SCIENTIFIC REP⚙RTS

OPEN

CT Slice Thickness and Convolution Kernel Affect Performance of a Radiomic Model for Predicting EGFR Status in Non-Small Cell Lung Cancer: A Preliminary Study

Yajun Li[1], Lin Lu[2], Manjun Xiao[1], Laurent Dercle[2,3], Yue Huang[5], Zishu Zhang[1], Lawrence H. Schwartz[2], Daiqiang Li[4] & Binsheng Zhao[2]

We evaluated whether the optimal selection of CT reconstruction settings enables the construction of a radiomics model to predict epidermal growth factor receptor (EGFR) mutation status in primary lung adenocarcinoma (LAC) using standard of care CT images. Fifty-one patients (EGFR:wildtype = 23:28) with LACs of clinical stage I/II/IIIA were included in the analysis. The LACs were segmented in four conditions, two slice thicknesses (Thin: 1 mm; Thick: 5 mm) and two convolution kernels (Sharp: B70f/B70s; Smooth: B30f/B31f/B31s), which constituted four groups: (1) Thin-Sharp, (2) Thin-Smooth, (3) Thick-Sharp, and (4) Thick-Smooth. Machine learning algorithms selected and combined 1,695 quantitative image features to build prediction models. The performance of prediction models was assessed by calculating the area under the curve (AUC). The best prediction model yielded AUC (95%CI) = 0.83 (0.68, 0.92) using the Thin-Smooth reconstruction setting. The AUC of models using thick slices was significantly lower than that of thin slices ($P < 10^{-3}$), whereas the impact of reconstruction kernel was not significant. Our study showed that the optimal prediction of EGFR mutational status in early stage LACs was achieved by using thin CT-scan slices, independently of convolution kernels. Results from the prediction model suggest that tumor heterogeneity is associated with EGFR mutation.

Lung cancer is the leading cause of cancer death for men and women in the U.S. and worldwide[1]. Adenocarcinoma is the main histological subtype of non–small cell lung carcinoma. A key mechanism of the tumorigenesis of adenocarcinoma is somatic mutations of the epidermal growth factor receptor (EGFR) gene, which leads to the overexpression of EGFR tyrosine kinase receptor in lung tumor tissue[2]. When ligands bind to the EGFR receptor, the molecule is phosphorylated and activates a downstream signaling pathway that inhibits apoptosis and mediates cancer cell growth, proliferation, and invasion. This creates autocrine and paracrine growth factor loops which promote tumor growth.

The diagnosis of EGFR mutational status on a per patient basis is a key biomarker for defining personalized treatment strategies. The mutation occurs frequently, especially in specific populations including non-smoking females and Asians (the reported frequency of EGFR mutations is 48% in China and 23% in the US[3]), and is therapeutically actionable. Gefitinib is a targeted molecular agent that inhibits EGFR. The EGFR mutation is, therefore, a strong predictor of prolonged progression-free survival and of higher response rate to Gefitinib[4–9]. The efficacy of Gefitinib treatment typically translates into a small magnitude of tumor size decrease and a symptomatic improvement[10,11].

[1]Department of Radiology, the Second Xiangya Hospital, Central South University, Changsha, Hunan, 410011, China. [2]Department of Radiology, Columbia University Medical Center, New York, NY, 10039, USA. [3]Gustave Roussy, Université Paris-Saclay, Inserm, UMR1015, Paris, France. [4]Department of Pathology, the Second Xiangya Hospital, Central South University, Changsha, Hunan, 410011, China. [5]Department of Surgery, University of Missouri, Columbia, MO, USA. Yajun Li and Lin Lu contributed equally. Correspondence and requests for materials should be addressed to Y.L. (email: liyajun9966@csu.edu.cn)

Individualized cancer treatment strategies are enabled by radiomic signatures associated with a specific gene mutation. Current results suggest that EGFR-mutant tumors have a unique imaging phenotype as compared to ALK-mutant[12–14] or EGFR-wildtype tumors[12–22], so that imaging could be used to select patients who will potentially benefit from Gefitinib and direct those without EGFR-mutant tumors to other therapies. Largely due to the lack of reliable software (e.g., tumor segmentation and characterization tools), preliminary data in this area have used qualitative imaging features visually assessed by radiologists which are observer dependent, require training, and are time-consuming to measure. Furthermore, current studies suffer from several limitations. First, most models used qualitative and basic imaging features or a limited set of radiomic features[17,19]. Second, most radiomic studies used retrospective imaging datasets that had heterogeneous imaging settings[16–19] (e.g., reconstruction kernel, slice thickness) which could affect radiomic features and thus critically alter the accuracy of radiomic signatures[16,18,19,23–26].

Therefore, in this study, we evaluated whether the optimization of reconstruction settings (i.e. Thin/Thick slice thicknesses, Sharp/Smooth convolution kernels) could allow the construction of a better radiomic signature, derived from a large number of quantitative image features, to predict the EGFR mutation status in primary lung adenocarcinoma (LAC) using standard of care CT imaging.

Results

Patient Characteristics.
A total of 51 patients were included in this study and scanned with all four CT imaging settings formed by the combinations of two slice thicknesses (1 mm and 5 mm) and two convolution kernels (Sharp and Smooth). Fifty-one primary lung adenocarcinoma tumors (one tumor per patient) were identified for analysis. The patient and tumor characteristics of the 51 patients are provided in Table 1 (The information about plerual invasion and tumor density (solid/partial-solid/GGO) were provided by Y.L., an experienced radiologist with 20-year experience of CT interpretation. The evaluation was done blinded from the tumor mutation status). We can see that female and non-smoker patients were more likely to have EGFR mutant tumors, suggesting that our study cohort was typical of Asian populations[27]. For tumor characteristics, such as size, density and pleural invasion, there were no significant difference between EGFR mutant and Wild Type groups (t-test for continuous data and chi-squared test for categorical data).

Reproducible Features.
For each lesion in each image group, 1,695 quantitative image features were extracted. By employing the same-day repeat CT dataset[25], 954, 1,182, 812, and 964 features were identified as reproducible features for the four image groups (Thin-Shp, Thin-Smo, Thick-Shp and Thick-Smo), respectively. In the coarse feature selection, a concordance correlation coefficient (CCC) threshold of 0.8 was applied to generate a compact list of candidate features, removing as redundant all features with correlation greater than 0.8. After applying the correlation threshold, the compact feature list contained 104, 98, 92, and 86 candidate features for the four image groups, respectively. In each compact feature list, only the top ten features were retained for the subsequent fine feature selection and model building.

Optimal EGFR Prediction Models at Different Imaging Settings.
The purpose of our study was to compare prediction models built upon images grouped according to the different acquisition parameters. Thus, we identified four optimal prediction models for the four imaging setting groups (Thin-Shp, Thin-Smo, Thick-Shp and Thick-Smo). In addition, we constructed a 'mixture' group of images for the comparison of prediction models built on homogenous image series vs. those built on heterogeneous acquisition parameters. In the mixture group, the image series for each patient was randomly collected from one of the four image groups and the image features were those reproducible across all the four image settings. For the sake of randomness, we studied ten random 'mixture' groups. For the 'mixture' model, the performance was the average of the ten randomly constructed models. The performance of the optimal model for each image groups is presented in Fig. 1. The features selected for building those optimal modes are presented in Table 2.

As shown in Fig. 1, Thin-Smo and Thin-Shp were identified as the best imaging settings to build EGFR prediction models, with the Thin-Smo model (AUC = 0.83) performing slightly better than the Thin-Shp model (AUC = 0.82) (P = 0.345). Thick-Smo and Thick-Shp were identified as the worst imaging settings to build EGFR prediction models, with the Thick-Smo model (AUC = 0.77) slightly worse than the Thick-Shp model (AUC = 0.79) (P = 0.0756). There were statistically significant differences (P < 0.001) between Thin and Thick models. Compared to the four homogenous imaging settings, the Mixture setting group was identified as the worst image group to build EGFR prediction model, significantly worse than the other four settings (P < 0.001). Moreover, the numbers of support vectors for the four optimal prediction SVM models, one for each of the four imaging settings (Thin-Shp, Thin-Smo, Thick-Shp and Thick-Smo), were 10, 12, 16 and 16, respectively, *i.e.*, 19.6%, 23.5%, 31.3% and 31.3% of patients were used as the support vectors for the final models, respectively. It is noted that models based on thin-section imaging settings were more generalizable than that based on thick-section imaging settings.

When applying the best model built using the Thin-Smo imaging setting to the other four imaging setting groups, the Thin-Shp group's performance stayed almost unchanged. The two Thick groups' performances dropped, but the Mixture group's performance increased (Fig. 2).

It is noticeable that the top two features selected to build the Thin-Smo and Thin-Shp models were the same, LoG_Entropy_Sigma2.5_2D and LoG_Z_Uniformity_Sigma2.5_2D. The higher the 'LoG_Entropy_Sigma2.5_2D' value, the more heterogeneous the lesion. As shown in Fig. 3, the median values of 'LoG_Entropy_Sigma2.5_2D' on EGFR mutant lesions were larger than those on EGFR wild-type lesions in all four imaging setting groups. Among the four groups, median values of 'LoG_Entropy_Sigma2.5_2D' on EGFR mutant lesions decreased from Thin-Shp to Thick-Smo groups, but remained relatively stable on EGFR wild-type lesions.
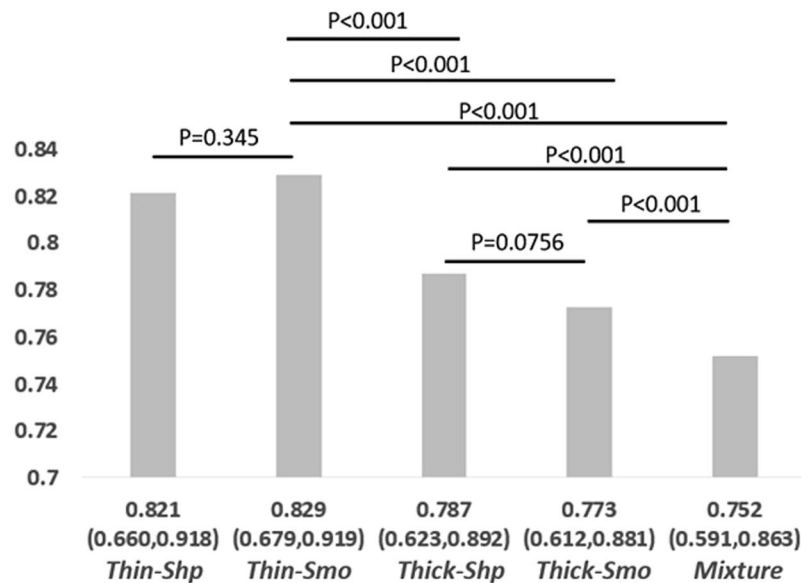
Figure 1. Performances of the optimal models built on different image groups (Thin-Shp, Thin-Smo, Thick-Shp, Thicl-Smo and Mixture).

Discussion

In this study, we demonstrated that the optimal selection of reconstruction parameters on CT-scan could enhance the predictive value of a radiomics signature to identify EGFR mutation status in early-stage lung adenocarcinoma. We also found that heterogeneous tumors were more likely to harbor the EGFR mutation. Our results showed that thin slices are the optimal reconstruction setting among the four most commonly used CT imaging parameters, as using thick slices significantly altered the predictive value of radiomics ($P < 10^{-3}$).

Our study demonstrated that the thin slices yielded AUCs of 0.82 and 0.83 for the prediction of EGFR mutation using sharp and smooth convolution kernels respectively. As a comparison, thick slices yielded AUCs of 0.79 and 0.77 using sharp and smooth convolution kernels respectively. The impact of convolution kernel was not found to be significant. This may be because the two most significant features selected for the thin-smo and thin-shp image series, Laplacian of Gaussian Entropy and Laplacian of Gaussian Uniformity, were computed from the images pre-processed by a smoothing filter (Gaussian-filter with a larger-sized filter length). This pre-processing procedure reduced the differences between smooth and sharp images. On a broader perspective, this report further demonstrates the importance of rigorous image acquisitions previously indicated by our group (i.e. impact of reconstruction settings[23], interobserver variability[28], and acquisition protocols[29]). Interestingly, a previous study reported that non-contrast, thin-slice, and standard convolution kernel-based CT in solitary pulmonary nodule was more informative and increased the diagnostic performance of a radiomics signature[25].

Laplacian of Gaussian entropy was a key feature selected in three out of four reconstruction settings. The Laplacian of Gaussian will smooth the image, which might explain why this feature was selected using both sharp and smooth reconstruction settings. This feature captures a heterogeneity pattern in pixel spatial distribution and, interestingly, has been previously found to be associated with tumor phenotype[30], tumor gene expression, tumor metabolism, tumor stage[31,32], patient prognosis[33–36], and treatment response. Laplacian of Gaussian entropy may be considered as a tumor-specific imaging biomarker that is also a function of the primary tumor type, the size of the tumor, and the metastatic site[30]. In this work we show that the difference in this biomarker between EGFR-wildtype and EGFR-mutant is influenced by the reconstruction settings, increasing when using Thin-Sharp reconstruction setting and decreasing when Thin-Smooth, Thick-Sharp and Thick-Smooth reconstruction settings are used. Two example cases, one EGFR mutant and one EGFR wild-type, were presented in Fig. 4.

Our study reported that EGFR tumors are heterogeneous. This is in line with the current literature indicating that EGFR-mutant tumors have a unique imaging phenotype[12–22], with differences in ground glass opacity, tumor size, pleural retraction and air bronchogram the most frequently reported imaging features. Other EGFR mutant-related features were reported anecdotally such as tumor shape, heterogeneous enhancement, calcification, peripheral fibrosis/emphysema, border definition, spiculation, pleural attachment/effusion, tumor location, nodules in primary tumor lobe, nodules in non-tumor lobes, N-stage, and M-stage[12–22]. Consequently, our study provides an external validation using quantitative image features that enhance the knowledge about the imaging phenotype associated with EGFR mutation in LACs. The independent validation of those EGFR mutant-related imaging features in our series is of crucial significance since type I errors and over-fitting is expected in radiomics studies.

More importantly, we validated our optimal radiomics-EGFR signature (AUC = 0.83) using both homogeneous and heterogeneous CT acquisition settings, a wide range of imaging features (n = 1695), and a machine learning approach. The improvement of the accuracy of the radiomics model in our series is of interest since most
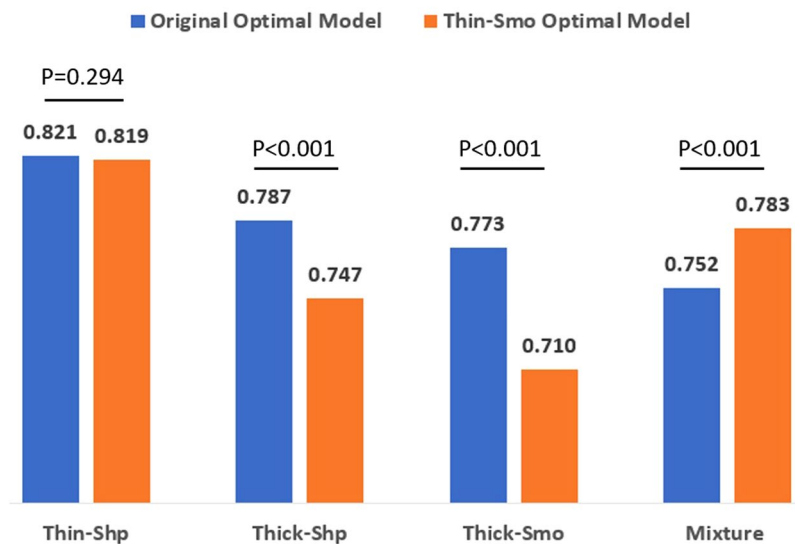
Figure 2. Performances of applying the Thin-Smo optimal model to the other four image groups.
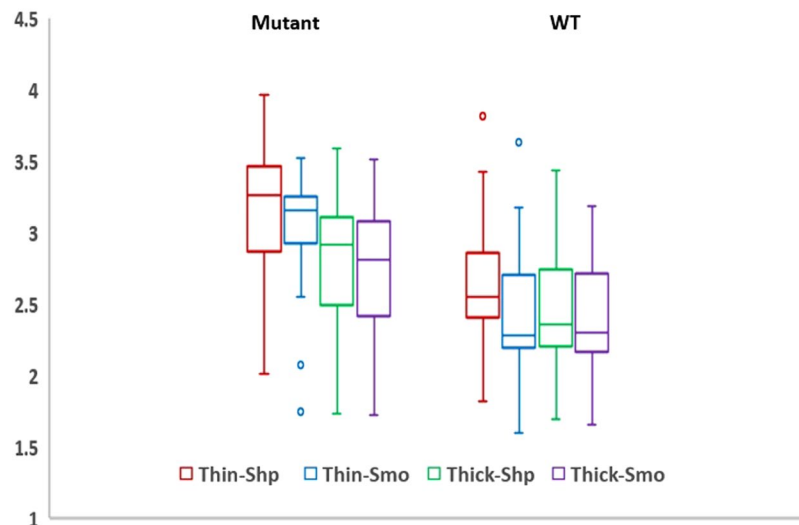

Figure 3. Feature value distributions of Laplacian of Gaussian-Entropy at different imaging setting groups.

previous models were built using basic imaging features, mostly qualitative rather than radiomic per se[12–15,18–22,37]. Prior radiomics studies used a limited number of imaging features (183[16], 5[17], 299[18], 30[19]) compared to the new imaging features implemented in our study. The AUC of previous models based on radiomic features were only 0.67[16] and 0.71[18], outperformed by our model (AUC = 0.83). Our study data included patients from a single Chinese institution, allowing them to be benchmarked against previous models which were also designed at a single center institution, mostly in Asia (China, Korea)[12–14,16–22].

We believe that these results could have major applications since CT-scans guide decision making throughout the course of NSCLC, including screening[38–40], characterization of lung nodules, TNM staging, biopsy guiding, radiation treatment planning, and response assessment. The outcome of TNM staging is defined by quantitative imaging metrics such as tumor size[41–43] or binary metrics derived from medical images (such as the involvement of the main bronchus or the presence of atelectasis, pneumonitis, or a diaphragm invasion)[44–46]. Furthermore, guiding personalized treatment by imaging biomarkers offers the prospect of a "virtual biopsy", which is attractive because conventional biopsies are limited to the sampling site and have a low negative predictive value (68%) and a significant false negative rate (9%)[47], especially in the case of a large lesion and a sub-solid nodule[48].

	EGFR Mutant	EGFR WildType	P
Number	23	28	
Age (yr, mean ± SD)	58.6 ± 9.42	57.6 ± 9.99	0.714
Gender			<0.01
Male	9	23	
Female	14	5	
Differentiation			<0.01
High/Mid	14	8	
Poor	6	20	
Not-known	3	/	
Stage			0.250
I	4	1	
II	4	5	
III	15	22	
Size	39.45 ± 17.94(mm)	43.33 ± 17.22(mm)	0.436
N-Staging			0.252
0	11	8	
1	1	2	
2	6	11	
3	5	7	
Density*			
Solid	17	16	0.185
Part-Solid	5	11	
GGO	1	1	
Pleural invasion			
Yes	11	14	0.899
No	12	14	
Smoking status			
Non-smoker	17	10	<0.01
Smoker	6	18	

Table 1. Distribution of patient and tumor characteristics. *The definition of tumor density is, solid - nodule has homogenous soft-tissue attenuation; partial-solid -nodule consists of both ground glass and solid soft-tissue attenuation components; GGO - nodule manifests as hazy increased attenuation in the lung that does not obliterate the bronchial and vascular margin.

Additionally, CT-guided lung biopsies are associated with complications such as pneumothorax and parenchymal hemorrhage[49,50].

One limitation of our model is that it was built in early stage lung adenocarcinoma, for which treatment is surgery and external beam radiation therapy rather than EGFR inhibitor. However, the definition of the radiomics signature in this model offers several advantages. First, because all patients included in our series had surgery, our reference standard for the identification of EGFR status is robust compared to the determination of EGFR mutational status using biopsies, which suffer from sampling bias. Second, the contours of early stage lung cancers are well defined compared to invasive and/or infiltrative advanced stage lung cancer, in which the determination of the border of the tumor is challenging due to atelectasia, pleuresia, and invasion of other structures by cancer. Third, because biopsies are not always performed in early stage treatment prior to external beam radiation therapies, this is a case in which virtual biopsy through imaging biomarkers is most likely to be useful. Another limitation of this study is the small number of patients, i.e., only 51 patients met the data inclusion criteria. We plan to continue the data collection, from multiple institutions, to validate our findings in the future.

We concluded that the optimal reconstruction setting on CT-scan to predict the presence of EGFR mutations in early stage LAC is thin slices. This could provide a noninvasive method to predict the genetic characteristics of LACs and help personalize patients care.

Method
Patients and Image Acquisition. Patient data were retrospectively collected from the Second Xiangya Hospital of Central South University, China. For retrospective study, the institutional review board approved the study before its commencement and waived the requirement for informed consent. Also, all experiments were performed in accordance with relevant guidelines and regulations of the institution. The primary patient cohort in this paper was collected by searching the institutional database for consecutive inpatients who met the following criteria: (1) underwent molecular examination from May 2014 to Dec 2016; (2) had complete histological and clinical information; and (3) were diagnosed for primary Stage I-III lung adenocarcinoma by surgical resection
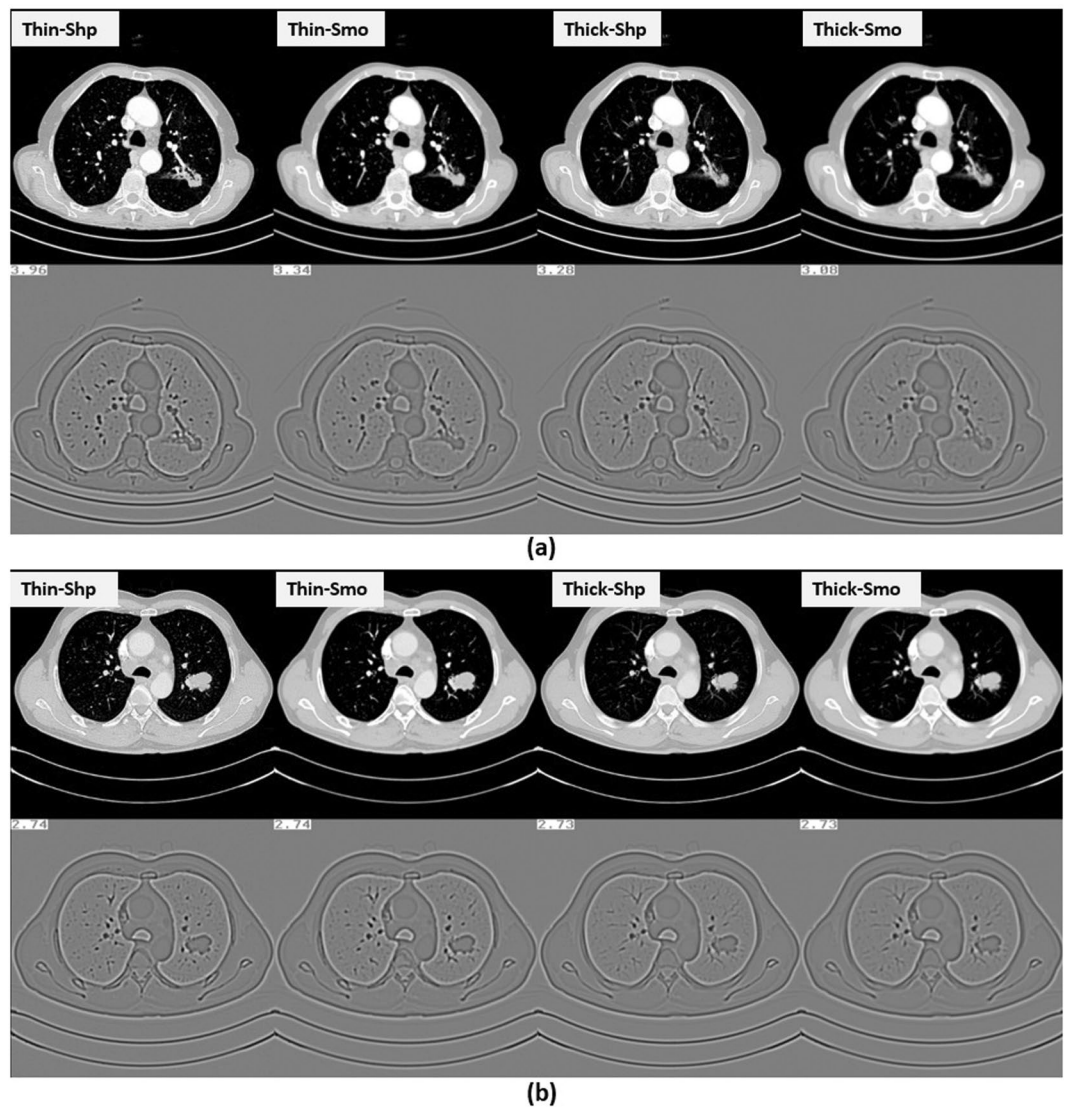
Figure 4. (**a**) An EGFR mutant case. (**b**) An EGFR wild-type case. The four setting images of the two cases were shown from left to right. Images shown at the second row of the sub-pictures were images processed by the Laplacian of Gaussian filter. The values of Laplacian of Gaussian-Entropy were shown at the top-left of the processed images.

Thin-Shp	Thin-Smo	Thick-Shp	Thick-Smo
LoG_Entropy_Sigma2.5_2D	LoG_Entropy_Sigma2.5_2D	GLCM_ASM_2D	Laplacian_Entropy_3D
LoG_Z_Uniformity_Sigma2.5_2D	LoG_Z_Uniformity_Sigma2.5_2D	GTDM_Contrast_2D	Sigmoid_Slope_Std_3D
GTDM_Coarseness_25D	Volume_3D	Sigmoid_Kurtosis_3D	GLCM_Homogeneity_2D
Gabor_dir135_2D	Shape_SI9_3D		Sigmoid_Amplitude_3D
	Gabor_dir45_3D		Laws_8_3D

Table 2. Selected features for building the optimal models.

(the 8th Edition of TNM in Lung Cancer). In total, 355 patients were collected, and 315 of them had complete histological and clinical information. Among the 315 patients, 74 patients were diagnosed for primary lung adenocarcinoma at Stage I-III.

Molecular examination was performed on all of these 74 patients using tumor specimens from surgical resection. EGFR wild-type and mutant status was determined by an amplification refractory mutation system real-time technology using Human EGFR Gene Mutations Fluorescence Polymerase Chain Reaction (PCR) Diagnostic Kit (Amoy Diagnostics Co., Ltd, Xiamen, China).

Image Group	Thin-Shp	Thin-Smo	Thick-Shp	Thick-Smo
Slice Thickness (mm)	1	1	5	5
Convolution Kernels	B70f / B70s	B30f/B31f/B31s	B70f/B70s	B30f/B31f/B31s
	EGFR Mutant		**EGFR Wild Type**	**p**
Manufacturer SIEMENS	23		28	/
Scan Model				0.396
Perspective	12		10	
Sensation 64	5		6	
SOMATOM Definition AS	2		1	
SOMATOM Definition Flash	4		9	
SOMATOM Force	0		2	
KVP				0.319
90	0		2	
110	11		10	
120	11		16	
130	1		0	
Exposure (mean ± SD mAs)	123 ± 60		99 ± 55	0.143
Pixel Spacing (mean ± SD mm)	0.655 ± 0.053		0.630 ± 0.055	0.107
Pitch (mean ± SD mm)	1.31 ± 0.56		1.11 ± 0.21	0.153
Contrast Agent				0.753
APPLIED	19		21	
Non-APPLIED	4		7	

Table 3. Distributions of CT scanners and scanning parameters in EGFR mutant and Wild Type groupes.

CT images of the 74 patients were searched approximately one month before their surgery. For each patient, multiple imaging series of CT data were collected. For the contrast-enhanced CT, the IV contrast was injected at a rate of 2.5 mL/sec via a pump injector, and the total amount of IV contrast were 60~70 ml. The scanning on thorax began on 30-second delay of the contrast injection. Imaging protocols used to scan the patients are provided in Table 3. Raw data of each patient's CT scan were reconstructed into four image series: the combinations of two slice thicknesses (Thin: 1 mm; Thick: 5 mm) and two convolution kernels (Sharp: B70f / B70s; Smooth: B30f/B31f/B31s). That is, each patient had four groups of images: 1) Thin-Shp, 2) Thin-Smo, 3) Thick-Shp, and 4) Thick-Smo. The four different image groups differ on levels of image spatial resolution and noise, e.g. Thin-Shp resulted in images with high spatial frequencies and noise preserved, while Thick-Smo resulted in images with low spatial frequencies and noise decreased. As can be seen in Table 3, since all the four image groups were reconstructed by using the exactly same raw scanning data, there was no bias on the comparison among the four groups. Finally, 51 patients having all four imaging settings were used as the study cohort in this paper.

Tumor Segmentation. Tumor segmentation is a procedure to define the lesion area from which radiomic features will be extracted. In our study, 51 lesions (one per patient) were segmented out from their surrounding background by using a semi-automated algorithm[51] based on watershed and active contours image processing techniques. The semi-automated segmentation was performed by an experienced radiologist (YL with 20 year experience of CT interpretation) on all four image groups. For the sake of consistency, tumor segmentation was first performed on the Thin-Shp group, and then duplicated to the other three image groups. During the duplication, slice re-sampling was used to guarantee the same voxel-resolution between two image series. The radiologist was allowed to edit the duplicated contours if slice re-sampling caused pixel shifting on images.

Feature Extraction. For each lesion in each image group, totally 1,695 well-defined quantitative image features were extracted by using an in-house feature extraction algorithm implemented in Matlab 2016b (Mathworks, Natick, USA). The 1,695 extracted features are able to characterize tumor phenotypes in terms of size (e.g., largest diameter, volume), shape (e.g., roundness, compactness), sharpness (e.g., Sigmoid slope), texture patterns (e.g., tumor heterogeneity quantified by Laplacian of Gaussian image filter, Gray-Level Co-occurrence Matrix). The 1,695 features represented an expansion of the set of 89 imaging features used in our previous work[23], achieved through increasing the scales of feature parameters. For instance, four scales of Laplacian of Gaussian filter, 0, 0.5, 1.5, and 2.5 sigma, were used in this study.

Reproducibility Analysis. A previously collected NSCLC test-retest dataset[52] was used to assess the reproducibility of the extracted quantitative image features. This test-retest dataset was a CT imaging dataset consisting of 32 NSCLC patients who underwent two repeat CT scans within 15 minutes. The 32 CT scans were reconstructed into six imaging setting groups, four of which were similar to the four reconstruction parameters used in our study. For each image group of the 32 patients, 1,695 radiomics features were extracted from each lesion on both test and re-test scans using the same method presented above. Concordance correlation coefficient (CCC)[53]

was used to evaluate the reproducibility of features for each image group. Features with a CCC larger than 0.9 were included for the subsequent analyses.

Model Building.   In our study, a 'coarse' to 'fine' strategy was employed to select informative and non-redundant candidate features from the large feature pool consisting of over a thousand features. The coarse selection was fast but only based on the properties of individual features, while the fine selection was time-consuming but based on the combination effect of multiple features.

The coarse selection included two steps, redundancy removal and feature ranking. In the step of redundancy removal, features with high correlation were regarded as redundant features and thus excluded from the following analysis. The procedure included first calculating correlation between features, then organizing all features into a hierarchical clustering tree according to their mutual correlations, and finally setting a correlation threshold to separate all features into a series of redundant groups (i.e. when setting correlation threshold as 0.5, it meant all candidate features are clustered into a series of redundant groups, within which correlation of all feature exceed 0.5). For each redundant group, feature ranking algorithms were applied to rank those correlated features, and only the top-ranked feature was selected for the following analysis. In our study, six feature ranking algorithms were employed, i.e., RELIEF[54], Chi-square score, Minimum redundancy maximum relevance[55], T-test score, Wilcoxon score, and Univariance accuracy. Through coarse selection, we created six compact candidate feature lists. Top ten features in each candidate feature list were used for the following analysis.

Fine selection was then applied to determine optimal features to be used to construct EGFR prediction modules. In this step, 'forward search' was adopted to evaluate features sequentially. Forward search initiated on an empty set and selected a feature if and only if the addition of the feature could increase the performance of prediction model. The procedure of forward search was repeated until all the candidate features in the compact candidate feature lists were evaluated. The Support Vector Machine algorithm[56] was used to construct models. As there were six feature lists, a total of six candidate prediction models were generated. Among the six prediction models, the model that achieved the highest performance was selected as the final optimal model for each imaging group.

In the implementation, all the algorithms were coded or download as packages on the Matlab 2016b (Mathworks, Natick, USA) platform. Parameters involved were all used default settings except the Box-Constraint[56] for the SVM algorithm. SVM algorithm is a machine-learning algorithm that performs classification by finding the hyperplane that maximizes the margin between two classes defined by the so-called support vectors, the percentage of the patient sample set[56]. Theoretically, SVM algorithm can fit any distribution of patients by using support vectors. However, the more the support vectors are used, the higher probability that the SVM model is overfitting. Therefore, SVM algorithm introduces the Box-Constraint, a parameter that controls the maximum penalty imposed on margin-violating support vectors, to help to prevent overfitting, *i.e.*, if Box-Constraint increases, then fewer support vectors will be used by the model. In our study, based on our previous experience, we empirically set the Box-Constraint $= 100$, one hundred times of the default Box-Constraint $= 1$ in Matlab, to prevent overfitting.

Performance Evaluation.   The performance of candidate prediction model was evaluated in terms of AUC (i.e. the area under the curve of receiver operating characteristic curve[57]). Due to the limited number of patients in the study cohort, we used three-fold cross-validation to estimate the performance of models instead of separating the study cohort into training and testing subsets. In the three-fold cross-validation, original data were randomly partitioned into three groups. When one group was used for testing, then the other groups was retained for training. The training-testing procedures were repeated three times, until each sample in the data set was assigned a prediction score. The final AUC as well as its confidence interval (95%) were estimated based on the prediction score by using bootstrapping (1000 times)[58]. Also, a bootstrap-based approach presented in the literature[59] was used to compare two models in terms of p-value.

Data Availability
The datasets generated and analyzed during the current study are available from the corresponding author.

References
1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7–30, https://doi.org/10.3322/caac.21332 (2016).
2. Rusch, V. *et al*. Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung. *Cancer Res* **53**, 2379–2385 (1993).
3. Midha, A., Dearden, S. & McCormack, R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). *Am J Cancer Res* **5**, 2892–2911 (2015).
4. Mok, T. S. *et al*. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**, 947–957, https://doi.org/10.1056/NEJMoa0810699 (2009).
5. Rosell, R. *et al*. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* **361**, 958–967, https://doi.org/10.1056/NEJMoa0904554 (2009).
6. Jackman, D. M. *et al*. Impact of epidermal growth factor receptor and KRAS mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: results of an online tumor registry of clinical trials. *Clin Cancer Res* **15**, 5267–5273, https://doi.org/10.1158/1078-0432.CCR-09-0888 (2009).
7. Sequist, L. V. *et al*. First-line gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations. *J Clin Oncol* **26**, 2442–2449, https://doi.org/10.1200/JCO.2007.14.8494 (2008).
8. Paez, J. G. *et al*. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500, https://doi.org/10.1126/science.1099314 (2004).
9. Lynch, T. J. *et al*. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129–2139, https://doi.org/10.1056/NEJMoa040938 (2004).
10. Kris, M. G. *et al*. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* **290**, 2149–2158, https://doi.org/10.1001/jama.290.16.2149 (2003).
11. Miller, V. A. *et al*. Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. *J Clin Oncol* **22**, 1103–1109, https://doi.org/10.1200/JCO.2004.08.158 (2004).

12. Kim, T. J., Lee, C. T., Jheon, S. H., Park, J. S. & Chung, J. H. Radiologic Characteristics of Surgically Resected Non-Small Cell Lung Cancer With ALK Rearrangement or EGFR Mutations. *Ann Thorac Surg* **101**, 473–480, https://doi.org/10.1016/j.athoracsur.2015.07.062 (2016).

13. Zhou, J. Y. *et al.* Comparative analysis of clinicoradiologic characteristics of lung adenocarcinomas with ALK rearrangements or EGFR mutations. *Eur Radiol* **25**, 1257–1266, https://doi.org/10.1007/s00330-014-3516-z (2015).

14. Choi, C. M., Kim, M. Y., Hwang, H. J., Lee, J. B. & Kim, W. S. Advanced adenocarcinoma of the lung: comparison of CT characteristics of patients with anaplastic lymphoma kinase gene rearrangement and those with epidermal growth factor receptor mutation. *Radiology* **275**, 272–279, https://doi.org/10.1148/radiol.14140848 (2015).

15. Rizzo, S. *et al.* CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. *Eur Radiol* **26**, 32–42, https://doi.org/10.1007/s00330-015-3814-0 (2016).

16. Aerts, H. J. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. *Sci Rep* **6**, 33860, https://doi.org/10.1038/srep33860 (2016).

17. Ozkan, E. *et al.* CT Gray-Level Texture Analysis as a Quantitative Imaging Biomarker of Epidermal Growth Factor Receptor Mutation Status in Adenocarcinoma of the Lung. *AJR Am J Roentgenol* **205**, 1016–1025, https://doi.org/10.2214/AJR.14.14147 (2015).

18. Liu, Y. *et al.* Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clin Lung Cancer* **17**, 441–448 e446, https://doi.org/10.1016/j.cllc.2016.02.001 (2016).

19. Liu, Y. *et al.* CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology* **280**, 271–280, https://doi.org/10.1148/radiol.2016151455 (2016).

20. Yang, Y. *et al.* EGFR L858R mutation is associated with lung adenocarcinoma patients with dominant ground-glass opacity. *Lung Cancer* **87**, 272–277, https://doi.org/10.1016/j.lungcan.2014.12.016 (2015).

21. Hsu, J. S. *et al.* Correlation between EGFR mutation status and computed tomography features in patients with advanced pulmonary adenocarcinoma. *J Thorac Imaging* **29**, 357–363, https://doi.org/10.1097/RTI.0000000000000116 (2014).

22. Lee, H. J. *et al.* Epidermal growth factor receptor mutation in lung adenocarcinomas: relationship with CT characteristics and histologic subtypes. *Radiology* **268**, 254–264, https://doi.org/10.1148/radiol.13112553 (2013).

23. Zhao, B. *et al.* Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Scientific reports* **6**, 23428 (2016).

24. Lu, L., Ehmke, R. C., Schwartz, L. H. & Zhao, B. Assessing agreement between radiomic features computed for multiple CT imaging settings. *PLoS One* **11**, e0166550 (2016).

25. He, L. *et al.* Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Scientific reports* **6**, 34921 (2016).

26. Huang, Q. *et al.* Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status. *J Med Imaging (Bellingham)* **5**, 011005, https://doi.org/10.1117/1.JMI.5.1.011005 (2018).

27. Ha, S. Y. *et al.* Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR. *Oncotarget* **6**, 5465 (2015).

28. Huang, Q. *et al.* Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status. *Journal of Medical Imaging* **5**, 011005 (2017).

29. Dercle, L. *et al.* Impact of Variability in Portal Venous Phase Acquisition Timing in Tumor Density Measurement and Treatment Response Assessment: Metastatic Colorectal Cancer as a Paradigm. *JCO Clinical Cancer Informatics* **1**, 1–8 (2017).

30. Dercle, L. *et al.* Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: ROI-area, acquisition protocol and tissue site exert substantial influence. *Sci Rep* **7**, 7952, https://doi.org/10.1038/s41598-017-08310-5 (2017).

31. Ganeshan, B., Skogen, K., Pressney, I., Coutroubis, D. & Miles, K. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival. *Clin Radiol* **67**, 157–164, https://doi.org/10.1016/j.crad.2011.08.012 (2012).

32. Ganeshan, B., Abaleke, S., Young, R. C., Chatwin, C. R. & Miles, K. A. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging* **10**, 137–143, https://doi.org/10.1102/1470-7330.2010.0021 (2010).

33. Ng, F., Kozarski, R., Ganeshan, B. & Goh, V. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol* **82**, 342–348, https://doi.org/10.1016/j.ejrad.2012.10.023 (2013).

34. Ganeshan, B., Panayiotou, E., Burnand, K., Dizdarevic, S. & Miles, K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol* **22**, 796–802, https://doi.org/10.1007/s00330-011-2319-8 (2012).

35. Ng, F., Ganeshan, B., Kozarski, R., Miles, K. A. & Goh, V. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology* **266**, 177–184, https://doi.org/10.1148/radiol.12120254 (2013).

36. Miles, K. A. *et al.* Multifunctional imaging signature for V-KI-RAS2 Kirsten rat sarcoma viral oncogene homolog (KRAS) mutations in colorectal cancer. *J Nucl Med* **55**, 386–391, https://doi.org/10.2967/jnumed.113.120485 (2014).

37. Shi, Z. *et al.* Radiological and Clinical Features associated with Epidermal Growth Factor Receptor Mutation Status of Exon 19 and 21 in Lung Adenocarcinoma. *Sci Rep* **7**, 364, https://doi.org/10.1038/s41598-017-00511-2 (2017).

38. National Lung Screening Trial Research, T. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395–409, https://doi.org/10.1056/NEJMoa1102873 (2011).

39. National Lung Screening Trial Research, T. *et al.* Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* **368**, 1980–1991, https://doi.org/10.1056/NEJMoa1209120 (2013).

40. Ma, J., Ward, E. M., Smith, R. & Jemal, A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* **119**, 1381–1385, https://doi.org/10.1002/cncr.27813 (2013).

41. Yoshizawa, A. *et al.* Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. *Mod Pathol* **24**, 653–664, https://doi.org/10.1038/modpathol.2010.232 (2011).

42. Tsutani, Y. *et al.* Prognostic significance of using solid versus whole tumor size on high-resolution computed tomography for predicting pathologic malignant grade of tumors in clinical stage IA lung adenocarcinoma: a multicenter study. *J Thorac Cardiovasc Surg* **143**, 607–612, https://doi.org/10.1016/j.jtcvs.2011.10.037 (2012).

43. Maeyashiki, T. *et al.* The size of consolidation on thin-section computed tomography is a better predictor of survival than the maximum tumour dimension in resectable lung cancer. *Eur J Cardiothorac Surg* **43**, 915–918, https://doi.org/10.1093/ejcts/ezs516 (2013).

44. Rami-Porta, R. *et al.* The IASLC Lung Cancer Staging Project: Proposals for the Revisions of the T Descriptors in the ForthcomingEighth Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* **10**, 990–1003, https://doi.org/10.1097/JTO.0000000000000559 (2015).

45. Goldstraw, P. *et al.* The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* **11**, 39–51, https://doi.org/10.1016/j.jtho.2015.09.009 (2016).

46. Rusch, V. W. *et al.* The IASLC Mesothelioma Staging Project: Proposals for the M Descriptors and for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Mesothelioma. *J Thorac Oncol* **11**, 2112–2119, https://doi.org/10.1016/j.jtho.2016.09.124 (2016).
47. Quint, L. E., Kretschmer, M., Chang, A. & Nan, B. CT-guided thoracic core biopsies: value of a negative result. *Cancer Imaging* **6**, 163–167, https://doi.org/10.1102/1470-7330.2006.0027 (2006).
48. Suh, Y. J. *et al.* Predictors of False-Negative Results from Percutaneous Transthoracic Fine-Needle Aspiration Biopsy: An Observational Study from a Retrospective Cohort. *Yonsei Med J* **57**, 1243–1251, https://doi.org/10.3349/ymj.2016.57.5.1243 (2016).
49. Rizzo, S. *et al.* Risk factors for complications of CT-guided lung biopsies. *Radiol Med* **116**, 548–563, https://doi.org/10.1007/s11547-011-0619-9 (2011).
50. Gupta, S., Wallace, M. J., Morello, F. A. Jr, Ahrar, K. & Hicks, M. E. CT-guided percutaneous needle biopsy of intrathoracic lesions by using the transsternal approach: experience in 37 patients. *Radiology* **222**, 57–62, https://doi.org/10.1148/radiol.2221010614 (2002).
51. Tan, Y., Schwartz, L. H. & Zhao, B. Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. *Medical physics* **40** (2013).
52. Zhao, B. *et al.* Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non–small cell lung cancer. *Radiology* **252**, 263–272 (2009).
53. Lawrence, I. & Lin, K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 255–268 (1989).
54. Kononenko, I. In *European conference on machine learning.* 171–182 (Springer).
55. Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology* **3**, 185–205 (2005).
56. Cortes, C. & Vapnik, V. Support-vector networks. *Machine learning* **20**, 273–297 (1995).
57. Devijver, P. A. & Kittler, J. *Pattern recognition: A statistical approach.* (Prentice hall, 1982).
58. Efron, B. & Tibshirani, R. J. *An introduction to the bootstrap.* (CRC press, 1994).
59. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications* **5**, 4006 (2014).

Acknowledgements

Author Contributions

B.Z, Y.L. and L.L. conceived the project. Y.L. and L.L. contributed equally to this study. Y.L. led the data collection, whereas L.L. led the data analysis. Y.L., M.X., Y.H., Z.Z., D.L. participated in the data collection. L.L., L.D., L.H.S., B.Z. participated in the data analysis. L.C., L.L. and B.Z. drafted the paper. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.