Research Article

Establishment and Evaluation of EGFR Mutation Prediction Model Based on Tumor Markers and CT Features in NSCLC

Hao Zhang,[1] Meng He,[1] Ren'an Wan,[1] Liangming Zhu,[2] and Xiangpeng Chu [1]

[1]*Department of Thoracic Surgery, Rizhao People's Hospital, Rizhao 276800, Shandong, China*
[2]*Department of Thoracic Surgery, Jinan Central Hospital, Jinan 250013, Shandong, China*

Correspondence should be addressed to Xiangpeng Chu; chuxiangpeng@sdrzph.cn

Background. Lung cancer has become one of the leading causes of cancer deaths worldwide. EGFR gene mutation has been reported in up to 60% of Asian populations and is currently one of the main targets for genotype-targeted therapy for NSCLC. *Objective.* The objective is to determine if a complex model combining serum tumor makers and computed tomographic (CT) features can predict epidermal growth factor receptor (EGFR) mutation with higher accuracy. *Material and Methods.* Retrospective analysis of the data of patients diagnosed with in nonsmall cell lung cancer (NSCLC) by EGFR gene testing was carried out in the Department of Thoracic Surgery, Jinan Central Hospital. Multivariate logistic regression analysis was used to determine the independent predictors of EGFR mutations, and logistic regression prediction models were developed. The subject operating characteristic curve (ROC) was plotted, and the area under the curve (AUC) was calculated to assess the accuracy and clinical application of the EGFR mutation prediction model. *Results.* Logistic regression analysis identified the predictive factors of EGFR mutation including nonsmoking, high expression level of Carcinoembryonic Antigen (CEA), low expression level of cytokeratin 19 fragments (CYFRA21-1), and subsolid density containing ground-glass opacity (GGO) component. Using the results of multivariate logistic regression analysis, we built a statistically determined clinical prediction model. The AUC of the complex prediction model increased significantly from 0.735 to 0.813 ($p = 0.014$) when CT features are added and from 0.612 to 0.813 ($p < 0.001$) when serum variables are added. When P was 0.441, the sensitivity was 86.7% and the specificity was 65.8%. *Conclusion.* A complex model combining serum tumor makers and CT features is more accurate in predicting EGFR mutation status in NSCLC patients than using either serum variables or imaging features alone. Our finding for EGFR mutation is urgently needed and helpful in clinical practice.

1. Introduction

Lung cancer has become one of the leading causes of cancer deaths worldwide [1]. EGFR gene mutation has been reported in up to 60% of Asian populations and is currently one of the main targets for genotype-targeted therapy for NSCLC [2]. The discovery of structural domain activating mutations in EGFR tyrosine kinase promoted the concept of targeted therapy [3]. EGFR tyrosine kinase inhibitors (TKIs) were the first targeted agents for the treatment of NSCLC [4]. Currently, definitive diagnosis of EGFR mutation status is mainly detected from genomic DNA samples obtained from tumor tissues. However, tissue samples are difficult to obtain for EGFR mutation analysis, and tumor heterogeneity has an impact on accurate detection of EGFR mutations [5–7]. Circulating tumor DNA (ctDNA) in plasma samples can be an alternative method to detect EGFR mutations, but the results do not always agree with those of biopsy samples due to tumor heterogeneity, the false-negative rate is relatively high, and the cost is very high [4, 7, 8].

Serum tumor markers are mainly used in clinical practice to screen high-risk groups, observe and evaluate the effect of tumor treatment, and monitor the progression and recurrence of tumors. The results of the value of different serum tumor markers are inconsistent including significantly higher expression levels of CEA and carbohydrate antigen 199 (CA199), and lower expression level of CYFRA21-1 [4, 9–11].

In the diagnosis of lung cancer, CT is a routinely used and relatively cost-effective method that exhibits a variety of imaging features and can provide free data for genomics. Some studies have shown that the tumors typically presented with GGO were correlated with EGFR mutation [12, 13], while others found an inverse relationship or the lack of correlation [14]. In addition, other CT image features that have been found to be related with EGFR mutation include the maximum tumor diameter, spiculated margins, and the air bronchial sign [15, 16].

Although serum tumor markers and CT features show values in evaluating EGFR mutation status, the accuracy of these two methods by themselves is low. To our knowledge, it remains unclear whether the combination of these two methods can better diagnose EGFR mutation, and there are few studies combining these features to build predictive models. Therefore, we aim to construct a comprehensive model and evaluate its clinical application by analyzing serum tumor markers and CT imaging features in NSCLC patients.

2. Materials and Methods

2.1. Patient Selection. We included patients with pathologically diagnosed NSCLC between January 2018 and June 2020 at Jinan Central Hospital. Inclusion criteria included (1) pathological diagnosis of NSCLC by surgical resection, (2) serum tumor marker testing for lung cancer was performed at our hospital, (3) preoperative thin-section CT images, and (4) complete clinical information. Exclusion criteria included (1) no documented EGFR mutation testing, (2) history of previous antitumor therapy, (3) difficulty in outlining tumor margins, and (4) incomplete clinical data. Clinical data were collected, including patient gender, age, smoking history, pathological type, and clinical stage.

2.2. EGFR Mutation Detection. EGFR mutation was detected by experienced clinicians in the Department of Pathology of Jinan Central Hospital Hospital using surgically resected specimens. If an exon mutation was detected in EGFR exons 18–21, the tumor was considered to be EGFR mutant.

2.3. Measurement of Serum Tumor Maker Levels. The reference intervals for each index were CEA: 0–5 ng/ml, CYFRA21-1: 0–3.3 ng/ml, NSE: 0–16.3 ng/ml, CA125 : 0–35 U/ml, and CA199: 0–27 U/ml, respectively.

2.4. Image Acquisition and Feature Extraction. Images are viewed and analyzed by 2 imaging physicians in a double-blinded fashion on a PACS system. All examinations are extended in an intracranial direction with or without the use of contrast media. All images are archived in digital format. The following data were recorded: (1) maximum diameter (mm) of the lesion; (2) margins; (3) lesion density; and (4) lesion site. When 2 physicians disagreed, a higher level physician was asked to perform the analysis and reach a consensus result.

2.5. Statistical Analysis. The association of clinical characteristics, serum tumor marker levels, and CT image features with EGFR mutation was investigated by univariate analysis. The predictive factors of EGFR mutation was identified by logistic regression analysis and then built a statistically determined clinical prediction model. ROC was produced, and AUC was calculated to assess the accuracy of the prediction model. $p < 0.05$ was considered a statistically significant difference.

3. Results

3.1. Relationship between EGFR Mutation Status and Clinical Characteristics. A total of 148 NSCLC patients, 70 men (47.3%) and 78 women (52.7%), with a mean age of 62.9 ± 10.28 years, were included in this study. The patient EGFR mutation rate was 50.6%. The results showed that the mutation rate was significantly higher in women, in non-smoking patients, and in adenocarcinoma patients, with statistically significant differences ($p < 0.05$), while the differences in age and clinical stage of tumors were not statistically significant ($p > 0.05$) when compared between the mutant and wild-type groups.

3.2. Correlation of EGFR Mutation with Serum Tumor Markers. The results showed that CYFRA21-1 levels were significantly higher in the wild-type group ($p < 0.001$); CA199 levels were higher in the mutant group, with statistically significant differences ($p < 0.05$), while CA125, CEA, and NSE levels in the mutant group were not statistically significant compared with those in the wild-type group ($p > 0.05$).

3.3. Correlation of EGFR Mutation with CT Features. The results showed that the maximum tumor diameter was larger in the wild-type group, and the difference was statistically significant ($p < 0.05$); the proportion of semisolid (with ground-glass density) was significantly higher in the mutant group ($p < 0.001$); and the differences in lesion location, lobulated sign, and spiculated margins were not statistically significant when compared between the mutant and wild-type groups ($p > 0.05$).

3.4. Possible Predictors and Prediction Model. Univariate logistic regression analysis identified independent predictors with statistical significance (Tables 1–3). Multifactorial analysis was performed using dichotomous logistic regression, and the results showed that nonsmoking ($p = 0.003$), high CA199 expression ($p = 0.001$), low expression of CYFRA21-1 ($p < 0.001$), and semisolid density containing GGO component ($P = 0.003$) were independent risk factors for the development of mutations in the EGFR gene ($p < 0.05$), as detailed in Table 4.

The clinical prediction model was established: $p = e^{x}/(1 + e^{x})$, $X = -0.985 + (1.294 * \text{Smoking history}) + (1.625 * \text{CA199}) - (1.522 * \text{CYFRA21-1}) + (1.602 * \text{Density})$, where smoking history was 1 (No) or 0 (Yes); CA199 and

TABLE 1: Relationship between clinical characteristics and EGFR mutation.

Characteristics	EGFR wild-type ($n = 73$)	EGFR mutations ($n = 75$)	p
Gender			<0.001
Female	25 (34.2%)	53 (70.7%)	
Male	48 (65.8%)	22 (29.3%)	
Age (years)	63.59 ± 10.47	60.63 ± 9.96	0.08
Smoking history			<0.001
Yes	40 (54.8%)	13 (17.3%)	
No	33 (45.2%)	62 (82.7%)	
Pathology type			0.037
AC	63 (86.3%)	72 (96.0%)	
Non-AC	10 (13.7%)	3 (4.0%)	
Clinical stage			0.101
I-II	33 (45.2%)	44 (58.7%)	
III-IV	40 (54.8%)	31 (41.3%)	

TABLE 2: Correlation of EGFR mutation with serum tumor markers.

Serum tumor markers	EGFR wild-type ($n = 73$)	EGFR mutations ($n = 75$)	$\chi2$	p
CA125 (U/ml)			2.733	0.098
<35.0	52	62		
≥35.0	21	13		
CA199 (U/ml)			6.435	0.011
<27.0	57	44		
≥27.0	16	31		
CEA (ng/ml)			2.176	0.140
<5.0	39	49		
≥5.0	34	26		
NSE (ng/ml)			0.116	0.734
<16.3	60	60		
≥16.3	13	15		
CYFRA21-1 (ng/ml)			18.247	<0.001
<3.3	25	52		
≥3.3	48	23		

TABLE 3: Correlation of EGFR mutation with CT features.

CT features	EGFR wild-type ($n = 73$)	EGFR mutations ($n = 75$)	$t/\chi2$	p
Maximum diameter (mm)	33.76 ± 20.85	26.83 ± 15.32	2.310	0.022
Density			10.689	0.001
Subsolid	8	25		
Solid	65	50		
Lesion location			3.080	0.079
Central	25	16		
Peripheral	48	59		
Lobulated sign			0.740	0.390
Yes	28	34		
No	45	41		
Spiculated margins			0.988	0.320
Yes	32	39		
No	41	36		

Note. Subsolid tumor contains GGO component.

CYFRA21-1 was 1 (positive) or 0 (negative); and density was 1 (subsolid) or 0 (solid). The Hosmer–Lemeshow test was used for goodness of fit ($\chi^2 = 7.530$; $p = 0.376$), indicating a good fit of the model.

3.5. ROC to Assess Prediction Model Effectiveness. The area under the curve for serum tumor marker predictor and imaging index predictor ROC was 0.735 (95% CI: 0.654–0.817, $p < 0.001$) and 0.612 (95% CI: 0.521–0.703,

TABLE 4: Result of multivariate logistic regression analysis.

Factors	B	S.E	Wals	p	OR	95%CI
Smoking history	1.294	0.430	9.042	0.003	3.649	1.569–8.483
CA199	1.625	0.473	11.799	0.001	5.077	2.009–12.829
CYFRA21-1	−1.522	0.432	12.417	<0.001	0.218	0.094–0.509
Density	1.602	0.547	8.595	0.003	4.964	1.701–14.488
Constant	−0.958	0.438	4.789	—	—	–

$p < 0.05$), respectively. When serum tumor markers were combined with CT image features, the area under the curve of the integrated model was 0.813 (95% CI: 0.743–0.883, $p < 0.001$). The difference in AUC between the integrated model and the serum tumor marker predictors or imaging index predictors was statistically significant ($p < 0.05$). When the cut-off value was 0.441, the sensitivity Se was 86.7% and specificity SP was 65.8%, as shown in Figure 1.

4. Discussion

Although TKIs can improve the prognosis of patients with EGFR mutations and have significant efficacy in patients with gene mutations, the detection rate of EGFR mutation is lower than expected [17]. Biopsy, the gold standard for EGFR mutation detection, may be limited by the lack of available tissue samples because biopsy and cytology specimens are first used for histological testing to confirm cancer type. In addition, patient refusal to undergo invasive biopsy, location or size of the tumor, difficulty in biopsy sampling, and potential risk of cancer metastasis also limit detection rates [18]. In this study, we attempted to evaluate the effectiveness of a complex prediction model wherein serum markers and CT features are combined.

Previous demographic analyses have shown that a high prevalence of EGFR mutation is associated with female, nonsmokers, adenocarcinoma tissue type, and East Asian populations [4, 19], which is consistent with the our study results. Furthermore, we found that smoking history was an independent predictor of EGFR mutation by multivariate analysis, consistent with the study by Sabri et al. [20].

Serum tumor markers can be tested quickly and accurately in the hospital at a low cost [11]. Preoperative serum tumor markers have been shown to correlate with EGFR mutation and the efficacy of EGFR-TKI therapy [21]. Therefore, it is practical to use STMs to predict EGFR mutation [22]. Our current study demonstrated that the serum CA199 level was significantly higher in the EGFR mutation group while CYFRA21-1 was significantly increased in the wild-type group. Zhang et al. found that serum CEA levels could be used as a predictive tumor marker for the efficacy of EGFR-TKI therapy [23]. Our study did not identify a significant difference in the CEA levels, which is worth to note that over half (77 out of 148 cases) of our selected NSCLC patients were at an early stage of NSCLC development. The CA125 and NSE levels were not significantly altered by mutations in EGFR gene, consistent with a previous report [20].

CT is a routinely used and relatively economical modality for diagnosing lung cancer, and it presents a variety of
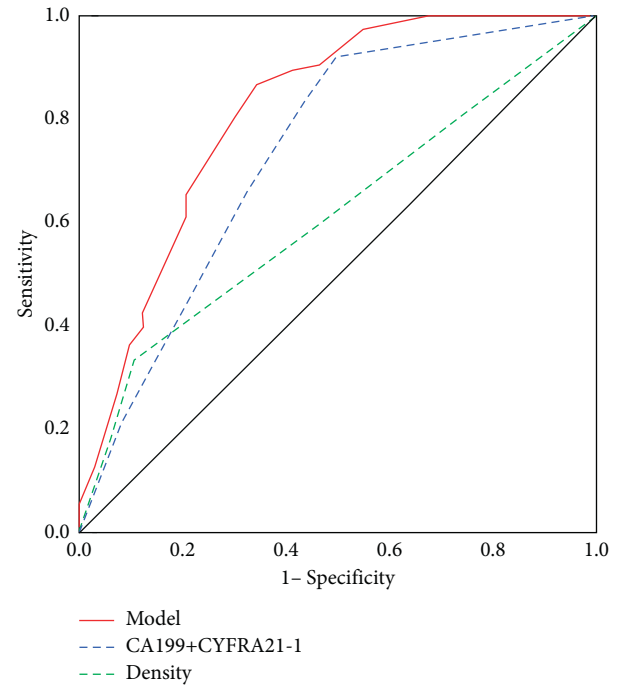

FIGURE 1: ROC curves for the complex model, serum predictors (CA199+CYFRA21-1), and imaging predictors (density) in differentiating EGFR mutation status.

imaging features that may be used to identify patients with NSCLC who are at risk for EGFR mutations. We found that EGFR mutation was associated with a smaller maximum diameter and subsolid density. Although univariate analysis found tumor size to be a factor associated with EGFR mutation, multifactorial analysis showed that tumor size was not a strong independent predictor, consistent with Rizzo et al. [24]. In addition, our study showed no significant differences in terms of lesion location, lobulated sign, and spiculated margins, which were consistent with some previous studies [20, 25] but contrary to the findings of Zhou et al. [26].

Although serum tumor markers and CT features can assess EGFR mutation status in NSCLC, the accuracy of predicting EGFR mutations by these two methods alone is not sufficient. To our knowledge, it remains unclear whether the combination of these two methods can better diagnose EGFR mutation status, and there are few clinical prediction models for EGFR mutations combining these features [20, 27]. We identified the independent predictors including nonsmoking, high CA199 expression, low CYFRA21-1 expression, and semisolid density containing a ground-glass opacity (GGO) component. The model was $p = e^x/(1 + e^x)$,

and $X = -0.985 + (1.294*\text{Smoking history}) + (1.625*\text{CA199}) - (1.522*\text{CYFRA21-1}) + (1.602*\text{Density})$. The AUC of the prediction model increased from 0.735 to 0.813 when CT feature predictors were added ($p = 0.014$) and from 0.612 to 0.813 when serum variables were added ($p < 0.001$). The cut-off point for 0.441 exhibited ideal sensitivity (86.7%) and acceptable specificity (65.8%).

5. Conclusion

In conclusion, EGFR mutation models constructed from serum tumor markers and CT features have good predictive efficacy. When properly combined, the complex model can have better predictive performance and higher diagnostic accuracy, facilitating clinical practice in identifying candidates for targeted therapy.

Our study has several limitations. First, it was a single-center retrospective study with a relatively small sample size and lack of external validation, which potentially compromise the generalization, sensitivity, and specificity of our model. Therefore, it needs to develop uniform standards for multicenter studies and to establish and test multicenter data. Second, squamous cell carcinoma (SCC) antigen was not included in this clinical prediction model due to incomplete documentation of these tumor markers in the HIS system. Therefore, we recommend refinement of our model prior to further validation. Finally, this study demonstrated that the integrated model has good predictive performance, but the accuracy is limited by its logistic regression method. Models built by different methods such as random forest and elastic network regression should be combined to develop the model.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [published correction appears in CA Cancer J Clin. 2020 Jul;70(4):313]," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] F. R. Hirsch, G. V. Scagliotti, J. L. Mulshine et al., "Lung cancer: current therapies and new targeted treatments," *The Lancet*, vol. 389, no. 10066, pp. 299–311, 2017.

[3] M. Zhou, D. M. Chen, L. Xia, Y. Y. Song, Y. K. Shi, and Z. M. Yang, "Clinical research and drug review of epidermal growth factor receptor tyrosine kinase inhibitors in advanced non-small cell lung cancer," *Zhonghua Zhongliu Zazhi*, vol. 41, no. 12, pp. 949–952, 2019.

[4] H. Chang, Y.-B. Liu, W. Yi, J.-B. Lu, and J.-X. Zhang, "Development and validation of a model to predict tyrosine kinase inhibitor-sensitive EGFR mutations of non-small cell lung cancer based on multi-institutional data," *Thoracic Cancer*, vol. 9, no. 12, pp. 1680–1686, 2018.

[5] B. Weber, P. Meldgaard, H. Hager et al., "Detection of EGFR mutations in plasma and biopsies from non-small cell lung cancer patients by allele-specific PCR assays," *BMC Cancer*, vol. 14, no. 1, p. 294, 2014.

[6] R. B. Shahi, S. De Brakeleer, J. De Grève, C. Geers, P. In't Veld, and E. Teugels, "Detection of EGFR-TK domain-activating mutations in NSCLC with generic PCR-based methods," *Applied Immunohistochemistry & Molecular Morphology*, vol. 23, no. 3, pp. 163–171, 2015.

[7] Y. Zhang, L. Chang, Y. Yang et al., "Intratumor heterogeneity comparison among different subtypes of non-small-cell lung cancer through multi-region tissue and matched ctDNA sequencing," *Molecular Cancer*, vol. 18, no. 1, p. 7, 2019.

[8] Q. Deng, Q. Fang, H. Sun et al., "Detection of plasma EGFR mutations for personalized treatment of lung cancer patients without pathologic diagnosis," *Cancer Medicine*, vol. 9, no. 6, pp. 2085–2095, 2020.

[9] J. B. Pan, Y. H. Hou, and G. J. Zhang, "Correlation between efficacy of the EGFR tyrosine kinase inhibitor and serum tumor markers in lung adenocarcinoma patients," *Clinical Laboratory*, vol. 60, pp. 1439–1447, 2014.

[10] S. Wang, P. Ma, G. Ma et al., "Value of serum tumor markers for predicting EGFR mutations and positive ALK expression in 1089 Chinese non-small-cell lung cancer patients: A retrospective analysis," *European Journal of Cancer*, vol. 124, pp. 1–14, 2020.

[11] L. X. Feng, J. Wang, Z. Yu et al., "Clinical significance of serum EGFR gene mutation and serum tumor markers in predicting tyrosine kinase inhibitor efficacy in lung adenocarcinoma," *Clinical and Translational Oncology*, vol. 21, no. 8, pp. 1005–1013, 2019.

[12] X. Han, J. Fan, J. Gu et al., "CT features associated with EGFR mutations and ALK positivity in patients with multiple primary lung adenocarcinomas," *Cancer Imaging*, vol. 20, no. 1, p. 51, 2020.

[13] Y. Liu, J. Kim, F. Qu et al., "CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma," *Radiology*, vol. 280, no. 1, pp. 271–280, 2016.

[14] Y. Chen, Y. Yang, L. Ma et al., "Prediction of EGFR mutations by conventional CT-features in advanced pulmonary adenocarcinoma," *European Journal of Radiology*, vol. 112, pp. 44–51, 2019.

[15] B. Sacconi, M. Anzidei, A. Leonardi et al., "Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: A correlation with EGFR mutations and survival rates," *Clinical Radiology*, vol. 72, no. 6, pp. 443–450, 2017.

[16] Z. Shi, X. Zheng, R. Shi et al., "Radiological and clinical features associated with epidermal growth factor receptor mutation status of exon 19 and 21 in lung adenocarcinoma," *Scientific Reports*, vol. 7, no. 1, p. 364, 2017.

[17] D. Que, H. Xiao, B. Zhao et al., "EGFR mutation status in plasma and tumor tissues in non-small cell lung cancer serves as a predictor of response to EGFR-TKI treatment," *Cancer Biology & Therapy*, vol. 17, no. 3, pp. 320–327, 2016.

[18] S. Wang, J. Shi, Z. Ye et al., "Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning," *European Respiratory Journal*, vol. 53, no. 3, Article ID 1800986, 2019.

[19] Y. Shi, J. S.-K. Au, S. Thongprasert et al., "A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER)," *Journal of Thoracic Oncology*, vol. 9, no. 2, pp. 154–162, 2014.

[20] A. Sabri, M. Batool, Z. Xu, D. Bethune, M. Abdolell, and D. Manos, "Predicting EGFR mutation status in lung cancer: Proposal for a scoring model using imaging and demographic characteristics," *European Radiology*, vol. 26, no. 11, pp. 4141–4147, 2016.

[21] R. Jiang, X. Wang, and K. Li, "Predictive and prognostic value of preoperative serum tumor markers is EGFR mutation-specific in resectable non-small-cell lung cancer," *Oncotarget*, vol. 7, no. 18, pp. 26823–26836, 2016.

[22] Z. Cai, "Relationship between serum carcinoembryonic antigen level and epidermal growth factor receptor mutations with the influence on the prognosis of non-small-cell lung cancer patients," *OncoTargets and Therapy*, vol. 9, pp. 3873–3878, 2016.

[23] Y. Zhang, B. Jin, M. Shao et al., "Monitoring of carcinoembryonic antigen levels is predictive of EGFR mutations and efficacy of EGFR-TKI in patients with lung adenocarcinoma," *Tumor Biology*, vol. 35, no. 5, pp. 4921–4928, 2014.

[24] S. Rizzo, F. Petrella, V. Buscarino et al., "CT radiogenomic characterization of EGFR, K-ras, and ALK mutations in non-small cell lung cancer," *European Radiology*, vol. 26, no. 1, pp. 32–42, 2016.

[25] M. Sugano, K. Shimizu, T. Nakano et al., "Correlation between computed tomography findings and epidermal growth factor receptor and KRAS gene mutations in patients with pulmonary adenocarcinoma," *Oncology Reports*, vol. 26, no. 5, pp. 1205–1211, 2011.

[26] J. Y. Zhou, J. Zheng, Z. F. Yu et al., "Comparative analysis of clinicoradiologic characteristics of lung adenocarcinomas with ALK rearrangements or EGFR mutations," *European Radiology*, vol. 25, no. 5, pp. 1257–1266, 2015.

[27] J. Zhang, X. Zhao, Y. Zhao et al., "Value of pre-therapy 18F-FDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 5, pp. 1137–1146, 2020.