IOP Publishing Journal Title
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

Impact of intensity discretization on textural indices of [18F]FDG-PET tumour heterogeneity in lung cancer patients

Attila Forgács[1,2], Monika Béresová[3], Ildikó Garai[1], Martin L. Lassen[4,5], Thomas Beyer[4], Matthew D. DiFranco[4], Ervin Berényi[3] and László Balkay[2]

[1] Scanomed Nuclear Medicine Center, Debrecen, Hungary

[2] Division of Nuclear Medicine and Translational Imaging, Department of Medical Imaging, Faculty of Medicine, University of Debrecen, Hungary

[3]Division of Radiology and Imaging Science, Department of Medical Imaging, Faculty of Medicine, University of Debrecen, Hungary

[4]QIMP Team, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria

[5]Cedars-Sinai Medical Center, AIM group, Los Angeles, USA

E-mail: attilaforgacsem@gmail.com

Received xxxxxx

Accepted for publication xxxxxx

Published xxxxxx

Abstract

Quantifying tumour heterogeneity from [18F]FDG-PET images promises benefits for treatment selection of cancer patients. Here, the calculation of texture parameters mandates an initial discretization step (binning) to reduce the number of intensity

IOP Publishing Journal Title
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

levels. Typically, three types of discrimination methods are used: lesion relative resampling (LRR) with fixed bin number, lesion absolute resampling (LAR) and absolute resampling (AR) with fixed bin widths. We investigated the effects of varying bin widths or bin number using 27 commonly cited local and regional texture indices (TIs) applied on lung tumour volumes. The data set were extracted from 58 lung cancer patients, with three different and robust tumour segmentation methods. In our cohort, the variations of the mean value as the function of the bin widths were similar for TIs calculated with LAR and AR quantification. The TI histograms calculated by LRR method showed distinct behaviour and its numerical values substantially effected by the selected bin number. The correlations of the AR and LAR based TIs demonstrated no principal differences between these methods. However, no correlation was found for the interrelationship between the TIs calculated by LRR and LAR (or AR) discretization method. Visual classification of the texture was also performed for each lesion. This classification analysis revealed that the parameters show statistically significant correlation with the visual score, if LAR or AR discretization method is considered, in contrast to LRR. Moreover, all the resulted tendencies were similar regardless the segmentation methods and the type of textural features involved in this work.

Keywords: texture analysis, radiomics, quantization, 18F-FDG-PET

1. Introduction

In-vivo tumour characterisation through non-invasive imaging and image-derived quantitative parameters promises a more accurate, personalized treatment planning in oncology (Gillies et al. 2016, Lambin et al. 2012, Aerts et al. 2014). Although quantifiable tumour properties such as size, shape and location have demonstrated clinical importance (Tschirhart et al. 2004, Mayr et al. 2006), there is a growing interest to also collect and analyse radiomics features (intensity, shape, or texture) from a range of imaging methods. Positron emission tomography (PET) imaging, in particular, is being used for tumour detection and delineation and for multi-parametric assessment of tumour characteristics (van Baardwijk et al. 2008, El Naqa et al. 2009, Chicklore et al. 2013). Recent research into PET tumour heterogeneity quantification has focused on [18F]FDG-PET imaging and the use of textural indices (TI) derived from the voxel-level standardised uptake value (SUV) in tumour volumes-of-interest (VOI) (Orlhac and Nioche 2017, Tixier et al. 2012, Hatt et al. 2017, Pyka et al. 2016, Bailly et al. 2016, Gillies et al. 2016, Lasnon et al. 2016, Forgacs et al. 2016, Cook et al. 2015, Leijenaar et al. 2015, Orlhac et al. 2016, Buvat et al. 2015, Yan et al. 2015, Tixier et al. 2011, Orlhac et al. 2018). One important family of TIs is the higher order texture features (including, but not limited to co-occurrence, size zone and run length matrix based methods), which numerically merge both the intensity values and spatial relation of the voxels. For this purpose, the original image intensities are first discretized to a reduced number of bins, and then grey-level matrices are generated, for example the grey level co-

IOP Publishing Journal Title
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

occurrence matrix (GLCM), the grey level size zone matrix (GLSZM), or the grey level run length matrix (GLRLM). These matrixes serve as a basis for the calculations of higher order statistics such as entropy, homogeneity, or short run emphasis (Haralick et al. 1973, Galloway 1975).

Regardless of the specific TI being used, discretization converts the voxel values and is a mandatory pre-processing step, with the aim of normalizing texture matrices sizes to a manageable value. It also usually reduces noise. In general, resampling converts a continuous variable (SUV values) into a discrete variable, which takes values from a finite set (Anil 1989). Typically, in medical imaging if given a digital image I where the intensity of a pixel I (i) ∈ {0 . . . N}, quantisation is accomplished by assigning each pixel I (i) ∈ {0 . . . D}, where D < N. As a result, the quantisation process prior to textural index calculation represents an irreversible scale reduction in the medical images, thus inducing distortion.

Three distinct resampling method were recently implemented and investigated, thereby presenting a challenge in defining the diagnostic value of each feature. Tixier et al. introduced a method to discretize the voxel-values into a fixed number of discrete bins (Tixier et al. 2011). A second method, introduced by Leijenaar et al., uses a constant SUV bin width (e.g., 0.3), whereas Orlhac et al. has proposed to use a fixed bin with, however discretizing the whole SUV scale, instead of individually for each lesion, resulting in a standard discretization level across the population (Orlhac et al. 2015).

Figure 1 show the differences between the quantisation processes by considering two different lesions (labelled with two dashed intervals), lesion 1 and lesion 2, with heterogeneous SUVmin and SUVmax values.

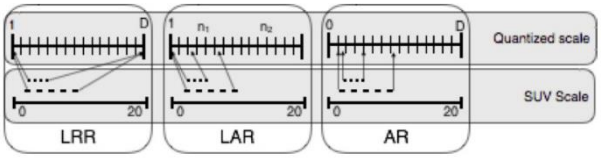

Figure 1. Comparison of quantisation algorithms: The lesion relative resampling method (LRR) normalises the range of SUV values for each VOI to a fixed number of bins, resulting in variable bin sizes across VOIs as well as a loss of consistent SUV scaling. Lesion absolute resampling (LAR) uses a fixed bin size for all VOIs, but assigns the first bin to the SUVmin of the lesion, leading to a variable number of bins as well as the loss of relative SUV values between lesions. Absolute resampling (AR) uses a fixed SUV scale, bin size and number of bins, meaning relative SUV values are preserved across VOIs after quantisation.

In Figure 1, the LRR method assigns the SUVmin and SUVmax of both lesions to the same range while disregarding the SUV range of the original lesions. As a result, the SUV range for each lesion is scaled differently, meaning that the same bin number in lesion 1 and 2 will not correspond to the same original SUV values. The LAR method does not scale the SUV scales differently for the two lesions, but the absolute values of the SUV scale are still not preserved since the SUV value of the first bin is determined individually for all lesions. In contrast, the AR method preserves the original

SUV scale and range by applying a simple quantisation of SUV values that maps the original values monotonically into a fixed scale from SUV 0 to 20. It has been demonstrated recently that different discretizations can have large impact on the calculated textural features and their related repeatability, and even on the correlative relationships between other variables such as SUVmax, SUVpeak and volume (Desseroit et al. 2017, van Velden et al. 2016, Shen et al. 2017). However, the results were contradictory, and the publication of Shen et al. does not even state clearly which quantization methods were applied.

In this work, we aimed to comprehensively express the effect of different discretization methods on texture evaluation using the same patient data set, focusing on several local (GLCM) and regional (GLSZM, GLRLM) based parameters. We present the variation in numerical values of textural indices introduced by different quantization and tissue segmentation processes, the cross-correlation of TIs calculated by these three methods, and the level of agreement between the numerical values and visual ranking.

2. Materials and Methods

2.1 Ethical Statement

The Regional and Institutional Ethics Committee, Clinical Center at the University Of Debrecen, Hungary approved this clinical study, which was then carried out in accordance with the relevant guidelines and regulations. As our study involves retrospective analysis of that data, informed consent was not required.

2.2 Patients and image acquisition

PET data of 58 patients (35 males, 23 females) with confirmed lung lesions were analysed retrospectively. Patients fasted for at least 6 hours before administration of [18F]FDG 325±73 MBq. Blood glucose levels were under 12 mmol/l. PET/CT images from all patients were acquired on the Philips Gemini TF 64 PET/CT system with emission scans beginning 60 minutes after radiotracer injection. Acquisition duration was 60-150 seconds per bed position, depending on the patient's weight. Image reconstructions were performed using the default Blob-OS-TF 3D ordered subset iterative TOF technique in 144 x 144 matrices with voxel size 4 x 4 x 4 mm3.

2.3 Quantisation Methods

Lesion relative resampling (LRR) as introduced by Tixier at al. (2012) were calculated according to:

$$I_{LRR}(i) = \begin{cases} 1 & SUV(i) = SUVmin \\ \left[Dx\frac{SUV(i)-SUV\,min}{SUVmax-SUVmin}\right] & otherwise \end{cases} \quad (1)$$

where $I_{LRR}(i)$ is the resampled i-th voxel of a lesion, D represents the predetermined number of bins (typical integer power of 2), $I_{LRR}(i)$ is a positive integer between 1 and D ($I_{LRR}(i) \in \{1 \ldots D\}$), and SUVmin and SUVmax are the minimum and maximum SUV values of the lesion. The square brackets designate the ceiling mathematical operation. We refer to this resampling method as lesion relative resampling, or LRR, since the SUV values for a given tumour are binned relative to the SUV values of each individual tumour. Five discrete bin numbers (16, 32, 64, 128 and 256) were selected for the LRR method.

Lesion absolute resampling (LAR) method, described by Leijenaar et al. is described as

$$I_{LAR(i)} = \left[\frac{SUV(i)}{B}\right] - \left[\frac{SUVmin}{B}\right] + 1 \quad (2)$$

where B represents the fixed bin width.

The absolute resampling (AR) method is defined in (Orlhac et al. 2015) as

$$I_{AR(i)} = round\left[\frac{C}{20}xSUV(i)\right] \quad (3)$$

where C represents a fixed bin number, typically 64 or 128. Because AR quantisation does not rely on lesion-specific SUVmin or SUVmax, the quantized SUV values retain their relationship to the original SUV scale and their proportionality for inter-subject comparisons, unlike the LRR method.

2.4 Tumour delineation

Lung lesion dataset was generated consisting of 63 lesions, which were segmented with three different methods. In a recent comprehensive review (Hatt, Lee et al, 2017) a large number of automatic methods were evaluated for PET studies. Based on this work we selected three inherently different, yet simple delineation approaches. In a first case, the tumours were manually delineated by an expert nuclear medicine physician defining free hand VOIs, without considering the anatomical information from the CT images. The necrotic areas within the tumour were not excluded in this study. The two additional methods were threshold based selecting the voxels in a larger predefined volume with the following criteria: i.) > 2.5 SUV value; ii.) > 40% of the maximal SUV value. We will refer these two threshold based methods as SUV_2.5 and 40%_SUVmax. Lesions below 10-25 mL were usually not considered, since heterogeneity measures of such small lesions have been shown to be unreliable (Brooks and Grigsby 2014, Hatt at al. 2015, Forgacs et al. 2016). However, the state-of-the-art recommendation is to include all lesions (Hatt, Tixier et al. 2017), therefore in this study we did not exclude lesions based on volume sizes.

2.5 Texture feature extraction

All feature calculations were performed in Matlab utilizing custom implementation of GLCM (Matlab script 1.), GLSZM (Matlab script 2.) and GLRLM (Matlab script 3.) algorithms. The heterogeneity parameters were computed by aggregating the different underlying directional matrices in the following way: 5 GLCM features were computed from a single matrix after merging all 3D directional matrices; 11 GLSZM features were computed from a 3D matrix; 11 GLRLM features were computed from each 2D directional matrix and averaged over 2D directions and slices. All the 27 features were defined according to the image biomarkers standardization initiative (IBSI) documentation (Zwanenburg et al. 2019). The full list can be found in the supplemental material (Table 1).

2.6 Influence of discretization parameters

Five textural indices of segmented lesions were calculated and plotted with three discretization methods, as a function of the parameters set (bin width, or bin number according to the method). While for the LRR method the number of bins (parameter D in Eq.1.) were varied as 16, 32, 64, 128, 256, in case the LAR the width of the bins was increased (parameter B in Eq.2.) as the $B \in \{0.05 \ldots 1\}$ in steps of 0.05. In case of the AR method, the number of bins denoted as C in Eq. 3 were varied (in the range from 20 to 400) in order to use the same bin widths as investigated with LAR method.

2.7 Correlation plots

Pairwise scatter plots of the numerical value of each lesion according to different discretization methods are presented (LRR 64 bin, LAR bin width 0.3, AR bin width 0.3

2.8 Visual assessment and parameter values

Visual classification of the texture was performed for each lesion by three experienced nuclear medicine physicians. There were no any exclusion criteria defined for the lesions (for instance excluding necrotic centre). An in-house Matlab software displayed the lesions one at a time in random order. The lesions were ranked from 1 to 5, reflecting the level of heterogeneity between absolute homogeneous (number 1) and highly heterogeneous (number 5). Prior to the scoring, a separate set of lesions was displayed to train the observers. The observers assigned visual scores for each lesion independently. Inter-rater reliability was assessed using two-way mixed, agreement, average measures intra-class correlation (ICC) to evaluate the degree that coders provided consistency in their ratings of heterogeneity across lesions (Hallgren et al. 2012). Finally, all the lesions received a visual score as the average of the three independent scores. Using IBM SPSS Statistics 22, Spearman correlation was calculated between visual scores and heterogeneity parameters derived from LRR (64 bin), LAR (bin width 0.3), AR (bin width 0.3) methods.

3. Results

Figure 2 shows how the discretization method and its parameter set affect the numerical value of investigated textural indices at cohort level after manually delineating all lesions and in case of the GLCM based five features. Each of the individual boxes represent the value of a given textural parameter extracted from the lesion dataset. The subplots are organized in three columns denoting the three investigated discretization methods, and the box-and-whisker subplot describes the distribution of the parameter values.

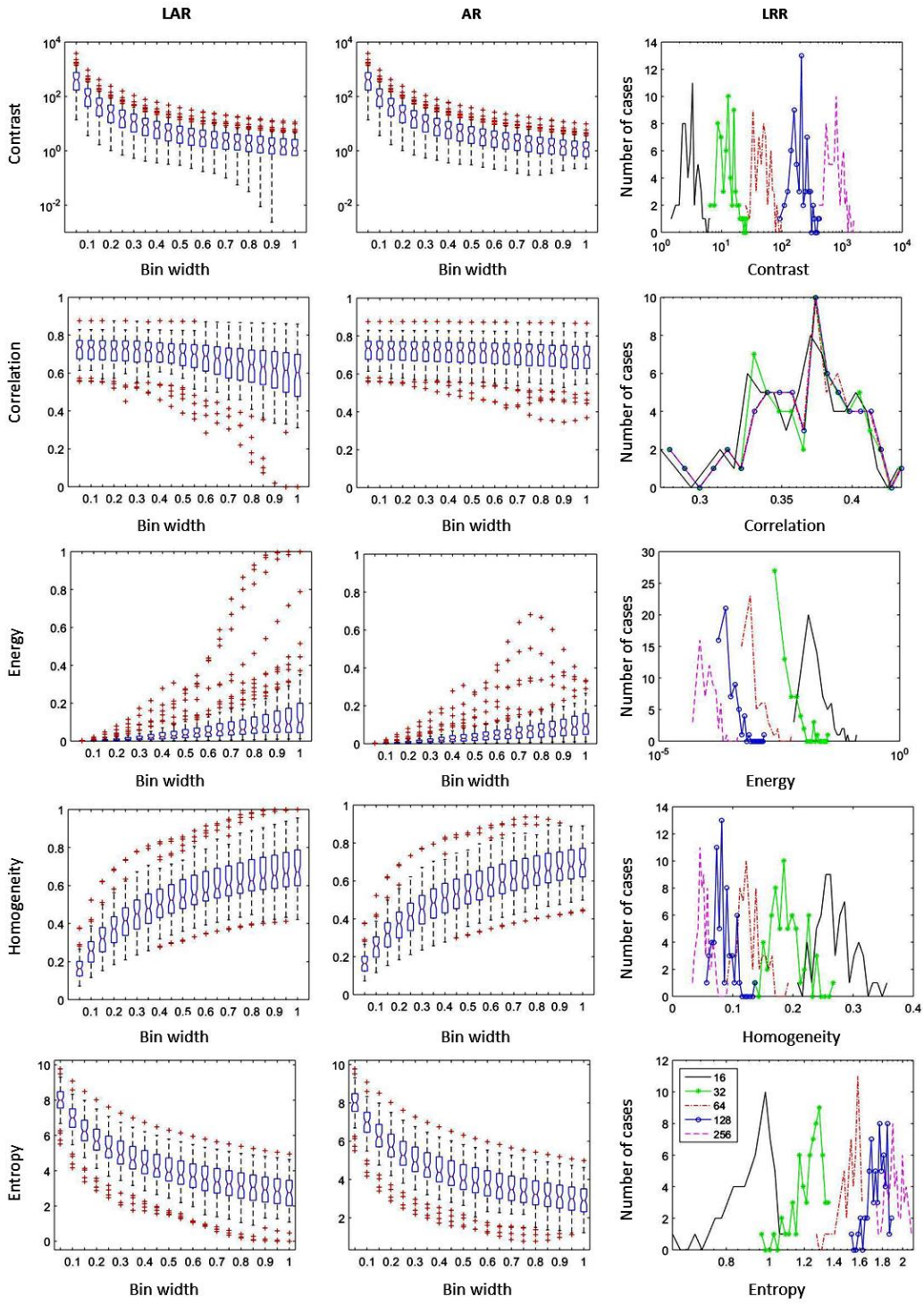

Figure 2. The distribution of GLCM based five features of the manually delineated lesions. The box-plots in the left and center column are related to the LAR and AR quantification methods (bin width from 0.05 to 1 SUV). The right column denoted for LRR method, applying 16, 32, 64, 128 and 256 number of bins.

The values of investigated textural parameters show two distinct trends: LAR and AR show similarly a monotonic trend with the applied bin width, but in case of the LRR method, the histogram of textural indices shows a shift on the x-axis according to the bin width. All the five GLCM texture parameters change dramatically in value (>100%) as the SUV bin width increased from 0.05 to 1 (LAR and AR method), or as the number of bins decreases (LRR method), except the Correlation which show high stability. Analysing the same manually delineated data set, the very same tendencies were found for the 11 GLSZM and the 11 GLRLM based features (supplemental material Fig. 3a and Fig. 4a), that is all the 22 texture features variate in high degree when the SUV bin width changes from 0.05 to 1 (LAR and AR), or as the number of bins alters (LRR method).

For a given lesion the bin width and the number of bins can be converted. Simplistically we can state that the application of LRR method on a group of data will result a lesion-dependent bin width. Since each of the lesions has different SUVmin and SUVmax voxel values, the usage of a fixed number of bins will result in a variation in bin width from lesion to lesion relative to the range of SUV values. Figure 3. displays how the bin width is distributed on the whole population of manually delineated lesions when applying a fixed number of bins.

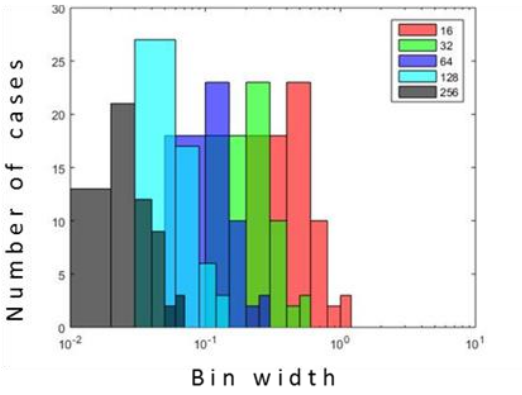

Figure 3. Bin width (B value) histograms in the case of the LRR quantization, for the dataset of manually delineated lesions.

Based on the data of Figure 2, pair-wise correlation scatter plots between the parameters calculated by three different quantization method are displayed in Figure 4.

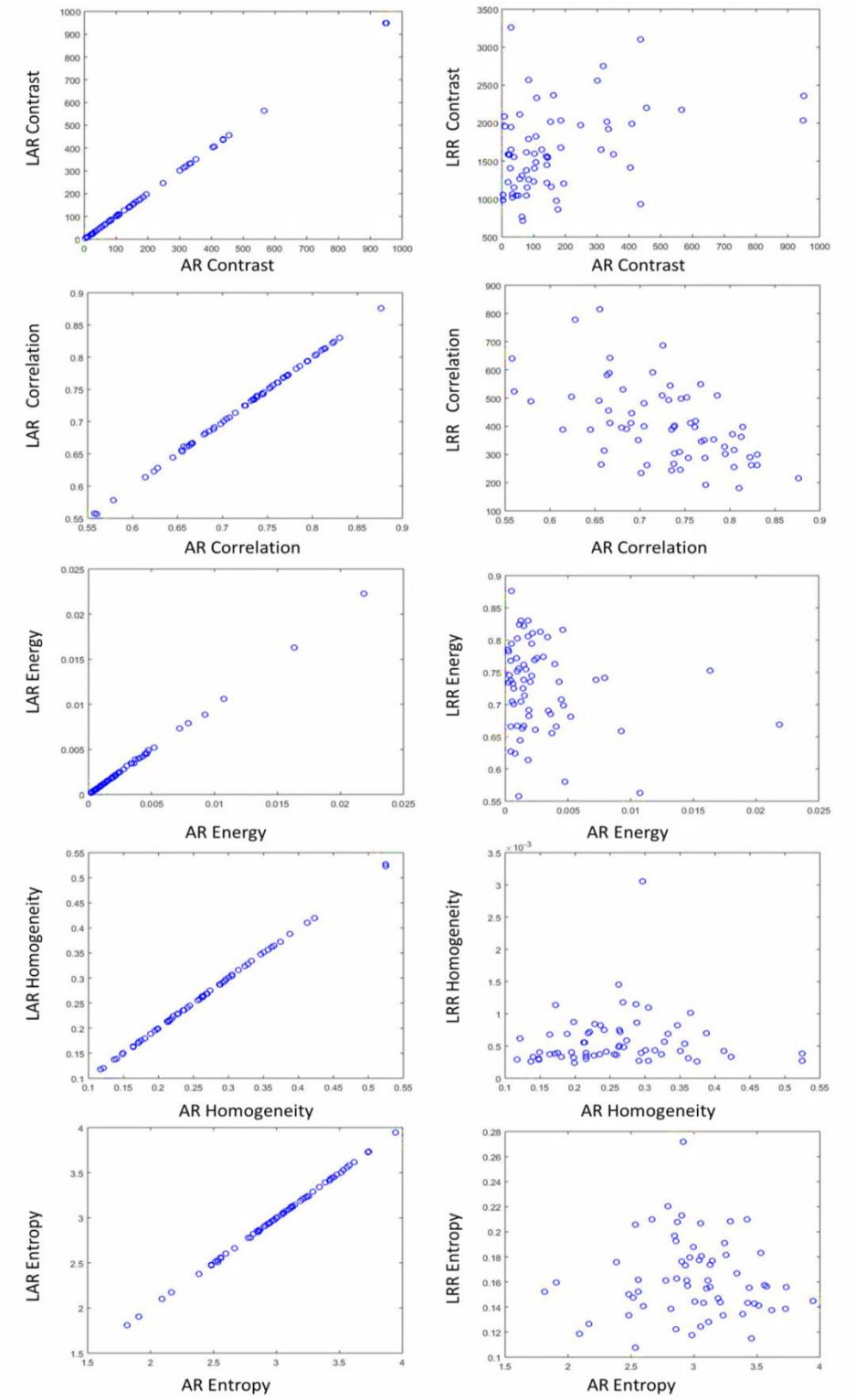

Figure 4. Pairwise scatter plots of the five calculated GLCM based texture parameters among the three different quantization methods including all manually segmented lesion VOIs. During these calculations the B was set to 0.1 for the AR and LAR method, and for the LRR quantization the bin number was 128.

Figure 4. clarifies that the LAR and AR methods result in highly correlated values, while the LRR method provide independent values. Considering the additional 22 regional GLSZM and GLRLM features, all related correlation plots show very similar tendency, the LAR and AR values correlate well, while the interrelationships between LRR-AR or LRR-LAR data are generally week in all plots (supplemental material Fig. 5 and Fig. 6). Interestingly, the Correlation parameter shows high similarity in the values calculated with the help of three different discretization methods.

We also investigated how the lesion segmentation method affects the results presented above. The resulted delineated volume distributions are presented in a box plot form in Figure 5.

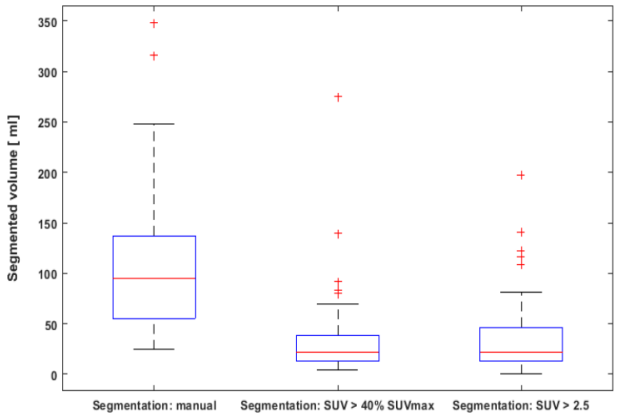

Figure 5. Box plot of the segmented volumes in different cases of delineation methods.

Considering all the segmented lesions the smallest and the largest volume was 0.77 mL and 347.0 mL, respectively. While several small lesions (< 20 mL) were generated during the segmentation process, in this study we did not exclude lesions based on volume sizes. To present the effect of the different delineation methods (manual, SUV_2.5 and 40%_SUVmax) we recalculated the heterogeneity distribution plots presented in Figure 2, supplementary Fig.3a and supplementary Fig.4a. The computed heterogeneity distribution plots for the SUV_2.5 and 40%_SUVmax segmentation methods are displayed in the supplemental material Fig 2b, Fig3b, Fig 4b and Fig 2c, Fig3c, Fig. 4c. All plots in Fig.2, Fig.3 and Fig.4 are related to the GLCM, GLRLM and GLSZM based features, respectively. In general, it can be observed that all feature distributions are very similar regardless of the delineation method.

Table 1 summarizes the results of a correlation analysis between visual score of heterogeneity and the GLCM based calculated values according to the discretization methods. The Correlation parameter does not show any correlation with the visual scores, regardless of the resampling method. The Contrast, Energy, Entropy and Homogeneity values show significant correlation with visual score if the LAR or AR method is applied. There was no correlation at all when the LRR method was applied. Similar results are achieved by analysing the same correlations in the case of the

GLSZM and GLRLM based features (see the Table 2 in the supplementary material). The calculated ICC shows good satisfactory level, ICC=0.871 (95% confidence interval, 0.804-0.918) indicating that readers had a high level of agreement and suggesting that heterogeneity was rated similarly across coders (Koo TK et al. 2016).

		Correlation			Contrast			Energy			Entropy			Homogeneity		
		LRR	LAR	AR	LRR	LAR	AR	LRR	LAR	AR	LRR	LAR	AR	LRR	LAR	AR
Score	Correlation Coefficient	.116	.146	.154	.004	.251*	.259*	.055	-.361**	-.361**	.096	.366**	.371**	.062	-.316*	-.305*
	Sig. (2-tailed)	.366	.255	.227	.0974	.047	.040	.667	.004	.004	.452	.003	.003	.627	.012	.015
	N	63	63	63	63	63	63	63	63	63	63	63	63	63	63	63

*Correlation is significant at the 0.05 level (2 tailed)

**Correlation is significant at the 0.01 level (2 tailed)

Table 1. Statistical results of Spearman correlation analysis between the visual score and texture indices

3. Discussion

In this study we focused on the role of three different discretization methods on the numerical value of GLCM, GLSZM and GLRLM based radiomics features in [18F] FDG images on the very same data set. We showed that a given texture feature can vary over a very wide range (a feature can change more than 100%) when changing the actual SUV bin width or number of bin values during the discretization step. Moreover, all the resulted tendencies were similar regardless the segmentation methods and the type of textural features involved in this work. The investigated discretization methods were: lesion relative resampling (LRR) introduced in 2011 by Tixier et al., lesion absolute resampling (LAR) introduced by Leijenaar et al. in 2015, and the absolute resampling method (AR) introduced in 2015 by Orlhac et al. These discretization methods transform the original voxel values according to a given number of bins (LRR), or according to a given bin width (LAR and AR). The comparative evaluation was performed on the same cohort, consisting of 63 delineated lung lesions in 58 clinical patients. For completeness, three different tumour segmentation approaches were involves, manual and 2 different threshold basis methods (SUV=2.5 and 40% of the SUVmax). Figure 2, supplementary Fig.3a and Fig.4a demonstrates the response of textural indices of manual segmented lesions as a function of bin width or number of bins for 27 frequently used GLCM, GLSZM and GLRLM based heterogeneity parameters. Each of the investigated parameters shows monotonic trend in the numerical value. Considering the range of each parameter, the actual numerical values (Y axis on each subplots) reflect more homogeneous or more heterogeneous patterns depending on the bin number or bin with (X axis). Interestingly, at a given bin width or given number of bins, the values of the lesions are distributed in a relatively narrow range. It is a striking fact, since in the lesion cohort (n=63) we can expect that some of the lesions represent highly homogeneous and some of show highly heterogeneous distributions. In contrast with this expectation, we found no parameter value (bin with or bin number), which resulted in these two extremes appearing in the box-plots. In addition, all the 27 texture parameters change dramatically in value (>100%) as the SUV bin width increased from 0.05 to 1 (LAR and AR method), or as the number of bins decreases (LRR method). It seems that the calculated values for given texture feature can be obtained in a wide range, selecting an appropriate SUV bin width or number of bin value. We found that this tendency did not alter, if the lesion delineation method is conceptually different. In principle, the delineation method essentially influences the volume and the texture of a lesion. Thus, to examine the relevance of the results in the aspect of delineation method, we extended the analysis with fixed-thresholding (2.5 SUV), and 40% of the SUVmax delineation methods (supplementary Fig.2, Fig.3 and Fig.4).

In case of the LRR method being applied, the actual bin width varies with the lesions according to the minima and maxima of each lesion. Figure 3 presents the distribution of

the bin widths seen in our cohort when using LRR for different numbers of bins. The pair-wise cross correlation scatter plots demonstrate differences between AR and LAR methods that are small but significant in the case of LRR quantification (Figure 4, supplementary Fig. 5 and Fig .6). This suggests that the results and previously published findings are more interchangeable between the AR and LAR methods. However, results based on LRR calculations are not transferable to studies, which use the LAR or AR method. The correlation study between visual score and the numerical values of the parameters confirm the previous results (Table 1. and Supplementary Table 2.). There are significant correlations if the AR or LAR method is applied, but no significant correlation is found if the discretization were done according to the LRR method. This outcome however does not indicate the goodness of one quantization method against the others, rather present the fact of the distinct behaviour. Beyond the visual judgement, the quantitative analysis of the texture has the potential to hold visually latent substance (Hatt, Tixier et al. 2017). Even so, in the aspects of the human thinking, the parameters showing correlation with visual expectations maybe more desired. The calculated 0.871 value of ICC indicate that a low amount of measurement error was introduced by the independent readers, and thus the statistical power is not substantially reduced.

The correlation parameter has a distinct nature in comparison with the other examined parameters. It is not sensitive to the bin number (Figure 2, supplemental material Fig. 3a and Fig. 4a) and it has no tendency as a function of bin width. It shows similar values independently from the discretization method (Figure 4, supplemental material Fig. 5 and Fig. 6), and moreover, it does not show any correlation with the visual score (Table 1., supplemental material Table 2.). The robustness of Correlation as a GLCM-based heterogeneity parameter was previously suggested by (Leijenaar et al. 2015). However, while it seems to be robust, it may on the other hand obscure an elementary error.

It has been demonstrated that LRR and LAR discretizations can have a large impact on the calculated textural features and their related repeatability (Desseroit et al. 2017, van Velden et al. 2016). Desseroit et al. found the LRR method superior in terms of the reliability at [18F]FDG PET studies, but the results of van Velden et al. showed better reliability in case of the LAR quantization. Recently, a study investigating the differences between the LRR and LAR methods based on ROC analysis and including 33 TIs, showed that more TIs will have better performance if the discretization method uses a fixed bin number (LRR) than a fixed bin size (Shen et al. 2017). Interestingly, the areas of the ROC curves were generally greater if they selected the least number of bins, 4, from the range of {4, 8, 16, 32, 48, 64, 80, 96, 112, or 128}. This is a strange result because dividing the original SUV scale to four bins only, the image information content will dramatically degrade. Unfortunately, this publication does not state clearly whether the LAR or AR method was applied. In addition,

Lucia et al. (2017) also showed that the predictive capability of a feature could depend greatly on the applied discretization method.

Image intensity quantisation is applied far beyond image texture analysis. From an information theory point of view, substantial quantisation can cause the loss of the information content. Conversely, finer quantisation is expected to improve the accuracy and separability of pixel values. In [18F]FDG-PET image processing, quantisation methods have also been implemented as morphological operators (Lucht et al. 1996) and non-uniform quantisation (root-squared) has also been studied (El Naqa et al. 2009). However, uniform quantisation is the de facto standard. The effects of the number of grey levels (Gómez et al. 2012), the impact of image reconstruction (Yan et al. 2015), the error of reproducibility (Tixier et al. 2012) and the effects of tumour volume (Brooks et al. 2014) have all been examined, but the search for an optimised quantisation method itself has yet to be fully addressed. Some examples of different types quantisation algorithms are equal probability (Haralick et al. 1973) or Gaussian weighted (Jobanputra and Clausi 2004), but an optimised quantisation method should be developed, which is specific to the scale of [18F]FDG SUV values. The SUV scale is an absolute scale in the PET field, and it is desirable to conserve this metabolic information for heterogeneity analysis. In principle, the smaller SUV bin widths should result in a lower quantisation error. However, the information content of the SUV value of a given voxel is affected not only by noise, but also by biological variations. SUV values might be reported to 5 or 6 decimal places, creating a false impression of high precision (Adams et al. 2010, Nahmias and Wahl 2008).

There are two main limitations of the current study. First, we investigated only the three most frequently applied discretization methods used for PET evaluations, therefore other approaches such as histogram equalization or Lloyd-max method could be involved, since these approaches have recently appeared in the literature (Vallières et al. 2015 and 2017, Lucia et al. 2018 and 2019). Secondly, to highlight the impact of discretization method on the numerical value of textural features we focused only on the group of GLCM, GLSZM and GLRLM based TIs without considering other features that have already been published. Finally, all images were delineated either by a single expert or using fixed thresholds methods segmentation before features calculation. The use of a semi-automatic application could reduce the inaccuracy of volumes segmentation.

3. Conclusion

In summary, we have highlighted the differences between three common quantisation methods on the stability of GLCM, GLSZM and GLRLM based tumour heterogeneity parameters in FDG-PET, in case of three different tumour delineation methods. We also showed that the calculated values of a given texture feature can vary over a wide range, selecting an appropriate SUV bin width or number of bin value in the discretization step. In order to preserve the

benefit of SUV for intra- and intersubject comparisons of tumour heterogeneity using [18F]FDG-PET, and to ensure the reliability of radiomics studies, future work should focus not only on the standardisation of all parameters involved in the eventual calculation (Image biomarker standardisation initiative, IBSI), but also on the harmonization of the discretization process as a latent step in the calculation of textural indices.

Acknowledgements

This work was supported by EFOP-3.6.3-VEKOP-16-2017-00009 co-financed (by EU and the European Social Found), and in addition by the Richter Gedeon Talentum Alapítvány (https://www.richter.hu).

References

[1] Adams M C, Turkington T G, Wilson J M, Wong T Z 2010 A systematic review of the factors affecting accuracy of SUV measurements Am. J. Roentgenol. 195 310–20

[2] Aerts H J W L, Velazquez E R, Leijenaar R T H, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen M M, Leemans C R, Dekker A, Quackenbush J, Gillies R J, Lambin P 2014 Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach Nat. Commun. 5:4006

[3] Anil K J 1989 Fundamentals of Digital Image Processing

[4] Bailly C, Bodet-Milin C, Couespel S, Necib H, Kraeber-Bodéré F, Ansquer C, Carlier T 2016 Revisiting the robustness of PET-based textural features in the context of multi-centric trials PLoS One 11 1–16

[5] Brooks F J, Grigsby P W 2014 The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake J. Nucl. Med. 55 37–42 Online: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.112.116715

[6] Buvat I, Orlhac F, Soussan M 2015 Tumor texture analysis in PET: Where do we stand? J. Nucl. Med. 56 1642–4 Online: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.115.163469

[7] Chicklore S, Goh V, Siddique M, Roy A, Marsden P K, Cook G J R 2013 Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis Eur. J. Nucl. Med. Mol. Imaging 40 133–40

[8] Cook G J , O'Brien M E, Siddique M, Chicklore S, Loi H Y, Sharma B, Punwani R, Bassett P, Goh V, Chua S 2015 Non-small cell lung cancer treated with erlotinib:heterogeneity of (18)F-FDG uptake at PET-association with treatment response and prognosis Radiology. 276 883-93

[9] Desseroit M C, Tixier F, Weber W A, Siegel B A, Cheze Le Rest C, Visvikis D, Hatt M 2017 Reliability of PET/CT Shape and heterogeneity features in functional and morphologic components of non-small cell lung cancer tumors: a repeatability analysis in a prospective multicenter cohort J Nucl Med. 58 406–411

[10] El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, Chaudhari S, Yang D, Schmitt M, Laforest R, Thorstad W, Deasy J O 2009 Exploring feature-based approaches in PET images for predicting cancer treatment outcomes Pattern Recognit. 42 1162–71

Online: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2701316&tool=pmcentrez&rendertype=abstract

[11] Forgacs A, Pall Jonsson H, Dahlbom M, Daver F, Difranco M D, Opposits G, Krizsan A K, Garai I, Czernin J, Varga J, Tron L, Balkay L 2016 A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images PLoS One 11 1–14

[12] Galloway M M 1975 Texture analysis using gray level run lengths Comput. Graph. Image Process. 4 172–9 Online: http://www.sciencedirect.com/science/article/pii/S0146664X75800086

[13] Gillies R J, Kinahan P E, Hricak H 2016 Radiomics: images are more than pictures, they are data Radiology 278 563–77 Online: http://pubs.rsna.org/doi/10.1148/radiol.2015151169

[14] Gómez W, Pereira W C A, Infantosi A F C 2012 Analysis of Co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound IEEE Trans. Med.

Imaging 31 1889-99

[15] Hallgren KA 2012 Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol. 8(1):23-34

[16] Haralick Robert M, Shanmugam K, Dinstein I 1973 Textural features for image classification IEEE Trans. Syst. Man, Cybern. SMC-3 6 610–21

[17] Hatt M, Lee JA, Schmidtlein CR, Naqa IE, Caldwell C, De Bernardi E, Lu W, Das S, Geets X, Gregoire V, Jeraj R, MacManus MP, Mawlawi OR, Nestle U, Pugachev AB, Schöder H, Shepherd T, Spezi E, Visvikis D, Zaidi H, Kirov AS 2017 Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211. Med Phys. 44(6):e1-e42. doi: 10.1002/mp.12124

[18] Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, Hindié E, Martineau A, Pradier O, Hustinx R, Perdrisot R, Guillevin R, El Naqa I, Visvikis D 2015 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(1):38-44. doi: 10.2967/jnumed.114.144055

[19] Hatt M, Tixier F, Pierce L, Kinahan P E, Cheze C, Rest L, Visvikis D 2017 Characterization of PET / CT images using texture analysis : the past , the present … any future ? Eur. J. Nucl. Med. Mol. Imaging 151–65 Online: http://dx.doi.org/10.1007/s00259-016-3427-0

[20] Hatt M, Tixier F, Visvikis D, Cheze Le Rest C 2017 Radiomics in PET/CT: More Than Meets the Eye? J Nucl Med. 58(3):365-366. doi: 10.2967/jnumed.116.184655

[21] Jobanputra R, Clausi D A 2004 Texture analysis using Gaussian weighted grey level co-occurrence probabilities Proceedings - 1st Canadian Conference on Computer and Robot Vision pp 51–7

[22] Koo T K, Li M Y 2016 A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr 15(2):155-63. doi:10.1016/j.jcm.2016.02.012.

[23] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout R G P M, Granton P, Zegers C M L, Gillies R, Boellard R, Dekker A and Aerts H J W L 2012 Radiomics: Extracting more information from medical images using advanced feature analysis Eur. J. Cancer 48 441–6

[24] Lasnon C, Majdoub M, Lavigne B, Do P 2016 F-FDG PET / CT heterogeneity quantification through textural features in the era of harmonisation programs : a focus on lung cancer Eur. J. Nucl. Med. Mol. Imaging Online: http://dx.doi.org/10.1007/s00259-016-3441-2

[25] Leijenaar R T H, Nalbantov G, Carvalho S, van Elmpt W J C, Troost E G C, Boellaard R, Aerts H J W, Gillies R J, Lambin P 2015 The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis Sci. Rep. 5 11075 Online: http://www.nature.com/articles/srep11075

[26] Lucia F, Visvikis D, Desseroit M C, Miranda O, Malhaire J P, Robin P, Pradier O, Hatt M, Schick U 2018 Prediction of outcome using pretreatment (18)F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy Eur J Nucl Med Mol Imaging. 45(5) 768-786. doi:10.1007/s00259-017-3898-7

[27] Lucia F, Visvikis D, Vallières M, Desseroit M C, Miranda O, Robin P, Bonaffini P A, Alfieri J, Masson I, Mervoyer A, Reinhold C, Pradier O, Hatt M, Schick U 2019 External validation of a combined PET and MRI radiomics model for prediction of

recurrence in cervical cancer patients treated with chemoradiotherapy Eur J Nucl Med Mol Imaging 46(4) 864-877. doi: 10.1007/s00259-018-4231-9

[28] Lucht R, Brix G, Lorenz W J 1996 Texture analysis of differently reconstructed PET images. Phys. Med. Biol. 41 2207–19 Online: https://www.ncbi.nlm.nih.gov/pubmed/8912391

[29] Mayr N A, Yuh W T C, Taoka T, Wang J Z, Wu D H, Montebello J F, Meeks S L, Paulino A C, Magnotta V A, Adli M, Sorosky J I, Knopp M V, Buatti J M 2006 Serial therapy-induced changes in tumor shape in cervical cancer and their impact on assessing tumor volume and treatment response Am. J. Roentgenol. 187 65–72

[30] Nahmias C, Wahl L M 2008 Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. J. Nucl. Med. 49 1804–8

[31] Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, Soussan M, Frouin F, Frouin V, Buvat I 2018 A post-reconstruction harmonization method for multicenter radiomic studies in PET J Nucl Med. pii: jnumed.117.199935

[32] Orlhac F, Nioche C 2017 Understanding changes in tumor texture indices in PET : J. Nucl. Med. 387–93 Online: https://www.ncbi.nlm.nih.gov/pubmed/27754906

[33] Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I 2015 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer PLoS One 10 1–16

[34] Orlhac F, Theze B, Soussan M, Boisgard R, Buvat I 2016 Multiscale texture analysis: from 18F-FDG PET images to histologic images J. Nucl. Med. 57 1823–8 Online: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.116.173708

[35] Pyka T, Gempt J, Hiob D, Ringel F, Schlegel J, Bette S, Wester H J, Meyer B, Förster S 2016 Textural analysis of pre-therapeutic [18F]-FET-PET and its correlation with tumor grade and patient survival in high-grade gliomas Eur. J. Nucl. Med. Mol. Imaging 43 133–41

[36] Shen W C, Chen S W, Liang J A, Hsieh T C, Yen K Y, Kao C H 2017 [18]Fluorodeoxyglucose positron emission tomography for the textural features of cervical cancer associated with lymph node metastasis and histological type. Eur. J. Nucl. Med. Mol. Imaging 44 1721-1731

[37] Tixier F, Hatt M, Le Rest C C, Le Pogam A, Corcos L, Visvikis D 2012 Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET J. Nucl. Med. 53 693–700 Online: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.111.099127

[38] Tixier F, Le Rest C C, Hatt M, Albarghach N, Pradier O, Metges J-P, Corcos L, Visvikis D 2011 Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer J. Nucl. Med. 52 369–78 Online: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.110.082404

[39] Tschirhart C E, Nagpurkar A, Whyne C M 2004 Effects of tumor location, shape and surface serration on burst fracture risk in the metastatic spine J. Biomech. 37 653–60

[40] Vallières M, Freeman C R, Skamene S R, El Naqa I 2015 A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities Phys Med Biol. 60(14) 5471-96. doi: 10.1088/0031-9155/60/14/5471

[41] Vallières M, Laberge S, Diamant A, El Naqa I 2017 Enhancement of multimodality texture-based prediction models via optimization of PET and MR image acquisition protocols: a proof of concept Phys Med Biol. 62(22) 8536-8565. doi:

10.1088/1361-6560/aa8a49

[42] van Baardwijk A, Bosmans G, van Suylen R J, van Kroonenburgh M, Hochstenbag M, Geskes G, Lambin P, De Ruysscher D 2008 Correlation of intra-tumour heterogeneity on 18F-FDG PET with pathologic features in non-small cell lung cancer: A feasibility study Radiother. Oncol. 87 55–8

[43] van Velden F H, Kramer G M, Frings V, Nissen I A, Mulder E R, de Langen A J, Hoekstra O S, Smit E F, Boellaard R 2016 Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: impact of reconstruction and delineation Mol Imaging Biol.18 788–95

[44] Yan J, Chu-Shern J L, Loi H Y, Khor L K, Sinha A K, Quek S T, Tham I W K, Townsend D 2015 Impact of image reconstruction settings on texture features in 18F-FDG PET J. Nucl. Med. 56 1667–73

Online:

http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.115.156927

[45] Zwanenburg A, Leger S, Valli`eres M, L¨ock S. Image biomarker standardisation initiative. arXiv preprint arXiv:1612.07003

[46] Matlab script 1.:

https://www.mathworks.com/matlabcentral/fileexchange/19058-cooc3d

[47] Matlab script 2.:

https://www.mathworks.com/matlabcentral/fileexchange/55587-ports-3d-image-texture-metric-calculation-package

[48] Matlab script 3.:

https://www.mathworks.com/matlabcentral/fileexchange/17482-gray-level-run-length-matrix-toolbox