Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Oncologia, Ematologia e Patologia

Ciclo XXXIII

Settore Concorsuale: 06/I1

Settore Scientifico Disciplinare: MED 36

RADIOMICS OF NON SMALL CELL LUNG CANCER: ASSOCIATION BETWEEN RADIOMIC FEATURES, LYMPH NODAL STATUS AND PROGNOSIS

Presentata da: dott.ssa Stefania Maria Rita Rizzo

Coordinatore Dottorato Prof.ssa Manuela Ferracin

Supervisore Prof. Alessio Giuseppe Morganti

Esame finale anno 2021

Radiomics of non small cell lung cancer: association between radiomics features, lymph nodal status and prognosis.

Summary

1. Abstract

Radiomics is an emerging translational field of research aiming to extract mineable high-dimensional data from clinical images able to offer information about prognosis of cancer patients. The radiomics process relies on a multi-step path that ends in the construction of a predictive model, tailored on specific outcomes. The main steps of the radiomics process are: image acquisition and reconstruction, segmentation, features extraction, model building. Each of these steps shows its own challenges to make the final model robust and reliable.

Patients with Non-small cell lung cancer (NSCLC) have baseline computed tomography (CT) and/or fluorodeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) imaging for diagnosis and staging.

The aim of this study was to evaluate whether a model based on radiomic and clinical features may be associated with lymph node (LN) status and overall survival (OS) in NSCLC patients. Patients with a pathological stage up to T3N1 were retrospectively selected and divided into training and validation sets. For the prediction of positive LNs and OS, the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model was used; univariable and multivariable logistic regression analysis assessed the association of clinical-radiomic variables and endpoints. All tests were repeated after dividing the groups according to the CT reconstruction algorithm. p-values < 0.05 were considered significant.

270 patients were included and divided into training (n = 180) and validation sets (n = 90). Transfissural extension was significantly associated with positive LNs. For OS prediction, high- and low-risk groups were different according to the radiomics score, also after dividing the two groups according to reconstruction algorithms. In conclusion, a combined clinical–radiomics model was not superior to a single clinical or single radiomics model to predict positive LNs. A radiomics model was able to separate high-risk and low-risk patients for OS.

2. Introduction

Radiomics is an emerging translational field of research aiming to extract mineable high-dimensional data from clinical images. The radiomic process is a multi-step process with definable inputs and outputs, such as image acquisition and reconstruction, image segmentation, features extraction and qualification, analysis, and model building [1]. Each step needs careful evaluation for the construction of robust and reliable models to be transferred into clinical practice for the purposes of prognosis, non-invasive disease tracking, and evaluation of disease response to treatment. Reproducibility and clinical value of parameters should be tested with internal cross-validation and then validated on independent external cohorts.

Studies published in the last few years have demonstrated that quantitative imaging features, extracted from routine medical imaging, have prognostic value and may predict clinical outcomes or allow for treatment monitoring in different cancer types.

Non-small cell lung cancer (NSCLC) accounts for more than 85 % of all lung cancer cases, with the main histological subtype consisting of adenocarcinoma.

Patients with NSCLC have baseline computed tomography (CT) and/or fluorodeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) imaging for diagnosis and staging. Regular follow-up imaging is also performed to evaluate treatment response and monitor for recurrence.

In clinical practice, the radiologic evaluation of treatment response largely relies on tumour size, supplemented with a qualitative assessment of other tumour characteristics such as homogeneity and shape. From a quantitative viewpoint, this approach is not comprehensive of a substantial amount of information within the medical image. The radiomics approach has the potential to identify quantitative markers of treatment response earlier in the course of treatment. This can enable treatment to be adapted, intensified or altered earlier in the course of disease in order to improve patient outcomes [2].

3. **Radiomics: a multi-step process**

The radiomics process can be divided into the following distinct steps: image acquisition and reconstruction; image segmentation; features extraction and qualification; analysis and model building [3,4].

CT Image acquisition and reconstruction

Routine clinical imaging shows a wide variation in acquisition parameters, such as image resolution (pixel size, matrix size, slice thickness), contrast-enhancement, as shown in **Figure 1**, energy (kV), tube current (mA) and exposure (mAs) (**Figure 2**).

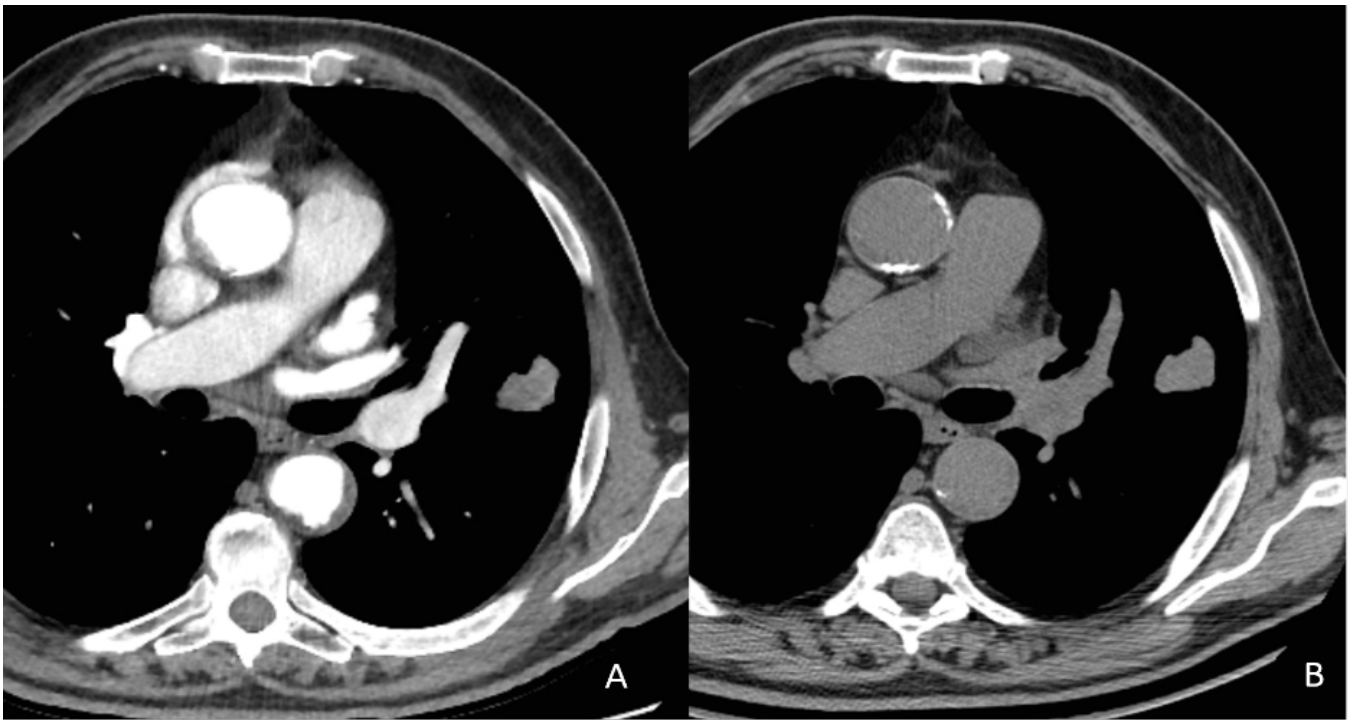


Figure 1.

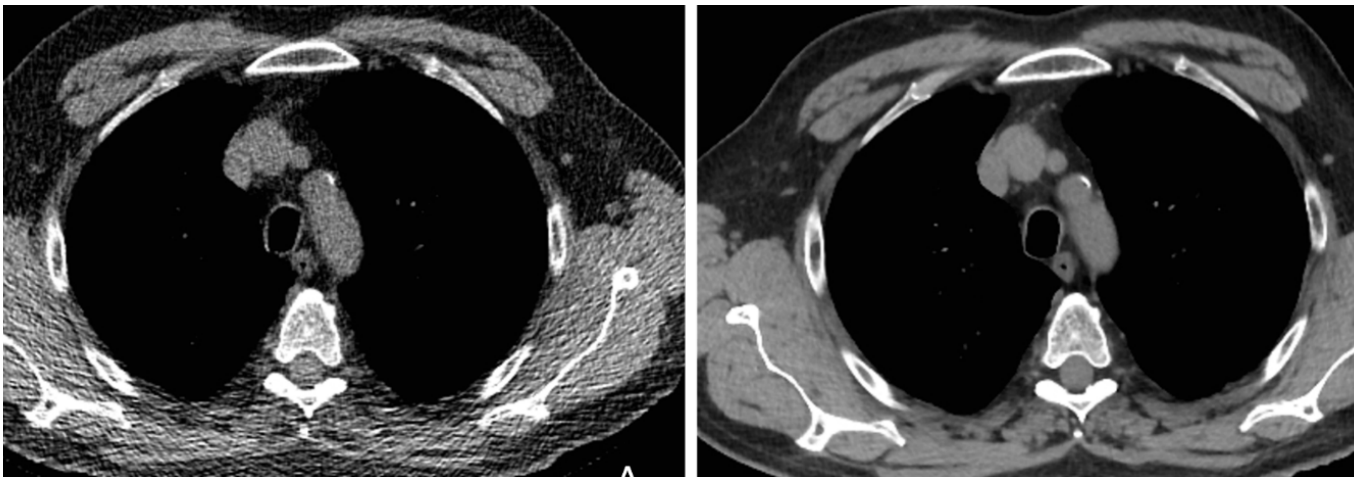


Figure 2.

Different combinations of these parameters may hinder comparisons of radiomic features obtained across institutions with different scanners and different patient populations, and within the same institution using different acquisition protocols [3]. Indeed, all the variables occurring during image acquisition and reconstruction affect image noise, and hence its texture and the value of radiomic features. As a result, differences in features extracted may not be due to different biology, but to different image parameters.

Standard CT phantoms, like the one proposed by the American Association of Physicists in Medicine, aim to evaluate imaging performance. There are sections for evaluating the true slice thickness and variation of Hounsfield Units (HU), looking at small variations in density (low contrast detectability) and assessing spatial resolution and HU variations showing high contrast detectability within a region of uniform medium. Such phantoms serve to assess how far image quality depends on the imaging technique. For example, decreasing the slice thickness reduces the photon statistics within a slice (unless mAs or kVp are increased accordingly), thereby increasing image noise. The axial field of view and reconstruction matrix size determine the pixel size and hence the spatial sampling in the axial plane, which has an impact on the description of heterogeneity. Reducing the pixel size increases image noise when the other parameters are kept unchanged, but increases spatial resolution. Pitch is a manufacturer-dependent parameter, so noise can only be compared between different scanners and vendors on images acquired using axial (not helical or spiral) acquisitions. Likewise, clinical conditions, such as the presence of artifacts due to a metallic prosthesis, may significantly affect image quality and impair quantitative analysis [5]. Furthermore, HU may also vary with the reconstruction algorithm [6] or scanner calibration. More sophisticated phantoms may therefore be required to match the effects of acquisition settings and reconstruction algorithms on radiomic features. For example, the Credence Cartridge Radiomics phantom, including different cartridges each exhibiting a different texture, was developed to test interscanner, intrascanner and multicentre variability in CT radiomic features [7], and the effect of different acquisition and reconstruction settings on feature robustness [8]. Many authors have investigated radiomic feature robustness and stability directly on clinical images by undertaking test-retest studies [9], or comparing the results obtained with different imaging settings or image processing algorithms [10]. All these studies call for dedicated investigations to select radiomic features with sufficient dynamic range among patients, intrapatient reproducibility and low sensitivity to image acquisition and reconstruction protocols [11].

Image segmentation.

Segmentation is a critical step of the radiomics process because data are extracted from the segmented volumes. This is challenging because many tumours show unclear borders. It is also contentious because there is no consensus on the need to seek either the ground truth or reproducibility of image segmentation [1]. Indeed, many authors consider manual segmentation by expert readers the ground truth despite high inter-reader variability. This method is also labour intensive (**Figure 3**) and hence not always feasible for radiomics analysis requiring very large data sets [3].

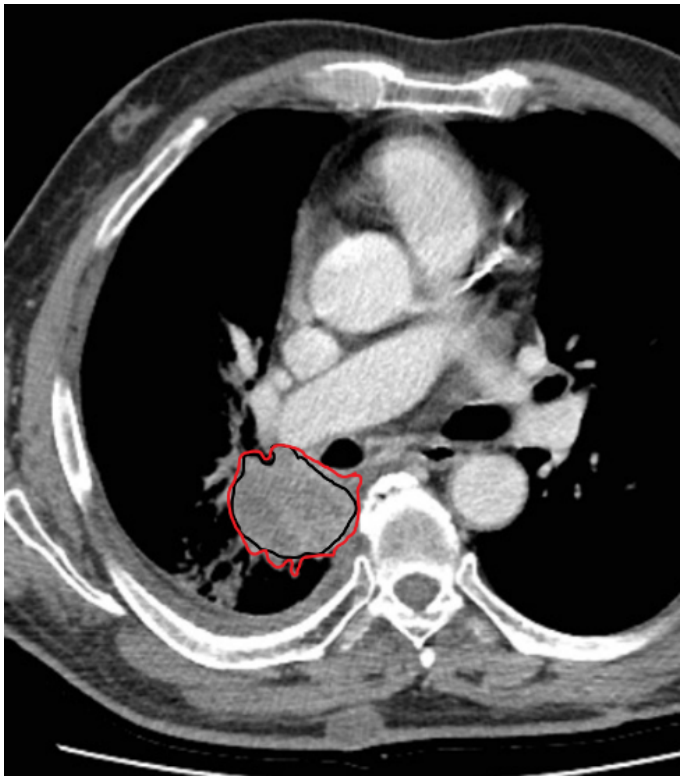


Figure 3

Many automatic and semiautomatic segmentation methods have been developed across imaging modalities and different anatomical regions. Common requirements include maximum automaticity with minimum operator interaction, time efficiency, accuracy and boundary reproducibility. Many segmentation algorithms rely on region-growing methods that require an operator to select a seed point within the VOI [12]. These methods work well for relatively homogeneous lesions, but show many limits and the need for intensive user correction for inhomogenous lesions. For example, most Stage I and Stage II lung tumours present as homogenous high-intensity lesions on a background of low-intensity lung parenchyma [13,14] and therefore can be automatically segmented with high reproducibility and accuracy. However, partially solid, ground glass opacities, nodules attached to vessels and nodules attached to the pleural wall or mediastinum remain difficult to segment automatically and show low reproducibility, especially for Stage III and Stage IV disease [15].

Other segmentation algorithms include level set methods that represent a contour as the zero level set of a higher dimensional function (level set function) then the method formulates the motion of the contour as the evolution of the level set function [16]. Graph cut methods construct an image-based graph and accomplish a globally optimal solution of energy minimization functions. Since graph cut algorithms try to identify a global optimum, they are

computationally expensive [17] and may lead to over-segmentation [18]. Active contour (snake) algorithms work like a stretched elastic band. The starting points are drawn around the lesion to then move through an iterative process to a point with the lowest energy function value. These algorithms may lead the snake to undesired locations because they depend on an optimal starting point and are sensitive to noise [19]. Semi-automatic segmentation algorithms, like livewires, do a graph search through local active contour analysis, while their cost function is minimized using dynamic programming. Nonetheless, the semi-automaticity still requires multiple human interactions [20].

As the above-mentioned examples show, there is still no universal segmentation algorithm that can work for all medical image applications, and some features may show stability and reproducibility using one segmentation method, but not another.

Feature extraction and qualification.

Classic semantic features are those commonly used in the radiology lexicon to describe a lesion [21]. Agnostic features are descriptors mathematically extracted by different software [8], with different levels of complexity, such as features describing lesion shape, intensity, texture and wavelet. First order statistics features describe the distribution of individual voxel values without concern for spatial relationships. These are generally histogram-based methods and reduce a region of interest to single values for mean, median, maximum, minimum, and uniformity or randomness (entropy) of the intensities on the image, as well as the skewness (asymmetry) and kurtosis (flatness) of the histogram of values. Shape features describe the shape of the traced ROI and its geometric properties, such as volume, maximum diameter or the three diameters along the three orthogonal directions, maximum surface, tumour compactness and sphericity. For example, the surface-to-volume ratio of a spiculated tumour will show a higher value than that of a round tumour of similar volume.

Second order statistics features generally describe textural features [22,23], meaning the statistical inter-relationships between voxels with similar (or dissimilar) values and take into account the spatial arrangement of the values. Texture analyses in radiomics can readily provide a measure of intratumoral heterogeneity. Examples of these features belong to the category of the grey level co-occurrence matrix (GLCM) and grey level run length matrix (GLRLM). GLCM quantifies the incidence of voxels with the same intensities by a predetermined distance and angle within the 3D structure. Within this category, features extracted may describe autocorrelation, contrast, correlation, cluster prominence, cluster shade, cluster tendency, dissimilarity, energy, homogeneity, maximum probability, sum of squares, sum average, sum variance, sum entropy or difference entropy. GLRLM indicates the length, in number of pixels, of consecutive pixels with the same grey level value on a raw. Within this category, features extracted may describe short and long run emphasis, grey level non-uniformity, run length non-uniformity, run percentage, low grey level run emphasis and high grey level run emphasis [24].

As a result, extraction of textural features may generate hundreds of variables, some of which may be redundant and require co-variance evaluation [14].

Analysis and model building.

Not all extracted features are useful for a particular task: most of them are in fact redundant. Therefore, it is crucial to select information useful for a specific purpose to ensure good radiomics performance. Initial efforts should focus on identifying appropriate findings with a potential clinical application.

Radiomics analysis usually includes two main steps: 1) dimensionality reduction and feature selection, usually obtained via unsupervised approaches, and 2) association analysis with one or more specific outcome(s) via supervised approaches. Different methods of dimensionality reduction/feature selection and model classification have been compared. The two most commonly used unsupervised approaches in radiomics studies are cluster analysis [11,25] and principal component analysis (PCA) [26,27]. Cluster analysis aims to create groups of similar features (clusters) with high intracluster redundancy and low intercluster correlation. A single feature could then be selected by each cluster as representative and used in the following association analysis with the outcome(s) [11]. PCA aims to create a smaller set of maximally uncorrelated variables from a large set of correlated variables, and to explain as much as possible of the total variation in the data set with the fewest possible principal components. All selected features considered reproducible, informative and non-redundant can then be associated with specific outcome(s) and/or phenotype(s). An important caveat for univariate analysis is multiple testing, which is often not accounted for in radiomics studies. The commonest way to overcome the multiple testing problem is to use Bonferroni or the less conservative false discovery rate corrections.

Supervised multivariate analysis consists of building a mathematical model to predict an outcome or response variable. The many different analysis approaches depend on the purpose of the study and the outcome category, ranging from statistical methods to data mining/machine learning approaches, such as random forests, to neural networks, linear regression, logistic regression, least absolute shrinkage and selection operator (LASSO), and Cox proportional hazards regression [28-32]. Unquestionably, the stability and reproducibility of the model must be assessed before applying a predictive model in a clinical setting. Indeed, it is well known that model fitting is optimal in the training set used to build the model, while validation in an external cohort provides more reliable fitting estimates [32]. The first step in model validation is internal cross-validation, but the reference standard to assess the reproducibility of a model is validation with prospectively collected independent cohorts, ideally within clinical trials [4].

Finally, it is important to establish whether potential radiomic predictors significantly increase the accuracy of existing clinical models. Addressing the influence of patient parameters in radiomics research using epidemiologic and biostatistical approaches will minimize spurious relationships and lead to more accurate and reproducible results [33].

4. Radiomics and radiogenomics of lung cancer: first evidences

Recent practice guidelines in oncology and pathology recommend that all locally advanced and metastatic NSCLC undergo testing for the most common targetable genetic abnormalities, such as epidermal growth factor receptor gene (EGFR) mutations, anaplastic lymphoma kinase gene (ALK) rearrangements, and non-targetable such as Kirsten rat sarcoma viral oncogene homolog (KRAS) mutations.

In order to assess the association between CT features and EGFR, ALK, KRAS mutations in NSCLC, we retrospectively selected patients who underwent chest CT and testing for the above gene mutations. Qualitative evaluation of CTs included: lobe; lesion diameter; shape; margins; ground-glass opacity; density; cavitation; air bronchogram; pleural thickening; intratumoral necrosis; nodules in tumour lobe; nodules in non-tumour lobes; pleural retraction; location; calcifications; emphysema; fibrosis; pleural contact; pleural effusion. Statistical analysis was performed to assess association of features with each gene mutation. ROC curves for gene mutations were drawn; the corresponding area under

the curve was calculated. The final cohort consisted of 285 patients, of which 60/280 (21.43 %) were positive for EGFR mutation; 31/270 (11.48 %) for ALK rearrangement;

64/240 (26.67 %) for KRAS mutation. EGFR mutation was associated with air bronchogram, pleural retraction, females, non-smokers, small lesion size, and absence of fibrosis. ALK rearrangements were associated with age and pleural effusion. KRAS mutation was associated with round shape, nodules in non-tumour lobes, and smoking [34].

5. **External validation**

As mentioned in the radiomics section, the validation of a model is able to test its reproducibility and reobustness. In order to validate the abovementioned associations between radiological features and clinical features with Epidermal Growth Factor Receptor (EGFR)/ Kirsten RAt Sarcoma (KRAS) alterations in an independent group of patients with Non-Small Cell Lung Cancer (NSCLC), we performed an external validation. A total of 122 patients with NSCLC tested for EGFR/KRAS alterations, retrospectrively collected at the Department of Radiation Oncology (MAASTRO), GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, were included. Univariate analysis of clinical and radiological features were performed to look at the associations of the studied features with EGFR/KRAS alterations. Previously calculated composite model parameters for each gene alteration prediction were applied to this validation cohort. ROC (Receiver Operating Characteristic) curves were drawn using the previously validated composite models, and also for each significant individual characteristic of the previous training cohort model.

At univariate analysis, EGFR+ confirmed an association with an internal air bronchogram, pleural retraction, emphysema and lack of smoking; KRAS+ with round shape, emphysema and smoking. The AUC (95%CI) in the new cohort was confirmed to be high for EGFR+ prediction, with a value of: 0.82 (0.69-0.95) vs. 0.82 in the previous cohort, whereas it was smaller for KRAS+ prediction, with a value of 0.60 (0.48-0.72) vs. 0.67 in the previous cohort. Looking at single features in the new cohort, we found that the AUC for the models including only smoking was similar to that of the full model (including radiological and clinical features) for both gene alterations [35].

Since it was demonstrated that CT radiomic features have prognostic information in stage I-III NSCLC patients [36], we also aimed to validate a prognostic radiomic signature in stage IV adenocarcinoma patients undergoing chemotherapy. For this purpose, we selected patients from two datasets of chemo-naive stage IV adenocarcinoma patients: the first one (n=285) was the one dataset used for the radiogenomics study [34]; the second (n=223) was the one of a multicenter clinical trial [37]. In total CT scans from 195 patients were eligible for analysis. Patients having a

prognostic index (PI) lower than the signature median (n=92) had a significantly better OS than patients with a PI higher than the

median (n=103, HR 1.445, 95% CI 1.07–1.95, p=0.02, c-index 0.576, 95% CI 0.527–0.624). Thus we showed that the radiomic signature, derived from daily practice CT scans, has prognostic value for stage IV NSCLC, although the performance was lower than the one previously described for stage I-III NSCLC stages [38].

6. Association of radiomics features with lymph nodal metastases and overall survival in patients with lung cancer staged up to pT3N1

Objectives.

The objective of this doctorate study was to evaluate whether a model based on quantitative CT radiomic and clinical features of lung cancer patients may be associated with LN status and with overall survival (OS).

Methods.

Patients' selection.

The Institutional Review Board approved this retrospective study (UID 2172), with waiver of informed consent. The study population was retrospectively selected from a database of patients with lung cancer staged up to T3 N1, operated on between 01/01/2012 and 01/08/2016. Preoperative staging was performed by whole body CT and Fluorodeoxyglucose Positron Emission Tomography (FDG PET) scan; in the event of suspected cN2 disease, preoperative endobronchial ultrasound trans bronchial needle aspiration (EBUS TBNA) was performed to rule out or to confirm lymph node involvement. Inclusion criteria were: availability of pre-surgical CT after contrast medium injection, with helical mode, 120 kVp, 2.5 mm slice thickness, 2.5 mm spacing; reconstruction performed with "Body" filter and "Standard" convolution kernel; surgery performed at our Institution; availability of histology, pathological node status (pN0; pN1), grading. Exclusion criteria were: CT performed with parameters different to those specified above; pre-operative chemotherapy.

CT imaging.

CT scans were randomly performed on the following CT scanners: Lightspeed Ultra, Lightspeed 16; Optima 660; Discovery CT750 HD (all GE Healthcare, Milwaukee, WI, USA). All scans were acquired in the portal venous phase and segmentations were performed on that series. All scanners implemented current modulation; Light speed 16 and Light Speed Ultra were equipped only with longitudinal z-axis modulation, while Optima 660 and Discovery CT750 also had angular xy modulation.

Clinical and radiological data recording.

The following clinical and radiological data were recorded: age; gender; grading; side; site (upper, medium, lower, mixed); axial nodule size; pT; pN; CT scanner; CT reconstruction algorithm (Filtered Back Projection, FBP, or Iterative Reconstruction, IR); exposure (mAs); status (alive or deceased); date of last contact or death.

Lesion segmentation.

On each axial CT image including the lung nodule, a radiologist traced free-hand 2D regions of interest, resulting in a 3D volume of interest (VOI) that was converted into a DICOM RT Structure format (AW Server 2.0 workstation, GE Healthcare).

Radiomic features extraction.

CT images and VOI were imported into the IBEX V 1.0 β tool (Imaging Biomarker Explorer Software [24]) for the extraction of radiomic features. The "Resample_VoxelSize" pre-processing was used to resample images to the same pixel size, chosen as the value most frequently observed in the dataset (0.7 × 0.7 mm2, the values in patient images ranging from 0.59 × 0.59 mm2 to 0.98 × 0.98mm2). "Threshold_Image_MaskXF" pre-processing, that excludes voxels on the ROI edge having intensity outside a user-defined range, was used to exclude parenchyma voxels erroneously included during manual delineation (threshold: −400 Hounsfield Units). All the features available in IBEX for the following categories were extracted: Shape, Intensity Histogram (IH), Intensity Direct (ID), Grey Level Co-occurrence Matrix (GLCM) 2.5, Grey Level Run Length Matrix (GLRLM) 2.5, Neighbour Intensity Difference (NID) 2.5, and Gradient Orient Histogram (GOH). GLCM and GLRLM were calculated comparing voxel intensities along 8 different directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) and with three different offsets between voxels (1, 4 and 7 voxels). The full list of features calculated by IBEX for each category is reported in [24], also including references for feature definition and formula. The [0-4096] HU interval was discretized in 256 bins for GLCM2.5 calculation, and 64 bins for the other categories.

Statistical analysis

Repeatability and reproducibility: In order to select the most stable and reproducible radiomic features, we first selected stable features with Overall Concordance Correlation Coefficient (OCCC) >0.95 based on the test-retest experiments on a phantom, where identical measures in different tests are expected. We then used one-way ANOVA to assess the features' reproducibility according to contrast medium, scanner, reconstruction algorithm, and exposure. Features with significantly different means according to at least one of the abovementioned parameters were considered not robust and excluded.

Training and validation datasets: We randomly selected 2/3 of the patients as a training dataset, and 1/3 as a validation dataset. This allocation proportion is commonly used to ensure the model is trained on a sufficient number of patients, in order to obtain precise parameter estimates. From a previously published paper, this commonly used strategy was demonstrated to be close to optimal for reasonably sized datasets (n ≥ 100) with strong signals (i.e., 85% or greater full dataset accuracy)

Positive LN prediction: The Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model was used to select radiomic features associated with positive LNs. We combined selected features into a radiomics score. We assessed the predictive accuracy of the radiomics score for positive LNs by calculating the Area Under the Curve (AUC). The association of clinical variables (age, gender, side, site, and nodule size) with positive LNs was assessed with univariable and multivariable logistic regression analysis. A clinical score was obtained as a linear combination of the selected clinical variables weighted by their respective coefficients; the corresponding AUC was then calculated for both datasets. Finally, a radiomics–clinical score was obtained by applying a logistic regression multivariable model to

the above-mentioned scores, and the corresponding AUC was calculated for both datasets. The AUC for the radiomics, clinical and clinical–radiomics models were compared with the DeLong test [25]. We replicated all the analyses separately on the two subgroups of patients with CT images reconstructed with FBP or IR algorithms. For each subgroup, AUCs for the radiomic, clinical and clinical–radiomics models were compared with the DeLong test [25].

Overall survival prediction: OS was calculated from the date of CT to the date of death or last follow-up, whichever occurred first. The LASSO Cox regression model was used to select the radiomic features predicting OS. We combined the selected features into a radiomics score as a linear combination of the selected features weighted by their respective coefficients. The association of the radiomics score with OS was assessed in the training and validation datasets by using Kaplan–Meier survival analysis. For this, the patients were classified into high-risk or low-risk groups according to the radiomics score, by using the third quartile as the threshold. The difference in the survival curves of the high-risk and low-risk groups was evaluated by using the Log–Rank test. The predictive accuracy of the radiomics score for OS was assessed in both datasets [26]. The association of clinical variables (age, gender, side, site, nodule size, histological type, grading, pT and pN) with OS was assessed with univariable and multivariable analysis. A clinical score was obtained as a linear combination of the clinical variables weighted by their respective coefficients. Finally, a combined radiomics–clinical score was obtained by applying a Cox regression multivariable model to the above-mentioned scores, and the corresponding C-index was calculated for the clinical–radiomics model in both datasets.

p-values < 0.05 were considered statistically significant. The analyses were performed using SAS software (SAS Institute Inc., Cary, USA), version 9.4 and R software (http://www.Rproject.org), version 3.5.3. (R Core Team, R Foundation for Statistical Computing, Vienna, Austria).

Results.

A total of 270 patients were enrolled and randomly divided into training (n= 180) and validation datasets (n= 90). Baseline characteristics of the cohort are shown in **Table 1**.

Table 1

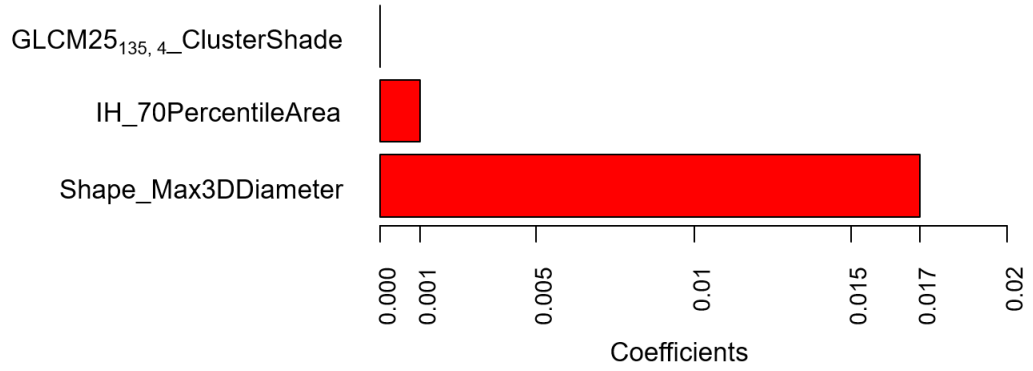
Characteristic	All Patients (N = 270) N (%)	Training Set (N = 180) N (%)	Validation Set (N = 90) N (%)
Age (years) ^	67.4 (61.0–72.6)	66.6 (60.7–72.2)	68.4 (62.3–72.8)
Gender			
Female	103 (38%)	74 (41%)	29 (32%)
Male	167 (62%)	106 (59%)	61 (68%)
Grading			
1	30 (13%)	21 (14%)	9 (12%)
2	82 (36%)	56 (37%)	26 (33%)
3	117 (51%)	74 (49%)	43 (55%)
Missing	41	29	12
Side			
Right	153 (57%)	102 (57%)	51 (57%)

Left	117 (43%)	78 (43%)	39 (43%)
Site			
Upper	154 (57%)	101 (56%)	53 (59%)
Medium	12 (4%)	9 (5%)	3 (3%)
Lower	93 (34%)	65 (36%)	28 (31%)
Mixed	11 (4%)	5 (3%)	6 (7%)
Nodule size (mm) ^	31 (18–45)	28 (17–45)	36 (22–46)
pT			
0	3 (1%)	1 (1%)	2 (2%)
1	97 (36%)	74 (41%)	23 (26%)
2	124 (46%)	76 (42%)	48 (53%)
3	46 (17%)	29 (16%)	17 (19%)
pN			
pN0	199 (74%)	139 (77%)	60 (67%)
pN1	71 (26%)	41 (23%)	30 (33%)
Algorithm type			
FBP	187 (69%)	130 (72%)	57 (63%)
IR	83 (31%)	50 (28%)	33 (37%)
Status			
Alive	202 (75%)	140 (78%)	62 (69%)
Deceased	68 (25%)	40 (22%)	28 (31%)
Follow-up (months) ^	46.1 (29.8–63.3)	47.0 (32.0–65.2)	45.5 (22.7–59.6)

A total of 881 radiomic features was calculated for each patient.

Lymph Nodes

The radiomics score obtained for the prediction of positive LNs on the training dataset consisted of three radiomic features: ClusterShade from GLCM25 category calculated along 135° direction with four voxels offset (GLCM25135, 4_ClusterShade), 70th percentile of the intensity values in the cumulative histogram (IH_70PercentileArea), and the maximum diameter evaluated on the 3D lesion volume (Shape_Max3DDiameter). Coefficients of the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model for the calculation of the radiomics score are shown in **Figure 4**.



GLCM=GrayLevelCooccurenceMatrix; IH=IntensityHistogram;

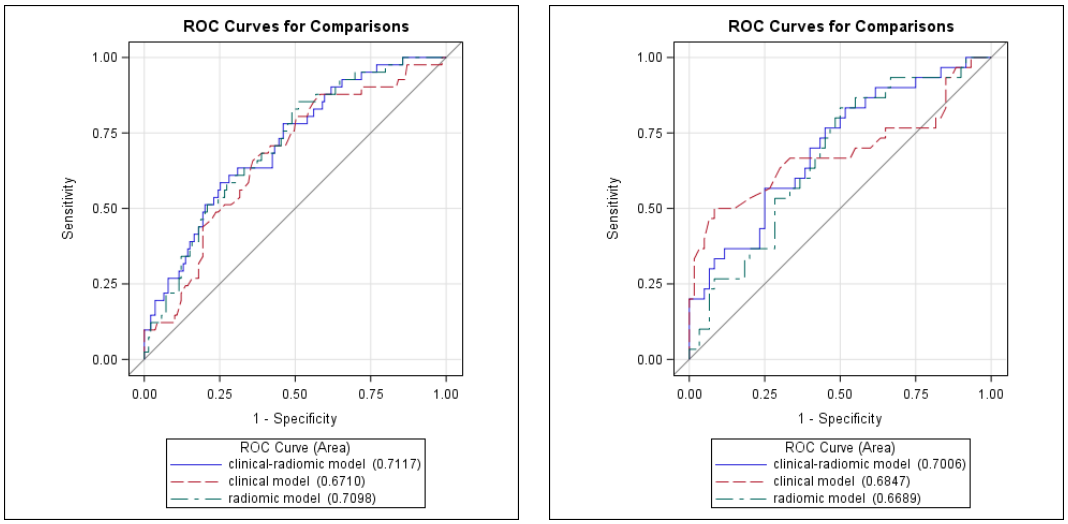
Figure 4

Among clinical variables, univariable analysis in the training set showed that mixed site (meaning that the tumour showed a transfissural growth into two lobes) and nodule size (the larger, the worse) were significantly associated with positive LNs. Multivariable analysis confirmed the importance of the site (**Table 2**).

Table 2

Variable	Univariable Analysis		Multivariable Analysis *	
	OR (95%CI)	***p*-value**	**OR (95%CI)**	***p*-value**
Age (years)	1.02 (0.98–1.06)	0.38	-	-
Gender (females vs males)	0.78 (0.38–1.61)	0.50	-	-
Side (left vs right)	0.61 (0.29–1.26)	0.18	-	-
Site				
Medium vs. Upper	0.42 (0.05–3.67)	0.14	0.35 (0.04–3.20)	0.13
Lower vs. Upper	0.85 (0.39–1.82)	0.22	0.79 (0.36–1.72)	0.27
Mixed vs. Upper	**13.57 (1.44–127.43)**	**0.01**	**10.94 (1.13–105.40)**	**0.02**
Nodule size (mm)	1.01 (1.00–1.02)	0.07	1.01 (1.00–1.02)	0.12

Receiving Operating Characteristic (ROC) curves for prediction of positive LNs showed no significant differences in performance of the combined clinical–radiomics model, as compared to a single radiomics or single clinical model in the training and validation sets (**Figure 5**).



(a) **(b)**

Figure 5

Overall Survival.

Looking at OS prediction, the Cox regression LASSO model selected six radiomic features for the radiomics score (**Figure 6**).

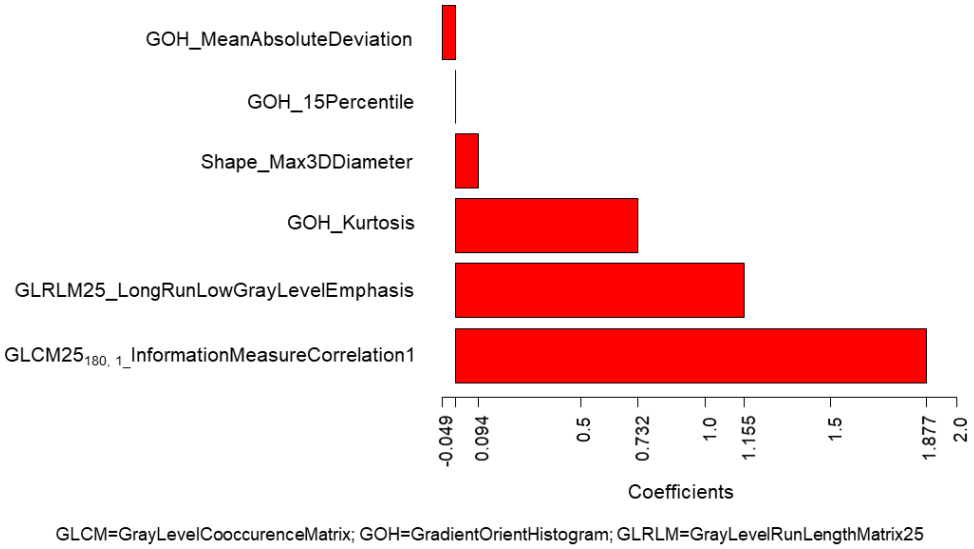


Figure 6

Parameters associated with OS in high-risk and low-risk patients, defined according to the third quartile of the radiomics score are shown in **Table 3**.

Table 3

	Training Data Set			Validation Data Set		
	High-Risk Group *	**Low-Risk Group**	**Total**	**High-Risk Group ***	**Low-Risk Group**	**Total**

No. of patients	45 (25%)	135 (75%)	180 (100%)	20 (22%)	70 (78%)	90 (100%)
Follow-up time						
Median (IQR)	37.0 (19.3-62.2)	49.0 (35.5–68.4)	47.0 (32.6–66.1)	33.0 (17.2–56.4)	47.0 (28.9–61.6)	46.0 (23.6-60.6)
Shortest (months)	0.8	0.9	0.8	1.6	0.2	0.2
No. of events						
At 1 year	8 (18%)	3 (2%)	11 (6%)	3 (15%)	3 (4%)	6 (7%)
At 3 years	17 (38%)	14 (10%)	31 (17%)	7 (35%)	13 (19%)	20 (22%)
At 5 years	20 (44%)	19 (13%)	39 (22%)	9 (45%)	16 (23%)	25 (29%)

The percentage of deaths at the second and third years in the validation set for high-risk patients was almost double that of the low-risk patients (respectively, 35% vs. 19% and 45% vs. 23%). As shown in **Figure 7**, a significant difference in OS for the high- and low-risk groups according to the radiomics score was observed in the training set and confirmed in the validation set.

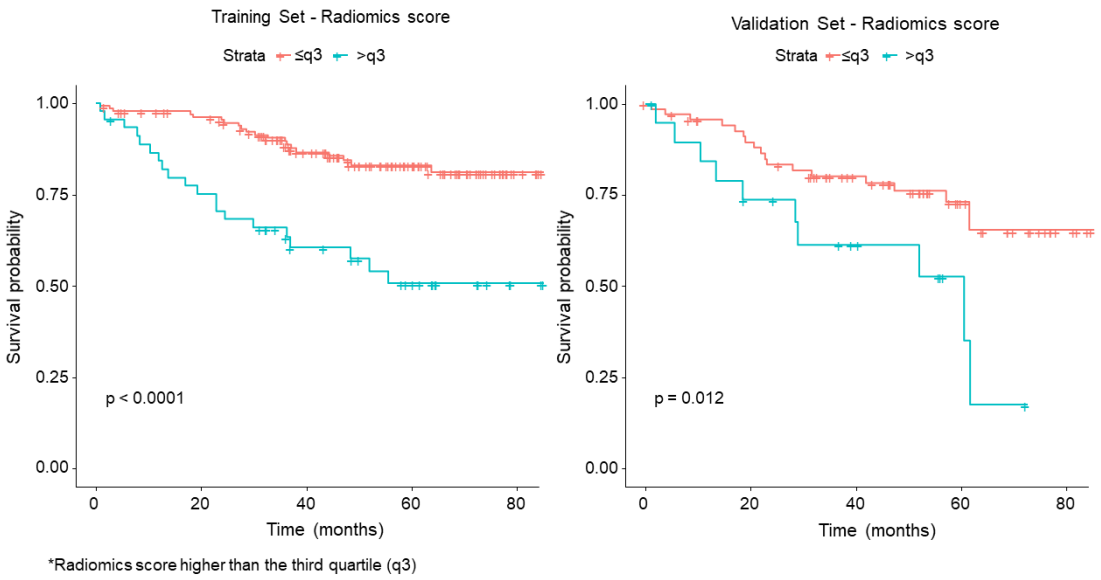


Figure 7

Discussion.

In this series, selected clinical and radiomic features showed association with the positivity of LN, although the combined radiomics–clinical model did not perform better than a single clinical or radiomics model. Among the clinical features, the mixed site was significant in the univariate and multivariate analysis. The importance of this feature is recognized by its role in Tumor-Nodes-Metastasis (TNM) staging, where the invasion of the visceral pleura makes the tumour belong to the T2 category, despite its size [41]. This result is concordant also with Li et al., who demonstrated

a significant association between adjacent lobe invasion through pleural fissure and LN positivity [42]. The size of the tumour, measured on an axial image, was borderline-significant in the univariable and multivariable analysis. Size is a well-known feature for prognosis, as demonstrated by the sub-division of tumours in the T category according to size [41]. Its importance is confirmed in our series by the better performance of the single clinical model, after the inclusion of size as a feature. Likewise, the importance of size as a prognostic factor for LN positivity is demonstrated by one radiomic feature included in the score, the Shape_Max3DDiameter. This feature represents a measure of the maximum dimension of the lesion, evaluated in 3D, not directly related to the volume and is more precise than the maximum axial diameter. This may also account for the slightly better performance of the radiomics model compared to the clinical model, where the Max3DDiameter radiomic feature may provide a more precise definition of size than the maximum axial diameter. Other significant features in the radiomics score for LN positivity prediction were IH_70PercentileArea and GLCM25135, 4_ClusterShadeClShade. IH_70PercentileArea belongs to the Intensity Histogram category and roughly indicates that the mean HU number within the lesion is associated with LN status. GLCM25135, 4_ClusterShade, belonging to the GLCM category, is a measure of the global skewness. When the Cluster Shade parameter is high, the distribution of HU values is asymmetric. In our series, low values of this feature, that may be associated with intratumoral necrosis, were more frequently associated with positive LNs, whereas high values, encountered in lesions with calcifications and ground-glass opacities, were associated with negative LNs.

Despite the abovementioned associations, the combined clinical and radiomics model did not perform better than a single clinical or radiomics model. This result is discordant with Tan et al., who demonstrated that, in patients with resectable oesophageal carcinoma, a radiomics nomogram provided a good risk estimation of LN metastasis and outperformed size criteria [43]. However, unlike these authors, we did not evaluate LN radiomic features, because our purpose was to predict the presence of positive LNs according to the characteristics of the primary tumour. Conversely, Yang et al. [44] demonstrated a good performance of a radiomics-based nomogram extracted from lung tumour to predict LN metastases. Although we used a similar method, by performing the radiomics analysis on the lung tumour volume, our results may be different because Yang et al. included patients with CT examinations acquired with the same parameters and reconstructed with the same algorithm [44]. In this regard, our entire sample is less homogeneous, because we included CTs obtained from different scanners and reconstructed with two different algorithms. However, it must be pointed out that from a methodological point of view, we tried to take this into account by performing a preliminary selection with ANOVA to choose features that were not significantly affected by the use of different scanners or reconstruction algorithms. This is important because reproducibility and differences in acquisitions and reconstructions are frequent issues in retrospective studies, which currently represent the majority of radiomics studies.

As a consequence, the poor performance of our predictive model, including radiomic and clinical parameters, as compared to the model based on clinical variables alone, could be due to the fact that feature selection with ANOVA may have eliminated features which potentially carry relevant predictive information, but are significantly affected by the scanner or reconstruction algorithm. It is possible that, in a more homogeneous dataset, such features would survive the feature selection process and would enter the predictive model with a significant improvement in terms of performance.

The analysis of association between clinical and radiomic features with OS showed a very good performance of the radiomics score, that significantly separated patients into high-risk and low-risk groups ($p < 0.0001$ in the training set; $p = 0.012$ in the validation set). Among radiomic features, GLCM25180, 1_InformationMeasureCorrelation1, GLRLM25_LongRunLowGrayLevel Emphasis, GOH_Kurtosis and Shape_Max 3DDiameter were included in the score. GLCM25180, 1_InformationMeasureCorrelation1 belongs to the GLCM category and quantifies the degree of randomness within the tumour, in terms of entropy and statistical disorder. GLRLM25_LongRunLowGrayLevel Emphasis belongs to the GLRLM category, and quantifies runs, intended as consecutive pixels with the same grey level. GOH_Kurtosis indicates the flatness of the curve of values, without concern for spatial relationships. A prevalent direction for the voxel intensity gradient indicates that structures within the 3D volume of interest (VOI) develop along a precise direction (e.g., an intratumoral vessel). Shape_Max 3DDiameter (see features of the radiomics score for LN) was associated with OS, with high values (large lesions) associated with worse OS.

This study has some limitations. The number of patients with malignant lymph nodes (pN1) in our group was relatively small (71/270; 26%) and this may account for different results compared to previous papers. However, the inclusion of patients with pN2 introduced a bias related to the neo-adjuvant chemotherapy. The CT examinations included in this study were performed over quite a long period of time (four years), and this may have affected the acquisition protocols. However, we specifically performed a preliminary ANOVA test to select stable and reproducible radiomic features.

Conclusions

In conclusion, a combined clinical–radiomics model was not superior to a single clinical or radiomics model in predicting LN metastases in lung cancer patients, whereas a radiomics score was able to significantly separate high-risk and low-risk patients for OS.

7. References

1. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. Radiology 278:563-577.

2. Chetan MR, Gleeson FV. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. Eur Radiol. 2020 Aug 18. doi: 10.1007/s00330-020-07141-9.

3. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30(9):1234-1248.

4. Lambin P, Leijenaar RTH, Deist T, et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14(12):749-762.

5. Dalal T, Kalra MK, Rizzo SM, et al. (2005) Metallic prosthesis: technique to avoid increase in CT radiation dose with automatic tube current modulation in a phantom and patients. Radiology. 236(2):671-675.

6. Rizzo SM, Kalra MK, Schmidt B, et al. (2005) CT images of abdomen and pelvis: effect of nonlinear three-dimensional optimized reconstruction algorithm on image quality and lesion characteristics. Radiology 237(1):309-315.

7. Mackin D, Fave X, Zhang L et al. (2015) Measuring Computed Tomography scanner variability of radiomic features. Investigative Radiology 50(11):757-765

8. Larue RTHM, Van Timmeren JE, De Jong EEC et al. (2017) Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncol 56(11): 1544-1553

9. Van Timmeren JE, Leijenaar RTH, Van Elmpt W et al. (2016) Test-retest data for radiomic feature stability analysis: generalizable or study-specific? Tomography 2(4):361-365

10. Solomon J, Mileto A, Nelson RC et al. (2016) Quantitative features of liver lesions, long nodules, and renal stones at multi-detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology 279(1):185-194

11. Rizzo S, Botta F, Raimondi S, et al. (2018) Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months. Eur Radiol. May 8. doi: 10.1007/s00330-018-5389-z

12. Hojjatoleslami S, Kittler J (1998) Region growing: a new approach. IEEE Trans Image Process 7(7):1079–1084.

13. Kalef-Ezra J, Karantanas A, Tsekeris P (1999) CT measurement of lung density. Acta Radiol. 40(3):333-337.

14. Sofka M, Wetzl J, Birkbeck N, et al. (2011) Multi-stage learning for robust lung segmentation in challenging CT volumes. Med Image Comput Comput Assist Interv 14(Pt 3):667-674.

15. Knollmann FD, Kumthekar R, Fetzer D, Socinski MA (2014) Assessing response to treatment in non--small-cell lung cancer: role of tumor volume evaluated by computed tomography. Clin Lung Cancer 15(2):103-109.

16. Gao H, Chae O (2010) Individual tooth segmentation from CT images using level set method with shape and intensity prior. Pattern Recognition 43(7):2406–2417.

17. Xu N, Bansal R, Ahuja N. (2003) Object segmentation using graph cuts based active contours. IEEE;42:II-46–53.

18. Ye X, Beddoe G, Slabaugh G. (2010) Automatic graph cut segmentation of lesions in CT using mean shift superpixels. J Biomed Imaging 2010:983963.

19. Tan Y, Schwartz LH, Zhao B (2013) Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. Med Phys 40(4):043502.

20. Sun S, Bauer C, Beichel R (2012) Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach. IEEE Trans Med Imaging 31(2):449-460.

21. Rizzo S, Petrella F, Buscarino V, et al (2016) CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. Eur Radiol. 26(1):32-42.

22. Lam SWC (1996) Texture feature extraction using gray level gradient based co-occurence matrices. IEEE;261:267–271.

23. Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 3(6):610–621.

24. Galloway MM (1975) Texture analysis using gray level run lengths. Comput Graph Image Process 4(2):172–179.

25. Wilkinson L, Friendly M. (2009) The history of the cluster heat map. Am Stat 63(2):179–184.

26. Jolliffe I.T. (2002) Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, XXIX, 487 p. 28 illus.

27. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 289–300.

28. Breiman L (2001) Random Forests. Machine Learning, 45, 5-32.

29. Eschrich S, Yang I, Bloom G, et al. (2005) Molecular staging for survival prediction of colorectal cancer patients. J Clin Oncol 23(15):3526–3535.

30. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso, in "Journal of the Royal Statistical Society. Series B", Vol. 58, No. 1, pp. 267-288

31. Shedden K, Taylor JM, Enkemann SA, et al (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. NatMed 14(8):822–827.

32. Harrell FE Jr, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 28;15(4):361-87.

33. Rizzo S, Botta F, Raimondi S et al. (2018) Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 14;2(1):36. doi: 10.1186/s41747-018-0068-z.

34. Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, Fumagalli C, Spitaleri G, Rampinelli C, De Marinis F, Spaggiari L, Bellomi M. CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer.

Eur Radiol. 2016 Jan;26(1):32-42. doi: 10.1007/s00330-015-3814-0.

35. Rizzo S, Raimondi S, de Jong EEC, et al. (2019) Genomics of non-small cell lung cancer (NSCLC): Association between CT-based imaging features and EGFR and K-RAS mutations in 122 patients-An external validation. Eur J Radiol; 110:148-155.

36. Aerts HJ, Velazquez ER, Leijenaar RT et al. (2014) Decoding tumour phenotype by

noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5, 4006.

37. Dingemans AM, Groen HJ, Herder GJ et al. (2015). A randomized phase II study

comparing paclitaxel-carboplatin-bevacizumab with or without nitroglycerin patches

in patients with stage IV nonsquamous nonsmall-cell lung cancer: NVALT12

(NCT01171170)dagger, Ann. Oncol. 26 (11) 2286–2293.

38. de Jong EEC, van Elmpt W, Rizzo S, et al. (2018) Applicability of a prognostic CT-based radiomic signature model trained on stage I-III non-small cell lung cancer in stage IV non-small cell lung cancer. Lung Cancer. 124:6-11.

39. DeLong ER, DeLong DM, Clarke-Pearson DL.(1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44, 837–45.

40. Gonen M, Heller G. (2005) Concordance probability and discriminatory power in proportional hazards regression. Biometrika 92, 965–970.

41. Goldstraw P, Chansky K, Crowley J, et al (2016) 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.

42. Li H, Wang R, Zhang D, et al. (2019) Lymph node metastasis outside of a tumor-bearing lobe in primary lung cancer and the status of interlobar fissures: The necessity for removing lymph nodes from an adjacent lobe. Medicine (Baltim.) 98, e14800.

43. Tan X, Ma Z, Yan L, et al (2019). Radiomics nomogram outperforms size criteria in discriminating lymph node metastasis in resectable esophageal squamous cell carcinoma. Eur. Radiol. 29, 392–400, doi:10.1007/s00330-018-5581-1.

44. Yang X, Pan X, Liu H, et al. (2018) A new approach to predict lymph node metastasis in solid lung adenocarcinoma: A radiomics nomogram. J. Thorac. Dis. 10 (Suppl. 7), S807–S819.

8. Tables and figures legends

Table 1. Baseline characteristics of the study population

Table 2. Univariable and multivariable Odds Ratios for the association between clinical variables with positive lymph nodes (training set).

Table 3. Overall Survival in high-risk and low-risk patients according to the radiomics score.

Figure 1. Axial CT images showing differences in the same acquisition plane between a contrast-enhanced and non-contrast enhanced image

Figure 2. Axial CT images showing the same acquisition plane with different radiation doses (lower in A, higher in B).

Figure 3. An example of manual segmentation of lung cancer. Although manual segmentation is often considered ground truth, this image shows red and black regions of interest delineated by two different readers on the same tumour

Figure 4. Values of the coefficients of the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model for the prediction of positive lymph nodes according to radiomic features (training set). The plot shows the model coefficients of the three radiomic features selected as significantly associated with lymph node status. These coefficients were used to calculate the radiomics score used to predict lymph node status in the validation set.

Figure 5. ROC curves for prediction of positive lymph nodes in (**a**) the training set and (**b**) the validation set according to clinical, radiomics and clinical–radiomics models. The plots show the ROC curves of the three models and the associated values of the Area under the Curves (AUC).

Figure 6. Values of the coefficients for Cox regression LASSO model for prediction of overall survival according to radiomic features (training set). The plot shows the model coefficients of the six radiomic features selected as significantly associated with overall survival. These coefficients were used to calculate the radiomic score used to predict overall survival in the validation set.

Figure 7. Kaplan–Meier curves and Log–Rank test for high-* and low-risk groups according to the radiomics score (* Radiomics score higher than the third quartile,q3).