HW, Shen DH, Chang WC, et al. Value of (1)(8)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(10):1889–97. doi: 10.1007/s00259-014-2802-y

17. Zhang Y, Chang L, Yang Y, Fang W, Guan Y, Wu A, et al. Intratumor heterogeneity comparison among different subtypes of non-small-cell lung cancer through multi-region tissue and matched ctDNA sequencing. *Mol Cancer* (2019) 18(1):7. doi: 10.1186/s12943-019-0939-9

18. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol* (2016) 61(13):R150–166. doi: 10.1088/0031-9155/61/13/R150

19. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* (2017) 77(14):3922–30. doi: 10.1158/0008-5472.CAN-17-0122

20. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* (2019) 53(3):1800986. doi: 10.1183/13993003.00986-2018

21. Yip SS, Kim J, Coroller TP, Parmar C, Velazquez ER, Huynh E, et al. Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer. *J Nucl Med* (2017) 58(4):569–76. doi: 10.2967/jnumed.116.181826

22. Li X, Yin G, Zhang Y, Dai D, Liu J, Chen P, et al. 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. doi: 10.3389/fonc.2019.01062

23. Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of pre-therapy (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* (2019) 47(5):1137–46. doi: 10.1007/s00259-019-04592-1

24. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* (1989) 17(7):2503–16. doi: 10.1093/nar/17.7.2503

25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* (2012) 9(4):357–9. doi: 10.1038/nmeth.1923

26. Chen JS, Huertas A, Medioni G. Fast Convolution with Laplacian-of-Gaussian Masks. *IEEE Trans Pattern Anal Mach Intell* (1987) 9(4):584–90. doi: 10.1109/TPAMI.1987.4767946

27. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339

28. Zwanenburg A, Vallières M, Abdalah M, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145

29. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinf* (2009) 10:213. doi: 10.1186/1471-2105-10-213

30. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, et al. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology* (2016) 280(1):271–80. doi: 10.1148/radiol.2016151431

31. Sacconi B, Anzidei M, Leonardi A, Boni F, Saba L, Scipione R, et al. Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with EGFR mutations and survival rates. *Clin Radiol* (2017) 72(6):443–50. doi: 10.1016/j.crad.2017.01.015

32. Shi Y, Au JS, Thongprasert S, Srinivasan S, Tsai CM, Khoa MT, et al. 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(2):154–62. doi: 10.1097/JTO.0000000000000033

33. Yu JY, Yu SF, Wang SH, Bai H, Zhao J, An TT, et al. Clinical outcomes of EGFR-TKI treatment and genetic heterogeneity in lung adenocarcinoma patients with EGFR mutations on exons 19 and 21. *Chin J Cancer* (2016) 35:30. doi: 10.1186/s40880-016-0086-2

34. Zhou J, Ben S. Comparison of therapeutic effects of EGFR-tyrosine kinase inhibitors on 19Del and L858R mutations in advanced lung adenocarcinoma and effect on cellular immune function. *Thorac Cancer* (2018) 9(2):228–33. doi: 10.1111/1759-7714.12568

35. Liu Q, Sun D, Li N, Kim J, Feng D, Huang G, et al. Predicting EGFR mutation subtypes in lung adenocarcinoma using (18)F-FDG PET/CT radiomic features. *Transl Lung Cancer Res* (2020) 9(3):549–62. doi: 10.21037/tlcr.2020.04.17

36. Papp L, Rausch I, Grahovac M, Hacker M, Beyer T. Optimized Feature Extraction for Radiomics Analysis of (18)F-FDG PET Imaging. *J Nucl Med* (2019) 60(6):864–72. doi: 10.2967/jnumed.118.217612

37. Tsujikawa T, Rahman T, Yamamoto M, Yamada S, Tsuyoshi H, Kiyono Y, et al. (18)F-FDG PET radiomics approaches: comparing and clustering features in cervical cancer. *Ann Nucl Med* (2017) 31(9):678–85. doi: 10.1007/s12149-017-1199-7