Role of intratumoral and peritumoral CT radiomics for the prediction of EGFR gene mutation in primary lung cancer

MOTOHIKO YAMAZAKI, TAKUYA YAGI, MASAKI TOMINAGA, KOJIRO MINATO and HIROYUKI ISHIKAWA

Department of Radiology and Radiation Oncology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan

Address correspondence to: Dr Motohiko Yamazaki
E-mail: xackey2001@gmail.com

INTRODUCTION

Epidermal growth factor receptor (EGFR) gene mutations are oncogenic driver mutations in primary lung cancer (PLC) and are frequently observed in non-smoking Asian females. The discovery of EGFR mutations has brought a revolutionary change to PLC treatment because mutated tumors respond well to EGFR tyrosine kinase inhibitors (TKIs). The tissue examination is required to confirm EGFR mutations; however, biopsy is invasive and carries a risk of complications. As such, CT examination, which is a non-invasive and readily available tool, plays an important role in predicting mutations.

Radiomics is an image analysis methodology that extracts a large number of quantitative features that cannot be assessed by the human eye, and has recently garnered considerable attention in predicting EGFR mutations. In oncology, radiomics is mainly divided into intra- and peritumoral radiomics. Nevertheless, many previous studies have focused on only intratumoral radiomic features in the prediction of EGFR mutations. Therefore, the importance of peritumoral radiomic features in detecting mutations is not well known. Given that peritumoral radiomics is associated with many aspects of PLC, such as prognosis, histology, and response to chemotherapy, we hypothesized that peritumoral as well as intratumoral features would provide useful information for predicting EGFR mutations.

Objectives

To determine the added value of combining intratumoral and peritumoral CT radiomics for the prediction of epidermal growth factor receptor (EGFR) gene mutations in primary lung cancer (PLC).

Methods

This study included 478 patients with PLC (348 adenocarcinomas and 130 other histological types) who underwent surgical resection and EGFR gene testing. Two radiologists performed segmentation of tumors and peritumoral regions using precontrast high-resolution CT images, and 398 radiomic features (212 intra- and 186 peritumoral features) were extracted. The peritumoral region was defined as the lung parenchyma within a distance of 3 mm from the tumor border. Model performance was estimated using Random Forest, a machine-learning algorithm.

Results

EGFR mutations were found in 162 tumors; 161 adenocarcinomas, and one pleomorphic carcinoma. After exclusion of poorly reproducible and redundant features, 32 radiomic features remained (14 intra- and 18 peritumoral features) and were included in the model building. For predicting EGFR mutations, combining intra- and peritumoral radiomics significantly improved the performance compared to intratumoral radiomics alone (AUC [area under the receiver operating characteristic curve], 0.774 vs 0.730; p < 0.001). Even in adenocarcinomas only, adding peritumoral radiomics significantly increased performance (AUC, 0.687 vs 0.630; p < 0.001). The predictive performance using radiomics and clinical features was significantly higher than that of clinical features alone (AUC, 0.826 vs 0.777; p = 0.005).

Conclusions

Combining intra- and peritumoral radiomics improves the predictive accuracy of EGFR mutations and could be used to aid in decision-making of whether to perform biopsy for gene tests.

Advances in knowledge

Adding peritumoral to intratumoral radiomics yields greater accuracy than intratumoral radiomics alone in predicting EGFR mutations and may serve as a non-invasive method of predicting the gene status in PLC.

FULL PAPER
The purpose of this study was to determine the added value of combining intra- and peritumoral CT radiomics for the prediction of EGFR mutations in PLC.

**METHODS AND MATERIALS**

**Patients**

This single institution retrospective study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Ethics Committee (approved number, 2020–0420). Requirement for informed consent was waived because of the retrospective nature of this study. A total of 599 consecutive patients with PLC who underwent surgical resection between November 2015 and December 2020 were considered for this study. We excluded 121 patients for the following reasons: (1) EGFR gene testing was not performed \( n = 66 \); (2) precontrast high-resolution CT (HRCT) was not performed within three months preoperatively at our institution \( n = 43 \); (3) received chemotherapy or radiotherapy preoperatively \( n = 10 \); and (4) unclear tumor margin on HRCT \( n = 2 \). Finally, 478 patients with 478 PLCs \( 288 \) men and 190 women; median age, 70 years; range, 29–87 years) were included in the present study. Reasons for detecting lung cancers were various, including health checkup, associated symptoms, and incidental detection. Before surgery, chest and abdominal CT, brain CT or MRI, and 18F-fluorodeoxyglucose PET/CT were carried out to determine the clinical stage.

**CT acquisition**

Imaging was performed using one of the following three CT scanners: (1) SOMATOM Force (Siemens, Germany), (2) Aquilion ONE (Canon, Japan), or (3) Ingenuity Elite (Philips, the Netherlands). Of the machines, the SOMATOM Force was mainly used \( 317/478 \), 66.3\%). After patients were placed in the supine position, scanning from the lung apex to the base was performed during the deep inspiratory phase. Most patients received both pre- and postcontrast CTs for preoperative staging. However, during the deep inspiratory phase. Most patients received both pre- and postcontrast CTs for preoperative staging. However, eight patients underwent only precontrast CT because of renal dysfunction or iodine allergy. Although postcontrast CT is the standard for lung cancer staging, we routinely perform both pre- and postcontrast CTs unless there is contraindication for contrast agents because adding precontrast CT allows to accurately evaluate the degree of lesions’ enhancement. When a preoperative biopsy cannot confirm the diagnosis of lung nodules, the degree of enhancement can be useful to characterize them.\cite{17} Adding precontrast CT also has the benefit in differentiating between high-density mediastinal cysts and lymph node enlargement which can be misleading on only postcontrast CT. In this study, precontrast HRCT images were used for the analysis. The scanning parameters for HRCT were as follows: tube current, automatic exposure control; tube voltage, 120 kVp; detector pitch, 0.813–1.172 mm; detector collimation, 64–96 × 0.3–0.625 mm; gantry rotation time, 0.4–0.5 s; field of view, 200 × 200 mm; pixel spacing, 512 × 512; and reconstruction slice thickness, 1.0 mm. A sharp algorithm was used for all the reconstruction factors.

**Gene testing**

Gene testing was performed on specimens acquired either by surgery \( n = 330 \) or transbronchoscopy \( n = 148 \). Various gene testing techniques were used in this retrospective study, but most \( 403/478 \), 84.3\% were polymerase chain reaction clamp-based test.\cite{18}

**Radiomic analysis**

The acquired CT images were anonymized and then transferred to a personal computer that was not connected to the network. One radiologist with 17 years of experience semi-automatically segmented the tumor and a peritumoral region using Segmentation Editor Module\cite{19} implemented in 3D Slicer\cite{20} version 4.13.0. The peritumoral region was defined as the lung parenchyma within a distance of 3 mm from the tumor border, which included air in the lung, pulmonary vessels, and bronchi, but not the thoracic wall and mediastinum. This criterion was based on a previous report that the peritumoral lung parenchyma within 3 mm without including the thoracic wall or mediastinum was significantly related to overall survival in lung cancer.\cite{20} A detailed procedure for segmentation is described in electronic Supplementary Document. After segmentation, a total of 398 radiomic features were computed using SlicerRadiomics,\cite{21} an extension for 3D Slicer, from the whole tumor (3D analysis) and the maximum cross-section of the tumor (2D analysis). The analyzed radiomic features comprised 212 intra- and 186 peritumoral features, including first order statistic, shape, and texture features. A fixed bin width of 25 Hounsfield units was used for the texture analysis, which is the default value of 3D Slicer. The description of each radiometric feature is available via Pyradiomics Documentation: https://pyradiomics.readthedocs.io/en/latest/features.html. To evaluate the interobserver agreement of radiomic feature values, another radiologist with 15 years of experience segmented 50 tumors randomly selected from the study group. Figure 1 illustrates the feature extraction process.

**Unsupervised clustering analysis**

Before feature selection and model building, unsupervised hierarchical clustering analysis was conducted to identify differences in the overall expression pattern of intra- and peritumoral radiomic features depending on the presence or absence of EGFR mutations. Ward’s method based on the SCC was used as the clustering algorithm.\cite{22} To adjust the feature scale, each radiomic feature was normalized to a mean of 0 and standard deviation of 1. ComplexHeatmap package\cite{23} implemented in R software\cite{24} version 4.0.4 was used to draw the heatmap.

**Building the machine-learning model**

As some radiomic features are unstable and similar to each other, poorly-reproducible and highly-correlated features were excluded using the criteria of the intraclass correlation coefficient (ICC)\( <0.5 \) and Spearman's correlation coefficient (SCC)\( ≥0.7 \). ICC calculation was based on the ICC \( (2, 1) \).\cite{25} After feature selection, the following four Random Forest classifiers\cite{26} were built for predicting EGFR mutations: (1) a model with intratumoral radiomics alone, (2) a model combining intra- and peritumoral radiomics, (3) a model combining intra- and peritumoral radiomics and clinical features, and (4) a model with clinical features alone. The ranger R package\cite{27} implemented in R software\cite{28} was used for model building. Clinical features included age, sex, Brinkman index, and clinical stage. The Brinkman index was...
calculated as the number of cigarettes smoked per day multiplied by the number of years smoked.²⁸ Random Forest is an ensemble machine-learning algorithm consisting of many decision trees;²⁶ a decision tree is a supervised learning technique with a tree-like structure. Random Forest integrates all predictions made by decision trees into one prediction. In this study, 500 decision trees were generated in Random Forest. For each tree, training data with the same number of the entire dataset (n = 478) were obtained by sampling with replacement from the entire dataset. Because of the resampling, about one-third of the patients (n ≈ 159) were not selected for each sampling. The unselected dataset is called out-of-bag (OOB) data and was used as test data to estimate the prediction performance for unseen data in this study. Predictions based on OOB data are reportedly an unbiased estimator of the performance on future data.²⁶ The number of features to be randomly selected when splitting the patient group was set to the square root of the total number of features; this is the default setting of the ranger R package. The other hyperparameters were also set to the default value. After constructing 500 trees, each tree’s prediction was combined, and the predicted probability of EGFR mutations was determined for each patient. Furthermore, we measured feature importance according to the mean decrease in the Gini index.²⁹ A decrease in the Gini index indicates how a feature contributes to discrimination between positive and negative EGFR mutations. The larger the mean decrease in the Gini index, the greater the feature contributes to the discrimination.

Statistical analysis
Fisher’s exact test and Mann–Whitney U-test were used to assess statistical differences in categorical and continuous variables, respectively. The performance of the predictive model using OOB data was evaluated using the area under the receiver operating characteristic curve (AUC). We investigated whether adding peritumoral radiomics significantly improved the AUC using the DeLong method.³⁰ Namely, a comparison of the AUCs was a one-sided hypothesis test, in which the null hypothesis was an assertion that a model with peritumoral radiomics was not superior to a model without. The other statistical analyses were two-tailed. A calibration curve was delineated to assess the agreement between the predicted and the actual probabilities of EGFR mutations on OOB data. The goodness-of-fit of the model was determined using the Hosmer-Lemeshow test. Statistical significance was set at p < 0.05.

RESULTS
Characteristics of patients and tumors
EGFR mutations were detected in 162 (33.9%) patients, and mutations were negative in 316 (66.1%) patients. As for race, one patient was a Caucasian, and the others were East Asian. Table 1 describes the characteristics of the patients and tumors based on their EGFR status. The percentage of EGFR mutations was significantly higher in younger patients (p = 0.010), female sex (p < 0.001), lower Brinkman Index (p < 0.001), and early clinical stage (p < 0.001). All mutated tumors were adenocarcinomas except for one case of pleomorphic carcinoma. Table 2 summarizes the histological types of adenocarcinomas. EGFR-mutated adenocarcinomas were lepidic predominant, papillary, or acinar types (p < 0.001); no other histological types showed mutations.

Clustering analysis
Figure 2 illustrates the radiomic heatmap constructed by using unsupervised clustering analysis. The algorithm divided the 478 patients into two clusters with similar radiomic features. These two clusters were found to have significantly different EGFR mutation rates (cluster 1, 56.3% [112/199] vs cluster 2, 17.9% [50/279]; p < 0.001), indicating that intra- and peritumoral radiomic features were significantly associated with EGFR mutations.
Radiomic feature selection

Figure 3 shows a flow chart of the feature selection. Of the 398 radiomic features, 92 were excluded because of their low reproducibility (ICC<0.5). Among the remaining 306 features, 274 were subsequently removed due to redundancy (SCC ≥ 0.7). Consequently, 32 radiomic features with 14 intra- and 18 peritumoral features remained and were used for model building. The selected 32 features consisted of three shape features, nine first order statistics, and 20 texture features, and they are presented in Table 3. All 398 radiomic features and ICCs are listed in Supplementary Table 1.

Model performance for EGFR mutation using OOB data

The model performance for predicting EGFR mutations on OOB data is presented in Table 4. Regarding the prediction among all cases (n = 478), intratumoral radiomics showed an AUC of 0.730 (95% confidence interval [CI]: 0.682–0.777). Compared to this performance, combined intra- and peritumoral radiomics showed a significantly higher performance, with an AUC of 0.774 (95% CI: 0.730–0.817; p < 0.001). The AUC for adding clinical features to radiomics was 0.826 (95% CI: 0.788–0.864), which was significantly higher than the AUC of 0.777 (95% CI: 0.734–0.821) for clinical features alone (p = 0.005).

In terms of performance for adenocarcinoma cases only (n = 348), the results were similar. The AUC of intratumoral radiomics was 0.630 (95% CI: 0.572–0.689), and the combination of intra- and peritumoral radiomics showed a significantly increased AUC of 0.687 (95% CI, 0.632–0.743; p < 0.001). The AUC by including radiomics and clinical features was 0.744 (95% CI, 0.693–0.796).

### Table 1. Patient and tumor characteristics (n = 478)

| Variables | EGFR-mutant (n = 162) | Wild-type (n = 316) | P |
|-----------|-----------------------|--------------------|---|
| Age       |                       |                    | 0.010 |
| < 65 y    | 52 (32.1)             | 67 (21.2)          |    |
| ≥ 65 y    | 110 (67.9)            | 249 (78.8)         |    |
| Sex       |                       |                    | < 0.001 |
| Males     | 51 (31.5)             | 237 (75.0)         |    |
| Females   | 111 (68.5)            | 79 (25.0)          |    |
| Brinkman index |                   |                    | < 0.001 |
| < 400     | 122 (75.3)            | 85 (26.9)          |    |
| ≥ 400     | 40 (24.7)             | 231 (73.1)         |    |
| Clinical stage |                  |                    | < 0.001 |
| 0 or I    | 154 (95.1)            | 246 (77.8)         |    |
| II or III | 8 (4.9)               | 70 (22.2)          |    |
| Histologic type |                |                    |    |
| Adenocarcinoma (n = 348) | 161 (99.4) | 187 (59.2) |    |
| Squamous cell carcinoma (n = 96) | 0 (0) | 96 (30.4) |    |
| Other types (n = 34) | 1 (0.6) | 33 (10.4) |    |

EGFR, epidermal growth factor receptor.
Data are presented as numbers (%).

### Table 2. Adenocarcinoma types (n = 348)

| Histologic type | EGFR-mutant (n = 161) | Wild-type (n = 187) | P |
|-----------------|-----------------------|--------------------|---|
| Lepidic predominant adenocarcinoma | 51 (31.7) | 45 (24.1) |    |
| Papillary or acinar adenocarcinoma | 110 (68.3) | 92 (49.2) |    |
| Other adenocarcinomas | 0 (0) | 50 (26.7) | < 0.001 |

EGFR, epidermal growth factor receptor.
Data are presented as numbers (%).

*aAdenocarcinoma in situ, minimally invasive adenocarcinoma, or lepidic adenocarcinoma

*bSolid, invasive mucinous, enteric, micropapillary, or colloid adenocarcinoma
This performance was significantly higher than the AUC of 0.703 (95% CI: 0.648–0.758) for the clinical features alone ($p = 0.045$).

The calibration curve using OOB data is shown in Figure 4. The Hosmer-Lemeshow test indicated that the model incorporating radiomics and clinical features showed proper calibration ($p = 0.456$).

**Model performance on identical CT scanner**

As this study used three different CT scanners, the model performance within patients who were investigated using SOMATOM Force only ($n = 317$) was evaluated to determine the generality of the study results. The AUC (95% CI) using intra- and peritumoral radiomics was significantly higher than that using intratumoral radiomics alone (0.780 [0.727–0.832] vs 0.740 [0.684–0.795]; $p = 0.003$), showing similar results.

**Feature importance**

The top 10 most important features were eight radiomic features, Brinkman index, and sex (Table 5). Brinkman index and sex were ranked as the first and third most important features, respectively. Among the radiomic features, the intratumoral 90th percentile was the most important, followed by peritumoral large area high gray level emphasis (LAHGLE); both were 3D.
features. Specifically, a low 90th percentile and high LAHGLE were significantly associated with EGFR mutations \((p < 0.001\) for each). The importance of other features is presented in Supplementary Table 2. Figures 5 and 6 show representative cases with and without EGFR mutations.

**DISCUSSION**

The present study constructed a prediction model for EGFR mutations in PLC and found that a combination of intratumoral and peritumoral radiomics improved accuracy compared to intratumoral radiomics alone. It has also been demonstrated that combining radiomics and clinical features yields greater accuracy than either in isolation in the prediction of EGFR mutations. To our knowledge, few studies on oncology, including but not limited to PLC, have revealed the added value of peritumoral radiomics in predicting cancer gene mutations.

Although biopsy is the gold standard for the diagnosis of gene mutations, understanding the pretest probability using radiomics is clinically useful. For example, if the pretest probability is high,
sampling of many tissues can be considered at the time of biopsy to reduce the false-negative rate. Conversely, in cases where the pretest probability is very low, physicians could determine not to perform biopsy intended for gene tests, considering disadvantages such as the risk of complications and increase in medical costs.

Figure 4. Calibration curve of the predictive model consisted of intratumoral, peritumoral, and clinical features. The x- and y-axes correspond to the probability of EGFR mutations predicted by the model and the actual probability, respectively; the probabilities are based on the out-of-bag data. The model fit well according to the Hosmer-Lemeshow test ($p = 0.456$). The black-dashed line represents perfect calibration.

Table 4. Model performance in EGFR mutation prediction using OOB data

|                                 | AUC (95% CI) | P for difference in AUC |
|---------------------------------|-------------|------------------------|
| All cases ($n = 478$)           |             |                        |
| Intratumor                      | 0.730 (0.682–0.777) |                      |
| Intra+peritumor                 | 0.774 (0.730–0.817) | < 0.001 (vs intratumor) |
| Intra+peritumor+clinical        | 0.826 (0.788–0.864) | 0.005 (vs clinical)    |
| Clinical                        | 0.777 (0.734–0.821) |                       |
| Adenocarcinoma only ($n = 348$) |             |                        |
| Intratumor                      | 0.630 (0.572–0.689) |                      |
| Intra+peritumor                 | 0.687 (0.632–0.743) | < 0.001 (vs intratumor) |
| Intra+peritumor+clinical        | 0.744 (0.693–0.796) | 0.045 (vs clinical)    |
| Clinical                        | 0.703 (0.648–0.758) |                       |

AUC, area under the curve; CI, confidence interval; EGFR, epidermal growth factor receptor; OOB, out-of-bag.

*AUC was calculated from OOB data, namely a cohort that was not used in building the model.*
All patients in this study had resectable PLC. Although the clinical significance of EGFR mutations has not been established in preoperative patients, gene examinations for such patients may be required for planning systemic drug therapy in case of postoperative recurrence. We believe that radiomic features extracted from primary tumors can be a useful predictor of EGFR mutations in recurrent lesions. Furthermore, a recent systematic review has reported that neoadjuvant EGFR-TKI therapy may be feasible for resectable PLC and that further clinical trials are ongoing. If strong evidence supports neoadjuvant EGFR-TKI therapy for resectable PLC, our model for predicting EGFR mutation would have a more extensive clinical applicability.

The association between peritumoral radiomics and EGFR mutations in our study is a reasonable result because peritumoral radiomics encompasses various aspects of PLC. For instance, Wu et al. showed that peritumoral radiomics accurately differentiated adenocarcinoma in situ (AIS) or minimally invasive adenocarcinoma (MIA) from invasive adenocarcinoma. Regarding the relationship between histology and EGFR status in lung adenocarcinoma, EGFR mutations are mainly observed in AIS, MIA, lepidic, and papillary types; solid and invasive mucinous types rarely show EGFR mutations. This is in line with our findings, where EGFR mutations were found in either lepidic predominant, papillary, or acinar types.

Table 5. Top 10 most important features

| Rank | Feature name         | Dimension | Feature family | Extracted region | Importancea |
|------|----------------------|-----------|----------------|------------------|-------------|
| 1    | Brinkman index       | N.A.      | Clinical       | N.A.             | 19.57       |
| 2    | 90th percentile      | 3D        | First order    | Intratumor       | 13.59       |
| 3    | Sex                  | N.A.      | Clinical       | N.A.             | 11.49       |
| 4    | LAHGLE                | 3D        | GLSZM          | Peritumor        | 10.25       |
| 5    | Maximal correlation coefficient | 3D | GLCM | Intratumor | 6.53 |
| 6    | Inverse difference moment normalized | 2D | GLCM | Peritumor | 6.46 |
| 7    | Cluster shade        | 2D        | GLCM           | Intratumor       | 6.21        |
| 8    | Inverse difference moment normalized | 3D | GLCM | Peritumor | 5.52 |
| 9    | Difference variance  | 2D        | GLCM           | Peritumor        | 5.47        |
| 10   | Coarseness           | 3D        | NGTDM          | Intratumor       | 4.91        |

GLCM, gray level co-occurrence matrix; GLSZM, gray level size zone matrix; LAHGLE, large area high gray level emphasis; NGTDM, neighboring gray tone difference matrix.

aMeasured using the mean decrease in the Gini index.

Figure 5. A 72-year-old female (non-smoker) with EGFR-mutant acinar adenocarcinoma. (a) Axial CT image shows a part-solid nodule in the apex of the left lung. (b) Heatmap that captures a peritumoral feature (large area high gray level emphasis) indicates high expression. The probability of an EGFR mutation predicted by the model was as high as 88.4%.
Only a few studies have used peritumoral radiomics to predict EGFR mutations in PLCs. Choe et al. developed a model using both intra- and peritumoral radiomics to predict EGFR mutations in lung adenocarcinoma. However, the accuracy was not significantly different from that of the model consisting of intra-tumoral radiomics alone. The discrepancy between the previous results and our own may be explained by differences in the population included in the predictive model. The previous study trained the predictive model using only lung adenocarcinoma cases, whereas the present study trained by including various histological types. Consequently, our model may have more effectively learned the differences in radiomic features between EGFR-mutated and wild-type tumors.

In our study, the most important peritumoral feature for predicting EGFR mutations was LAHGLE, with high values correlating with mutations. This radiomic feature measures the extent of a large area with a high CT value and can be elevated when high-density structures, such as large pulmonary vessels, are present around tumors. In contrast, LAHGLE can decrease if low-density structures, such as honeycombing and emphysema, exist. According to previous studies, EGFR-mutated adenocarcinomas were more likely to show vascular convergence but less frequently accompanied fibrosis (related to the presence of honeycombing) and emphysema around tumors compared to wild-type adenocarcinomas. These CT features might have affected the LAHGLE values in our study. From a biological perspective, peritumoral radiomics on breast cancer is reported to be associated with the density of tumor-infiltrating lymphocytes, an immune response marker against tumors. It should be further investigated whether such association is also observed in PLC.

In our study, the intratumoral 90th percentile was the most important radiomic feature in predicting EGFR mutations; a low value was associated with EGFR mutations. A lower 90th percentile value denotes a lower density; therefore, EGFR-mutated tumors in our study may contain CT features with a low density, such as ground-glass opacity and air. This assumption is supported by previous findings, where EGFR-mutated adenocarcinomas had a higher frequency of ground-glass opacity and air-bronchogram than wild-type tumors. The intratumoral 90th percentile is also reported to be useful in the previous study by Yagi et al., who assessed the histological types of lung adenocarcinoma using texture analysis.

Our study has several limitations. First, it was retrospective and included only surgically resected cases. In addition, all patients except one were East Asian. These can contribute to selection bias. Therefore, the prediction model constructed in this study should be validated for inoperable PLCs and different race groups to clarify the robustness of our model. Second, the present study did not determine whether a distance within 3 mm from the tumor border is optimal for evaluating peritumoral features. Although the previous study also used 3 mm, ideally, peritumoral radiomic features should be extracted from a variety of ranges to compare their predictive performance. A performance comparison between peritumoral radiomics with and without the chest wall or mediastinum can also be considered. Third, the present study did not assess the performance using postcontrast HRCT. Use of both pre- and postcontrast HRCTs might increase predictive performance, and it should be investigated in future studies.

In conclusion, the combined use of intra- and peritumoral radiomic features improved the performance for EGFR mutation prediction by intra- and peritumoral CT radiomics in PLC.
prediction in PLCs. This result may help to consider the indication of EGFR gene tests and to predict candidates for EGFR TKIs.

**ACKNOWLEDGEMENTS**

The authors would like to thank Dr. Tatsuya Goto (a) and Mr. Yoshiyuki Noto (b) for patient data collection. (a)Division of Thoracic and Cardiovascular Surgery, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. (b) Division of Radiology, Niigata University Medical and Dental Hospital, Niigata, Japan.

**REFERENCES**

1. Tomizawa Y, Iijima H, Sunaga N, Sato K, Takise A, Otani Y, et al. Clinopathologic significance of the mutations of the epidermal growth factor receptor gene in patients with non-small cell lung cancer. *Clin Cancer Res* 2005; 11: 6816–22. https://doi.org/10.1158/1078-0432.CCR-05-0441
2. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res* 2004; 64: 8919–23. https://doi.org/10.1158/0008-5472.CAN-04-2818
3. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; 350: 2129–39. https://doi.org/10.1056/NEJMoa040938
4. Gillies RJ, Knahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563–77. https://doi.org/10.1148/radiol.2015151169
5. Liu G, Xu Z, Ge Y, Jiang B, Groen H, Vliegenthart R, et al. Perinodular and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening. *Sci Rep* 2020; 10(1): 10528. https://doi.org/10.1038/s41598-020-67378-8
6. Liu K, Li K, Wu T, Liang M, Zhong Y, Xu X, et al. Improving the accuracy of prognosis for clinical stage I solid lung adenocarcinoma by radiomics models covering tumor per se and peritumoral changes on CT. *Eur Radiol* 2022; 32: 2742–50. https://doi.org/10.1007/s00330-021-08366-y
7. Khorrami M, Jain P, Bera K, Alilou M, Thawani R, Patel P, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features. *Lung Cancer* 2019; 135: 1–9. https://doi.org/10.1016/j.lungcan.2019.06.020
8. Choe J, Lee SM, Kim W, Do K-H, Kim S, Choi S, et al. CT radiomics-based prediction of anaplastic lymphoma kinase and epidermal growth factor receptor mutations in lung adenocarcinoma. *Eur J Radiol* 2021; 139: 109710. https://doi.org/10.1016/j.ejrad.2021.109710
9. Yu P, Liu G, Xu Z, Ge Y, Jiang B, Groen H, Vliegenthart R, et al. Peritumoral and intratumoral radiomic features predict distant metastasis in locally advanced NSCLC. *PLoS One* 2018; 13(11): e0206108. https://doi.org/10.1371/journal.pone.0206108
10. Pérez-Morales J, Tunali I, Stringfield O, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res* 2004; 64: 8919–23. https://doi.org/10.1158/0008-5472.CAN-04-2818
11. Liu K, Li K, Wu T, Liang M, Zhong Y, Xu X, et al. Improving the accuracy of prognosis for clinical stage I solid lung adenocarcinoma by radiomics models covering tumor per se and peritumoral changes on CT. *Eur Radiol* 2022; 32: 2742–50. https://doi.org/10.1007/s00330-021-08366-y
12. Akinci D’Antonoli T, Fachioni A, Lenkovicz J, Chiappetta M, Cicchetti G, Martino A, et al. CT radiomics signature of tumor and peritumoral lung parenchyma to predict nonsmall cell lung cancer postsurgical recurrence risk. *Acad Radiol* 2020; 27: 497–507. https://doi.org/10.1016/j.acra.2019.05.019
13. Beig N, Khorrami M, Alilou M, Prasanna P, Brama N, Oorooi M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. *Radiology* 2019; 290: 783–92. https://doi.org/10.1148/radiol.2018180910
14. Wu L, Gao C, Ye J, Tao J, Wang N, Pang P, et al. The value of various peritumoral radiomic features in differentiating the invasiveness of adenocarcinoma manifesting as ground-glass nodules. *Eur Radiol* 2021; 31: 9030–37. https://doi.org/10.1007/s00330-021-07948-0
15. Khorrami M, Jain P, Bera K, Alilou M, Thawani R, Patel P, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features. *Lung Cancer* 2019; 135: 1–9. https://doi.org/10.1016/j.lungcan.2019.06.020
16. Khorrami M, Khunger M, Zagouras A, Patil P, Thawani R, Bera K, et al. Combination of peri- and intratumoral radiomic features on baseline CT scans predicts response to chemotherapy in lung adenocarcinoma. *Radiofert Intell* 2019; 1: e180012. https://doi.org/10.1148/radiol.2019180012
17. Swensen SJ, Viggiano RW, Midhun DE, Müller NL, Sherrick A, Yamashita K, et al. Lung nodule enhancement at CT: multicenter study. *Radiology* 2000; 214: 73–80. https://doi.org/10.1148/radiol.214.1.Lr00ai473
18. Nagai Y, Miyazawa H, Tanaka T, Udagawa K, Kato M, et al. Genetic heterogeneity of the epidermal growth factor receptor in non-small cell lung cancer cell lines revealed by a rapid and sensitive detection system, the peptide nucleic acid-locked nucleic acid PCR clamp. *Cancer Res* 2005; 65: 7276–82. https://doi.org/10.1158/0008-5472.CAN-05-0331
19. Pinter C, Lasso A, Fichtinger G. Polymorph segmentation representation for medical image computing. *Comput Methods Programs Biomed* 2019; 171: 19–26. https://doi.org/10.1016/j.cmpb.2019.02.011
20. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D slicer as an image computing platform for the FUNDING

This work was supported by JSPS KAKENHI grant number 20K16693.

**PATIENT CONSENT**

Written informed consent was waived by the Institutional Review Board.

**ETHICS APPROVAL**

Institutional Review Board approval was obtained.

**ETHICS APPROVAL**

Written informed consent was waived by the Institutional Review Board.

**FUNDING**

This work was supported by JSPS KAKENHI grant number 20K16693.

**PATIENT CONSENT**

Written informed consent was waived by the Institutional Review Board.

**ETHICS APPROVAL**

Institutional Review Board approval was obtained.
quantitative imaging network. Magn Reson Imaging 2012; 30: 1323–41. https://doi.org/10.1016/j.mri.2012.05.001

21. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Auscoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017; 77: e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339

22. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963; 58: 236–44. https://doi.org/10.1080/01621459.1963.10500845

23. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016; 32: 2847–49. https://doi.org/10.1093/bioinformatics/btw313

24. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org

25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016; 15: 155–63. https://doi.org/10.1016/j.jcm.2016.02.012

26. Breiman L. Random forests. Mach Learn 2001; 45: 5–32. https://doi.org/10.1023/A:1010933404324

27. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 2017; 77: 1–17. https://doi.org/10.18637/jss.v077.i01

28. Brinkman GL, Coates EO. The effect of bronchitis, smoking, and occupation on ventilation. Am Rev Respir Dis 1963; 87: 684–93. https://doi.org/10.1164/arrd.1963.87.5.684

29. Tangirala S. Evaluating the impact of gini index and information gain on classification using decision tree classifier algorithm*. IJACSA 2001; 11: 612–19. https://doi.org/10.14569/IJACSA.2020.0110277

30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44: 837–45. https://doi.org/10.2307/2531595

31. Sun L, Guo Y-J, Song J, Wang Y-R, Zhang S-L, Huang L-T, et al. Neoadjuvant EGFR-TKI therapy for EGFR- mutant NSCLC: A systematic review and pooled analysis of five prospective clinical trials. Front Oncol 2020; 10: 586596. https://doi.org/10.3389/fonc.2020.586596

32. Yoshizawa A, Sumiyoshi S, Sonobe M, Kobayashi M, Fujimoto M, Kawakami F, et al. Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: analysis of 440 Japanese patients. J Thorac Oncol 2013; 8: 52–61. https://doi.org/10.1097/JTO0b013e3182769aa8

33. Zhang H, Cai W, Wang Y, Liao M, Tian S. CT and clinical characteristics that predict risk of EGFR mutation in non-small cell lung cancer: a systematic review and meta-analysis. Int J Clin Oncol 2019; 24: 649–59. https://doi.org/10.1007/s10147-019-01403-3

34. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, et al. CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. Radiology 2016; 280: 271–80. https://doi.org/10.1148/radiol.2016151455

35. Braman N, Prasanna P, Whitney J, Singh S, Beig N, Etesami M, et al. Association of peritumoral radiomics with tumor biology and pathologic response to preoperative targeted therapy for HER2 (ERBB2)- positive breast cancer. JAMA Netw Open 2019; 2: e192561. https://doi.org/10.1001/jamanetworkopen.2019.2561

36. Yagi T, Yamazaki M, Ohashi R, Ogawa R, Ishikawa H, Yoshimura N, et al. HRCT texture analysis for pure or part-solid ground-glass nodules: distinguishability of adenocarcinoma in situ or minimally invasive adenocarcinoma from invasive adenocarcinoma. Ipn J Radiol 2018; 36: 113–21. https://doi.org/10.1007/s11604-017-0711-2