Development and Validation of Machine Learning Models to Predict Epidermal Growth Factor Receptor Mutation in Non-Small Cell Lung Cancer: A Multi-Center Retrospective Radiomics Study

Yafeng Liu, MD[1,†] , Jiawei Zhou, MD[1,†], Jing Wu, PhD[1,2,*], Wenyang Wang, PhD[1], Xueqin Wang, MD[1], Jianqiang Guo, MD[1], Qingsen Wang, MD[1], Xin Zhang, MD[1], Danting Li, MD[1], Jun Xie, MD[3], Xuansheng Ding, PhD[1,4,5], Yingru Xing, PhD[1,6,**], and Dong Hu, PhD[1,2,3,***]

Abstract

Objective: To develop and validate a generalized prediction model that can classify epidermal growth factor receptor *(EGFR)* mutation status in non–small cell lung cancer patients.

Methods: A total of 346 patients (296 in the training cohort and 50 in the validation cohort) from four centers were included in this retrospective study. First, 1085 features were extracted using IBEX from the computed tomography images. The features were screened using the intraclass correlation coefficient, hypothesis tests and least absolute shrinkage and selection operator. Logistic regression (LR), decision tree (DT), random forest (RF), and support vector machine (SVM) were used to build a radiomics model for classification. The models were evaluated using the following metrics: area under the curve (AUC), calibration curve (CAL), decision curve analysis (DCA), concordance index (C-index), and Brier score.

Results: Sixteen features were selected, and models were built using LR, DT, RF, and SVM. In the training cohort, the AUCs was .723, .842, .995, and .883; In the validation cohort, the AUCs were .658, 0567, .88, and .765. RF model with the best AUC, its CAL, C-index (training cohort=.998; validation cohort=.883), and Brier score (training cohort=.007; validation cohort=0.137) showed a satisfactory predictive accuracy; DCA indicated that the RF model has better clinical application value.

Conclusion: Machine learning models based on computed tomography images can be used to evaluate *EGFR* status in patients with non–small cell lung cancer, and the RF model outperformed LR, DT, and SVM.

[1]School of Medicine, Anhui University of Science and Technology, Huainan, P.R. China
[2]Anhui Province Engineering Laboratory of Occupational Health and Safety, Anhui University of Science and Technology, Huainan, P.R. China
[3]Key Laboratory of Industrial Dust Prevention and Control & Occupational Safety and Health of the Ministry of Education, Anhui University of Science and Technology, Huainan, P.R. China
[4]Cancer Hospital of Anhui University of Science and Technology, Huainan, P.R. China
[5]School of Pharmacy, China Pharmaceutical University, Nanjing, China
[6]Department of Clinical Laboratory, Anhui Zhongke Gengjiu Hospital, Hefei, P.R. China

[†]These authors contributed equally to this work.

Corresponding Authors:
*Wu Jing, PhD, Anhui University of Science and Technology, No. 168 Taifeng Street, Huainan 232001, P.R. China.
Email: wujing8008l@126.com

**Xing Yingru, PhD, Anhui University of Science and Technology, No. 168 Taifeng Street, Huainan 232001, P.R. China.
Email: austxingyr@126.com

***Hu Dong, PhD, Anhui University of Science and Technology, No. 168 Taifeng Street, Huainan 232001, P.R. China.
Email: austhudong@126.com

Introduction

Approximately 85% of lung cancers are non–small cell lung cancers (NSCLC), which have high recurrence rates and poor prognosis.[1,2] In the treatment of NSCLC, first-line chemotherapy regimens are only 30% effective,[3] whereas the effectiveness of epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in patients with EGFR-sensitive mutations reaches 70%.[4] The presence of EGFR-sensitive mutations is a major predictor of the effectiveness of drugs with EGFR.[5]

Tissue biopsy to determine the *EGFR* gene status in NSCLC patients is extremely accurate; however, it has some limitations, such as difficulties in obtaining tissue samples and high economic costs.[6,7] With the rapid development of the most advanced artificial intelligence technology and radiomics,[8] high-throughput extraction of radiomics features from medical images is required to quantify the shape, intensity, and texture of tumors to comprehensively characterize the tumor phenotype,[9] and noninvasive radiomics models have shown great potential in diagnosis, prognosis, and genetic information.[10-12]

In recent years, the use of positron emission tomography/computed tomography (PET/CT) or enhanced CT images to forecast the status of *EGFR* mutations has promoted the progress of relevant studies.[13-17] However, due to differences in population distribution, living area, economy, and medical institution equipment capacity involved in separate studies, the research results in economically developed regions may not be suitable for the region where this research team is located. Therefore, in this retrospective study, we collected radiographic data from four centers involving populations with different demographic factors. Applying machine learning to radiomics constructs a strong generalization model to predict *EGFR* mutations in patients with NSCLC, providing a reference for clinical practice.

Data and Methods

Patient Imaging and Clinical Data. NSCLC radiogenomics data[18] were obtained from the Cancer Imaging Archive portal and included 211 patients. Among these, 129 patients had wild-type *EGFR*, 43 had *EGFR* mutations, and 39 had unknown genes. We included all patients who underwent chest CT scans and had known *EGFR* mutation status; 39 patients with alien genes and two patients in whom IBEX generated errors during feature extraction were excluded, leading to a total of 168 patients to be included in the study. Supplementary Data 1 (S1) contains information regarding the scanning parameters. The personal information of patients in medical materials has been anonymized. This study was conducted in accordance with the STROBE guidelines.[19]

In addition, we collected clinical and imaging data of patients with primary NSCLC between January 2016 and December 2020 at the Cancer Hospital of Anhui University of Science and Technology, the Eastern Hospital of Anhui University of Science and Technology, and the Huainan Chaoyang Hospital of Anhui University of Science and Technology, using the following inclusion criteria: (1) patients with pathologically proven NSCLC, (2) *EGFR* gene status testing performed on biopsy tissues, and (3) CT scans performed within 2 weeks before treatment. The exclusion criteria were as follows: (1) patients who received radiotherapy, chemotherapy, concurrent radiotherapy, or traditional Chinese medicine treatment before CT imaging and (2) incomplete image information of the patient. 86 patients from the Cancer Hospital of Anhui University of Science and Technology, 50 from the Eastern Hospital of Anhui University of Science and Technology, and 41 from Huainan Chaoyang Hospital of Anhui University of Science and Technology were included in compliance with the above conditions.

To improve the generalization ability of the model constructed from the heterogenous and complex dataset, 296 patients from the NSCLC radiogenomics data, Cancer Hospital of Anhui University of Science and Technology, and Huainan Chaoyang Hospital of Anhui University of Science and Technology were used as the training cohort, and 50 patients from the Eastern Hospital of Anhui University of Science and Technology were used as the validation cohort.

This retrospective study was conducted in accordance with the principles of the Helsinki Declaration. The Ethics Committee of Anhui University of Science and Technology (approval no. L2022001) conducted an ethical review of the three medical institutions involved (Cancer Hospital of Anhui University of Science and Technology, Eastern Hospital of Anhui University of Science and Technology, and Huainan Chaoyang Hospital of Anhui University of Science and Technology). Oral consent was obtained, and data were processed anonymously before conducting the study. The research flow is illustrated in Figure 1.

Image Segmentation, Image Pre-processing, and Feature Extraction

The collected CT images were uploaded to IBEX in Digital Imaging and Communication in Medicine (DICOM), and regions of interest (ROIs) were manually outlined layer-by-layer
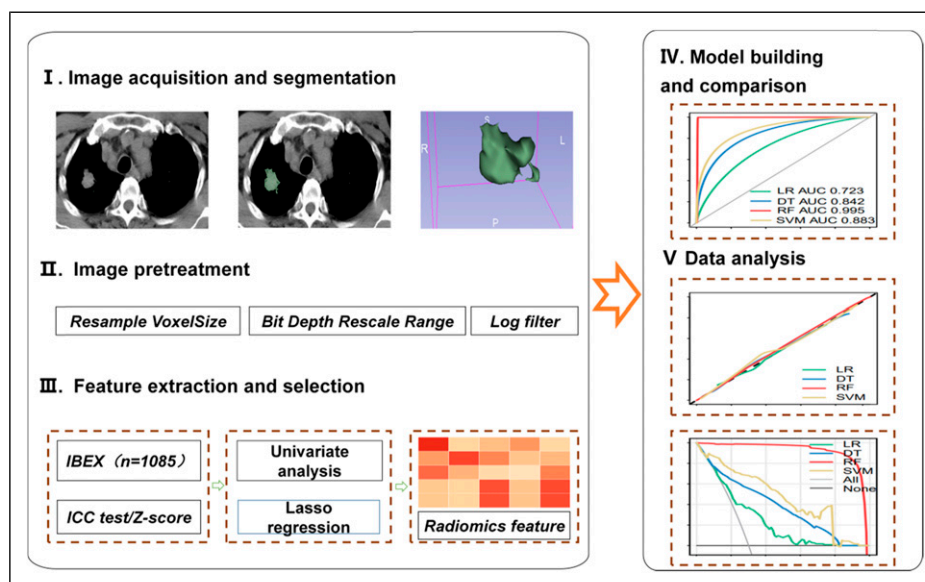
Figure 1. The overall framework of data analysis and model integration.

by two highly qualified diagnostic cardiothoracic disease imaging physicians (one 8 years and one 10 years working experience) without knowledge of the *EGFR* test results (lung window: 1500 HU, −500 HU; mediastinal window: 300 HU, 30 HU). After the sketch was completed, the images were preprocessed using resample voxel size, bit depth rescale range, and log filter in IBEX to achieve image-scale uniformity, correction of grayscale inhomogeneity, and image denoting.

Five types of radiomics features were extracted from the ROIs: (1) intensity histogram (n = 49), (2) shape (n = 18), (3) texture-based features including grayscale co-occurrence matrix (n = 840) features and gray level run length matrix (n = 33); (4) grayscale intensity (n = 135); and (5) neighborhood intensity difference (n = 10). Supplementary data 2 (S2) shows the kinds of features extracted in the 3D image.

Radiomics Feature Selection

Feature selection is important to improve model generalization and optimize the model.[20] The two physicians performed independent ROI delineation and feature extraction on all data. The features extracted by the two physicians were subjected to the ICC test to select features with stability and repeatability (ICC < .5, poor reliability; .5 < ICC < .75, medium reliability; .75 < ICC < .9, good reliability; and ICC > .9, excellent reliability).[21] Second, features with ICC > .75 were standardized using the Z-score method.

Third, the Shapiro–Wilk test ($P > .05$) and Bartlett's test ($P > .05$) were used to test the normality and homogeneity of variance of the features with ICC > .75. An independent sample T-test ($P < .05$) was used for the data in accordance with the normal distribution and homogeneity of variance, and the Mann–Whitney U test ($P < .05$) was used for the data.

Finally, to avoid overfitting or selection bias, LASSO regression verified following 10-fold cross-validation was used to screen out the radiomics features of the constructed model.

Machine Learning Model Construction and External Validation

After screening the core radiomics features, the four most popular machine learning classifiers (logistic regression (LR), decision tree (DT), random forest (RF), and radius-based function support vector machines (SVM)) were applied to construct imaging histology models in the training and validation cohorts. We applied an exhaustive grid search approach was applied to identify the values of the hyperparameters that optimize the model prediction performance. Supplementary data 3 (S3) shows the setting of hyperparameters of different machine learning classifiers. The area under the curve (AUC), calibration curve (CAL), decision curve analysis (DCA), concordance index (C-index), and Brier score were used to estimate the discrimination, calibration, and clinical applicability of models constructed using different classifiers. The C-index ranges from .5 to 1, with a C-index <.5 reflecting complete inconsistency, and the model has no predictive value and C-index = 1, reflecting complete consistency. The Brier score was used to measure the overall performance of the model; if the Brier score=0, the model was considered to have perfect overall performance, and the predicted and actual values were in perfect agreement. If the Brier score is >.25, the model was considered to have no value.

Statistical Analysis

All statistical analyses were performed using Empower Stats (version 2.2) and R software (version 4.0.5). Quantitative data

Table 1. Patients in the Training and Validation Cohorts.

Characteristic	Training				Validation n = 50[d]	P-value
	n = 169[a]	n = 86[b]	n = 41[c]	total = 296		
Age (y, mean ± SD)	67.65 ± 10.33	64.65 ± 11.65	67.93 ± 14.89	66.82 ± 11.49	66.56 ± 9.44	.675
EGFR status						.029
Wild type	126 (74.56%)	34 (39.53%)	19 (46.34%)	179 (60.47%)	22 (44.00%)	
Mutant	43 (25.44%)	52 (60.47%)	22 (53.66%)	117 (39.53%)	28 (56.00%)	
Sex						.007
Female	62 (36.69%)	31 (36.05%)	19 (46.34%)	112 (37.84%)	29 (58.00%)	
Male	107 (63.31%)	55 (63.95%)	22 (53.66%)	184 (62.16%)	21 (42.00%)	
Smoking status						.002
Never smoker	40 (23.67%)	33 (38.37%)	21 (51.22%)	94 (31.76%)	27 (54.00%)	
Smoker	129 (76.33%)	53 (61.63%)	20 (48.78%)	202 (68.24%)	23 (46.00%)	
TYPE						<.001
Luad	149 (88.17%)	70 (81.40%)	34 (82.93%)	253 (85.47%)	32 (64.00%)	
Lusc	17 (10.06%)	15 (17.44%)	7 (17.07%)	39 (13.18%)	18 (36.00%)	
Other	3 (1.78%)	1 (1.16%)	0 (.00%)	4 (1.35%)	0 (.00%)	

Note: Luad, Lung adenocarcinoma.

Lusc, lung squad cell carcinoma.

Other, 3 Large cell carcinoma and 1 pulmonary sarcomatoid carcinoma.

[a]The Cancer Imaging Archive.

[b]Cancer Hospital of Anhui University of Science and Technology.

[c]Huainan Chaoyang Hospital of Anhui University of Science and Technology.

[d]Eastern Hospital of Anhui University of Science and Technology.

are described as the mean ± standard deviation (SD), and qualitative data are described as frequencies (percentages). The "glmnet" package was used to implement the LASSO. CAL, DCA, C-index, and Brier scores were used to evaluate the performance of the machine learning classifier models. Differences between the AUC values of the models were compared using the Delong test. Statistical significance was set at $P < .05$.

Results

Clinical Data Analysis

The patients were divided into training and validation cohorts (Table 1). The training cohort consisted of 296 patients (184 men and 112 women; mean age: 66.82 ± 11.49 years; range: 24–89 years) from three centers. Of these, 117 (39.53%) had *EGFR* mutations, 179 (60.47%) had wild-type *EGFR*, and 253 (85.47%) had adenocarcinoma, 39 (13.18%) had squamous cell carcinoma, and 4 (1.35%) had other types of cancer (3 Large cell carcinoma;1 pulmonary sarcomatoid carcinoma). There were 202 (68.24%) smokers and 94 non-smokers (31.76%). The validation cohort included 50 patients (21 men and 29 women; mean age 66.56 ± 9.44 years; range, 43–85 years). There were 28 (56.00%) patients had *EGFR* mutations, 22 (44.00%) had wild-type *EGFR*, 32 (64.00%) had adenocarcinoma, and 18 (36.00%) had squamous cell carcinoma. There were 23 (46.00%) smokers and 27 non-smokers (54.00%).

There were no significant differences in age between the training and validation cohorts. However, there were

significant differences in *EGFR* mutation rates, sex, smoking status, and tumor type (Table 1).

Feature Extraction and Selection

A total of 1085 radiomics features were successfully extracted from each patient's ROI. First, 376 features with an ICC value < .75 were eliminated (Figure 2A). Second, 191 features were eliminated following hypothesis testing. Finally, the remaining 518 features were analyzed using 10-fold cross-validated LASSO regression and a standard error rule (Figures 2B and 2C). Sixteen core features were screened based on optimal $\lambda = .03202$ and standard error = .05841 (Table 2).

Radiomics Model Performance

According to the 16 screened radiomics features, the LR, DT, RF, and SVM classifiers were used to construct the model in the training cohort and validated in the validation cohort. The specific performances of the four classifier prediction models are shown in Figure 3 and Table 3.

In the training cohort, Figure 3A shows that the RF classifier performed the best (AUC=.995; 95% confidence interval [CI], .98–.996; sensitivity, 99.2%; specificity, 98.9%; accuracy, 99%). The remaining three classifiers were applied as follows (LR: AUC=.723, DT: AUC=.842, SVM: AUC=.883). The calibration curve (Figure 3B) shows excellent agreement between the predicted and actual values for
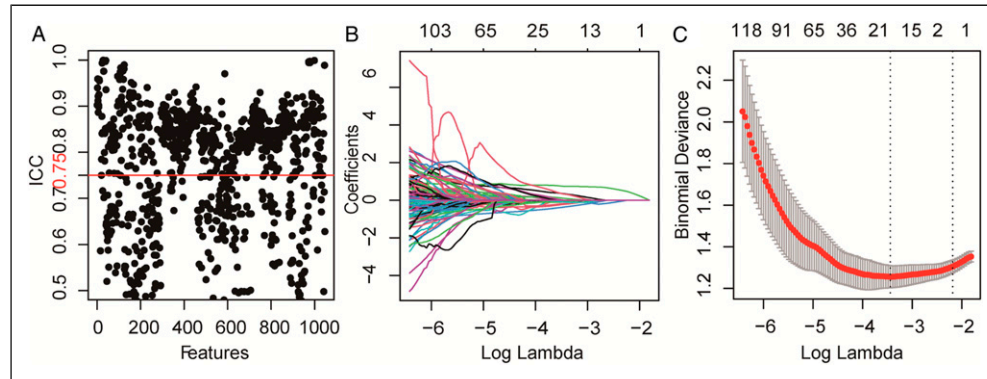
Figure 2. Selection of radiomics features. (A): ICC histogram of radiomics features; (B/C): LASSO method for screening of radiomics features.

Table 2. Texture Features Selection for Radiomics Models.

Parameters	Parameter category	Importance
Mean absolute deviation	Intensity histogram	−.065652535
60 Percentile area	Intensity histogram	−.027231004
Convex	Shape	.737731264
Correlation	Gray level cooccurence matrix 3	−.364074713
Dissimilarity	Gray level cooccurence matrix 3	.338776007
5-1 Homogeneity 2	Gray level cooccurence matrix 3	−.176750873
10-4 Homogeneity 2	Gray level cooccurence matrix 3	−.048826405
-333-7 Information measure corr 1	Gray level cooccurence matrix 3	−.053598886
8-1 Information measure corr 1	Gray level cooccurence matrix 3	−.217087549
9-7 Information measure corr 1	Gray level cooccurence matrix 3	.095533464
12-4 Inverse diff norm	Gray level cooccurence matrix 3	−.013040947
6-4 Inverse variance	Gray level cooccurence matrix 3	.322764285
8-4 Inverse variance	Gray level cooccurence matrix 3	−.092512731
8-1 Max Probability	Gray level cooccurence matrix 3	−.024734566
12-7 Max Probability	Gray level cooccurence matrix 3	−.102971831
−333 Run length nonuniformity	Gray level Run length matrix 25	−.002072575

the four machine learning classifiers. DCA (Figure 3C) indicated that the four machine learning classifiers provided more benefits than all treatments or no treatments.

In the validation cohort, Figure 3D shows that the RF classifier performed better (AUC=.88, 95% CI: .75-.946; sensitivity=96.5%; specificity=95.5%; accuracy=96%) than the other three classifiers (LR: AUC=.658, DT: AUC=.567, SVM: AUC=.765). The calibration curve (Figure 3E) shows a trend in which the predicted values for the RF classifier are closer to the 45°standard line, indicating that consistency of the RF model is more desirable. DCA (Figure 3F) also indicated that the RF classifier could achieve more clinical net benefits at almost all threshold probabilities.

In this study, the C-index of the RF model (training: RF=.998; validation: RF=.883) was higher than that of the other models (training: LR=.725, DT=.855, SVM=.905; validation: LR=.664, DT=.605, SVM=.773) in both the training and validation cohorts (Table 3).

In this study, the Brier score of the RF model (training: RF=.007; validation: RF=.137) was lower than that of the other models (training: LR=.203, DT=.153, SVM=.119; validation: LR=.235, DT=.244, SVM=.162) in both the training and validation cohorts (Table 3).

In this study, the Delong test showed that the AUC of the RF classifier in the training cohort was not significantly different from that of the RF classifier in the validation cohort ($P > .05$). However, there were significant differences ($P < .001$) between the LR, DT and SVM classifiers in the validation cohort (Table 4).

Discussion

This study aims to construct a predictive model with strong generalizability. We hope that this radiomics model can be used to determine the *EGFR* status of patients with NSCLC and provide a reference for guiding personalized targeted
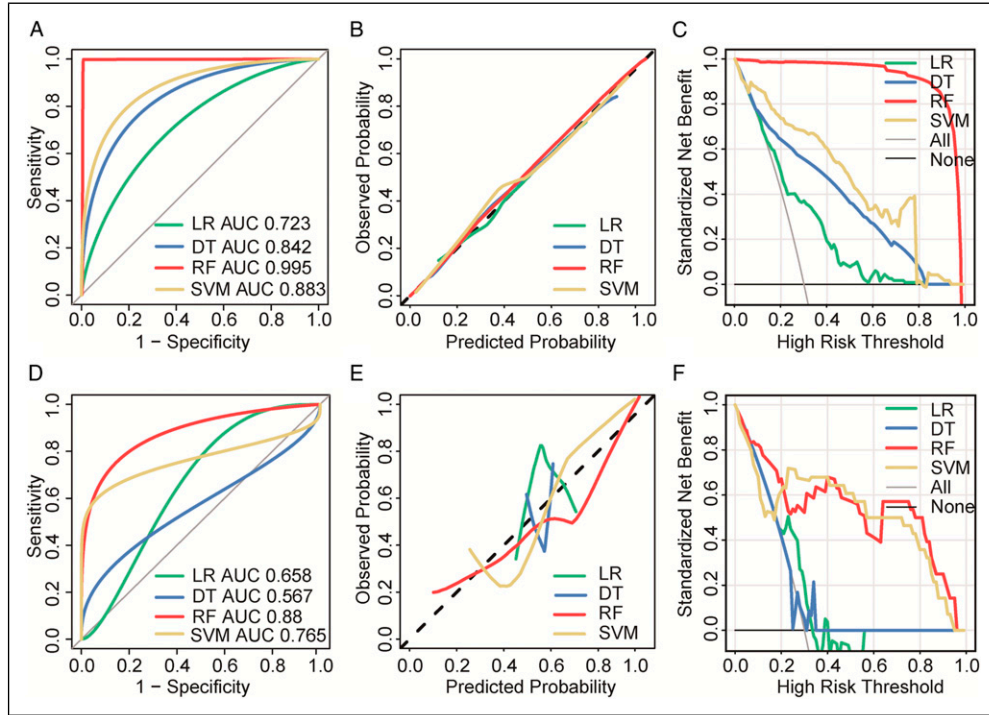
Figure 3. Building and performance of four machine learning classifier models. Receiver operating characteristic curves (3A), Calibration curves (3B), and Decision curves (3C) of different classifiers and models generated from the development cohorts; Receiver operating characteristic curves (3D), Calibration curves (3E), and Decision curves (3F) of different classifiers and models generated from the validation cohorts.

Table 3. Performance of the Radiomics Signature.

	AUC	Accuracy	Sensitivity	Specificity	C-index	Brier score
Training						
LR	.723	.69	.53	.794	.725	.203
DT	.842	.794	.701	.855	.837	.153
RF	.995	.99	.992	.989	.998	.007
SVM	.883	.845	.838	.85	.905	.119
Validation						
LR	.658	.68	.75	.591	.664	.235
DT	.567	.46	.358	.591	.605	.244
RF	.88	.96	.965	.955	.883	.137
SVM	.765	.84	.715	.99	.773	.162

therapies. Finally, we obtained 16 radiomic features with accurate prediction ability, including intensity histograms (n = 2), shape (n = 1), and GLCM (n = 13). These features encompass the description of intensity distribution, spatial relationships between different intensity levels, shape of texture patterns, and tumor heterogeneity. The intensity histogram is related to the gray level frequency distribution within the ROI, relies on single-voxel values rather than adjacent interacting voxels, and may be obtained from the voxel intensity histogram.[22] Morphological features are used to describe tumor characteristics by calculating the ROI, providing information on the size of the lesion tissue.[23] The correlation of some features with *EGFR* mutations has been confirmed in other studies related to the prediction of *EGFR* status using imaging histology.[24,25]

Diverse machine learning algorithms have their own advantages and disadvantages. Currently, the most common machine learning methods are LR, SVM, RF, and DT. In this study, the performance of the radiomics models was evaluated using the four different classifiers mentioned above, and the RF classifier with the highest diagnostic performance and good calibration and stability in the validation cohort was selected. In similar studies, Yang et al[26] applied an RF classifier to construct a model for predicting *EGFR* mutation status in patients with lung adenocarcinoma based on CT radiomics features; the AUC of the training cohort was .826,

Table 4. Delong Test of Machine Learning Classifier Model.

Training- AUC	Validation- AUC	*P*-value
RF	RF	>.05
RF	LR	<.001
RF	DT	<.001
RF	SVM	<.001

while that of the validation cohort was .779; however, this was only a single-center study. Velazquez et al[27] used CT radiomics features combined with clinical variables to predict *EGFR* mutations, with an AUC of .75 and lacked external data validation, the clinical applicability of which was limited.

Histological examination, the gold standard for *EGFR* detection, may provide additional support in clinical practice. However, if the puncture position is unavailable or the basic conditions are poor, multiple aspiration biopsies are required. Imaging examinations can provide a reference regarding *EGFR* gene status while understanding tumorigenesis and progression through imaging. Similarly, in patients with multiple tumors, radiography is beneficial for selecting the most suspicious tumor for biopsy. Thus, when histopathological examination is difficult, radiomics may play a useful role in clinical practice.

In this study, data from three centers were mixed to construct a training cohort, and core radiomics features that reflected EGFR status were screened. The model was verified in a validation cohort and the results were stable, which could reflect the generalizability of the model to a certain extent. However, the limited dataset cannot include all information reflecting *EGFR* status; therefore, the test results may not fully reflect the generalizability of the model. Future research will focus on verifying the generalizability of this model. In addition, there are the following limitations. (1) Radiomics analysis of histological features was mainly performed using a retrospective study design, which is still different from the actual predictive clinical need, leading to the need for further validation in prospective studies. (2) Different CT imaging protocols in different hospitals and radiomics features are influenced by CT scanner parameters (e.g., reconstruction kernel or slice thickness). Although resampling and pre-processing were performed to limit the differences between them, undiscovered differences may still exist. (3) For ROI outlining, manual and automatic outlinings offer unique advantages. The difference between the two approaches in terms of image alignment and contour generation may affect the calculation of radiomic features.

Conclusion

By comparing the four machine learning models, the RF model had a satisfactory performance for predicting the *EGFR* status of NSCLC. However, these results are preliminary and need to be validated using prospective datasets to assess their potential clinical applications.

ORCID iDs

Liu Yafeng https://orcid.org/0000-0003-3930-4872
Hu Dong https://orcid.org/0000-0002-8995-4729

Supplemental Material

Supplemental material for this article is available online.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA A Cancer J Clin*. 2020;70(1):7-30.
2. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA A Cancer J Clin*. 2019;69(5):363-385.
3. Bittner N, Ostoros GL. New treatment options for lung adenocarcinoma–in view of molecular background. *Pathol Oncol Res*. 2014;20(1):11-25.
4. Novello S. Epidermal growth factor receptor tyrosine kinase inhibitors as adjuvant therapy in completely resected Non-small-cell lung cancer. *J Clin Oncol*. 2015;33(34):3985-3986.
5. Goepel L, Jacobi A, Augustin M, et al. Rapid improvement of psoriasis in a patient with lung cancer after treatment with erlotinib. *J Eur Acad Dermatol Venereol*. 2018;32(8):e311-e313.
6. Bedard PL, Hansen AR, Ratain MJ, et al. Tumor heterogeneity in the clinic. *Nature*. 2013;501(7467):355-364.
7. Swanton C. Intratumor heterogeneity: evolution through space and time. *Cancer Res*. 2012;72(19):4875-4882.
8. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762.

9. Yuan M, Zhang YD, Pu XH, et al. Comparison of a radiomic biomarker with volumetric analysis for decoding tumor phenotypes of lung adenocarcinoma with different disease-specific survival. *Eur Radiol*. 2017;27(11):4857-4865.

10. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446.

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *radiology*. 2016;278(2):563-577.

12. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol*. 2016;2(12):1636-1642.

13. Mu W, Jiang L, Zhang J, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun*. 2020;11(1):5228.

14. Hong D, Xu K, Zhang L, et al. Radiomics signature as a predictive factor for EGFR mutations in advanced lung adenocarcinoma. *Front Oncol*. 2020;10:28.

15. Koyasu S, Nishio M, Isoda H, et al. Usefulness of gradient tree boosting for predicting histological subtype and EGFR mutation status of non-small cell lung cancer on 18F FDG-PET/CT. *Ann Nucl Med*. 2020;34(1):49-57.

16. Ninatti G, Kirienko M, Neri E, et al. Imaging-based prediction of molecular therapy targets in NSCLC by radiogenomics and AI approaches: a systematic review. *Diagnostics*. 2020;10(6):359.

17. Wang S, Shi J, Ye Z, et al. Predicting of EGFR mutation status in lung adenocarcinoma computed tomography image using deep learning. *Eur Respir J*. 2019;53(3):1800986.

18. Bakr S, Gevaert O, Echegaray S, et al. A radiogenomic dataset of non-small cell lung cancer. *Sci Data*. 2018;5(1):180-202.

19. Von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Prev Med*. 2007;147:573-577.

20. Zhang L, Chen B, Liu X, et al. Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell Lung Cancer. *Translat Oncol*. 2018;11(1):94-101.

21. Koo TK, Li MY. Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Med*. 2016;15(2):155–163.

22. Nyflot MJ, Yang F, Byrd D, et al. Quantitative radiomics: impact of stochastic effects on Textural feature analysis implies a need for standards. *J Med Imag*. 2015;2(4):041002.

23. Li M, Wang L, Deng S, et al. Color image segmentation using adaptive hierarchical-histogram thresholding. *PLoS One*. 2020;15(1):e0226345.

24. Shen TX, Liu L, Li WH, et al. CT imaging-based histogram features for prediction of EGFR mutation status of bone metastases in patients with primary lung adenocarcinoma. *Cancer Imag*. 2019;19(1):34.

25. Lu L, Sun SH, Yang H, et al. Radiomics prediction of EGFR status in lung cancer-our experience in using multiple feature extractors and the cancer imaging archive data. *Tomography*. 2020;6(2):223-230.

26. Yang X, Dong X, Wang J, et al. Computed tomography-based radiomics signature: a potential indicator of epidermal growth factor receptor mutation in pulmonary adenocarcinoma appearing as a subsolid nodule. *Oncologist*. 2019;24(11):e1156-e1164.

27. Rios Velazquez E, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in Lung Cancer. *Cancer Res*. 2017;77(14):3922-3930.