



Three-Dimensional Convolutional Neural Network-Based Prediction of Epidermal Growth Factor Receptor Expression Status in Patients With Non-Small Cell Lung Cancer

Xuemei Huang¹, Yingli Sun¹, Mingyu Tan¹, Weiling Ma¹, Pan Gao¹, Lin Qi¹, Jinjuan Lu¹, Yuling Yang¹, Kun Wang¹, Wufei Chen¹, Liang Jin¹, Kaiming Kuang², Shaofeng Duan³ and Ming Li^{1*}

¹ Department of Radiology, Huadong Hospital Affiliated With Fudan University, Shanghai, China, ² Dianei Technology, Shanghai, China, ³ Precision Health Institution, GE Healthcare, Shanghai, China

OPEN ACCESS

Edited by:

Xue Qin Yu,
The University of Sydney, Australia

Reviewed by:

Stephen Yip,
Janssen Pharmaceuticals, Inc.,
United States
Xuan Wu,
Peking University Shenzhen Hospital,
China

*Correspondence:

Ming Li
minli77@163.com

Specialty section:

This article was submitted to
Thoracic Oncology,
a section of the journal
Frontiers in Oncology

Received: 08 September 2021

Accepted: 10 January 2022

Published: 02 February 2022

Citation:

Huang X, Sun Y, Tan M, Ma W,
Gao P, Qi L, Lu J, Yang Y, Wang K,
Chen W, Jin L, Kuang K, Duan S and
Li M (2022) Three-Dimensional
Convolutional Neural Network-Based
Prediction of Epidermal Growth
Factor Receptor Expression
Status in Patients With
Non-Small Cell Lung Cancer.
Front. Oncol. 12:772770.
doi: 10.3389/fonc.2022.772770

Objectives: EGFR testing is a mandatory step before targeted therapy for non-small cell lung cancer patients. Combining some quantifiable features to establish a predictive model of EGFR expression status, break the limitations of tissue biopsy.

Materials and Methods: We retrospectively analyzed 1074 patients of non-small cell lung cancer with complete reports of *EGFR* gene testing. Then manually segmented VOI, captured the clinicopathological features, analyzed traditional radiology features, and extracted radiomic, and deep learning features. The cases were randomly divided into training and test set. We carried out feature screening; then applied the light GBM algorithm, Resnet-101 algorithm, logistic regression to develop sole models, and fused models to predict EGFR mutation conditions. The efficiency of models was evaluated by ROC and PRC curves.

Results: We successfully established Model_{clinical}, Model_{radiomic}, Model_{CNN} (based on clinical-radiology, radiomic and deep learning features respectively), Model_{radiomic+clinical} (combining clinical-radiology and radiomic features), and Model_{CNN+radiomic+clinical} (combining clinical-radiology, radiomic, and deep learning features). Among the prediction models, Model_{CNN+radiomic+clinical} showed the highest performance, followed by Model_{CNN}, and then Model_{radiomic+clinical}. All three models were able to accurately predict EGFR mutation with AUC values of 0.751, 0.738, and 0.684, respectively. There was no significant difference in the AUC values between Model_{CNN+radiomic+clinical} and Model_{CNN}. Further analysis showed that Model_{CNN+radiomic+clinical} effectively improved the efficacy of Model_{radiomic+clinical} and showed better efficacy than Model_{CNN}. The inclusion of clinical-radiology features did not effectively improve the efficacy of Model_{radiomic}.

Conclusions: Either deep learning or radiomic signature-based models can provide a fairly accurate non-invasive prediction of EGFR expression status. The model combined both features effectively enhanced the performance of radiomic models and provided

marginal enhancement to deep learning models. Collectively, fusion models offer a novel and more reliable way of providing the efficacy of currently developed prediction models, and have far-reaching potential for the optimization of noninvasive EGFR mutation status prediction methods.

Keywords: NSCLC, EGFR, tomography, radiogenomics, deep learning, machine learning

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths, with incidence and mortality rates of approximately 11.4% and 18%, respectively, and is the second-highest incidence rate in the world (1). Non-small cell lung cancer is the main pathological form and accounts for approximately 80-90% of all lung cancers (2). Targeted therapy has become one of the first-line standard treatments for non-small cell lung cancer patients; because this form of treatment can effectively improve their prognosis, prolong the PFS and OS, compared with traditional means of treatment, like chemotherapy (3–6). In patients with non-small cell lung cancer, EGFR is responsible for approximately 10-20% of all and is the most predominant driver mutations target for targeted therapy (7). As a consequence, EGFR-TKI therapy plays a pivotal role in the targeted therapy of patients with non-small cell lung cancer.

Prior to EGFR-TKI treatment, it is essential to perform EGFR genetic testing to clarify the presence of EGFR mutations. There are several methods that can be used to detect EGFR mutations, including tissue biopsy, liquid biopsy, and radiogenomics.

Histopathological biopsy has been the gold standard in terms of high sensitivity and specificity in clinical disease and genetic diagnosis. However, it still has the following restrictions: 1. High sample size threshold, requiring at least 20% of tumor cells in the sample to be detectable (8). 2. As the tumor genotype itself possesses heterogeneity (9–11), while part of the samples are taken from puncture biopsies, so there is a risk of sampling bias, which means that the gene mutation status detection result may not correspond to the authentic condition and is not representative of the whole gene expression profile of the cancer spot. 3. Because of heterogeneity of neoplastic cell genetic status, disease progression or drug resistance commonly occurs in terminal period of the disease, so that re-biopsy is necessitated to evaluate the disease and clarify if a drug resistant mutation such as T790M (12) has evolved to instruct subsequent treatment, yet the biopsy is an invasive operation with complications including pneumothorax and bleeding, and

often not feasible due to the patient's physiological issues in terminal course of the disease, thus blocking the personalized health maintenance strategy. 4. More expensive, with higher standards of material storage and instrumentation, which is not conducive to applying and promoting in certain impoverished and remote areas.

Liquid biopsy refers to the extraction of tumor gene-carrying agents from body fluids, such as Circulating tumor-derived DNA, cell-free DNA, etc., for detecting the relevant genetic alterations, and it has the merits of real-time detection and minor invasion, however, due to the existence of tumor spatial heterogeneity, it may not be capable of accurate localization or representing the true mutation level in the whole tumor; besides, in the early stage, there are often no circulating tumor cells in body fluids, and their concentration is susceptible to influence, resulting in an insufficient sample size. Presently, cell-free DNA is the only liquid biopsy marker recommended for insufficient volume of pathology biopsies or to monitor the presence of EGFR T790M mutations with disease progression or drug resistance (13, 14). Moreover, a recent study (15) indicates that the sensitivity and specificity of this technique are poor and that the practical use of this method remains controversial.

Hence enabling a holistic and comprehensive analysis of the lesion by surmounting the obstacle posed by genetic heterogeneity is now a much desirable claim.

Regarding the aforementioned downsides of tissue biopsy pathology and liquid biopsy, researchers have exploited the advancing artificial intelligence to provide a technology with promising clinical applicability - radiogenomics (16–19). It is a group of imaging biomarkers that can offset the constraints of tissue biopsies and liquid biopsies by effectively and non-invasively projecting the mutational status of genes such as EGFR and ALK *via* artificial intelligence methodology, enabling high-throughput molecular biological information, as tumor heterogeneity and genotype, which is not visible to the naked eye, and converting them into digital signals (deep learning features or radiomics features), quantifying and characterizing them to facilitate disease diagnosis as well as monitoring and guiding targeted therapy decision-making. Several researchers have reported that radiogenomics represents a promising application for EGFR gene detection. Both deep learning models (20, 21) and radiomic models (22–24) have been shown to be more precise in predicting the mutational status of EGFR. However, most studies have applied deep learning and radiomic features in an independent manner; far fewer studies have attempted to combine these two features. A previous study reported the successful creation of an EGFR mutation prediction model based on the fusion of these two

Abbreviations: NSCLC, Non-Small Cell Lung Cancer; EGFR, Epidermal Growth Factor Receptor; TKI, Tyrosine Kinase Inhibitors; PFS, Progression-Free Survival; OS, Overall Survival; CT, Computed Tomography; MM, Millimeter; WHO, World Health Organization; PCR, Polymerase Chain Reaction; DICOM, Digital Imaging Communications in Medicine; VOI, Volume Region of Interest; ICC, Intraclass Correlation Coefficient; Light GBM, Light Gradient Boosting Machine; HU, Hounsfield Unit; Res-Net, Residual Network; ROC, Receiver Operating Characteristic Curve; PRC, Precision Recall Curve; AUC, Area Under the Curve; GLCM, Gray-Level Co-occurrence Matrix; ALK, Anaplastic Lymphoma Kinase; KRAS, Kirsten Rat Sarcoma Viral Oncogene Homolog.

features (25). However, this study only included patients with solid lung adenocarcinoma. Furthermore, some of the images used were thick; this may have led to the loss of valuable features. Moreover, the EGFR mutation sites described in this previous study only contained exons 19 and 21. This is a concern because the ground-glass component within a cancer site maybe can provide more heterogeneous information than the solid component (26).

In the present study, we aimed to investigate and validate whether a prediction model incorporating deep learning features and radiomic features can improve the performance of the current mainstream models for the non-invasive prediction of EGFR mutations. To expand the application of radiomic features and deep learning features for non-invasive gene detection, we recruited a large number of patients with ground glass non-small cell lung cancers and used thin-layer images to avoid or minimize the loss of effective features.

MATERIALS AND METHODS

Figure 1 shows a schematic for how the models were constructed.

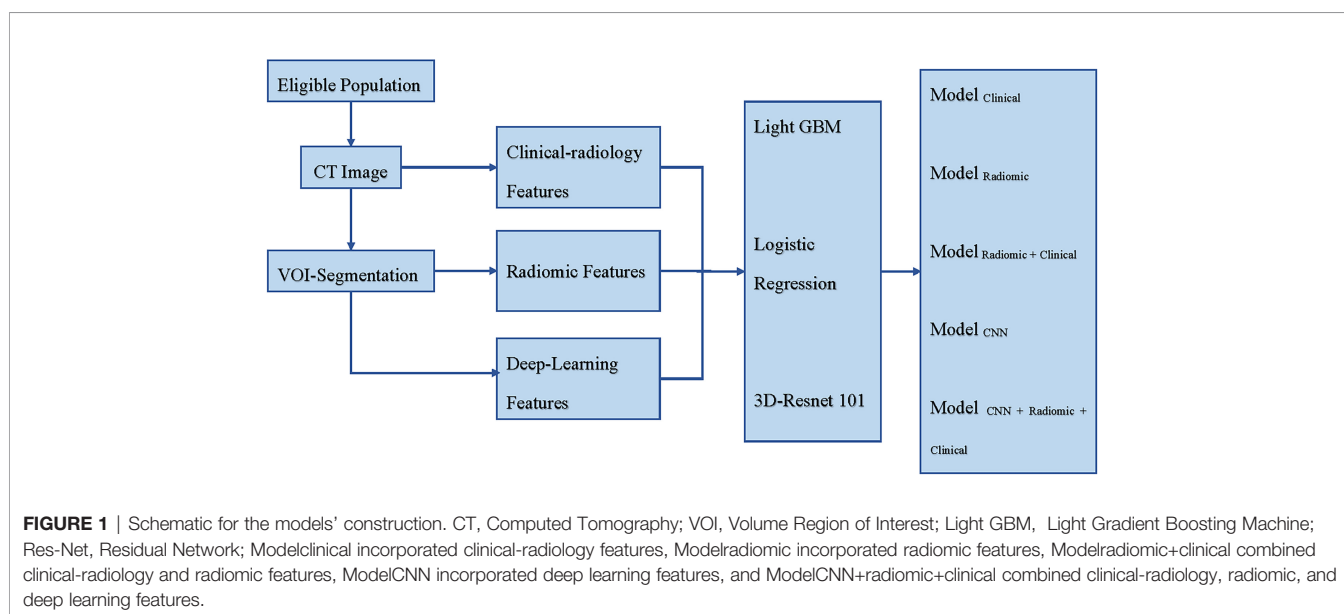
Population and Clinicopathological Data

Before initiating the research, we derived the AUC value of the radiogenomic model from that of several previous studies, which was about 0.70-0.95, and made a sample size estimation based on this data, which resulted in a predicted maximum number of 104 people needed. Later, after reminded by deep learning experts, and given the demand for large data samples for deep learning, it was decided to extend the sample on the pre-estimated sample size. We ultimately retrospectively recruited patients with pathologically confirmed primary non-small cell lung cancer between 4th June 2019 and 21st January 2021 at the Huadong

Hospital, Fudan University, Shanghai, China. All patients were screened according to strict inclusion and exclusion criteria; this process led to the inclusion of 1074 eligible patients. The inclusion criteria were as follows: (1) detailed EGFR gene test reports were available, (2) the interval between chest CT examination and surgery was within 1 month, and (3) pathological samples were obtained from surgically resected specimens. The exclusion criteria were as follows: (1) image layer thickness greater than 1.5 mm, (2) images with severe motion artifacts or conditions such as pleural effusion or obstructive pneumonia that may affect detailed observation, (3) preoperative history of tumors or a history of lung surgery, and (4) an inability to convert image format or extract features for unknown reasons. For each patient, we collated a complete range of clinicopathological data, including age, gender, smoking history, invasive degree, and EGFR mutation status. The basic principle of the training/test split is to maintain a general fraction of positive samples in each subset. We used the `train_test_split` function in Scitkit-learn 0.24.2 to perform a random selection of training/test data while maintaining roughly the same proportion of positives/negatives in both subsets, and to guarantee reproducibility, we kept the seed of the random number generator fixed at 42, which is a prevalent alternative among deep learning researchers. All cases were randomly divided into a training set (770 cases) and a test set (304 cases).

CT Instrument and Parameters

All patients were scanned with a GE Discovery CT750HD or LightSpeed VCT or Somatom Sensation 16 CT system, operating with the following parameters: tube voltage: 120 kV; tube current: 200 mA; reconstruction algorithm: STND/medium sharp; and layer thickness: 1.00/1.25/1.5 mm. Three apparatus distribution for Discovery: VCT: Somatom (training set-340:184:246; test set-135:83:86) The scan phase was set to the



deep inspiratory phase and the patient was scanned in the supine position. Images were acquired in the DICOM format. Further details of the parameters used for CT are shown in **Supplementary Table 1**.

Histopathology and the Diagnosis of EGFR Status

The histopathological type of non-small cell lung cancer was identified by our diagnostic pathologists for secondary diagnosis using the 2011 international and multidisciplinary classification guidelines proposed by the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society (27) and the World Health Organization (WHO) 2015 guidelines for lung cancer classification (28). The mutation status of EGFR exons 18, 19, 20 and 21 (which are associated with drug targets) was detected using a real-time fluorescent PCR-based amplification refractory mutation system and a human EGFR gene mutation real-time reverse transcription-polymerase chain reaction diagnostic kit (AmoyDx, Xiamen, China).

VOI Segmentation and Radiology Features

First, the pixels in the raw DICOM images were uniformly transformed to a layer thickness of 1 mm. Then, the VOI of the cancer was manually segmented by a junior diagnostician (Reader 1) using the open-source software 3D-slicer (<https://www.slicer.org/>) ensuring that large blood vessels and fibrous connective tissue was avoided during contouring. A secondary manual correction was performed by a senior physician (Reader 2). Another senior diagnostician (Reader 3) analyzed and recorded the CT radiology features of the tumor while remaining blinded to the EGFR mutation status and pathological subtypes. Reader 3 recorded a range of data, including location, cancer density, border, vacuole sign, air bronchogram sign, spiculation sign, lobulation sign, halo sign, vascular alteration, pleural indentation, and umbilicated indentation. In case of disagreement, a second evaluation was performed by another senior diagnostician (Reader 4); the results were recorded after discussion and agreement. All images were observed with a window position of -500 HU and a window width of 1500 HU. In the following features description, for the sake of brevity, we merge the radiology features with the clinical features, and use the description of the clinical features uniformly.

Analysis of Radiomic Features

The outlined VOIs were placed into Pyradiomics (29) (version 3.0 software) to extract radiomic features. Pyradiomics is an open-source python package for extracting radiomic features from medical imaging.

Reproducibility Analysis

To assess the reproducibility and stability of the radiomic features, 60 patients were randomly selected by the diagnostician (Reader 1) for secondary manual segmentation of the tumor VOI after one month. The radiomic features were

extracted and subjected to ICC analysis; features with an ICC index ≥ 0.95 were selected for subsequent model construction.

Clinical and Radiomic Models

To further identify redundant features and improve the performance of the radiomic model, we re-screened the initial radiomic features by considering mutual information between each feature and the mutation status of the EGFR gene. The mutual information between two random variables is a non-negative value that measures the dependence between the two variables (30). This function relies on a non-parametric approach based on entropy estimation from K-nearest neighbor distances and can be used for the univariate selection of features. Ultimately, we filtered the top 10% of features with the highest mutual information in the training set to develop the model. Then, we retained the same 10% of features in the test set to evaluate model performance. Based on the screened radiomic features and clinical features, we established Model_{radiomic} and a fusion model (Model_{radiomic+clinical}) using the Light GBM algorithm (31). To avoid overfitting, during model construction, we adjusted several hyperparameters, including learning rate, data down-sampling ratio, feature down-sampling ratio, and L1/L2 regularization strength. The learning rate was tuned before the steady convergence of the training and validation losses of the model was observed. Intensity of overfitting prevention enhances when we decrease the data down-sampling rate, feature down-sampling rate, or augment the L1/L2 regularization strength.

Deep Learning Model

Both the original CT images and the mask of the VOI were resampled to a space-occupying 1 mm \times 1 mm \times 1 mm. Next, we counted the spanning distribution of the cancer in three dimensions, and selected 64 mm \times 64 mm \times 64 mm as the input size for deep learning to ensure that the cropped input size could cover the extent of all lung nodules. The HU values of this patch were processed using the clip of the lung window [(-1000, 400)] and subjected to the minimum-maximum normalization process. Next, the resultant data were imported into the Ampyx 3D ResNet101 network to facilitate the creation of Model_{CNN}, a model that featured only deep learning features. 3D ResNet101 (32) is a well-characterized and broad applicable neural network in the field of deep learning, and remains considered as a strong comparative baseline in computational vision research. Compared to its successor, its network is relatively simplistic, which further alleviates overfitting and thus enables a more robust model ultimately. The model was optimized with AdamW (33) with a maximum learning rate of 0.001. We also used a cosine annealing schedule (34) to gradually reduce the model to 10^{-6} within 500 epochs. To further suppress overfitting and enhance the robustness of the model, we performed data augmentation using random rotation, random flip, and mix-up (35) with an α of 0.2. Since the objective of this study was not to innovate new neural network structures, the hyperparameters of this ResNet101 model were adjusted following the configuration given in the Torch Vision Python package.

Fusion of Clinical-Radiomic-Deep Learning Features Model

Since the deep learning features and clinical/radiomic features are totally different in terms of both data distribution and expressed meaning, and the number of filtered clinical/radiomic features is larger than that of deep learning features, the weight of clinical/radiomic features tends to be greater if the features are simply combined, and the model performance is poor. Therefore, we finally opted to model the prediction probability of $\text{Model}_{\text{CNN}}$ and that of $\text{Model}_{\text{radiomic+clinical}}$ and constructed a metamodel $\text{Model}_{\text{CNN+radiomic+clinical}}$ using logistic regression. Essentially, we perform 5-fold cross-validation on the $\text{Model}_{\text{CNN}}$ and the $\text{Model}_{\text{radiomic+clinical}}$ respectively in the training set, and build a logistic regression $\text{Model}_{\text{CNN+radiomic+clinical}}$ by weighting the probabilities calculated from the two models.

Model Evaluation

Next, the ROC curve, AUC value, and PRC curve were used to evaluate the predictive performance of each model. To verify whether the fusion model performs better than the sole model and whether the improvement in model performance is statistically significant, the De-long test is applied to compare the performance variation of each model.

Statistical Analysis

This research was carried out with Python (version 3.8.10). Modeling of radiomics features, clinical features, and the concatenation of both was done using Light GBM (version3.2.1). CNN experiments were conducted using PyTorch (version1.8.1). The logistic regression model fusing clinical, radiomic, and deep learning features were provided by Scikit-Learn (version0.24.2.). DeLong tests were done in MedCalc (version20.0009). The sample size was calculated in PASS 15 (Power: 0.90; Alpha: 0.05; AUC1:0.7; Two-Sided). Univariate analysis and multivariate logistic analysis using SPSS (version23.0). The normality distribution of the continuous variables was verified with the Kolmogorov-Smirnov test ($P < 0.001$). Continuous variables were analyzed using Mann-Whitney U test. Categorical variables were analyzed using chi-square tests or Fisher's exact test. p-values less than 0.05 were considered statistically significant.

RESULTS

A total of 1074 eligible non-small cell lung cancer cases were enrolled in this study, including 527 wild-type EGFR cases and 547 EGFR mutant cases; there were 443 males and 631 females. Analysis of between-group discrepancy showed that there was no significant difference in the clinical-radiology characteristics when compared between the training and test sets, as detailed in **Supplementary Table 2**. The distribution of the clinical-radiology characteristics of EGFR mutant-type and wild-type cases within the training set is shown in **Table 1**. Screening of the training set revealed that six items (gender, age, invasive degree, cancer density, vacuole sign, and smoking history) were all

independent predictors for EGFR mutation. Detailed statistics of clinical characteristics are shown in **Table 2**. In contrast, location, border, air bronchogram sign, spiculation sign, lobulation sign, halo sign, vascular alteration, pleural indentation, and umbilicated indentation, could not specifically identify EGFR mutation. For each case, 1218 radiomic features were extracted from the VOI; ICC analysis yielded a mean correlation coefficient of 0.96 ± 0.07 . Subsequently, 243 radiomic features with coefficients < 0.95 were excluded, and the top 10% of the radiomic features with the highest mutual information were identified, and used to build the model. Finally, six clinical features and 108 radiomic features were used to build the predictive models. The top 20 radiomic features selected are shown in **Supplementary Table 3**.

Next, we successfully built five prediction models: $\text{Model}_{\text{CNN+radiomic+clinical}}$, $\text{Model}_{\text{CNN}}$, $\text{Model}_{\text{radiomic+clinical}}$, $\text{Model}_{\text{radiomic}}$, and $\text{Model}_{\text{clinical}}$. The performance of each model was verified in the test set, as shown in **Figure 2**. In the test set, the most effective prediction model, as based on the ROC curve, was $\text{Model}_{\text{CNN+radiomic+clinical}}$ with an AUC of 0.751; this was followed by $\text{Model}_{\text{CNN}}$, $\text{Model}_{\text{radiomic+clinical}}$, and finally $\text{Model}_{\text{clinical}}$. Our analysis showed that deep learning models and radiomic models both can predict EGFR mutations with the best levels of accuracy. $\text{Model}_{\text{CNN+radiomic+clinical}}$ which featured both deep learning and radiomic features, showed more effective improvement than the mainstream radiomic models ($\text{Model}_{\text{radiomic+clinical}}$ and $\text{Model}_{\text{radiomic}}$), with p-values of 0.0067 and 0.0063, respectively. Although the Delong Test revealed that the difference in efficacy between the two models was not statistically significant, detailed analysis of the ROC and PRC curves showed that the fusion model ($\text{Model}_{\text{CNN+radiomic+clinical}}$) was slightly more effective than the deep learning model ($\text{Model}_{\text{CNN}}$). The Delong Test also showed that the difference in efficacy between $\text{Model}_{\text{radiomic+clinical}}$ and $\text{Model}_{\text{radiomic}}$ was also not statistically significant, and that the addition of clinical information did not enhance the efficacy of $\text{Model}_{\text{radiomic}}$ ($p = 0.876$).

DISCUSSION

In this study, we developed a fusion model for predicting EGFR mutation levels in 1074 patients with non-small cell lung cancer by analyzing the clinical, radiology, radiomic, and deep learning features. The value of the combined model ($\text{Model}_{\text{CNN+radiomic+clinical}}$) was more efficient than models based on radiomic or deep learning features alone, particularly those based on radiomic features. The general objectives of this study were to investigate the feasibility of improving the efficacy of prevalent models to date (predictive models based on radiomic or deep learning features alone) and to provide a new approach for constructing models for non-invasive detection of EGFR mutations, and there may be a promise for future extensions to develop models for predicting other genotypes or other tasks.

Tumor heterogeneity (36–38) is the leading driver of drug resistance and disease progression in the post-EGFR-TKI treatment course, and the underlying factor that liquid biopsy and puncture pathology may not reflect the overall truly mutated

TABLE 1 | The distribution of clinical-radiology features for EGFR mutant and wild type cases in the training set.

| Characteristics | EGFR wild | EGFR mutation | <i>p</i> -value |
|---------------------------------|-------------|---------------|-------------------|
| Gender | | | 0.004* |
| Male | 176 (46.6) | 142 (36.2) | |
| Female | 202 (53.4) | 250 (63.8) | |
| Age | 57.5 (18.0) | 60.0 (16.0) | 0.004* |
| Invasive Degree | | | <0.001* |
| Non-invasive | 64 (16.9) | 27 (6.9) | |
| Micro-invasive | 158 (41.8) | 131 (33.4) | |
| Invasive | 156 (41.3) | 234 (59.7) | |
| Location | | | 0.248 |
| RUL | 137 (36.2) | 145 (37.0) | |
| RML | 23 (6.1) | 39 (9.9) | |
| RLL | 64 (16.9) | 59 (15.1) | |
| LUL | 98 (25.9) | 103 (26.3) | |
| LLL | 56 (14.8) | 46 (11.7) | |
| Cancer density | | | <0.001* |
| Pure GGO | 72 (19.0) | 29 (7.4) | |
| Mixed GGO | 221 (58.5) | 313 (79.8) | |
| Solid | 85 (22.5) | 50 (12.8) | |
| Border | | | 0.121 |
| Well-define | 271 (71.7) | 254 (64.8) | |
| Less-define | 61 (16.1) | 79 (20.2) | |
| Ill-define | 46 (12.2) | 59 (15.1) | |
| Vacuolation | | | <0.001* |
| Present | 112 (29.6) | 202 (51.5) | |
| Absent | 266 (70.4) | 190 (48.5) | |
| Air Bronchogram | | | <0.001* |
| Present | 120 (31.7) | 194 (49.5) | |
| Absent | 258 (68.3) | 198 (50.5) | |
| Spiculation | | | 0.025* |
| Short | 77 (20.4) | 99 (25.3) | |
| Deep | 26 (6.9) | 32 (8.2) | |
| Mixed | 60 (15.9) | 81 (20.7) | |
| Absent | 215 (56.9) | 180 (45.9) | |
| Lobulation | | | 0.133 |
| Shallow | 125 (33.1) | 114 (29.1) | |
| Deep | 4 (1.1) | 11 (2.8) | |
| Mixed | 248 (65.6) | 263 (67.1) | |
| Absent | 1 (0.3) | 4 (1.0) | |
| Halo | | | 0.006* |
| Present | 46 (12.2) | 76 (19.4) | |
| Absent | 332 (87.8) | 316 (80.6) | |
| Vascular- Alteration | | | 0.675 |
| Present | 191 (50.5) | 204 (52.0) | |
| Absent | 187 (49.5) | 188 (48.0) | |
| Pleural- Indentation | | | <0.001* |
| Present | 133 (35.2) | 187 (47.7) | |
| Absent | 245 (64.8) | 205 (52.3) | |
| Umbilicated- Indentation | | | 0.001* |
| Present | 29 (7.7) | 59 (15.1) | |
| Absent | 349 (92.3) | 333 (84.9) | |
| Smoke History | | | 0.020* |
| Yes | 172 (45.5) | 146 (37.2) | |
| No | 206 (54.5) | 246 (62.8) | |

RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; GGO, ground glass opacity; Categorical variables (e.g. gender) are expressed by a number (percentage), continuous variables (e.g. age) are expressed by the Median (interquartile range). * $p < 0.05$ (significant), *P*-values taken with three decimal places equal to 0.000 are expressed as <0.001.

The bolded values in the left column refer to the clinical-radiological features included in the statistical analysis of this study, and the bolded values in the right column refer to the *P* values, with *P* less than 0.05 as the criterion to evaluate whether they are statistically significant and whether they are included in the subsequent statistical sub-analysis. The data are bolded for the purpose of making them more prominent and clear only.

status of the lesion in the process of disease genetic identification, therapeutic efficacy monitoring and follow-up.

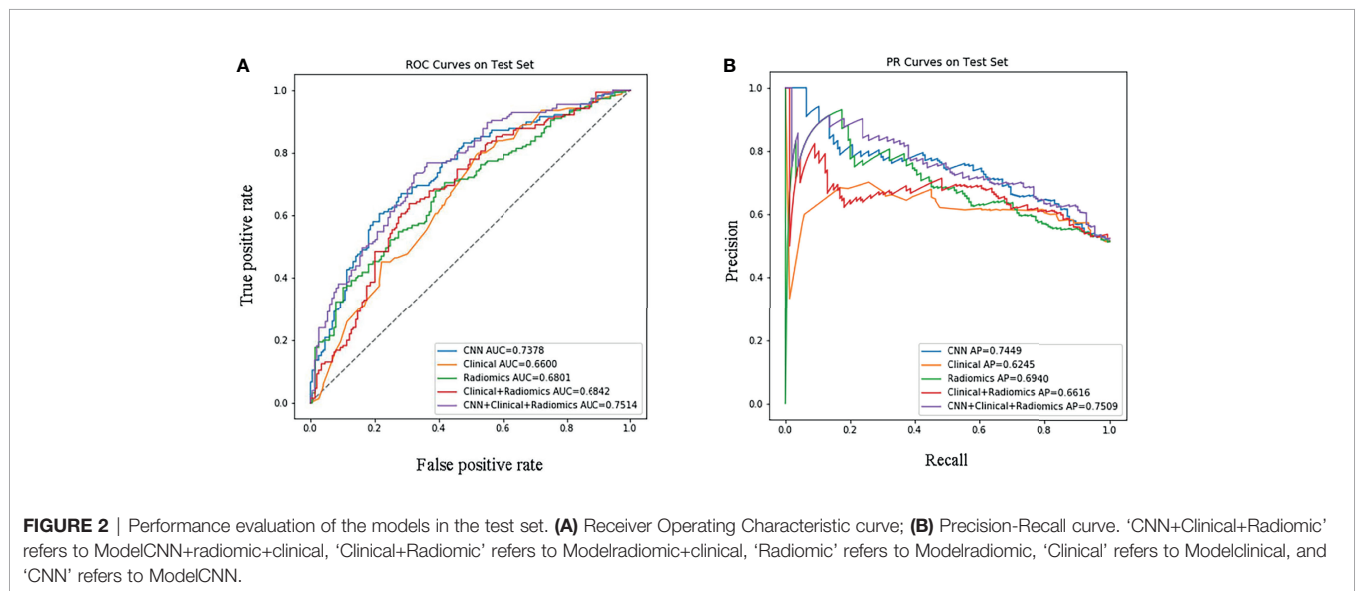
However, Radiogenomics can effectively discern the heterogeneous patterns within tumors through artificial

intelligence and mathematical statistics, bridging the limitations of pathological biopsies and liquid biopsies and assisting clinicians in conducting more precise clinical decisions. For remote and impoverished area and countries,

TABLE 2 | Statistical analysis outcome of clinical-radiology characteristics.

| Selected Features | Univariate Analysis | | Multivariate Analysis | |
|---------------------------------|---------------------|--------|------------------------|------------------|
| | Z or χ^2 | P | Regression coefficient | P |
| Gender | 8.481 | 0.004 | -0.649 | <0.001 |
| Age | -2.826 | 0.004 | 0.015 | 0.037 |
| Invasive Degree | 32.923 | <0.001 | -1.158 | <0.001 |
| Cancer density | 42.991 | <0.001 | 1.510 | <0.001 |
| Vacuolation | 38.221 | <0.001 | 0.571 | 0.001 |
| Air-Bronchogram | 25.088 | <0.001 | 0.251 | 0.165 |
| Spiculation | 9.348 | 0.025 | 0.313 | 0.253 |
| Halo | 7.520 | 0.006 | -0.506 | 0.051 |
| Pleural- Indentation | 12.418 | <0.001 | 0.145 | 0.493 |
| Umbilicated- Indentation | 10.352 | 0.001 | 0.481 | 0.093 |
| Smoke History | 5.413 | 0.020 | -0.335 | 0.038 |

Univariate Analysis: Continuous variables were analyzed using Mann-Whitney U test. Categorical variables were analyzed using chi-square tests or Fisher's exact test. Features with bolded numbers of the P-value column are independent predictors.



this is an inexpensive, low-cost and efficient genetic diagnostic weapon if the radiogenomic model can be brought to clinical practice successfully by future.

The results of this study confirmed the reliability of radiomic and deep learning models for the non-invasive prediction of EGFR mutation status in lung adenocarcinoma with a high degree of accuracy. In lung adenocarcinoma patients, two previous studies (39, 40) combined both radiomic and clinical features to successfully build a radiomic-clinical model that could efficiently identify EGFR mutant phenotypes from wild types with good AUCs of 0.779 and 0.823. However, two other studies (41, 42) also successfully built a combined radiomic-clinical prediction model but also found that a deep learning feature-based model could also predict EGFR gene mutation status in patients with lung adenocarcinoma in a more accurate manner, achieving AUCs of 0.810 and 0.758. These previous findings are consistent with the results of our current study. However, our present differs from these previous studies in that they predominantly applied radiomic and deep learning features separately to build

radiomic-clinical models or deep learning models. In this study, we innovatively developed a fusion prediction model to diagnose EGFR mutations in patients with non-small cell lung cancer by fusing the most widely accepted clinicopathological, radiology, and radiomic features with deep learning features. A previous study published findings for a fusion model that were similar to our present results; the efficacy of this previous fusion model was also more efficient than the radiomic model (AUC: 0.831 vs 0.758) (25). Comparing to this study, which enrolled only solid lung adenocarcinoma cases, had incomplete coverage of the mutant site, and used thick layers of images, our study also included a significant number of ground glass type non-small cell lung cancer cases and new radiology features. All of the images used in the present study had a layer thickness of <1.5 mm, thus making our models more realistic to the actual clinical scenario, thus providing more applicable data that could support the wider use of these models clinically.

Our current findings confirm the concept of fusing multiple features to build prediction models to enhance the efficacy of

individual models. We prove that this strategy is feasible and may be applied to the prediction of other genetic targets in the future, and even to other fields, including the identification of benign and malignant nodes, prediction of the degree of infiltration, as well as the prognosis of survival analysis.

Both Clinicopathology features that gender and smoking history, degree of invasion, and morphology features like cancer density and the vacuole sign, were independent predictors for the EGFR mutant phenotype. The present study reconfirmed that EGFR mutant phenotype is more prevalent in women and non-smoking patients (43–45). In addition, the tumor invasion degree and density are highly associated with the EGFR mutation status. The higher the degree of tumor infiltration and density, the more likely the mutation of EGFR will occur. A greater degree of invasion indicates more heterogeneous cells, faster gene duplication and an increased mutation frequency. This is in line with prior research (46) study 1 where the mutation frequency of EGFR was observed to be much larger and more distinct in IAC, than in MIA, AIS, and AAH. Compared to pure ground glass nodules, mixed ground glass nodules and solid nodules with greater density had significantly better EGFR mutation rates, which also is aligned with previous studies (46, 47) posting that the solid component is remarkably sensitive for diagnosing invasiveness and has a superior EGFR mutation profile. Both vacuole sign and age were correlated with EGFR mutation condition, yet unfortunately, this discovery was not in accordance with the results of earlier studies (48–50), probably because our research center specializes in geriatrics, so the population enrolled is mostly elderly, so there might be a sample error, while the studies correlating vacuole sign and EGFR are fewer, both of which have to be further verified by subsequent research.

Two previous studies incorporated two EGFR-related predictors, gender and smoking history, into the construction of a fused clinical-radiomic model; however, the efficacy of the final separate radiomic model was not improved (51, 52). We also found that several radiology features were not significantly correlated with EGFR mutations, including air bronchogram sign, spiculation sign, and lobulation sign. The involvement of relevant features in model construction did not effectively augment the efficacy of the radiomic model. These highly subjective and time-consuming features should be considered carefully in future studies; deletion of these features may help to streamline the development procedures of radiogenomic predictive models.

Some limitations need to be considered. First, this was a retrospective study. Firstly, EGFR frequently merges with tumor suppressor genes mutations (53), like TP53 (incidence >5%), but in the clinical setting, tumor suppressor genes testing is not routinely conducted, thus the genetic data in this study only contains EGFR synapses, and there is no investigation yet to elucidate whether the effect of the remaining co-alteration mutations upon the radiogenomic model, so more information should be collected on combined mutations for rigorous prospective trials in the future. Second, EGFR mutation prevalence is varying across ethnics, such that it is generally of a higher rate for the Asian population than that of the American and European ones (54), hence the model may be

more generalizable to Asia; also, there are large regional diversity in lifestyle practices, which may sometimes change the structuring composition of the model, such as clinical features smoking history. This is why in coming future, multi-center, multi-ethnic studies are expected to validate the robustness and generalization power of radiogenomics models. And lastly: in this study, a time-consuming manual segmentation pattern was implemented; the future semi-automatic or fully automatic segmentation mode with deep learning algorithms should be applied to streamline the whole process.

CONCLUSION

Both radiomic models and deep learning models can predict EGFR gene mutation status relatively efficiently and non-invasively. By integrating radiomics and deep learning features, it is possible to build prediction models that can significantly upgrade the performance of the basic radiomic models and help to improve the performance of deep learning models. Models featuring deep learning techniques have the potential for broader application in the non-invasive diagnosis of lung cancer genes mutation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

This study involved human participants and was reviewed and approved by the Ethics Committee of the Huadong Hospital Affiliated with Fudan University (number:2019K134). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ML and YS conceived the concept of this study. YY and KW collected the data. PG, MT, WM, LQ, JL, WC, LJ, KK, and SD analyzed the data. XH and YS drafted and revised the manuscript. All authors reviewed the manuscript, and ML made corrections to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Nature Science Foundation of China (Reference: 61976238; ML), and the

Shanghai “Rising Stars of Medical Talent” and Youth Development Program “Outstanding Youth Medical Talents” (Reference: SHWSRS(2021)_99), and Scientific Research Program of Shanghai Science and Technology Commission (Reference: 20Y11902900).

REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, et al. Non–Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2017) 15(4):504–35. doi: 10.6004/jncn.2017.0050
- Ramalingam SS, Yang JC-H, Lee CK, Kurata T, Kim D-W, John T, et al. Osimertinib as First-Line Treatment of EGFR Mutation–Positive Advanced Non–Small-Cell Lung Cancer. *J Clin Oncol* (2018) 36(9):841–9. doi: 10.1200/jco.2017.74.7576
- Yoshida K, Yamada Y. Erlotinib Alone or With Bevacizumab as First-Line Therapy in Patients With Advanced non-Squamous non-Small-Cell Lung Cancer Harboring EGFR Mutations (JO25567): An Open-Label, Randomized, Multicenter, Phase II Study. *Transl Lung Cancer Res* (2015) 4(3):217–9. doi: 10.3978/j.issn.2218-6751.2015.03.04
- Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib Versus Standard Chemotherapy as First-Line Treatment for European Patients With Advanced EGFR Mutation-Positive Non-Small-Cell Lung Cancer (EURTAC): A Multicentre, Open-Label, Randomised Phase 3 Trial. *Lancet Oncol* (2012) 13(3):239–46. doi: 10.1016/s1470-2045(11)70393-x
- Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or Chemotherapy for Non-Small-Cell Lung Cancer With Mutated EGFR. *N Engl J Med* (2010) 362(25):2380–8. doi: 10.1056/NEJMoa0909530
- Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and Biological Features Associated With Epidermal Growth Factor Receptor Gene Mutations in Lung Cancers. *J Natl Cancer Inst* (2005) 97(5):339–46. doi: 10.1093/jnci/dji055
- Kalemkerian GP, Narula N, Kennedy EB, Biermann WA, Donington J, Leigh NB, et al. Molecular Testing Guideline for the Selection of Patients With Lung Cancer for Treatment With Targeted Tyrosine Kinase Inhibitors: American Society of Clinical Oncology Endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology Clinical Practice Guideline Update. *J Clin Oncol* (2018) 36(9):911–9. doi: 10.1200/jco.2017.76.7293
- Skoulidis F, Heymach JV. Co-Occurring Genomic Alterations in Non-Small-Cell Lung Cancer Biology and Therapy. *Nat Rev Cancer* (2019) 19(9):495–509. doi: 10.1038/s41568-019-0179-8
- Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* (2012) 366(10):883–92. doi: 10.1056/NEJMoa1113205
- Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* (2017) 376(22):2109–21. doi: 10.1056/NEJMoa1616288
- Imyanitov EN, Iyevleva AG, Levchenko EV. Molecular Testing and Targeted Therapy for non-Small Cell Lung Cancer: Current Status and Perspectives. *Crit Rev Oncol/Hematol* (2021) 157:103194. doi: 10.1016/j.critrevonc.2020.103194
- Maheswaran S, Sequist LV, Nagrath S, Ullus L, Brannigan B, Collura CV, et al. Detection of Mutations in EGFR in Circulating Lung-Cancer Cells. *N Engl J Med* (2008) 359(4):366–77. doi: 10.1056/NEJMoa0800668
- Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating Liquid Biopsies Into the Management of Cancer. *Nat Rev Clin Oncol* (2017) 14(9):531–48. doi: 10.1038/nrclinonc.2017.14
- Marquette CH, Boutros J, Benzaquen J, Ferreira M, Pastre J, Pison C, et al. Circulating Tumour Cells as a Potential Biomarker for Lung Cancer Screening: A Prospective Cohort Study. *Lancet Respir Med* (2020) 8(7):709–16. doi: 10.1016/s2213-2600(20)30081-3
- Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, et al. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships Between Molecular and Imaging Phenotypes With Prognostic Implications. *Radiology* (2018) 286(1):307–15. doi: 10.1148/radiol.2017161845
- Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. Quantitative Imaging of Cancer in the Postgenomic Era: Radio(Geno)Mics, Deep Learning, and Habitats. *Cancer* (2018) 124(24):4633–49. doi: 10.1002/cncr.31630
- Iwatate Y, Hoshino I, Yokota H, Ishige F, Itami M, Mori Y, et al. Radiogenomics for Predicting P53 Status, PD-L1 Expression, and Prognosis With Machine Learning in Pancreatic Cancer. *Br J Cancer* (2020) 123(8):1253–61. doi: 10.1038/s41416-020-0997-1
- Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* (2017) 77(14):3922–30. doi: 10.1158/0008-5472.Can-17-0122
- Yin G, Wang Z, Song Y, Li X, Chen Y, Zhu L, et al. Prediction of EGFR Mutation Status Based on (18)F-FDG PET/CT Imaging Using Deep Learning-Based Model in Lung Adenocarcinoma. *Front Oncol* (2021) 11:709137. doi: 10.3389/fonc.2021.709137
- Qin R, Wang Z, Qiao K, Hai J, Jiang L, Chen J, et al. Multi-Type Interdependent Feature Analysis Based on Hybrid Neural Networks for Computer-Aided Diagnosis of Epidermal Growth Factor Receptor Mutations. *IEEE Access* (2020) 8:38517–27. doi: 10.1109/access.2020.2971281
- Rossi G, Barabino E, Fedeli A, Ficarra G, Coco S, Russo A, et al. Radiomic Detection of EGFR Mutations in NSCLC. *Cancer Res* (2021) 81(3):724–31. doi: 10.1158/0008-5472.CAN-20-0999
- Zhang L, Chen B, Liu X, Song J, Fang M, Hu C, et al. Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in non-Small Cell Lung Cancer. *Transl Oncol* (2018) 11(1):94–101. doi: 10.1016/j.tranon.2017.10.012
- Lu X, Li M, Zhang H, Hua S, Meng F, Yang H, et al. A Novel Radiomic Nomogram for Predicting Epidermal Growth Factor Receptor Mutation in Peripheral Lung Adenocarcinoma. *Phys Med Biol* (2020) 65(5):055012. doi: 10.1088/1361-6560/ab6f98
- Li XY, Xiong JF, Jia TY, Shen TL, Hou RP, Zhao J, et al. Detection of Epithelial Growth Factor Receptor (EGFR) Mutations on CT Images of Patients With Lung Adenocarcinoma Using Radiomics and/or Multi-Level Residual Convolutionary Neural Networks. *J Thorac Dis* (2018) 10(12):6624–35. doi: 10.21037/jtd.2018.11.03
- Wu G, Woodruff HC, Shen J, Refaee T, Sanduleanu S, Ibrahim A, et al. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study. *Radiology* (2020) 297(2):451–8. doi: 10.1148/radiol.2020192431
- Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: International Multidisciplinary Classification of Lung Adenocarcinoma: Executive Summary. *Proc Am Thorac Soc* (2011) 8(5):381–5. doi: 10.1513/pats.201107-042ST
- Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/JTO.0000000000000630
- van Griethuysen JMM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.772770/full#supplementary-material>

- Phenotype. *Cancer Res* (2017) 77(21):e104–e7. doi: 10.1158/0008-5472.CAN-17-0339
30. Kraskov A, Stgbauer H, Grassberger PJP. Estimating Mutual Information. *Phys Rev E* (2004) 69(6 Pt 2):066138. doi: 10.1103/PhysRevE.69.066138
 31. Bentéjac C, Csörgő A, Martínez-Muñoz G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif Intell Rev* (2020) 54(3):1937–67. doi: 10.1007/s10462-020-09896-5
 32. Kaiming He XZ, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proc Comput Vis Pattern Recognit (CVPR)* (2018) 19:770–8. doi: 10.1109/CVPR.2016.90
 33. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. *Proc ICLR* (2019), 1–15.
 34. Loshchilov I, Hutter F. Sgdr: Stochastic Gradient Descent With Warm Restarts. *ProcICLR* (2017), 1–16.
 35. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. Mixup: Beyond Empirical Risk Minimization. *Proc ICLR* (2018), 1760–72. Vancouver, BC, Canada.
 36. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* (2017) 168(4):613–28. doi: 10.1016/j.cell.2017.01.018
 37. Lim ZF, Ma PC. Emerging Insights of Tumor Heterogeneity and Drug Resistance Mechanisms in Lung Cancer Targeted Therapy. *J Hematol Oncol* (2019) 12(1):134. doi: 10.1186/s13045-019-0818-2
 38. Shibue T, Weinberg RA. EMT, Cscs, and Drug Resistance: The Mechanistic Link and Clinical Implications. *Nat Rev Clin Oncol* (2017) 14(10):611–29. doi: 10.1038/nrclinonc.2017.44
 39. Yang X, Dong X, Wang J, Li W, Gu Z, Gao D, et al. Computed Tomography-Based Radiomics Signature: A Potential Indicator of Epidermal Growth Factor Receptor Mutation in Pulmonary Adenocarcinoma Appearing as a Subsolid Nodule. *Oncologist* (2019) 24(11):e1156–e64. doi: 10.1634/theoncologist.2018-0706
 40. Zhang B, Qi S, Pan X, Li C, Yao Y, Qian W, et al. Deep CNN Model Using CT Radiomics Feature Mapping Recognizes EGFR Gene Mutation Status of Lung Adenocarcinoma. *Front Oncol* (2020) 10:598721. doi: 10.3389/fonc.2020.598721
 41. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR Mutation Status in Lung Adenocarcinoma on Computed Tomography Image Using Deep Learning. *Eur Respir J* (2019) 53(3). doi: 10.1183/13993003.00986-2018
 42. Zhao W, Yang J, Ni B, Bi D, Sun Y, Xu M, et al. Toward Automatic Prediction of EGFR Mutation Status in Pulmonary Adenocarcinoma With 3D Deep Learning. *Cancer Med* (2019) 8(7):3532–43. doi: 10.1002/cam4.2233
 43. Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, et al. CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in non-Small Cell Lung Cancer. *Eur Radiol* (2016) 26(1):32–42. doi: 10.1007/s00330-015-3814-0
 44. Bell DW, Lynch TJ, Hasserlat SM, Harris PL, Okimoto RA, Brannigan BW, et al. Epidermal Growth Factor Receptor Mutations and Gene Amplification in Non-Small-Cell Lung Cancer: Molecular Analysis of the IDEAL/INTACT Gefitinib Trials. *J Clin Oncol* (2005) 23(31):8081–92. doi: 10.1200/jco.2005.02.7078
 45. Hong SJ, Kim TJ, Choi YW, Park JS, Chung JH, Lee KW. Radiogenomic Correlation in Lung Adenocarcinoma With Epidermal Growth Factor Receptor Mutations: Imaging Features and Histological Subtypes. *Eur Radiol* (2016) 26(10):3660–8. doi: 10.1007/s00330-015-4196-z
 46. Li Y, Li X, Li H, Zhao Y, Liu Z, Sun K, et al. Genomic Characterisation of Pulmonary Subsolid Nodules: Mutational Landscape and Radiological Features. *Eur Respir J* (2020) 55(2):1901409. doi: 10.1183/13993003.01409-2019
 47. Aherne EA, Plodkowski AJ, Montecalvo J, Hayan S, Zheng J, Capanu M, et al. What CT Characteristics of Lepidic Predominant Pattern Lung Adenocarcinomas Correlate With Invasiveness on Pathology? *Lung Cancer (Amsterdam Netherlands)* (2018) 118:83–9. doi: 10.1016/j.lungcan.2018.01.013
 48. Nie Y, Liu H, Tan X, Wang H, Li F, Li C, et al. Correlation Between High-Resolution Computed Tomography Lung Nodule Characteristics and EGFR Mutation in Lung Adenocarcinomas. *OncoTargets Ther* (2019) 12:519–26. doi: 10.2147/ott.S184217
 49. Lu L, Sun SH, Yang H, Linning E, Guo P, Schwartz LH, et al. E LRadiomics Prediction of EGFR Status in Lung Cancer—Our Experience in Using Multiple Feature Extractors and the Cancer Imaging Archive Data. *Tomography (Ann Arbor Mich)* (2020) 6(2):223–30. doi: 10.18383/j.tom.2020.00017
 50. Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, et al. Radiomics Signature: A Potential and Incremental Predictor for EGFR Mutation Status in NSCLC Patients, Comparison With CT Morphology. *Lung Cancer (Amsterdam Netherlands)* (2019) 132:28–35. doi: 10.1016/j.lungcan.2019.03.025
 51. Xiong JF, Jia TY, Li XY, Wen Y, Xu ZY, Cai XW, et al. Identifying Epidermal Growth Factor Receptor Mutation Status in Patients With Lung Adenocarcinoma by Three-Dimensional Convolutional Neural Networks. *Br J Radiol* (2018) 91(1092):20180334. doi: 10.1259/bjr.20180334
 52. Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of Pre-Therapy (18)F-FDG PET/CT Radiomics in Predicting EGFR Mutation Status in Patients With Non-Small Cell Lung Cancer. *Eur J Nucl Med Mol Imaging* (2020) 47(5):1137–46. doi: 10.1007/s00259-019-04592-1
 53. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic Mutations Affect Key Pathways in Lung Adenocarcinoma. *Nature* (2008) 455(7216):1069–75. doi: 10.1038/nature07423
 54. Dearden S, Stevens J, Wu YL, Blowers D. Mutation Incidence and Coincidence in Non Small-Cell Lung Cancer: Meta-Analyses by Ethnicity and Histology (Mutmap). *Ann Oncol: Off J Eur Soc Med Oncol* (2013) 24(9):2371–6. doi: 10.1093/annonc/mdt205

Conflict of Interest: Author KK was employed by Diane Technology. Author SD was employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Huang, Sun, Tan, Ma, Gao, Qi, Lu, Yang, Wang, Chen, Jin, Kuang, Duan and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.