


Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery

Margarita Kirienko¹ · Luca Cozzi² · Lidija Antunovic³ · Lisa Lozza⁴ · Antonella Fogliata¹ · Emanuele Voulaz⁵ · Alexia Rossi^{1,6} · Arturo Chiti^{1,3} · Martina Sollini¹ 

Received: 14 August 2017 / Accepted: 14 September 2017 / Published online: 24 September 2017
© Springer-Verlag GmbH Germany 2017

Abstract

Purpose Radiomic features derived from the texture analysis of different imaging modalities show promise in lesion characterisation, response prediction, and prognostication in lung cancer patients. The present study aimed to identify an images-based radiomic signature capable of predicting disease-free survival (DFS) in non-small cell lung cancer (NSCLC) patients undergoing surgery.

Methods A cohort of 295 patients was selected. Clinical parameters (age, sex, histological type, tumour grade, and stage) were recorded for all patients. The endpoint of this study was DFS. Both computed tomography (CT) and fluorodeoxyglucose positron emission tomography (PET) images generated from the PET/CT scanner were analysed. Textural features were calculated using the LifeX package. Statistical analysis was performed

using the R platform. The datasets were separated into two cohorts by random selection to perform training and validation of the statistical models. Predictors were fed into a multivariate Cox proportional hazard regression model and the receiver operating characteristic (ROC) curve as well as the corresponding area under the curve (AUC) were computed for each model built.

Results The Cox models that included radiomic features for the CT, the PET, and the PET+CT images resulted in an AUC of 0.75 (95%CI: 0.65–0.85), 0.68 (95%CI: 0.57–0.80), and 0.68 (95%CI: 0.58–0.74), respectively. The addition of clinical predictors to the Cox models resulted in an AUC of 0.61 (95%CI: 0.51–0.69), 0.64 (95%CI: 0.53–0.75), and 0.65 (95%CI: 0.50–0.72) for the CT, the PET, and the PET+CT images, respectively.

Conclusions A radiomic signature, for either CT, PET, or PET/CT images, has been identified and validated for the prediction of disease-free survival in patients with non-small cell lung cancer treated by surgery.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00259-017-3837-7>) contains supplementary material, which is available to authorized users.

✉ Martina Sollini
martina.sollini@cancercenter.humanitas.it;
martinasollini@msn.com

¹ Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini, 20090 Pieve Emanuele, Milan, Italy

² Radiotherapy and Radiosurgery, Humanitas Clinical and Research Center, Via Manzoni 56, 20089 Rozzano, Milan, Italy

³ Nuclear Medicine, Humanitas Clinical and Research Center, Via Manzoni 56, 20089 Rozzano, Milan, Italy

⁴ Orobix Srl, Via Camozzi 144, 24121 Bergamo, Italy

⁵ Thoracic Surgery, Humanitas Clinical and Research Center, Via Manzoni 56, 20089 Rozzano, Milan, Italy

⁶ Radiology, Humanitas Clinical and Research Center, Via Manzoni 56, 20089 Rozzano, Milan, Italy

Keywords Radiomics · Lung cancer · Prognosis · CT · PET/CT · Texture analysis

Introduction

Lung cancer is the fourth most frequently diagnosed malignancy in Europe, with an incidence of 41.9 per 100,000, and is the leading cause of cancer-related death, with a mortality of 35.2 per 100,000, in both sexes [1]. Non-small cell lung cancer (NSCLC) accounts for 85% of all primary lung cancers [2]. The tumour node metastasis (TNM) staging system is the most important postoperative prognostic tool and guides patient treatment [3–5]. However, survival times vary widely among NSCLC patients at the same disease stage who

undergo complete tumour resection, demonstrating the need for novel prognostic methods [3, 6].

Computed tomography (CT) and fluorodeoxyglucose positron emission tomography/CT (FDG-PET/CT) are routinely used for lesion detection, lesion characterisation and clinical staging of NSCLC patients. Such imaging provides “phenotype” information, is minimally or non-invasive, and can be repeated, allowing a personalised assessment of the disease. A quantitative radiomic approach, using features derived from the texture analysis of different imaging modalities, including FDG-PET and CT, has been tested for lesion characterisation, response prediction and prognostication, with promising results [7–9]. However, most of these studies have been affected by methodological drawbacks such as limited patient datasets and lack of validation. Moreover, the combination of information derived from both PET and CT has not been validated [10]. The present study aimed to identify a CT- and PET-based radiomic signature capable of predicting disease-free survival (DFS) in NSCLC patients undergoing surgery.

Methods

Study design and patient selection

In this retrospective single-centre investigation, the following inclusion/exclusion criteria were applied to select patients from the institutional database. The inclusion criteria were: a) age > 18 years old, b) surgical intervention for a lung lesion between 01/01/2011 and 30/11/2016, c) FDG-PET/CT scan performed in our institution to characterise a lung lesion or to stage a lung tumour within 45 days before surgery. The exclusion criteria were: a) histology other than lung adenocarcinoma or squamous cell carcinoma, b) concomitant or previous (within 3 years from lung cancer diagnosis) other cancer type. A total of 522 patients (M:F = 343:179) were selected from the institutional database. A cohort consisting of 295 patients was identified applying the above-mentioned inclusion and exclusion criteria. For all patients, available clinical parameters such as age, sex, histological type, tumour grade, and stage were recorded. Smoking habits and performance status were not considered in this analysis. Tumour staging was defined on the basis of the American Joint Committee on Cancer TNM Staging Manual, 7th Edition [11]. T and N status were defined using the information provided in the histopathological report of the surgical intervention. Identification of the presence of distant metastases was based on histology and/or imaging.

Treatment and follow-up were performed according to our internal standard procedures after discussion at the multidisciplinary lung tumour board. In particular, for patients with pT1–2a pN0 disease only follow-up was indicated. Patients with pT1–2 N1, pT3 N0, and any T N2 disease received adjuvant chemotherapy; the last-mentioned group also

underwent radiation treatment to the mediastinum. For patients with stage III disease with pN1 involvement, adjuvant chemotherapy was indicated.

The endpoint of this study was DFS, which was defined as the time between the date of surgery and either the date of relapse (event), which refers to tumour recurrence or tumour-related death, or the date of last patient access (censored). The study was approved by the institutional ethics committee.

FDG-PET/CT image acquisition

FDG-PET/CT image acquisition was performed according to the European Association of Nuclear Medicine (EANM) guidelines version 1.0 and from February 2015, 2.0 [12]. Briefly, appropriate patient preparation (fasting from at least 4 h) and adequate blood glucose levels (< 200 mg/dL) were requested. Images were acquired 60 ± 5 min after FDG injection, using an integrated PET/CT scanner, either a Siemens Biograph 6 LSO (Siemens, Erlangen, Germany) ($n = 159$ patients) or a General Electric Discovery 690 (General Electric Healthcare, Waukesha, WI, USA) ($n = 136$ patients). Image acquisition parameters are reported in Supplementary Table 1. All PET images were corrected for attenuation using the acquired CT data.

Texture analysis

The volume of interest (VOI) of the primary tumour lesion was (semi-)automatically defined on PET images with a threshold of 40% of the maximum standardised uptake value (SUV_{max}) using a commercial software (PET VCAR, GE Healthcare, Waukesha, WI, USA) by three nuclear medicine certified physicians, all together. Partial volume effect correction was not performed.

Both CT and PET images generated from the PET/CT scanner were analysed. Textural features were calculated on both CT and PET images within the same VOI. The position of the VOI on the CT images was manually adjusted to identify the correct position of the lesion when respiratory movements determine a mismatch between CT and PET images. A consensus among reviewers was found concerning the adjustment of the position of the VOI on CT images. Features were extracted using the LifeX package (<http://www.lifexsoft.org>) [13]. Tonal discretisation for PET data was performed to reduce the continuous scale to 64 bins with absolute scale bounds between 0 and 25. Similarly, the discretisation for CT data was done with 10 Hounsfield unit (HU) increments from –1000 to 3000 HU. The features included conventional and histogram-based parameters, shape and size, and second and high order features, as detailed in Supplementary Table 2. LifeX calculates texture features only for VOI of at least 64 voxels. For some patients, the PET VOI did not reach the

minimum number of 64 voxels related to the image matrix resolution. Therefore, radiomic features derived from CT images were studied within the entire cohort of 295 patients (CT dataset), while those from PET were studied within the subsample of 259 patients (PET dataset) (Table 1).

Statistical analysis

Statistical analysis was performed using the open source R platform. Median DFS and 1-year, 2-, and 3-year DFS rates were calculated. Kaplan-Meier curves were used to describe DFS for the CT dataset and the PET dataset, and to describe the DFS stratified according to histology (adenocarcinoma versus squamous cell carcinoma) and tumour stage (I–II versus III–IV). Univariate analysis was used to test the clinical predictors (age, sex, histology, and stage) in both CT

and PET (PET/CT) datasets. Tumour differentiation grade was not included among clinical predictors since it was not available for patients who underwent neo-adjuvant chemotherapy. Mann–Whitney U test was used to investigate texture feature differences between scanners (Siemens versus GE) for both CT and PET images. In addition, CT and PET datasets were studied regardless of the scanner used for image acquisition in order to assess the reproducibility of the radiomic analysis.

The following analyses were performed on CT and PET data. Each radiomic feature was scrutinised for its univariate significance. The datasets were then separated into two cohorts by random selection to perform training and validation of the statistical models. The mutual correlation between the preliminarily selected features was evaluated with the Spearman correlation coefficient in order to assess potential

Table 1 Baseline patient characteristics

Characteristic	CT dataset			PET dataset		
	Overall <i>n</i> = 295	Training <i>n</i> = 195	Validation <i>n</i> = 100	Overall <i>n</i> = 259	Training <i>n</i> = 169	Validation <i>n</i> = 90
Age (years; mean±SD)	68.8 ± 8.8	68.7±8.8	68.9±8.6	70.1 ± 8.4	69.6 ± 8.7	70.9 ± 8.4
	N pts. (%)	N pts. (%)	N pts. (%)	N pts. (%)	N pts. (%)	N pts. (%)
Sex						
Male	197 (67)	135 (69)	62 (62)	176 (68)	116 (69)	60 (67)
Female	98 (33)	60 (31)	38 (38)	83 (32)	53 (31)	30 (33)
Histology						
Adenocarcinoma	193 (65)	132 (68)	61 (61)	162 (63)	110 (65)	52 (58)
Squamous cellular carcinoma	102 (35)	63 (32)	39 (39)	97 (37)	59 (35)	38 (42)
Pathological stage						
I	109 (37)	71 (36)	38 (38)	93 (36)	60 (35)	33 (38)
II	75 (25)	53 (27)	22 (22)	71 (27)	49 (29)	22 (24)
III	96 (33)	61 (31)	35 (35)	83 (32)	52 (31)	31 (34)
IV ^a	15 (5)	10 (5)	5 (5)	12 (5)	8 (5)	4 (4)
Differentiation grade						
G1	7 (2)	5 (2)	2 (2)	7 (3)	5 (3)	2 (2)
G2	136 (46)	86 (44)	50 (50)	120 (46)	75 (44)	45 (50)
G3	120 (41)	85 (44)	35 (35)	105 (41)	71 (42)	34 (38)
Not specified ^b	32 (11)	19 (10)	13 (13)	27 (10)	18 (11)	9 (10)
Scanner type						
Discovery 690 – General Electric	136 (46)	91 (47)	45 (45)	118 (46)	78 (46)	40 (44)
Biograph 6 - Siemens	159 (54)	104 (53)	55 (55)	141 (54)	91 (54)	50 (56)
Outcome						
No evidence of disease	151 (51)	105 (54)	46 (46)	130 (50)	90 (53)	40 (44)
Recurrence	144 (49)	90 (46)	54 (54)	129 (50)	79 (47)	50 (56)
Follow-up, mean±SD (months)	22.7±17.2	22.9±17.1	22.5±17.2	23.5±17.6	23.5±17.6	23.7±17.5
Follow-up, median	20.1	20.1	20.0	20.5	20.6	20.3

N, number; *pts.*, patients; *SD*, standard deviation

^a Patients with brain metastases at diagnosis that were treated radically before lung surgery

^b Grading was not specified for those patients who underwent neo-adjuvant chemotherapy

redundancy between these features [14]. A threshold of 0.90 was used when testing correlations between features. All uncorrelated predictors identified as significant ($p < 0.05$; p values were corrected for false discovery rate) after multiple testing correction (with the Holm-Bonferroni method), in both the training and the validation dataset, were fed into a multivariate Cox proportional hazard regression model. The choice to include only significant predictors in the multivariate Cox proportional hazard regression model was based on the wish to test reproducible features within the training and validation datasets. Clinical predictors were included in the multivariate analysis regardless of their significance at univariate analysis. The Cox proportional hazard regression model was further used to predict the DFS in the training and validation datasets. Different models were constructed: a) including clinical predictors only (age, sex, histological type, and stage), b) including radiomic parameters only – the radiomic signature (b1 - not including and b2 - including the features found to be significantly different between scanners by the Mann–Whitney U test), and c) including both the radiomic signature (including all the significant features) and clinical predictors.

To evaluate the model performance of the validation set, the receiver operating characteristic (ROC) curve was determined and the corresponding area under the curve (AUC) computed for each model. The 95% confidence interval for the AUC was computed with the default of 2000 stratified bootstrap replicates.

All the patients were dichotomised into low- and high-risk groups and the actuarial DFS curves were determined. For each radiomic feature found to be significant at univariate analysis, a search for the threshold that could better split the population into low- and high-risk groups was performed. This was achieved by dividing the population into groups using a continuously moving covariate value in the range of all available values. The best threshold was defined as the value with the lowest log-rank p value in the Kaplan-Meier statistic. To overcome the noisy distribution of the p values, a simplified bootstrap procedure was implemented. The threshold search procedure was repeated 500 times, each time leaving out 20% of the sample and retaining the remaining 80% randomly chosen. The radiomic feature was then retained if it was significant on 80% of the tests and its dichotomising threshold was defined as the average over the runs.

Table 2 Results of univariate analysis and Cox regression models for CT images

CT dataset ($n = 295$) Parameter	Univariate p value	Cox regression p value			
		Clinical model (a)	Radiomic models (b)		Radiomic and clinical model (c)
			b1	b2	
Clinical predictor					
Age	0.09	n.s.	n.a.	n.a.	n.s.
Sex	0.1	0.07	n.a.	n.a.	n.s.
Histological type	0.2	n.s.	n.a.	n.a.	n.s.
Tumour stage	<0.001	<0.001	n.a.	n.a.	<0.001
CT feature					
Compacity_GLCM	0.02	n.a.	Not included	n.s.	n.s.
Homogeneity 1_GLCM	0.05	n.a.	0.05	n.s.	n.s.
SRE_GLRLM	0.001	n.a.	0.01	0.006	0.001
LRE_GLRLM	0.04	n.a.	Not included	0.02	0.05
RP_GLRLM	0.05	n.a.	Not included	n.s.	n.s.
AUC (95%CI) for the validation cohort		0.58 (0.50–0.67)	0.62 (0.51–0.72)	0.75 (0.65–0.85)	0.61 (0.51–0.69)
DFS low-risk group (median±SD) [months]					
Training		26.0 ± 2.7	28.0 ± 2.5	36.8 ± 3.7	35.0 ± 3.5
Validation		34.6 ± 3.1	Not reached	Not reached	30.2 ± 2.8
DFS high-risk group (median±SD) [months]					
Training		11.1 ± 1.2	14.2 ± 2.1	16.1 ± 2.1	11.2 ± 1.1
Validation		13.2 ± 1.4	10.7 ± 1.3	10.8 ± 1.4	11.8 ± 1.1

Model b1: included features significant at univariate analysis and not significantly different between scanners. Model b2: included all features significant at univariate analysis. AUC, area under the curve; 95%CI, 95% confidence interval; CT, computed tomography; DFS, disease-free survival; GLCM, grey-level co-occurrence matrix; LRE_GLRLM, long-run emphasis computed within the grey-level run-length matrix; n.a., not applicable; n.s., not significant (the exact value was not provided by R when >0.05); RP_GLRLM, run percentage computed within the grey-level run-length matrix; SD, standard deviation; SRE_GLRLM, short-run emphasis computed within the grey-level run-length matrix

Results

Patients’ clinical characteristics are reported in Table 1. The median DFS was 19.8 ± 3.2 months (95%CI: 13.4–26.1 months) and 18.9 ± 2.4 months (95%CI: 14.1–23.7 months) for the CT and PET datasets, respectively. One-, 2-, and 3-year DFS rates for the CT dataset were $65.0\% \pm 3.0\%$, $46.2\% \pm 3.0\%$, and $36.1\% \pm 3.6\%$, respectively. One-, 2-, and 3-year DFS rates for the PET dataset were $63.5\% \pm 3.0\%$, $44.2\% \pm 4.0\%$, and $36.1\% \pm 4.0\%$, respectively. The actuarial overall DFS curve and the DFS curves for the subgroups stratified according to the histological type (adenocarcinoma versus squamous cell carcinoma) and tumour stage (I–II versus III–IV) are shown in Supplementary Fig. 1 for both datasets. At univariate analysis, among the clinical predictors, tumour stage was found to be significant in both CT and PET datasets, as reported in Tables 2 and 3, respectively.

No statistically significant difference was found for 25/41 CT features and 32/43 PET features extracted from images

acquired using the two different scanners (Supplementary Figs. 2 and 3).

CT dataset

The results of the univariate analysis and of the Cox models are detailed in Table 2. The Cox model (b1), built using only the radiomic features that were significant at univariate analysis and excluding the features that were significantly different between the two scanners, resulted in an AUC of 0.62 (95%CI: 0.51–0.72) with a dichotomising threshold of 0.03. Adding into the Cox model (b2) also the features that were significantly different between the two scanners resulted in an AUC of 0.75 (95%CI: 0.65–0.85) with a dichotomising threshold of 0.06. The Cox model (c), which included both the radiomic signature (all the significant radiomic features at univariate analysis) and clinical predictors, resulted in an AUC of 0.61 (95%CI: 0.51–0.69). Figure 1 shows the results for all models from the CT dataset. For all the five CT

Table 3 Results of univariate analysis and Cox regression models for PET images

PET dataset (n = 259) Parameter	Univariate p value	Cox regression p value			
		Clinical model (a)	Radiomic models (b)		Radiomic and clinical model (c)
			b1	b2	
Clinical predictor					
Age	n.s.	n.s.	n.a.	n.a.	n.s.
Sex	n.s.	n.s.	n.a.	n.a.	n.s.
Histological type	n.s.	n.s.	n.a.	n.a.	n.s.
Tumour stage	<0.001	0.004	n.a.	n.a.	<0.001
PET feature					
SUV _{mean}	0.008	n.a.	0.001	0.007	0.06
SUV _{max}	0.008	n.a.	0.001	0.008	0.01
Energy_GLCM	0.02	n.a.	Not included	n.s.	n.s.
LRHGE_GLRLM	0.04	n.a.	0.01	n.s.	n.s.
HGRE_GLRLM	0.05	n.a.	n.s.	n.s.	n.s.
LZHGE_GLZLM	0.02	n.a.	Not included	0.004	0.03
AUC (95%CI) for the validation cohort		0.58 (0.50–0.69)	0.66 (0.56–0.78)	0.68 (0.57–0.80)	0.64 (0.53–0.75)
DFS low-risk group (median±SD) [months]					
Training		30.2 ± 3.6	30.5 ± 3.3	36.1 ± 4.0	33.5 ± 3.8
Validation		58.0 ± 4.7	22.7 ± 3.1	30.0 ± 4.3	26.0 ± 3.1
DFS high-risk group (median±SD) [months]					
Training		11.1 ± 1.0	13.5 ± 1.4	12.3 ± 1.0	13.2 ± 1.6
Validation		13.1 ± 1.2	15.0 ± 1.6	10.5 ± 1.0	12.1 ± 1.3

Model b1: included features significant at univariate analysis and not significantly different between scanners. Model b2: included all features significant at univariate analysis. AUC, area under the curve; 95%CI, 95% confidence interval; DFS, disease-free survival; GLCM, grey-level co-occurrence matrix; HGRE_GLRLM, high grey-level run emphasis computed within the grey-level run-length matrix; LRHGE_GLRLM, long-run high grey-level emphasis computed within the grey-level run-length matrix; LZHGE_GLZLM, long-zone high grey-level emphasis computed within the grey-level zone-length matrix; n.a., not applicable; n.s., not significant (the exact value was not provided by R when >0.05); PET, positron emission tomography; SD, standard deviation; SUV, standardised uptake value

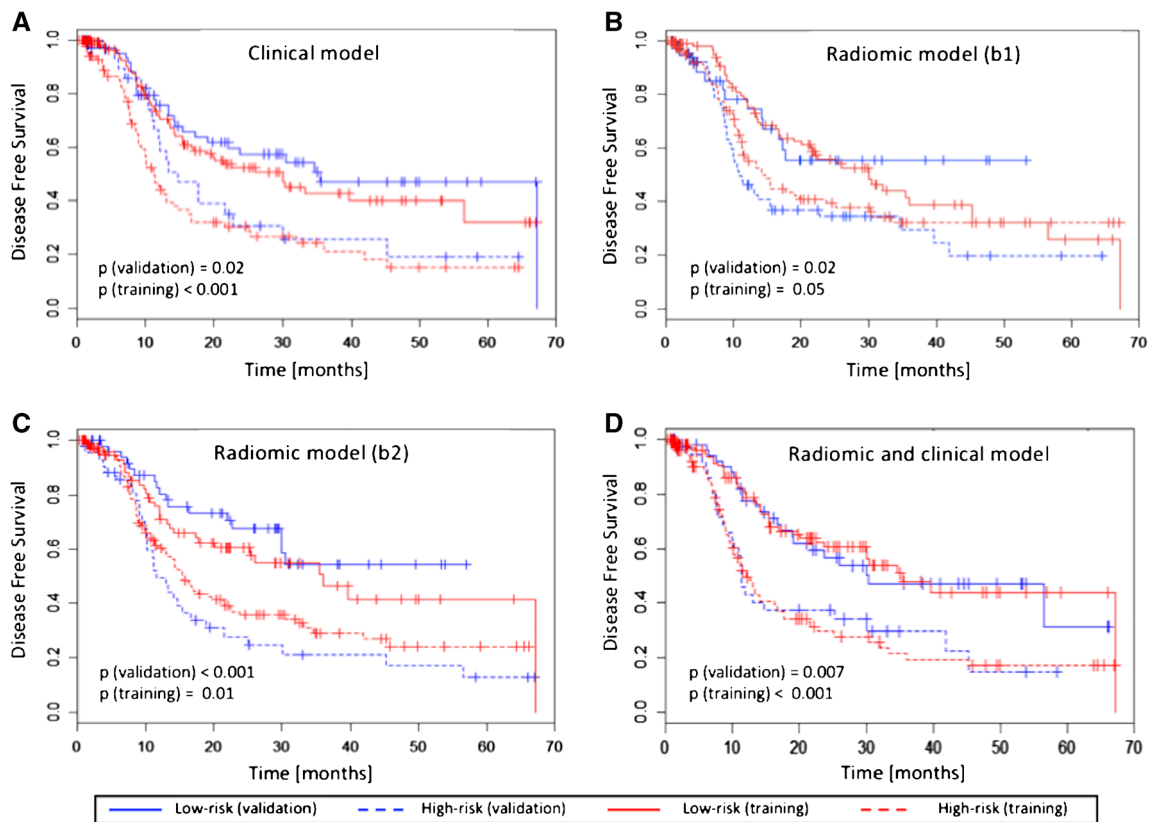


Fig. 1 Disease-free survival prediction results for the CT data. Kaplan-Meier curves for the DFS resulting from the Cox regression models built using clinical variables (a), the radiomic signature (b and c), and their combination (d) for the training and validation groups within the CT dataset

radiomic features selected, it was possible to identify repeatable thresholds in both the training and the validation group (Supplementary Fig. 4).

PET dataset

The results of the univariate analysis and of the Cox models are detailed in Table 3. The Cox model (b1), which took into account the features that were significant at univariate analysis, but excluded those that were significantly different between the two scanners, had an AUC of 0.66 (95%CI: 0.56–0.78) with a dichotomising threshold of 0.09. The Cox model (b2), built using the radiomic signature including all the significant features at univariate analysis, had an AUC of 0.68 (95%CI: 0.57–0.80) with a dichotomising threshold of 0.09. The Cox model built using both radiomic features and clinical predictors (c) had an AUC of 0.64 (95%CI: 0.53–0.75). Figure 2 shows the results for all the multivariate models for the PET dataset. No repeatable thresholds were found for the selected PET-based features.

PET+CT dataset

Table 4 shows the results of the univariate analysis and of the Cox regression models for the combined PET+CT analysis

using both CT and PET signatures. Figure 3 shows the results for all the multivariate models for the PET+CT data. The AUC for the different Cox regression models which included radiomic features were (b1) 0.62 (95%CI: 0.52–0.70), (b2) 0.68 (95%CI: 0.58–0.74), and (c) 0.65 (95%CI: 0.50–0.72), respectively. The dichotomising threshold was 0.09 and 0.07 for the radiomic b1 and b2 models, respectively.

Figure 4 shows PET and CT images of patients classified as low and high risk according to the PET+CT radiomic signature.

Discussion

The present study assessed the performance of a radiomic approach for outcome prediction in NSCLC patients undergoing surgery. The main finding of the study was that image-derived parameters outperformed common clinical predictors, including TNM staging.

Currently, the TNM staging system is the reference method for risk estimation and decision-making on treatment approach. However, the TNM staging system has some limitations. First, patients with disease at the same stage exhibit wide variations in the incidence of recurrence after curative resection [15]. Additionally, the pathological TNM stage may

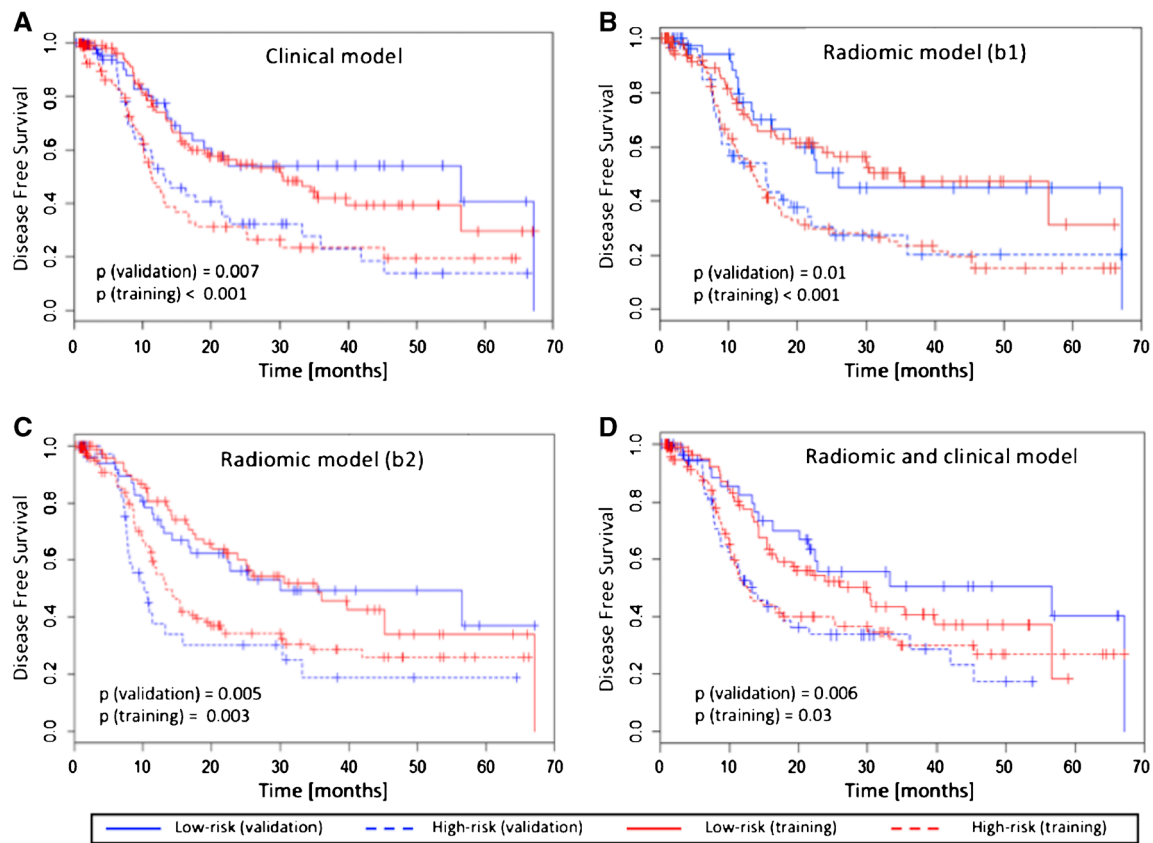


Fig. 2 Disease-free survival prediction results for the PET data. Kaplan-Meier curves for the DFS resulting from the Cox regression models built using clinical variables (a), the radiomic signature (b and c), and their combination (d) for the training and validation groups within the PET dataset

differ from the stage at diagnosis in patients who have undergone neo-adjuvant chemotherapy since the TNM staging system is based on the pathological assessment after resection [16]. Moreover, the N parameter may be affected by the number of lymph nodes removed during surgery, which is related to the surgeon's ability and the surgical approach [17, 18]. Finally, pathological TNM staging is not feasible in patients with advanced disease.

The results of the present study show that the radiomic signature of CT, PET, and PET+CT images is able to predict DFS in patients with NSCLC treated with surgery with curative intent. Improvement in the accuracy of prediction of patient outcome at baseline has multiple positive implications for medical and personal decision making. The identification of those patients in whom disease is likely to recur can impact on the therapeutic approach and patient management after primary treatment [15, 19]. There is growing evidence that radiomics can provide accurate risk stratification in NSCLC that may allow for individualised patient treatment [8, 20]. Moreover, this approach presents several logistic advantages: firstly, it is almost simultaneous with the baseline image acquisition and does not entail any additional procedure for the patient; secondly, it provides data on CT and PET images at the same time; finally, thanks to automated algorithms, it is relatively easy and fast. Texture analysis is non-invasive and

objective when automatic/semi-automatic approaches for segmentation are used; moreover, it seems not to be hampered by use of different scanners and is potentially cost-effective in predicting DFS in NSCLC.

Literature on the role of radiomics of both PET and CT images in predicting outcome in NSCLC after curative resection is relatively scarce [21–24]. Textural parameters of pre-treatment FDG-PET/CT proved to be predictive of recurrence [21] and survival [22, 23]. More interestingly, in a cohort of 119 patients (including 59 treated with surgery), Desseroit et al. [24] retrospectively built a nomogram combining the best features of different categories (clinical variables, volume and standard metrics, PET and CT texture features) in order to improve patient stratification. The nomogram including texture features on both CT and PET images provided higher stratification power than staging alone. However, none of the aforementioned studies performed a validation of the results. The goal of radiomic analysis is to obtain a prognostic or predictive model with a high accuracy and efficiency. Therefore, an external validation, or at least a cross-validation (internal validation), should always be performed to assess the accuracy of the model [8, 25]. Conventional image-derived parameters (e.g., SUV_{max}) have been used as reference models in radiomic studies [26]. Incorporation of the most

Table 4 Results of univariate analysis and Cox regression models for PET+CT images

PET dataset (<i>n</i> = 259) Parameter	Univariate <i>p</i> value	Cox regression <i>p</i> value			
		Clinical model (a)	Radiomic models (b)		Radiomic and clinical model (c)
			b1	b2	
Clinical predictor					
Age	0.09	n.s.	n.a.	n.a.	n.s.
Sex	0.1	n.s.	n.a.	n.a.	n.s.
Histological type	0.2	0.005	n.a.	n.a.	n.s.
Tumour stage	<0.001	<0.001	n.a.	n.a.	<0.001
CT feature					
Compacity_GLCM	0.02	n.a.	Not included	n.s.	n.s.
Homogeneity_GLCM	0.05	n.a.	n.s.	n.s.	n.s.
SRE_GLRLM	0.001	n.a.	0.05	0.006	0.01
LRE_GLRLM	0.04	n.a.	Not included	n.s.	n.s.
RP_GLRLM	0.05	n.a.	Not included	n.s.	n.s.
PET feature					
SUV _{mean}	0.008	n.a.	0.008	0.008	0.06
SUV _{max}	0.008	n.a.	0.05	n.s.	n.s.
Energy_GLCM	0.02	n.a.	Not included	n.s.	0.02
LRHGE_GLRLM	0.04	n.a.	n.s.	0.01	0.005
HGRE_GLRLM	0.05	n.a.	n.s.	n.s.	n.s.
LZHGE_GLZLM	0.02	n.a.	Not included	0.01	0.06
AUC (95%CI) for the validation cohort		0.61 (0.50–0.73)	0.62 (0.52–0.70)	0.68 (0.58–0.74)	0.65 (0.50–0.72)
DFS low-risk group (median±SD) [months]					
Training		33.2 ± 3.8	26.0 ± 3.8	36.7 ± 3.8	36.2 ± 3.6
Validation		21.7 ± 4.1	24.3 ± 4.3	27.3 ± 4.5	28.1 ± 4.2
DFS high-risk group (median±SD) [months]					
Training		11.6 ± 1.6	11.7 ± 1.6	12.4 ± 1.7	13.0 ± 1.7
Validation		12.1 ± 1.7	14.5 ± 2.1	11.3 ± 1.8	10.8 ± 1.3

Model b1: included features significant at univariate analysis and not significantly different between scanners. Model b2: included all features significant at univariate analysis. *AUC*, area under the curve; *95%CI*, 95% confidence interval; *CT*, computed tomography; *DFS*, disease-free survival; *GLCM*, grey-level co-occurrence matrix; *HGRE_GLRLM*, high grey-level run emphasis computed within the grey-level run-length matrix; *LRE_GLRLM*, long-run emphasis computed within the grey-level run-length matrix; *LRHGE_GLRLM*, long-run high grey-level emphasis computed within the grey-level run-length matrix; *LZHGE_GLZLM*, long-zone high grey-level emphasis computed within the grey-level zone-length matrix; *n.a.*, not applicable; *n.s.*, not significant (the exact value was not provided by *R* when >0.05); *PET*, positron emission tomography; *SD*, standard deviation; *RP_GLRLM*, run percentage computed within the grey-level run-length matrix; *SRE_GLRLM*, short-run emphasis computed within the grey-level run-length matrix; *SUV*, standardised uptake value

informative radiomic features into these reference models provides an indication of the gain in model performance. Patient and clinical characteristics might be added to the input variable list, as they may influence the outcome variable and the extracted radiomic features [25]. In our datasets, the model including both clinical predictors and the radiomic signature performed worse than the radiomic signature alone in CT, PET, and PET+CT analyses, in both training and validation groups.

Our findings are in line with recent studies that evaluated texture analysis of CT images alone in predicting patient

outcome. These investigations pointed out the prognostic role of radiomic features [27, 28], and validation of the results was also provided [28–30]. Moreover, texture features extracted from CT and cone-beam CT images can be considered comparable since they show high concordance values [30]. Textural parameters were found to be associated with survival [31] and recurrence [32] also in NSCLC patients treated with external beam radiotherapy. Considering these very promising data, efforts aiming at standardisation of methods for feature extraction and identification of reproducible radiomic features among centres are warranted.

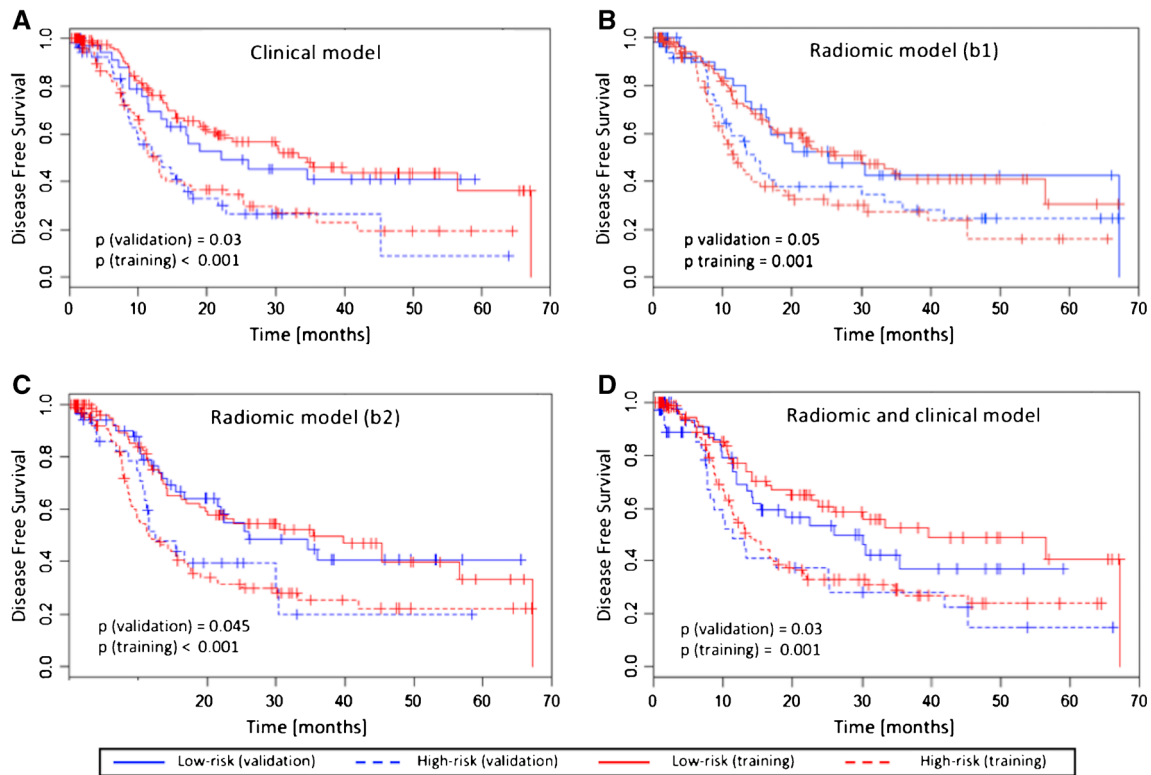


Fig. 3 Disease-free survival prediction results for the PET+CT data. Kaplan-Meier curves for the DFS resulting from the Cox regression models built using clinical variables (a), the radiomic signature (b and c), and their combination (d) for the training and validation groups within the PET+CT dataset

Recently, reconstruction methods have been reported to have an impact on the repeatability of radiomic features extracted from PET/CT images acquired using the same scanner in NSCLC [33, 34]. In particular, texture features proved more sensitive to a change in the segmentation than

to a change in the reconstruction method [33]. In our cohorts (CT and PET datasets) we tested feature variability between images acquired using two scanners and different acquisition parameters and reconstruction methods (Supplementary Table 1). Only 39% and 25% of the

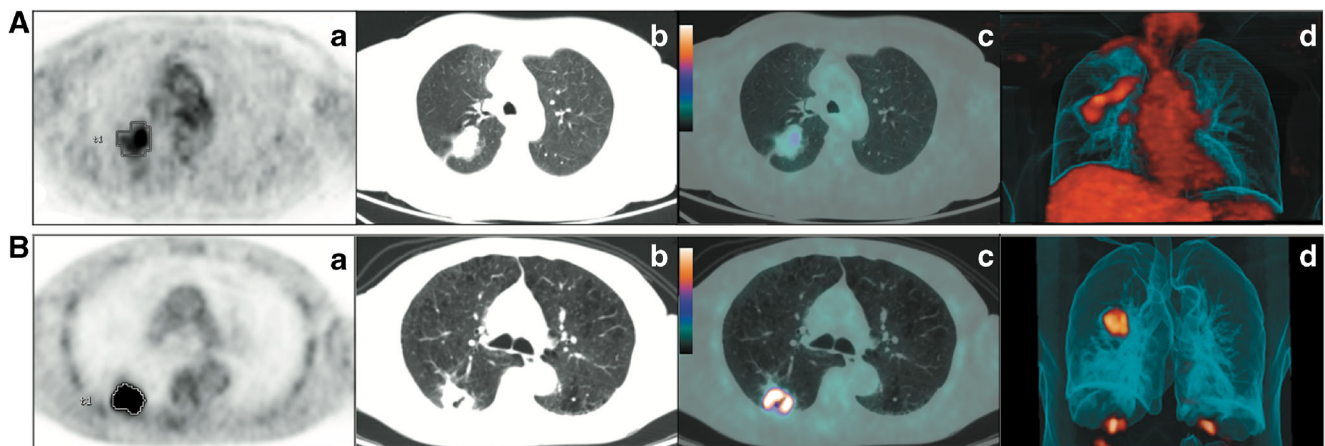


Fig. 4 Clinical cases PET/CT images of a low-risk (A) and a high-risk (B) patient. **A** Low-risk patient: axial PET, CT, PET/CT and a three-dimensional reconstruction of PET/CT images (a, b, c, and d, respectively) of a 64-year-old female with adenocarcinoma, pathological stage 2a (T2N0M0), with no evidence of disease 31 months after surgery. PET parameters: (\uparrow and \downarrow : above and below the median value) SUV_{mean} 2.0 \downarrow , $SUV_{maximum}$ 5.1 \downarrow , $Energy_{GLCM}$ 0.028 \uparrow , $LRHGE$ 75.6 \downarrow , $HGRE$ 46.7 \downarrow , $LZHGE$ 21298 \uparrow . CT parameters: $Compacity$ 3.10 \uparrow , $Homogeneity1$ 0.214 \downarrow , $Short-run\ emphasis$ 0.966 \uparrow ,

$Long-run\ emphasis$ 1.17 \downarrow and $Run\ percentage$ 0.953 \uparrow . **B** High-risk patient: axial PET, CT, PET/CT, and a three-dimensional reconstruction of PET/CT images (a, b, c and d, respectively) of a 60-year-old male, with squamous cell carcinoma, pathological stage 2a (T2N0M0), who experienced disease recurrence 11 months after surgery. PET parameters: SUV_{mean} 8.8 \uparrow , $SUV_{maximum}$ 15.4 \uparrow , $Energy_{GLCM}$ 0.002 \downarrow , $LRHGE$ 991.7 \uparrow , $HGRE$ 864.3 \uparrow , $LZHGE$ 5372 \uparrow . CT parameters: $Compacity$ 2.98 \uparrow , $Homogeneity1$ 0.281 \uparrow , $Short-run\ emphasis$ 0.938 \downarrow , $Long-run\ emphasis$ 1.34 \uparrow and $Run\ percentage$ 0.912 \downarrow

features were statistically different for the CT and PET images, respectively. Moreover, in the overall CT dataset, a reproducible threshold was found for each texture feature of the radiomic signature (Supplementary Fig. 4) for all bootstrap replicates in both the training and the validation dataset. This was not possible for PET images, likely because texture features from PET datasets are more sensitive to the acquisition and reconstruction parameters than CT features, resulting in fluctuating thresholds. Nonetheless, reproducible models were found in both the training and the validation set for CT, PET, and PET+CT images, also using features apparently different between scanners. Furthermore, the model with the best discrimination power between low- and high-risk patients was the one that included all the significant texture features, resulting in the highest AUC [0.75 (95%CI: 0.65–0.85) for the CT dataset].

Just focussing on the primary lesion evaluation, the added value of PET appears limited when comparing the AUC of the radiomic signature among CT, PET, and PET+CT (Tables 2, 3, and 4) in the prediction of DFS. In the CT analysis, radiomic features were more reliable since, as mentioned above, a stable threshold was identified for each texture feature in both the training and the validation groups. One might speculate that these findings, if confirmed in future investigations, would impact on the diagnostic and staging work-up of patients with lung tumours. The role of FDG-PET/CT could remain crucial in the assessment of metastatic spread of the disease.

Some limitations of this study have to be acknowledged. First, the study analysed in a retrospective fashion images acquired during clinical routine, using two different PET/CT scanners. Nonetheless, radiomic features were found to be significant and reproducible in the training and validation groups. This is in favour of the applicability of texture analysis in different centres using different scanners. An external validation is planned to confirm these results. Second, texture analysis of CT images was performed using the CT component of the PET/CT instead of high-resolution diagnostic CT images. This choice was related to the fact that even if all patients performed a high-resolution diagnostic CT as part of their pre-operative work-up, these images were not available in our institutional database for a large percentage of patients since in many cases they performed this examination in other hospitals and it was not repeated to avoid an additional radiation dose to the patients. Third, smoking habit and differentiation grading are recognised prognostic factors but they were not tested in the present study due to the unavailability of these data for the whole cohort of patients. However, while in clinical practice these factors have some influence on the choice of treatment approach, TNM stage, and comorbidities are the main determinants of treatment strategy.

Conclusions

The radiomic signature derived from baseline PET/CT images was predictive of disease-free survival in non-small cell lung cancer patients undergoing surgery. Our experience confirms the rationale for initiating studies in very large cohorts of patients with the aim of standardising and implementing this non-invasive approach in the clinical setting in order to facilitate imaging-based personalisation of treatment strategy.

Acknowledgements We thank Elena Vanni for support in patient selection; Paola Bossi and Dahoud Rahal for collaboration in pathological analyses; Marco Alloisio, Giulia Veronesi and the Thoracic Surgery Unit for close collaboration in patient selection and follow-up; Lorenzo Leonardi for image processing; and Riccardo Muglia, Nicolò Gennaro and Orazio Giuseppe Santonocito for their help in patient selection.

M.K. is supported by the AIRC (Italian Association for Cancer Research) scholarship funded by the grant won by A.C. (IG-2016-18585).

Author contributions M.S., M.K. and A.C. conceived the idea of the study; L.C., L.L. and A.F. performed the statistical analysis; E.V. collected the data and selected the patients; M.S., M.K. and L.A. reviewed and segmented the images; L.C. performed image analysis; M.S., M.K. and L.C. wrote the manuscript; A.R. edited and reviewed the manuscript.

All the authors discussed the results and commented on the manuscript.

Compliance with ethical standards

Disclosure of potential A. Chiti received speaker honoraria from General Electric and Sirtex Medical System, acted as scientific advisor for Blue Earth Diagnostics and benefited from an unconditional grant from Sanofi to Humanitas University. All honoraria and grants are outside the scope of the submitted work.

L. Cozzi acts as Scientific Advisor to Varian Medical Systems. All honoraria are outside the scope of the submitted work.

M. Kirienko is supported by the AIRC (Italian Association for Cancer Research) scholarship funded by the grant won by A.C. (IG-2016-18,585).

Conflict of interest All other authors have no conflicts of interest.

Research involving human participants The study was approved by the institutional Ethics Committee. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent For this type of study formal consent was not required.

References

1. EUCAN | Home page [Internet]. Available from: <http://eco.iarc.fr/EUCAN/Default.aspx>.

2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc.* 2008;83:584–94.
3. Vansteenkiste J, Crino L, Dooms C, Douillard JY, Faivre-Finn C, Lim E, et al. 2nd ESMO consensus conference on lung cancer: early-stage non-small-cell lung cancer consensus on diagnosis, treatment and follow-up. *Ann Oncol.* 2014;25:1462–74.
4. Zhang M, Zhang Z, Garmestani K, Schultz J, Axworthy DB, Goldman CK, et al. Pretarget radiotherapy with an anti-CD25 antibody-streptavidin fusion protein was effective in therapy of leukemia/lymphoma xenografts. *Proc Natl Acad Sci U S A.* 2003;100:1891–5.
5. Deterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest.* 2017;151:193–203.
6. Ost D, Goldberg J, Rolnitzky L, Rom WN. Survival after surgery in stage IA and IB non-small cell lung cancer. *Am J Respir Crit Care Med.* 2008;177:516–23.
7. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48:441–6.
8. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. *Sci Rep.* 2017;7:358.
9. Bashir U, Siddique MM, Mclean E, Goh V, Cook GJ. Imaging heterogeneity in lung cancer: techniques, applications, and challenges. *Am J Roentgenol.* 2016;207:534–43.
10. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One.* 2015;10:1–18.
11. Rami-Porta R, Crowley J, Goldstraw P. The revised TNM staging system for lung cancer. *Ann Thorac Cardiovasc Surg.* 2009;15:4–9.
12. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2014;42:328–54.
13. Orlhac F, Nioche C, Buvat I. Technical Appendix — Local Image Features Extraction — LIFEx —. 2016.
14. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: where do we stand? *J Nucl Med.* 2015;56:1642–4.
15. Uramoto H, Tanaka F. Recurrence after surgery in patients with NSCLC. *Transl Lung Cancer Res.* 2014;3:242–9.
16. Hellmann MD, Chaft JE, William WN, Rusch V, Pisters KMW, Kalhor N, et al. Pathological response after neoadjuvant chemotherapy in resectable non-small-cell lung cancers: proposal for the use of major pathological response as a surrogate endpoint. *Lancet Oncol.* 2014;15:1–17.
17. Velez-Cubian FO, Rodriguez KL, Thau MR, Moodie CC, Garrett JR, Fontaine JP, et al. Efficacy of lymph node dissection during robotic-assisted lobectomy for non-small cell lung cancer: retrospective review of 159 consecutive cases. *J Thorac Dis.* 2016;8:2454–63.
18. Korasidis S, Menna C, Andretti C, Maurizi G, D'Andrilli A, Ciccone AM, et al. Lymph node dissection after pulmonary resection for lung cancer: a mini review. *Ann Transl Med.* 2016;4:368.
19. Brundage MD. Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest.* 2002;122:1037–57.
20. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev.* 2016;1:207–26.
21. Kim D-H, Jung J, Son SH, Kim C-Y, Hong CM, Oh J-R, et al. Prognostic significance of intratumoral metabolic heterogeneity on 18F-FDG PET/CT in pathological N0 non-small cell lung cancer. *Clin Nucl Med.* 2015;40:708–14.
22. Tixier F, Hatt M, Valla C, Fleury V, Lamour C, Ezzouhri S, et al. Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. *J Nucl Med.* 2014;55:1235–41.
23. Apostolova I, Rogasch J, Buchert R, Wertzel H, Achenbach HJ, Schreiber J, et al. Quantitative assessment of the asphericity of pretherapeutic FDG uptake as an independent predictor of outcome in NSCLC. *BMC Cancer.* 2014;14:896.
24. Desseroit M-C, Visvikis D, Tixier F, Majdoub M, Perdrisot R, Guillemin R, et al. Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I-III. *Eur J Nucl Med Mol Imaging.* 2016;43:1477–85.
25. Larue RTHM, Defraene G, De Ruyscher D, Lambin P, Van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br J Radiol.* 2017;90:20160665.
26. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
27. Coroller T, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung. *Radiother Oncol.* 2015;114:345–50.
28. Yuan M, Zhang Y-D, Pu X-H, Zhong Y, Li H, Wu J-F, et al. Comparison of a radiomic biomarker with volumetric analysis for decoding tumour phenotypes of lung adenocarcinoma with different disease-specific survival. *Eur Radiol.* 2017;In press.
29. Parmar C, Leijenaar RT, Grossmann P, Rios-Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *Sci Rep.* 2015;5:11044.
30. van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiother Oncol.* 2017;123:363–9.
31. Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Int J Radiat.* 2016;120:258–66.
32. Li Q, Kim J, Balagurunathan Y, Liu Y, Latifi K, Stringfield O, et al. Imaging features from pre-treatment CT scans are associated with clinical outcomes in non-small-cell lung cancer patients treated with stereotactic body radiotherapy. *Med Phys.* 2017;44:4341–9.
33. Van Velden FHP, Kramer GM, Frings V, Nissen IA, Mulder ER, De Langen AJ, et al. Repeatability of radiomic features in non-small-cell lung cancer [(18F)]FDG-PET/CT studies: impact of reconstruction and delineation. *Mol Imaging Biol.* 2016;18:788–95.
34. Yan J, Chu-shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of image reconstruction settings on texture features in 18 F-FDG PET. *J Nucl Med.* 2015;56:1667–74.