


SCIENTIFIC REPORTS



OPEN

PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology

M. Sollini¹, L. Cozzi^{1,2}, L. Antunovic³, A. Chiti^{1,3}  & M. Kirienko¹

Received: 25 August 2016

Accepted: 23 February 2017

Published online: 23 March 2017

Imaging with positron emission tomography (PET)/computed tomography (CT) is crucial in the management of cancer because of its value in tumor staging, response assessment, restaging, prognosis and treatment responsiveness prediction. In the last years, interest has grown in texture analysis which provides an “*in-vivo*” lesion characterization, and predictive information in several malignancies including NSCLC; however several drawbacks and limitations affect these studies, especially because of lack of standardization in features calculation, definitions and methodology reporting. The present paper provides a comprehensive review of literature describing the state-of-the-art of FDG-PET/CT texture analysis in NSCLC, suggesting a proposal for harmonization of methodology.

Positron emission tomography (PET)/computed tomography (CT) using the radiopharmaceutical ¹⁸F-fluoro-deoxy-glucose (FDG) has a paramount role in the management of cancer patients owing to its value in tumor staging, response assessment, and restaging as well as in prognosis and prediction of treatment response. The standardized uptake value (SUV) obtained from FDG-PET scans is the most widely used parameter for lesion characterization and it has been shown to have a prognostic value¹. More recently, volumetric parameters, including metabolic tumor volume (MTV) and total lesion glycolysis (TLG), have also been proposed for assessment of prognosis^{2–5}. Moreover, in recent years there has been emerging evidence that the heterogeneity of density values on CT and of FDG uptake within the primary tumor can permit *in vivo* lesion characterization and provide predictive information in malignancies, including non-small cell lung cancer (NSCLC)^{6–10}.

The term “heterogeneity” conveys different meanings depending on the imaging modality: in FDG-PET it refers to the variability in the distribution of radiopharmaceutical uptake, while in CT it refers to the variability in tissue density. Lesion “heterogeneity” can be described by a multitude of mathematical methods that, taken together, constitute the “texture analysis” which provides numerous quantitative and semiquantitative indices, termed “features”^{11–13}. This approach as a whole is named “radiomics”.

Textural features seem to perform better than the conventional uptake parameters used for image interpretation in clinical routine, such as SUV measurements (e.g., SUV_{max}, SUV_{mean}), which are subject to several limitations¹⁴. Numerous studies have explored the additional information that can be extracted by texture analysis, with the aim of characterizing tumor lesions. However, these investigations have had multiple drawbacks and limitations, especially with respect to lack of standardization in feature calculation, definitions, and reporting methodology¹⁵.

Texture analysis has the potential to impact on patient management if its ability to characterize lesions *in vivo* and to provide predictive information is demonstrated in prospective studies. As lung cancer is the fourth most frequently diagnosed malignancy in Europe and the the leading cause of cancer mortality (<http://eco.iarc.fr/eucan>), texture analysis in such patients, when validated, will strongly impact on patient management and health-care systems. The present article provides a comprehensive review of the literature describing the state-of-the-art in FDG-PET/CT texture analysis methods in lung cancer. It also reports on the ability of textural features to identify tumor phenotype and to provide additional predictive and prognostic information in patients with NSCLC. Moreover, a comprehensive review of calculation methods, feature names, and definitions is performed, and a scheme for harmonization of methodology and reporting of results is proposed.

¹Department of Biomedical Sciences, Humanitas University, via Manzoni, 113-20089, Rozzano, (Milan), Italy.

²Radiotherapy and Radiosurgery Unit, Humanitas Clinical and Research Center, via Manzoni, 56-20089, Rozzano, (Milan), Italy. ³Nuclear Medicine Unit, Humanitas Clinical and Research Center, via Manzoni, 56-20089, Rozzano, (Milan), Italy. Correspondence and requests for materials should be addressed to M.S. (email: martina.sollini@cancercenter.humanitas.it)

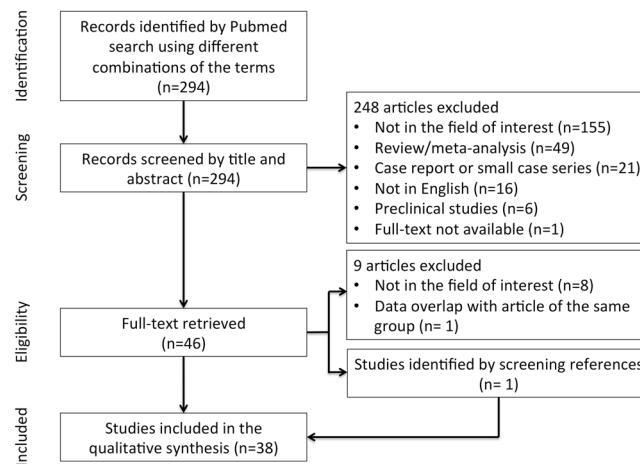


Figure 1. Schematic representation of the process of selection of literature data included in the review.

Methods

From the PubMed/MEDLINE database a search algorithm based on a combination of the following terms was used: (a) “texture” or “textural” or “radiomics” or “heterogeneity” or “heterogeneous” or “features” or “histogram” and (b) “lung cancer” or “NSCLC” and (c) “PET” or “PET/CT”. No start date limit was used and the search extended to 23 April 2016. To expand our search, references of the retrieved articles were also screened. Two authors independently searched articles and performed an initial screening of identified titles and abstracts. All studies or subsets in studies investigating the role of PET or PET/CT radiomics in patients with suspected/definite NSCLC were considered eligible. The exclusion criteria were: (a) articles not within the field of interest; (b) review articles, editorials or letters, comments, and conference proceedings; (c) articles not in the English language; (d) case reports or small case series (<10 patients); and (e) *in vitro* or animal studies.

Among the 294 studies identified by reviewing titles and abstracts, 247 were excluded by applying the criteria mentioned above. One paper not retrievable in the full-text version was excluded. Forty-six articles, retrieved in full-text version, were assessed for eligibility. Nine articles were excluded after reading the full text version. One further article was identified after screening of the references. Overall, 38 articles were selected and used for the qualitative synthesis (Fig. 1). Considering the aim of this review (i.e. texture analysis in NSCLC) which takes into account a variety of heterogeneous papers in terms of aim(s), methods, and results; a systematic review according to the PRISMA algorithm¹⁶ was considered not feasible. Selected papers were grouped into two different sets based on their aims: technical/methodological and clinical studies. The technical/methodological group included 16 papers (568 patients) that tested specific algorithms, different approaches for segmentation and tumor volume delineation, modalities of image acquisition, attenuation correction, or reconstruction. The clinical set comprised 22 papers (2306 patients), and we separately reviewed the results in respect of the diagnostic and the prognostic or predictive role of the textural features in NSCLC. Although prognosis and treatment prediction should be reported separately, in order to avoid an overlap of contents we treated these topics in the same section since many of the analyzed studies have evaluated them simultaneously. In approaching this review, the main difficulties were related to the differences in textural feature nomenclature and to the comparison of results regarding features obtained using different approaches. Therefore, we started with a reclassification of each feature (Fig. 2). For each feature we specified the order (i.e., first, second, or superior), the matrix (e.g., histogram, gray-level co-occurrence), the definition, and/or the formula, when available, as reported in the Supplementary material. Hereafter the features that are denominated in the identical way but can be derived using different approaches (e.g., histogram, gray-level co-occurrence matrix) are reported by adding the matrix from which they have been derived in subscript.

Results

Nomenclature and methods in texture analysis. Texture analysis refers to a variety of mathematical methods that may be applied to describe the relationships between the intensity of pixels or voxels and their position within an image. An advantage of measuring textural parameters is that it is a post-processing technique that can be applied to data acquired during routine clinical imaging protocols, thereby maximizing the information that can be derived from standard clinical images¹⁴. Distinct approaches (statistics based^{17,18}, model-based^{19–21}, transform-based^{22–24}, and structural²⁵) may be used to analyze functional imaging information, resulting in numerous radiomics features, such as descriptors of the image intensity histogram, “shape and size” features, descriptors of the relationships between image voxels (e.g., GLCM-, and NGTDM-derived features), textures extracted from wavelet and Laplacian of Gaussian filtered images, and fractal features²⁶, as shown in Fig. 2.

The first approach consists in the summarizing of 3D functional imaging data into a single curve – histogram – representing the voxel intensity values contained within the volume of interest (VOI), allowing for a simplified interpretation. Intensity-volume histograms (IVH) or cumulative SUV-volume histograms (CSH) have been proposed by El Naqa *et al.*²⁷ as a novel way to characterize heterogeneity in tumor tracer uptake. These histograms are similar to dose-volume histograms frequently used in radiotherapy²⁸. A set of metrics are derived from IVH

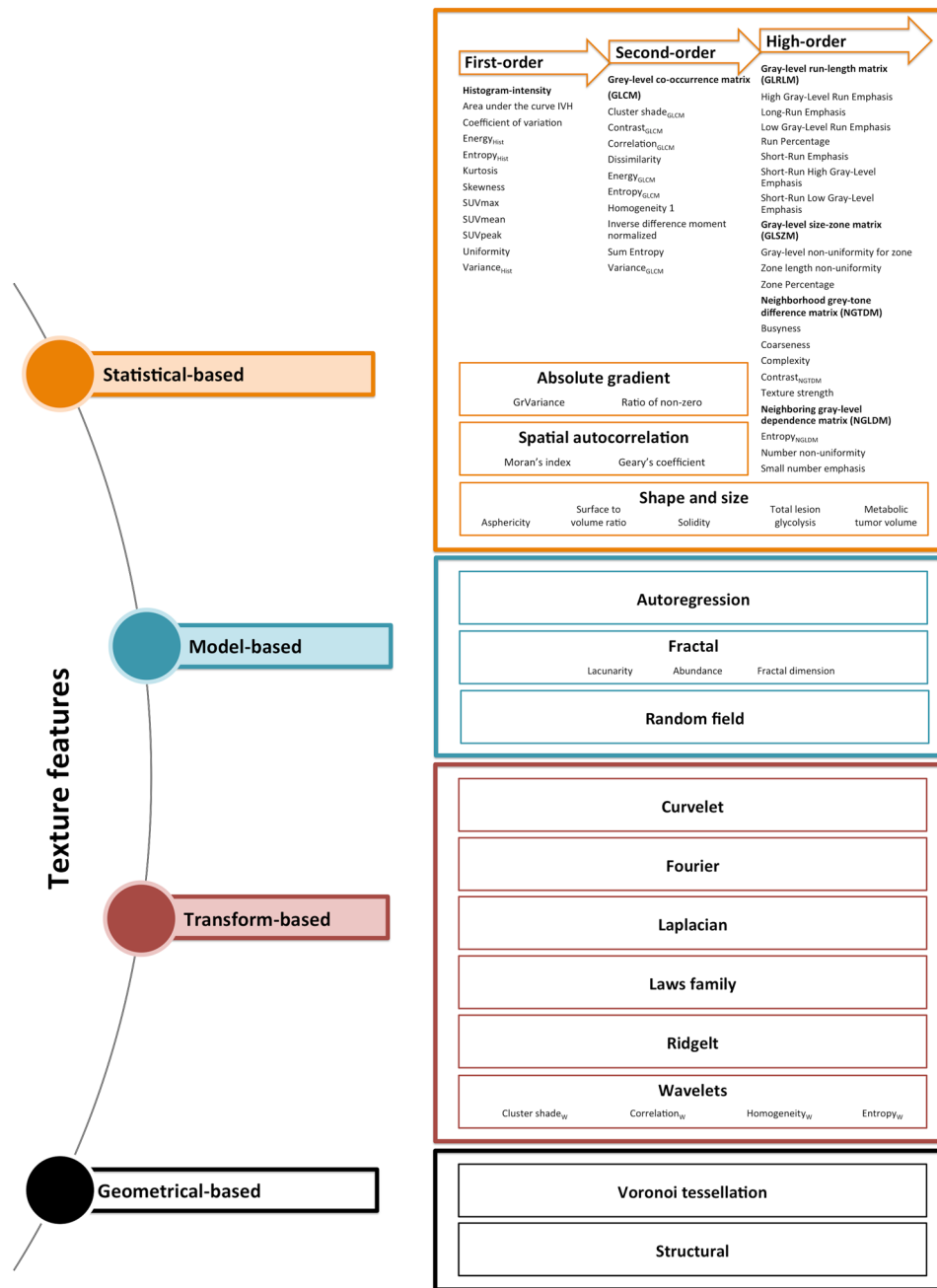


Figure 2. Methodological approaches in image texture analysis (the most frequently evaluated PET features in lung cancer patients are reported as examples).

representations that reflect the voxel value frequency distribution²⁷: I_x (minimum intensity to $x\%$ highest intensity volume); V_x (percentage volume having at least $x\%$ intensity value); and descriptive statistics (mean, minimum, maximum, standard deviation, etc.)²⁹. In CSH the percent volume of a tumor (derived from CT or from PET-based (semi-)automatic tumor delineation methods³⁰) with an SUV above a certain threshold is plotted against that threshold value, which is varied from 0 to 100% of SUV_{max} . The area under the CSH (AUC-CSH) may be a quantitative index of tracer uptake heterogeneity and/or heterogeneous tumor response³¹. Any method to characterize heterogeneity, however, will treat both partial volume effect and noise as heterogeneity. Therefore partial volume correction (PVC) and image denoising should be applied prior to calculating AUC-CSH³². Despite this consideration, PVC has not been routinely applied. To overcome this potential limitation on the quantitative measurements, most studies have considered relatively large lesions (generally volumes $>3-5\text{ cm}^3$), assuming that PET cannot characterize heterogeneity in smaller volumes because of its limited spatial resolution.

IVH and other first-order approaches are limited by their spatial insensitivity. To overcome this drawback, textural features and “shape and size” attributes may be extracted that contain embedded spatial and topological information. In fact, second-order and high-order statistics (i.e., based on gray-level matrix, nearest neighbor

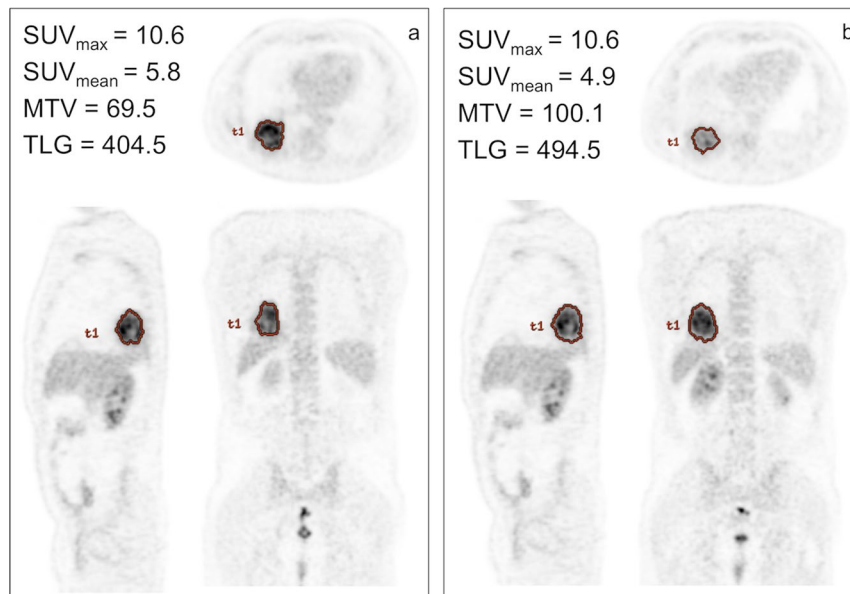


Figure 3. Example of tumor contouring using in (a) a threshold method at 50% of SUV_{max} and (b) a method based on an absolute SUV cut-off of 2.5. The ROI identified by using the absolute SUV cut-off of 2.5 is greater than that identified by the threshold method, as shown by axial (top), sagittal (right), and coronal (left) images (same slices).

spatial dependence matrices, etc.) provide information from the spatial relationship of image voxels³³. The gray-level co-occurrence matrix features may be used to represent texture information because of its relatively simple and intuitive structure. Surface plots of the co-occurrence matrix give a pictorial representation of the spatial-intensity distribution, which is typically masked by first-order histogram analyses. Several other matrices are also used, including the neighborhood gray-tone difference matrix, which provides information regarding how each voxel value differs from the neighboring voxel values; the gray-level run length matrix, which stores the number of voxels with identical values in each direction; and the gray-level size zone matrix, which stores the size of the 3D region that includes a given voxel value^{27, 29, 33–37}. Therefore, a multitude of metrics may be derived from the gray-level matrices to characterize the structure of interest²⁷. These metrics are independent of tumor position, orientation, size, and brightness and take into account the local intensity-spatial distribution^{33, 38, 39}. Hence, the combination of these features can provide an intensity-spatially dependent map of the tumor metabolic uptake that can potentially be used as a signature to characterize the tumor phenotype and response to treatment.

Texture feature extraction requires the voxel intensity values within the VOI to be discretized. This discretization step not only reduces image noise, but also normalizes intensities across all patients, allowing for a direct comparison of all calculated textural features between patients. Shape and size features are calculated, describing the 3D shape and size of the lesions⁴⁰. However, it should be considered that volumetric indexes (e.g., sphericity) may also be extracted from the IVH^{40–42}, allowing for a simplified interpretation²⁷.

The majority of texture features that have been used in PET medical imaging to date fall into one of the following three categories: (a) first-order features derived from statistical moments of the image intensity histogram, (b) second-order features derived from the GLCM, and (c) higher order features derived from analysis of the NGTDM, NGLDM, or GLSZM⁴³. Despite the difficulties in generalization due to the variability in textural PET features among studies, related to the methodology used, we can summarize that IVH features tend to depend on the tumor delineation method^{44–46} and that features derived from GLSZM have been reported to be the most susceptible to variability^{40, 46, 47}, while the GLCM-derived features tend to be the most robust. In particular, “entropy_{Hist}” and “entropy_{GLCM}” have been reported to be features less dependent on the tumor segmentation method⁴⁶, reconstruction settings, iteration numbers, and voxel size⁴⁷, and type of acquisition (3D versus 4D)^{42, 48}.

Texture analysis and technical/methodological investigations. The accuracy and precision of texture analysis derived from PET images depends significantly on scanning protocols. Factors such as image acquisition, reconstruction, and inherent image quality parameters (noise, motion artifacts, and slice thickness) may be important. It is to be expected that all texture analysis methods are influenced to some extent by these factors and the sensitivity of various textural features may be based on different image models. Further aspects that require careful assessment are the methods used for region of interest (ROI) definition on PET images and the intraobserver and interobserver variations¹⁴. Figure 3 shows examples of different methods used for ROI definition. All of these aspects have been evaluated in the determination of PET features in NSCLC patients (Table 1).

Hatt *et al.*⁴⁴ evaluated the impact of five different methods of segmentation on anatomic tumor volume, MTV, and heterogeneity (“coefficient of variation” histogram-based) in a small group of NSCLC patients. They found

| Reference | Type of study | Patients, n | Setting, stage | Aspect evaluated | Lesion segmentation method | PET parameter and textural index matrix | Main results |
|-------------------------|---------------|--|-------------------------|--|--|--|--|
| Cheng ⁴⁸ | R | 56 | Staging, I–III (only T) | Impact of respiration-averaged CT on PET texture parameters | Adaptive threshold, threshold uptake 45% of the SUV_{max}^* | FOS/IVH = 3 SS = 1 GLCM = 4 GLRLM = 3 NGTDM = 4 | Texture parameters obtained with helical and respiration-averaged PET/CT showed a high degree of agreement (SUV entropy and entropy had the lowest levels of variation) |
| Cui ⁵⁰ | n.r. | 20 | n.r. | Impact of the segmentation method on tumor volume estimation (validation of DM algorithm) | Automatic (DM), fuzzy C-means, threshold uptake 40% of the SUV_{max} , threshold uptake 50% of the SUV_{max} , tumor-customized downhill, watershed [§] | FOS/IVH = 1 NGTDM = 1 Gr = 1 | DM algorithm was able to segment the tumor (also when adjacent to mediastinum or chest wall) and outperformed the other lung segmentation methods in terms of overlapping measure |
| Cui ⁵¹ | n.r. | 40 | n.r. | Impact of the segmentation method on tumor volume estimation (validation of topo-poly algorithm) | Threshold uptake 40% of the SUV_{max} , threshold uptake 50% of the SUV_{max} , adaptive threshold, fuzzy C-means, tumor-customized downhill, random walks, high-order interactive learning segmentation, PET/CT tumor-background likelihood model, topo-poly [§] | NGTDM = 1 | Topo-poly algorithm was able to delineate tumor margins better than other methods |
| Dong ⁴⁵ | R | 50 | Staging, I–IV | Impact of the segmentation method on tumor volume estimation | Absolute SUV cut-off of 2.5, manual (2 observers), threshold at 40% of the SUV_{max}^* | FOS/IVH = 1 SS = 1 GLCM = 1 + visual score | Intratumor heterogeneity significantly correlated with differences in the GTV definition (high heterogeneity corresponded to a larger GTV) |
| Gao ⁵⁷ | n.r. | 132 | Staging, I–III | Impact of computer-based algorithm on diagnosis of mediastinal lymph node metastases (validation of computer-based algorithm) | Manual [†] | FOS/IVH = 3 GLCM = 5 + visual score | Diagnostic ability of computer-based algorithm and visual experience was similar |
| Hatt ⁴⁴ | n.r. | 25, only 17 analyzed | Staging, Ib–IIIb | Impact of the segmentation method on the tumor volume estimation | Adaptive threshold, fully automatic method (FLAB), manual, threshold at 50% of the maximum [*] | FOS/IVH = 1 SS = 1 | All delineation methods except the manual one resulted in underevaluation of MTV. Anatomic tumor size and heterogeneity were correlated (larger lesions were more heterogeneous) |
| Hofheinz ⁴⁹ | n.r. | 30 | n.r. | Impact of the segmentation method on tumor volume estimation (validation of voxel-specific threshold algorithm) | Lesion-specific threshold, manual, voxel-specific threshold ^{**} | FOS/IVH = 2 SS = 1 | Voxel-specific threshold method was able to reproduce tumor boundaries accurately, independent of the heterogeneity |
| Leijenaar ⁴⁰ | n.r. | 11 (test-retest cohort) + 23 (inter-observer cohort) | | Features' test-retest reliability and interobserver stability among multiple tumor delineation methods | Manual (by 5 observers), threshold at 50% of the maximum | FOS/IVH = 54 SS = 8 GLCM = 22 GLRLM = 11 GLSZM = 11 | The majority of features had high test-retest (71%) and interobserver (91%) stability in terms of ICC |
| Leijenaar ⁵² | P | 35 | Staging, I–III | Comparison of different discretization methods for textural features | Manual (SUV discretization using a fixed bin size and a fixed number of bins) | GLCM = 22 GLRLM = 11 GLSZM = 11 | SUV discretization had a crucial effect on textural features |
| Oliver ⁴² | R | 23 | | Sensitivity of texture features to tumor motion by comparison of static (3D) and respiratory-gated (4D) PET imaging | Adaptive threshold (background-adapted thresholding method) [*] | FOS/IVH SS GLCM GLRLM (total 56) | Quantitative analysis using a 3D versus 4D acquisition provided notably different image feature values, mainly due to the impact of respiratory motion |
| Orlhac ⁴⁶ | P | 24 | Staging, III | Impact of the segmentation method on the tumor volume estimation | Threshold at 40% of the maximum, adaptive threshold ^{**} | FOS/IVH = 8 SS = 1 GLCM = 6 GLRLM = 11 GLSZM = 11 NGLDM = 3 | IVH-based indices strongly depended on the tumor delineation method; 17/31 second- or high-order statistic features were robust with respect to tumor segmentation. Several texture indices included similar information. Some texture indices were highly correlated with MTV |
| Orlhac ⁵³ | R | 48 | Staging, I–III | Impact of resampling step on textural features and on the ability of textural features to reflect tissue-specific patterns of metabolic activity | Adaptive threshold (relative resampling approach and absolute resampling approach) ^{**} | FOS/IVH = 1 SS = 1 GLCM = 2 GLRLM = 3 GLSZM = 2 | Textural features computed using an absolute resampling method varied as a function of the tissue type and cancer subtype more than when using the usual relative resampling approach |
| Tixier ⁵⁵ | P | 20 | Staging, I–II | Impact of static and parametric acquisition on PET features | Fully automatic method (FLAB) ^{**} | FOS/IVH = 2 SS = 3 GLCM = 3 GLSZM = 2 | Compared with static SUV images, parametric images did not provide significant complementary information concerning heterogeneity quantification |

Continued

| Reference | Type of study | Patients, n | Setting, stage | Aspect evaluated | Lesion segmentation method | PET parameter and textural index matrix | Main results |
|--------------------------|---------------|-------------|------------------|---|--|--|---|
| van Velden ⁴¹ | P | 11 | Staging, IIIb–IV | Repeatability of texture features using different reconstruction settings and delineation methods | Threshold uptake 50% of the 3D SUV _{peak} on EANM-compliant (reconstruction method 1) and PSF-based (reconstruction method 2) images [°] | FOS/IVH = 29 FF = 3 SS = 10 GLRLM = 22 GLCM = 44 L = 1 SA = 2 | The majority of features had a high level of repeatability (ICC ≥ 0.90 for 63 features). Features were more sensitive to a change in delineation method (n = 25) than a change in reconstruction method (n = 3) |
| Yan ⁴⁷ | R | 17 | n.r., I–IV | Variability of PET textural features using different reconstruction methods, iteration numbers, and voxel size | Threshold uptake 40% of the SUV _{max} ^{°°} | FOS/IVH = 6 GLCM = 21 GLRLM = 11 GLSZM = 13 NGLDM = 5 NGTDM = 5 | Image features had different sensitivities to reconstruction settings (entropy _{HIST} difference entropy, inverse difference normalized, inverse difference moment normalized, low gray-level run emphasis, high gray-level run emphasis, and low gray-level zone emphasis were the most robust features; skewness, cluster shade, and zone percentage exhibited large variations) |
| Yip ⁵⁶ | R | 26 | Staging, n.r. | Sensitivity of texture features to tumor motion by comparing static (3D) and respiratory-gated (4D) PET imaging | Threshold uptake 40% of the SUV _{max} | GLCM = 1 GLRLM = 1 NGTDM = 4 | 4D-PET derived textures were less susceptible to tumor motion and may have greater prognostic value |

Table 1. Publications reporting methodological investigations on texture analysis in NSCLC patients. FF: fractal features; FLAB: fuzzy locally adaptive Bayesian; FOS/IVH: first-order statistics/intensity-volume histogram; GLCM: gray-level co-occurrence matrix; GLRLM: gray-level run-length matrix; GLSZM: gray-level size-zone matrix; Gr: absolute gradient; ICC: intra-class correlation coefficient; L: Laplacian; LF: Laws family; n.a.: not available; n.r.: not reported; NGLDM: neighboring gray-level dependence matrix; NGTDM: neighborhood gray-tone difference matrix; P: prospective; R: retrospective; SA: spatial autocorrelation; SS: shape and size; W: wavelet *Segmentation of only primary lung lesion. [°]Segmentation of lymph nodes. ^{°°}Segmentation of primary lung lesion and other tissues (e.g. lymph nodes). [°]Included in the analysis only lung lesion with a volume > of a minimum cut-off (e.g. 3 mL).

that all delineation methods except the manual one resulted in an underestimation of MTV, and that larger lesions were more heterogeneous.

Similar results were obtained in a larger population (n = 50) with NSCLC. Tumor volume was observed to be significantly diverse using different approaches (manual or automatic) on CT and on fused PET/CT images (volumes delineated on CT were larger than those defined on PET images). Intratumor heterogeneity, defined by visual scoring, “coefficient of variation”, or “entropy_{GLCM}” (gray-level co-occurrence matrix – GLCM) significantly correlated with differences in the target volume [tumors with a high heterogeneity showed a larger gross tumor volume (GTV)], suggesting that caution should be exercised when applying relatively simple threshold-based segmentation to define the target volume for tumors with high heterogeneity⁴⁵.

Hofheinz *et al.*⁴⁹ developed and tested a voxel-specific threshold algorithm as a delineation method for heterogeneous tumors. This method, which can be considered as an extension of an adaptive threshold method, proved able to reproduce the true tumor boundaries accurately, without being influenced by the heterogeneity (“coefficient of variation”).

Cui *et al.*⁵⁰ developed an automatic algorithm that used the PET SUV volume and the CT volume to localize and segment tumor lesions. This algorithm outperformed other (semi-)automatic methods in terms of overlapping measure, and they found that the feature “contrast_{NGTDM}” (NGTDM: neighbor gray-tone difference matrix based) was valuable in automatic tumor localization. The same group developed and tested a “topo-poly” algorithm (which incorporated an intensity graph and a topology graph) in two groups of patients defined as having ‘isolated’ (i.e., lung tumor located in the lung parenchyma and away from associated structures/tissues in the thorax) or ‘complex’ (i.e., tumor abutted/involving a variety of adjacent structures, where the tumor margins were indistinct and/or had heterogeneous regions of FDG uptake) disease. This method provided better anatomic and functional boundary delineations for both small and large tumors and for ‘complex’ cases. Again, “contrast_{NGTDM}” was valuable in automatic tumor localization⁵¹.

Leijenaar *et al.*⁴⁰ tested more than 100 PET features (first-order statistics and intensity volume histogram – FOS/IVH, “shape and size” features – SS, GLCM, gray-level run-length matrix – GLRLM, and gray-level size-zone matrix – GLSZM) to evaluate their test–retest reliability and interobserver stability among different tumor delineation methods in 34 NSCLC patients. Considering all features, a good overall similarity in feature stability was observed, based on rankings in terms of test–retest and interobserver intra-class correlation coefficient (ICC, $p \ll 0.001$). Comparing stability rankings per feature group, a high similarity was found for both the first-order statistics ($p \ll 0.001$) and other textural features ($p \ll 0.001$). Features based on GLSZM had the overall lowest ranks, indicating that these features have the highest variability. For the IVH features the observed similarity was more moderate ($p \ll 0.001$). Comparison of the rankings for the geometric features proved non-significant ($p = 0.086$). Overall, more stable features on repeated PET scans were also more robust against interobserver variability. In a similar number of patients, the same group evaluated prospectively different discretization methods

| Reference | Type of study | Patients, n | Setting, stage | Aspect investigated | Lesion segmentation method | PET features and textural index matrix | Main results |
|--------------------------|---------------|-------------|------------------|---|--|---|--|
| Apostolova ⁷³ | R | 60 | Staging, I–III | Prognostic value of asphericity | Adaptive threshold method** | FOS/IVH = 2 SS = 4 | Asphericity was a predictor of progression-free survival and overall survival |
| Budiawan ⁵⁹ | R | 44 | Staging, I–IV | Ability of PET features to predict lymph node metastases | Manual* ^o | FOS/IVH = 4 + visual score | Metastatic lymph nodes had higher heterogeneity (coefficient of variation) than inflammatory ones |
| Carvalho ⁸⁰ | n.r. | 220 | Staging, I–IIIb | Prognostic value of heterogeneity based on PET textural features | Absolute SUV cut-off values of 2.5, 3, and 4, threshold at 40% and 50% of SUV _{max} | FOS/IVH = 8 SS = 1 | Best prognostic value for overall survival was found for relative portions of the tumor above higher uptakes (80% SUV) |
| Cook ⁷⁷ | R | 53 | Staging, I–III | Ability of PET features to predict prognosis and disease progression after concurrent chemoradiotherapy | Threshold at 45% of the SUV _{max} * | FOS/IVH = 3 SS = 2 NGTDM = 4 | Coarseness, contrast, and busyness were associated with response to chemoradiotherapy and prognosis |
| Cook ⁷⁸ | P | 47 | Staging, IIIb–IV | Ability of PET features to predict prognosis and disease progression after erlotinib | Threshold at 40% of the SUV _{max} * | FOS/IVH = 8 SS = 2 NGTDM = 4 | Heterogeneity predicted response to erlotinib. Changes in entropy _{ITM} (baseline and 6 weeks) were independently associated with overall survival and treatment response |
| Desseroit ⁸³ | R | 116 | Staging, I–III | Develop a nomogram by exploiting intratumor heterogeneity (PET and CT features) to identify patients with the poorest prognosis | Fully automatic method (FLAB) | FOS/IVH = 3 SS = 1 GLCM = 2 GLSZM = 2 (+ 35 on CT images) | Intratumor heterogeneity could be used to create a nomogram with a higher stratification power than staging alone (poorest prognosis: stage III, large tumor volume, high PET heterogeneity, and low CT heterogeneity) |
| Fried ⁸¹ | R | 195 | Staging, III | Ability of PET features to enhance overall survival risk stratification | Manual ⁵ | FOS/IVH = 8 SS = 3 GLCM = 4 | Imaging features (solidity and primary tumor energy) improved risk stratification |
| Fried ⁸² | R | 225 | Staging, III | Ability of PET features to identify patients who might benefit from a higher radiation dose compared with that for the entire stage III | Semiautomatic gradient based ⁸ | FOS/IVH = 1 SS = 3 GLCM = 1 | Imaging features were found to be capable of isolating subgroups of patients who received a benefit or detriment from dose escalation |
| Ha ⁶⁰ | R | 30 | Diagnostic, n.r. | Correlation between metabolic heterogeneity and histopathologic characteristics | Adaptive threshold* | FOS/IVH = 1 GLCM = 21 Gr = 2 | The majority of texture features analyzed (including SUV _{max}) differed between Adk and Sqc |
| Hatt ⁷⁴ | R | 101 | Staging, I–III | Relationship between tumor MTV and derived heterogeneity measurements | Fully automatic method (FLAB)** | FOS/IVH = 3 SS = 1 GLCM = 2 GLSZM = 2 | Correlation between MTV and textural features varied greatly depending on the MTV (reduced correlation for increasing volumes) |
| Kang ⁷⁵ | R | 116 | Staging, III | Ability of PET features to predict disease progression after concurrent chemoradiotherapy | Absolute SUV cut-off value of 3.0* | FOS/IVH = 2 SS = 1 | Intratumoral heterogeneity predicted disease progression after chemoradiotherapy in inoperable stage III NSCLC |
| Kim ⁶¹ | R | 119 | Staging, I | Ability of PET features to predict prognosis after curative surgical resection in pathologically N0 tumor | Absolute SUV cut-off value of 2.5* | FOS/IVH = 2 SS = 2 | Heterogeneity of primary tumor was predictive of recurrence in pN0 Adk but not in Sqc |
| Lovinfosse ³⁶ | R | 63 | Staging, I | Ability of PET features to predict prognosis after radiotherapy | Fully automatic method (FLAB)* | FOS/IVH = 7 SS = 2 GLCM = 6 GLSZM = 2 NGTDM = 3 | Intratumoral heterogeneity (dissimilarity) appeared to be a strong independent outcome predictor after radiotherapy |
| Miwa ³⁷ | R | 54 | Diagnostic, n.a. | Ability of PET and CT features to differentiate malignant from benign pulmonary nodules | Threshold at 40–100% (intervals of 2%) of SUV _{max} | FOS/IVH = 1 FF = 1 (+1 on CT images) | Intratumoral heterogeneity could help to differentiate malignant and benign pulmonary nodules (better diagnostic ability of density fractal dimension on PET than morphological fractal dimension on CT) |

Continued

| Reference | Type of study | Patients, n | Setting, stage | Aspect investigated | Lesion segmentation method | PET features and textural index matrix | Main results |
|-------------------------------|---------------|---|---|---|--|---|--|
| Nair ⁷¹ | R | 172 (study cohort = 25, external cohort = 63, validation cohort = 84) | Staging, I–IV (study cohort) and I–II (validation cohort) | Identify individual genes and gene expression signatures associated with prognostically relevant PET features | Adaptive threshold method* | FOS/IVH = 10 SS = 3 | Four genes (LY6E, RNF149, MCM6, FAP) associated with textural features were also associated with survival |
| Ohri ⁷⁶ | P | 250, only 201 analyzed | Staging, IIB–III | Prognostic value of heterogeneity based on PET textural features | Semiautomatic gradient-based | FOS/IVH SS GLCM GLRLM GLSZM NGTDM NGLDM (total 45) + visual score | SumAver _g was an independent predictor of overall survival |
| Pyka ⁷⁹ | R | 45 | Staging, I | Ability of PET features to predict prognosis and disease progression after radiotherapy | Absolute SUV cut-off values of 2.0 and 2.5* | FOS/IVH = 3 SS = 1 GLCM = 2 NGTDM = 3 | Tumor heterogeneity was associated with response to radiation therapy |
| Tixier ³⁵ | R | 108, only 102 analyzed | Staging, I–III | Prognostic value of heterogeneity | Fully automatic method (FLAB) ^{***} | FOS/IVH = 3 SS = 2 GLCM = 3 GLSZM = 3 + visual score | High SUV, large metabolic volumes, and high heterogeneity were associated with poorer overall survival and recurrence-free survival |
| Vaidya ²⁹ | R | 27 | Staging, I–IV | Ability of PET and CT features to predict disease progression after radiotherapy | Manual | FOS/IVH = 12 SS = 2 GLCM = 4 (+ 32 on CT images) | IVH parameters (I_x metrics for PET and V_x metrics for CT) yielded the highest association with locoregional control |
| van Gómez López ⁶² | R | 38 | Staging, I–IIIa | Correlation between metabolic heterogeneity and pathologic staging | Absolute SUV cut-off value of 2.5* | FOS/IVH = 2 SS = 2 GLCM = 5 | Tumor heterogeneity was correlated with global metabolic parameters, and both were associated with macroscopic tumor diameter and, under special conditions (exclusion of a small tumor with high AJCC stage), with the AJCC stage |
| Win ¹³ | P | 122 (study cohort = 56, validation cohort = 66) | Staging, I–IV | Ability of PET and CT features to predict survival | Threshold at 42% of the SUV_{max} * | FOS/IVH = 2 (+ 1 on CT images) | PET-derived heterogeneity was predictive of survival at univariate analysis; at multivariate analysis only CT-derived heterogeneity, stage, and permeability were independent predictors of survival |
| Wu ⁸⁴ | R | 101 (study cohort = 70, validation cohort = 31) | Staging, I | Ability of PET features to predict distant metastases | Fully automatic method* | FOS/IVH = 11 SS = 2 GLCM = 3 W = 24 LF = 30 | The optimal prognostic model for identifying groups at risk of developing distant metastasis included $SUV_{peak2mL}$ and Gauss cluster shade _{1aws} |

Table 2. Publications reporting studies on the diagnostic, prognostic and predictive role of texture analysis in NSCLC patients. Adk: adenocarcinoma type; FF: fractal features; FLAB: fuzzy locally adaptive Bayesian; FOS/IVH: first-order statistics/intensity-volume histogram; GLCM: gray-level co-occurrence matrix; GLRLM: gray-level run-length matrix; GLSZM: gray-level size-zone matrix; Gr: absolute gradient; ICC: intra-class correlation coefficient; L: Laplacian; LF: Laws family; n.a.: not available; n.r.: not reported; NGLDM: neighboring gray-level dependence matrix; NGTDM: neighborhood gray-tone difference matrix; P: prospective; R: retrospective; SA: spatial autocorrelation; Sqc: squamocellular types; SS: shape and size; W: wavelet. *Segmentation of only primary lung lesion. #Segmentation of lymph nodes. §Segmentation of primary lung lesion and other tissues (e.g. lymph nodes). ^Application of partial volume correction. °Included in the analysis only lung lesion with a volume > of a minimum cut-off (e.g. 3 mL).

(fixed bin size *versus* fixed number of bins) for textural feature extraction in the context of treatment response assessment. Textural feature values were shown to depend on the intensity resolution used for SUV discretization. Discretizing SUVs using a fixed number of bins was found to be less appropriate for inter- and intrapatient comparison of textural feature values in a clinical setting. Additionally, results obtained for the features could not be directly compared when different intensity resolutions were used, suggesting that their interpretation (e.g., prognostic or predictive value) depended on the intensity resolution. It is noteworthy that the “ $correlation_{GLCM}$ ” was the only feature observed to have highly similar patient rankings over the course of treatment, regardless of the discretization method or discretization value used⁵².

Orlhac *et al.*⁴⁶ investigated a consistent number of texture indices on a variety of tumors (including 24 NSCLC) to gain a better insight into how they relate to one another and to conventional indices such as SUV, MTV, and TLG and to determine the extent of their robustness with respect to the gray-level resampling scheme and formula and to the tumor delineation method. All histogram indices strongly depended on the tumor delineation

method. Similarly, “contrast_{NGTDM}”, “busyness” (NGTDM-based), “low gray-level run emphasis”, “short-run low gray-level emphasis”, “long-run low gray-level emphasis” (GLRLM-based), “low gray-level zone emphasis”, and “short-zone low gray-level emphasis” (GLSZM-based) were highly sensitive to the segmentation method, while “homogeneity_{GLCM}” and “entropy_{GLCM}” were found to be robust with respect to tumor segmentation. The same group investigated, in 48 treatment-naïve NSCLC patients, the effect of the resampling approach on the ability of textural features to reflect tissue-specific patterns of metabolic activity. An adaptive threshold method was used to delineate tumors. The relative resampling approach (RR) was compared with the absolute resampling (AR) approach. Seven features (from GLCM, GLRLM, and GLSZM) were calculated and correlated with tissue types and cancer subtypes. AR-based “entropy_{GLCM}” could differentiate between tumor and healthy tissue ($p < 0.0001$). Using the AR method, tumor tissue exhibited higher “high gray-level zone emphasis” than healthy tissue, while tumors had lower “homogeneity_{GLCM}” and “low gray-level zone emphasis”. AR-based textural features differed adenocarcinoma (Adk) and squamocellular carcinoma (Sq) ($p \leq 0.05$)⁵³.

Yan *et al.*⁴⁷ tested the variability of more than 60 PET-textural features using different reconstruction settings, different iteration numbers, and different voxel size in 17 NSCLC patients. “Skewness” (IVH-based), “cluster shade_{GLCM}”, and “zone percentage” (GLSZM-based) were the least robust with respect to reconstruction algorithms using default settings and were the most sensitive to iteration number. Among all the features evaluated, “entropy_{Hist}”, “difference entropy”, “inverse difference normalized”, “inverse difference moment normalized”, “low gray-level run emphasis”, “high gray-level run emphasis”, and “low gray-level zone emphasis” proved to be the most robust.

Recently, repeatability of more than 100 radiomics features using different reconstruction settings, first using the point spread function and secondly complying with the European Association of Nuclear Medicine (EANM) guidelines for tumor PET imaging⁵⁴, and using different delineation methods, first on PET and then on CT images, was evaluated in 11 NSCLC patients. The best performance was seen using CT-based delineation (32%), followed by EANM-compliant reconstruction (17%), PET-based delineation (17%), and point spread function-based reconstruction (10%). The majority of PET features (98%) had a repeatability comparable to that reported for simple SUV measures (e.g., SUV_{max}) in the literature. Sixty-three features showed a very high ICC (≥ 0.90) independent of delineation or reconstruction. The performance of radiomics features depended more on the delineation method than on the applied reconstruction algorithm (changes in 25 and 3 features, respectively). CT-based delineation showed favorable repeatabilities and ICCs for most radiomics features, an exception being shape-based features, for which PET-based delineation performed better⁴¹.

Compared with static images, parametric images don't provide significant complementary information concerning standard parameters (SUV_{max}, SUV_{mean}, and metabolically active tumor volume - MATV) and heterogeneity quantification (histogram-based)⁵⁵. Differences in quantitative analysis using three-dimensional (3D) versus respiratory-gated (4D) acquisition have been reported. According to Oliver *et al.*⁴², the features with the least variability were “sphericity”, “spherical disproportion”, “entropy_{Hist}”, “entropy_{GLCM}”, “sum entropy”, “information measure of correlation 2”, “short run emphasis”, “long run emphasis”, and “run percentage”, while the features with the largest differences (>50%) were “kurtosis”, “low gray-level run emphasis”, “short run low gray-level emphasis”, and “long run low gray-level emphasis”.

Yip *et al.*⁵⁶ found significant differences in “maximal correlation coefficient”, “long run low gray-level emphasis”, “coarseness”, and “busyness” (NGTDM-based) between 3D and 4D PET imaging. When measuring tumor heterogeneity characteristics, reduced motion blurring by 4D PET acquisition was found to offer significantly better spatial resolution of textural features. 3D PET textures may lead to inaccurate prediction of treatment outcome, hindering optimal management of lung cancer patients. 4D PET textures may have a better prognostic value as they are less susceptible to tumor motion^{42, 56}. Different results have been reported by Cheng *et al.*⁴⁸, who compared the attenuation correction of PET images with helical CT (PET/HCT) and respiration-averaged CT (PET/ACT) in 56 NSCLC patients. PET/ACT yielded significantly higher SUV_{max}, SUV_{mean}, and TLG while significant differences between PET/HCT and PET/ACT were not observed with regard to other features, including “entropy_{Hist}”, “entropy_{GLCM}”, “dissimilarity”, “homogeneity_{GLCM}”, and “uniformity” (GLCM-based), “gray-level non-uniformity”, “zone-size non-uniformity”, and “high gray-level large zone emphasis” (GLSZM-based), and “coarseness”, “busyness”, “contrast_{NGTDM}”, and “complexity” (NGTDM-based).

Textural features have also been used to develop a computer-based algorithm which supported a vector machine; combined image parameters, derived from CT, PET, and PET/CT images, were found to improve diagnosis of mediastinal lymph node metastases by PET/CT⁵⁷.

Texture analysis and clinical applications. *Diagnosis.* Imaging texture analysis has been evaluated in order to determine which type and level of tissue heterogeneity can be captured and quantified through PET and to bridge the gap between *in vivo* and *ex vivo* tumor characterization⁵⁸. Histologic characteristics and PET features have been compared to identify whether texture analysis can help in differentiating between benign and malignant lesions or in classifying NSCLC subtypes (Table 2).

As mentioned above, Orlach *et al.*⁵³ compared the relative with the absolute resampling approach (RR and AR, respectively) and calculated the correlations of seven features with tissue types (tumor versus healthy tissue) and cancer subtypes (Adk versus Sq). RR-based “entropy_{GLCM}” didn't distinguish between tumor and healthy tissue ($p = 0.7621$) whereas the same index computed with the AR method was able to differentiate between these tissue types ($p < 0.0001$). Using the AR method, tumor tissue exhibited higher “high gray-level zone emphasis” than healthy tissue, while tumors had lower “homogeneity_{GLCM}” and “low gray-level zone emphasis”. Comparing textural indices in Adk versus Sq, all RR-based textural features were not significant ($p > 0.07$), in contrast to the AR-based textural features ($p \leq 0.05$). According to these results, features computed using an AR method vary as a function of the tissue type and cancer subtype and might be useful for tumor characterization.

Miwa *et al.*³⁷ evaluated whether morphological complexity (“morphological fractal dimension” derived from CT) and intratumoral heterogeneity (“density fractal dimension” derived from PET) assessed by fractal analysis improved the differential diagnosis between benign and malignant lung nodules in 54 patients with suspected NSCLC. Both fractal dimensions assessed by PET and CT were lower in malignant than in benign nodules ($p < 0.05$). SUV_{max} was higher in malignant than in benign nodules ($p < 0.05$). Tumor size significantly correlated with SUV_{max} ($p < 0.0001$), but not with either “morphological fractal dimension” ($p = 0.61$) or “density fractal dimension” ($p = 0.09$). The diagnostic accuracy of “density fractal dimension” tended to be higher than SUV_{max} (78% versus 68%, respectively) and was better than that for “morphological fractal dimension” (65%).

Heterogeneity has also been evaluated to determine whether it can help in differentiating between metastatic and inflammatory lymph nodes in lung Adk, as assessed by visual analysis, and other standard parameters of PET and CT (size and Hounsfield units). In this study, heterogeneity was assessed as “coefficient of variation” of lymph nodes in 44 patients (with a total of 94 biopsy-proven lymph nodes). Visual assessment for malignancy had high sensitivity (81%) but a relatively low specificity (67%), with an accuracy of 75%. The diagnostic performance of PET/CT using the cut-offs commonly employed for standard PET and CT parameters ($SUV_{max} = 2.5$, size of lymph nodes = 1 cm, and Hounsfield units = 120) was not satisfactory (accuracy of 56%, 60%, and 68%, respectively). Using an optimal cut-off determined by this study ($SUV_{max} = 5.96$, size of lymph nodes = 1.5 cm, and Hounsfield units = 136), the accuracy increased for SUV_{max} and size but not for Hounsfield units (81%, 84%, and 65%, respectively). Heterogeneity measured as “coefficient of variation” (using a cut-off = 0.2) yielded good sensitivity, specificity, and accuracy (88%, 76%, and 82%, respectively). The accuracy of “coefficient of variation” was slightly higher than that of SUV_{max} and size when using optimal cut-offs, but significantly higher than that of visual assessment and Hounsfield units. “Coefficient of variation”, SUV_{max} and size were significantly higher in metastatic lymph nodes than in benign ones ($p < 0.0001$), while the Hounsfield unit value was significantly lower in metastatic than in benign lymph nodes ($p = 0.0249$). Univariate analysis showed that all parameters except visual assessment were significant predictors, while using multivariate logistic regression only “coefficient of variation” and size proved statistically significant ($p = 0.032$ and 0.023 , respectively)⁵⁹.

Yip *et al.*⁵⁶, in order to evaluate whether texture features may be affected differently in Adk (21 lesions) versus Sqc (13 lesions) by motion, calculated the relative difference in each texture between 3D and 4D PET. The relative difference in each texture between 3D and 4D PET was not found to be significantly different between histologies ($p = 0.26$).

Ha *et al.*⁶⁰ analyzed differences in 24 textural features between Adk and Sqc (17 and 13 patients, respectively). The majority of texture parameters that showed a significant difference between Adk and Sqc were derived from GLCM (93%). SUV_{max} showed the most significant association with tumor pathology ($p = 0.001$). Upon autoclustering by linear discriminant analysis with those texture parameters that showed a significant difference between tumor subtypes ($n = 15$), the classification accuracy was found to be 83% (25/30 lesions were correctly clustered to their own tumor subtype). When analyzing with all parameters ($n = 24$), linear discriminant analysis clustered the lesions accurately according to their pathology, i.e., Adk versus Sqc, with a classification accuracy of 100% (linear separability of this autoclustering = 0.90).

Similarly, Kim *et al.*⁶¹ found that SUV_{max} , MTV, TLG, and heterogeneity (defined as the derivative of the volume-threshold function from 20 to 80%), were significantly higher in Sqc than in Adk.

van Gómez López *et al.*⁶² evaluated the correlation between conventional metabolic parameters (SUV_{max} , SUV_{mean} , MTV, and TLG) and heterogeneity (“energy_{GLCM}”, “contrast_{GLCM}”, “correlation”, “entropy_{GLCM}”, and “homogeneity_{GLCM}”), histology, tumor size, and AJCC stage in 38 NSCLC patients (24 Sqc and 14 Adk). There was a positive relationship for all metabolic parameters with “entropy_{GLCM}”, “correlation”, and “homogeneity_{GLCM}” and a negative relationship with “energy_{GLCM}” and “contrast_{GLCM}”. No statistically significant differences were found between the mean values of tumor size, AJCC stage, and standard metabolic parameters in Adk versus Sqc tumors. Concerning textural features, “energy_{GLCM}” was lower in Adk than in Sqc ($p = 0.027$) while “homogeneity_{GLCM}” was higher in Adk than in Sqc ($p = 0.047$). Tumor size was correlated with “energy_{GLCM}”, “contrast_{GLCM}”, “correlation”, “entropy_{GLCM}”, “MTV”, and “TLG” ($p < 0.01$). A statistical correlation between the pT and “energy_{GLCM}”, “contrast_{GLCM}”, “entropy_{GLCM}”, and “MTV” ($p \leq 0.05$) was found, but not between remaining AJCC subgroups and the other textural or metabolic parameters.

Prognosis and treatment response prediction. The Warburg effect, first described over 80 years ago, postulates that tumors undergo glycolysis preferentially despite adequate intracellular oxygen tension^{63,64}. While Warburg believed this to be a consequence of mitochondrial dysfunction, tumor glycolysis can proceed with functional cellular mitochondria and may be an adaptive response for tumor survival^{65–67}. Furthermore, studies have recently linked glycolysis in cancer to more widespread deregulation of cell bioenergetics^{68–70}, suggesting that FDG uptake may be a surrogate for more than glycolysis alone and perhaps a lens through which one can view global tumor bioenergetics⁷¹. Therefore, texture features have been introduced as imaging biomarkers on the assumption that they are an index of the degree of tumor heterogeneity, and that biologic tumor heterogeneity is associated with poor prognosis in cancer patients and can contribute to treatment failure and drug resistance⁷². The prognostic value of texture analysis has been evaluated in different NSCLC clinical settings (Table 2).

Nair *et al.*⁷¹ evaluated a possible association between textural features, gene expression signatures, and survival in a computational study (172 NSCLC patients). Fourteen PET features were extracted within the study cohort ($n = 25$). Individual genes associated with PET features in the study cohort were directly analyzed in the external cohort ($n = 63$) for their association with clinical outcomes. Lastly, PET features associated with prognostic gene signatures from the external cohort were tested in a validation cohort ($n = 84$). Four genes (LY6E, RNF149, MCM6, FAP) associated with textural features were found also to be associated with survival. Histogram-based and “shape and size” features together provided a more accurate prognostic model than each feature alone,

| Diagnostic role |
|--|
| Compared with non-malignant lesions, malignant lung nodules are characterized by higher SUV_{max} and lower morphological and density fractal dimensions ³⁷ . |
| Metastatic lymph nodes are characterized by higher heterogeneity (coefficient of variation) than inflammatory ones ⁵⁹ . |
| Large lesions are characterized by high heterogeneity (i.e., visual score, $entropy_{GLCM}$, coefficient of variation) ^{44, 45, 74} . |
| Each subtype of NSCLC tumor has different metabolic heterogeneity characteristics. Compared with Adk, Sqc is characterized by higher SUV_{max} , AUC-IVH, $energy_{GLCM}$, $entropy_{GLCM}$, sum entropy, difference entropy, and inverse different moment and by lower homogeneity $_{GLCM}$, sum of squares, angular second moment, ratio of non-zero $_{Gr}$ and difference variance ^{60–62} . |
| Prognostic and predictive role |
| Heterogeneity (i.e., AUC-CSH) can predict recurrence in pN0 Adk patients who have undergone curative surgery but not in Sqc patients (high heterogeneity is associated with a shorter DFS) ⁶¹ . |
| Best prognostic value for overall survival is found for relative portions of the tumor above higher uptakes defined as $SUV_{max} > 80\%$ (i.e., V_{80}) in patients who received radiation therapy (sequential chemoradiation, concurrent chemoradiation, or only radiation). The higher the portion above higher uptake (V_{80}), the better the prognosis ^{29, 80} . |
| Heterogeneity (i.e., low AUC-CSH) identifies patients with inoperable stage III NSCLC with poor PFS ⁷⁵ . |
| High SUV_{max} , large MTV, and high heterogeneity (i.e., high $entropy_{GLCM}$, high asphericity, homogeneity $_{GLCM}$ and high dissimilarity, size-zone variability, and low zone percentage) are associated with poorer OS and RFS in stage I–III NSCLC ^{35, 73, 74, 83} . |
| Tumor heterogeneity (i.e., $entropy_{GLCM}$) is associated with response to radiation therapy in NSCLC (DSS is lower for patients with high $entropy_{GLCM}$) ⁷⁹ . |
| Lesions in responders (complete or partial response) to chemoradiotherapy are characterized by lower coarseness, contrast $_{NGTDM}$, and busyness than non-responders (stable or progressive disease). High coarseness values are associated with an increased risk of progression (increased risk of death), whereas high contrast $_{NGTDM}$ and busyness values are associated with a lower risk of progression (PFS and LPFS) ¹⁴ . |
| Large primary tumors with low SumAverage (i.e., more heterogeneous) have a poor prognosis following chemoradiotherapy ⁷⁶ . |
| Lesions in responders to erlotinib are characterized by lower heterogeneity than those in non-responders. Specifically, lower heterogeneity after 6 weeks of treatment, as measured by contrast $_{NGTDM}$, is independently associated with longer survival, and a larger reduction in heterogeneity between baseline and 6 weeks of treatment, as measured by $entropy_{Hist}$, is independently associated with longer survival and with treatment response ⁷⁸ . |
| Tumor heterogeneity (i.e., dissimilarity) appears to be a strong independent outcome predictor (DSS and DFS) after radiotherapy. Low dissimilarity is associated with a higher risk of recurrence ³⁶ . |
| The optimal prognostic model for identification of groups of NSCLC patients at risk for developing distant metastasis includes $SUV_{peak2ml}$ and Gauss cluster shade $_{laws}$. High $SUV_{peak2ml}$ and Gauss cluster shade $_{laws}$ are associated with an increased risk of distant metastases ⁸⁴ . |
| Solidity (which quantifies the dispersion of primary and nodal disease in a local region, with high values corresponding to disease that is compact and in close proximity, and low values corresponding to disease that is dispersed) and primary tumor $energy_{GLCM}$ (higher level for tumors that are more heterogeneous) improve risk stratification compared with a model with conventional prognostic factors alone in stage III NSCLC. Solidity and primary tumor $energy_{GLCM}$ are capable of isolating subgroups of patients who will receive a benefit or detriment from dose escalation (i.e., as disease solidity and primary co-occurrence matrix $energy$ increase, patients receiving higher dose radiation therapy have improved OS and PFS compared with those receiving lower doses) ^{81, 82} . |

Table 3. Summary of clinically relevant results in investigations assessing the diagnostic, prognostic and predictive role of FDG-PET/CT texture analysis. Adk: adenocarcinoma type; AUC-IVH: area under the curve within the intensity volume histogram; DFS: disease-free survival; DSS: disease-specific survival; GLCM: gray-level co-occurrence matrix; GLRLM: gray-level run-length matrix; GLSZM: gray-level size-zone matrix; Gr: absolute gradient; LPFS: local progression-free survival; MTV: metabolic tumor volume; NGTDM: neighborhood gray-tone difference matrix; NSCLC: non-small cell lung cancer; OS: overall survival; PFS: progression-free survival; Sqc: squamocellular types; SUV: standardized uptake value.

suggesting that leveraging tumor genomics with an expanded collection of PET features may enhance understanding of the value of FDG uptake as an imaging biomarker beyond its association with glycolysis.

Win *et al.*¹³ compared the prognostic value of texture analysis with tumor staging and other imaging prognostic factors (i.e., metabolism assessed by PET/CT and permeability assessed by dynamic contrast-enhanced CT) in 122 NSCLC patients treated with curative or palliative approach. Tumor heterogeneity (“ $entropy_{Hist}$ ”) was calculated from both attenuation-corrected CT images and SUV images without image filtration. “ $Entropy_{Hist}$ ” (derived from both CT and PET images), permeability, and stage were found to be survival predictors at univariate analysis ($p \leq 0.003$), in contrast to SUV_{max} ($p = 0.948$). At multivariate analysis, “ $entropy_{Hist}$ ” derived from CT ($p = 0.021$), stage ($p = 0.001$), and permeability ($p < 0.001$) were identified as independent survival predictors, irrespective of the treatment objective (curative or palliative). In the study by Cheng *et al.*, “ $Entropy_{Hist}$ ”, “ $entropy_{GLCM}$ ”, and “coarseness”, derived from both PET/HCT and PET/ACT, were able to predict disease-specific survival at univariate ($p \leq 0.01$) and multivariate analysis ($p < 0.05$) in stage I–III NSCLC patients⁴⁸.

Similarly, the “shape and size” features “asphericity” ($p < 0.001$) and “solidity” ($p = 0.05$), as well as “primary surgical treatment” ($p = 0.05$), were found to be significant independent predictors of progression-free survival in 60 NSCLC patients treated with different approaches. Concerning overall survival, only “asphericity” and “primary surgical treatment” ($p = 0.02$ and $= 0.01$, respectively) proved to be independent predictors, and none of the other PET parameters (including SUV_{max} , TLG, and MTV) showed a significant predictive value in this series of patients⁷³.

The GLCM- and GLSZM-derived features “ $entropy_{GLCM}$ ”, “ $homogeneity_{GLCM}$ ”, “dissimilarity”, “size-zone variability”, and “zone percentage”, but not “high intensity emphasis” (GLSZM-based), have also been reported to be independent prognostic factors with respect to stage (although not independently of each other) in patients treated with different approaches. Nonetheless, the addition of risk factors allowed a better differentiation of patient outcome. High SUV, large metabolic volumes, and high heterogeneity were associated with a poorer overall survival and recurrence-free survival³⁵, suggesting that heterogeneity quantification and volume (i.e., MTV)

| All the following technical aspects should be provided for PET texture features calculation | |
|---|---|
| a) | scanner |
| b) | method of images acquisition (e.g. respiratory motion, dynamic) |
| c) | parameters used to acquire images |
| d) | parameters used to reconstruct images |
| e) | type of images used to extract features (i.e., PET or both PET and CT) |
| f) | “target” of texture analysis (e.g., primary tumor, lymph nodes, metastases) |
| g) | application of PVC and/or a minimum lesion size/volume |
| h) | method of segmentation (e.g., threshold uptake 40% of the SUV_{max}) |
| i) | discretization method (e.g., fixed number of bins) |
| j) | software |
| k) | features and matrix computation method* |
| An appropriate statistical analysis should be used | |
| Datasets of 10–15 patients per feature have been recommended to test the prognostic power of texture features | |
| Textural features selection and validation | |
| The use of the radiomics features insensitive to acquisition modes and reconstruction parameters is recommended. A correlation of conventional metrics (SUV, MTV, etc.) and texture features should be assessed to evaluate the potential complementary value of the measures. Independent validation datasets are needed to confirm the results. | |

Table 4. Summary of relevant methodological issues in calculating and reporting FDG-PET/CT texture analysis. *A proposal for a consistent terminology is reported within the Supplementary material.

may provide valuable complementary information with respect to prognosis, although the complementary information increases substantially with larger volumes⁷⁴.

Heterogeneity of primary tumor (evaluated by the area under the curve of cumulative SUV histograms: AUC-CSH) was observed to be an independent predictor of recurrence in pathologically N0 Adk but not in Sqc ($p = 0.03$ and 0.13 , respectively) after curative surgical resection⁶¹ as well as a predictor of disease progression after concurrent chemoradiotherapy in patients with inoperable stage III NSCLC⁷⁵.

Interestingly, “Sum Average” (GLCM-based) was strongly associated with overall survival in a multi-institutional dataset of locally advanced NSCLC patients with large tumors who were treated with definitive chemoradiotherapy, suggesting its robustness as a prognostic factor⁷⁶.

NGTDM-derived features (“coarseness”, “contrast_{NGTDM}” and “busyness”) have also been reported to be associated with response to chemoradiotherapy and prognosis⁷⁷ in locally advanced NSCLC. The same group, testing FOS and high-order features as predictors of response or survival in patients treated with erlotinib, found that response to erlotinib was associated with reduced heterogeneity and that the percentage of changes in “entropy_{Y_{Hist}}” (between baseline and 6-week PET/CT) was independently associated with overall survival and treatment response⁷⁸.

Similarly, PET features have been reported to be able to predict outcome and/or treatment response in NSCLC patients treated with definitive radiotherapy^{29, 36, 79, 80}. In this specific clinical setting, “entropy_{GLCM}” has been reported to be an independent predictor of disease-specific survival ($p = 0.016$)⁷⁹, while “dissimilarity” has been found to be associated with both disease-specific survival ($p = 0.037$) and disease-free survival ($p < 0.01$)³⁶.

Initial attempts have been made to determine whether quantitative imaging features from pretreatment PET can enhance overall survival risk stratification beyond what can be achieved with conventional prognostic factors in NSCLC. In patients with stage III NSCLC, linear predictors of overall survival generated with both quantitative imaging features (histogram-derived, GLCM-derived, and “shape and size” features) and conventional prognostic factors (age, sex, histologic findings and stage, Karnofsky performance status, smoking status and estimated pack-years, treatment type) have demonstrated improved risk stratification compared with those generated with conventional prognostic factors alone in terms of log-rank statistics ($p = 0.18$ versus $= 0.0001$, respectively)⁸⁰. The use of quantitative imaging features selected during cross-validation improved the model using conventional prognostic factors alone ($p = 0.007$). Disease “solidity” and primary tumor “energy_{GLCM}” were found to be selected in all folds of cross-validation⁸¹. Additionally, these features were found to be capable of isolating subgroups of patients who received a benefit or detriment from dose escalation⁸².

Similarly, PET “entropy_{GLCM}” and CT “zone percentage” have been found to have the highest complementary values with clinical stage and functional volume in stage I–III NSCLC⁸³. Desseroit *et al.*⁸³ provided a nomogram able to improve stratification amongst patients with stage II and III disease, allowing identification of those with the poorest prognosis (clinical stage III, large MTV, high PET heterogeneity, and low CT heterogeneity).

In early-stage NSCLC the optimal prognostic model for prediction of distant metastases in patients treated with stereotactic ablative radiation therapy included two image features that allowed quantification of intratumor heterogeneity and SUV_{peak} . A significant improvement ($p = 0.0001$) in predicting freedom from distant metastasis was seen when histologic information was added compared with a prognostic model based solely on imaging features⁸⁴.

Discussion

Texture features are of growing interest for tumor characterization in imaging. Nevertheless, on the basis of results published to date on FDG PET, it is unclear which indices should be used, what they represent, and how they are related to conventional parameters such as SUVs, MTV, and TLG⁴⁶. We summarize the results of the available studies within Table 3. PET features differed significantly in malignant and non-malignant tissues (considering either primary lung tumors or lymph nodes)^{37, 46, 59} and also in Adk and Sqc^{60–62}. However, literature data are really heterogeneous in this setting and, despite promising results, it isn't possible to suggest for use a reproducible feature or a combination of features able to characterize definitely malignant tissues or lung cancer subtypes.

Again, different metrics, matrices, and methods (e.g., tumor segmentation, survival endpoints) have been reported in the evaluation of NSCLC prognosis based on PET features. Commonly, the term “heterogeneous” is used with different meanings. Concerning texture analysis, “heterogeneity” may result from one or more PET features positively or negatively related to treatment and/or outcome. Therefore clinical texture papers often use the term “heterogeneity” to summarize specific tumor characteristics expressed by PET features. Among these, “entropy_{Hist}” and “entropy_{GLCM}” are most frequently reported to have an independent prognostic role, able to predict treatment outcome and/or survival in NSCLC patients^{13, 29, 35, 48, 78, 79, 83}.

It is important for textural feature values to be directly comparable, both between and within patients, in order to derive meaningful conclusions from radiomic analysis and allow their use in clinical routine. However, the lack of a standardized method to calculate textural features prevents comparison between literature data and meta-analysis. Additionally, crucial information about texture extraction is not always available within the published articles. In order to try to simplify the complexity of texture analysis and to facilitate comparison among different series, some mandatory information concerning specific methodological aspects should be reported. Specifically, in the drafting of a paper on PET texture analysis, the following should be indicated in the Methods section (Table 4). Specifically, in the drafting of a paper on PET texture analysis, the following should be indicated in the Methods section: (a) the scanner, (b) the method (e.g., respiratory motion, dynamic) and parameters used to (c) acquire and (d) reconstruct images, (e) the type of images used to extract features (i.e., PET or both PET and CT), (f) the “target” of texture analysis (e.g., primary tumor, lymph nodes), (g) the application of PVC and/or a minimum lesion size/volume, (h) the method of segmentation (e.g., threshold uptake 40% of the SUV_{max}), (i) the discretization method (e.g., fixed number of bins), (j) the software, and (k) the feature(s) and matrix computation method. As the published results are very preliminary, a preferred method for reconstruction, discretization, or segmentation cannot be recommended. Obviously, consistent terminology in respect of features (a proposal is reported within the Supplementary material) and an appropriate statistical analysis are mandatory. Datasets of 10–15 patients per feature have been recommended to test the prognostic power of texture features. Moreover, use of the radiomics features insensitive to acquisition modes and reconstruction parameters is recommended. The correlation of conventional metrics (SUV, MTV, etc.) and texture features should be assessed to evaluate the potential complementary value of the measures. In addition, independent validation datasets are needed to confirm the results. Finally, there is a need for easy-to-use software tools for feature extraction (their main characteristics have been very recently summarized by Hatt *et al.*⁸⁵) since those available are handled only by non-clinician experts, are time consuming, and are able to produce a lot of textures from different matrices, many of which are probably unnecessary since their biologic significance is encapsulated within others.

In conclusion, standardization is mandatory to prove the value of the information that can be derived from medical images, enabling non-invasive *in vivo* characterization of lung lesions and accurate risk stratification for the purpose of decision making regarding treatment strategy.

Ethical approval. This article does not contain any study with human participants or animals. The patient imaged in the Fig. 3, signed an inform consent to use his personal data including imaging, also for publication; however the figure is completely anonymized, preventing the possibility of discovering the identity of the individual.

References

1. Tixier, F. *et al.* Reproducibility of Tumor Uptake Heterogeneity Characterization Through Textural Feature Analysis in 18F-FDG PET. *J. Nucl. Med.* **53**, 693–700 (2012).
2. Sauter, A. W., Schwenzler, N., Divine, M. R., Pichler, B. J. & Pfannenberger, C. Image-derived biomarkers and multimodal imaging strategies for lung cancer management. *Eur. J. Nucl. Med. Mol. Imaging* **42**, 634–643 (2015).
3. Meignan, M., Itti, E., Gallamini, A. & Younes, A. FDG PET/CT imaging as a biomarker in lymphoma. *Eur. J. Nucl. Med. Mol. Imaging* **42**, 623–633 (2015).
4. Differding, S., Hanin, F.-X. & Grégoire, V. PET imaging biomarkers in head and neck cancer. *Eur. J. Nucl. Med. Mol. Imaging* **42**, 613–622 (2015).
5. Picchio, M. *et al.* Predictive value of pre-therapy (18)F-FDG PET/CT for the outcome of (18)F-FDG PET-guided radiotherapy in patients with head and neck cancer. *Eur. J. Nucl. Med. Mol. Imaging* **41**, 21–31 (2014).
6. Guo, W. *et al.* Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. *J. Med. Imaging (Bellingham, Wash.)* **2**, 41007 (2015).
7. Wang, J. *et al.* Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study. *PLoS One* **10**, e0143308 (2015).
8. Yip, S. S. F. *et al.* Relationship between the Temporal Changes in Positron-Emission-Tomography-Imaging-Based Textural Features and Pathologic Response and Survival in Esophageal Cancer Patients. *Front. Oncol.* **6**, 72 (2016).
9. Hyun, S. H. *et al.* Intratumoral heterogeneity of (18)F-FDG uptake predicts survival in patients with pancreatic ductal adenocarcinoma. *Eur. J. Nucl. Med. Mol. Imaging*. doi:10.1007/s00259-016-3316-6 (2016).
10. Rahim, M. K. *et al.* Recent Trends in PET Image Interpretations Using Volumetric and Texture-based Quantification Methods in Nuclear Oncology. *Nucl. Med. Mol. Imaging (2010)*. **48**, 1–15 (2014).
11. O'Connor, J. P. B. *et al.* Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin. Cancer Res.* **21**, 249–257 (2015).

12. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**, 4006 (2014).
13. Win, T. *et al.* Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer. *Clin. Cancer Res.* **19**, 3591–3599 (2013).
14. Chicklore, S. *et al.* Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur. J. Nucl. Med. Mol. Imaging* **40**, 133–140 (2013).
15. Buvat, I., Orlhac, F. & Soussan, M. Tumor Texture Analysis in PET: Where Do We Stand? *J. Nucl. Med.* **56**, 1642–1644 (2015).
16. Zorzela, L. *et al.* PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ* **352**, i157 (2016).
17. Al-Kadi, O. S. & Watson, D. Texture analysis of aggressive and nonaggressive lung tumor CE CT images. *IEEE Trans. Biomed. Eng.* **55**, 1822–1830 (2008).
18. Ganeshan, B., Miles, K. A., Young, R. C. D. & Chatwin, C. R. Hepatic entropy and uniformity: additional parameters that can potentially increase the effectiveness of contrast enhancement during abdominal CT. *Clin. Radiol.* **62**, 761–768 (2007).
19. Brown, R. a. & Frayne, R. A comparison of texture quantification techniques based on the Fourier and S transforms. *Med. Phys.* **35**, 4998–5008 (2008).
20. Craciunescu, O. I., Das, S. K. & Clegg, S. T. Dynamic contrast-enhanced MRI and fractal characteristics of percolation clusters in two-dimensional tumor blood perfusion. *J. Biomech. Eng.* **121**, 480–486 (1999).
21. Goh, V., Sanghera, B., Wellsted, D. M., Sundin, J. & Halligan, S. Assessment of the spatial pattern of colorectal tumour perfusion estimated at perfusion CT using two-dimensional fractal analysis. *Eur. Radiol.* **19**, 1358–1365 (2009).
22. Dettori, L. & Semler, L. A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Comput. Biol. Med.* **37**, 486–498 (2007).
23. Sanghera, B. *et al.* Reproducibility of 2D and 3D Fractal Analysis Techniques for the Assessment of Spatial Heterogeneity of Regional Blood Flow in Rectal Cancer. <http://dx.doi.org/10.1148/radiol.12111316> (2012).
24. Al-Kadi, O. S. Assessment of texture measures susceptibility to noise in conventional and contrast enhanced computed tomography lung tumour images. *Comput. Med. Imaging Graph.* **34**, 494–503 (2010).
25. Tuceryan, M., Tuceryan, M., Jain, A. K. & Jain, A. K. The Handbook of Pattern Recognition and Computer Vision (2nd Edition), Texture Analysis. *Pattern Recognit.* 207–248, doi:10.1097/RCT.0b013e3181ec05e4 (1998).
26. Yip, S. S. F. & Aerts, H. J. W. L. Applications and limitations of radiomics. *Phys. Med. Biol.* **61**, R150–R166 (2016).
27. El Naqa, I. *et al.* Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* **42**, 1162–1171 (2009).
28. Drzymala, R. E. *et al.* Dose-volume histograms. *Int. J. Radiat. Oncol. Biol. Phys.* **21**, 71–78 (1991).
29. Vaidya, M. *et al.* Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother. Oncol.* **102**, 239–245 (2012).
30. Nestle, U. *et al.* Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J. Nucl. Med.* **46**, 1342–1348 (2005).
31. Van Velden, F. H. P. *et al.* Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur. J. Nucl. Med. Mol. Imaging* **38**, 1636–1647 (2011).
32. Boellaard, R. *et al.* A novel cumulative SUV- volume histogram method for parameterizing heterogeneous tumour tracer uptake in oncology FDG PET studies. *Eur. J. Nucl. Med. Mol. Imaging* **37**, S261 (2010).
33. Haralick, R., Shanmugam, K. & Dinstein, I. Texture Features for Image Classification. *IEEE Trans Sys Man Cyb SMC.* **3**, 610–621 (1973).
34. Yan, R. *et al.* Detection of Myocardial Metabolic Abnormalities by 18F-FDG PET/CT and Corresponding Pathological Changes in Beagles with Local Heart Irradiation. *Korean J. Radiol.* **16**, 919–928 (2015).
35. Tixier, F. *et al.* Visual Versus Quantitative Assessment of Intratumor 18F-FDG PET Uptake Heterogeneity: Prognostic Value in Non-Small Cell Lung Cancer. *J. Nucl. Med.* **55**, 1235–1241 (2014).
36. Lovinfosse, P. *et al.* FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. *Eur. J. Nucl. Med. Mol. Imaging*, doi:10.1007/s00259-016-3314-8 (2016).
37. Miwa, K. *et al.* FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur. J. Radiol.* **83**, 715–719 (2014).
38. Grigsby, P. W., Siegel, B. A., Dehdashti, F., Rader, J. & Zoberi, I. Posttherapy [18F] fluorodeoxyglucose positron emission tomography in carcinoma of the cervix: response and outcome. *J. Clin. Oncol.* **22**, 2167–2171 (2004).
39. Greven, K. M. Positron-emission tomography for head and neck cancer. *Semin. Radiat. Oncol.* **14**, 121–129 (2004).
40. Leijenaar, R. T. H. *et al.* Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol.* **52**, 1391–1397 (2013).
41. Velden, F. H. P. V. *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.*, doi:10.1007/s11307-016-0940-2 (2016).
42. Oliver, J. A. *et al.* Variability of Image Features Computed from Conventional and Respiratory-Gated PET/CT Images of Lung Cancer. *Transl. Oncol.* **8**, 524–534 (2015).
43. Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D. & Müller, H. Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Med. Image Anal.* **18**, 176–196 (2014).
44. Hatt, M. *et al.* Impact of Tumor Size and Tracer Uptake Heterogeneity in 18F-FDG PET and CT Non-Small Cell Lung Cancer Tumor Delineation. *J. Nucl. Med.* **52**, 1690–1697 (2011).
45. Dong, X. *et al.* Intra-tumour (18) F-FDG uptake heterogeneity decreases the reliability on target volume definition with positron emission tomography/computed tomography imaging. *J. Med. Imaging Radiat. Oncol.* **4**, 338–345 (2015).
46. Orlhac, F. *et al.* Tumor Texture Analysis in 18F-FDG PET: Relationships Between Texture Parameters, Histogram Indices, Standardized Uptake Values, Metabolic Volumes, and Total Lesion Glycolysis. *J. Nucl. Med.* **55**, 414–422 (2014).
47. Yan, J. *et al.* Impact of Image Reconstruction Settings on Texture Features in 18 F-FDG PET. *J Nucl med* **56**, 1667–1674 (2015).
48. Cheng, N. M., Dean Fang, Y. H., Tsan, D. L., Hsu, C. H. & Yen, T. C. Respiration-averaged CT for attenuation correction of PET Images - Impact on PET Texture Features in Non-Small Cell Lung Cancer Patients. *PLoS One* **11**, 1–15 (2016).
49. Hofheinz, F. *et al.* An automatic method for accurate volume delineation of heterogeneous tumors in PET. *Med. Phys.* **40**, 82503 (2013).
50. Cui, H., Wang, X. & Feng, D. Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, 5384–5387 doi:10.1109/EMBC.2012.6347211 (2012).
51. Cui, H., Wang, X., Zhou, J. & Eberl, S. Topology polymorphism graph for lung tumor segmentation in PET-CT images. *Phys. Med. Biol.* **4893**, 4893 (2015).
52. Leijenaar, R. T. H. *et al.* The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci. Rep.* **5**, 11075 (2015).
53. Orlhac, F., Soussan, M., Chouahnia, K., Martinod, E. & Buvat, I. 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. *PLoS One* **10**, 1–16 (2015).
54. Boellaard, R. *et al.* FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur. J. Nucl. Med. Mol. Imaging* **42**, 328–354 (2014).
55. Tixier, F. *et al.* Comparison of tumor uptake heterogeneity characterization between static and parametric 18F-FDG PET images in Non-Small Cell Lung Cancer. *J. Nucl. Med.* **31** (2016).

56. Yip, S. *et al.* Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. *PLoS One* **9**, e115510 (2014).
57. Gao, X. *et al.* The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur. J. Radiol.* **84**, 312–317 (2015).
58. Buvat, I., Orhac, F. & Soussan, M. Tumor Texture Analysis in PET: Where Do We Stand? *J. Nucl. Med.* **56**, 1642–1644 (2015).
59. Budiawan, H. *et al.* Heterogeneity Analysis of 18F-FDG Uptake in Differentiating Between Metastatic and Inflammatory Lymph Nodes in Adenocarcinoma of the Lung: Comparison with Other Parameters and its Application in a Clinical Setting. *Nucl. Med. Mol. Imaging* (2010). **47**, 232–241 (2013).
60. Ha, S. *et al.* Autoclustering of Non-small Cell Lung Carcinoma Subtypes on 18F-FDG PET Using Texture Analysis: A Preliminary Result. *Nucl. Med. Mol. Imaging* (2010). **48**, 278–286 (2014).
61. Kim, D.-H. *et al.* Prognostic Significance of Intratumoral Metabolic Heterogeneity on 18F-FDG PET/CT in Pathological N0 Non-Small Cell Lung Cancer. *Clin. Nucl. Med.* **40**, 708–714 (2015).
62. van Gómez López, O. *et al.* Heterogeneity in [18 F] fluorodeoxyglucose positron emission tomography/computed tomography of non – small cell lung carcinoma and its relationship to metabolic parameters and pathologic staging. *Mol. Imaging* **13**, 1–12 (2014).
63. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
64. Warburg, O. On respiratory impairment in cancer cells. *Science* **124**, 269–270 (1956).
65. Mathupala, S. P., Ko, Y. H. & Pedersen, P. L. Hexokinase-2 bound to mitochondria: cancer's stygian link to the "Warburg Effect" and a pivotal target for effective therapy. *Semin. Cancer Biol.* **19**, 17–24 (2009).
66. Frezza, C. & Gottlieb, E. Mitochondria in cancer: Not just innocent bystanders. *Semin. Cancer Biol.* **19**, 4–11 (2009).
67. Robey, R. B. & Hay, N. Is Akt the "Warburg kinase" and "Akt-energy metabolism interactions and oncogenesis. *Semin. Cancer Biol.* **19**, 25–31 (2009).
68. Dang, C. V. Glutaminolysis: supplying carbon or nitrogen or both for cancer cells? *Cell Cycle* **9**, 3884–3886 (2010).
69. Dang, C. V. Rethinking the Warburg effect with Myc micromanaging glutamine metabolism. *Cancer Res.* **70**, 859–862 (2010).
70. Dang, C. V. Enigmatic MYC Conducts an Unfolding Systems Biology Symphony. *Genes Cancer* **1**, 526–531 (2010).
71. Nair, V. S. *et al.* Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res.* **72**, 3725–3734 (2012).
72. Chalkidou, A., O'Doherty, M. J. & Marsden, P. K. False discovery rates in PET and CT studies with texture features: A systematic review. *PLoS One* **10**, 1–18 (2015).
73. Apostolova, I. *et al.* Quantitative assessment of the asphericity of pretherapeutic FDG uptake as an independent predictor of outcome in NSCLC. *BMC Cancer* **14**, 896 (2014).
74. Hatt, M. *et al.* 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J. Nucl. Med.* **56**, 38–44 (2015).
75. Kang, S.-R. *et al.* Intratumoral Metabolic Heterogeneity for Prediction of Disease Progression After Concurrent Chemoradiotherapy in Patients with Inoperable Stage III Non-Small-Cell Lung Cancer. *Nucl. Med. Mol. Imaging* (2010). **48**, 16–25 (2014).
76. Ohri, N. *et al.* Pretreatment 18FDG-PET Textural Features in Locally Advanced Non-Small Cell Lung Cancer: Secondary Analysis of ACRIN 6668/RTOG 0235. *J. Nucl. Med.* 1–30, doi:10.2967/jnumed.115.166934 (2016).
77. Cook, G. J. R. *et al.* Are Pretreatment 18F-FDG PET Tumor Textural Features in Non-Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy? *J. Nucl. Med.* **54**, 19–26 (2013).
78. Cook, G. J. R. *et al.* Non-Small Cell Lung Cancer Treated with Erlotinib: Heterogeneity of (18)F-FDG Uptake at PET-Association with Treatment Response and Prognosis. *Radiology* **276**, 883–893 (2015).
79. Pyka, T. *et al.* Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat. Oncol.* **10**, 100 (2015).
80. Carvalho, S. *et al.* Prognostic value of metabolic metrics extracted from baseline PET images in NSCLC in non small cell lung cancer. *Acta Oncol* **52**, 1398–1404 (2013).
81. Fried, D. V. *et al.* Stage III Non-Small Cell Lung Cancer: Prognostic Value of FDG PET Quantitative Imaging Features Combined with Clinical Prognostic Factors. *Radiology* **278**, 214–222 (2016).
82. Fried, D. V. *et al.* Potential Use of 18F-fluorodeoxyglucose Positron Emission Tomography-Based Quantitative Imaging Features for Guiding Dose Escalation in Stage III Non-Small Cell Lung Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **94**, 368–376 (2016).
83. Desseroit, M.-C. *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*, doi:10.1007/s00259-016-3325-5 (2016).
84. Wu, J. & Rubin, D. L. Early-Stage Non – Small Cell Lung Cancer : Quantitative Imaging Characteristics of 18 F Fluorodeoxyglucose PET/CT Allow. *Radiology* (2016).
85. Hatt, M. *et al.* Characterization of PET/CT images using texture analysis: the past, the present ... any future? *Eur. J. Nucl. Med. Mol. Imaging*, doi:10.1007/s00259-016-3427-0 (2016).

Author Contributions

The manuscript has been seen and approved by all authors, whose individual contributions were as follows: K.M. and S.M. conceptualized the paper, K.M. and L.A. contributed to data selection; K.M. and S.M., wrote the paper; K.M. and S.M. provided figures and tables; L.C. revised and checked the nomenclature and made uniform the formulas of the supplemental material; C.A. supervised the project and paper writing.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-00426-y

Competing Interests: The authors declare that they have no competing interests.

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017