Original Article

Detection of *EGFR* mutations in early-stage lung adenocarcinoma by machine learning-based radiomics

Kenshiro Omura[1#], Yu Murakami[2,3#], Kohei Hashimoto[1], Hikaru Takahashi[2], Ryoko Suzuki[3], Yasuo Yoshioka[3], Masahiko Oguchi[2,3], Junji Ichinose[1], Yosuke Matsuura[1], Masayuki Nakao[1], Sakae Okumura[1], Mingyon Mun[1]

[1]Department of Thoracic Surgical Oncology, the Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan; [2]AI Medical Center, the Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan; [3]Radiation Oncology Department, the Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan

Contributions: (I) Conception and design: K Hashimoto, Y Murakami; (II) Administrative support: M Mun; (III) Provision of study materials or patients: K Omura, K Hashimoto, Y Murakami, H Takahashi; (IV) Collection and assembly of data: K Omura, K Hashimoto, Y Murakami, H Takahashi, R Suzuki; (V) Data analysis and interpretation: K Omura, K Hashimoto, Y Murakami, H Takahashi, R Suzuki; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Kohei Hashimoto, MD, PhD. Department of Thoracic Surgical Oncology, the Cancer Institute Hospital, Japanese Foundation for Cancer Research, 3-8-31 Ariake, Koto, Tokyo 135-8550, Japan. Email: kohei.hashimoto@jfcr.or.jp.

Background: We hypothesized that epidermal growth factor receptor (*EGFR*) mutations could be detected in early-stage lung adenocarcinoma using radiomics.

Methods: This retrospective study included consecutive patients with clinical stage I/II lung adenocarcinoma who underwent curative-intent pulmonary resection from March–December 2016. Using preoperative enhanced chest computed tomography, 3,951 radiomic features were extracted in total from the tumor (area within the tumor boundary), tumor rim (area within ±3 mm of the tumor boundary), and tumor exterior (area between +10 mm outside the tumor and tumor boundary). A machine learning-based radiomics model was constructed to detect *EGFR* mutations. The combined model incorporated both radiomic and clinical features (gender and smoking history). The performance was validated with five-fold cross-validation and evaluated using the mean area under the curve (AUC).

Results: Of 99 patients (mean age, 66±11 years; female, 66.6%; clinical stage I/II, 89.9%/10.1%), *EGFR* mutations in the surgical specimen were detected in 46 (46.5%). A median of 4 (range, 2 to 8) radiomic features was selected for each validation session. The mean AUCs in the radiomics and combined models were 0.75 and 0.83, respectively. The two top-ranked features in the combined model were the radiomic features extracted from the tumor exterior and the tumor, indicating a higher impact of radiomic features over relevant clinical features.

Conclusions: Radiomic features, including those in the peri-tumoral area, may help detect *EGFR* mutations in lung adenocarcinomas in preoperative settings. This non-invasive image-based technology could help guide future precision neoadjuvant therapy.

Keywords: Neoadjuvant therapy; machine-learning; radiomics

Submitted Nov 22, 2022. Accepted for publication Mar 10, 2023. Published online Apr 11, 2023.
doi: 10.21037/tcr-22-2683
View this article at: https://dx.doi.org/10.21037/tcr-22-2683

Introduction

Molecular targeting therapy has revolutionized the landscape of the treatment of advanced-stage lung adenocarcinoma (1-5). Genetic mutations in the epidermal growth factor receptor (*EGFR*) domain are now established biomarkers to predict the treatment response using the *EGFR* tyrosine kinase inhibitor (TKI). The incidence of these mutations is approximately 50% in the Asian population and fifteen% in the Caucasian population (6). In the adjuvant settings, osimertinib, a third-generation *EGFR*-TKI, has recently been shown to prolong disease-free survival in patients with completely resected non-squamous non-small cell lung cancer (NSCLC) with *EGFR* mutations (7). Currently, the efficacy of neoadjuvant osimertinib is being tested to further improve the prognosis for resectable primary non-squamous NSCLC harboring *EGFR* mutations (8). *EGFR* mutation is also a key exclusion criterion in a clinical trial with neoadjuvant immunotherapy (Impower 30, NCT03456063). For inclusion in these clinical trials, invasive tissue biopsy is mandatory at this moment. Although computed tomography (CT)-guided biopsy can be effective in the case of peripheral lung tumors, tissue biopsy can be challenging in patients with early-stage operable lung cancers in clinical practice, where the sensitivity of the transbronchial tissue biopsy for peripheral tumors is around 57% (9) while that of the endoscopic ultrasound-guided needle aspiration of the mediastinum is around 89% (10).

Validated clinical predictors of *EGFR* mutations are reported (gender, smoking, ethnicity). A prediction model based on these three factors performed reasonably with area under the curve (AUC) of 0.75 (11), which could be further improved. Radiomics is defined as the quantitative analysis of multiple high-throughput data derived from medical imaging examinations. Previously, radiomic features, in combination with clinical or pathological data, can be used to develop models that predict tumor characteristics in the field of lung cancers (12).

We hypothesized that *EGFR* mutations could be detected by radiomics of primary early-stage resectable lung adenocarcinoma using preoperative CT. We present the following article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2683/rc).

Highlight box

Key findings

- A machine learning-based model using radiomic features, extracted from the early-stage lung adenocarcinoma preoperative image, along with clinical features, predicted EGFR mutations with AUC of 0.83.

What is known and what is new?

- A machine learning method can help create prediction models from a large amount of extracted data and clinical features. This manuscript adds what kind of content is important for machine learning to create prediction models.

What is the implication, and what should change now?

- This non-invasive image-based technology in detecting actionable mutations may guide future precision neoadjuvant therapy.

Methods

Patient selection

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Institutional Review Board of the Cancer Institute Hospital (No. 2021-GA-1064) on June 28, 2021, and the patient consent requirement was waived because of its retrospective nature. This retrospective study included consecutive patients with clinical stage I/II lung adenocarcinoma who underwent curative-intent pulmonary resection from March 2016 to December 2016. Patients with pure ground-glass opacity or lobar atelectasis or those who received preoperative treatment were excluded. Patients who underwent only non-contrast CT or CT taken more than 2 months before surgery were also excluded. *EGFR* gene mutation status was analyzed from paraffin-embedded surgical specimens using a polymerase chain reaction (PCR) kit (cobas® *EGFR* Mutation Test v2, Roche, Basel, Switzerland) targeting *exons 18*, *19*, *20*, and *21*. The overall workflow of this study is shown in *Figure 1*.

The segmentation

All patients underwent contrast-enhanced chest CT using Discovery CT750 HD (GE Healthcare, Chicago, IL, USA). The acquisition parameters were as follows: tube voltage, 120 kVp; tube current, 100–649 mA; exposure time, 400–699 ms; milliampere-seconds, 50–329.9 mAs, slice thickness, 1.25 mm (Table S1). In axial CT images under lung field view, tumor boundary was segmented into slices containing primary tumor using Eclipse software ver. 8.6 or ver. 10.0 (Varian Medical Systems, Palo Alto, CA, USA). We then

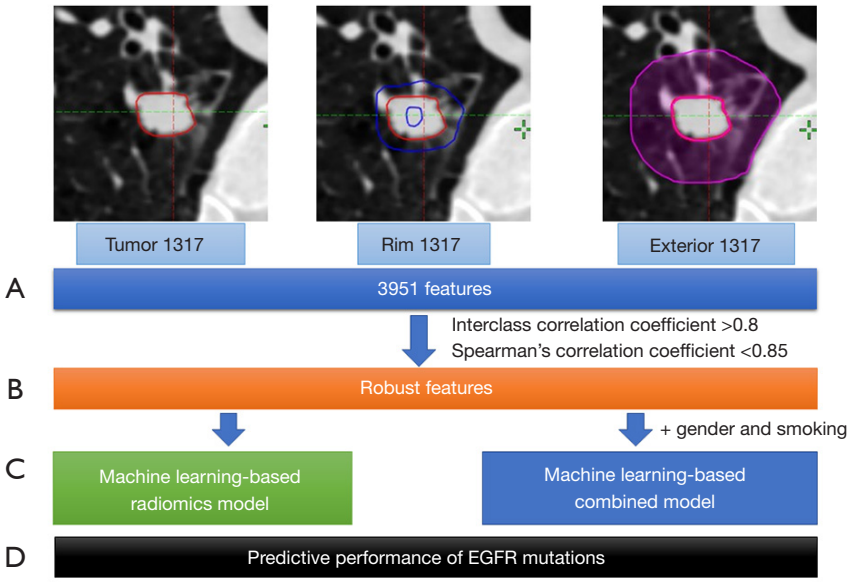

Figure 1 Study design. A: radiomic features were extracted from the preoperative enhanced CT scan. B: an interclass correlation coefficient >0.8 between two physicians was considered reproducible features. Features showing an Spearman's correlation coefficient ≥0.85 were excluded. C: machine learning-based models were created for prediction of EGFR mutations. D: AUC was calculated through five-fold cross-validation to evaluate predictive performance. EGFR, epidermal growth factor receptor; CT, computed tomography; AUC, area under the curve.

created volume of interests (VOIs) as follows (*Figure 1*): the tumor (VOI within the tumor boundary), tumor rim (VOI within ±3 mm of the tumor boundary), and tumor exterior (VOI between +10 mm outside the tumor and the tumor boundary). The demarcation of the tumor rim and tumor exterior areas was calculated using the automated function of the software. If the created VOI went outside the lungs (i.e., the VOI partially included mediastinum or chest wall), the contour was manually modified to fit within the lung field.

The radiomics feature extraction

Prior to feature extraction, all CT images were resampled to an isotropic grid of 1×1×1 mm^3 using B-splines to consistently calculate the three-dimensional features. The Laplacian of Gaussian (LoG) filter (σ =0.5, 1.0, 1.5, 2.0, and 2.5 mm) and wavelet filter that computes eight decompositions (HHH, HHL, HLH, HLL, LHH, LHL, LLH, LLL) per level were used for transforming CT images. A total of 3,951 radiomic features, including 108 original features (first-order statistics, n=18; shape, n=15; texture, n=75) and 1,209 filtered features (LoG, n=465; wavelet, n=744), were extracted from each VOI (n=1,317×3) using PyRadiomics version 3.0.1 (13). The texture features were categorized as gray-level co-occurrence matrix (n=24), gray-level run length matrix (n=16), gray-level size zone matrix (n=16), gray-level dependence matrix (n=14), and neighboring gray-tone difference matrix (n=5). The bin width was set to 25 Hounsfield units (14).

The feature selection and model building

To assess the inter-observer reproducibility for VOI segmentations, segmentation of the tumor was performed using 30 randomly selected patients by two physicians with experience in lung cancer diagnostics (KH and RS) (15,16). Then, the interclass correlation coefficient (ICC) of >0.8 calculated between two physicians was considered a reproducible feature. To eliminate redundant features, Spearman's correlation coefficient (SCC) among the features of all possible two combinations was calculated, and features showing an SCC of ≥0.85 were excluded.

Prior to model building, each feature was standardized using Z-scores. The machine learning-based radiomics model [using the Boruta algorithm (17-19) (BorutaPy

version 0.3) for feature selection and Random Forest (Scikit-learn version 1.0.2) for prediction] was constructed to detect *EGFR* mutation. Suitable hyperparameters for Random Forest were determined by Optuna version 2.10.0 (20), while those for Boruta were empirically determined. The tuned hyperparameters by Optuna were max_depth, max_leaf_nodes, min_samples_leaf, and n_estimators. The meanings of these parameters are described in Kursa *et al.* (17). The number of maximum iterations to select the optimal features for Boruta was set to 500. The combined model incorporated both radiomic as well as clinical features (gender and smoking history). Gender and smoking history were selected as these are clinically validated predictor of *EGFR* mutation (21). As this study was performed in monotonous population (Asians), ethnicity was not included. Other clinical variables in this study were not included to avoid overfitting. The current study is based on TRIPOD Type 1b (22).

Assessment of performance of the models and statistical analyses

The performance was validated with non-nested five-fold cross-validation, and the mean area under the receiver operating characteristic (ROC) curve was used for evaluating the prediction models. Accuracy, recall, precision, specificity, and F1 score were also calculated. Comparison of the clinical characteristics was performed for continuous and categorical variables using *t*-test or Mann-Whitney test, and Fisher's exact test, respectively. A comparison of AUCs was performed using Delong's test. Statistical analyses were performed using R software and GraphPad Prism version 8 (GraphPad Software Inc., La Jolla, CA, USA). All P values were two-sided, and P<0.05 was considered statistically significant.

Results

Clinical characteristics of patients

The patient characteristics are summarized in *Table 1*. This study enrolled 99 patients (mean age, 66±11 years; female, 66.6%; clinical stage I/II 89.9%/10.1%). There was no lung cancer that invaded mediastinum or chest wall. *EGFR* mutations in the surgical specimen were detected in 46 (46.5%) patients. Compared to wild type, patients with EGFR mutation included significantly more female (female: *EGFR* wild group, 54.7% *vs.* *EGFR* mutation group, 78.3%; P=0.02) and less smoking history (current or previous) (smoking history: *EGFR* wild group, 60.4% *vs.* *EGFR* mutation group, 19.6%; P=0.01) as expected. Further, a significant difference in the histologic subtype (P=0.005) between the groups existed. In the *EGFR* wild group, the most common histologic subtype was the papillary type (39.6%), followed by the solid subtype (30.2%). In the *EGFR* mutant group, the most common histologic subtype was the papillary type (71.7%), followed by the adenocarcinoma in situ (AIS)/minimally invasive adenocarcinoma (MIA) pattern (13.0%). There were no significant differences in mode of surgery and nodal status between the groups.

Among the *EGFR* mutant patients, 26 (54.2%) presented *exon 21 L858R*, 19 (39.5%) had *exon 19 deletions*, and 1 (2.1%) exhibited *exon 21 L861Q* as an uncommon mutation. Of 26 patients with *exon 21 L858R* mutation, two (4.2%) had *de novo exon 20 T790M* as well.

Feature selection and diagnostic performance

After assessing the ICC, 1,317 features extracted from VOI of the tumor were reduced to 1,267 features (in total, 3,955 features were reduced to 3,901 features). Further, 3,901 features were reduced to 453 robust radiomic features using SCC analysis (Table S2). Of 453 features that cleared the ICC and SCC, a median of 4 (range, 2 to 8) radiomic features from each validation session were selected using the Boruta algorithm.

The predictive performance of radiomics and combined models are summarized in Table S3. In the training sets, the mean AUC of the radiomics and combined models were 0.78 (range, 0.64–0.94) and 0.74 (range, 0.70–0.80), respectively. In the validation set, the combined model showed acceptable predictive performance with a mean AUC of 0.83 (range, 0.74–0.95), which tended to be higher than that of the radiomics model (mean AUC, 0.75; range, 0.65–0.82). However, Delong's test revealed that the difference in AUC did not reach statistical significance (P=0.052). The ROC of each validation set in the radiomics and combined model are presented in *Figure 2*.

Importance of features selected using machine learning models

A total of 15 radiomic features from the radiomics model and 17 features, including two clinical features, from the combined model were selected through a five-fold cross-

Table 1 Clinicopathological characteristics of patients in *EGFR* wild group and *EGFR* mutation group

Variables	Total (n=99)	*EGFR* wild type (n=53)	*EGFR* mutation (n=46)	P value
Age (years)	66±11	67±10	63±11	0.06
Gender: female (%)	65 (66.6)	29 (54.7)	36 (78.3)	0.02
Smoking: yes (%)	41(41.4)	32 (60.4)	9 (19.6)	0.01
Clinical stage (%)				0.76
Stage IA	80 (80.8)	42 (79.2)	38 (82.6)	
Stage IB	9 (9.1)	4 (7.5)	5 (10.9)	
Stage IIA	4 (4.0)	3 (5.7)	1 (2.2)	
Stage IIB	6 (6.1)	4 (7.5)	2 (4.3)	
Mode of surgery (%)				0.36
Sublobar resection	14 (14.1)	9 (17.0)	5 (10.9)	
Lobectomy	83 (83.8)	42 (79.2)	41 (89.1)	
Pneumonectomy	2 (2.0)	2 (3.8)	0 (0.0)	
Pathological diagnosis (%)				0.09
pN0	83 (83.8)	43 (81.1)	40 (87.0)	
pN1	8 (8.1)	7 (13.2)	1 (2.2)	
pN2	8 (8.1)	3 (5.7)	5 (10.9)	
Histology (%)				0.005
AIS/MIA	12 (12.1)	6 (11.3)	6 (13.0)	
Papillary	54 (54.5)	21 (39.6)	33 (71.7)	
Lepidic	14 (14.1)	10 (18.9)	4 (8.7)	
Solid	19 (19.2)	16 (30.2)	3 (6.5)	

Values are presented as n (%) or mean ± SD. *EGFR*, epidermal growth factor receptor; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; SD, standard deviation.

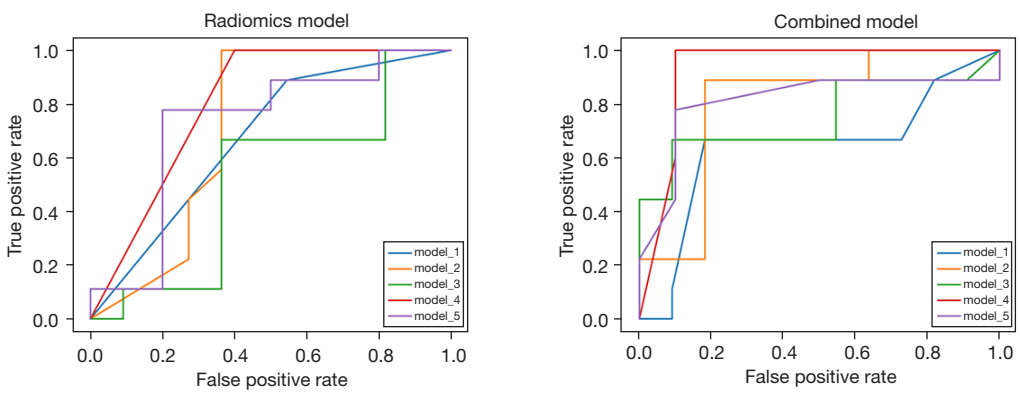

Figure 2 Receiver operating curves of each validation session in each model. The mean AUCs of five cross-validation sessions in the radiomics and combined models were 0.75 (range, 0.65–0.82) and 0.83 (range, 0.74–0.95), respectively. AUC, area under the curve.

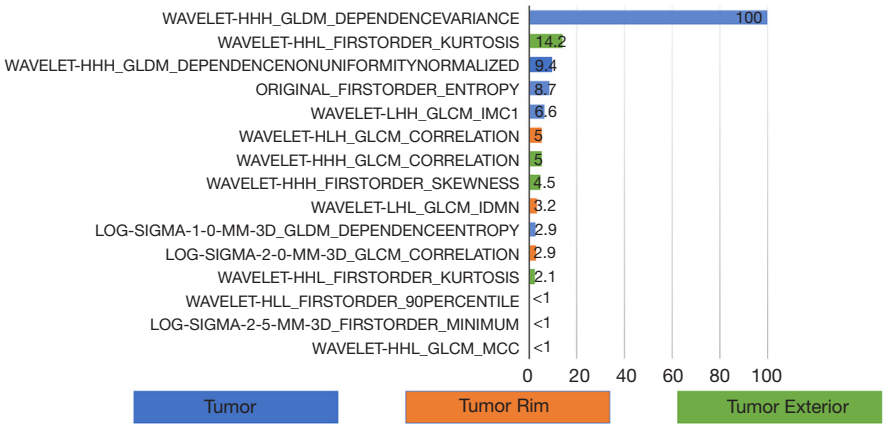

Figure 3 Ranking of features selected using Boruta algorithm in the radiomics model (each number represents a percentage of coefficient of the selected feature with reference to the top-ranked feature). HHH, HLL, HHL, LHH, LHL, HLH: wavelet filter.

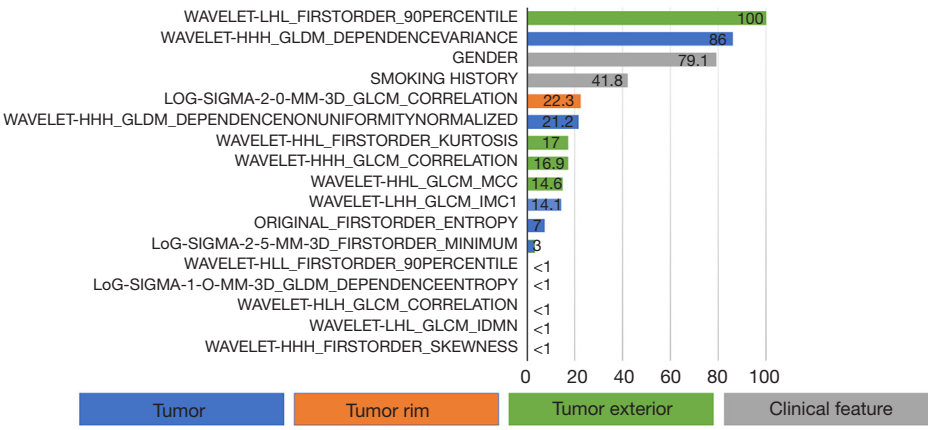

Figure 4 Ranking of features selected using Boruta algorithm in the combined model (each number represents a percentage of the coefficient of the selected feature with reference to the top-ranked feature). HHH, HHL, LHH, LHL, HLH: wavelet filter.

validation session. The selected radiomic and clinical features and the coefficients in each cross-validation session are listed in Table S4.

Figure 3 represents the ranking of each selected radiomic feature based on the average coefficient through five sessions in the radiomics model. *Figure 4* shows the rankings of radiomic and clinical features in the combined model. Each feature was converted to a percentage value relative to the average coefficient of the top-ranked feature in each model. It was noted that the average coefficient of two of the radiomic features (wavelet-LHL_firstorder_90Percentile from the tumor exterior and wavelet-HHH_gldm_DependenceVariance from the tumor) were superior to those of the relevant clinical features (gender and smoking history) in the combined model, suggesting that the radiomic features might have a higher impact on detecting *EGFR* mutations over the clinical features.

Significance of top-ranked radiomic features

Representative images showing the top three and bottom three cases in the two features that were ranked above clinical features in the combined model are demonstrated in *Figure 5*. The first order describes the distribution of voxel intensities within the image region. A voxel intensity of 90 percentile in the tumor exterior was selected in this feature.

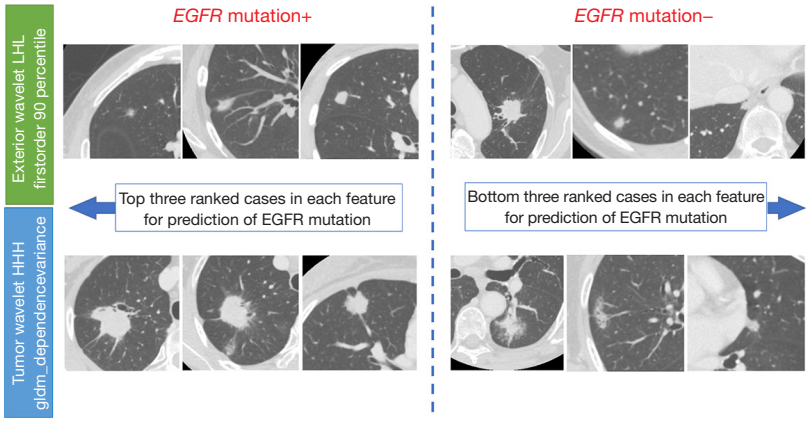

Figure 5 Left side: representative images of top three ranked cases in two radiomic features which was ranked above relevant clinical features in the combined model. Allow indicates cases with higher-ranked feature within *EGFR* mutation+. Right side: cases with bottom three ranked cases in each radiomic feature. Allow indicates cases with lower-ranked feature within *EGFR* mutation–. LHL, HHH: wavelet filter. *EGFR*, epidermal growth factor receptor.

Dependence variance measures the variance in dependence size in the image (within the tumor in this feature). No apparent recognizable pattern was observed, particularly in these images, indicating that radiomics could detect subtle changes unrecognizable to human eyes.

Discussion

In this study, we demonstrated the feasibility of predicting *EGFR* mutations in early-stage operable adenocarcinomas using radiomic features with a mean AUC of 0.75 in the validation settings. Although this was not statistically different, mean AUC improved up to 0.83 by combining with established clinical predictive factors of *EGFR* mutations (gender and smoking history). To quantify the contribution of these clinical variables more clearly, we further created the Clinical model and the all-combined model which includes all clinical variables (age, gender, smoking history, and clinical stage) and radiomic features (Table S5). Consequently, the mean AUC of the clinical model was identical to that of the radiomics model (0.77 *vs.* 0.75). However, the predictive performance in the all-combined model was limited (i.e., the mean AUC of 0.82), implying the selection of gender and smoking history for the Combined model is reasonable. We utilized a machine learning model to select important features and generate a predictive model out of a large number of features extracted from radiological images. Classically, the existence of air bronchogram or ground glass opacity is a known semantic CT feature associated with EGFR mutations (23). However, subtle changes in the portion of ground glass opacity may not be distinguishable to the human eye. Radiomics can translate these tiny changes into the probability of mutation. Other unknown semantic or unrecognizable features, as suggested in *Figure 3*, were also integrated into the computation.

Previous studies have shown the possibility of *EGFR* mutation status in NSCLC utilizing CT images (24-29). However, only few of them comprised the surgical population. In this study, we focused on the preoperative CT images from early-stage NSCLC, aiming for potential usage in neoadjuvant settings. Given the ongoing research and the rapid evolution in clinical practice of advanced-stage NSCLC, as well as the adjuvant setting driven by *EGFR*-TKI (7,30) and immunotherapy (31), neoadjuvant precision medicine can plausibly become a reality with preoperative treatment guided by actionable biomarkers. *EGFR* mutation should play a pivotal role in this setting, considering its prevalence, the existence of a series of effective *EGFR*-TKIs (1-5), and its association with a low response rate to immunotherapy (32). However, biopsies for relatively small peripheral targets in early-stage operable patients can be challenging. Occasionally, the quantity or quality of samples is not enough for DNA testing. Moreover, even in the liquid biopsy (ctDNA), there is the possibility of false-negative result in NSCLC (33), thereby considering the tissue-based analysis for *EGFR* mutations as the reference standard (34). Our radiomics-based non-

invasive prediction of *EGFR* mutations could help guide the choice of preoperative treatment in this setting as additional information to the existing methods. Potentially, integration with other non-invasive predictive measures such as liquid biopsy (35) can decrease the need for biopsies in the future.

Several potential benefits of radiomic biomarkers exist in the neoadjuvant setting. First, this could avoid biopsy-related complications. Although transbronchial and transthoracic biopsies are considered established procedures, rare but serious complications hampering curative operations can still occur (36,37). Second, our technology could save time by skipping biopsy and waiting time for results. Waiting time before surgery can impact long-term outcomes (38); hence, saving this time could benefit patients from not just avoiding complications but also potentially improving the oncological outcomes.

Our findings suggested the importance of the peritumoral region to detect *EGFR* mutations. There are several explanations. First, the peritumoral region may have detected the skewed field of the lung parenchyma or broncho-vasculature structures, possibly due to the related mutation status of the tumor. Also, cancerous islets, known to be found adjacent to tumors (39), may be subtle changes to the human eye field but could have been detected as a radiomic feature. Second, manual demarcation may have missed some tumor edge which could have been picked as the peritumoral region. The top-ranked radiomic feature in the combined model and the second-ranked feature in the radiomics model were from the tumor exterior region, suggesting the importance of the signal outside the visible tumor margin for predicting the mutation profile.

There are several limitations to this study. First, this is a single institutional retrospective study with relatively small sample size. Although we performed cross-validation according to TRIPOD guidelines (22), our findings need further validation with a large sample set. Moreover, because it is well-known that an *EGFR* mutation profile can differ across ethnicities (40), our model may perform differently in other regions or ethnicities. Therefore, the models should be further validated in a multi-national and multi-ethnicity study. Second, although we focused only on adenocarcinomas considering the clinically proven utility of *EGFR*-TKIs against the disease and relative rarity and uncertainty of the predictive value of the treatment efficacy of squamous cell lung carcinoma *EGFR* mutation (in fact recommendation of the *EGFR* mutation testing for squamous cell carcinoma is somewhat mixed among society guidelines), we are aware that adenocarcinoma can be confirmed only after invasive biopsy at this moment. Nevertheless, adenocarcinoma is the most relevant type of NSCLC, and this prediction model perhaps could be combined with the radiomics prediction model of the histological type of NSCLC. Besides, sometimes biopsy of a tumor only provides histology diagnosis but not enough sample for molecular testing result in daily practice. Our image biomarker may be useful in that situation. Third, we utilized manual demarcation of the tumor boundary to extract radiomic features that can be a source of errors. Although we performed ICC to diminish inter-observer variability, this could have affected the results as radiomics is sensitive to slight changes. Incorporating automated detection technology could reduce the inter-observer variation. Fourth, we chose non-nested cross validation for evaluating the prediction performance, thus optimal hyperparameters for Random Forest were not tuned using independent validation sets. Finally, we did not consider other machine-learning techniques other than Random Forest for prediction. Jia *et al.* used Random Forest classifier to predict *EGFR* mutations using radiomic features and clinical variables (29), and demonstrated that relatively higher predictive performance was observed (AUC =0.828), which is consistent with our results (AUC =0.83).

Conclusions

Radiomic features may help detect *EGFR* mutations in lung adenocarcinomas in preoperative settings. Important signals even outside the recognizable tumor boundary may exist. This non-invasive image-based technology in detecting actionable mutations could help guide future precision neoadjuvant therapy.

Acknowledgments

This paper was presented as a poster in the 102nd annual meeting of the American Association for Thoracic Surgery, Boston, MA, May 14–17, 2022.
Funding: This study was funded by the Japanese Foundation for Multidisciplinary Treatment of Cancer and the divisional funding.

Footnote

Reporting Checklist: The authors have completed the

TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2683/rc

Data Sharing Statement: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2683/dss

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2683/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2683/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Institutional Review Board of the Cancer Institute Hospital (No. 2021-GA-1064) on June 28, 2021, and the patient consent requirement was waived because of its retrospective nature.

References

1. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947-57.
2. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011;12:735-42.
3. Ramalingam SS, Vansteenkiste J, Planchard D, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. N Engl J Med 2020;382:41-50.
4. Sequist LV, Yang JC, Yamamoto N, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 2013;31:3327-34.
5. Wu YL, Cheng Y, Zhou X, et al. Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. Lancet Oncol 2017;18:1454-66.
6. Han B, Tjulandin S, Hagiwara K, et al. EGFR mutation prevalence in Asia-Pacific and Russian patients with advanced NSCLC of adenocarcinoma and non-adenocarcinoma histology: The IGNITE study. Lung Cancer 2017;113:37-44.
7. Wu YL, Tsuboi M, He J, et al. Osimertinib in Resected EGFR-Mutated Non-Small-Cell Lung Cancer. N Engl J Med 2020;383:1711-23.
8. Tsuboi M, Weder W, Escriu C, et al. Neoadjuvant osimertinib with/without chemotherapy versus chemotherapy alone for EGFR-mutated resectable non-small-cell lung cancer: NeoADAURA. Future Oncol 2021;17:4045-55.
9. Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e142S-65S.
10. Silvestri GA, Gonzalez AV, Jantz MA, et al. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e211S-50S.
11. Aye PS, Tin Tin S, McKeage MJ, et al. Development and validation of a predictive model for estimating EGFR mutation probabilities in patients with non-squamous non-small cell lung cancer in New Zealand. BMC Cancer 2020;20:658.
12. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. Eur J Radiol 2017;86:297-307.
13. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:e104-7.
14. Kakino R, Nakamura M, Mitsuyoshi T, et al. Application and limitation of radiomics approach to prognostic

prediction for lung stereotactic body radiotherapy using breath-hold CT images with random survival forest: A multi-institutional study. Med Phys 2020;47:4634-43.
15. Nakanishi R, Akiyoshi T, Toda S, et al. Radiomics Approach Outperforms Diameter Criteria for Predicting Pathological Lateral Lymph Node Metastasis After Neoadjuvant (Chemo)Radiotherapy in Advanced Low Rectal Cancer. Ann Surg Oncol 2020;27:4273-83.
16. Murakami Y, Soyano T, Kozuka T, et al. Dose-Based Radiomic Analysis (Dosiomics) for Intensity Modulated Radiation Therapy in Patients With Prostate Cancer: Correlation Between Planned Dose Distribution and Biochemical Failure. Int J Radiat Oncol Biol Phys 2022;112:247-59.
17. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw 2010;36:1-13.
18. Otani S, Himoto Y, Nishio M, et al. Radiomic machine learning for pretreatment assessment of prognostic risk factors for endometrial cancer and its effects on radiologists' decisions of deep myometrial invasion. Magn Reson Imaging 2022;85:161-7.
19. Kursa MB. Robustness of Random Forest-based gene selection methods. BMC Bioinformatics 2014;15:8.
20. Akiba T, Sano S, Yanase T, et al. A Next-generation Hyperparameter Optimization Framework. arXiv:1907.10902v1. 2019.
21. Zhang YL, Yuan JQ, Wang KF, et al. The prevalence of EGFR mutation in patients with non-small cell lung cancer: a systematic review and meta-analysis. Oncotarget 2016;7:78985-93.
22. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.
23. Cheng Z, Shan F, Yang Y, et al. CT characteristics of non-small cell lung cancer with epidermal growth factor receptor mutation: a systematic review and meta-analysis. BMC Med Imaging 2017;17:5.
24. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology 2016;281:947-57.
25. Rossi G, Barabino E, Fedeli A, et al. Radiomic Detection of EGFR Mutations in NSCLC. Cancer Res 2021;81:724-31.
26. Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur Respir J 2019;53:1800986.
27. Liu Y, Kim J, Balagurunathan Y, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. Clin Lung Cancer 2016;17:441-448.e6.
28. 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.
29. Jia TY, Xiong JF, Li XY, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. Eur Radiol 2019;29:4742-50.
30. Zhong WZ, Wang Q, Mao WM, et al. Gefitinib Versus Vinorelbine Plus Cisplatin as Adjuvant Treatment for Stage II-IIIA (N1-N2) EGFR-Mutant NSCLC: Final Overall Survival Analysis of CTONG1104 Phase III Trial. J Clin Oncol 2021;39:713-22.
31. Felip E, Altorki N, Zhou C, et al. Adjuvant atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial. Lancet 2021;398:1344-57.
32. Gainor JF, Shaw AT, Sequist LV, et al. EGFR Mutations and ALK Rearrangements Are Associated with Low Response Rates to PD-1 Pathway Blockade in Non-Small Cell Lung Cancer: A Retrospective Analysis. Clin Cancer Res 2016;22:4585-93.
33. Rolfo C, Mack PC, Scagliotti GV, et al. Liquid Biopsy for Advanced Non-Small Cell Lung Cancer (NSCLC): A Statement Paper from the IASLC. J Thorac Oncol 2018;13:1248-68.
34. Malapelle U, Sirera R, Jantus-Lewintre E, et al. Profile of the Roche cobas® EGFR mutation test v2 for non-small cell lung cancer. Expert Rev Mol Diagn 2017;17:209-15.
35. Merker JD, Oxnard GR, Compton C, et al. Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. J Clin Oncol 2018;36:1631-41.
36. Asano F, Aoe M, Ohsaki Y, et al. Deaths and complications associated with respiratory endoscopy: a survey by the Japan Society for Respiratory Endoscopy in 2010. Respirology 2012;17:478-85.
37. Tomiyama N, Yasuhara Y, Nakajima Y, et al. CT-guided needle biopsy of lung lesions: a survey of severe complication based on 9783 biopsies in Japan. Eur J Radiol 2006;59:60-4.
38. Ponholzer F, Kroepfl V, Ng C, et al. Delay to surgical treatment in lung cancer patients and its impact on survival in a video-assisted thoracoscopic lobectomy cohort. Sci Rep 2021;11:4914.

39. Koukourakis MI, Giatromanolaki A, Thorpe PE, et al. Vascular endothelial growth factor/KDR activated microvessel density versus CD31 standard microvessel density in non-small cell lung cancer. Cancer Res 2000;60:3088-95.

40. Shi Y, Au JS, Thongprasert S, 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;9:154-62.

Cite this article as: Omura K, Murakami Y, Hashimoto K, Takahashi H, Suzuki R, Yoshioka Y, Oguchi M, Ichinose J, Matsuura Y, Nakao M, Okumura S, Mun M. Detection of *EGFR* mutations in early-stage lung adenocarcinoma by machine learning-based radiomics. Transl Cancer Res 2023;12(4):837-847. doi: 10.21037/tcr-22-2683

Supplementary

Table S1 Acquisition parameters

Patient	Manufacturer	Model name	Slice thickness (mm)	kVp (KV)	Tube current (mA)	Exposure time (ms)	Milliampere-seconds (mAs)
1	GE Medical Systems	Discovery CT750 HD	1.25	120	195	500	97.5
2	GE Medical Systems	Discovery CT750 HD	1.25	120	317	699	221.6
3	GE Medical Systems	Discovery CT750 HD	1.25	120	406	400	162.4
4	GE Medical Systems	Discovery CT750 HD	1.25	120	288	500	144.0
5	GE Medical Systems	Discovery CT750 HD	1.25	120	115	600	69.0
6	GE Medical Systems	Discovery CT750 HD	1.25	120	149	699	104.2
7	GE Medical Systems	Discovery CT750 HD	1.25	120	175	600	105.0
8	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
9	GE Medical Systems	Discovery CT750 HD	1.25	120	274	699	191.5
10	GE Medical Systems	Discovery CT750 HD	1.25	120	153	500	76.5
11	GE Medical Systems	Discovery CT750 HD	1.25	120	356	500	178.0
12	GE Medical Systems	Discovery CT750 HD	1.25	120	152	600	91.2
13	GE Medical Systems	Discovery CT750 HD	1.25	120	219	600	131.4
14	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
15	GE Medical Systems	Discovery CT750 HD	1.25	120	133	500	66.5
16	GE Medical Systems	Discovery CT750 HD	1.25	120	484	500	242.0
17	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
18	GE Medical Systems	Discovery CT750 HD	1.25	120	129	500	64.5
20	GE Medical Systems	Discovery CT750 HD	1.25	120	296	500	148.0
21	GE Medical Systems	Discovery CT750 HD	1.25	120	400	600	240.0
22	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
23	GE Medical Systems	Discovery CT750 HD	1.25	120	180	600	108.0
24	GE Medical Systems	Discovery CT750 HD	1.25	120	162	500	81.0
25	GE Medical Systems	Discovery CT750 HD	1.25	120	345	500	172.5
26	GE Medical Systems	Discovery CT750 HD	1.25	120	183	500	91.5
27	GE Medical Systems	Discovery CT750 HD	1.25	120	119	500	59.5
28	GE Medical Systems	Discovery CT750 HD	1.25	120	467	699	326.4
29	GE Medical Systems	Discovery CT750 HD	1.25	120	378	600	226.8
30	GE Medical Systems	Discovery CT750 HD	1.25	120	219	500	109.5
31	GE Medical Systems	Discovery CT750 HD	1.25	120	256	500	128.0
32	GE Medical Systems	Discovery CT750 HD	1.25	120	276	699	192.9
33	GE Medical Systems	Discovery CT750 HD	1.25	120	178	699	124.4
34	GE Medical Systems	Discovery CT750 HD	1.25	120	210	500	105.0
35	GE Medical Systems	Discovery CT750 HD	1.25	120	143	500	71.5
37	GE Medical Systems	Discovery CT750 HD	1.25	120	421	500	210.5
38	GE Medical Systems	Discovery CT750 HD	1.25	120	472	500	236.0
39	GE Medical Systems	Discovery CT750 HD	1.25	120	115	500	57.5
40	GE Medical Systems	Discovery CT750 HD	1.25	120	205	699	143.3
41	GE Medical Systems	Discovery CT750 HD	1.25	120	419	500	209.5
42	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
43	GE Medical Systems	Discovery CT750 HD	1.25	120	248	500	124.0
44	GE Medical Systems	Discovery CT750 HD	1.25	120	283	500	141.5
45	GE Medical Systems	Discovery CT750 HD	1.25	120	210	500	105.0
46	GE Medical Systems	Discovery CT750 HD	1.25	120	570	500	285.0
47	GE Medical Systems	Discovery CT750 HD	1.25	120	356	500	178.0
48	GE Medical Systems	Discovery CT750 HD	1.25	120	455	500	227.5
49	GE Medical Systems	Discovery CT750 HD	1.25	120	229	500	114.5
50	GE Medical Systems	Discovery CT750 HD	1.25	120	314	500	157.0
51	GE Medical Systems	Discovery CT750 HD	1.25	120	499	500	249.5
52	GE Medical Systems	Discovery CT750 HD	1.25	120	649	500	324.5
53	GE Medical Systems	Discovery CT750 HD	1.25	120	196	500	98.0
54	GE Medical Systems	Discovery CT750 HD	1.25	120	342	500	171.0
55	GE Medical Systems	Discovery CT750 HD	1.25	120	198	500	99.0
56	GE Medical Systems	Discovery CT750 HD	1.25	120	228	500	114.0
57	GE Medical Systems	Discovery CT750 HD	1.25	120	308	500	154.0
58	GE Medical Systems	Discovery CT750 HD	1.25	120	239	500	119.5
59	GE Medical Systems	Discovery CT750 HD	1.25	120	219	500	109.5
60	GE Medical Systems	Discovery CT750 HD	1.25	120	444	500	222.0
61	GE Medical Systems	Discovery CT750 HD	1.25	120	292	500	146.0
62	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
63	GE Medical Systems	Discovery CT750 HD	1.25	120	444	500	222.0
64	GE Medical Systems	Discovery CT750 HD	1.25	120	130	500	65.0
65	GE Medical Systems	Discovery CT750 HD	1.25	120	118	500	59.0
66	GE Medical Systems	Discovery CT750 HD	1.25	120	469	500	234.5
67	GE Medical Systems	Discovery CT750 HD	1.25	120	184	500	92.0
68	GE Medical Systems	Discovery CT750 HD	1.25	120	472	699	329.9
69	GE Medical Systems	Discovery CT750 HD	1.25	120	290	500	145.0
70	GE Medical Systems	Discovery CT750 HD	1.25	120	530	500	265.0
71	GE Medical Systems	Discovery CT750 HD	1.25	120	141	500	70.5
72	GE Medical Systems	Discovery CT750 HD	1.25	120	121	500	60.5
73	GE Medical Systems	Discovery CT750 HD	1.25	120	126	500	63.0
74	GE Medical Systems	Discovery CT750 HD	1.25	120	175	500	87.5
75	GE Medical Systems	Discovery CT750 HD	1.25	120	186	500	93.0
76	GE Medical Systems	Discovery CT750 HD	1.25	120	305	500	152.5
77	GE Medical Systems	Discovery CT750 HD	1.25	120	148	500	74.0
78	GE Medical Systems	Discovery CT750 HD	1.25	120	102	500	51.0
79	GE Medical Systems	Discovery CT750 HD	1.25	120	129	500	64.5
80	GE Medical Systems	Discovery CT750 HD	1.25	120	246	500	123.0
81	GE Medical Systems	Discovery CT750 HD	1.25	120	458	500	229.0
82	GE Medical Systems	Discovery CT750 HD	1.25	120	119	500	59.5
83	GE Medical Systems	Discovery CT750 HD	1.25	120	464	500	232.0
84	GE Medical Systems	Discovery CT750 HD	1.25	120	225	500	112.5
85	GE Medical Systems	Discovery CT750 HD	1.25	120	320	500	160.0
86	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
87	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0
88	GE Medical Systems	Discovery CT750 HD	1.25	120	116	500	58.0
89	GE Medical Systems	Discovery CT750 HD	1.25	120	292	500	146.0
90	GE Medical Systems	Discovery CT750 HD	1.25	120	468	500	234.0
91	GE Medical Systems	Discovery CT750 HD	1.25	120	201	500	100.5
92	GE Medical Systems	Discovery CT750 HD	1.25	120	101	600	60.6
93	GE Medical Systems	Discovery CT750 HD	1.25	120	133	500	66.5
94	GE Medical Systems	Discovery CT750 HD	1.25	120	359	500	179.5
95	GE Medical Systems	Discovery CT750 HD	1.25	120	361	600	216.6
96	GE Medical Systems	Discovery CT750 HD	1.25	120	183	500	91.5
97	GE Medical Systems	Discovery CT750 HD	1.25	120	316	500	158.0
98	GE Medical Systems	Discovery CT750 HD	1.25	120	404	500	202.0
99	GE Medical Systems	Discovery CT750 HD	1.25	120	287	500	143.5
100	GE Medical Systems	Discovery CT750 HD	1.25	120	146	500	73.0
101	GE Medical Systems	Discovery CT750 HD	1.25	120	100	500	50.0

https://dx.doi.org/10.21037/tcr-22-2683

Table S2 The list of robust radiomic features

No.	Feature name
1	GTVp_original_shape_VoxelVolume
2	GTVp_original_shape_Compactness1
3	GTVp_original_shape_Flatness
4	GTVp_original_firstorder_10Percentile
5	GTVp_original_firstorder_90Percentile
6	GTVp_original_firstorder_Energy
7	GTVp_original_firstorder_Entropy
8	GTVp_original_firstorder_InterquartileRange
9	GTVp_original_firstorder_Maximum
10	GTVp_original_firstorder_Minimum
11	GTVp_original_firstorder_Range
12	GTVp_original_glcm_ClusterShade
13	GTVp_original_glcm_Contrast
14	GTVp_original_glcm_Correlation
15	GTVp_original_glcm_Idmn
16	GTVp_original_glcm_Imc1
17	GTVp_original_glrlm_LongRunLowGrayLevelEmphasis
18	GTVp_original_glszm_SizeZoneNonUniformityNormalized
19	GTVp_original_gldm_LargeDependenceLowGrayLevelEmphasis
20	GTVp_original_gldm_SmallDependenceHighGrayLevelEmphasis
21	GTVp_log-sigma-0-5-mm-3D_firstorder_10Percentile
22	GTVp_log-sigma-0-5-mm-3D_firstorder_90Percentile
23	GTVp_log-sigma-0-5-mm-3D_firstorder_Kurtosis
24	GTVp_log-sigma-0-5-mm-3D_firstorder_Maximum
25	GTVp_log-sigma-0-5-mm-3D_firstorder_Mean
26	GTVp_log-sigma-0-5-mm-3D_firstorder_Minimum
27	GTVp_log-sigma-0-5-mm-3D_firstorder_Skewness
28	GTVp_log-sigma-0-5-mm-3D_glcm_Correlation
29	GTVp_log-sigma-0-5-mm-3D_glcm_Imc1
30	GTVp_log-sigma-0-5-mm-3D_glrlm_LongRunHighGrayLevelEmphasis
31	GTVp_log-sigma-0-5-mm-3D_glrlm_RunEntropy
32	GTVp_log-sigma-0-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
33	GTVp_log-sigma-0-5-mm-3D_glszm_ZoneEntropy
34	GTVp_log-sigma-0-5-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis
35	GTVp_log-sigma-1-0-mm-3D_firstorder_Kurtosis
36	GTVp_log-sigma-1-0-mm-3D_firstorder_Maximum
37	GTVp_log-sigma-1-0-mm-3D_firstorder_Minimum
38	GTVp_log-sigma-1-0-mm-3D_firstorder_Skewness
39	GTVp_log-sigma-1-0-mm-3D_glcm_Correlation
40	GTVp_log-sigma-1-0-mm-3D_glcm_Idmn
41	GTVp_log-sigma-1-0-mm-3D_glcm_Imc1
42	GTVp_log-sigma-1-0-mm-3D_glcm_MCC
43	GTVp_log-sigma-1-0-mm-3D_glrlm_LongRunHighGrayLevelEmphasis
44	GTVp_log-sigma-1-0-mm-3D_glrlm_RunEntropy
45	GTVp_log-sigma-1-0-mm-3D_glszm_SizeZoneNonUniformityNormalized
46	GTVp_log-sigma-1-0-mm-3D_gldm_DependenceEntropy
47	GTVp_log-sigma-1-5-mm-3D_firstorder_Kurtosis
48	GTVp_log-sigma-1-5-mm-3D_firstorder_Minimum
49	GTVp_log-sigma-1-5-mm-3D_glcm_Autocorrelation
50	GTVp_log-sigma-1-5-mm-3D_glcm_Imc1
51	GTVp_log-sigma-1-5-mm-3D_glszm_GrayLevelNonUniformityNormalized
52	GTVp_log-sigma-1-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
53	GTVp_log-sigma-2-0-mm-3D_firstorder_Minimum
54	GTVp_log-sigma-2-0-mm-3D_glcm_Autocorrelation
55	GTVp_log-sigma-2-0-mm-3D_glrlm_RunEntropy
56	GTVp_log-sigma-2-0-mm-3D_glszm_SizeZoneNonUniformityNormalized
57	GTVp_log-sigma-2-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
58	GTVp_log-sigma-2-0-mm-3D_ngtdm_Complexity
59	GTVp_log-sigma-2-5-mm-3D_firstorder_Minimum
60	GTVp_log-sigma-2-5-mm-3D_firstorder_Range
61	GTVp_log-sigma-2-5-mm-3D_glcm_Autocorrelation
62	GTVp_log-sigma-2-5-mm-3D_glszm_LowGrayLevelZoneEmphasis
63	GTVp_log-sigma-2-5-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
64	GTVp_wavelet-LLH_firstorder_10Percentile
65	GTVp_wavelet-LLH_firstorder_90Percentile
66	GTVp_wavelet-LLH_firstorder_Kurtosis
67	GTVp_wavelet-LLH_firstorder_Maximum
68	GTVp_wavelet-LLH_firstorder_Mean
69	GTVp_wavelet-LLH_firstorder_Minimum
70	GTVp_wavelet-LLH_firstorder_Skewness
71	GTVp_wavelet-LLH_glcm_Correlation
72	GTVp_wavelet-LLH_glcm_Imc1
73	GTVp_wavelet-LLH_glcm_MCC
74	GTVp_wavelet-LLH_glrlm_LongRunHighGrayLevelEmphasis
75	GTVp_wavelet-LLH_glszm_SizeZoneNonUniformityNormalized
76	GTVp_wavelet-LHL_firstorder_10Percentile
77	GTVp_wavelet-LHL_firstorder_90Percentile
78	GTVp_wavelet-LHL_firstorder_Kurtosis
79	GTVp_wavelet-LHL_firstorder_Maximum
80	GTVp_wavelet-LHL_firstorder_Skewness
81	GTVp_wavelet-LHL_glcm_Correlation
82	GTVp_wavelet-LHL_glcm_Imc1
83	GTVp_wavelet-LHL_glcm_MCC
84	GTVp_wavelet-LHL_glszm_SizeZoneNonUniformityNormalized
85	GTVp_wavelet-LHL_gldm_LargeDependenceLowGrayLevelEmphasis
86	GTVp_wavelet-LHH_firstorder_Kurtosis
87	GTVp_wavelet-LHH_firstorder_Maximum
88	GTVp_wavelet-LHH_firstorder_Mean
89	GTVp_wavelet-LHH_firstorder_Median
90	GTVp_wavelet-LHH_firstorder_Minimum
91	GTVp_wavelet-LHH_firstorder_Skewness
92	GTVp_wavelet-LHH_glcm_Correlation
93	GTVp_wavelet-LHH_glcm_Imc1
94	GTVp_wavelet-LHH_glcm_MCC
95	GTVp_wavelet-LHH_glrlm_RunEntropy
96	GTVp_wavelet-HLL_firstorder_10Percentile
97	GTVp_wavelet-HLL_firstorder_90Percentile
98	GTVp_wavelet-HLL_firstorder_Kurtosis
99	GTVp_wavelet-HLL_firstorder_Maximum
100	GTVp_wavelet-HLL_firstorder_Minimum
101	GTVp_wavelet-HLL_firstorder_Skewness
102	GTVp_wavelet-HLL_glcm_Correlation
103	GTVp_wavelet-HLL_glcm_MCC
104	GTVp_wavelet-HLL_glszm_SizeZoneNonUniformityNormalized
105	GTVp_wavelet-HLH_firstorder_Kurtosis
106	GTVp_wavelet-HLH_firstorder_Maximum
107	GTVp_wavelet-HLH_firstorder_Mean
108	GTVp_wavelet-HLH_firstorder_Median
109	GTVp_wavelet-HLH_firstorder_Minimum
110	GTVp_wavelet-HLH_glcm_ClusterShade
111	GTVp_wavelet-HLH_glcm_Correlation
112	GTVp_wavelet-HLH_glcm_Imc1
113	GTVp_wavelet-HLH_glcm_MCC
114	GTVp_wavelet-HLH_glrlm_RunEntropy
115	GTVp_wavelet-HLH_glszm_SizeZoneNonUniformityNormalized
116	GTVp_wavelet-HHL_firstorder_Kurtosis
117	GTVp_wavelet-HHL_firstorder_Maximum
118	GTVp_wavelet-HHL_firstorder_Minimum
119	GTVp_wavelet-HHL_firstorder_Skewness
120	GTVp_wavelet-HHL_glcm_ClusterShade
121	GTVp_wavelet-HHL_glcm_Correlation
122	GTVp_wavelet-HHL_glcm_MCC
123	GTVp_wavelet-HHL_glrlm_RunEntropy
124	GTVp_wavelet-HHL_glszm_SizeZoneNonUniformityNormalized
125	GTVp_wavelet-HHL_gldm_LargeDependenceLowGrayLevelEmphasis
126	GTVp_wavelet-HHH_firstorder_Kurtosis
127	GTVp_wavelet-HHH_firstorder_Maximum
128	GTVp_wavelet-HHH_firstorder_Median
129	GTVp_wavelet-HHH_firstorder_Minimum
130	GTVp_wavelet-HHH_firstorder_Skewness
131	GTVp_wavelet-HHH_glcm_ClusterShade
132	GTVp_wavelet-HHH_glcm_Correlation
133	GTVp_wavelet-HHH_glcm_Idmn
134	GTVp_wavelet-HHH_glcm_Imc1
135	GTVp_wavelet-HHH_glcm_InverseVariance
136	GTVp_wavelet-HHH_glrlm_RunEntropy
137	GTVp_wavelet-HHH_glszm_SizeZoneNonUniformityNormalized
138	GTVp_wavelet-HHH_glszm_ZoneEntropy
139	GTVp_wavelet-HHH_gldm_DependenceEntropy
140	GTVp_wavelet-HHH_gldm_DependenceNonUniformityNormalized
141	GTVp_wavelet-HHH_gldm_DependenceVariance
142	GTVp_wavelet-LLL_firstorder_Maximum
143	GTVp_wavelet-LLL_glcm_MCC
144	GTVp_wavelet-LLL_glszm_LargeAreaLowGrayLevelEmphasis
145	GTVp_wavelet-LLL_glszm_SizeZoneNonUniformityNormalized
146	Rim_original_shape_Compactness1
147	Rim_original_shape_Elongation
148	Rim_original_firstorder_10Percentile
149	Rim_original_firstorder_Entropy
150	Rim_original_firstorder_InterquartileRange
151	Rim_original_firstorder_Maximum
152	Rim_original_firstorder_Minimum
153	Rim_original_glcm_Contrast
154	Rim_original_glcm_Correlation
155	Rim_original_glcm_Id
156	Rim_original_glcm_Idmn
157	Rim_original_glcm_Imc1
158	Rim_original_glszm_LargeAreaLowGrayLevelEmphasis
159	Rim_original_glszm_SizeZoneNonUniformityNormalized
160	Rim_original_glszm_ZoneEntropy
161	Rim_original_ngtdm_Busyness
162	Rim_original_ngtdm_Complexity
163	Rim_log-sigma-0-5-mm-3D_firstorder_Kurtosis
164	Rim_log-sigma-0-5-mm-3D_firstorder_Mean
165	Rim_log-sigma-0-5-mm-3D_firstorder_Minimum
166	Rim_log-sigma-0-5-mm-3D_firstorder_Skewness
167	Rim_log-sigma-0-5-mm-3D_glcm_Correlation
168	Rim_log-sigma-0-5-mm-3D_glcm_Idmn
169	Rim_log-sigma-0-5-mm-3D_glcm_Imc1
170	Rim_log-sigma-0-5-mm-3D_glcm_MCC
171	Rim_log-sigma-0-5-mm-3D_glszm_LargeAreaEmphasis
172	Rim_log-sigma-0-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
173	Rim_log-sigma-0-5-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
174	Rim_log-sigma-1-0-mm-3D_firstorder_Maximum
175	Rim_log-sigma-1-0-mm-3D_firstorder_Minimum
176	Rim_log-sigma-1-0-mm-3D_glcm_Correlation
177	Rim_log-sigma-1-0-mm-3D_glcm_Idmn
178	Rim_log-sigma-1-0-mm-3D_glcm_Imc1
179	Rim_log-sigma-1-0-mm-3D_glcm_MCC
180	Rim_log-sigma-1-0-mm-3D_glszm_GrayLevelNonUniformityNormalized
181	Rim_log-sigma-1-0-mm-3D_glszm_SizeZoneNonUniformityNormalized
182	Rim_log-sigma-1-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis
183	Rim_log-sigma-1-5-mm-3D_firstorder_Maximum
184	Rim_log-sigma-1-5-mm-3D_firstorder_Minimum
185	Rim_log-sigma-1-5-mm-3D_firstorder_Range
186	Rim_log-sigma-1-5-mm-3D_glcm_Correlation
187	Rim_log-sigma-1-5-mm-3D_glcm_Idmn
188	Rim_log-sigma-1-5-mm-3D_glcm_Imc1
189	Rim_log-sigma-1-5-mm-3D_glszm_GrayLevelNonUniformityNormalized
190	Rim_log-sigma-1-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
191	Rim_log-sigma-2-0-mm-3D_firstorder_Minimum
192	Rim_log-sigma-2-0-mm-3D_firstorder_Range
193	Rim_log-sigma-2-0-mm-3D_glcm_Correlation
194	Rim_log-sigma-2-0-mm-3D_glszm_GrayLevelNonUniformityNormalized
195	Rim_log-sigma-2-0-mm-3D_glszm_HighGrayLevelZoneEmphasis
196	Rim_log-sigma-2-0-mm-3D_glszm_LowGrayLevelZoneEmphasis
197	Rim_log-sigma-2-0-mm-3D_glszm_SizeZoneNonUniformityNormalized
198	Rim_log-sigma-2-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
199	Rim_log-sigma-2-0-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
200	Rim_log-sigma-2-5-mm-3D_firstorder_Range
201	Rim_log-sigma-2-5-mm-3D_glszm_GrayLevelNonUniformityNormalized
202	Rim_log-sigma-2-5-mm-3D_glszm_HighGrayLevelZoneEmphasis
203	Rim_log-sigma-2-5-mm-3D_glszm_LowGrayLevelZoneEmphasis
204	Rim_log-sigma-2-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
205	Rim_log-sigma-2-5-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
206	Rim_wavelet-LLH_firstorder_10Percentile
207	Rim_wavelet-LLH_firstorder_Kurtosis
208	Rim_wavelet-LLH_firstorder_Mean
209	Rim_wavelet-LLH_firstorder_Median
210	Rim_wavelet-LLH_firstorder_Minimum
211	Rim_wavelet-LLH_firstorder_Skewness
212	Rim_wavelet-LLH_glcm_ClusterShade
213	Rim_wavelet-LLH_glcm_Correlation
214	Rim_wavelet-LLH_glcm_Idmn
215	Rim_wavelet-LLH_glcm_Imc1
216	Rim_wavelet-LLH_glcm_MCC
217	Rim_wavelet-LLH_glszm_LargeAreaEmphasis
218	Rim_wavelet-LLH_glszm_SizeZoneNonUniformityNormalized
219	Rim_wavelet-LLH_gldm_LargeDependenceLowGrayLevelEmphasis
220	Rim_wavelet-LHL_firstorder_10Percentile
221	Rim_wavelet-LHL_firstorder_Kurtosis
222	Rim_wavelet-LHL_firstorder_Maximum
223	Rim_wavelet-LHL_firstorder_Mean
224	Rim_wavelet-LHL_firstorder_Minimum
225	Rim_wavelet-LHL_firstorder_Skewness
226	Rim_wavelet-LHL_glcm_ClusterShade
227	Rim_wavelet-LHL_glcm_Correlation
228	Rim_wavelet-LHL_glcm_Idmn
229	Rim_wavelet-LHL_glcm_Imc1
230	Rim_wavelet-LHL_glcm_MCC
231	Rim_wavelet-LHL_glszm_LargeAreaEmphasis
232	Rim_wavelet-LHL_glszm_SizeZoneNonUniformityNormalized
233	Rim_wavelet-LHL_gldm_LargeDependenceLowGrayLevelEmphasis
234	Rim_wavelet-LHH_firstorder_10Percentile
235	Rim_wavelet-LHH_firstorder_Kurtosis
236	Rim_wavelet-LHH_firstorder_Maximum
237	Rim_wavelet-LHH_firstorder_Mean
238	Rim_wavelet-LHH_firstorder_Median
239	Rim_wavelet-LHH_firstorder_Skewness
240	Rim_wavelet-LHH_glcm_ClusterShade
241	Rim_wavelet-LHH_glcm_Correlation
242	Rim_wavelet-LHH_glcm_Idmn
243	Rim_wavelet-LHH_glcm_Imc1
244	Rim_wavelet-LHH_glszm_SizeZoneNonUniformityNormalized
245	Rim_wavelet-HLL_firstorder_Kurtosis
246	Rim_wavelet-HLL_firstorder_Mean
247	Rim_wavelet-HLL_firstorder_Median
248	Rim_wavelet-HLL_firstorder_Minimum
249	Rim_wavelet-HLL_firstorder_Skewness
250	Rim_wavelet-HLL_glcm_ClusterShade
251	Rim_wavelet-HLL_glcm_Correlation
252	Rim_wavelet-HLL_glcm_Idmn
253	Rim_wavelet-HLL_glcm_MCC
254	Rim_wavelet-HLL_gldm_LargeDependenceLowGrayLevelEmphasis
255	Rim_wavelet-HLH_firstorder_Kurtosis
256	Rim_wavelet-HLH_firstorder_Maximum
257	Rim_wavelet-HLH_firstorder_Mean
258	Rim_wavelet-HLH_firstorder_Median
259	Rim_wavelet-HLH_firstorder_Skewness
260	Rim_wavelet-HLH_glcm_ClusterShade
261	Rim_wavelet-HLH_glcm_Correlation
262	Rim_wavelet-HLH_glcm_Idmn
263	Rim_wavelet-HLH_glcm_Imc1
264	Rim_wavelet-HLH_glszm_SizeZoneNonUniformityNormalized
265	Rim_wavelet-HHL_firstorder_Kurtosis
266	Rim_wavelet-HHL_firstorder_Maximum
267	Rim_wavelet-HHL_firstorder_Mean
268	Rim_wavelet-HHL_firstorder_Median
269	Rim_wavelet-HHL_firstorder_Minimum
270	Rim_wavelet-HHL_firstorder_Skewness
271	Rim_wavelet-HHL_glcm_ClusterShade
272	Rim_wavelet-HHL_glcm_Correlation
273	Rim_wavelet-HHL_glcm_Idmn
274	Rim_wavelet-HHL_glcm_Imc1
275	Rim_wavelet-HHL_glcm_MCC
276	Rim_wavelet-HHL_glszm_SizeZoneNonUniformityNormalized
277	Rim_wavelet-HHL_ngtdm_Contrast
278	Rim_wavelet-HHH_firstorder_Kurtosis
279	Rim_wavelet-HHH_firstorder_Maximum
280	Rim_wavelet-HHH_firstorder_Mean
281	Rim_wavelet-HHH_firstorder_Median
282	Rim_wavelet-HHH_firstorder_Minimum
283	Rim_wavelet-HHH_firstorder_Skewness
284	Rim_wavelet-HHH_glcm_ClusterShade
285	Rim_wavelet-HHH_glcm_Correlation
286	Rim_wavelet-HHH_glcm_Idmn
287	Rim_wavelet-HHH_glcm_Imc1
288	Rim_wavelet-HHH_glszm_SizeZoneNonUniformityNormalized
289	Rim_wavelet-HHH_glszm_ZoneEntropy
290	Rim_wavelet-HHH_glszm_ZonePercentage
291	Rim_wavelet-HHH_gldm_DependenceVariance
292	Rim_wavelet-LLL_glcm_Imc1
293	Rim_wavelet-LLL_glszm_LargeAreaLowGrayLevelEmphasis
294	Exterior_original_shape_Compactness1
295	Exterior_original_shape_SurfaceVolumeRatio
296	Exterior_original_shape_Flatness
297	Exterior_original_firstorder_10Percentile
298	Exterior_original_firstorder_90Percentile
299	Exterior_original_firstorder_Maximum
300	Exterior_original_firstorder_Minimum
301	Exterior_original_glcm_Idmn
302	Exterior_original_glszm_LargeAreaEmphasis
303	Exterior_original_glszm_SizeZoneNonUniformityNormalized
304	Exterior_original_ngtdm_Busyness
305	Exterior_log-sigma-0-5-mm-3D_firstorder_Mean
306	Exterior_log-sigma-0-5-mm-3D_firstorder_Median
307	Exterior_log-sigma-0-5-mm-3D_firstorder_Minimum
308	Exterior_log-sigma-0-5-mm-3D_firstorder_Skewness
309	Exterior_log-sigma-0-5-mm-3D_glcm_ClusterShade
310	Exterior_log-sigma-0-5-mm-3D_glcm_Correlation
311	Exterior_log-sigma-0-5-mm-3D_glcm_Idmn
312	Exterior_log-sigma-0-5-mm-3D_glcm_Imc1
313	Exterior_log-sigma-0-5-mm-3D_glcm_Imc2
314	Exterior_log-sigma-0-5-mm-3D_glcm_MCC
315	Exterior_log-sigma-0-5-mm-3D_glrlm_LongRunHighGrayLevelEmphasis
316	Exterior_log-sigma-0-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
317	Exterior_log-sigma-0-5-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
318	Exterior_log-sigma-1-0-mm-3D_firstorder_Kurtosis
319	Exterior_log-sigma-1-0-mm-3D_firstorder_Minimum
320	Exterior_log-sigma-1-0-mm-3D_firstorder_Skewness
321	Exterior_log-sigma-1-0-mm-3D_glcm_ClusterShade
322	Exterior_log-sigma-1-0-mm-3D_glcm_Correlation
323	Exterior_log-sigma-1-0-mm-3D_glcm_Idmn
324	Exterior_log-sigma-1-0-mm-3D_glcm_Imc1
325	Exterior_log-sigma-1-0-mm-3D_glcm_Imc2
326	Exterior_log-sigma-1-0-mm-3D_glcm_MCC
327	Exterior_log-sigma-1-0-mm-3D_glrlm_LongRunHighGrayLevelEmphasis
328	Exterior_log-sigma-1-0-mm-3D_glrlm_LongRunLowGrayLevelEmphasis
329	Exterior_log-sigma-1-0-mm-3D_glszm_SizeZoneNonUniformityNormalized
330	Exterior_log-sigma-1-0-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
331	Exterior_log-sigma-1-5-mm-3D_firstorder_Minimum
332	Exterior_log-sigma-1-5-mm-3D_glcm_Correlation
333	Exterior_log-sigma-1-5-mm-3D_glcm_Idmn
334	Exterior_log-sigma-1-5-mm-3D_glcm_InverseVariance
335	Exterior_log-sigma-1-5-mm-3D_glrlm_LongRunHighGrayLevelEmphasis
336	Exterior_log-sigma-1-5-mm-3D_glszm_SizeZoneNonUniformityNormalized
337	Exterior_log-sigma-1-5-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
338	Exterior_log-sigma-2-0-mm-3D_firstorder_Minimum
339	Exterior_log-sigma-2-0-mm-3D_glcm_Idmn
340	Exterior_log-sigma-2-0-mm-3D_glcm_InverseVariance
341	Exterior_log-sigma-2-0-mm-3D_glrlm_LongRunHighGrayLevelEmphasis
342	Exterior_log-sigma-2-0-mm-3D_glszm_GrayLevelNonUniformityNormalized
343	Exterior_log-sigma-2-0-mm-3D_glszm_SizeZoneNonUniformityNormalized
344	Exterior_log-sigma-2-0-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
345	Exterior_log-sigma-2-5-mm-3D_firstorder_Minimum
346	Exterior_log-sigma-2-5-mm-3D_firstorder_Range
347	Exterior_log-sigma-2-5-mm-3D_glcm_Idmn
348	Exterior_log-sigma-2-5-mm-3D_glcm_InverseVariance
349	Exterior_log-sigma-2-5-mm-3D_glrlm_LongRunLowGrayLevelEmphasis
350	Exterior_log-sigma-2-5-mm-3D_glszm_GrayLevelNonUniformityNormalized
351	Exterior_log-sigma-2-5-mm-3D_glszm_LowGrayLevelZoneEmphasis
352	Exterior_log-sigma-2-5-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
353	Exterior_log-sigma-2-5-mm-3D_gldm_SmallDependenceHighGrayLevelEmphasis
354	Exterior_log-sigma-2-5-mm-3D_gldm_SmallDependenceLowGrayLevelEmphasis
355	Exterior_wavelet-LLH_firstorder_90Percentile
356	Exterior_wavelet-LLH_firstorder_Kurtosis
357	Exterior_wavelet-LLH_firstorder_Mean
358	Exterior_wavelet-LLH_firstorder_Median
359	Exterior_wavelet-LLH_firstorder_Minimum
360	Exterior_wavelet-LLH_firstorder_Skewness
361	Exterior_wavelet-LLH_glcm_ClusterShade
362	Exterior_wavelet-LLH_glcm_Correlation
363	Exterior_wavelet-LLH_glcm_Idmn
364	Exterior_wavelet-LLH_glcm_Imc1
365	Exterior_wavelet-LLH_glszm_LargeAreaLowGrayLevelEmphasis
366	Exterior_wavelet-LLH_glszm_SizeZoneNonUniformityNormalized
367	Exterior_wavelet-LLH_ngtdm_Busyness
368	Exterior_wavelet-LLH_ngtdm_Contrast
369	Exterior_wavelet-LHL_firstorder_90Percentile
370	Exterior_wavelet-LHL_firstorder_Kurtosis
371	Exterior_wavelet-LHL_firstorder_Maximum
372	Exterior_wavelet-LHL_firstorder_Minimum
373	Exterior_wavelet-LHL_firstorder_Skewness
374	Exterior_wavelet-LHL_glcm_ClusterShade
375	Exterior_wavelet-LHL_glcm_Correlation
376	Exterior_wavelet-LHL_glcm_Idmn
377	Exterior_wavelet-LHL_glcm_Imc1
378	Exterior_wavelet-LHL_glcm_Imc2
379	Exterior_wavelet-LHL_glcm_MCC
380	Exterior_wavelet-LHL_glszm_SizeZoneNonUniformityNormalized
381	Exterior_wavelet-LHL_glszm_ZoneEntropy
382	Exterior_wavelet-LHL_ngtdm_Busyness
383	Exterior_wavelet-LHH_firstorder_10Percentile
384	Exterior_wavelet-LHH_firstorder_Kurtosis
385	Exterior_wavelet-LHH_firstorder_Maximum
386	Exterior_wavelet-LHH_firstorder_Mean
387	Exterior_wavelet-LHH_firstorder_Median
388	Exterior_wavelet-LHH_firstorder_Minimum
389	Exterior_wavelet-LHH_firstorder_Skewness
390	Exterior_wavelet-LHH_glcm_ClusterShade
391	Exterior_wavelet-LHH_glcm_Correlation
392	Exterior_wavelet-LHH_glcm_Idmn
393	Exterior_wavelet-LHH_glcm_Imc1
394	Exterior_wavelet-LHH_glcm_MCC
395	Exterior_wavelet-LHH_glszm_LargeAreaLowGrayLevelEmphasis
396	Exterior_wavelet-LHH_glszm_SizeZoneNonUniformityNormalized
397	Exterior_wavelet-LHH_gldm_DependenceNonUniformityNormalized
398	Exterior_wavelet-LHH_gldm_DependenceVariance
399	Exterior_wavelet-HLL_firstorder_Maximum
400	Exterior_wavelet-HLL_firstorder_Mean
401	Exterior_wavelet-HLL_firstorder_Minimum
402	Exterior_wavelet-HLL_firstorder_Skewness
403	Exterior_wavelet-HLL_glcm_ClusterShade
404	Exterior_wavelet-HLL_glcm_Correlation
405	Exterior_wavelet-HLL_glcm_Idmn
406	Exterior_wavelet-HLL_glcm_Imc1
407	Exterior_wavelet-HLL_glcm_Imc2
408	Exterior_wavelet-HLL_glcm_MCC
409	Exterior_wavelet-HLL_gldm_SmallDependenceLowGrayLevelEmphasis
410	Exterior_wavelet-HLL_ngtdm_Busyness
411	Exterior_wavelet-HLH_firstorder_Kurtosis
412	Exterior_wavelet-HLH_firstorder_Maximum
413	Exterior_wavelet-HLH_firstorder_Mean
414	Exterior_wavelet-HLH_firstorder_Median
415	Exterior_wavelet-HLH_firstorder_Minimum
416	Exterior_wavelet-HLH_firstorder_Skewness
417	Exterior_wavelet-HLH_glcm_ClusterShade
418	Exterior_wavelet-HLH_glcm_Correlation
419	Exterior_wavelet-HLH_glcm_Idmn
420	Exterior_wavelet-HLH_glcm_Imc1
421	Exterior_wavelet-HLH_glcm_MCC
422	Exterior_wavelet-HLH_glszm_LargeAreaLowGrayLevelEmphasis
423	Exterior_wavelet-HLH_glszm_SizeZoneNonUniformityNormalized
424	Exterior_wavelet-HLH_gldm_DependenceNonUniformityNormalized
425	Exterior_wavelet-HLH_gldm_DependenceVariance
426	Exterior_wavelet-HHL_firstorder_Kurtosis
427	Exterior_wavelet-HHL_firstorder_Maximum
428	Exterior_wavelet-HHL_firstorder_Mean
429	Exterior_wavelet-HHL_firstorder_Median
430	Exterior_wavelet-HHL_firstorder_Minimum
431	Exterior_wavelet-HHL_firstorder_Skewness
432	Exterior_wavelet-HHL_glcm_ClusterShade
433	Exterior_wavelet-HHL_glcm_Correlation
434	Exterior_wavelet-HHL_glcm_Idmn
435	Exterior_wavelet-HHL_glcm_Imc1
436	Exterior_wavelet-HHL_glcm_Imc2
437	Exterior_wavelet-HHL_glcm_MCC
438	Exterior_wavelet-HHL_glszm_SizeZoneNonUniformityNormalized
439	Exterior_wavelet-HHL_gldm_DependenceNonUniformityNormalized
440	Exterior_wavelet-HHH_firstorder_Kurtosis
441	Exterior_wavelet-HHH_firstorder_Maximum
442	Exterior_wavelet-HHH_firstorder_Mean
443	Exterior_wavelet-HHH_firstorder_Median
444	Exterior_wavelet-HHH_firstorder_Minimum
445	Exterior_wavelet-HHH_firstorder_Skewness
446	Exterior_wavelet-HHH_glcm_ClusterShade
447	Exterior_wavelet-HHH_glcm_Correlation
448	Exterior_wavelet-HHH_glcm_Idmn
449	Exterior_wavelet-HHH_glcm_Imc1
450	Exterior_wavelet-HHH_glcm_MCC
451	Exterior_wavelet-HHH_glszm_SizeZoneNonUniformityNormalized
452	Exterior_wavelet-LLL_glszm_LargeAreaHighGrayLevelEmphasis
453	Exterior_wavelet-LLL_glszm_SizeZoneNonUniformityNormalized

Table S3 Performance of radiomics model and combined model in prediction of EGFR mutation

Random Forest	Radiomics model						Combined model					
	S1	S2	S3	S4	S5	Mean	S1	S2	S3	S4	S5	Mean
Training set (mutation+)	79 (37)	79 (37)	79 (37)	79 (36)	80 (37)		79 (37)	79 (37)	79 (37)	79 (36)	80 (37)	
Validation set (mutation+)	20 (9)	20 (9)	20 (9)	20 (10)	19 (9)		20 (9)	20 (9)	20 (9)	20 (10)	19 (9)	
Training												
Accuracy	0.65	0.79	0.87	0.61	0.94	0.77	0.75	0.80	0.76	0.72	0.71	0.75
Recall	0.97	0.81	0.97	0.94	0.92	0.92	0.78	0.78	0.62	0.69	0.51	0.68
Precision	0.57	0.75	0.80	0.54	0.94	0.72	0.71	0.78	0.88	0.74	0.88	0.76
Specificity	0.36	0.76	0.79	0.33	0.95	0.64	0.71	0.81	0.88	0.74	0.88	0.81
F1	0.72	0.78	0.88	0.69	0.93	0.80	0.74	0.78	0.71	0.69	0.62	0.71
AUC	0.67	0.79	0.88	0.64	0.94	0.78	0.75	0.80	0.75	0.72	0.70	0.74
Validation												
Accuracy	0.65	0.80	0.65	0.80	0.79	0.74	0.75	0.85	0.80	0.95	0.84	0.84
Recall	0.89	1.00	0.67	1.0	0.78	0.87	0.67	0.89	0.67	1.00	0.78	0.80
Precision	0.57	0.69	0.60	0.71	0.78	0.67	0.75	0.80	0.86	0.91	0.88	0.84
Specificity	0.46	0.64	0.64	0.60	0.80	0.63	0.82	0.82	0.91	0.90	0.90	0.87
F1	0.70	–0.82	0.63	0.83	0.78	0.75	0.71	0.84	0.75	0.95	0.82	0.81
AUC	0.67	0.82	0.65	0.80	0.79	0.75	0.74	0.85	0.79	0.95	0.84	0.83

EGFR, epidermal growth factor receptor; AUC, area under the curve.

https://dx.doi.org/10.21037/tcr-22-2683

Table S4 Selected features and coefficient in each prediction models

Features	S1	S2	S3	S4	S5
Radiomics model					
GTVp_log-sigma-2-5-mm-3D_firstorder_Minimum	<0.0001	n.a	n.a	n.a	n.a
GTVp_wavelet-HLL_firstorder_90Percentile	<0.0001	n.a	n.a	n.a	n.a
GTVp_wavelet-HHH_gldm_DependenceVariance	1	0.69	0.25	1	0.10
Rim_wavelet-HLH_glcm_Correlation	<0.0001	n.a	n.a	n.a	0.15
Exterior_wavelet-LHL_firstorder_90Percentile	<0.0001	0.23	n.a	n.a	0.21
Exterior_wavelet-HHL_firstorder_Kurtosis	<0.0001	n.a	n.a	n.a	0.064
Rim_log-sigma-2-0-mm-3D_glcm_Correlation	n.a	0.088	n.a	n.a	n.a
GTVp_original_firstorder_Entropy	n.a	n.a	0.26	n.a	n.a
GTVp_wavelet-LHH_glcm_Imc1	n.a	n.a	0.20	n.a	n.a
GTVp_wavelet-HHH_gldm_DependenceNonUniformityNormalized	n.a	n.a	0.29	n.a	n.a
Exterior_wavelet-HHL_glcm_MCC	n.a	n.a	n.a	<0.0001	n.a
GTVp_log-sigma-1-0-mm-3D_gldm_DependenceEntropy	n.a	n.a	n.a	n.a	0.088
Rim_wavelet-LHL_glcm_Idmn	n.a	n.a	n.a	n.a	0.098
Exterior_wavelet-HHH_firstorder_Skewness	n.a	n.a	n.a	n.a	0.14
Exterior_wavelet-HHH_glcm_Correlation	n.a	n.a	n.a	n.a	0.15
Combined model					
GTVp_log-sigma-2-5-mm-3D_firstorder_Minimum	0.037	n.a	n.a	n.a	n.a
GTVp_wavelet-HLL_firstorder_90Percentile	<0.0001	n.a	n.a	n.a	n.a
GTVp_wavelet-HHH_gldm_DependenceVariance	0.059	0.26	0.17	0.33	0.20
Rim_wavelet-HLH_glcm_Correlation	<0.0001	n.a	n.a	n.a	<0.0001
Exterior_wavelet-LHL_firstorder_90Percentile	0.59	0.20	n.a	n.a	0.40
Exterior_wavelet-HHL_firstorder_Kurtosis	0.0019	n.a	n.a	n.a	0.20
Sex	0.20	0.026	0.21	0.50	<0.0001
Smoking	0.12	0.25	0.13	<0.0001	<0.0001
Rim_log-sigma-2-0-mm-3D_glcm_Correlation	n.a	0.26	n.a	n.a	n.a
GTVp_original_firstorder_Entropy	<0.0001	<0.0001	0.083	<0.0001	<0.0001
GTVp_wavelet-LHH_glcm_Imc1	n.a	n.a	0.17	n.a	n.a
GTVp_wavelet-HHH_gldm_DependenceNonUniformityNormalized	n.a	n.a	0.25	n.a	n.a
Exterior_wavelet-HHL_glcm_MCC	n.a	n.a	n.a	0.17	n.a
GTVp_log-sigma-1-0-mm-3D_gldm_DependenceEntropy	n.a	n.a	n.a	n.a	<0.0001
Rim_wavelet-LHL_glcm_Idmn	n.a	n.a	n.a	n.a	<0.0001
Exterior_wavelet-HHH_firstorder_Skewness	n.a	n.a	n.a	n.a	<0.0001
Exterior_wavelet-HHH_glcm_Correlation	n.a	n.a	n.a	n.a	0.2

Each coefficient was standardized so that the sum of the selected coefficients in each model was exactly 1. By performing the standardization, we were able to estimate and compare the importance of each feature by considering the average of five sessions. Not applicable (n.a.) was indicated when the feature was not selected in each session but was selected in other sessions, which was considered 0 when the average was calculated. HLL, HHL, LHL, HLH: wavelet filter. GTV, gross tumor volume; log, Laplacian-of-Gaussian.

https://dx.doi.org/10.21037/tcr-22-2683

Table S5 Performance of clinical model and all-combined model in prediction of EGFR mutation

Random Forest	Clinical model						All-combined model					
	S1	S2	S3	S4	S5	Mean	S1	S2	S3	S4	S5	Mean
Training set (mutation+)	79 (37)	79 (37)	79 (37)	79 (36)	80 (37)		79 (37)	79 (37)	79 (37)	79 (36)	80 (37)	
Validation set (mutation+)	20 (9)	20 (9)	20 (9)	20 (10)	19 (9)		20 (9)	20 (9)	20 (9)	20 (10)	19 (9)	
Training												
Accuracy	0.72	0.73	0.57	0.71	0.66	0.68	0.76	0.79	0.77	0.70	0.78	0.76
Recall	0.78	0.65	0.62	0.78	0.81	0.73	0.73	0.84	0.70	0.69	0.70	0.73
Precision	0.67	0.75	0.54	0.65	0.60	0.64	0.75	0.74	0.79	0.66	0.79	0.74
Specificity	0.67	0.81	0.52	0.65	0.54	0.64	0.79	0.74	0.83	0.70	0.84	0.78
F1	0.73	0.70	0.58	0.71	0.69	0.68	0.74	0.79	0.74	0.68	0.74	0.76
AUC	0.73	0.73	0.57	0.71	0.67	0.68	0.76	0.79	0.79	0.70	0.77	0.74
Validation												
Accuracy	0.80	0.85	0.75	0.80	0.63	0.77	0.70	0.85	0.80	0.95	0.79	0.82
Recall	0.78	0.89	0.78	0.80	0.67	0.78	0.67	1.00	0.67	1.00	0.89	0.84
Precision	0.78	0.80	0.70	0.80	0.60	0.74	0.67	0.75	0.86	0.91	0.73	0.78
Specificity	0.82	0.82	0.73	0.80	0.60	0.75	0.73	0.73	0.91	0.95	0.70	0.79
F1	0.78	0.84	0.74	0.80	0.63	0.76	0.67	0.86	0.86	0.95	0.80	0.81
AUC	0.80	0.85	0.75	0.80	0.63	0.77	0.70	0.86	0.79	0.95	0.79	0.82

EGFR, epidermal growth factor receptor; AUC, area under the curve.

https://dx.doi.org/10.21037/tcr-22-2683