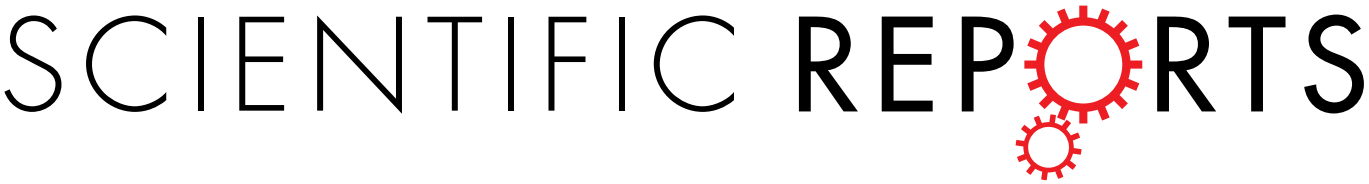
The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis



OPEN

The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis

Received: 04 November 2014
Accepted: 13 May 2015
Published: 05 August 2015

Ralph T.H. Leijenaar[1], Georgi Nalbantov[1], Sara Carvalho[1], Wouter J.C. van Elmpt[1], Esther G.C. Troost[1], Ronald Boellaard[2], Hugo J.W.L Aerts[1,3], Robert J. Gillies[4] & Philippe Lambin[1]

FDG-PET-derived textural features describing intra-tumor heterogeneity are increasingly investigated as imaging biomarkers. As part of the process of quantifying heterogeneity, image intensities (SUVs) are typically resampled into a reduced number of discrete bins. We focused on the implications of the manner in which this discretization is implemented. Two methods were evaluated: (1) R_D, dividing the SUV range into D equally spaced bins, where the intensity resolution (i.e. bin size) varies per image; and (2) R_B, maintaining a constant intensity resolution B. Clinical feasibility was assessed on 35 lung cancer patients, imaged before and in the second week of radiotherapy. Forty-four textural features were determined for different D and B for both imaging time points. Feature values depended on the intensity resolution and out of both assessed methods, R_B was shown to allow for a meaningful inter- and intra-patient comparison of feature values. Overall, patients ranked differently according to feature values–which was used as a surrogate for textural feature interpretation–between both discretization methods. Our study shows that the manner of SUV discretization has a crucial effect on the resulting textural features and the interpretation thereof, emphasizing the importance of standardized methodology in tumor texture analysis.

In recent years, oncological research has increasingly focused on the prediction of treatment outcome based on individual patient and tumor characteristics[1], aiming to avoid the one-size-fits-all treatment approach that under- and over-treats a large number of patients. Imaging can play a crucial role here, as it allows for a non-invasive identification and characterization of the tumor[2,3]. Positron emission tomography (PET) is a valuable tool for detecting and staging cancer[4]. In recent years, PET imaging has also been increasingly used for decision support[5], treatment planning[6,7] and response monitoring during radiotherapy[8]. The most widely used PET tracer is [18F] fluoro-2-deoxy-D-glucose (FDG), commonly quantified by standardized uptake values (SUVs)[9]. Easily derived SUV measurements, such as the maximum

[1]Department of Radiation Oncology (MAASTRO), GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre (MUMC+), Maastricht, the Netherlands. [2]Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands. [3]Departments of Radiation Oncology and Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [4]Department of Cancer Imaging and Metabolism, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. Correspondence and requests for materials should be addressed to R.T.H.L. (email: ralph.leijenaar@maastro.nl)

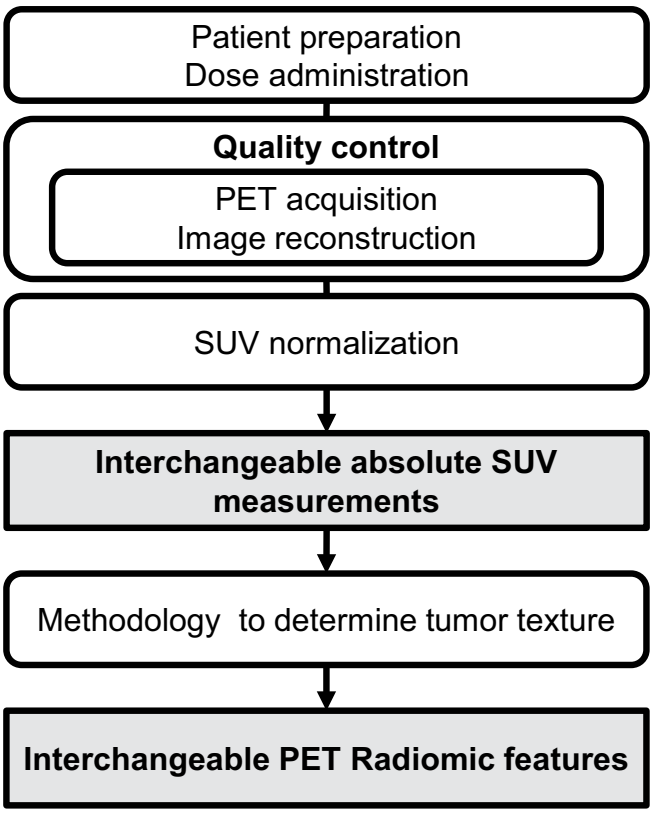


Figure 1. Levels of standardization in PET Radiomics. Interchangeable absolute SUV measurements are obtained by standardizing patient preparation, dose administration, image acquisition, image reconstruction and SUV normalization[26]. Standardization of the methodology used for tumor texture analysis ensures interchangeable PET Radiomic features and their ascribed values

or peak SUV[10], are described as predictors for treatment outcome[11–13,39]. Additionally, more advanced quantitative imaging features describing tumor image texture (i.e. the spatial arrangement of intensities within the image), which reflect intra-tumor heterogeneity of metabolic activity, are increasingly being investigated as potential imaging biomarkers in lung[14,15], head and neck[16,17], cervical[16,18], esophageal[19–21] and other cancers[22,23] — a field of research often referred to as 'Radiomics'[2,3,24,25,40,41,42].

Efforts have been made to provide guidelines for quality control measures in PET imaging and to standardize patient preparation, dose administration, image acquisition, image reconstruction and SUV normalization, in such a way that absolute SUV measurements are interchangeable in multicenter studies[26]. Interchangeable SUV measurements are very important in PET Radiomics, but the methodology used to determine textural features is also subject to variability. Standardization is therefore needed[27–29] (Fig. 1).

One important methodological factor is SUV discretization (i.e. resampling image intensity values). Discretization reduces the otherwise infinite possible number of intensity values to a finite set and effectively reduces image noise. Most recent literature describes using a fixed number (e.g. 8, 16) of discrete resampled values or 'bins' to divide the image SUV range into equally spaced intervals before calculating textural features[14–16,18–22,28,30–32]. Consequently, this results in discretized images with varying bin sizes or 'intensity resolutions,' depending on the SUV range. An alternative discretization method is to resample the image SUVs with a fixed bin size in units of SUV (e.g. 0.1, 0.5), maintaining a constant intensity resolution across all tumor images[33].

When aiming to identify imaging biomarkers in cohort and multicenter studies or trials, it is important that textural features and their ascribed values be directly comparable, both inter- and intra-patient, in order to derive meaningful conclusions. To our knowledge, the effect of the SUV discretization method in this respect has not been previously evaluated and we hypothesize that the aforementioned intensity resolution used for SUV discretization plays a key role in this regard.

The general objectives of our study are to compare both aforementioned conceptually different discretization methods for several popular textural features and to identify which of these methods is most appropriate for texture quantification in a clinical setting. We will specifically investigate the role of the intensity resolution and use a clinical case study to demonstrate the effect of the SUV discretization methodology on the interpretation of the assessed textural features.

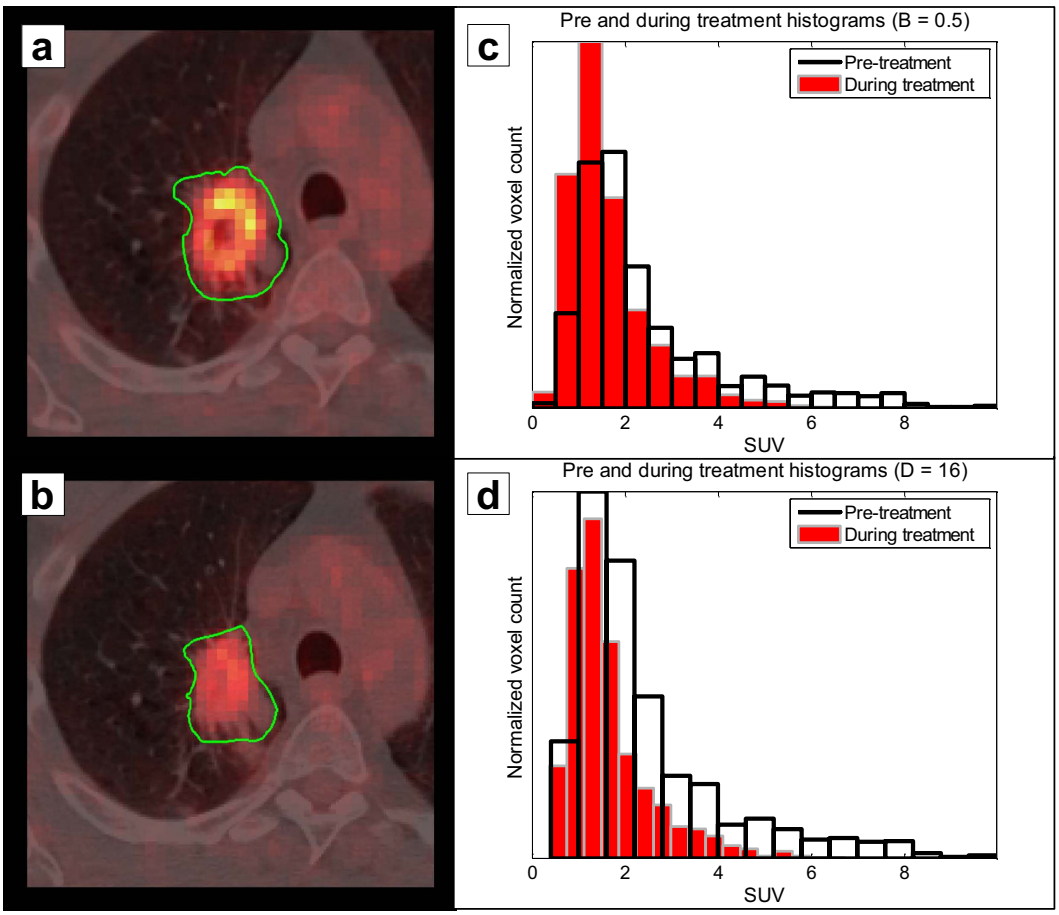


Figure 2. ***Left column:*** Representative images of sequential imaging for one patient, showing pre-treatment imaging (**a**) and imaging during the second week of radiotherapy (**b**). The tumor delineation is outlined in green. Both images are displayed with the same window/level settings. ***Right column:*** Histograms of the pre-treatment and during treatment images, resampled with a fixed bin size (i.e. intensity resolution) (**c**) or a predefined number of bins (**d**). In (**d**), one can appreciate the difference in resulting intensity resolution when resampling with a fixed number of bins. Pre-treatment and during treatment intensity resolutions were 0.6 and 0.37 [SUV], respectively

Materials and Methods

Patients and PET imaging. This study comprised 35 non-small cell lung cancer (NSCLC) patients who were prospectively included in a clinical trial (NCT00522639) and scheduled for radiotherapy and/or chemotherapy between July and December 2008[11]. 18F-FDG-PET/CT imaging was performed on a Biograph 40 PET/CT scanner (Siemens Medical Solutions) twice: (1) after induction chemotherapy but before radiotherapy and (2) during the second week of radiotherapy (Fig. 2a,b). Patients fasted for at least six hours before imaging. The injected amount of 18F-FDG was (4 × body weight [kg] + 20) MBq. Patients rested 60 minutes before image acquisition. Patients' blood glucose levels were below 10 mmol/L, so no correction for blood glucose level was applied.

PET images were iteratively reconstructed using normalization- and attenuation-weighted OSEM using 4 iterations, 8 subsets and a 5 mm Gaussian filter. The resulting images had an in-plane pixel size of 4 × 4 mm and a 3 mm slice thickness. PET images were converted into units SUV, normalized by patient body weight[9]. Tumor volumes of interest (VOIs) were manually delineated on fused PET/CT images for treatment planning purposes. Further details are described elsewhere[11]. This study was conducted according to national laws and guidelines and approved by the appropriate local trial committee at Maastricht University Medical Center (MUMC+), Maastricht, The Netherlands. All included patients signed an informed consent form.

Image processing and feature extraction. SUVs within the VOI were first discretized using: (1) a fixed bin size (B), or intensity resolution, in units of SUV (Fig. 2c) and (2) a fixed number of bins (D), or discrete resampling values (Fig. 2d). For image I, let $I(x)$ represent the SUV of voxel x, SUV_{min} the minimum SUV in I and SUV_{max} the maximum SUV in I. Resampling SUVs into bins with an intensity resolution of B was performed using:

$$I_B(x) = \left\lceil \frac{I(x)}{B} \right\rceil - \min\left(\left\lceil \frac{I(x)}{B} \right\rceil \right) + 1 \quad (1)$$

Where term $\lceil \min(I(x)/B) + 1 \rceil$ ensures that the bin count starts at 1. We use the shorthand notation R_B for this resampling method. Resampling SUVs into D bins was performed using:

$$I_D(x) = \begin{cases} 1 & I(x) = SUV_{min} \\ \left\lceil D \times \frac{I(x) - SUV_{min}}{SUV_{max} - SUV_{min}} \right\rceil & \text{otherwise} \end{cases} \quad (2)$$

Where the intensity resolution equals $(SUV_{max} - SUV_{min})/D$. This resampling method is denoted by R_D. Discretization using R_B and R_D was performed for different discretization values B (0.05, 0.1, 0.2, 0.5 and 1 [SUV]) and D (8, 16, 32, 64 and 128), respectively.

Textural features describing the spatial distribution of voxel intensities were calculated from gray-level co-occurrence (GLCM)[34], gray-level run-length (GLRLM)[35] and gray-level size-zone texture matrices (GLSZM)[21]. Texture matrices were determined by considering 26 connected voxels (i.e. voxels were considered to be neighbors in all 13 directions in three dimensions) at a distance of 1 voxel. Features derived from GLCM and GLRLM were calculated by averaging their value over all 13 directions. In total, 44 textural features (22 GLCM, 11 GLRLM and 11 GLSZM) were calculated. Changes in feature values between the pre-treatment and during treatment imaging time points were described as delta features, defined as:

$$\Delta X = X_{during\ treatment} - X_{pretreatment} \quad (3)$$

Image analysis was performed in Matlab R2012b (The Mathworks, Natick, MA) using an adapted version of CERR[36] and software developed in-house to extract textural features. Mathematical definitions for features assessed in this study are described elsewhere[33].

Statistical analysis. For both R_B and R_D, the pairwise intra-class correlation coefficient (ICC)[37] was calculated for each feature for all possible pairwise combinations of B (ICC_B) and D (ICC_D), to assess whether pre-treatment feature values were consistent for different discretization values. The ICC was defined as:

$$ICC = \frac{BMS - WMS}{BMS + WMS} \quad (4)$$

Where BMS and WMS are the between-subjects and within-subjects mean squares, respectively, obtained by Kruskal-Wallis one-way ANOVA. An ICC of 1 indicates perfect agreement (i.e. identical feature values).

Patient rankings according to feature value were created to serve as a surrogate for textural feature interpretation. Pairwise correlations between patient rankings were evaluated with Spearman's rank correlation coefficient (ρ). We compared patient rakings according to pre-treatment feature values and patient rankings according to delta feature values (ΔX), between all possible pairwise combinations of B (ρ^{BB}), D (ρ^{DD}) and B and D (ρ^{BD}). We considered a pairwise ρ to indicate acceptable concordance between rankings when $\rho > 0.9$. Statistical analysis was performed in Matlab R2012b.

Results

Consistency of feature values for varying intensity resolutions. To assess whether feature values (using either R_B or R_D) were consistent for different discretization values, we calculated the pairwise ICCs for each feature between different values of B (ICC_B) and D (ICC_D), respectively. This analysis was performed on the pre-treatment images. For each feature, we reported the range and median of all pairwise ICCs (Fig. 3). None of the observed pairwise ICCs was higher than 0.85, meaning that textural features and their ascribed value depend on the intensity resolution used for SUV discretization.

Variability of intensity resolution when resampling with a fixed number of bins. Using R_D, we determined the pre-treatment and during treatment bin sizes, as well as their difference, for each lesion. We observed a significant variation in both inter- and intra-lesional intensity resolution, which is directly proportional to the SUV range. The ratio of the largest with the smallest observed intensity resolution was 1254% for pre-treatment imaging and 1038% for during treatment imaging. Absolute percentage differences in intensity resolution between pre-treatment and during treatment images ranged between 0.5% and 56%, with a median of 21%.

Comparing patient rankings based on pre-treatment feature values. For each feature we determined the patient ranking according to feature value, using R_B and R_D for different resampling

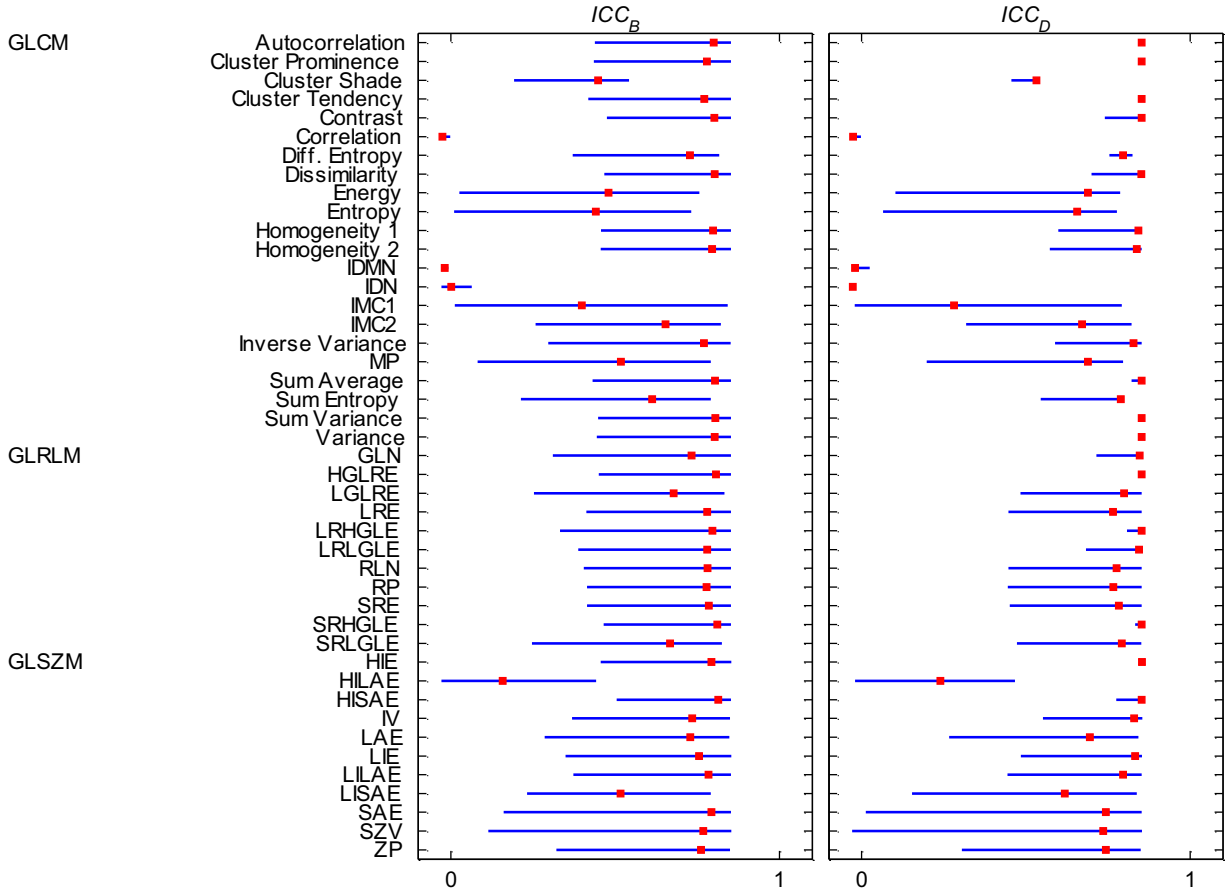


Figure 3. Graphical representation of pairwise ICCs for each feature for different values of B (ICC_B) and D (ICC_D), based on pre-treatment imaging. Blue lines extend from the minimum to the maximum observed ICC value. Median ICC values are represented by the red markers. ***Abbreviations of feature groups:*** gray-level co-occurrence (GLCM), gray-level run-length (GLRLM) and gray-level size-zone (GLSZM). ***Abbreviations of feature names:*** Difference Entropy (Diff. Entropy), Inverse difference moment normalized (IDMN), Inverse difference normalized (IDN), Informational measure of correlation 1 (IMC1), Informational measure of correlation 2 (IMC2), Maximum probability (MP), Gray-Level Nonuniformity (GLN), High Gray-Level Run Emphasis (HGLRE), Low Gray-Level Run Emphasis (LGLRE), Long Run Emphasis (LRE), Long Run High Gray-Level Emphasis (LRHGLE), Long Run Low Gray-Level Emphasis (LRLGLE), Run-Length Nonuniformity (RLN), Run Percentage (RP), Short Run Emphasis (SRE), Short Run High Gray-Level Emphasis (SRHGLE), Short Run Low Gray-Level Emphasis (SRLGLE), High Intensity Emphasis (HIE), High Intensity Large Area Emphasis (HILAE), High Intensity Small Area Emphasis (HISAE), Intensity Variability (IV), Large Area Emphasis (LAE), Low Intensity Emphasis (LIE), Low Intensity Large Area Emphasis (LILAE), Low Intensity Small Area Emphasis (LISAE), Small Area Emphasis (SAE), Size-Zone Variability (SZV), Zone Percentage (ZP)

values B and D, respectively. We then calculated pairwise ρ of patient rankings between different B (ρ^{BB}), different D (ρ^{DD}) and between different B and D (ρ^{BD}). For each feature, we reported the range and median of all pairwise ρ (Fig. 4).

We identified 14 GLCM and 6 GLRLM features to give reliable patient rankings for both discretization methods (i.e. all pairwise $\rho^{BB} > 0.9$ and all pairwise $\rho^{DD} > 0.9$), meaning that patient rankings were nearly not affected by changes in intensity resolution. GLCM 'Difference entropy,' GLRLM 'Gray-Level Non-uniformity (GLN)' and GLSZM 'High Intensity Emphasis (HIE)' were only found to provide robust patient rankings for different resampling values when using R_B. GLCM features 'Correlation,' 'Inverse Difference Moment Normalized (IDMN)' and 'Inverse Difference Normalized (IDN)' provided very similar patient rankings between both discretization methods, regardless of the value of either B or D (i.e. all pairwise $\rho^{BD} > 0.9$). All other features presented dissimilar patient rankings between both discretization methods.

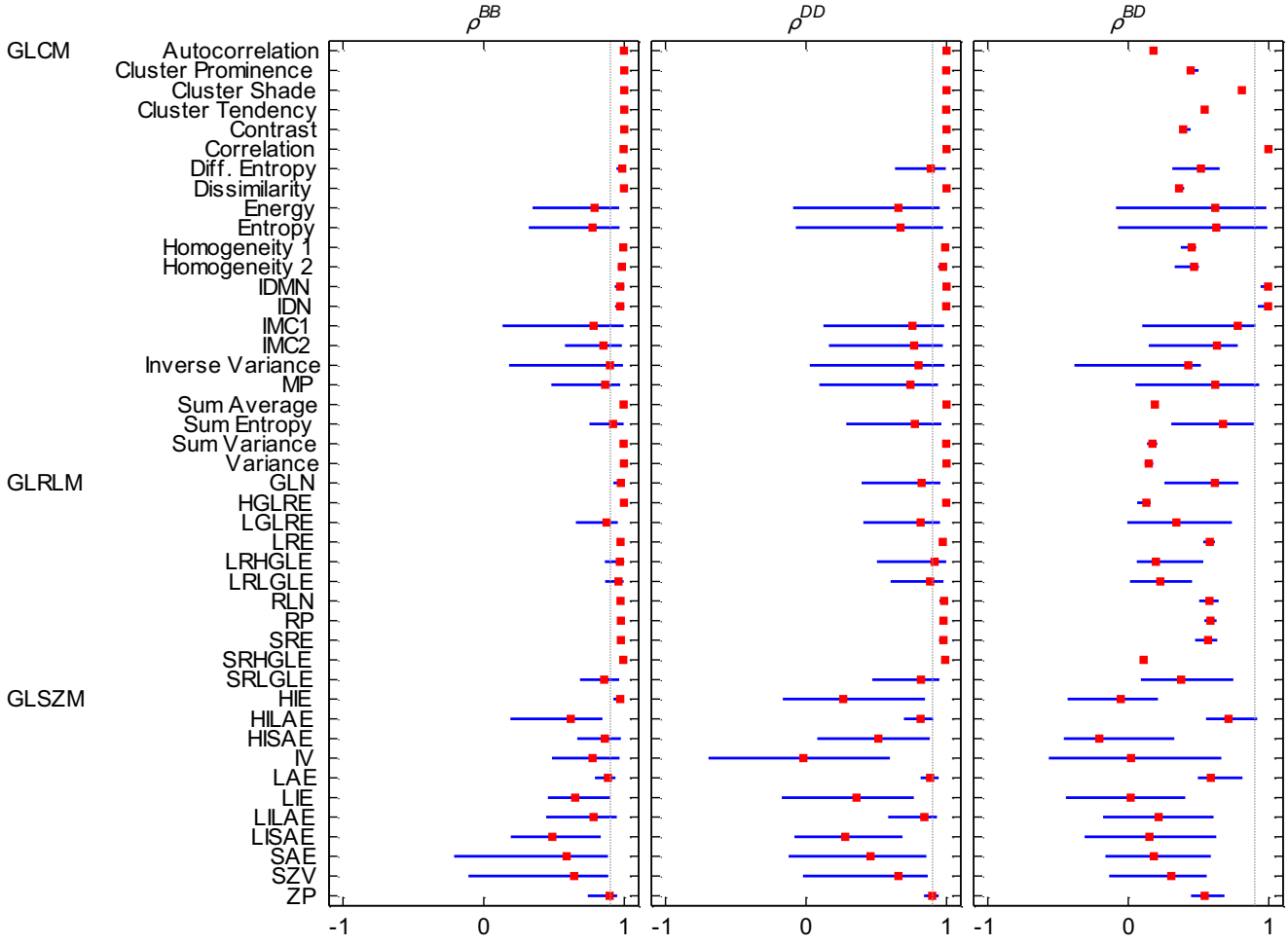


Figure 4. Graphical representation of pairwise Spearman rank correlations between patient rankings according to feature value for different B(ρ^{BB}), different D (ρ^{DD}) and between different B and D (ρ^{BD}), based on pre-treatment imaging. Blue lines extend from the minimum to the maximum observed pairwise ρ. Median ρ values are represented by the red markers. The gray vertical line represents $\rho = 0.9$. For abbreviations, see the caption of Fig. 3

Comparing patient rankings based on delta feature values. We also performed pairwise comparisons of patient rankings for each ΔX between different B (ρ_{Δ}^{BB}), D (ρ_{Δ}^{DD}) and different B and D (ρ_{Δ}^{BD}). For each feature, we reported the range and median of all pairwise ρ (Fig. 5). ρ_{Δ}^{BB} and ρ_{Δ}^{DD} were both higher than 0.9 for of 10 GLCM features and 2 GLRLM features. ΔX of GLCM features 'Difference Entropy,' 'Homogeneity 1' and 'Sum Entropy' were only found to give similar patient rankings for different resampling values when using R_B. For ΔX of GLCM features 'IDMN' and 'IDN,' this was the case when using R_D for different D. The high ρ_{Δ}^{BD} (0.95–1.00) for all pairwise comparisons for GLCM feature 'Correlation' indicated highly similar patient rankings based on ΔX between both discretization methods, regardless of the value of B or D. Some pairwise ρ_{Δ}^{BD} for GLCM features 'IDMN' and 'IDN' indicated similar patient rankings for ΔX as well, but with a large range for ρ_{Δ}^{BD} (0.55–0.98 and 0.57–0.99, respectively). For all other ΔX, assessed patient rankings were found to be discordant between both discretization methods.

Discussion

We compared tumor texture analysis based on SUV discretization using either a fixed number of bins (R_D) or a fixed bin size in units SUV (R_B), in the context of clinical treatment response assessment. Textural feature values were shown to depend on the intensity resolution used for SUV discretization. Overall, both resampling methods gave discordant results in terms of interpreting textural features. In the following section, we will discuss which method may be appropriate for use in a clinical setting.

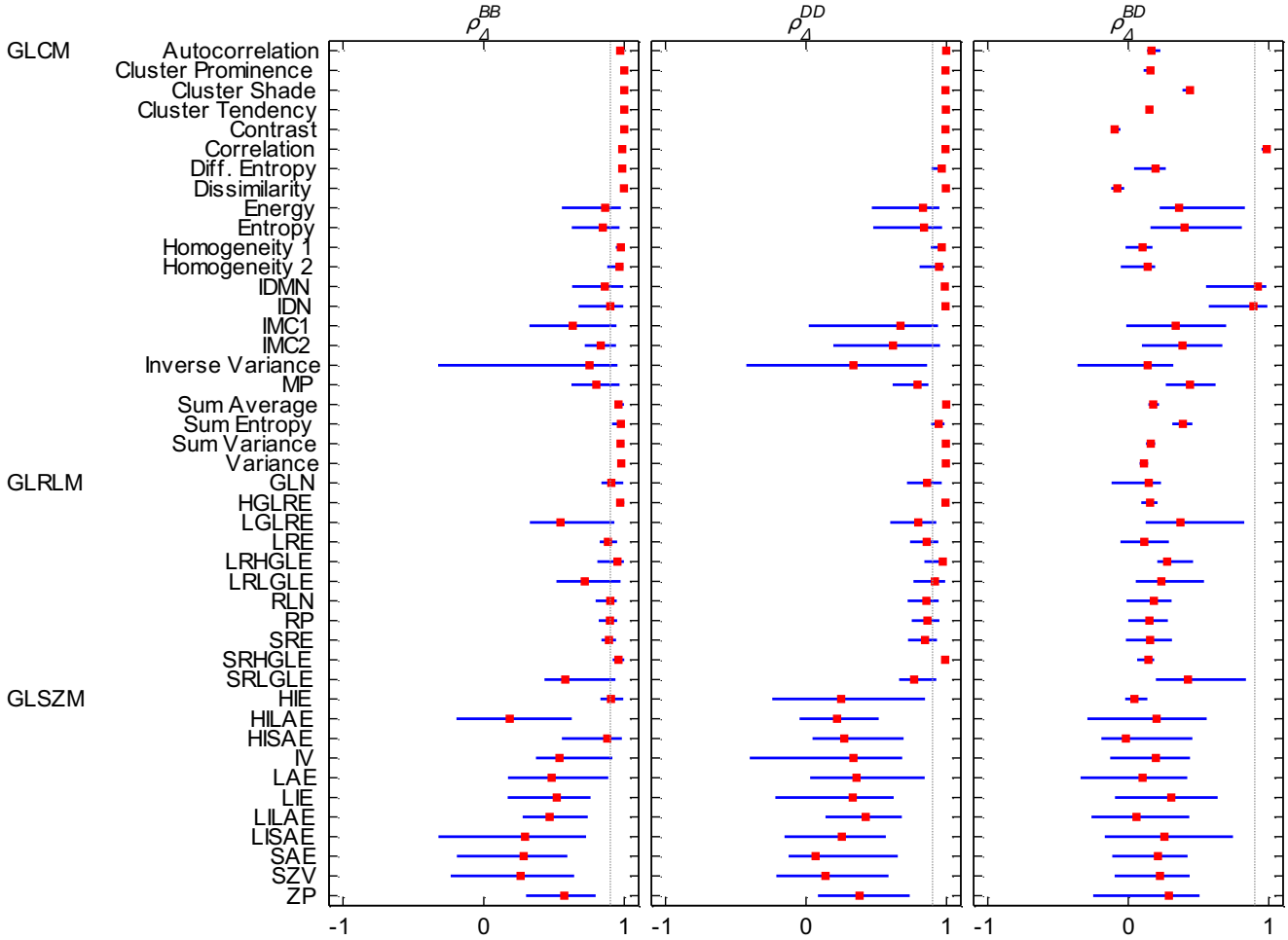


Figure 5. Graphical representation of pairwise Spearman rank correlations between patient rankings according to ΔX for different B (ρ_{Δ}^{BB}), D (ρ_{Δ}^{DD}), and different B and D (ρ_{Δ}^{BD}). Blue lines extend from the minimum to the maximum observed pairwise ρ_{Δ}. Median ρ_{Δ} values are represented by the red markers. The gray vertical line represents $\rho = 0.9$. For abbreviations, see the caption of Fig. 3

Correct comparison of textural feature values. As pointed out earlier, it is important that textural feature values be directly comparable, both inter- and intra-patient, in order to derive meaningful conclusions from tumor texture analysis. The key role of the intensity resolution in this respect can be illustrated by the mathematical background of histogram bin probabilities. Let X be a continuous random variable, such as SUVs in a tumor image, with probability density function $f(x)$. The bin probabilities $P(i)$ of the first order histogram, considering equally spaced and non-overlapping bins, are then defined as:

$$P(i) = \int_{t(i)}^{t(i)+w} f(x)\,dx \tag{5}$$

Where w represents the histogram bin size (i.e. the intensity resolution) and $t(i)$ denotes the left-hand endpoint of bin i. Analogous to $P(i)$, textural matrices are essentially histograms of joint probability densities that describe the probability of a voxel assigned to bin i being either (1) adjacent to a voxel assigned to bin j ($P_{GLCM}(i, j)$), (2) part of a consecutive run of l voxels assigned to bin i ($P_{GLRLM}(i, l)$) or (3) part of a connected neighborhood of v voxels ($P_{GLSZM}(i, v)$).

The aim is to compare textural feature values calculated from these histograms between tumor images. For all tumor images, the image intensities (x) are not dimensionless, but measured in SUV units. Maintaining a constant intensity resolution (w, in SUV units) across tumor images yields identical histogram probability definitions ($P(i)$) for each image, and hence directly comparable numerical values of each calculated feature. Using a non-constant intensity resolution across images causes one to quantify patterns (i.e. texture) on a different intensity scale (in terms of SUV) in each image.

By calculating pairwise ICCs, we observed that feature values indeed depend on the intensity resolution used for SUV discretization (Fig. 3). More importantly, there was a significant inter- and

intra-lesional variation in intensity resolution during the course of treatment when using R_D for image intensity resampling. Effectively, R_D discards the absolute radiotracer uptake information (i.e. metabolic activity), by considering each tumor image to have the same dimensionless range of intensity. In this respect, we consider R_D to be a less appropriate choice for SUV discretization in a clinical setting, as it results in textural feature values that are not defined on the same SUV scale for each tumor image. In contrast, a constant intensity resolution is maintained across resampled images when using R_B for SUV discretization, which we believe makes it a more suitable method for tumor texture analysis.

Impact of SUV discretization method and intensity resolution on the interpretation of textural features. Several textural features were found to provide reliable patient rankings using either R_B or R_D for discretization, suggesting that results based on these features may be compared between studies if they exclusively use R_B (where studies may use a different intensity resolution B) or R_D (where studies may use a different number of bins D). However, as discussed in the previous section, we find R_D to be less appropriate in a clinical setting considering that tumor image intensities are measured in SUV units and that tumor images generally do not have the same SUV range. We therefore illustrated the implications of SUV discretization with R_D instead of R_B on the interpretation of textural features in our clinical case study. Both discretization methods resulted overall in patients being ranked differently according to their feature value (Figs 4,5). These results show that the manner of SUV discretization can affect the interpretation of textural features and should therefore be carefully considered in tumor texture analysis.

We furthermore observed that when R_B was used, patient rankings for several features were affected by the choice of intensity resolution (B). For those features, at least one pairwise ρ^{BB} was found to be lower than 0.9 (Fig. 4). This suggests that results obtained for those features cannot be directly compared when different intensity resolutions are used and also suggests that their interpretation (e.g. prognostic or predictive value) depends on the intensity resolution.

It is noteworthy that the GLCM feature 'Correlation' was the only feature observed to have highly similar patient rankings over the course of treatment, regardless of the discretization method or discretization value used (Figs 4,5). This suggests that results obtained for this particular feature might be reliably compared between studies, provided the same discretization method is used throughout each specific study.

We used different arbitrary values for B and D throughout our study, where we kept the ratio between the smallest and largest B or D approximately the same and reasonably large. Although other values may be used as well, we found this selection to be sufficient to study our objectives. In terms of R_B however, an optimal intensity resolution cannot be straightforwardly determined. A value of 0.5 [SUV] has been described earlier, but without substantial motivation[33]. Methods for estimating an optimal intensity resolution could be performed[38]. It should then be emphasized that the same intensity resolution needs to be maintained throughout the entire study, as determining a separate bin size for each individual patient results in non-comparable feature values. However, estimating an optimal intensity resolution does not take into account the aforementioned effect the intensity resolution has on feature interpretation, as well as the fact that using different intensity resolutions may result in complementary information[17]. In this respect, clinical validation including outcome measures is necessary to identify optimal settings that lead to meaningful results in tumor texture analysis.

Standardization in texture analysis. FDG-PET quantification is affected by several factors, including for instance breathing motion in lung[26]. Recent studies have investigated several technical aspects of FDG-PET-derived textural parameters in different cancer sites, including their test-retest repeatability and robustness regarding tumor delineation or partial volume correction[31–33], or their variability due to image acquisition and reconstruction parameters[30]. In order to provide a complete overview and acknowledging that feature stability may as well be dependent on the methodology used for SUV discretization in tumor texture analysis, we did not exclude textural features previously reported to have limited repeatability or robustness. Reliability analyses should however be performed at specific settings used in tumor texture analysis, in order to identify those features suitable for treatment assessment. The aforementioned studies point to the importance of robust and standardized PET protocols in terms of reliable quantification of tumor heterogeneity with textural features, especially when the SUV is considered to be an interchangeable quantity[26,29]. This becomes even more essential when using fixed intensity resolutions for SUV discretization, as shown in this paper. Our study confirms that using standardized methodology for tumor texture analysis is also an important aspect of identifying and validating imaging biomarkers related to a certain outcome or underlying biology[43,44], between different studies or trials[27,28].

Conclusion

When aiming to identify and validate imaging biomarkers with tumor texture analysis of FDG-PET, it is important that the textural feature values be directly comparable, both inter- and intra-patient, in order to derive meaningful conclusions. We focused on the effect of SUV discretization and compared tumor texture analysis based on SUV discretization using a fixed intensity resolution (i.e. bin size) in units SUV (R_B) with using a fixed number of bins (R_D). We showed that maintaining a constant intensity resolution for SUV discretization across tumor images (R_B) yields textural feature values that are defined on the

same SUV scale, allowing for a meaningful comparison of texture between images. Discretizing SUVs using R_D was found to be less appropriate for inter- and intra-patient comparison of textural feature values in a clinical setting. The interpretation of textural features was overall different between both discretization methods and, for several features, affected by the choice of intensity resolution. Our study shows that the manner of SUV discretization has a crucial effect on the resulting textural features and the interpretation thereof and should therefore be carefully considered, underlining the importance of standardized methodology in tumor texture analysis.

References

1. Lambin, P. *et al.* Predicting outcomes in radiation oncology–multifactorial decision support systems. *Nat Rev Clin Oncol* **10,** 27–40, doi:10.1038/nrclinonc.2012.196 (2013).
2. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* **48,** 441–446, doi:10.1016/j.ejca.2011.11.036 (2012).
3. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5,** doi:10.1038/ncomms5006 (2014).
4. Lin, P. *et al.* Diagnostic and staging impact of radiotherapy planning FDG-PET-CT in non-small-cell lung cancer. *Radiother Oncol* **101,** 284–290, doi:10.1016/j.radonc.2011.06.030 (2011).
5. Lambin, P. *et al.* 'Rapid Learning health care in oncology' - An approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncol* **109,** 159–164, doi:10.1016/j.radonc.2013.07.007 (2013).
6. De Ruysscher, D., Nestle, U., Jeraj, R. & Macmanus, M. PET scans in radiotherapy planning of lung cancer. *Lung Cancer* **75,** 141–145, doi:10.1016/j.lungcan.2011.07.018 (2012).
7. Troost, E. G. *et al.* Innovations in radiotherapy planning of head and neck cancers: role of PET. *J Nucl Med* **51,** 66–76, doi:10.2967/jnumed.108.061499 (2010).
8. Van Elmpt, W., Pottgen, C. & De Ruysscher, D. Therapy response assessment in radiotherapy of lung cancer. *Q J Nucl Med Mol Imaging* **55,** 648–654 (2011).
9. Thie, J. Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med* **45,** 1431–1434 (2004).
10. Wahl, R. L., Jacene, H., Kasamon, Y. & Lodge, M. A. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* **50,** 122S–150S, doi:10.2967/jnumed.108.057307 (2009).
11. van Elmpt, W., Ollers, M., Dingemans, A. M., Lambin, P. & De Ruysscher, D. Response assessment using 18F-FDG PET early in the course of radiotherapy correlates with survival in advanced-stage non-small cell lung cancer. *J Nucl Med* **53,** 1514–1520, doi:10.2967/jnumed.111.102566 (2012).
12. Takeda, A. *et al.* The maximum standardized uptake value (SUVmax) on FDG-PET is a strong predictor of local recurrence for localized non-small-cell lung cancer after stereotactic body radiotherapy (SBRT). *Radiother Oncol* **101,** 291–297, doi:10.1016/j.radonc.2011.08.008 (2011).
13. Velazquez, E. R., Aerts, H. J., Oberije, C., De Ruysscher, D. & Lambin, P. Prediction of residual metabolic activity after treatment in NSCLC patients. *Acta Oncol* **49,** 1033–1039, doi:10.3109/0284186X.2010.498441 (2010).
14. Vaidya, M. *et al.* Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol* **102,** 239–245, doi:10.1016/j.radonc.2011.10.014 (2012).
15. Cook, G. J. *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, doi:10.2967/jnumed.112.107375 (2013).
16. El Naqa, I. *et al.* Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit* **42,** 1162–1171, doi:10.1016/j.patcog.2008.08.011 (2009).
17. Cheng, N. M. *et al.* Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. *J Nucl Med* **54,** 1703–1709, doi:10.2967/jnumed.112.119289 (2013).
18. Yang, F., Thomas, M. A., Dehdashti, F. & Grigsby, P. W. Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. *Eur J Nucl Med Mol Imaging* **40,** 716–727, doi:10.1007/s00259-012-2332-4 (2013).
19. Tan, S. *et al.* Spatial-Temporal [(18)F]FDG-PET Features for Predicting Pathologic Response of Esophageal Cancer to Neoadjuvant Chemoradiation Therapy. *Int J Radiat Oncol Biol Phys*, doi:10.1016/j.ijrobp.2012.10.017 (2012).
20. Dong, X. *et al.* Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor 18F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage. *Nuclear medicine communications* **34,** 40–46, doi:10.1097/MNM.0b013e32835ae50c (2013).
21. Tixier, F. *et al.* Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med* **52,** 369–378, doi:10.2967/jnumed.110.082404 (2011).
22. Tixier, F. *et al.* Correlation of Intra-Tumor 18F-FDG Uptake Heterogeneity Indices with Perfusion CT Derived Parameters in Colorectal Cancer. *PLoS One* **9,** e99567, doi:10.1371/journal.pone.0099567 (2014).
23. 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, doi:10.1007/s00259-012-2247-0 (2013).
24. Cook, G. J. R. *et al.* Radiomics in PET: principles and applications. *Clinical and Translational Imaging*, doi:10.1007/s40336-014-0064-0 (2014).
25. Naqa, I. E. The role of quantitative PET in predicting cancer treatment outcomes. *Clinical and Translational Imaging*, doi:10.1007/s40336-014-0063-1 (2014).
26. Boellaard, R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* **50 Suppl 1,** 11S–20S, doi:10.2967/jnumed.108.057182 (2009).
27. Brooks, F. J. On some misconceptions about tumor heterogeneity quantification. *Eur J Nucl Med Mol Imaging* **40,** 1292–1294, doi:10.1007/s00259-013-2430-y (2013).
28. 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, doi:10.2967/jnumed.113.129858 (2014).
29. Cheng, N. M., Fang, Y. H. & Yen, T. C. The promise and limits of PET texture analysis. *Ann Nucl Med*, doi:10.1007/s12149-013-0759-8 (2013).
30. Galavis, P. E., Hollensen, C., Jallow, N., Paliwal, B. & Jeraj, R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol* **49,** 1012–1016, doi:10.3109/0284186X.2010.498437 (2010).
31. Tixier, F. *et al.* Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. *J Nucl Med* **53,** 693–700, doi:10.2967/jnumed.111.099127 (2012).

32. Hatt, M., Tixier, F., Cheze Le Rest, C., Pradier, O. & Visvikis, D. Robustness of intratumour F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging*, doi:10.1007/s00259-013-2486-8 (2013).
33. 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, doi:10.3109/0284186X.2013.812798 (2013).
34. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural Features of Image Classification. *IEEE T Syst Man Cyb* **SMC-3,** 610–621 (1973).
35. Galloway, M. Texture analysis using gray level run lengths. *Comput Vision Graph* **4,** 172–179 (1975).
36. Deasy, J. O., Blanco, A. I. & Clark, V. H. CERR: a computational environment for radiotherapy research. *Med Phys* **30,** 979–985 (2003).
37. Shrout, P. E. & Fleiss, J. L. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychol Bull* **86,** 420–428 (1979).
38. Izenman, A. J. Recent Developments in Nonparametric Density Estimation. *Journal of the American Statistical Association* **86,** 205–224 (1991).
39. Carvalho, S. *et al.* Prognostic value of metabolic metrics extracted from baseline positron emission tomography images in non-small cell lung cancer. *Acta Oncol* **52,** 1398–1404, doi:10.3109/0284186X.2013.812795 (2013).
40. Parmar, C. *et al.* Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Scientific reports* **4,** 11044, doi:10.1038/srep11044 (2015).
41. Coroller, T. P. *et al.* CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol*, doi:10.1016/j.radonc.2015.02.015 (2015).
42. Parmar, C. *et al.* Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* **9,** e102107, doi:10.1371/journal.pone.0102107 (2014).
43. Panth, K. M. *et al.* Is there a causal relationship between genetic changes and radiomics-based image features? An *in vivo* preclinical experiment with doxycycline inducible GADD34 tumor cells. *Radiother Oncol*, doi:10.1016/j.radonc.2015.06.013 (2015). In press.
44. Hoeben, B. A. *et al.* Systematic analysis of 18F-FDG PET and metabolism, proliferation and hypoxia markers for classification of head and neck tumors. *BMC cancer* **14,** 130 doi:10.1186/1471-2407-14-130 (2014).

Acknowledgements

The authors acknowledge the support of the QuIC-ConCePT project, partly funded by EFPIA companies and the Innovative Medicine Initiative Joint Undertaking (IMI JU) under Grant Agreement No. 115151. Authors also acknowledge financial support from the National Institute of Health (NIH-USA U01 CA 143062–01, Radiomics of NSCLC), the CTMM framework (AIRFORCE project, grant 030–103), EU 6th and 7th framework program (EUROXY, METOXIA, EURECA, ARTFORCE), euroCAT (IVA Interreg - www.eurocat.info), SME Phase 2 (EU proposal 673780 – RAIL), Alpe d'HuZes-KWF (DESIGN), Kankeronderzoekfonds Limburg from the Health Foundation Limburg and the Dutch Cancer Society (KWF UM 2011-5020, KWF UM 2009-4454).

Author Contributions

R.T.H.L. and P.L. conceived of the project. R.T.H.L. analyzed the data and wrote the paper. G.N., S.C., W.J.C.E., E.G.C.T., R.B., H.J.W.L.A., R.J.G. and P.L. provided expert guidance, data, or analysis tools. All authors reviewed the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: 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; doi: 10.1038/srep11075 (2015).