Diagnostic Accuracy Study

Medicine®

OPEN

Can CT radiomic analysis in NSCLC predict histology and EGFR mutation status?

Subba R. Digumarthy, MD[a,*], Atul M. Padole, MD[a], Roberto Lo Gullo, MD[c], Lecia V. Sequist, MD[b], Mannudeep K. Kalra, MD[a]

Abstract

To assess the role of radiomic features in distinguishing squamous and adenocarcinoma subtypes of nonsmall cell lung cancers (NSCLC) and predict EGFR mutations.

Institution Review Board-approved study included chest CT scans of 93 consecutive patients (43 men, 50 women, mean age 60 ± 11 years) with biopsy-proven squamous and adenocarcinoma lung cancers greater than 1 cm. All cancers were evaluated for epidermal growth factor receptor (EGFR) mutation. The clinical parameters such as age, sex, and smoking history and standard morphology-based CT imaging features such as target lesion longest diameter (LD), longest perpendicular diameter (LPD), density, and presence of cavity were recorded. The radiomics data was obtained using commercial CT texture analysis (CTTA) software. The CTTA was performed on a single image of the dominant lung lesion. The predictive value of clinical history, standard imaging features, and radiomics was assessed with multivariable logistic regression and receiver operating characteristic (ROC) analyses.

Between adenocarcinoma and squamous cell carcinomas, ROC analysis showed significant difference in 3/11 radiomic features (entropy, normalized SD, total) [AUC 0.686–0.744, $P = .006$ to $< .0001$], 1/3 clinical features (smoking) [AUC 0.732, $P = .001$], and 2/3 imaging features (LD and LPD) [AUC 0.646–0658, $P = .020$ to .032]. ROC analysis for probability variables showed higher values for radiomics (AUC 0.800, $P < .0001$) than clinical (AUC 0.676, $P = .017$) and standard imaging (AUC 0.708, $P < .0001$). Between EGFR mutant and wild-type adenocarcinoma, ROC analysis showed significant difference in 2/11 radiomic features (kurtosis, K2) [AUC 0.656–0.713, $P = .03$ to .003], 1/3 clinical features (smoking) [AUC 0.758, $P < .0001$]. The combined probability variable for radiomics, clinical and imaging features was higher (AUC 0.890, $P < .0001$) than independent probability variables.

The radiomics evaluation adds incremental value to clinical history and standard imaging features in predicting histology and EGFR mutations.

Abbreviations: AUC = area under curve, EGFR = epidermal growth factor receptor, HU = Hounsfield unit (HU), LD = longest diameter, LPD = longest perpendicular diameter, MPP = mean positive pixels, NCCN = National Comprehensive Cancer Network, NSCLC = nonsmall cell lung cancer, PPP = percent positive pixels, ROC = receiver operating characteristic, ROI = region of interest, SSF = spatial scaling factor.

Keywords: adenocarcinoma, EGFR mutation, NSCLC, radiomics, squamous cell carcinomas

1. Introduction

Lung cancer is the leading cause of cancer-related death worldwide, with a dismal 5-year survival rate of 15% in men and 21% in women, according to the American Cancer Society.[1] Over the last two decades, progress has been made in understanding the genetic and molecular basis of lung cancer in the hope that a genotype-driven targeted treatment approach to lung cancer will improve the survival and quality of life of patients with lung cancer.[2–5] These new targeted therapies are efficacious and selective in their action. Consequently, it is now standard clinical practice to genotype advanced nonsmall cell lung cancer (NSCLC) at the time of diagnosis to help choose the best therapy.[2]

Early initiation of targeted therapies is associated with improved outcome, prolonged progression-free survival, and lower loco-regional recurrence rates but they should not be considered until tumor histology and molecular genetic analysis have been confirmed.[5,6] The National Comprehensive Cancer Network (NCCN) has described clinical practice guidelines for molecular genetic analysis for which there are FDA-approved targeted therapies.[7] Though genotyping is essential for choosing the best treatment, there are barriers in some practice settings such as ability to get sufficient tissue for testing, cost of genotyping, and turn-around time to receive the genotyping results.

A noninvasive technique to obtain information regarding histology and mutations associated with NSCLC could be transformative for enabling targeted therapy, primarily if the technique can be used as an adjunct to a commonly used imaging technique such as CT. Recent publications have highlighted the role of radiomics in various malignancies including lung cancer.[8–18] Radiomics involves histogram-based analyses of distribution and spatial variation of pixel values within a region

Editor: Raffaele Pezzilli.

The authors have no funding and conflicts of interest to disclose.

[a] Department of Radiology, [b] Department of Medicine, Massachusetts General Hospital, Boston, MA, [c] Department of Radiology, University of Milano, Milan, Italy.

[*] Correspondence: Subba R. Digumarthy, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 75 Blossom Court, Suite 236, Boston, MA 02114 (e-mail: sdigumarthy@mgh.harvard.edu).

Medicine (2019) 98:1(e13963)

Received: 2 November 2018 / Received in final form: 6 December 2018 / Accepted: 10 December 2018

http://dx.doi.org/10.1097/MD.0000000000013963

of interest (ROI) to obtain information about tumor heterogeneity. Few published studies have evaluated the role of radiomics in predicting epidermal growth factor receptor (EGFR) mutation in adenocarcinoma, the most common mutation that has an approved targeted therapy as a first-line treatment. To the best of our knowledge, there are no publications on the incremental value of radiomics when combined with clinical history and standard imaging features. The purpose of our study was therefore to assess if radiomics can distinguish lung cancers based on histology and EGFR genetic mutations.

2. Materials and methods

The Human Research Committee of our institutional review board approved the study. The study was compliant with the Health Insurance Portability and Accountability Act (HIPAA). Two studies of co-authors (LVS, MKK) have received research grants or consultation fee for unrelated projects. None of the other authors have any financial conflict of interest concerning the study.

2.1. Patients

Our retrospective study included patients with NSCLC who had known histologic diagnosis and genotyping analysis of at least EGFR between January 2008 and December 2013. Patients were identified from a lung cancer database managed by our Medical Thoracic Oncology Program registry. Patients with histologic subtypes other than adenocarcinoma and squamous cell carcinoma (such as those with small cell and large cell lung cancer and metastatic cancers from nonlung primary sites) were excluded (Fig. 1). We included 93 patients with nonsmall lung cancer (total 94 lung nodules/masses: 69 adenocarcinomas and 25 squamous cell carcinoma). Out of 69 adenocarcinomas, 25 were EGFR mutation positive and the remaining 44 were EGFR wild-type. The mean age of patients was 60 ± 11 years (range: 26–96 years). There were 43 men and 50 women (Table 1). The tissue diagnosis of NSCLC was established with mediastinoscopy, bronchoscopy, or CT-guided biopsy of primary or metastatic sites. All biopsy specimens were tested for EGFR mutations by multiplex PCR-based assay (Snapshot; Applied Biosystems, Foster City, CA). The smoking history was gathered from electronic medical records and was classified as current, former, and never smokers.

2.2. CT scanning

The CT examinations were performed at a single hospital but on multiple different CT scanners (8, 16, 64, and 128- slice multidetector-row CT). All included CT examinations represented contrast-enhanced CT performed using helical acquisition mode, 100–120 kV, automatic exposure control, reconstructed slice thickness of 2.5 to 3 mm and standard soft tissue reconstruction kernel. Only transverse CT images were used for CT texture analysis.

Analyzed images included one chest CT examination per eligible patient performed before any therapy (surgery, radiation, or systemic therapy). In patients with multiple potential imaging studies, the examination closest to the date of tissue sampling was selected. The images were evaluated on PACS system (Impax, software version 6.5, Agfa Healthcare) by fellowship trained thoracic radiologist (SRD, 17 years of experience). The size was recorded in two dimensions: the longest diameter (LD) and longest perpendicular diameter (LPD). The density of the lung lesion was characterized as solid, ground glass, and part solid nodule based on Fleischner society guidelines of thoracic radiology nomenclature.

2.3. Radiomics

DICOM images from CT examinations were imported to a secure offline server for radiomics analysis with commercially available software (TexRAD limited, UK). For each CT examination, a radiologist co-investigator (RLG) identified the transverse image with the maximal dimension of the malignant lung nodule or mass. This co-investigator (RLG) was blinded to the results of the histology and molecular genetic testing. If a motion, contrast streaking or beam hardening artifacts were noted on the image with maximal dimension, another "artifact-free" image demonstrating the lesion was chosen. The longest and its orthogonal dimensions were measured for each lesion on the image used for radiomics. A ROI was carefully drawn in the nodule or mass avoiding contact with its edges using a semi-automated process (Fig. 2). All lesions in the pleura, chest wall, mediastinum, and bones were excluded. For cavitary lesions, a threshold of greater than or equal to −50HU was applied to exclude air component and include only the solid portions of the lesions. Macroscopic calcifications if any were also excluded from the ROI.

The radiomic analysis comprised an initial image filtration step followed by quantification of texture within the lesion. The software highlights the size of the image features with the spatial scaling factor (SSF), which ranges between object radii of 0, 2, 3, 4, 5, and 6 mm. The filtration step is important to remove image heterogeneity that is due to photon noise and to highlight biologically important heterogeneity.[8,10]

2.4. Quantitative analysis

The following features were analyzed using the CTTA software. The mean value of pixels with in a ROI as mean Hounsfield unit (HU) values. The degree of dispersion of the HU value is expressed as standard deviation (SD). The SD was also expressed as logarithmic expression as "normalized SD": ln(SD)/ln(n) [= ln (SD)/ln (total pixels in ROI)] to represent the variability of region size. The number of pixels having positive value are expressed as percent positive pixels (PPP), and the mean positive pixels (MPP) indicates value greater than 0. The asymmetry of histogram is expressed as skewness, which can be negative or positive. The peak of distribution in histogram represents kurtosis and is a marker of vascularity and angiogenesis in the tumor. The kurtosis can be positive with peak greater than normal Gaussian distribution and negative with smaller peak. The complexity or heterogeneity of the structure is represented by entropy. The vascular structures are represented by kurtosis and distribution of nonvascular structures by skewness. The software enables selection of SSFs (0, 2, 3, 4, 5, 6) based on the lesion size to alter the filtration threshold.

2.5. Statistical analysis

The data were analyzed using SPSS 21 statistical software (IBM, Armonk, NY). A Pearson correlation analysis was performed to compare the radiomics features, clinical history (age, gender, and smoking) and standard imaging features among different cancer types (adenocarcinoma vs squamous cell carcinoma, EGFR positive vs EGFR wild-type). The P-value of less than or equal to 0.05 with a 95% confidence interval was considered significant.

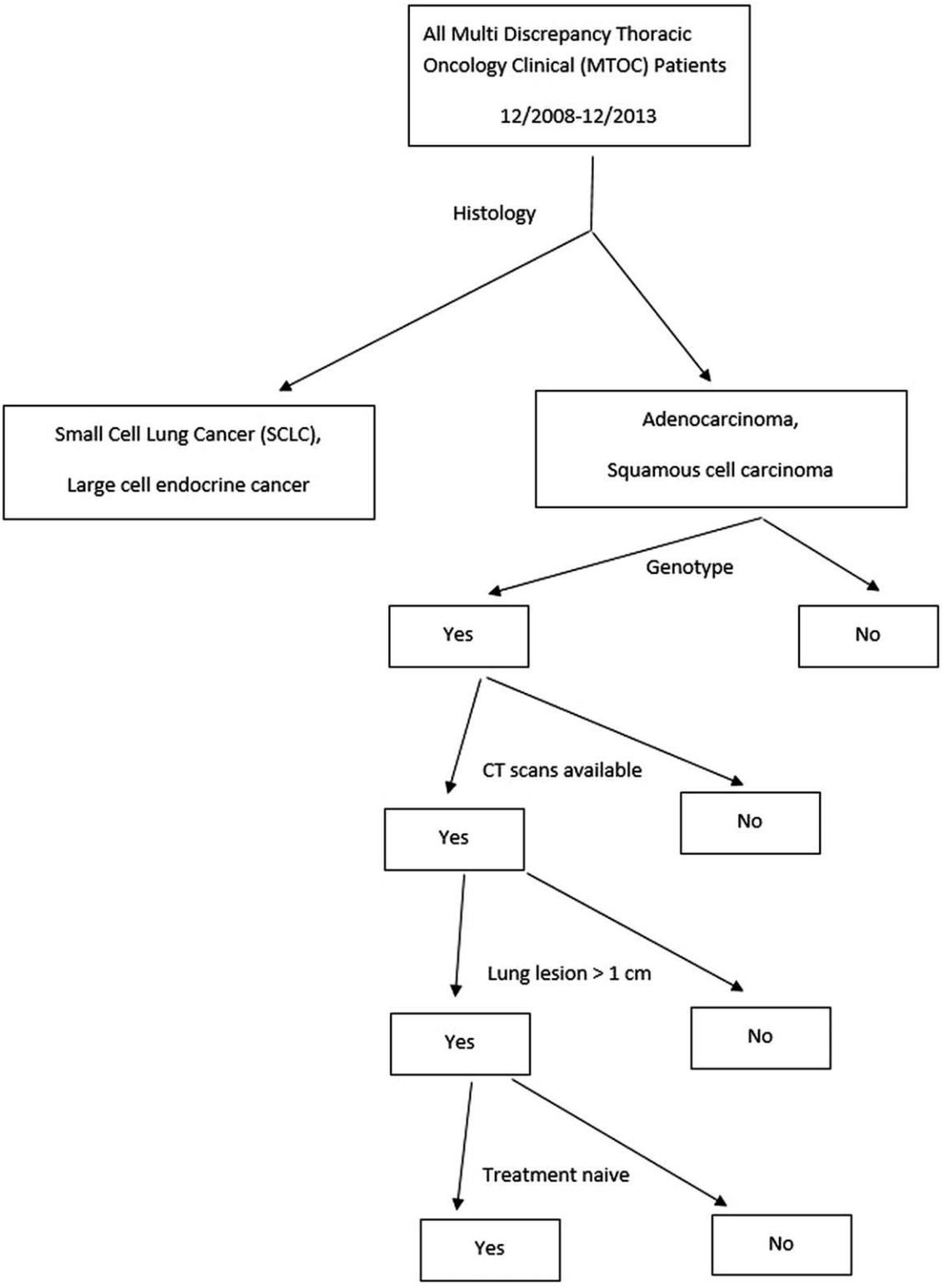

Figure 1. CONSORT diagram of patient selection.

Receiver operating characteristic (ROC) curves and area under curve (AUC) were generated for various radiomic features. The probability variables for each group and all group together were derived from the binary logistic regression analysis. The ROC curves were generated for these probability variables to see the difference.

3. Results

3.1. Differentiating adenocarcinoma vs squamous cell carcinoma

Pearson correlation analysis showed that there were significant correlations for 4/11 radiomics features (entropy, log of SD,

Table 1

Patient demographics.

Number of patients	93
Male: Female	43:50
Mean age ± SD	60 ± 11 years
Smoker: Nonsmoker	61:32
Number of nonsmall lung cancer (Adenocarcinoma: Squamous cell carcinoma)	94 (69:25)
Adenocarcinoma (EGFR mutation: Wild type)	69 (27:42)

EGFR = epidermal growth factor receptor, SD = standard deviation.

normalized SD, total), 2/3 clinical (smoking, gender), and 1/3 imaging features (LD) among adenocarcinomas and squamous cell carcinomas ($r = -0.354$–0.112, n = 94, $P = .033$ to $<.0001$). The separate ROC analysis showed that 3/11 radiomic features [AUC 0.686–0.744, $P = .006$ to $<.0001$], 1/3 clinical features (smoking) (AUC 0.732, $P = .001$), and 2/3 imaging features (LD and LPD) [AUC 0.646–0658, $P = .020$ to .032] were significantly different. The entropy was the only predictor on logistic regression ($P = .015$, Nagelkerke $R^2 = 0.30$). The entropy explained the 30% variance and correctly identified 79.0% carcinomas. After adding all major clinical, imaging and radiomics features in the regression model, the entropy ($P = .030$), gender ($P = .049$), and smoking ($P = .007$) were the predictors (Nagelkerke $R^2 = 0.57$). The entropy, gender, and smoking explained the 57% variance and correctly identified 85.0% carcinomas.

The separate ROC analysis (for radiomics, clinical, imaging) showed that 3/11 radiomic features (entropy, normalized SD, total) [AUC 0.686–0.744 (0.545–0.859), $P = .006$ to <0.0001], 1/3 clinical features (smoking) [AUC 0.732 (0.633–0.831), $P = .001$], and 2/3 imaging features (LD and LPD) [AUC 0.646–0658 (0.508–0.795), $P = .020$ to 0.032] were significantly different between the adenocarcinomas and squamous cell carcinomas (Table 2).

For probability variables (radiomics, clinical, imaging), ROC analysis showed higher AUC value for radiomics (AUC 0.800, $P < .0001$) than clinical (AUC 0.780, $P < .0001$) and imaging (AUC 0.694, $P = .004$) for differentiating adenocarcinomas and squamous cell carcinomas (Fig. 3). There was a significant difference among radiomics and imaging probability variables ($P < .0001$), and no difference among radiomics and clinical features ($P = .13$). The AUC value for combined (radiomics, clinical, imaging) probability variable was significantly higher (AUC 0.923, $P < .0001$) than separate probability variables.

3.2. Differentiating EGFR mutant vs EGFR wild-type adenocarcinoma

ROC analyses (for radiomics, clinical, imaging) showed that 2/11 radiomic features (kurtosis, K2) [AUC 0.656–0.713 (0.522–0.845), $P = .03$ to .003], 1/3 clinical features (smoking) [AUC 0.758 (0.638–0.878), $P < .0001$] (Fig. 4, Table 3), and none of the imaging features were significantly different between EGFR mutant and EGFR wild-type adenocarcinomas. The kurtosis was the only predictor for differentiating FGFR positive and EGFR wild-type adenocarcinomas on logistic regression analysis ($P = .037$, Nagelkerke $R^2 = 0.15$). The kurtosis explained the 15% variance and correctly identified 70.0% carcinomas.

For probability variables (radiomics, clinical, imaging), ROC analysis showed slightly higher AUC value for clinical (AUC 0.794, $P < .0001$) than radiomics (AUC 0.725, $P = .002$) and significantly higher than standard imaging (AUC 0.553, $P = .461$). The AUC value for combined (radiomics, clinical, imaging) probability variable was significantly higher (AUC 0.863, $P < .0001$) than separate probability variables.

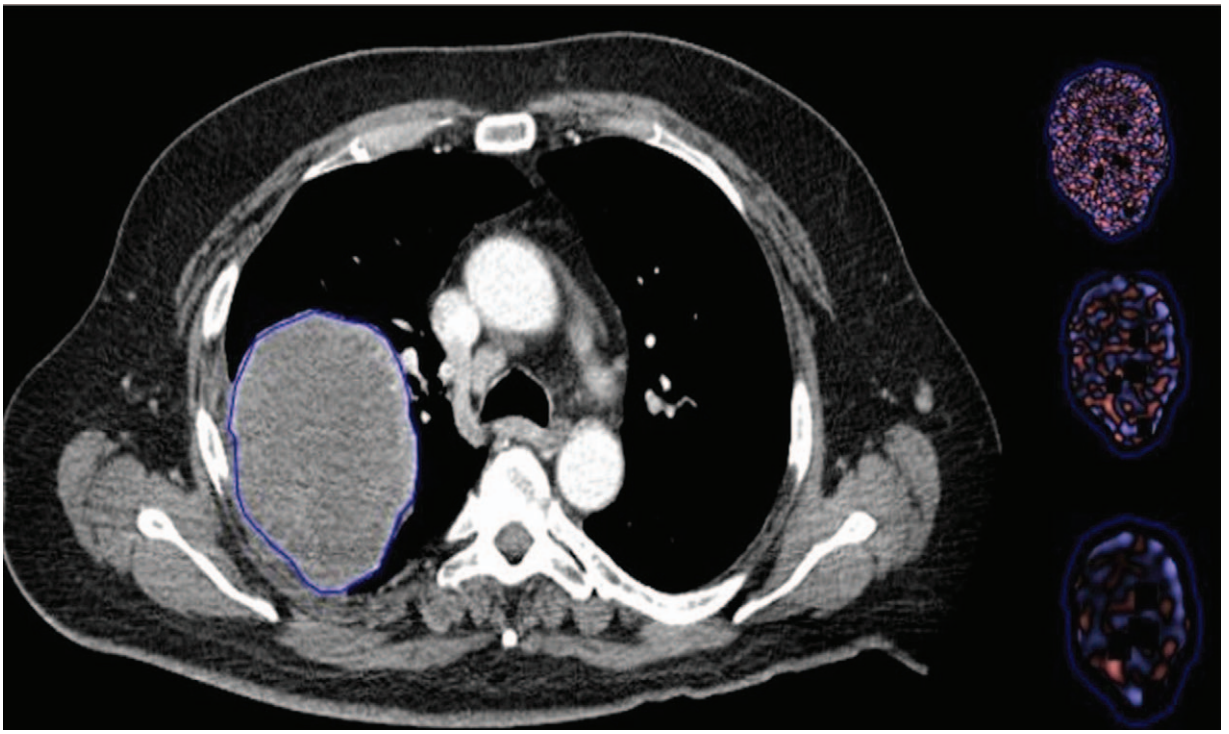

Figure 2. Segmentation of tumor by outlining the region of interest for extracting radiomic features.

Table 2

AUC values for radiomic features for adenocarcinoma vs squamous cell carcinoma.

Test Result Variable	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Clinical	0.780	0.047	0.000	0.687	0.873
Imaging	0.694	0.065	0.004	0.567	0.821
Radiomics	0.800	0.052	0.000	0.698	0.902
Clinical, imaging & Radiomics	0.923	0.028	0.000	0.868	0.978
Smoking (Clinical)	0.732	0.050	0.001	0.633	0.831
Gender (Clinical)	0.617	0.066	0.084	0.489	0.746
Longest Diameter (Imaging)	0.646	0.070	0.032	0.508	0.783
Density (Imaging)	0.464	0.065	0.593	0.336	0.592
Skewness (Radiomics)	0.512	0.071	0.864	0.373	0.650
Kurtosis (Radiomics)	0.577	0.074	0.253	0.432	0.722
Entropy (Radiomics)	0.744	0.059	0.000	0.629	0.859
Mean Positive Pixel (Radiomics)	0.571	0.067	0.295	0.439	0.703
Normalized SD (Radiomics)	0.686	0.072	0.006	0.545	0.827

AUC = area under curve, SD = standard deviation.

3.3. Differentiating EGFR wild-type adenocarcinoma vs squamous cell carcinoma

The separate ROC analysis (for radiomics, clinical, imaging) showed that only 3/11 radiomic features (entropy, normalized SD, total) [AUC 0.673–0.759 (0.532–0.879), $P = .018$ to $< .0001$], 1/3 of the clinical features, and 2/3 imaging features (LP and LPD) [AUC 0.652–667 (0.513–0.804), $P = .039$ to .023] were significantly different (Table 4).

For probability variables (radiomics, clinical, imaging), ROC analysis showed higher AUC value for radiomics (AUC 0.800, $P < .0001$) than clinical (AUC 0.676, $P = .017$) and standard imaging (AUC 0.708, $P < .0001$) for differentiating EGFR wild-type adenocarcinoma and squamous cell carcinomas. The AUC value for combined (radiomics, clinical, imaging) probability variable was significantly higher (AUC 0.890, $P < .0001$) than separate probability variables.

3.4. Differentiating EGFR mutant adenocarcinoma vs squamous cell carcinoma

ROC analyses (for radiomics, clinical, imaging) showed that 6/11 radiomic features (entropy, kurtosis, log of SD, normalized SD, total, K2) [AUC 0.664–0.721 (0.505–0.862), $P = .011$ to $< .0001$], 2/3 clinical features (smoking, gender) [AUC 0.672–0.889 (0.523–0.987), $P = .034$ to $< .0001$], and none of the imaging features were significantly different between the two groups.

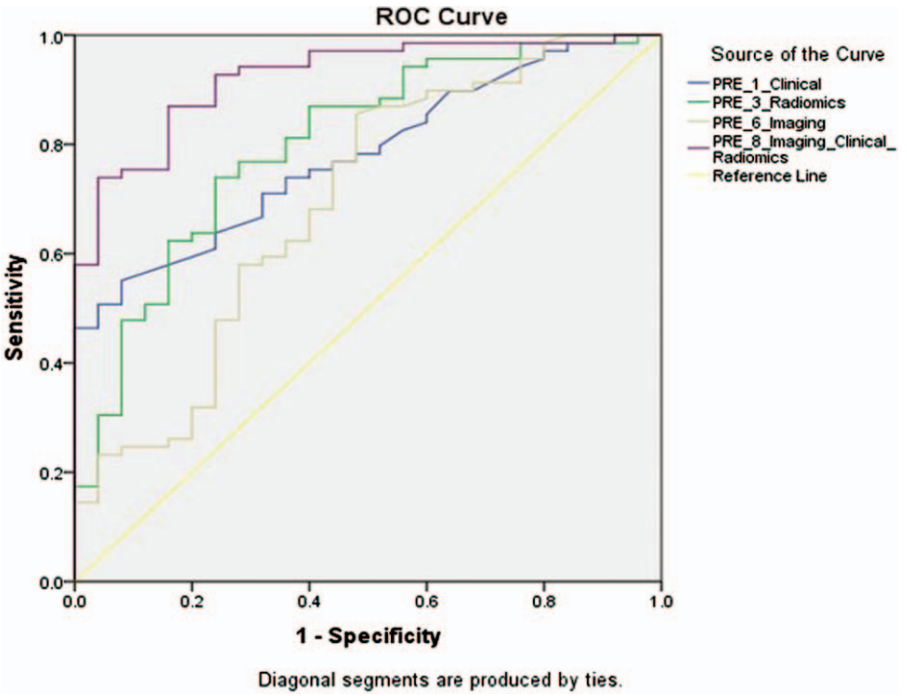

Figure 3. Receiver operating characteristic curves with AUC values for probability variables radiomics, clinical, imaging and combined differentiate adenocarcinomas and squamous cell carcinomas of the lung. The AUC value for combined (radiomics, clinical, imaging) probability variable was improved (AUC 0.923, $P < .0001$) and higher than separate probability variables. AUC = area under curve.

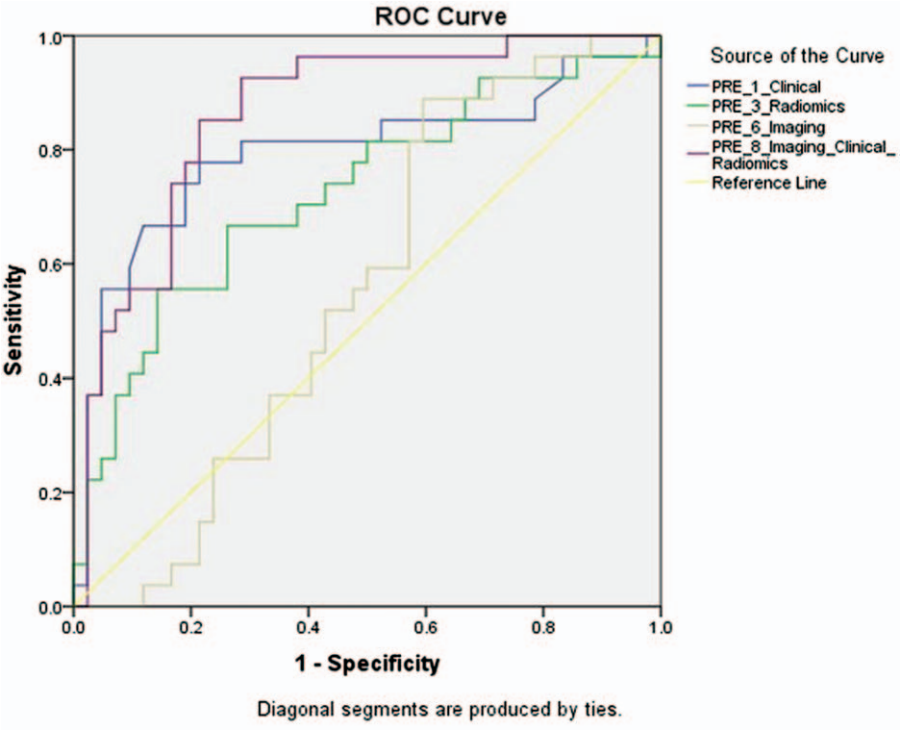

Figure 4. Receiver operating characteristic curves with AUC values for probability variables radiomics, clinical, imaging and combined differentiate fibroblast growth factor receptor positive and epidermal growth factor receptor wild-type adenocarcinomas of the lung. The AUC value for combined (radiomics, clinical, imaging) probability variable was improved (AUC 0.863, $P < .0001$) and higher than separate probability variables. AUC = area under curve.

ROC analysis for probability variables (radiomics, clinical, imaging) showed higher AUC value for clinical (most squamous cell carcinomas patients were male smokers and most EGFR mutant patients were female nonsmokers) (AUC 0.936, $P < .0001$) than radiomics (AUC 0.815, $P < .0001$) and imaging (AUC 0.666, $P = .040$) for differentiating EGFR mutant adenocarcinoma and squamous cell carcinomas (Table 5). The AUC value for combined (radiomics, clinical, imaging) probability variable was close to 1 ($P < .0001$) and significantly higher than independent probability variables.

4. Discussion

Recent studies on value of radiomics for predicting histopathology have not considered the significance of clinical history and imaging features which may have overemphasized the significance of radiomics.[19–21] We found that accuracy for radiomics features (AUC 0.800) was higher than clinical (AUC 0.780) and standard imaging (AUC 0.694) in differentiating adenocarcinoma from squamous cell carcinomas. After combining radiomics, clinical, and imaging features, the accuracy increased (AUC 0.923). Similar accuracy has been reported with few previous studies as well.[22–33] This is important due to increased overall

Table 3

AUC values for radiomic features for EGFR mutant vs EGFR wild type adenocarcinoma.

Test Result Variable	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Clinical	0.794	0.062	0.000	0.671	0.916
Imaging	0.553	0.069	0.461	0.418	0.688
Radiomics	0.725	0.065	0.002	0.598	0.852
Clinical, imaging & Radiomics	0.863	0.045	0.000	0.775	0.951
Smoking (Clinical)	0.758	0.061	0.000	0.638	0.878
Gender (Clinical)	0.590	0.070	0.210	0.453	0.727
Longest Diameter (Imaging)	0.558	0.070	0.417	0.422	0.695
Density (Imaging)	0.529	0.071	0.685	0.390	0.668
Skewness (Radiomics)	0.517	0.076	0.811	0.367	0.667
Kurtosis (Radiomics)	0.713	0.067	0.003	0.581	0.845
Entropy (Radiomics)	0.582	0.070	0.253	0.445	0.720
Mean Positive Pixel (Radiomics)	0.530	0.072	0.680	0.389	0.671
Normalized SD (Radiomics)	0.525	0.069	0.731	0.390	0.660

AUC = area under curve, EGFR = epidermal growth factor receptor, SD = standard deviation.

Table 4

AUC values for radiomic features for EGFR wild type adenocarcinoma vs squamous cell carcinoma.

Test Result Variable	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Clinical	0.676	0.065	0.017	0.549	0.802
Imaging	0.708	0.065	0.005	0.581	0.834
Radiomics	0.800	0.057	0.000	0.689	0.911
Clinical, imaging & Radiomics	0.890	0.039	0.000	0.814	0.965
Smoking (Clinical)	0.631	0.067	0.075	0.499	0.763
Gender (Clinical)	0.582	0.072	0.265	0.441	0.723
Longest Diameter (Imaging)	0.652	0.071	0.039	0.513	0.790
Density (Imaging)	0.548	0.072	0.517	0.407	0.688
Skewness (Radiomics)	0.517	0.078	0.815	0.365	0.669
Kurtosis (Radiomics)	0.519	0.082	0.800	0.359	0.678
Entropy (Radiomics)	0.759	0.061	0.000	0.639	0.879
Mean Positive Pixel (Radiomics)	0.581	0.072	0.270	0.439	0.723
Normalized SD (Radiomics)	0.673	0.072	0.018	0.532	0.814

AUC = area under curve, EGFR = epidermal growth factor receptor, SD = standard deviation.

incidence of adenocarcinoma and those in the central locations.[31] The radiomics (entropy) along with clinical (smoking and gender) and imaging features can correctly identify 85% adenocarcinoma and squamous cell carcinomas. This predictive value is even higher between EGFR mutant adenocarcinoma and squamous cell carcinomas, where the accuracy reaches up to 95%. This is mainly due to the fact that most squamous cell carcinomas patients were male smokers and most EGFR mutant patients were female nonsmokers. The combined radiomics (kurtosis), clinical (smoking), and imaging can correctly differentiate only up to 75% EGFR positive and EGFR wild-type adenocarcinoma. The relatively similar clinical features can partly explain the slightly decreased accuracy in this group, but even in this cohort addition of radiomics improved the accuracy of detection. In our study, none of the imaging features (density and size of the tumor) was a predictor for predicting the mutation status.

Weiss et al have reported that radiomics can differentiate NSCLC with KRAS mutation from pan-wildtype.[30] Our study demonstrates that radiomics features can detect differences in tumor heterogeneity (kurtosis and entropy) between adenocarcinomas with and without EGFR mutations and between adenocarcinomas and squamous cell carcinoma. The accuracy of differentiating these carcinomas improved with combining radiomics features along with clinical and imaging features noted in our study. Mak and Digumarthy et al reported lower 18F-fluorodeoxyglucose (FDG) uptake on PET examinations in patients with EGFR-positive adenocarcinoma compared to those with EGFR wild-type.[32] This supports observation from our study that lesions with higher tumor heterogeneity (entropy) have higher metabolic activity with greater 18-FDG uptake. Future studies will be needed to assess if radiomics in combination with the 18-FDG PET can enable better identification of EGFR mutation in patients with adenocarcinoma.

There was a significant difference in kurtosis between EGFR mutant and wild-type adenocarcinoma. Kurtosis is a marker for tumor angiogenesis, which in turn is an essential factor determining tumor aggressiveness and overall survival. Recent studies have reported improved survival upon addition of antiangiogenic therapy to patients with EGFR mutant adenocarcinomas receiving erlotinib.[27] This implies that as a surrogate marker of tumor angiogenesis, kurtosis might be useful in predicting and assessing response to antiangiogenic treatment in patients with EGFR mutant adenocarcinomas.

Table 5

AUC values for radiomic features for EGFR mutation adenocarcinoma vs squamous cell carcinoma.

Test Result Variable	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Clinical	0.936	0.034	0.000	0.870	1.000
Imaging	0.666	0.080	0.040	0.509	0.823
Radiomics	0.815	0.060	0.000	0.698	0.932
Clinical, imaging & Radiomics	1.000	0.000	0.000	1.000	1.000
Smoking (Clinical)	0.889	0.050	0.000	0.791	0.987
Gender (Clinical)	0.672	0.076	0.034	0.523	0.821
Longest Diameter (Imaging)	0.636	0.082	0.094	0.475	0.797
Density (Imaging)	0.519	0.081	0.819	0.360	0.797
Skewness (Radiomics)	0.503	0.081	0.971	0.344	0.662
Kurtosis (Radiomics)	0.669	0.077	0.037	0.518	0.820
Entropy (Radiomics)	0.721	0.072	0.006	0.579	0.862
Mean Positive Pixel (Radiomics)	0.556	0.081	0.492	0.397	0.714
Normalized SD (Radiomics)	0.706	0.080	0.011	0.548	0.863

AUC = area under curve, EGFR = epidermal growth factor receptor, SD = standard deviation.

Implication of our study is that by integrating clinical and imaging features with radiomics features can improve the distinction of tumors with squamous and adenocarcinoma histology and EGFR mutation. There are established guidelines for genetic testing for EGFR mutations and ALK rearrangements in patients with lung adenocarcinomas[7] but issues related to cost, invasiveness, and logistics of genetic profiling are not trivial. In this context, our study demonstrates that differentiation of lung cancer based on histology and/or presence of EGFR positive with noninvasive technique such as radiomics could assist the multidisciplinary lung cancer treatment team in the decision making regarding molecular testing options by triaging patients with lung cancer, separating them based on histology and likelihood of presence of mutation, in addition to other predictors of presence of genetic mutation such as smoking history and ethnicity. This could potentially affect therapy selection for patients with cancer and would allow the pathologist to standardize genetic profiling, prioritizing tests, saving tissues from biopsy sample allowing for a more thoughtful use of time, available sampled tissue and test and labor-related expenses. Although less likely in the developed nations, tools such as radiomics could be helpful in nations with limited resources. Indeed, such a cost-effective model based on radiomics has been reported for colorectal cancer surveillance[34] and in chemotherapy selection for patients with NSCLC.[35]

Our study has limitations. Since this was a retrospective study, we did not perform sample size estimation. However, our study has a higher sample size compared to some other publications on radiomics. Ng et al have reported that measurement of radiomics over the entire tumor volume is more accurate than from a single image with largest axial dimensions (as in our study) since the former better represents tumor heterogeneity and has a stronger correlation with overall survival compared to a single slice analysis.[36] However, most investigations have reported promising results for assessment of tumor biology[37,38] and prognosis[24,39] with radiomics from a single image with the largest cross-sectional tumor area. Another limitation of our study involves the exclusion of other histological types of primary and metastatic lung cancers. We also did not assess the effect of different forms of treatment on radiomics since all post-treatment CT examinations were excluded from our study.

In conclusion, radiomics features of NSCLC can help distinguish between adenocarcinoma and squamous cell carcinoma and EGFR positive and wild-type adenocarcinomas. Entropy and Kurtosis are the two most important distinguishing radiomic features. The radiomics evaluation adds incremental value to clinical history and standard imaging features in predicting histology and EGFR mutations and therefore, incorporating radiomics in CT evaluation of nonsmall lung cancers has potential clinical application.

Author contributions

Conceptualization: Subba R. Digumarthy, Lecia V. Sequist, Mannudeep K. Kalra.
Data curation: Atul M. Padole, Roberto Lo Gullo.
Formal analysis: Atul M. Padole, Roberto Lo Gullo.
Investigation: Atul M. Padole, Roberto Lo Gullo, Mannudeep K. Kalra.
Methodology: Subba R. Digumarthy, Lecia V. Sequist, Mannudeep K. Kalra.
Project administration: Subba R. Digumarthy.
Resources: Roberto Lo Gullo, Lecia V. Sequist.
Software: Atul M. Padole, Roberto Lo Gullo.
Supervision: Subba R. Digumarthy.
Validation: Subba R. Digumarthy.
Writing – original draft: Subba R. Digumarthy, Atul M. Padole, Roberto Lo Gullo, Mannudeep K. Kalra.
Writing – review & editing: Subba R. Digumarthy, Atul M. Padole, Lecia V. Sequist, Mannudeep K. Kalra.

References

[1] American Cancer SocietyCancer Facts and Figures 2017. 2017;American Cancer Society, Atlanta, GA:Available at: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf. Accessed on March 17, 2017

[2] Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small cell lung cancer with mutated EGFR. N Engl J Med 2010;362:2380–8.

[3] Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129–39.

[4] Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.

[5] Pao W, Miller V, Zakowski M, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci USA 2004;101:13306–11.

[6] Cardarella S, Johnson BE. The impact of genomic changes on treatment of lung cancer. Am J Respir Crit Care Med 2013;188:770–5.

[7] National Comprehensive Cancer Network (NCCN) Clinical practice guidelines in oncology. Available at: https://www.nccn.org/professionals/physician_gls/f_guidelines.asp. Accessed on March 17, 2017.

[8] Ganeshan B, Panayiotou E, Burnand K, et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol 2012;22:796–802.

[9] Sacconi B, Anzidei M, Leonardi A, et al. Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with EGFR mutations and survival rates. Clin Radiol 2017;72:443–50.

[10] Ganeshan B, Goh V, Mandeville HC, et al. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. Radiology 2013;266:326–36.

[11] McNitt-Gray MF, Wyckoff N, Sayre JW, et al. The effects of the co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. Comput Med Imaging Graph 1999;23:339–48.

[12] Kido S, Kuriyama K, Higashiyama M, et al. Fractal analysis of small peripheral pulmonary nodules in thin-section CT: evaluation of the lung-nodule interfaces. J Comput Assist Tomogr 2002;26:573–8.

[13] Kido S, Kuriyama K, Higashiyama M, et al. Fractal analysis of internal and peripheral textures of small peripheral bronchogenic carcinomas in thin-section computed tomography: comparison of bronchioalveolar cell carcinomas with non-bronchioalveolar cell carcinomas. J Comput Assist Tomogr 2003;27:56–61.

[14] Farag A, Elhabian S, Graham J, et al. Toward precise pulmonary nodule descriptors for nodule type classification. Med Image Comput Assist Interv 2010;13:626–33.

[15] Way TW, Sahiner B, Chan HP, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features. Med Phys 2009;36:3086–98.

[16] Digumarthy SR, De Man R, Canellas R, et al. Multifactorial analysis of mortality in screening detected lung cancer. J Oncol 2018;2018:

[17] McNitt-Gray MF, Hart EM, Wyckoff N, et al. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. Med Phys 1999;26:880–8.

[18] Digumarthy SR, Padole AM, Lo Gullo R, et al. CT texture analysis of histologically proven benign and malignant lung lesions. Medicine (Baltimore) 2018;97:e11172.

[19] Ferreira Junior JR, Koenigkam-Santos M, Cipriano FEG, et al. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. Comput Methods Programs Biomed 2018;159:23–30.

[20] Wu W, Parmar C, Grossmann P, et al. Exploratory study to identify radiomics classifiers for lung cancer histology. Front Oncol 2016;30: 6:71.
[21] Zhu X, Dong D, Chen Z, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. Eur Radiol 2018;28:2772–8.
[22] Shac YZ, Liu LZ, Wang FH, et al. Quantitative evaluation of CT-MRI images of various tumors with expansive or infiltrative growth pattern. Zhonghua Yi Xue Za Zhi 2008;88:1503–6.
[23] Skogen K, Ganeshan B, Good C, et al. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. J Neurooncol 2013;111:213–9.
[24] Win T, Miles KA, James SM, et al. Tumor heterogeneity and permeability as measured on the CT component of PET/CT predicts survival in patients with non-small cell lung cancer. Clin Cancer Res 2013;19: 3591–9.
[25] Ganeshan B, Skogen K, Pressney I, et al. Tumor heterogeneity in esophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumor metabolism, stage, and survival. Clin Radiol 2012;67:157–64.
[26] Ng F, Ganeshan B, Kozarski R, et al. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5 year survival. Radiology 2013;266:177–84.
[27] Goh V, Ganeshan B, Nathan P, et al. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. Radiology 2011;261:165–71.
[28] Barbone F, Bovenzi M, Cavallieri F, et al. Cigarette smoking and histologic type of lung cancer in men. Chest 1997;112:1474–9.
[29] Khuder SA, Dayal HH, Mutgi AB, Willey JC, Dayal G. Effect of cigarette smoking on major histological types of lung cancer in men. Lung Cancer 1998;22:15–21.
[30] Weiss GJ, Ganeshan B, Miles KA, et al. Noninvasive image texture analysis differentiates K-ras mutation from pan-wild type NSCLC and is prognostic. PLoS One 2014;9:e100244.
[31] Moon Y, Lee KY, Sung SW, et al. Differing histopathology and prognosis in pulmonary adenocarcinoma at central and peripheral locations. J Thorac Dis 2016;8:169–77.
[32] Mak RH, Digumarthy SR, Muzikansky A, et al. Role of 18F-fluorodeoxyglucose positron emission tomography in predicting epidermal growth factor receptor mutations in non-small cell lung cancer. Oncologist 2011;16:319–26.
[33] Li YX, Wang JL, Gao M, et al. Celecoxib-erlotinib combination delays growth and inhibits angiogenesis in EGFR-mutated lung cancer. Am J Cancer Res 2016;6:1494–510.
[34] Miles KA, Ganeshan B. Potential for texture analysis of hepatic CT to cost-effectively modify post-operative surveillance of patients with colorectal cancer. Presented at the European Congress of Radiology 2011, Vienna, Austria. Available at: http://posterng.netkey.at/esr/viewing/index.php?module=viewing_poster&doi=10.1594/ecr2011/C-0685. Accessed on March 17, 2017.
[35] Miles KA, Ganeshan B. Selection of patients with advanced non small cell lung cancer for chemotherapy: potential cost-effectiveness of CT texture analysis. Presented at European Congress of Radiology 2012, Vienna, Austria. Available at: http://posterng.netkey.at/esr/viewing/index.php?module=viewing_poster&task=viewsection&ti=352155. Accessed on March 17, 2017.
[36] Ng F, Kozarski R, Ganeshan B, et al. 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 2013;82:342–8.
[37] Ganeshan B, Abaleke S, Young R, et al. 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 2010;10:137–43.
[38] Ganeshan B, Miles KA, Young RC, et al. In search of biological correlates for liver texture on portal-phase CT. Acad Radiol 2007;14:1058–68.
[39] Ganeshan B, Miles KA, Young RC, et al. Hepatic enhancement in colorectal cancer: texture analysis correlates with hepatic haemodynamics and patient survival. Acad Radiol 2007;14:1520–30.