Università degli Studi di Torino

Dipartimento di Fisica

Dottorato di ricerca in Fisica

XXXV Ciclo

Thesis Title:

“Radiomic features in positron emission tomography of patients with neuroendocrine tumors: impact of different acquisition and processing parameters”

Tesi presentata da: Dott. Osvaldo Rampado

Tutor: Prof. Vincenzo Monaco

Coordinatore del dottorato: Prof. Paolo Olivero

Anni accademici: 2019-20, 20-21, 21-22, 22-23

Settore scientifico-disciplinare di afferenza: FIS/07

Index

Abstract

The use of quantitative techniques for the extraction of information from medical images is increasing in recent years. Among these, the so-called radiomics has assumed particular importance. Radiomics consists in the calculation of a series of quantitative characteristics from portions of medical images for which it is possible to evaluate the correlation with variables of clinical interest (classifications, histological parameters, prognosis, response to therapy). The intent is therefore to transform the largely subjective elements that physicians use to extract information from the images into objective numerical data, as a base to formulate decision-making models.

To reach this goal radiomics uses quantitative parameters, named radiomic features, which are extracted from region of interest (ROI) of medical images. These features express various properties of the ROI, from simple statistical indexes of pixel distributions, to complex matrix calculations of the heterogeneity texture. Several software programs have been developed to extract hundreds of different features. Radiomics analysis is often implementable in retrospective studies on normal images already acquired in clinical routine for staging and monitoring of the disease. If the predictive value of specific radiomic features will be demonstrated in the future, their use can easily be extended to all diagnostic centers without additional costs and without increases in invasiveness for the patient.

There are different research fields in the development of radiomics approach. Several works are aiming at evaluating the relative impact of acquisition and processing parameters in the calculation of radiomic features on the values obtained and on the ability to discriminate the results of the clinical variables of interest. The medical physicist can deal with this type of investigation starting from his own skills in image formation, acquisition protocols, segmentation and post processing methods, that could affect the quality of the final image. The research work of this thesis also fits into this context, in particular in the field of nuclear medicine images for the diagnosis of neuroendocrine tumors (NET).

The main purpose of this work is the characterization of the robustness of features, extracted from patient and phantom images, with respect to different acquisition and pre-processing methods. The use of phantoms is particularly useful for being able to carry out acquisitions with different equipment in the same conditions and for evaluating the repeatability and reproducibility of the feature values. In the context of nuclear medicine, the development of phantoms presents a greater degree of complexity than in MRI or CT imaging, as the phantom must be filled with the radioactive substance with the correct timing taking into account the physical decay. All the experimental work was developed at the AOU Città della Salute e della Scienza in Turin.

In the first chapter of the thesis the basics of radiomics analysis are presented, including the radiomic workflow which comprises different steps from image acquisition to data analysis and potential model development. A particular focus is dedicated to the pre-processing techniques, which are the procedures usually applied before the feature extraction, and to the description of the categories of features used in the thesis. Nuclear medicine positron emission tomography (PET) is also introduced, in particular describing the quantitative content of images and discussing the main parameters involved in their formation. Finally, a short review of the literature of radiomics in NET tumours is presented, with examples of applications for various clinical outcomes.

Chapter 2 presents a study on the impact of different segmentation approaches (manual and semi-automatic) and preprocessing discretization in a cohort of NET patients. Different statistical indexes and data mining approaches, such as principal component analysis, have been used to investigate the feature robustness and potential predictive value.

The implementation and the use of a phantom with specific inserts produced by a 3D printing approach is described in chapter 3. In a first part the different strategies to develop phantoms for nuclear medicine radiomics and the relative manufacturing processes are described. The phantom was acquired with different radioisotopes and different parameters were used in the reconstruction of the images, and the relative variability in radiomics features was analyzed.

In chapter 4 a summary of lessons learned during the doctoral work and future perspectives in radiomics studies are presented.

1. Radiomics in nuclear medicine

1.1 1.1 What is radiomics

In recent years there has been an enormous diffusion of artificial intelligence techniques and big data analyses in many fields, including the medical one. In particular, a great deal of attention is paid to various methodologies for analyzing diagnostic images and, among them, radiomics is proving to be particularly promising for clinical applications.

Radiomics is defined as a quantitative approach to medical images, based on the extraction of a large number of quantities, called radiomic features (RF) or radiomics characteristics, to evaluate possible correlations with clinical variables of interest.

The basic idea of radiomics is that the images contain information reflecting biological processes which can be derived from the relationships between voxels (or pixels), intensity distributions and texture properties of the image. Typically, the ultimate goal is to correlate the results of the radiomic analysis with clinical data to develop diagnostic, prognostic or prediction models for the response to a therapy using artificial intelligence techniques and big data. A particular field of radiomics is the so-called "*delta radiomics*", or study of how the values of radiomic features vary over time (e.g. during a therapy) and how these changes are related to biological processes or ongoing physiological reactions.

1.2 Radiomics workflow

The workflow in a radiomics study is particularly complicated and presents many critical passages. Figure 1.1 depicts the typical procedures followed in a radiomic analysis.

Starting from the acquired medical image, a processing step is performed to improve the quality: for example, filters can be applied to reduce noise or artifacts.

Segmentation corresponds to the definition of a partial volume of the image, often indicated with ROI (Region of Interest), within which the radiomic features are calculated. For example, in the case of images of oncologic patients, the ROI is typically the volume of the tumor mass. In the case of medical images, segmentation can be performed manually by the medical staff, semi-automatically (a software helps the medical staff, but it is not able to segment independently) or automatically (the software acts completely independent).

Interpolation of images consists of resampling them to change the original resolution. This is an important step to standardize samples of images obtained with different machines or from different centers. The general indication is to provide cubic voxel images, while there is no consensus on the choice of a downsampling or an upsampling procedure: in the first case it is expected a loss of information, in the other there is the risk of introducing artifacts in the image. Moreover, currently there is no consensus on the choice of the optimal interpolation algorithms in various situations.

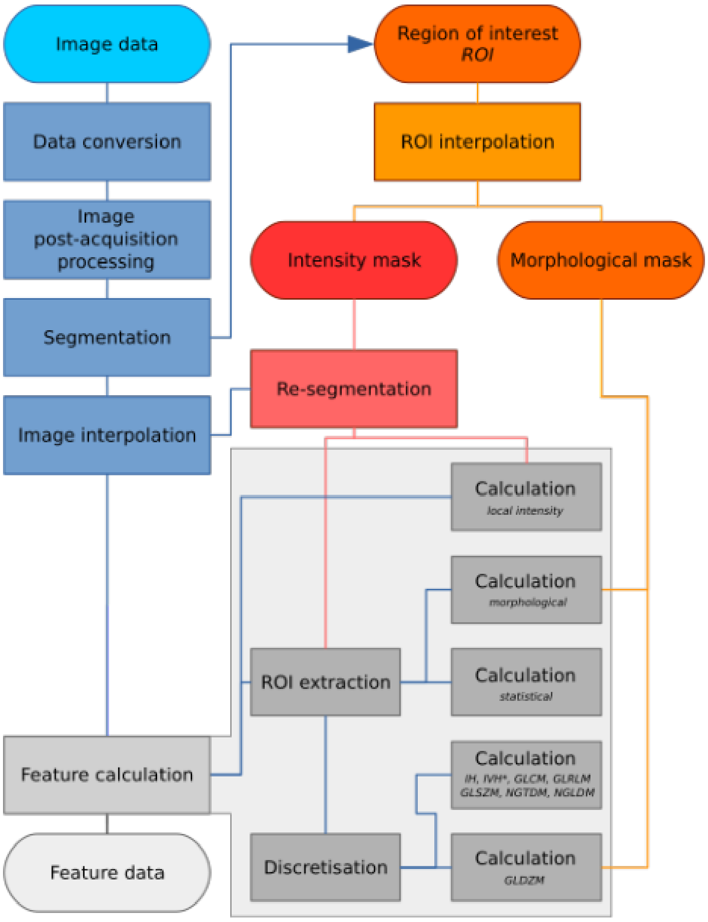


Figure 1.1: Example of workflow for radiomic feature extraction [IBSI 2019]

Starting from the ROI, a morphological mask and a mask of intensities are created. The first corresponds to the original ROI, while the second mask is the result of a process called “*re-segmentation*”. Re-segmentation corresponds to the elimination of voxels or image pixels whose intensity is outside a predefined range of values: for example, in computed tomography (CT) images voxels with values compatible with air or bone tissues are often eliminated from the ROI regions. In the pictures obtained from magnetic resonance the intensities are in arbitrary values and therefore typically no re-segmentation is performed.

The discretization of the gray levels consists in grouping them into classes (typically indicated as “bins”), allowing to create a histogram of the frequencies of gray levels. The intensities of the voxels are then changed by assigning the same value to all those that fall within the range corresponding to a certain class. It is a fundamental step for the calculation of the radiomic features related to the texture of the image and it can also be useful to eliminate at least part of the noise of the images.

There are two approaches for the discretization of the voxel values: fixed number of bins or bins with fixed width. In the first approach (fixed bin number, FBN) the same number of classes is used for all the images, and the voxel values are scaled according to the following formula:

$$I_{BN}(x) = \begin{cases} 1 \; if \; I(x) = \min\,(I(x)) \\ int\left\{BN * \frac{I(x) - \min\,(I(x))}{\max(I(x)) - \min\,(I(x))}\right\} otherwise \end{cases} \qquad (1)$$

where $I(x)$ is the intensity of voxel x, BN the number of bins, and $I_{BN}(x)$ a integer corresponding to the discretized gray-level of the voxel x. After this discretization approach, the maximum voxel value will be BN for all images, regardless of their previous maximum value. In this discretization method the relationship between the intensity levels of the image and their physiological meaning is lost, but a normalization effect is obtained which is useful in cases where the voxel values are expressed in arbitrary units, as in the case of magnetic resonance images.

The second approach consists in using the same width for the bins of all images (fixed bin size FBS). The value of the voxel $I_{BS}(x)$ is calculated as:

$$I_{BS}(x) = int\left\{\left[\frac{I(x)}{BS}\right] - min\left(\left[\frac{I(x)}{BS}\right]\right) + 1\right\} \qquad (2)$$

where BS is the bin size. In this discretization method the relationship between the intensities and their physiological significance is maintained. This is the method used for the discretization of Computed Tomography (CT) and often for Positron Emission Tomography (PET) images. In the case of CT images the intensities are expressed in Hounsfield Units (HU), which is a transformation of the coefficient of linear attenuation, while in the case of PET images the voxel level is expressed in SUV (Standardized Uptake Value), corresponding to the activity of the radiopharmaceutical in the voxel normalized to the total injected activity.

Figure 1.2 shows an example of discretization of voxel intensities of an image with different number of bins. Finally, after the interpolation and discretization of the image gray levels, the actual calculation of the radiomic features is performed on the voxels of the ROIs.

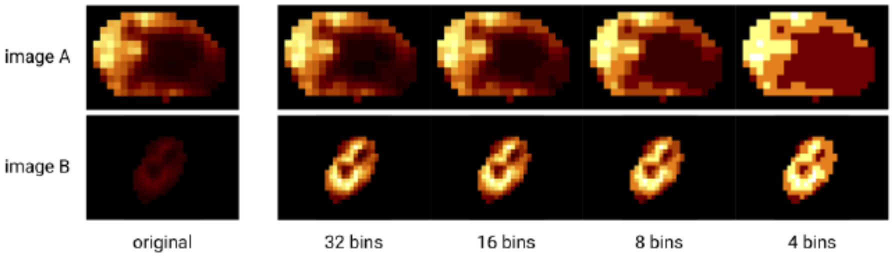


Figure 1.2: Example of discretization of the pixel intensities of an image with different number of bins [IBSI 2019]

1.3 Radiomics features

1.3.1 Types of radiomics features

The radiomic features are a series of quantities extracted from the voxel values of medical images. There are a large number of radiomic features, related to morphological properties, to the intensity distributions of the image pixels or voxels or to the properties of the image texture.

The features can be divided into seven groups:

- morphological features: linked to the shape of the ROI (for example volume, surface, sphericity, …);
- first order: calculated starting from the histogram of the frequency of intensities (e.g. mean, entropy, skewness, standard deviation);
- GLCM (Gray Level Co-occurrence Matrix): related to texture properties of the image, in particular to the distribution of gray levels in neighboring voxels (e.g. contrast, correlation, cluster tendency);
- GLSZM (Gray Level Size Zone Matrix): linked to texture properties of the image, in particular to the size of clusters of neighboring voxels with the same intensity (e.g. gray level variance, gray level non-uniformity);
- GLRLM (Gray Level Run Length Matrix): related to texture properties of the image, in particular to the length of sequences (called “*runs*”) of consecutive voxels along a given direction with the same intensity (e.g. gray level non-uniformity, run variance);
- NGTDM (Neighboring Gray Tone Difference Matrix): related to the difference between the intensity of a voxel with the average intensity of its neighbors (e.g. contrast, coarseness);
- GLDM (gray level dependence matrix): linked to the homogeneity of the image texture (e.g. gray level non-uniformity, large dependence emphasis).

1.3.2 Examples of first order radiomics features

The first order features are calculated, after the discretization of the gray levels, starting from the intensities of the voxels and from the histogram of the frequency of the intensities. We consider X defined as the set of intensity values of N_p voxels within the ROI, X(i) the value of the i-th voxel, P the histogram of the frequency of the discretized gray levels with N_g bins, P(i) the frequency of the voxels with values belonging to the i-th bin and $p(i) = P(i)/N_p$ the normalized probability for the same bin.

"*Entropy*" indicates the randomness of the intensity values of the image and measures the average amount of information to encode them. It is defined as:

$$entropy = -\sum_{i=1}^{N_g} p(i) \cdot log_2(p(i)) \qquad (3)$$

"*Skewness*" measures the asymmetry with respect to the mean of the intensity distribution of the image voxels:

$$skewness = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^3}{\sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2}^{\,3}} \qquad (4)$$

1.3.3 Examples of second order radiomics features

GLCM features are calculated from the gray level co-occurrence matrix and reflect texture properties of the image. The co-occurrence matrix of the gray levels, denoted by P, has dimension N_g × N_g (with N_g the number of gray levels, or bins, following the discretization). The element P(i,j) represents the number of times that an intensity voxel belonging to bin i and one to bin j are adjacent. For example, from an image represented by the matrix I we obtain the matrix of co-occurrence P.

$$I = \begin{bmatrix} 1 & 2 & 5 & 2 & 3 \\ 3 & 2 & 1 & 3 & 1 \\ 1 & 3 & 5 & 5 & 2 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 4 & 3 & 5 \end{bmatrix} \qquad P = \begin{bmatrix} 5 & 6 & 8 & 1 & 2 \\ 6 & 2 & 2 & 1 & 4 \\ 8 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 0 & 0 \\ 2 & 4 & 2 & 0 & 1 \end{bmatrix}$$

In this example, there are 5 pairs of adjacent pixels with value 1 in matrix I and so the value in the P matrix of the first row and first column is 5; there are 8 pairs of adjacent pixels with values 1 and 3 and as a consequence the value of the third column and first row (or third row and first column, the matrix is symmetric) is 8, and so on.

A series of other useful quantities can therefore be defined for the calculation of the features:

- p: the normalized co-occurrence matrix, with

$$p(i,j) = \frac{P(i,j)}{\sum P(i,j)} \tag{5}$$

- N_g: number of discretized gray levels
- the marginal probability of row i

$$p_x(i) = \sum_{j=1}^{N_g} p(i,j) \tag{6}$$

- the marginal probability of column j

$$p_y(j) = \sum_{i=1}^{N_g} p(i,j) \tag{7}$$

- the average intensity μ_x of p_x

$$\mu_x = \sum_{i=1}^{N_g} p_x(i)\, i \tag{8}$$

- the average intensity μ_y of p_y

$$\mu_y = \sum_{j=1}^{N_g} p_y(j)\, j \tag{9}$$

- σ_x: standard deviation of p_x
- σ_y: standard deviation of p_y

“*Contrast*” measures the local variation in intensity favoring values away from the diagonal: as the contrast increases, the difference in intensity between neighboring voxels increases.

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 p(i,j) \tag{10}$$

“*Correlation*” is defined as:

$$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij - \mu_x\mu_y}{\sigma_x(i)\sigma_y(j)} \tag{11}$$

The correlation values range from 0 (no correlation) to 1 (perfect correlation) and show the linear dependence of gray levels with their respective voxels in the GLCM.

“*Cluster tendency*” measures the tendency of the voxels to form groups with similar values of gray levels:

$$cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(i + j - \mu_x - \mu_y\right)^2 p(i,j) \tag{12}$$

"*Cluster shade*" is a measure of GLCM skewness and uniformity: the more it is large, the greater the asymmetry with respect to the mean.

$$cluster\ shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(i + j - \mu_x - \mu_y\right)^3 p(i,j) \qquad (13)$$

1.4 Robustness of radiomics features

1.4.1 Concept of robustness

The concept of robustness of radiomic features is of particular importance and is the basis of a large number of publications [Suter 2020,Granzier 2020,Zwanenburg 2019].

As seen previously, radiomics workflow involves a sequence of choices (how to apply different filters to the image, segmentation of the region of interest, the discretization of gray levels and image interpolation) that may have an influence on the values of the extracted features: in fact, different choices can modify the values of the voxel intensity or the number of voxels of the image that will be analyzed. Being the intensity of the voxels and the relationships between their positions at the basis of the calculation of the features, repeating the analysis under different conditions can lead to different feature values.

As there is currently no strong consensus on which is the optimal choice in each workflow step, to have reproducible results and reliable models it is necessary to investigate the effect of the different processing conditions on the values of the radiomic features.

It is therefore fundamental to assess the robustness, i.e. independence of feature values from the conditions under which they were calculated, such as discretization, resampling or the methodology used to define ROI (e.g different programs or different operators who have carried out the segmentation).

In addition to the independence of feature values from processing conditions, it is also of interest to evaluate whether the latter have an effect on the capability of the feature to discriminate patients who respond positively to a particular therapy from those that do not respond: in this case we speak of robustness if the features retain the ability to distinguish the two groups even as their values change. To evaluate the robustness, two statistical tests are typically used: the "*intraclass correlation coefficient*" (ICC) and the "*coefficient of variation*" (COV).

1.4.2 Intraclass correlation coefficient

The intraclass correlation coefficient (ICC) is widely used in radiomics studies to evaluate the robustness of features with respect to parameter variations [Koo 2016].

The ICC has a value between 0 and 1 and is used to measure the reliability of an experimental method or the replicability of a measurement. The ICC allows to evaluate the correlation and agreement of the measurements and for this reason it is preferred in the radiomics field to the Student's t-test (which measures only the agreement) and the Pearson correlation coefficient (which measures only the correlation).

In the case of this study it is used to verify the effect of the different conditions (ROI segmentation, discretization of the gray levels, image acquisition and processing parameters) on the values of the radiomic features. There are different definitions of ICC to choose from, on the basis of the specifics of the study in which to apply it. In our case the form of ICC "*two-way random-effects model, single rater, consistency*" was used. Two-way random-effects model means that we want to generalize the conclusions obtained in a limited set of conditions, while single rater indicates the fact that each value of the radiomic features has been calculated only once (and is not the result of an average). The choice of consistency is particularly important: it is made when the interest is not in the absolute concordance of the values of the features extracted under different conditions, but in the maintenance of the correlation between the values. In this way we want to verify that, even when the value of a feature varies in different conditions, it maintains for example its ability to discriminate patients responding to therapy from those unresponsive.

To calculate the ICC in the form chosen for this study, we consider the matrix X whose elements x(i, j) are the values of a feature calculated for the i-th patient in the j-th condition. Let N be the number of patients and K the number of conditions. We define

$$S_i = \frac{1}{K}\sum_{j=1}^{K} x(i,j) \tag{14}$$

the mean of the values measured for patient i for all the conditions, and

$$M_j = \frac{1}{N}\sum_{i=1}^{N} x(i,j) \tag{15}$$

the mean of the values measured in condition j across all patients.

From these two terms we can derive the following quantities:

$$\bar{x} = \frac{1}{NK}\sum_{i=1}^{N}\sum_{j}^{K} x(i,j) \tag{16}$$

$$SST(sum\ of\ square\ total) = \sum_{i=1}^{N}\sum_{j=1}^{K}(x(i,j)-\bar{x})^2 \tag{17}$$

$$SSBS(sum\ of\ square\ between\ subjects) = \sum_{i=1}^{N}\sum_{j=1}^{K}(S_i-\bar{x})^2 \tag{18}$$

$$SSBM(sum\ of\ square\ between\ measures) = \sum_{i=1}^{N}\sum_{j=1}^{K}(M_j-\bar{x})^2 \tag{19}$$

$$SSE(sum\ of\ square\ errors) = SST - SSBS - SSBM \tag{20}$$

$$MSBS(mean\ square\ between\ subjects) = \frac{SSBS}{K-1} \tag{21}$$

$$MSE(mean\ square\ error) = \frac{SSE}{(N-1)(K-1)} \tag{22}$$

And finally the ICC is defined as:

$$ICC = \frac{MSBS-MSE}{MSBS+(N-1)MSE} \tag{23}$$

The correspondance between ICC values and the "robustness level" varies in different studies, but generally a value of ICC greater than 0.9 is associated to an excellent robustness and a value between 0.75 and 0.9 to a good robustness.

1.4.3 Use of COV in radiomics robustness studies

The coefficient of variation (COV) or relative standard deviation is widely used in radiomics studies to evaluate the robustness of a feature. In particular, it is used to quantify the change in the value of a feature when calculated under different conditions.

In general, it is defined as the ratio of the standard deviation of the values of a feature extracted under different conditions over its mean value and it is an indication of the dispersion of the data. It is typically indicated in percentage terms.

In this study, for each feature, a coefficient of variation COV_i for the patient with index i is defined as:

$$COV_i = 100 * \frac{\sqrt{\frac{1}{K-1}\sum_{j=1}^{K}(x(i,j)-\overline{x(\imath)})^2}}{\overline{x(\imath)}} \tag{24}$$

where x(i, j) is the feature value calculated for patient i for a choice j of a parameter among K possibilities (for example the different discretizations of the levels of gray or the different resampling of the image), $\overline{x(\imath)}$ is the average value of the feature calculated in the different conditions for the patient i.

To have a single COV value for each radiomic feature, the mean value calculated on N patients is used:

$$COV = \frac{1}{N}\sum_{i=1}^{N} COV_i \tag{25}$$

Unlike the ICC, the COV provides an indication of the absolute agreement between the feature values calculated in different situations. There is no univocal interpretation of the value of the coefficient of variation, but in the literature a value of COV below 10% is often considered as optimal robustness.

1.5 The needs of standardization in radiomics analysis

Radiomics at the moment has limits that do not yet allow the passage from the research phase to the clinical application. In particular, we have not yet arrived at a consensus on methodologies and workflow choices that are optimal for different applications. Furthermore, limitations regarding reproducibility are generally observed in validation studies. The lack of a standard procedure is particularly significant because the values of radiomic features are influenced by workflow choices: the use of different filters during image processing or different gray level discretizations or of different resolutions in the interpolation phase can have an impact on the values of the radiomic features and, consequently, on the model to be developed.

For this reason, initiatives such as [IBSI 2019] (image biomarker standardization initiative) or RQS (radiomics quality score) [Lambin 2017] are fundamental to improve the standardization of procedural aspects in radiomics studies. IBSI is a collaboration aimed at standardizing the extraction procedures of radiomic features: it takes care of standardizing feature definitions, workflow for the extraction, tools to validate the radiomics software and how to report the results. RQS, on the other hand, is a score (with a maximum value of 36) that can be assigned to studies of radiomics to evaluate their quality by assigning points to various workflow steps: for example, points are awarded if the image acquisition protocol is described, if a study was also performed using a phantom, if multiple ROIs were used (for example obtained with different software) or if the training of the model was done with a data sample independent of the validation sample.

1.6 General characteristics of PET images

Positron emission tomography (PET) is increasingly used for the diagnosis, staging and follow-up of various malignancies. PET is based on the detection of annihilation photons released when radionuclides injected into patients, such as F-18, carbon-11 and Ga-68, emit positrons which undergo electron annihilation. Two photons are produced in opposite directions each with an energy of 511 keV (corresponding to the electron mass energy) and are detected in coincidence as they are absorbed by scintillation crystals consisting of germinated bismuth (BGO), lutetium oxyorthosilicate (LSO), or gadolinium silicate (GSO). A PET image reconstruction is based on the fact that each annihilation occurs in a point along the line connecting the detection points of a coincident photon pair and the annihilation activity in each voxel will be proportional to the radiotracer uptake.

FDG is a radiopharmaceutical analogue of glucose which is absorbed metabolically by active tumor cells using facilitated transport similar to that used by glucose. The uptake rate of FDG by tumor cells is proportional to their metabolic activity. Positrons emitted by a radionuclide have enough kinetic energy to travel a small distance (called positron mean range) before annihilating with an electron. The distance is less in dense structures such as bone but greater for lungs and air; for example, for FDG in water, the average positron gap is 1.4 mm. This change in position between the origin of the positron and its site of annihilation causes a blurring in the PET image, limiting the spatial resolution of the PET (typically 5 mm on current scanners). Another source of decrease in spatial resolution is non-linearity. When the positrons and electrons annihilate each other, they are in motion and the actual angle between the two emitted photons is not exactly 180°. This effect introduces a random variation with a rms of $0{,}5^{o}$ with respect to the 180^{o} angle assumed in the PET reconstruction, causing a further degradation in spatial resolution also called "*annihilation angle blur*". The effect of the image increases as the distance between the two detectors detecting coincident events increases; it can therefore be decreased by reducing the diameter of the scanner.

Many annihilation photons are lost because of their absorption in the body. The attenuation is lower in the body surface than deep in the body, low in lungs, and especially high in dense tissues such as bones, leading to severe artifacts in PET image, unless it is corrected during image reconstruction. For this reason, CT images are acquired together with PET images and CT voxel values are used to correct for attenuation effects on the annihilating photons.

There are several methods to assess differences in radiotracer uptake by normal and pathological tissues, such as visual inspection, "*standardized uptake value*" (SUV), and glucose metabolic rate. In most cases visual inspection is used in the analysis of PET-CT results by comparing PET and CT data, as well as viewing fused PET-CT images. SUVs are used for semiquantification of FDG uptake. Another method for quantifying dynamic PET results is the more complex calculation of glucose metabolic rate.

The SUV of a local region of interest is calculated according to the following formula:

$$SUV = \frac{detected\ activity\ concentration}{\frac{injected\ activity}{body\ mass}} \qquad (26)$$

This parameter expresses the ratio between the concentration of radiopharmaceutical accumulated in a target lesion (in kBq/g) and the average concentration of tracer in the patient body (in MBq/kg). The SUV of a tissue is usually represented as its minimum, maximum or average value in the region of interest. The average SUV is the mathematical mean of all pixels in the region of interest, while the minimum and maximum SUV are the values of the pixelz with the lowest and highest SUV values, respectively. Typically, malignant tumors have SUVs greater than 2.5-3.0, while normal tissues such as liver, lung, and marrow have SUVs between 0.5 and 2.5. Variations of the SUV definition have been also introduced, such as the SUV corrected for glucose and the SUV normalized for the surface area or the lean body mass. It is useful to know the SUV of the tumor before the initiation of a therapy to evaluate the grade of the tumor and to evaluate the treatment response after radiotherapy or chemotherapy [Thompson 2002]. It is important to standardize the time interval between radiotracer injection and the PET study because SUV variability over time has been well documented [Thie 2004].

The use of SUVs as a measure of relative tissue/organ uptake facilitates comparisons between patients and has been suggested as a basis for diagnosis. However, the practice of using SUV thresholds for diagnosis is not widely accepted There are at least two general reasons for the inconsistent use of SUVs in practice. The first is that accurate diagnostic and staging information need not depend on accurate image quantification, as the related image content (i.e., image appearance) is often sufficient for such purposes. Second, the measured SUVs exhibit a large degree of variability due to physical and biological sources of error, as well as inconsistent and non-optimized image acquisition, processing, and analysis. More specifically, the use of SUV thresholds (e.g. SUV > 2.5), to characterize a nodule or mass as benign or malignant, has been repeatedly shown to be often invalid. Thus, many benign infectious/inflammatory processes will have substantial FDG uptake with high SUV values, and conversely, many sluggish or slow-growing malignant processes may have minimal uptake and low SUV values. That's not to say, however, that using SUV thresholds for diagnosis is of no value. In circumstances where a nodule or tissue mass has no greater uptake than adjacent reference tissue and the pretest probability of malignancy is low, a decision to develop a "watch and wait" management strategy can often be made. with confidence. In this situation, the very low false negative rate of negligible FDG uptake can help in the decision to avoid unnecessary invasive procedures for tissue diagnosis. This has often been referred to as using FDG-PET as a "molecular imaging probe". Thus, PET/CT with FDG may aid in the decision to avoid an unnecessary invasive tissue biopsy, as well as guide such a procedure to a tissue location where a valid diagnostic biopsy specimen can be obtained.

1.7 Guidelines and recommendations in nuclear medicine radiomics

In November 2022 the Society of Nuclear Medicine and Molecular Imaging (SNMMI) and the European Association of Nuclear Medicine (EANM) have published a guideline on radiomics in nuclear medicine, in order to provide comprehensive best practice information for robust radiomics analyses, including study design, quality, data collection, the impact of acquisition and reconstruction, discovery and segmentation, standardization and implementation of features, as well as appropriate modeling schemes and assessments [Hatt 2023]. The list of recommendations is summarized in table 1 of the guidelines and here reported in figure 1.3.

It is in particular of interest to analyze more deeply the recommendations about three aspects investigated in this phd research: the consequences of different acquisition parameters, segmentation methods and preprocessing (resampling and voxel contents discretization).

In the guidelines it is highlighted that it not possible at present to recommend clinical acquisition/reconstruction settings optimized specifically for the purpose of radiomics studies. They therefore recommend relying on the current harmonization guidelines for PET/CT imaging which were developed to make PET imaging as reliable and reproducible as possible between centres, as this can definitely improve the robustness and reproducibility of the derived radiomic characteristics [Pfaehler 2020]. Future harmonization guidelines are expected to pay more attention to radiomics applications suggesting parameters and acquisition settings to minimize the variability of resulting radiomic features [Papp 2019]. It is well known that the variability with respect to different acquisition and reconstruction factors (including but not limited to scanner model and/or generation, absorption time, scan duration, reconstruction algorithm and parameters, post-filtering settings) can influence the values of radiomic features in PET, as reported in several studies [Galavis 2010, Yan 2015], although the resulting impact on their clinical relevance and power of differentiation may not necessarily be strongly influenced [Tankyevych 2021, Ly 2018]. However, these studies have also highlighted the fact that the sensitivity of radiomic features can vary widely, with some studies showing greater robustness to various factors than others.

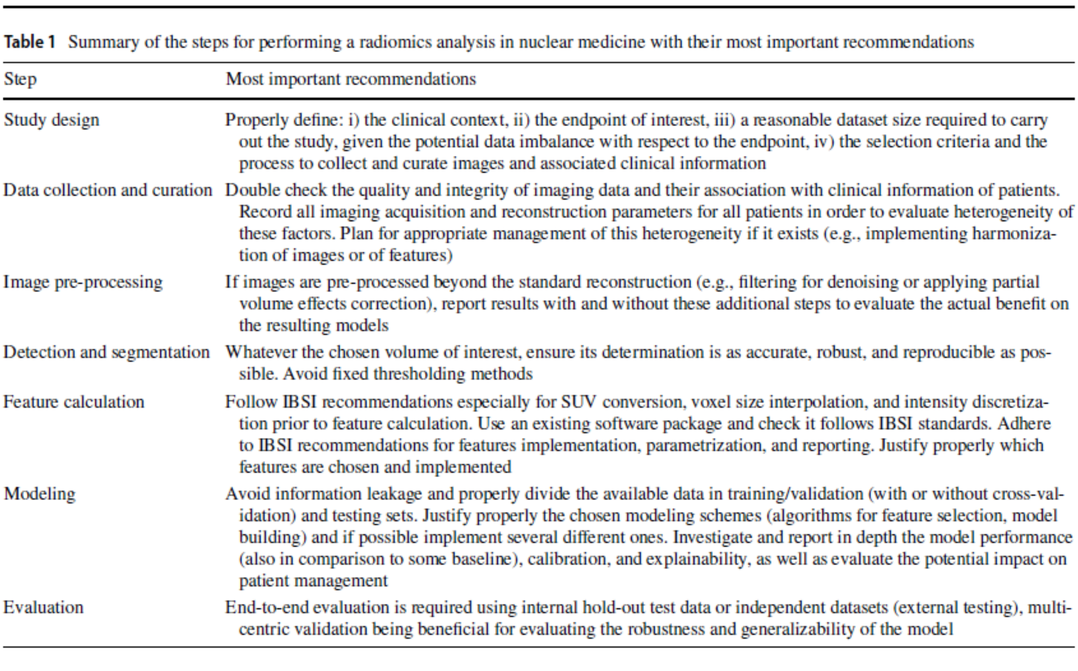


Table 1 Summary of the steps for performing a radiomics analysis in nuclear medicine with their most important recommendations

Step	Most important recommendations
Study design	Properly define: i) the clinical context, ii) the endpoint of interest, iii) a reasonable dataset size required to carry out the study, given the potential data imbalance with respect to the endpoint, iv) the selection criteria and the process to collect and curate images and associated clinical information
Data collection and curation	Double check the quality and integrity of imaging data and their association with clinical information of patients. Record all imaging acquisition and reconstruction parameters for all patients in order to evaluate heterogeneity of these factors. Plan for appropriate management of this heterogeneity if it exists (e.g., implementing harmonization of images or of features)
Image pre-processing	If images are pre-processed beyond the standard reconstruction (e.g., filtering for denoising or applying partial volume effects correction), report results with and without these additional steps to evaluate the actual benefit on the resulting models
Detection and segmentation	Whatever the chosen volume of interest, ensure its determination is as accurate, robust, and reproducible as possible. Avoid fixed thresholding methods
Feature calculation	Follow IBSI recommendations especially for SUV conversion, voxel size interpolation, and intensity discretization prior to feature calculation. Use an existing software package and check it follows IBSI standards. Adhere to IBSI recommendations for features implementation, parametrization, and reporting. Justify properly which features are chosen and implemented
Modeling	Avoid information leakage and properly divide the available data in training/validation (with or without cross-validation) and testing sets. Justify properly the chosen modeling schemes (algorithms for feature selection, model building) and if possible implement several different ones. Investigate and report in depth the model performance (also in comparison to some baseline), calibration, and explainability, as well as evaluate the potential impact on patient management
Evaluation	End-to-end evaluation is required using internal hold-out test data or independent datasets (external testing), multicentric validation being beneficial for evaluating the robustness and generalizability of the model

Figure 1.3 Summary of the steps for performing a radiomics analysis in nuclear medicine with their most important recommendations (from [Hatt 2023])

Based on existing findings and previously published recommendations, including AAPM Working Group Report 211 [Hatt 2017] and MICCAI PETseg Challenge [Hatt

2018], for the segmentation methods in radiomics studies, the main recommendations are the following:

i) Methods that favor positive predictive value over sensitivity in segmentation performance should be preferred. For example, including parts of the absorption at the edges is likely to introduce greater bias in the resulting measurements of characteristics (particularly absorption medium but also specific structural characteristics) due to partial volume effects (if no compensation/correction has been applied to the images before) [Hatt 2013, Hatt 2017];

ii) Methods based on a fixed threshold (for example, 40 or 50% of the maximum SUV) have the advantage of being fairly reproducible by multiple readers. However, they should not be used without heavy adjustments or expert corrections for the purpose of radiomics studies, as they have been shown to perform poorly especially in heterogeneous lesions [Hatt 2017].

iii) One should rely on (semi) automated methods, rather than manual delineation. Ideally, a consensus between several methods should be considered to improve the performance. If no automated algorithm but only manual delineation is available, ideally a consensus of at least three expert delineations should be obtained, for example using approaches such as the Simultaneous Performance Level and Performance Estimation Technique (STAPLE) [Warfield 2004]. However, this would likely limit the analysis to small datasets. Alternatively, if only manual delineation by a single observer is possible for the entire study, appropriate study design should include a patient subgroup analysis to investigate the potential impact of between-user variability on patient performance. For example, a model trained on patients delineated by one expert could be applied to test patients for which delineation was performed by another expert.

iv) There is growing interest in using deep learning approaches to segmentation. Current state-of-the-art methods to achieve fully automated PET image detection and segmentation are almost all based on DNN like U-Net architecture. The generalizability and performance of these algorithms are however questionable as they may fail in new, previously unseen cases. Fully automated segmentation should therefore always keep human experts in the loop for quality assurance.

Regarding the two discretization methods described before (FBN and FBS), FBS in PET has shown to produce features with a lower correlation with the corresponding number of voxels involved in the calculation (i.e., tumor volume) than FBN [Leijenaar 2015]. However, since FBS instead introduces a spurious correlation with SUV [Orhlac 2015], there is no consensus on the superiority of one over the other method in terms of modeling performance. For comparison purposes, it may be useful to systematically implement and report both.

For the FBN binning method, the recommended number of bins should be between 4 and 64 bins. A higher number of bins typically leads to very sparse and uninformative texture matrices [Hatt 2015]. For the FBS discretization method, the lower bound should generally be set to 0.0 SUV. The recommended container size depends on the problem. Typical bin sizes lead to a ROI with a number of bins between 8 and 64. Too small bin sizes, such as 0.01 SUV, should be avoided, as this will lead to very sparse and uninformative texture matrices. Typically, texture matrices should be calculated with the default parameters listed in the IBSI reference document. It is also necessary to be careful with the software, as different programs may not use the same default settings.

2. Impact of segmentation and discretization on radiomic features in PET/CT images of neuroendocrine tumor patients

2.1 Neuroendocrine tumor imaging and purpose of the study

Neuroendocrine tumors (NET) are a heterogeneous group of malignancies, characterized by different subtypes and several possible primary locations [Oronsky 2017]. The subtypes are classified on the base of some histological indexes derived by the analysis of portions of biological tissues extracted with ordinary biopsy. As an example the histological index Ki-67 is usually considered to define the NET tumor grading. The Ki67 index represents the fraction of Ki-67 antigen-positive tumor cells expressed as a percentage, where this antigen is a nuclear protein closely associated with cell proliferation. However, according to recent studies the assessment of tumor aggressiveness by only lesion biopsy is subject to potential grading underestimation [Gerlinger 2019]. It is important to consider the heterogeneity of each NET case, in terms of spatial distribution (inter- and intra-tumoral heterogeneity) and also in terms of temporal evolution (in particular how the more aggressive cells clone during the observation period and follow up). Moreover, NET often appear as multiple-lesions and in this case biopsy sampling for each single lesion is obviously not feasible, with a consequent absence of complete characterization of the primary and the secondary multiple-lesions [Grillo 2016].

Medical imaging and in particular PET imaging might offer a contribution in the characterization of tumor heterogeneity, as already outlined in the first chapter. The importance of PET imaging is also related to the role of somatostatin (SST), a small peptide that exerts inhibitory effects on a wide range of neuroendocrine cells. Due to the fact that somatostatin regulates cell growth and hormone secretion, somatostatin receptors (SSTRs) have become valuable targets for the treatment of different types of NETs. At present, PET imaging with ^{68}Ga-DOTA-peptides analogue to the somatostatin is considered the state of the art to quantify SSTR in vivo [Giovannini 2018], while 18F-fluorodeoxyglucose (18F-FDG) PET-CT is used to metabolically characterize more aggressive and higher-grade NET lesions [Carideo 2019]. This dual approach has been recently evaluated leading to the development of a nuclear medicine “score”, the NETPET [Chan 2017]. The NETPET scoring scheme was devised based on quantitative and qualitative assessments made by a number of nuclear medicine physicians and medical oncologists. The strategy adopted was to identify the single lesion that was more avid for FDG than its SSTR-emulating radiopharmaceutical uptake (specifically [68Ga]-DOTATATE), as this is likely to represent the more aggressive disease phenotype present in the subject. The NETPET grade uses a categorical scale from 0 to 5 and was based largely on the characteristics of the individual initial lesion, with a grade of P1 indicating purely SSTR-avid disease with no FDG uptake in any lesion and P5 indicating the presence of FDG with negative SSTR, considered a very poor prognostic marker.

Nevertheless, the simple in vivo quantification of receptor expression is not sufficient to fully characterize the biology of the tumor and the intra patients and intratumor heterogeneity. This drawback might be solved with a better characterization of tumor heterogeneity by the extraction of radiomic features (RFs), as a surrogate biomarker for NET lesions characterization, from the ^{68}Ga-DOTA-peptide PET-CT. As stated in chapter 1, scientific interest in radiomics applied to PET imaging is rapidly increasing, but the methodological approach needs to be validated and standardized and, thus, harmonization among protocols is needed. In particular, the approach used for lesion segmentation and the voxel value discretization for second order features are two procedural parameters that may affect significantly the final results.

Most of the studies about RFs robustness are focused on 18F-FDG PET/CT imaging [Orhlac 2016], whereas at present there is only one study evaluating the impact on RFs of different image acquisition and reconstruction parameters for 68Ga-DOTA-peptides PET/CT, which howevever does not consider the effect of different segmentation approaches [Bailly 2016]. There are several reasons to evaluate the RFs robustness specifically in 68Ga-DOTA-peptide tracers: a diverse range of positrons compromising the resolution in PET in a different way comparing to 18F-FDG; a different physiological distribution of 68Ga-DOTA-peptide; an high inter-patient and intra-patient heterogeneity for both physiological and pathological uptake (the SUV maximum and relative range) compared to 18F-FDG, leading to the need of providing different segmentation methods and discretization settings. For all these reasons the results obtained with 18F-FDG cannot be directly transposed to 68Ga-DOTA-peptide and specific additional investigation must be performed for this radiotracer.

RFs could also be used to evaluate the response to Peptide receptor radionuclide therapy (PRRT), that is an effective treatment for gastroenteropancreatic (GEP) NET [Strosberg 2017]. PRRT is a treatment that targets some types of neuroendocrine tumors, binding to specific receptors on the tumor cells and destroying them with a small, but powerful dose of radioactivity. It is not a cure, but PRRT can effectively slow or stop tumor growth. This helps to improve the length and quality of life for people with neuroendocrine tumors. PRRT belong to the family of “theranostic”, a term derived from a combination of the words therapeutics and diagnostics. In this emerging field of medicine, drugs and/or techniques are uniquely combined to simultaneously or sequentially diagnose and treat medical conditions. Based on the results of a recent trial [Oberg 2017], 177Lu-DOTATATE was recently approved in Europe for the treatment of inoperable or metastatic GEP-NET in progression disease. Usually the response assessment to PRRT is based on imaging monitoring of the disease, even if the Delphic consensus assessment for GEP-NET [Bozkurt 2016] consider the current strategies suboptimal due to limitations of parameters evaluated on both anatomic (lesion dimensions and measurements evaluation) and functional imaging (standardize uptake value (SUV) interpretation due to the heterogeneity in somatostatin receptor expression) and also due to the heterogeneity of NET in histology, localization and behavior.

The main objective of this study is to evaluate the robustness of RFs as a function of segmentation methods and discretization settings in 68Ga-DOTATOC PET/CT images.

2.2 Material and methods

2.2.1 Patients selection

This study was performed in collaboration with nuclear physicians, who individuates 270 examinations of patients affected by NET that performed a ^{68}Ga-DOTA-TOC PET/CT in the hospital A.O.U. Città della Salute e della Scienza between February 2017 and July 2019. All the examinations were reviewed in order to select cases suitable for this radiomics studies. The inclusion criteria were the following: 1) histologically proven NET; 2) ^{68}Ga-DOTA-TOC PET/CT were acquired for staging in treatment-naïve patients or restaging after surgery; 3) willing to sign an informed consent form (ICF). Exclusion criteria were 1) age < 18 years; 2) previous systemic therapies (e.g. somatostatin analogues, chemotherapy, everolimus, peptide receptor radionuclide therapy, etc.). 49 patients with a total of 60 lesions matched these inclusion criteria and were considered in this analysis. Primary tumor sites were GEP-NET, lung NET and others NET in 77.5% (38/49), 18.4% (9/49) and 4.1% (2/49) of cases, respectively.

For two patients with a similar background disease and a different outcome of PRRT, RF distributions on a total of 18 lesions were evaluated with different intensity discretization. Considering the novelty of this type of therapy, it is very difficult to obtain data from multiple patients from a single center, but we decided to investigate the RF distributions of the lesions of these two patients with different therapeutic outcomes with the aim of analyzing the absolute RF intervals in these situations and the presence of a difference between the two subjects warranting further investigation of potential predictive value. Both patients went under surgery of the primitive tumor and started the therapy with somatostatin analogue (lanreotide). Both patients developed a metastatic disease to the liver, treated with several radiofrequence ablations (RF). The two patients were treated with a first line chemotherapy after progression, evaluated on the biopsy of a liver metastasis (Ki67 25% for patient A and Ki67 22% for patient B), and with everolimus after an extent of progression evaluated with imaging. Despite increasing grading, due to a further disease progression and after a carefully evaluation, both patients were treated with six cycles of PRRT therapy. The NETPET grade [8] before PRRT treatment was 2a for patient A and 1 for patient B and PRRT response was assessed according to RECIST 1.1 criteria, that is a standard way to measure how well a cancer patient responds to treatment. It is based on whether tumors shrink, stay the same, or get bigger. During the follow-up, Patient A died for complications related to the underlying disease 13 months after PRRT, while Patient B responded positively to the therapy.

2.2.2 PET/CT acquisition and image reconstruction

All patients underwent PET/CT on an analog 3-dimensional (3D) PET scanner (Philips Gemini Dual-slice EXP scanner – PET AllegroTM system with Brilliance CT scanner – Philips Medical Systems, Cleveland, OH). In accordance with the procedure guidelines for PET imaging [Bozkurt 2017], the injected tracer average activity was 145.1 MBq of 68Ga-DOTA-TOC with a relative standard deviation of 25.3 MBq (range: minimum 100 Mbq, maximum 212 MBq). After 60 minutes of uptake and following free-breathing CT acquisition for attenuation correction from the vertex of the skull to the mid-thighs (5 mm slice, 40 mAs and 120 kVp), PET data were acquired in 3-dimensional (3D) mode, covering the same anatomical region of the CT, with 2.5 min per bed position and 6-8 bed positions per patient. The PET scans were reconstructed by ordered subset expectation maximization (OSEM) algorithm (3D-RAMLA) [Ortuno 2006], with the following settings: 4 iterations, 8 subsets and field of view (FOV) of 576 mm. For all reconstructions, matrix size was 144×144 voxels, resulting in isotropic voxels of 4.0×4.0×4.0 mm^3. All acquisitions were corrected for photon attenuation (using the corresponding CT image), as well as for scatter and random coincidences.

2.2.3 Lesion segmentation

The lesion volumes of interest (VOI) is defined using two different approaches: manual and semiautomatic segmentation. For both methods the application of a threshold of voxel values to redefine the VOIs was also investigated, including in the VOIs only the voxel above a specified percentage of the SUV max.

In the manual method, for each lesion, a three-dimensional VOI was manually delineated (VOI_m), slice-by-slice, in the OSEM PET images by four independent nuclear medicine physicians (FC, VL, GP and BD with 10, 7, 5 and 3 years of expertise respectively), by using the software LIFEx v. 4.81 (IMIV/CEA, Orsay, France - www.lifexsoft.org) [Nioche 2018].

Each lesion was also contoured using a semi-automatic edge-based (SAEB) algorithm (VOISAEB), homemade implemented in MATLAB (MathWorks) code, based on the active contour model proposed by Chan and Vese [Chan 2001]. The algorithm is semi-automatic since, the operator intervention is required in order to insert the central point of the lesion (Figure 2.1).

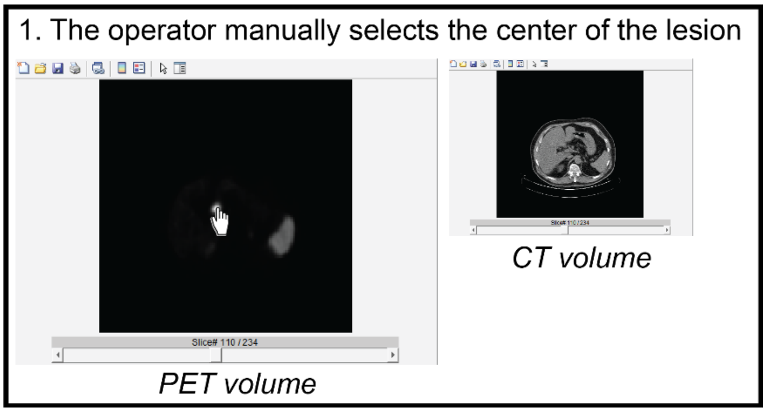


Fig. 2.1 First step of semiautomatic segmentation

The developed MATLAB graphical user interface allowed the operator to view both the PET and the CT images separately. Edge enhancement filters were applied to emphasize the edges of the lesion (Figure 2.2) and, subsequently, a curve was evolved iteratively on both the original and the edge-enhanced image in order to match the lesion contours by using a level-set formulation (Figure 2.3). The iteration 0 of the level-set, which is the initialization, was the center of the lesion indicated by the operator. The final contour of the lesion ($\mathrm{VOI}_{\mathrm{SAEB}}$) was achieved at the end of the iterative level-set.

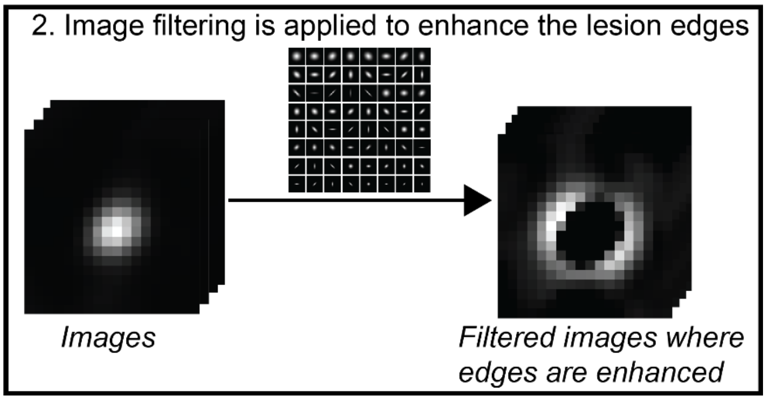


Fig. 2.2 Second step of semiautomatic segmentation

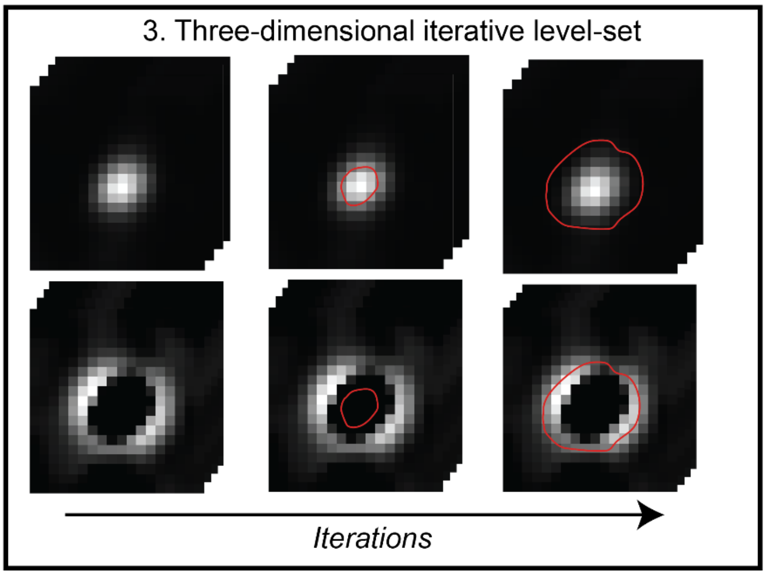


Fig. 2.3 Third step of semiautomatic segmentation

The threshold-based segmentation approach was implemented applying three different thresholds on both manual VOI_m and VOI_{SAEB}, defined as 20, 30 and 40% of the SUV_{max} (VOI_{20}, VOI_{30} and VOI_{40}, respectively). This approach could represent another step towards standardization, but it is crucial to define the proper threshold in order to include all the voxel representing the tumor heterogeneity and at the same time to discard surrounding physiological uptake, especially for the segmentation of liver metastases which are very prevalent in metastatic neuroendocrine tumors.

2.2.4 Intensity discretization

To perform RFs calculation, in particular of textural features, voxels values were redefined considering a limited numbers of SUV intensity values (grey-level intensity discretization process). In order to investigate the effect of this preprocessing step, the analysis was performed with two different settings of intensity rescaling (as already defined in paragraph 1.2 of the first chapter), according to the IBSI nomenclature:

- absolute intensity rescaling factors with 64 number of grey levels (bins) between different ranges of SUV units: 0 and 60 SUV (abbreviated AR60) in most cases (“fixed bin size” equal to 0.95), and also AR80 and AR100, the last in particular for the lesions of the two patients with different therapy outcome;
- relative intensity rescale factor (min-max of the SUV), using a “fixed bin number” equal to 64 number of grey levels and different size of bin, according to the uptake characteristic of each lesion/VOI (relative resampling (RR)).

Number of grey levels was set to 64, that is the typical value adopted in most of the PET radiomics studies [Leijenaar 2015, Orlhac 2015].

2.2.5 Radiomic features extraction

Radiomics features were extracted from PET images in all the VOI_m segmented by each nuclear medicine physician, in all the VOI_{SAEB} and in all the corresponding VOI_{20}, VOI_{30} and VOI_{40} using the two intensity rescaling factors (AR60 and RR). As a consequence, a total of 40 combinations of VOI, threshold and intensity rescaling factors were tested for the sample of 49 NET patients.

The radiomics features calculated by LIFEx (www.lifexsoft.org) software (in agreement with IBSI definitions, see appendix 1) were 51, subdivided in different RF families as follow:

- ten conventional PET parameters: such as SUV_{max}, SUVmean, SUVmin, SUVpeak, SUVstd;
- five descriptors of the image intensity histogram: HISTO_Skewness (asymmetry), HISTO_Kurtosis (flatness), HISTO_Energy (uniformity), HISTO_Entropy_log2 and HISTO_Entropy _log10 (randomness);
- four shape-based features, that describe shape and size of VOI: SHAPE_Volume (mL), SHAPE_Volume (voxels), SHAPE_Sphericity and SHAPE_Compacity;
- thirty-two textural features: a) seven features from grey-level co-occurrence matrix (GLCM), describing the correlation between pair of voxels in 13 directions of a three-dimensional space; b) eleven features from grey-level run length matrix (GLRLM), describing the number and length of run with a certain level of grey in 13 directions of a three-dimensional space (a gray level run can be described as a line of pixels in a certain direction with the same intensity value); c) eleven features from grey-level zone length matrix (GLZLM), describing the number and size of zones with a certain level of grey in 13 directions of a three-dimensional space; d) three features from neighborhood grey-level different matrix (NGLDM), describing the difference between a voxel and its connected neighbors.

Detailed descriptions of these features can be found in the Imaging Biomarker Standardization Initiative reference manual [IBSI 2019] and in the LifeX documentation (www.lifexsoft.org) [39]. In appendix 1 a comparison between IBSI features definitions and Lifex features is provided.

2.2.6 Statistical Analysis

Quantitative comparisons between VOI_m and VOI_{SAEB} were evaluated through the Dice Similarity Coefficient (DSC), which measures spatial overlap between two different volumes of the same lesions:

$$DSC(V_1, V_2) = 2\frac{|V_1 \cap V_2|}{|V_1| + |V_2|}$$

where $|V_1|$ and $|V_2|$ are the volumes of the two segmentations to be compared, $|V_1 \cap V_2|$ is the volume of the overlap between V_1 and V_2. DSC values can range from 0, when the two segmentations have no overlap, to 1 when the two segmentations are coincident.

The algorithm for simultaneous truth and performance level estimation (STAPLE), [Warfield 2004], was also used to compare VOI_{SAEB} with the "true" segmentation (VOI_{STAPLE}) derived by the four VOI_m. STAPLE algorithm considers a set of segmentations and calculates a probabilistic estimate of the true segmentation and a measure of the level of performance represented by each segmentation. The source of each segmentation in the collection can be one or more trained human raters, or it can be an automated segmentation algorithm. The probabilistic estimate of true segmentation is formed by estimating an optimal combination of the segmentations, weighting each segmentation by the estimated level of performance (in this study all equals), and incorporating a previous model for the spatial distribution of the segmented structures and spatial homogeneity constraints. STAPLE is simple to apply to clinical imaging data, readily allows performance evaluation of an automated image segmentation algorithm, and allows for direct comparisons between the human evaluator and algorithm performance.

Robustness of RFs was assessed by Intra-class Correlation Coefficients (ICC) to evaluate RFs consistency and Coefficient of Variance (COV) for each lesion to evaluate RFs reproducibility in the various settings.

The Intra-class Correlation Coefficients (ICC) was defined in chapter 1. RFs were considered highly robust in case of ICC> 0.9, robust if ICC > 0.8, moderate robust if ICC was between 0.5 and 0.8 and poorly robust if ICC was < 0.5. Also the COV was defined in chapter 1.

To investigate the relationship between RFs and lesion volume and SUV_{max}, a Pearson's correlation analysis was carried out.

All analyses were performed using statistical R software (R Foundation, Vienna , Austria).

2.3 Results

2.3.1 Impact of different segmentation approaches on VOI dimensions

Some representative examples of segmentation of three different lesions are shown in Figure 2.4.

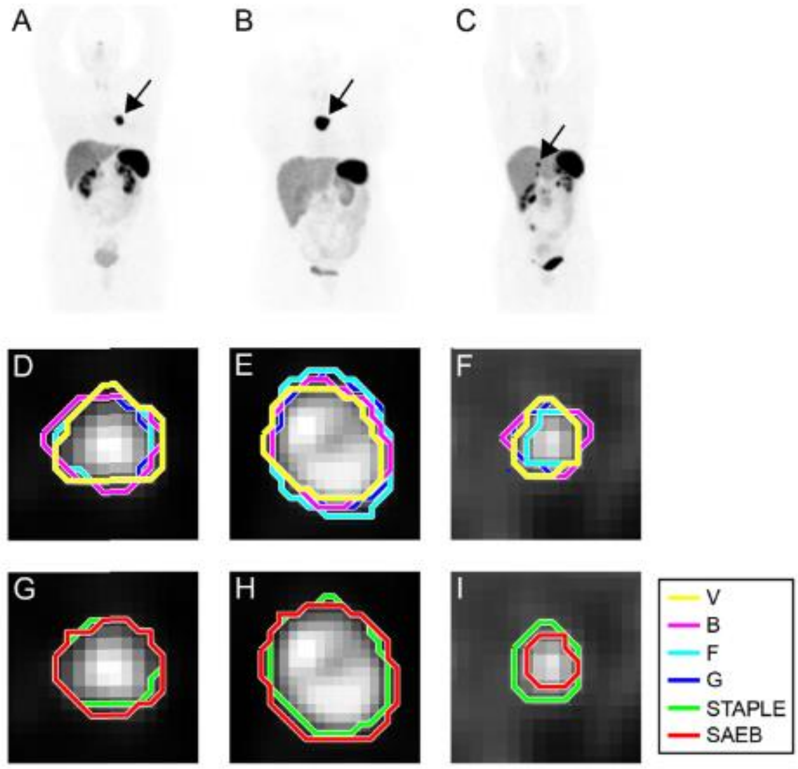


Fig. 2.4 Example of segmentation (single slice images) of lesions extracted from three different examined patients. (A) lung primary NET maximum intensity projection (MIP) of ^{68}Ga-DOTATOC PET; (D) same lung primary NET lesion contoured by different observers and (G) by the SAEB and STAPLE algorithms. In B, E, and H the same representations of a metastatic lesion in a mediastinal lymph node are shown; in C, F, and I a metastatic liver lesion.

The DSC were calculated between pairs of different segmentation approaches: between VOI_{SAEB} and the different VOI_m and between manual operators for each lesion (letters V, B, F and G indicate the different operators). DSC distributions are reported as boxplot in Figure 2.5 and 2.6. The mean value of DSC comparing VOI_{SAEB} with VOI_{STAPLE} was 0.75 ± 0.11 (0.45 – 0.92), while the mean value of DSC among VOI_m was 0.78 ± 0.03 (0.75 – 0.83). Comparison between operators showed a perfect segmentation matching (DSC = 1) for 24 out of 60 lesions with the 40% threshold; mean DSC using different SUV_{max} threshold are reported in Figure 2.7. Mean DSC index value improved as the applied SUV_{max} threshold increases.

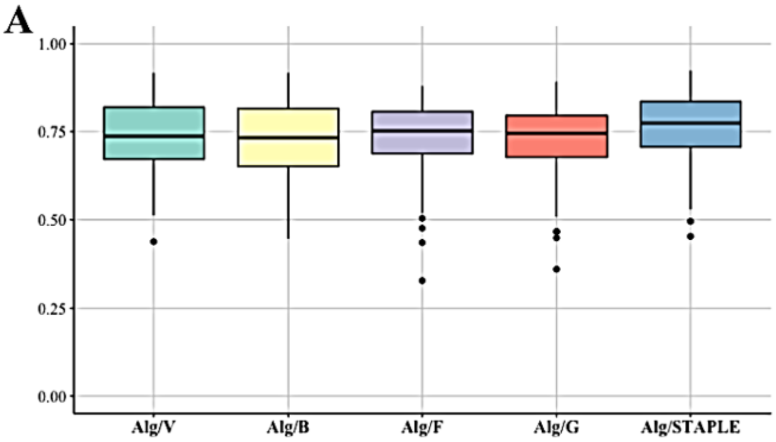


Fig. 2.5 Figure 4 – Boxplots of mean DSC comparing the semi-automatic edge-based algorithm (Alg) with manual segmentations (letters V, B, F and G indicate the different operators) and with STAPLE (A).

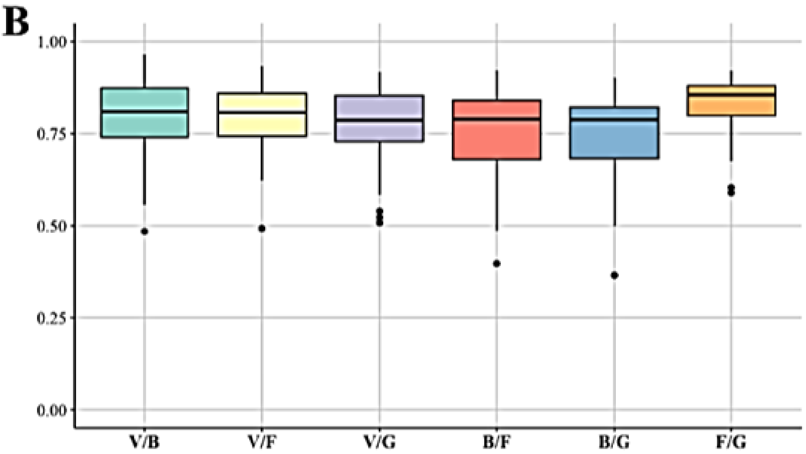


Fig. 2.6 Boxplots of mean DSC comparing manual segmentations by different operators.

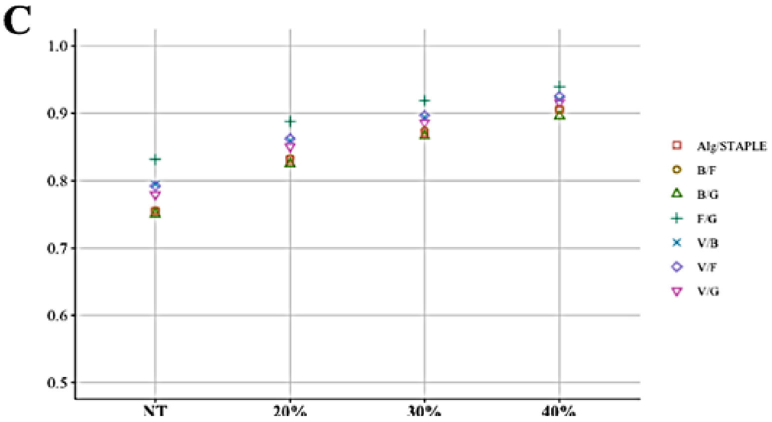


Fig. 2.7 Boxplots of mean DSC comparing different manual segmentations and Alg with STAPLE for different SUV$_{max}$ thresholds (NT = no threshold applied).

2.3.2 Analysis of the consequences of different segmentation methods on RF values

ICC and COV values calculated for each RF obtained with different segmentation methods (manual and semiautomatic) without threshold and with a threshold of 40% are shown in Figure 2.8. For this analysis the discretization method was fixed with a SUV maximum of 60 (AR60).

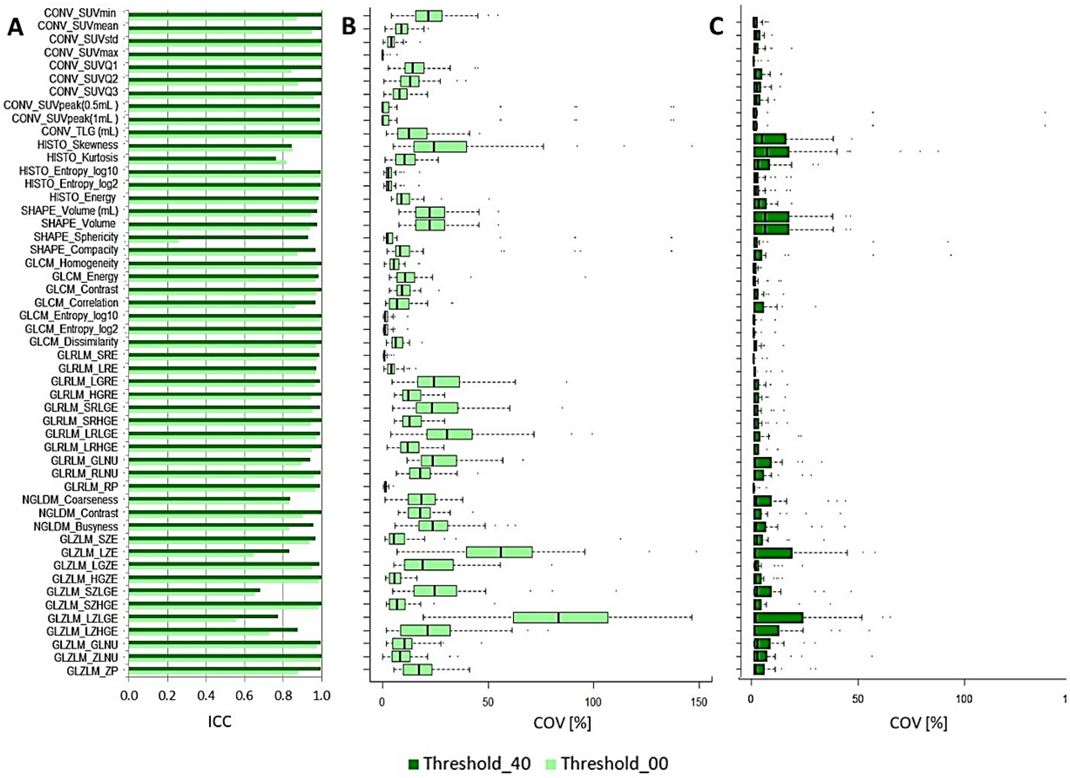


Figure 2.8 – (A) Bar diagrams of intra-class correlation coefficient (ICC) values of RFs for robustness to segmentation (different operators and semi-automatic algortihm), using AR60 and applying different SUV_{max} thresholds (no SUV_{max} threshold and 40% SUV_{max} threshold). (B) Boxplot of percentage COV for segmentation (different operators) for each RFs, using AR60 and applying no SUV_{max} thresholds. (C) Boxplot of percentage COV (for segmentation (different operators) for each RFs, using AR60 and applying 40% SUV_{max} thresholds.

Using no threshold and applying the AR60 intensity rescale factors, 65.3% of RFs showed high robustness (ICC > 0.9) to segmentation (7/10 conventional, 3/6 histogram, 2/4 shape and 22/32 textural). With a 40% SUV_{max} threshold, the fraction of RFs with high robustness increased to 86.5%. An increase of the SUV_{max} threshold produced a substantial increase of ICC of the following features: SHAPE_Sphericity, SHAPE_Compacity, GLCM_Correlation, NGLDM_Contrast, NGLDM_Busyness, GLZLM_LZE, GLZLM_LZLGE, GLZLM_LZHGE and GLZLM_ZP. A slightly increase was observed for all GLRLM features and for the remaining GLZLM ones (Figure 2.8 A).

About the corresponding COV analysis, when no threshold was applied, the grade of dispersion of the majority of RFs was rather low: median COV values were below 10%

for 47% of RFs and below 20% for 75% of RFs (Figure 2.8 B). Only two RFs (GLZLM_LZE and GLZLM_LZLGE) showed a COV >50%. Using a 40% SUV_{max} threshold, median values of COV were lower with median values well below of 10% for all the RFs (Figure 2.8 C).

ICC and COV analysis about the different SUV_{max} thresholds (no threshold, 20%, 30%, 40%) for AR60 and the first manual operator (the results obtained for the other VOI_m and VOI_{SAEB} were similar) are shown in figure 2.9. 51.9% of RF_s (5/10 conventional, 3/6 histogram, 0/4 shape and 19/32 textural) have high robustness (ICC > 0.9). The results of COV showed a high variability of the majority of RFs as a function of different SUV_{max} thresholds. Median value of COV was <10% for few RFs, namely SUV_{max}, SUV_{peak} (0.5ml and 1ml), HISTO_Entropy GLCM_Homogeneity, GLCM_Contrast, GLCM_Entropy, GLCM_Dissimilarity, GLRLM_SRE, GLRLM_LRE, GLRLM_RP, GLZLM_SZE and GLZM_ZP.

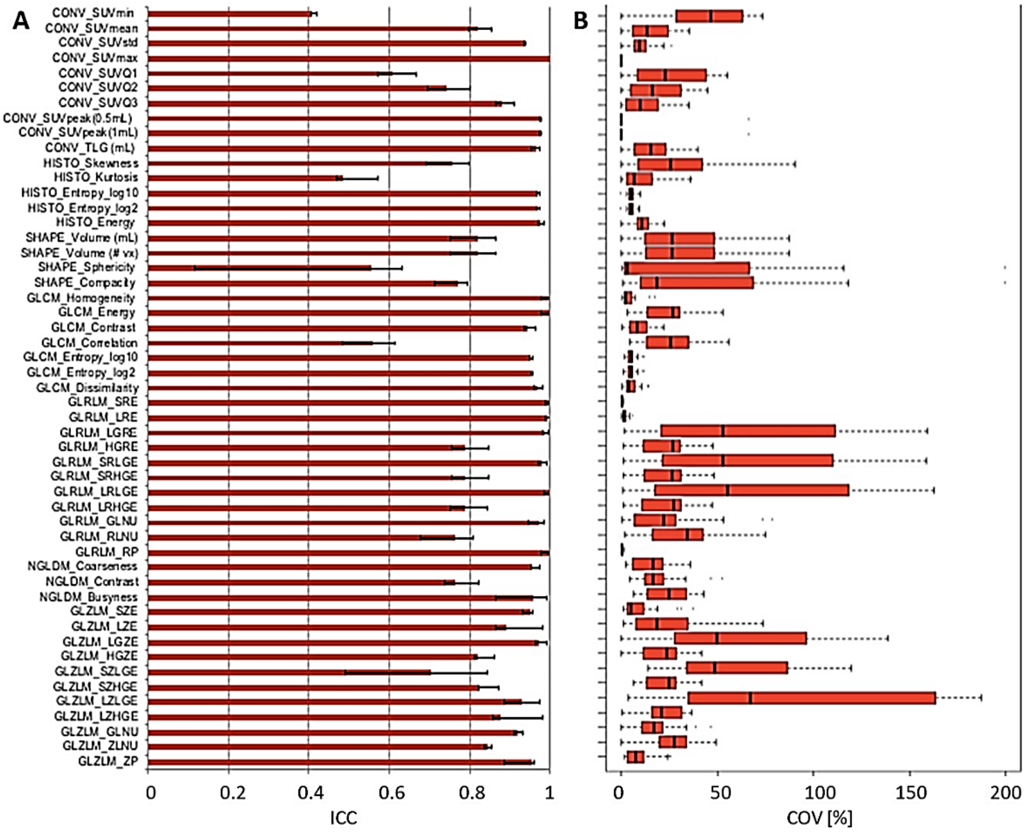


Figure 2.9 - (A) Bar diagrams of intra-class correlation coefficient (ICC) values of RFs for robustness to SUV_{max} thresholding. Bars show the median ICC between the different segmentations for the absolute intensity rescale factor AR60. Range error bars (in black) encompass the lowest and highest values for different operators. (B) Bar diagrams of mean Boxplot of COV (CV) for different threshold (20, 30, 40%) for each RFs, for the first operator (results superposable for the other operators). Mean COV (CV) has been calculated between all operators. NOTE: LIFEx software calculate the TLG (total lesion glycolysis) conventional parameter, that in this study corresponds to the TLSRE (total lesion somatostatin receptor expression).

2.3.3 Analysis of the consequences of different discretization approaches on RF values

Figure 2.10 A show ICC values for RFs evaluated with the two intensity discretization approaches, for the first operator and without applying a SUV_{max} threshold (the results for the other operators are very similar). Median values of ICC for intensity rescale factors were > 0.9 for all the conventional and shape features (not affected by the discretization), for HISTO_Skewness and HISTO_Kurtosis, and for only three texture features, namely GLCM_Correlation, GLRLM_RLNU and GLZLM_GLNU. Overall, the percentage of highly robust features (ICC>0.9) was 37% (10/10 conventional, 2/5 histogram, 4/4 shape and 3/32 textural). The majority of the remaining textural features showed a very poor robustness to discretization settings except for NGLDM Coarseness which had a median ICC > 0.7. The COV analysis (Figure 2.10 B) highlights in general low COV values for all the RFs with high ICC.

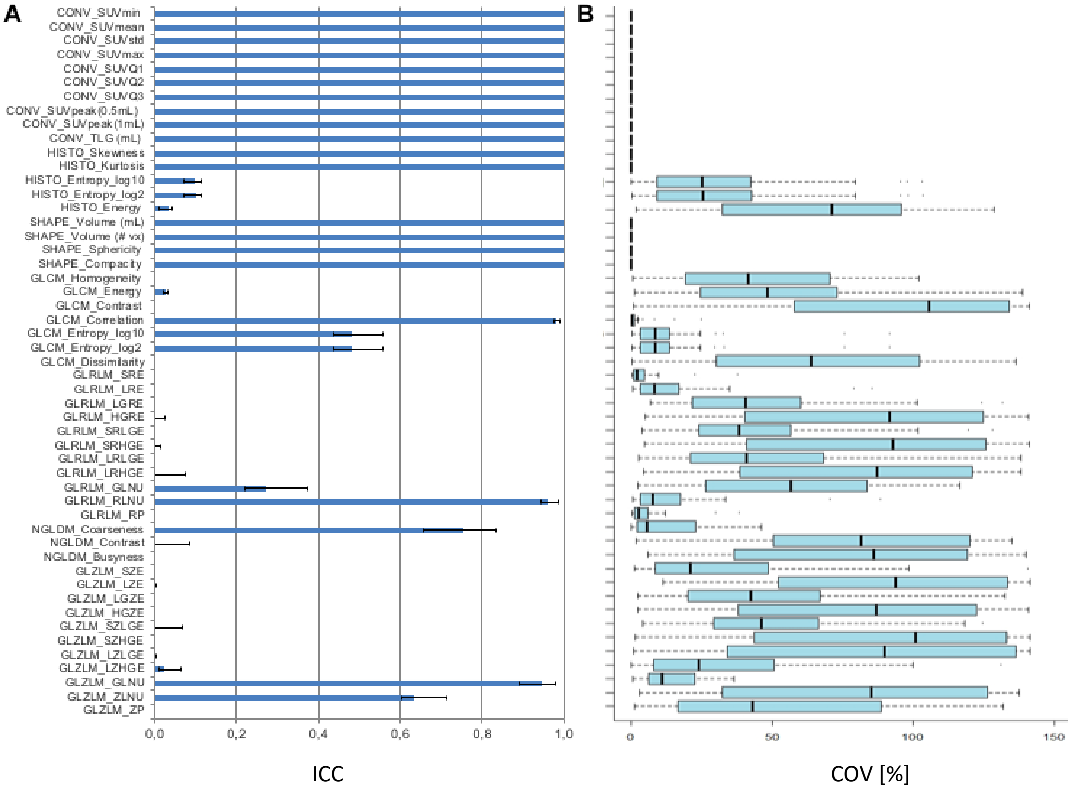
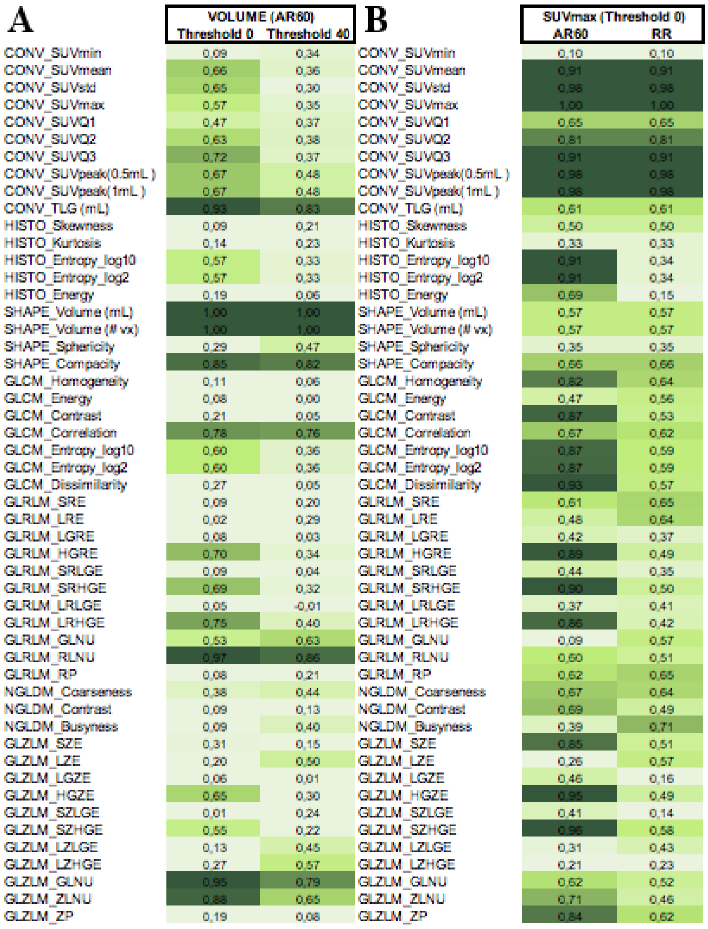


Figure 2.10 - (A) Bar diagrams of intra-class correlation coefficient (ICC) values of RFs for robustness to different intensity rescale factors (RR and AR60) when no threshold was applied for a single operator. Bars in blue show the median ICC between different segmentations, applying no threshold. Range error bars (in black) encompass the lowest and highest values for different operators. (B) Boxplot of COV for different intensity rescale factors (RR and AR60), applying no threshold, for the first operator (results superposable for the other operators). Mean COV (CV) has been calculated between all operators.

Most of the other textural features are characterized by a very high dispersion, corresponding to a high percentage of COV value. Only GLCM_Entropy (log10 and log2), GLRLM_SRE, GLRLM_LRE, GLRLM_RP and NGLDM_coarseness presented a COV <10%, despite a corresponding low ICC for these RFs.

2.3.4 Analysis of the correlation of RFs with SUV_{max} and lesion volume

An important aspect of the information provided by RFs is the independence of their values from conventional SUV_{max} and lesion volumes, since RF analysis is expected to provide additional information beyond simple baseline characteristics. If an RF is strongly correlated with SUV_{max} or with volume, any subsequent evaluation may be redundant. Figure 2.11 showed Pearson correlation coefficients between RFs values and volume (for AR60, without SUV_{max} threshold and with 40% SUV_{max} threshold applied, respectively) and SUV_{max} of the ROI (AR60 and RR, no threshold applied). Few RFs show high correlation with volume, both without threshold and with a 40% threshold. Instead, several RFs showed a high correlation with SUV_{max} when AR60 was employed (higher than 0.9 for HISTO_Entropy_log10, HISTO_Entropy_lo2, GLCM_Dissimilarity, GLRLM_SRHGE, GLZLM_SZHGE). When RR was used, no RFs were high correlated with SUV_{max} (except obviously for the “SUV” distribution features).



A

	VOLUME (AR60)	
	Threshold 0	Threshold 40
CONV_SUVmin	0,09	0,34
CONV_SUVmean	0,66	0,36
CONV_SUVstd	0,65	0,30
CONV_SUVmax	0,57	0,35
CONV_SUVQ1	0,47	0,37
CONV_SUVQ2	0,63	0,38
CONV_SUVQ3	0,72	0,37
CONV_SUVpeak(0.5mL)	0,67	0,48
CONV_SUVpeak(1mL)	0,67	0,48
CONV_TLG (mL)	0,93	0,83
HISTO_Skewness	0,09	0,21
HISTO_Kurtosis	0,14	0,23
HISTO_Entropy_log10	0,57	0,33
HISTO_Entropy_log2	0,57	0,33
HISTO_Energy	0,19	0,06
SHAPE_Volume (mL)	1,00	1,00
SHAPE_Volume (# vx)	1,00	1,00
SHAPE_Sphericity	0,29	0,47
SHAPE_Compacity	0,85	0,82
GLCM_Homogeneity	0,11	0,06
GLCM_Energy	0,08	0,00
GLCM_Contrast	0,21	0,05
GLCM_Correlation	0,78	0,76
GLCM_Entropy_log10	0,60	0,36
GLCM_Entropy_log2	0,60	0,36
GLCM_Dissimilarity	0,27	0,05
GLRLM_SRE	0,09	0,20
GLRLM_LRE	0,02	0,29
GLRLM_LGRE	0,08	0,03
GLRLM_HGRE	0,70	0,34
GLRLM_SRLGE	0,09	0,04
GLRLM_SRHGE	0,69	0,32
GLRLM_LRLGE	0,05	-0,01
GLRLM_LRHGE	0,75	0,40
GLRLM_GLNU	0,53	0,63
GLRLM_RLNU	0,97	0,86
GLRLM_RP	0,08	0,21
NGLDM_Coarseness	0,38	0,44
NGLDM_Contrast	0,09	0,13
NGLDM_Busyness	0,09	0,40
GLZLM_SZE	0,31	0,15
GLZLM_LZE	0,20	0,50
GLZLM_LGZE	0,06	0,01
GLZLM_HGZE	0,65	0,30
GLZLM_SZLGE	0,01	0,24
GLZLM_SZHGE	0,55	0,22
GLZLM_LZLGE	0,13	0,45
GLZLM_LZHGE	0,27	0,57
GLZLM_GLNU	0,95	0,79
GLZLM_ZLNU	0,88	0,65
GLZLM_ZP	0,19	0,08

B

	SUVmax (Threshold 0)	
	AR60	RR
CONV_SUVmin	0,10	0,10
CONV_SUVmean	0,91	0,91
CONV_SUVstd	0,98	0,98
CONV_SUVmax	1,00	1,00
CONV_SUVQ1	0,65	0,65
CONV_SUVQ2	0,81	0,81
CONV_SUVQ3	0,91	0,91
CONV_SUVpeak(0.5mL)	0,98	0,98
CONV_SUVpeak(1mL)	0,98	0,98
CONV_TLG (mL)	0,61	0,61
HISTO_Skewness	0,50	0,50
HISTO_Kurtosis	0,33	0,33
HISTO_Entropy_log10	0,91	0,34
HISTO_Entropy_log2	0,91	0,34
HISTO_Energy	0,69	0,15
SHAPE_Volume (mL)	0,57	0,57
SHAPE_Volume (# vx)	0,57	0,57
SHAPE_Sphericity	0,35	0,35
SHAPE_Compacity	0,66	0,66
GLCM_Homogeneity	0,82	0,64
GLCM_Energy	0,47	0,56
GLCM_Contrast	0,87	0,53
GLCM_Correlation	0,67	0,62
GLCM_Entropy_log10	0,87	0,59
GLCM_Entropy_log2	0,87	0,59
GLCM_Dissimilarity	0,93	0,57
GLRLM_SRE	0,61	0,65
GLRLM_LRE	0,48	0,64
GLRLM_LGRE	0,42	0,37
GLRLM_HGRE	0,89	0,49
GLRLM_SRLGE	0,44	0,35
GLRLM_SRHGE	0,90	0,50
GLRLM_LRLGE	0,37	0,41
GLRLM_LRHGE	0,86	0,42
GLRLM_GLNU	0,09	0,57
GLRLM_RLNU	0,60	0,51
GLRLM_RP	0,62	0,65
NGLDM_Coarseness	0,67	0,64
NGLDM_Contrast	0,69	0,49
NGLDM_Busyness	0,39	0,71
GLZLM_SZE	0,85	0,51
GLZLM_LZE	0,26	0,57
GLZLM_LGZE	0,46	0,16
GLZLM_HGZE	0,95	0,49
GLZLM_SZLGE	0,41	0,14
GLZLM_SZHGE	0,96	0,58
GLZLM_LZLGE	0,31	0,43
GLZLM_LZHGE	0,21	0,23
GLZLM_GLNU	0,62	0,52
GLZLM_ZLNU	0,71	0,46
GLZLM_ZP	0,84	0,62

Figure 2.11 Pearson correlation coefficients (mean values between different operators): (A) between RFs values and volume (for no SUV_{max} threshold and 40% SUV_{max} threshold applied); (B) between RFs and SUV_{max} (for AR60 and RR intensity rescale factors) without threshold.

2.3.5 Impact of discretization on RFs for different therapy outcomes

For the two patients with different therapy outcome, textural analysis was performed on pre-PRRT PET/CT data, on a total of 8 liver metastases in Patient A and 10 liver metastases in Patient B. Patient A presented a progression disease after PRRT, while patient B showed a very good partial response. P-values of Mann-Whitney test comparing the radiomic features values extracted by the liver lesions of the two patients are given in Table 2.1.

RADIOMIC FEATURES	**p-value**	**RADIOMIC FEATURES**	**p-value**
CONVENTIONAL-SUVmin	0.536	**GLRLM_LRE**	0.001
CONVENTIONAL-SUVmean	0.230	GLRLM_LGRE	0.351
CONVENTIONAL-SUVstd	0.025	GLRLM_HGZE	0.071
CONVENTIONAL-SUVmax	0.036	GLRLM_SRLGE	0.174
CONVENTIONAL-SUVpeak sphere 1mL	0.222	**GLRLM_SRHGE**	0.042
CONVENTIONAL-TLG (mL)	0.025	GLRLM_LRLGE	0.408
HISTO_Skewness	0.001	GLRLM_LRHGE	0.918
HISTO_Kurtosis	0.007	**GLRLM_GLNU**	0.001
HISTO_ExcessKurtosis	0.007	**GLRLM_RLNU**	0.023
HISTO_Entropy_log10	0.230	**GLRLM_RP**	0.002
HISTO_Entropy_log2	0.230	NGLDM_Coarseness	0.142
HISTO_Energy (uniformity)	0.417	**NGLDM_Contrast**	0.001
SHAPE_Volume (mL)	0.003	**NGLDM_Busyness**	0.001
SHAPE_Volume (voxel)	0.003	**GLZLM_SZE**	0.001
SHAPE_Sphericity	0.070	**GLZLM_LZE**	0.001
SHAPE_Compacity	0.003	GLZLM_LGZE	0.174
GLCM_Homogeneity	0.002	**GLZLM_HGZE**	0.008
GLCM_Energy	0.210	GLZLM_SZLGE	0.252
GLCM_Contrast	0.001	**GLZLM_SZHGE**	0.003
GLCM_Correlation	0.008	**GLZLM_LZLGE**	0.002
GLCM_Entropy_log10	0.091	**GLZLM_LZHGE**	0.003
GLCM_Entropy_log2	0.091	GLZLM_GLNU	0.210
GLCM_Dissimilarity	0.001	GLZLM_ZLNU	0.174
GLRLM_SRE	0.001	**GLZLM_ZP**	0.001

Table 2.1 – P-values of Mann-Whitney U-test comparing the radiomic features value extracted by the liver lesions VOIms of the pre-PRRT PET/CT of the two patients (p values lower than 0.05 are highlighted in grey).

25 features show a p-value of the Mann-Whitney test below 0.05 (three conventional included the $\mathrm{SUV}_{\mathrm{max}}$, three histograms, three shape features included the volume and nineteen textural features). Considering only the features resulted significant at the Mann-Whitney test, a Pearson correlation analysis was performed in order to identify the non-redundant features. The results are showed in Figure 2.12 as heatmap. It is appreciable a cluster of features with Pearson correlation close to -1 with $\mathrm{SUV}_{\mathrm{max}}$ and volume and some features with positive correlation.

Non è possibile visualizzare l'immagine.

Figure 2.12 – Graphic representation of radiomic-feature clustering identified with the Pearson Correlation matrix heat map, considering only the 26 radiomic features resulted significant at the Mann-Whitney test ($p<0.05$). This graphic displays the absolute value of the correlation coefficient between each pair of radiomic features, ranging from + 1 (positive linear correlation, in red) to – 1 (negative linear correlation, in green). The correlation coefficient of 0 is represented in yellow and identified radiomic features that are not correlated.

The features with a poor correlation (range between 0.8 and -0.8) with SUV_{max} and/or volume of the VOI are showed in Figure 2.13, and the boxplot of the seven features of second-order (GLCM_Correlation, GLRLM_LRE, GLRLM_RP, GLRLM_SRE, GLZLM_LZLGE, GLZLM_ZP and NGLDM_Busyness) not correlated with both SUV_{max} and volume are showed in Figure 2.14. Applying the Bonferroni correction for 7 tests a significance level of $0.05/7= 0.007$ was considered. Six of the seven features show a significant difference between the lesions of the responding patient and the lesions of the not responding.

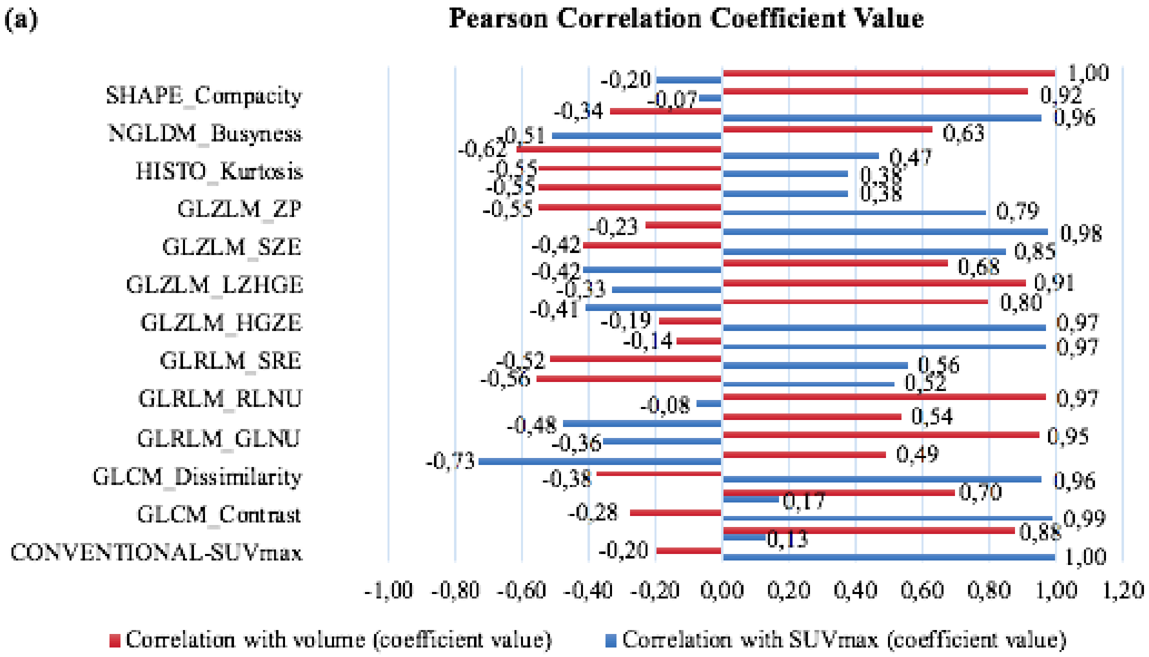


Figure 2-13 – Pearson correlation coefficient values between radiomic features and SUV_{max} and volume (mL), respectively.

Another type of analysis that can be applied to samples of multivariate observations is the principal component analysis (PCA), a technique that allows to increase the interpretability of the data while preserving the maximum amount of information and allowing the visualization of multidimensional data. Formally, PCA is a statistical technique for reducing the dimensionality of a dataset. This is accomplished by linearly transforming the data into a new coordinate system where (most) of the variation in the data can be described with smaller dimensions than the initial data.

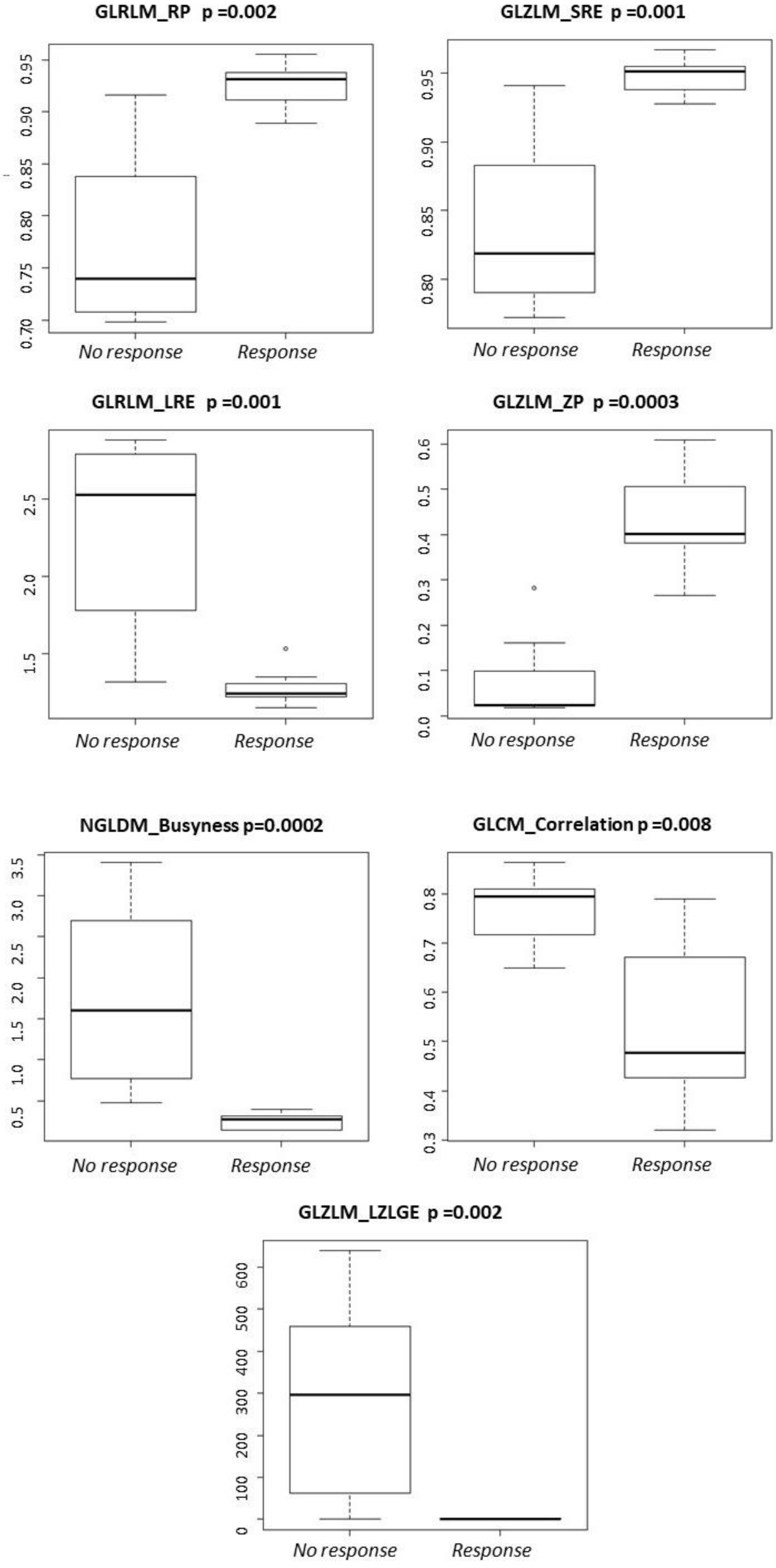


Figure 2.14 Boxplot of the second-order features not correlated with both SUV_{max} and volume.

The Principal Component Analysis first two components, performed on the 26 features, explain 83.8% of total variance. Figure 2.15 shows the coefficients of the two first principal components relative to the RF. Positive correlated RF point to the same side of the plot and negative correlated RF point to opposite sides of the graph. The SUV-related RFs are visible in the lower left quadrant and the volume-related RFs in the lower right.

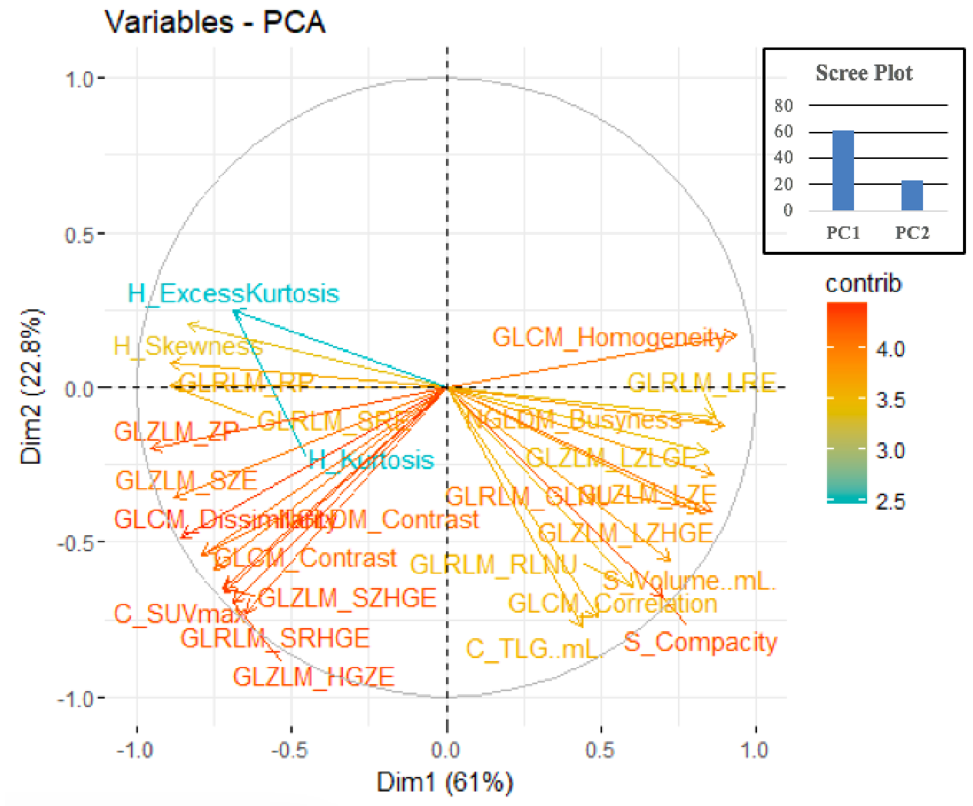


Figure 2.15 – Graphic representation of the features correlation plots resulted from the Principal Principal Component Analysis (PCA), performed on the 26 radiomic features resulted significant at the Mann-Whitney test (a).

The impact of different intensity rescale choices on the relative differences between responding and not responding lesions was also investigated (table 2.2, figures 2.16-2.19). In general, significative differences were observed for more RF when the intensity discretization was equal (fixed bin size) or similar for the two patients. A p-value lower than 0.05 for all the combinations of intensity discretization was observed for the following texture RF: GLCM_CORRELATION, GLRLM_GLNU (figure 2.18), GLRLM_RLNU and GLZLM_LZHGE (figure 2.19). These texture RF showed both robustness with respect to all the different discretization approaches and potential predictivity added value on the two patients with different PRRT outcome. GLCM_CORRELATION is also not correlated with SUV_{max}.

Feature	NR030 vs R060	NR030 vs R080	NR030 vs R100	NR060 vs R060	NR060 vs R080	NR060 vs R100	NR080 vs R060	NR080 vs R080	NR080 vs R100
CONVENTIONAL_SUVmax	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
GLCM_Homogeneity	0.05	0.17	0.61	0.00	0.01	0.02	0.00	0.00	0.01
GLRLM_SRHGE	0.54	0.47	0.21	0.04	0.35	1.00	0.01	0.03	0.14
GLZLM_HGZE	0.84	0.61	0.47	0.01	0.07	0.17	0.00	0.02	0.03
GLRLM_GLNU	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02
GLRLM_RLNU	0.04	0.02	0.04	0.04	0.02	0.01	0.02	0.01	0.01
GLZLM_LZHGE	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
GLCM_Correlation	0.003	0.005	0.005	0.008	0.012	0.012	0.012	0.016	0.016
GLRLM_SRE	0.091	0.142	0.681	0.001	0.008	0.016	0.000	0.001	0.003
GLRLM_LRE	0.142	0.142	0.299	0.001	0.005	0.023	0.000	0.000	0.003
GLRLM_RP	0.114	0.114	0.408	0.002	0.008	0.016	0.000	0.001	0.003
NGLDM_Coarseness	0.142	0.055	0.055	0.142	0.055	0.055	0.091	0.042	0.042
GLZLM_LZLGE	0.408	0.470	0.016	0.002	0.016	0.055	0.000	0.002	0.008
GLZLM_ZP	0.042	0.091	0.210	0.000	0.003	0.008	0.000	0.000	0.001

Table 2.2 – P-values of Mann-Whitney comparing the radiomic features value extracted by the liver lesions VOI$_{ms}$ of the pre-PRRT PET/CT of the two patients, for different intensity discretization choices. Features in italic character are not correlated with SUV$_{max}$ and volume.

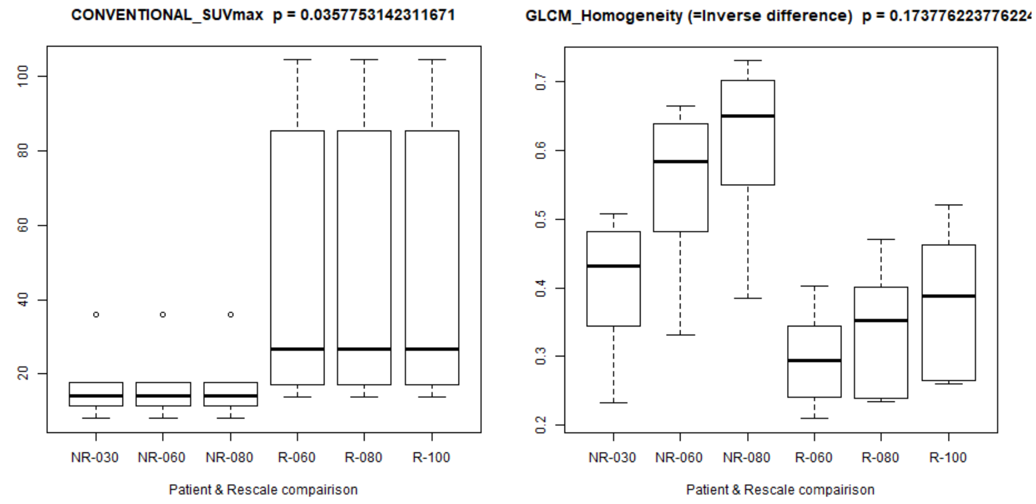


Figure 2.16 – Distributions of Conventional SUV$_{max}$ and GLCM_homogeneity for the two patients and for different intensity discretization.

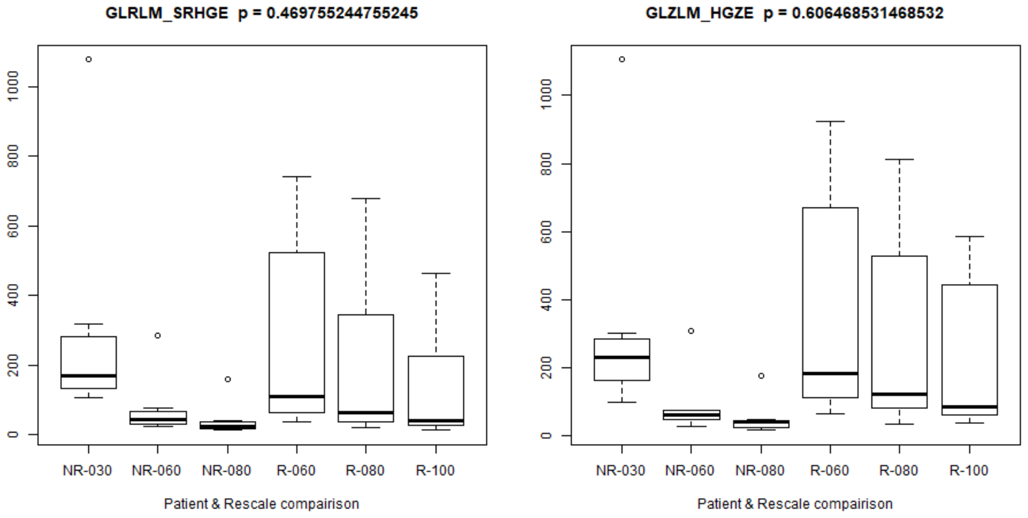


Figure 2.17 – Distributions of GLRLM_SRHGE and GLZLM_HGZE for the two patients and for different intensity discretization.

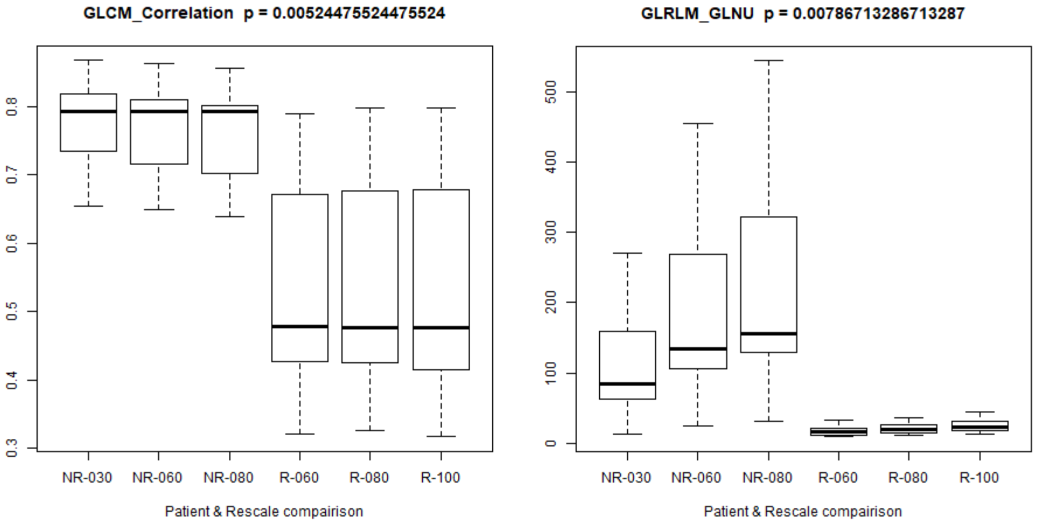


Figure 2.18 – Distributions of GLCM_correlation and GLRLM_GLNU for the two patients and for different intensity discretization.

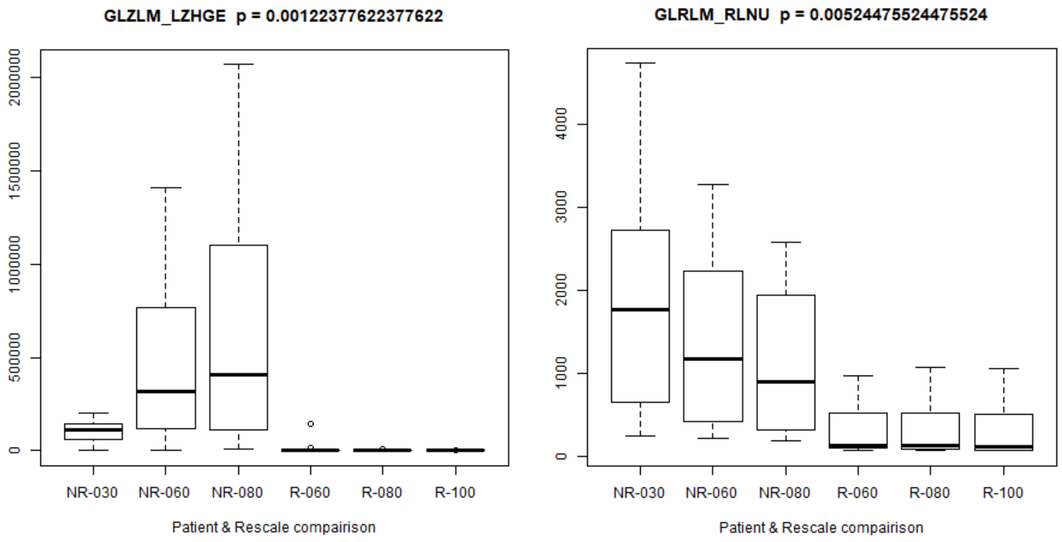


Figure 2.19 – Distributions of GLZLM_LZHGE and GLRLM_RLNU for the two patients and for different intensity discretization.

2.4 Discussion

This study aims to investigate the effects of different segmentation methods and grey-level intensity discretization on RFs in 68Ga-DOTA-TOC PET/CT. The major findings can be summarized as follows: (a) 65.3% of RFs (7/10 conventional, 3/6 histogram, 2/4 shape and 22/32 textural) showed high robustness (ICC > 0.9) and low relative dispersion (low mean COV value) to different operators without SUV_{max} threshold applied (manual segmentation); (b) increasing the SUV_{max} isocontouring threshold had a positive effect on RFs robustness to segmentation, but with the possible loosing of important biological information on the population under study; (c) quantitative comparison between a semi-automatic edge-based (SAEB) algorithm and manual segmentation showed a dice-coefficient similarity (DSC) of 0.75 ± 0.11 comparable to the DSC between operators (0.78 ± 0.03). These results suggest that a semi-automatic algorithm might be able to substitute manual segmentation to solve operator variability; (d) the use of absolute intensity rescaling factor (AR60) achieved higher robustness of RFs to segmentation and isocontouring thresholding than relative (RR) intensity rescaling factor; (e) fixed bin size with constant bin for all patients should highlight better different clinical variables such as therapy response.

In monocentric studies, image segmentation is one of the first step to take into account in radiomic analysis, since it is a possible source of RFs variability. Overall, RFs robustness to segmentation results quite good applying no SUV_{max} threshold (ICC >0.9 for 65.3% RFs), using AR60 absolute intensity rescale factor. The less robust RFs was SHAPE Sphericity, which describes the closeness of the VOI shape to a sphere and therefore is strongly sensitive to operator delineation. On the contrary, SHAPE_Compacity which is also a shape descriptor that reflects how compact is the VOI resulted to be robust. In line with previous studies applied to 18-FDG PET [Altazi 2017, Bashir 2017], GLZLM (also called GLSZM) features and in particular the ones measuring zones with low grey-level (SZLGE and LZLGE) resulted to have a moderate robustness (ICC from 0.5 to 0.8) to segmentation. This is likely related to the lower uptake in lesions edges, where operator and SAEB segmentation showed a higher variability (Figure 2.4).

When using AR60 absolute intensity rescale factor, the SUV_{max} thresholding had no considerable impact on inter-segmentation ICC values of the most part of textural features. Instead, increasing SUV_{max} threshold increased robustness of conventional, shape and GLRLM features. At the same time, however, it is clear that the SUV_{max} thresholding has an important impact on the dispersion of RFs, reducing it consistently, with a progressive decrease of all RFs COV values towards zero. These results may lead to choose segmentation with a 40% SUV_{max} threshold as preferable from the point of view of robustness, increasing the similarity between segmentations, but it should also be considered that the erosion of the external part of the lesion volume could be associated with a loss of relevant information. [Bashir 2017] already reported that segmentation with 40% SUV_{max} threshold yields superior inter-observer reproducibility of texture features in 18F-FDG PET/CT images, despite the loss of information related to exclusion of voxel intensities below the fixed 40% threshold, such as those arising from low-activity tumor regions or tumor boundaries. However, as early demonstrated by the study of Biehl et al with 18F-FDG PET/CT images [Biehl 2006], there is no consensus in the use of SUV_{max} threshold, also because this parameter can lead to an over- and/or underestimation of the lesion PET volume compared with CT volume related to low resolution, inherent noise, high uncertainties in lesion boundaries and motion blurring of the lesion related to the tomography characteristics [Foster 2014]. In my opinion, shared with the nuclear medicine physicians participating to this study, the concern related to the use of threshold-based segmentation methods in 68Ga-DOTA-peptide PET-CT images must be even higher due to the aforementioned heterogeneity of somatostatin receptors expression in neuroendocrine tumors which may explain that currently the few studies of radiomics with 68Ga-DOTA-peptide in the literature has been made on a manual-based segmentation [Werner 2019, Weber 2019].

In this context, it is evident that the use of automatic and semi-automatic segmentation methods is increasing in radiomics studies. The results of this study regarding tumor segmentation accuracy are promising since DSC comparing the SAEB segmentation with the manual segmentations was 0.75 ± 0.11 (median 0.77). The SAEB algorithm is characterized by its hybrid nature: the curve evolves in the image looking to both the original image (important for homogeneity) and the edge-enhanced image (important for discontinuities detection). This feature allows the SAEB algorithm to behave reproducibly for lesions in different locations, with different image appearances and different background uptakes, such as blob-like lesions with dark background (Figure 2.4A, 2.4D and 2.4G), heterogeneous lesions (Figure 2.4B, 2.4E and 2.4H) and liver lesions where background present SUV values comparable to the lesion (Figure 2.4C, 2.4F and 2.4I). Thus, the use of automatic and semi-automatic segmentation methods, such as SAEB algorithm, appears to be an important procedural aspect in future studies about analysis of RFs in 68Ga-DOTA SSTR-peptide PET/CT imaging, in order to reduce inter- and intra-reader variability of manual segmentation methods, which is also time consuming, and in order to reduce relevant information loss due to the application of 40% SUV_{max} threshold. However, further studies are needed to validate the semi-automatic segmentation method implemented for this study, comparing it with other similar methods already implemented, especially for PET/CT images, such as another gradient-based method commercially available named GRADIENT (MIM Software, Cleveland, OH) or the Fuzzy locally adaptive Bayesian (FLAB) method propped by [Halt 2009].

Another interesting result of this study is related to the negative impact of the relative intensity rescaling factor on the robustness to segmentation of the majority of textural features analyzed. As the rescaling is carried out according to the minimum and maximum values of the VOI, the same image is rescaled differently depending on the segmentation operator/method used, leading to high variability also in RFs values. Anyway, it must be highlighted that the use of relative resampling is not recommended in PET images, as already observed in 18F-FDG PET/CT [Leijenar 2015, Orlhac 2017]. In clinical cases the “fixed bin number” (RR) intensity rescaling factor is intuitively less appropriate: it is based on the range of SUV intensities found in the volume of interest, with low SUVs corresponding to low bin numbers and high SUVs corresponding to high bin numbers; hence bin width (in SUV) and SUV range may vary between images in a cohort, even though the number of bins is consistent. For the identification of RFs as new cancer-specific biomarker (e.g. NET), it is important that the textural features values would be directly comparable, both inter- and intra-patient, in order to derive meaningful conclusions. Moreover, NETs are characterized by an extremely variable expression of somatostatin receptors in 68Ga-DOTA-peptides with a corresponding broader range for SUV values (from close to 0 up to higher than 100) compared to 18F-FDG PET/CT, causing a greater impact on the RR compared to AR. This concept has been already emphasized by several studies: beyond all the variabilities related to the tomographs, segmentations and post-processing settings, the robustness of RFs is also related to the tumor characteristic and behavior [Messerli 2019] and to the radiotracer analyzed, as recently demonstrated by [Lu 2016] that studied the stability of RFs for nasopharyngeal carcinoma on both 11C-choline and 18F-FDG PET/CT images, with different results. In accordance with previous studies [Zwanenburg 2019, Traverso 2018], the impact of intensity discretization on textural features was stronger than segmentation. This is exactly the case of textural features, where differentiation between high and low-gray levels is needed, and thus, the choice of the discretization setting is relevant. When using no SUV_{max} threshold, only four textural features resulted to be robust: GLCM_Correlation, GLRLM_RLNU, NGLDM_Coarseness and GLZLM_GLNU (Figure 2.10).

ICC and COV provide in general complementary information: commonly it was observed for several features low values of ICC and high COV or the opposite, but there are also cases with low ICC and low COV or high ICC and high COV. These last cases can be explained considering that COV is sometimes high when the average values of features are close to zero. On the other hand, ICC is sometimes high even if COV is also high because, despite the great variability, the RFs remain correlated by changing the parameters, as we can observed in particular for GLRLM_LGRE, GLRLM_SRLGE, GLRLM_LRLGE and GLZLM_LZLGE for the threshold variability (Figure 2.10). This behavior can be due to the very small, close to zero, values of these RFs.

The extraction of robust RFs from 68Ga-DOTA SSTR-peptide PET/CT might contribute to solve some limitations related to the clinical evaluation of the SSTR expression in NET. At present, no consensus has been already reached regarding the assessment of patients who need to be investigated with both 18F-FDG and 68Ga-DOTA-peptide PET/CT (even if it is more probable that NET neoplasms with Ki67> 15% will have positive lesions in 18F-FDG PET/CT [Binderup 2010, Oberg 2016]). Furthermore, an early detection of more aggressive disease with 18F-FDG PET/CT does not necessary reflect a change in the therapeutic strategy. Finally, conventional semi-quantitative PET parameters showed a sub-optimal feasibility to select patients for receptor radionuclide therapy (PRRT) and to evaluate response to PRRT. In this scenario, this innovative analysis might be applied as prognostic biomarker and predictor of tumor heterogeneity in NET.

Regarding the analysis performed on the two patients with different PRRT outcome, the Mann-Whitney test performed on the pre-PRRT PET/CT data of the responder patient versus the non-responder patient showed high significant difference for the following features: GLCM (homogeneity, contrast, correlation and dissimilarity), GLRLM (SRE, GLNU, RLNU, LRE and RP), GLZLM (LZHGE, SZE, LZE, HGZE, SZHGE and LZLGE). Out of these features, seven second-order radiomic features have been identified as poor correlated (range between 0.8 and -0.8) with SUV_{max} and PET volume parameters: GLCM-Correlation, GLRLM-LRE, GLRLM-RP, GLRLM-SRE, GLZLM-LZLGE, GLZLM-ZP and NGLDM-Busyness. These radiomic features could potentially offer added predictive value, comparing with SUV_{max} and PET volume, and be more accurate in prediction of therapy (PRRT) response.

This study is not exempt from limitations. First, we did not initially perform a phantom study. To overcome this limitation, as recommended in the radiomics quality score (RQS) proposed by [Lambin 2017], in chapter 3 the results of a study on an anthropomorphic phantom filled with synthetic lesions as obtained from PET/CT images and created by the 3D printer are presented. This model will reflect realistic tumor shapes and heterogeneity uptakes for a prospective evaluation of RFs robustness. Second, NET are rare tumors and G3 NET and G3 NEC are rarely evaluated with 68Ga-DOTA SSTR-peptide; for this reason our sample size is yet too small and heterogeneous to evaluate the possible correlation between robust RFs and the histological NET grading system. Third, the difficulty of edge-based algorithms in the segmentation of lesions with an uptake similar to the background, or with small (<16 cm^3) or large dimension (>160 cm^3) [Pfaehler 2020] has been previously highlighted. However, we acknowledge that the number of cases analyzed in this study is not enough to quantitatively validate the algorithm in the aforementioned conditions.

In conclusion, the results of this study indicate that the use of RF is also feasible in 68Ga-DOTA-TOC PET/CT.The manual delineation of the VOI had an impact on the RF type dependent RF values, preserving the correlation with high ICC values in most cases, although some relatively high COV values. The 40% SUV$_{max}$ threshold increased RF robustness, but with a potential information loss of scanable lesions. A semi-automatic segmentation algorithm could be useful to solve both the impact of different manuals RF robustness segmentations and loss of valuable information due to SUV$_{max}$ threshold segmentation method. Finally, the discretization of the gray level affects the robustness of the RFs, which vary depending on whether relative or absolute are used resampling. In my opinion, an absolute resampling is more suitable for evaluation of NET with functional imaging (68Ga-DOTA-TOC PET/CT). These results suggest the need to standardize the methodology used in radiomics PET studies in 68Ga-DOTA-TOC PET/CT.

3. Implementation of a 3D printed phantom to study the impact of reconstruction parameters and used radioisotopes on PET radiomics features

3.1 The use of phantoms in nuclear medicine for radiomics analysis

In the literature there are a series of studies showing various implementation strategies of phantoms designed to simulate different patterns of signal heterogeneity in nuclear medicine images. Compared to the case of CT or MR, the fabrication of a nuclear medicine phantom has several unique features. The phantoms must in fact possess the following characteristics:

- ability to simulate a spatial distribution of the radioisotope similar to that of a real patient, for which the result strongly depends on the physiology and metabolism of the radiopharmaceutical as well as on the anatomical conformation;
- ease of management with contained risks of contamination and exposure of the operators, considering that the phantom must normally be packaged before each acquisition with handling of unsealed sources (in liquid, gel, powder form);
- reduced preparation times both for radioprotection and practical reasons;
- repeatability of the radioisotope insertion process: being the primary purpose often that of evaluating the repeatability and reproducibility of the heterogeneity measurements, the repeatability of the phantom packaging process for each acquisition is an essential requirement;
- traceability of the amount of radioisotope contained;
- containment materials should have attenuation coefficients similar to those of biological tissues.

An interesting review of the different approaches that can be used is present in [Valladares 2020] and the advantages and disadvantages of the different solutions adopted are summarized below.

In [Kadrmas 2009] the use of a cellular foam inside a phantom is proposed. Foams were inserted into the thoracic and pelvic compartment of a full-body phantom (figure 3.1). When phantoms are filled with a radioisotope, the pores create slightly non-uniform regions. This solution allows the use of a single radioactive solution to simulate different textures and concentrations, has limited preparation times and allows the use of anthropomorphic phantoms. On the other hand, it is rather difficult to reuse the same expanded materials after a first acquisition of the phantom and therefore it is possible to predict a poor reproducibility in multi-center studies, which has not yet been evaluated with this solution.

In [Carles 2017] the use of alginate to create heterogeneous phantom inserts is reported and the effects of respiratory motion in the analysis of texture features was evaluated. Combinations of alginate with four different activity concentrations of [18F]FDG and different geometric arrangements were used to construct lesions. During the phantom assembling process, an activity concentration similar to that observed in lung cancer patients was ensured for the lesions and the background. This approach certainly allows a wide flexibility and possibility of modeling the shape and concentration of radioisotope, with geometric shapes of various types made by hand. The main disadvantages are the limited repeatability, a non-negligible risk of personnel contamination during the construction phase and rather high preparation times.

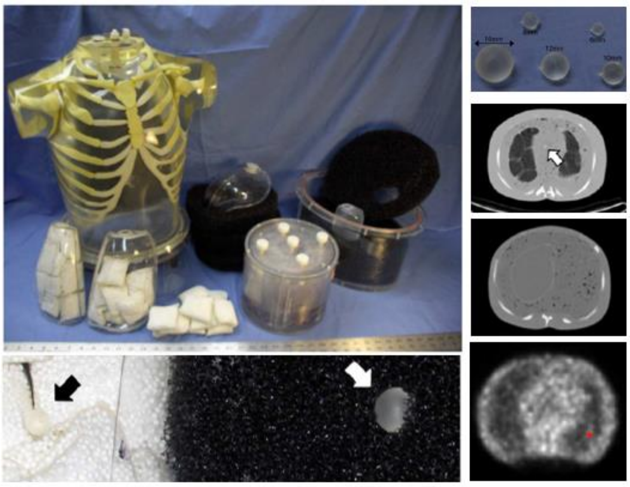


Figure 3.1 – Use of cellular foam to simulate heterogeneity in anthropomorphic phantoms [Kadrmas 2009].

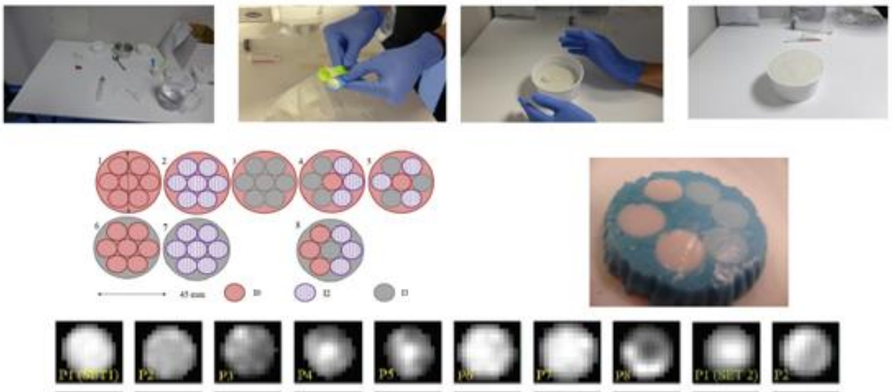


Figure 3.2 – Alginate phantom inserts [Carles 2017].

In [Forgacs 2016] a heterogeneous insert, made by seven syringes disposed as the bullets in the drum of a revolver gun, was placed in the NEMA IQ phantom for PET scans (figure 3.3). The sensitivity of the heterogeneity parameters was also evaluated by filling four and three insert syringes with 11C and 18F solutions, respectively. 18F has a half time of 110 minutes and C11 of about 20 minutes. Hence, different PET scans were acquired in the same system over time, resulting in different activity ratios due to the difference in half-life between 11C and 18F and as a consequence different heterogeneities pattern can be obtained with a single phantom preparation. The values of the corresponding features were plotted against time. This method therefore presents excellent ease of preparation and combination of different concentrations of radiopharmaceutical in the same insert, high reproducibility and contained risks of contamination and external exposure. The main disadvantages can be identified in the limited heterogeneity which is represented only on the axial plane and the presence of a simulation distant from the anatomical reality.

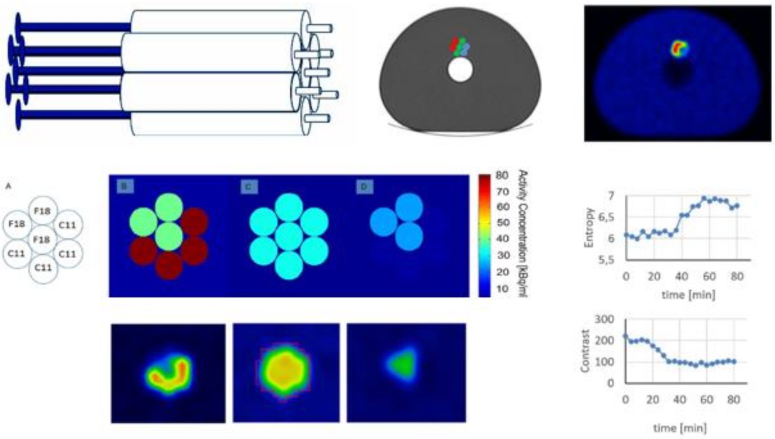


Figure 3.3 – Inserts of nuclear medicine phantom with syringes disposed as revolver gun. Pictures indicated as a,b, and c show activity distribution inside the revolver at t=0, 45 and 80 minutes of F18 and C11 radionuclides [Forgacs 2016].

In another study [Presotton 2018], tumor-like heterogeneous radioactivity distributions are simulated using cylindrical vessels filled with silica gel molecular sieves (Figure 3.4).

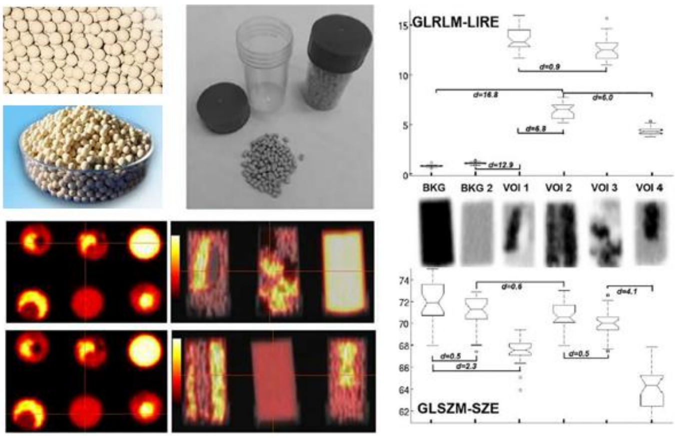


Figure 3.4 – Inserts of nuclear medicine phantom with silica gel molecular sieves [Presotto 2018].

The same quantity of molecular sieves was incorporated into three vessels of aqueous solution with different concentrations of 18F activity. Once the sieves absorbed the liquid, the mean activity concentrations were 3.6, 7.1 and 21.5 kBq/mL. The sieve mixes were placed in four cylindrical probes to create heterogeneous regions with different spatial radioactivity distributions. Of the 39 radiomic features extracted and evaluated, 21 were recommended for discrimination of heterogeneous patterns. These include 6 features from Gray Level Run Length Matrix (GLLRM), 8 from Gray Level Dimension Zone Matrix (GLSZM), 6 from Gray Level Co-Occurrence Matrix gray scale (GLCM), and 1 from Gray Level Difference Matrix (GLDM). This solution makes it possible to create multiple and complex heterogeneities with relative ease of preparation. On the other hand, the method has poor reproducibility, non-negligible contamination risks and a simulation that is rather distant from anatomical reality.

A completely different method has recently been proposed to mimic heterogeneous distributions of PET activity. It consists of a robotic arm capable of producing precise and reproducible movements along a 3D trajectory, with a long half-life sealed point radioactive source (Na22) mounted at the end of the robotic arm. Using controlled movements, different activity distribution patterns can be traced (figure 3.5).

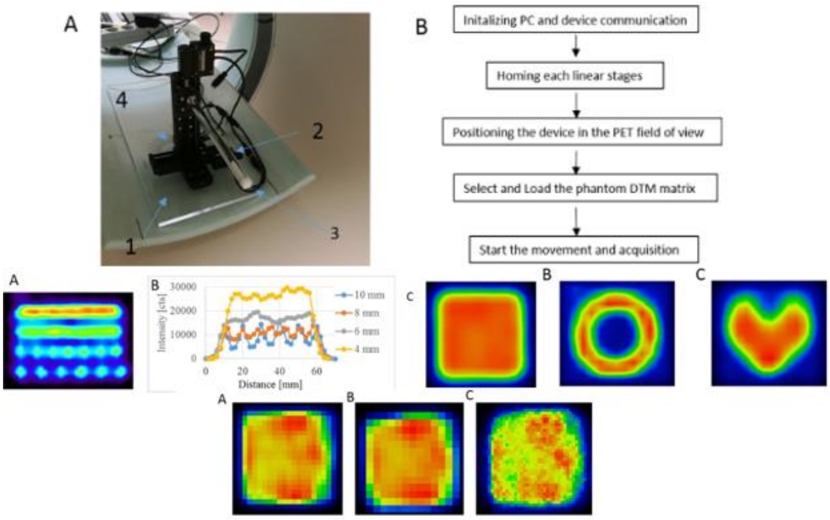


Figure 3.5 – Use of a robotic arm with a sealed source to simulate heterogeneity in PET imaging [Forgacs 2019].

This solution has numerous advantages, including practically zero preparation times and risks of contamination, low exposure risks, high reproducibility and heterogeneity in the images created. On the other hand, the acquisition times are very long to allow the progressive movement of the robotic arm, it is difficult to simulate different acquisition times and furthermore the attenuating material similar to human tissue is absent.

Berthon realized a three dimensional phantom stacking paper sheets printed with radioactive [^{18}F]FDG ink by means of a modified conventional printer [Berthon 2015]. They made in this way a sandwich phantoms to model the uptake of heterogeneous and irregular radiotracers in head and neck cancer patients. The prints were stacked between 2 mm thick oval sheets of PMMA. They were assembled all together using a plastic stand (Figure 3.6). Several irregular and spheroidal lesions, including heterogeneous models, were printed to test the performance of automatic segmentation algorithms for tumor delineation. In this study the tumor heterogeneity was not assessed. This method is the only one that presents the possibility of simulating the complete radionuclide distribution of a real case. The main disadvantages are linked to the presence of a homogeneous attenuation of the PMMA sheets and to non-negligible contamination risks. The repeatability of the method has not yet been evaluated for possible multicenter studies.

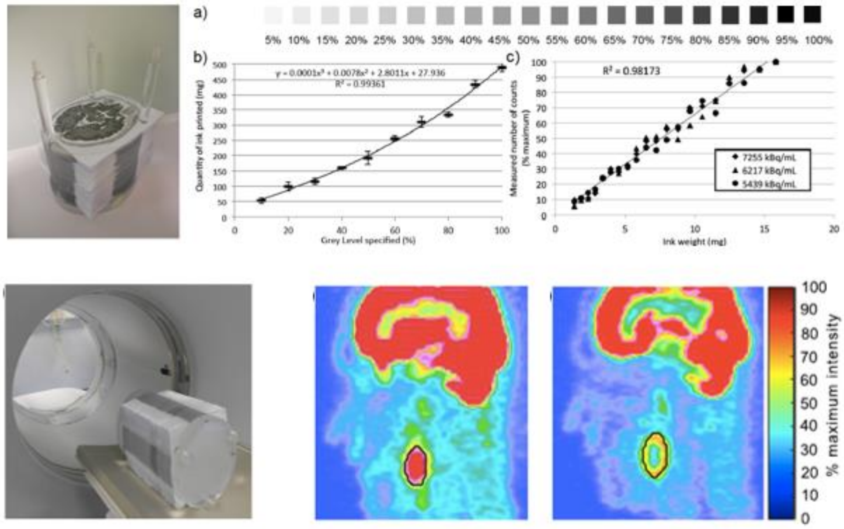


Figure 3.6 – 3D phantom implemented by sheets of paper printed with radioactive ink [Berthon 2015].

Cervino reported on a simple process of designing and manufacturing a 3D printed phantom insert to simulate the heterogeneous uptake of radiotracers as observed in lung tumors [Cervino 2017]. The insert was composed of two regions; the external part was formed by four porous wedges filling 50% of the total external volume. The four porous parts together form a cylinder having an internal hollow region. This arrangement provides a 2:1 activity concentration ratio between the inner porous region and the outer portion (figure 3.7). The insert was inserted into a QUASAR™ multipurpose phantom body.

Pfäehler report on the use of 3D printed shells obtained by the segmentation of non-small cell lung cancer (NSCLC) from real patients [Pfaehler 2020]. Three simulated lesions (homogeneous and heterogeneous absorption) are filled with different activity concentrations of an aqueous solution of [^{18}F]FDG. For PET scans, 3D printed lesions were placed in the NEMA IQ phantom to be filled with a specific concentration of background activity.

In general, the use of 3D printers offers a wide variety of phantom modeling, high repeatability, practicality of use and contained radioprotection risks. The main disadvantages are associated with limited heterogeneity, while attention must be paid to the thickness and density of the materials used to make the inserts.

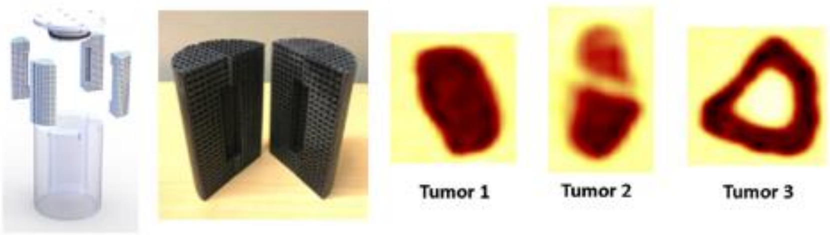


Figure 3.7 – Phantom produced by 3D printing with geometric structure [Cervino 2017]. and modelled on the base of real lung tumors (Pfaheler et al).

In this study, we implemented a phantom with 3D-printed inserts, simulating necrotic NET lesions, aimed to analyze the ability to simulate real lesions and obtain RF values similar to those of real patients, to investigate the reproducibility of RFs for different reconstruction techniques on the same PET/CT scanner, and to assess the impact of different radionuclides on RFs for identical inserts.

3.2 Material and methods

3.2.1 Phantom implementation

The basic idea for the implementation of a 3D phantom was a modification of a classic PET/CT phantom used for quality control, consisting of a hollow PMMA cylinder with spherical inserts of different sizes. For routine testing, the cylinder is filled with a low-concentration of radionuclide liquid solution to simulate the patient's fundus of normal tissue that is not radiopharmaceutical-avid and with a higher concentration of radionuclides in the spheres (usually at a 4:1 ratio). For the implementation of the phantom we decided to replace the regular sphere inserts with irregular shapes more similar to heterogeneous lesions of the patient, maintaining the same practical approach of filling management of the different inserts with ad hoc solutions prepared just before the acquisition of the phantom.

A tumoral lesion with a necrotic core extracted from a real PET/CT image served as a model to create the phantom inserts. The lesion was segmented manually using LifeX software, and the NIFTI-file of the segmented lesion was converted to an STL-file using ITK-snap software. STL is a file format commonly used for 3D Printing and computer aided design. 3D models of the completed inserts, including a stem and screw-on end for screwing the inserts into the Phantom, were created using the free Tinkercard web app. Six different insert sizes were envisioned: the largest size was approximately 54 x 47 x

44 mm with a necrotic core of 42 x 23 x 25 mm, while the others were smaller with each size reduced to 90%, 80%, 70%, 60% and 50%, respectively (Fig. 3.8).

Materials used for 3D printing of phantoms in nuclear medicine should be transparent (to monitor the filling of the radioactive solution), waterproof and with densities similar to biologic tissues. In this study we made some attempts with different 3D printing materials and 3D printers in order to find the best solution. In a first attempt we printed the inserts using a polylactic acid (Ultimaker PLA), that is cheap and compatible with the printer model available in our nuclear medicine department (Ultimaker S5). The results were totally unsatisfied because the obtained inserts did not allow to see the liquid inside and they were not waterproof (Fig. 3.9).

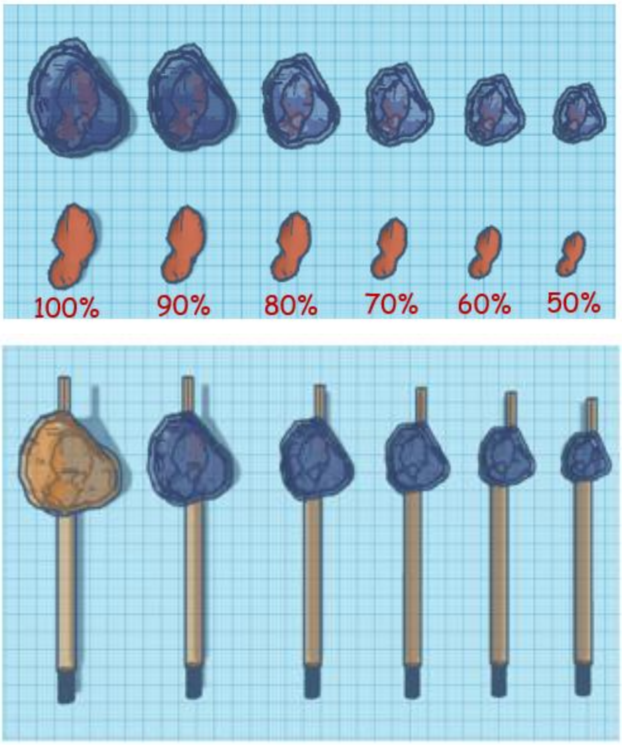


Figure 3.8 – Inserts simulating NET lesions with necrotic core and with different sizes.

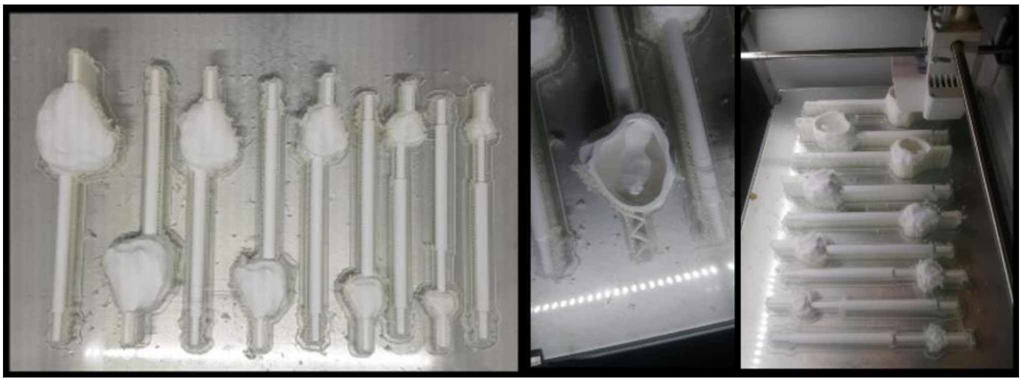


Figure 3.9 – First attempt of 3D printed inserts production with PLA material.

In the search for alternative solutions, the possibility of producing inserts with a totally transparent material (Polymaker Polysmooth) compatible with a treatment after printing to make it waterproof was identified. The post-processing treatment was implemented with a special machine (Polymaker Polysher) designed to polish extrusion-based 3D printed parts with isopropyl alcohol (also known as isopropanol) or ethanol to obtain a shiny, layer-free surface finish. No heating is involved, as the object is coated in liquid via nebulization technology. The total printing time for the six lesions was of about 25 hours. Unfortunately, the post-polishing process heavily reduced the transparency of the material and consequently this second attempt also failed.

Eventually we found another material with nominal properties compatible with our purposes, consisting of a transparent resin (FormLabs Standard Resin Cartridges Clear), compatible with a different printer available from a specialized commercial service (Form3+ 3D-printer, FormLabs). The production cost for six lesions was around 200 Euros and the result was satisfactory in terms of transparency and impermeability (Fig. 3.10).

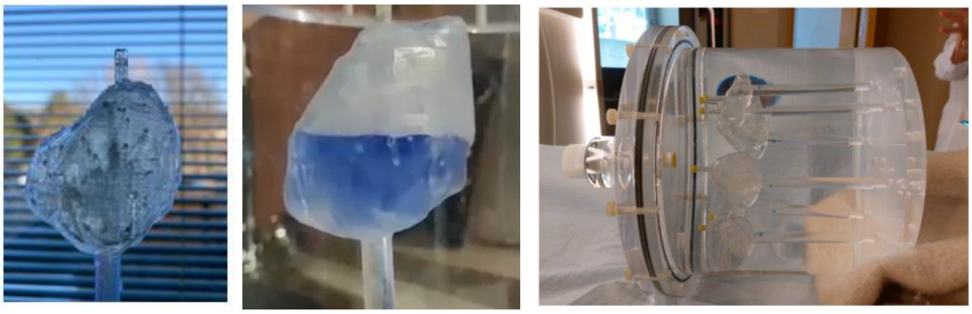


Figure 3.10 – Insert printed with clear resin, example of filling with liquid and 3d printed inserts included in the phantom.

3.2.2 Phantom acquisition

The phantom was scanned four times on a Philips Vereos digital PET/CT, using two different radioisotopes (^{18}F and ^{68}Ga) with two different solution concentrations, as shown in table 3.1. A ratio of concentrations of about 4:1 between the inserts and the background of the phantom is normally used for quality control purposes, whereas a ratio of about 16:1 was also used in this study in order to have values more similar to real patients SUV distributions.

Phantom acquisition	Radionuclide	Background concentration [kBq/ml]	Insert concentration [kBq/ml]
1	^{18}F	2.6	10.3
2	^{18}F	1.8	29.4
3	^{68}Ga	2.8	11
4	^{68}Ga	1.6	27.0

Table 3.1 – Phantom acquisitions with different radionuclides and concentrations.

All the scans were performed using the list-mode (recording of raw data with additional information such as detector pair address of the line of response) in order to allow subsequent reconstructions with different acquisition times. Phantom scans were reconstructed with different parameters, starting from the common reconstruction approach used for quality control (first row of table 3.2):

- the acquisition time was retrospectively changed using values of 1, 1.5, 2.5 and 5 minutes. The relative images were reconstructed considering only the relative time of signal collection, resulting in different signal to noise ratios.

- PSF reconstruction: a reconstruction option (present or absent) that improve the spatial resolution reducing the point spread function of the image production.

- number of subsets: in the *ordered-subsets expectation-maximization* (OSEM) algorithm [Ortuno 2006], the projection data are divided into subsets which are handled sequentially during each OSEM iteration. The number of subsets gives the approximate acceleration factor: one iteration of OSEM with N subsets gives a picture roughly similar to that of N iterations of other reconstruction algorithms such as the *maximum likelihood expectation maximization* (MLEM) method. There is a trade-off between the number of subsets and image quality. When the number of subsets is large, the size of each subset is small and each contains less tomographic and statistical information. This can result in enhanced noise textures and other subset-related artifacts in the final image. In this study the number of subsets was varied from 1 to a maximum of 30.

- number of iterations: the standard number of 3 iterations was changed from 1 to 10.

Table 3.2 shows the different combinations of the reconstruction parameters: the first column shows a code consisting of a letter which will be used in the following paragraphs and in the figures to indicate the results obtained for the corresponding combinations of parameters.

Code	Reconstruction description	Time [min]	PSF	Iteration	Subsets
A	Pet original 5 mm	5	No	3	15
B	Pet with PSF	5	Yes	3	15
C	Pet 4 it 15 sub	5	No	4	15
D	PET 3 it 1sub	5	No	3	16
E	PET 2p5 min	5	No	3	1
F	PET 1p5 min	2.5	No	3	15
G	PET 1 min	1.5	No	3	15
H	PET 1 it 15 sub	1	No	3	15
I	PET 1 it 15 sub	1	No	1	15
L	PET 5 it 15 sub	5	No	5	15
M	PET 7 it 15 sub	5	No	7	15
N	PET 10 it 15 sub	5	No	10	15
O	PET 3 it 25 sub	5	No	3	25
P	PET 3 it 35 sub	5	No	3	35
Q	PET 3 it 50 sub	5	No	3	50

Table 3.2 – Different combinations of parameters used for the reconstructions of phantom images.

3.2.3 Features calculation

52 radiomic features (Rfs), including first order (histogram) and second order (texture) features, were extracted with LifeX software. Relative trends of RFs and coefficients of variation (COV) were analyzed. Wilkoxon tests were applied to assess the differences between the RFs associated with the two isotopes. The RFs' correlation with volume was analyzed by calculating Pearson's correlation coefficients.

In addition, we have also selected 9 necrotic tumour lesions with different volumes from 9 different real PET/CT images (6 FDG and 3 DOTATOC, respectively) in order to compare the phantom RFs values with actual patients RFs.

3.2.4 Statistical analysis

The comparison between actual patient RFs and phantom RFs was done for the four combinations of radionuclides and concentrations, using the feature values extracted after the standard reconstruction. Mann Whitney p-values were also calculated.

In order to investigate the influence of the used radionuclide on RFs, a Mann-Whitney test with paired data was performed considering for each RF, radionuclide and reconstruction parameter combination the 6 values of the 6 lesions obtained with concentration ratio 4:1 and the 6 values obtained with the concentration 16:1.

The impact of different reconstruction techniques was explored analyzing relative plots of RFs towards the parameter values and implementing ICC tests considering each lesion with his radionuclide and concentration as a subject and different parameters as observers. Relative COV were also assessed.

Finally, the correlation of phantom RFs with lesion volumes were assessed by Pearson correlation coefficients.

All the analysis were performed using the "R" package (https://www.r-project.org/).

3.3 Results

3.3.1 Phantom dataset

The dataset of RFs consisted of a total of 360 image volumes and 18720 RFs values: 52 features per 6 lesions per 4 acquisition conditions (2 radionuclides and 2 concentrations) per 15 combinations of the reconstruction parameters.

Figure 3.12 show an example of phantom images obtained with CT, ^{18}F PET and ^{68}Ga PET. Voxel dimensions were 1.2 x 1.2 x 2 mm for the CT and 4 x 4 x 4 mm for PET. Hounsfield units of the lesion walls and necrotic cores were about 115 HU, close to PMMA values and slightly higher than biological soft tissues (usually in the range 40 – 80 HU). In figures 3.13 and 3.14 screenshots of PET images obtained with different reconstruction parameters for the two radionuclides are shown. The images are different in terms of noise and spatial resolution. In most cases slight appreciable differences are visible but for some reconstruction parameters the images are heavily modified, such as image E obtained with only one subset which is very smooth or the H image which is noisier due to the reduced acquisition time.

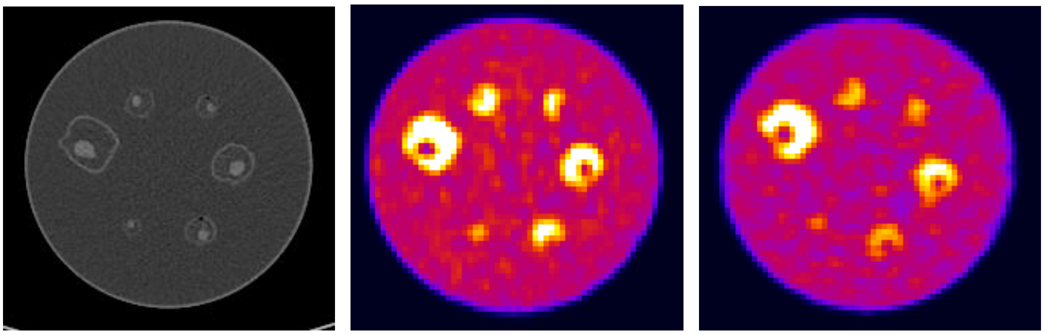


Figure 3.11 – Images of the phantom: CT central section on the left, a ^{18}F PET section in the middle and a ^{68}Ga PET section on the right.

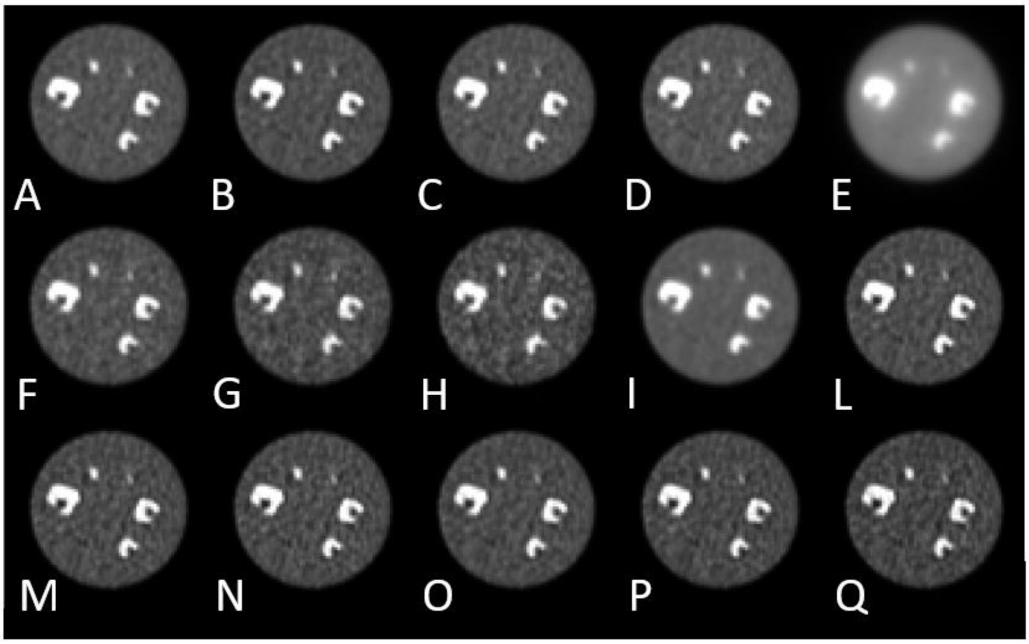


Fig. 3.12 Example of ^{18}F PET phantom images obtained with different reconstruction parameters (the letters correspond to the code of processing parameters combination reported in table 3.2).

In figure 3.14 some examples of lesions of real patients are shown, with sizes and shapes similar to the ones of the phantom inserts.

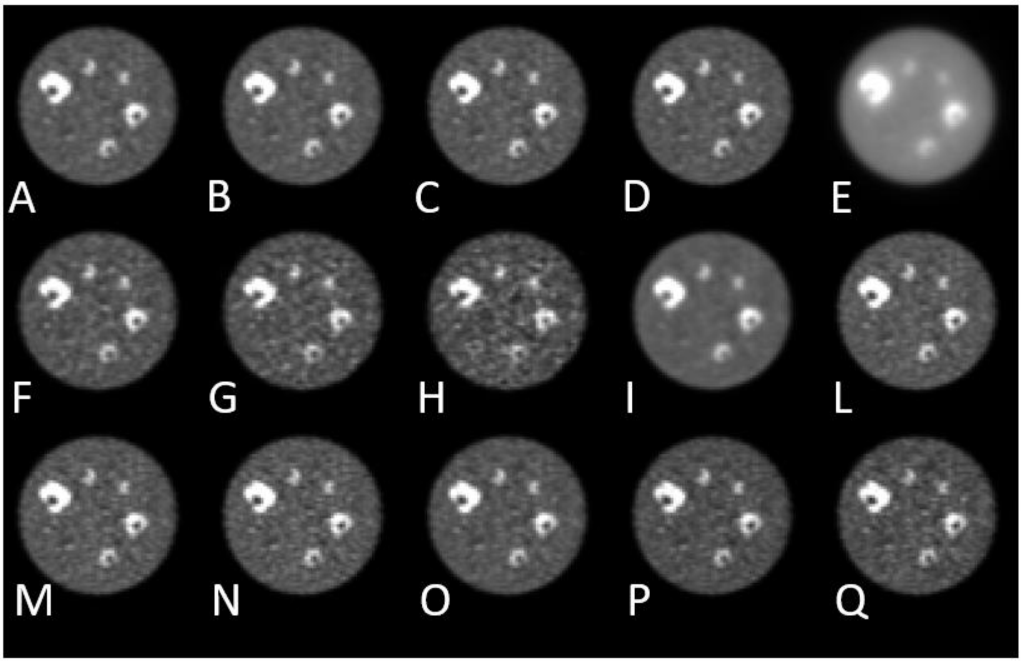


Fig. 3.13 Example of ^{68}Ga PET phantom images obtained with different reconstruction parameters (the letters correspond to the code of processing parameters combination reported in table 3.2).

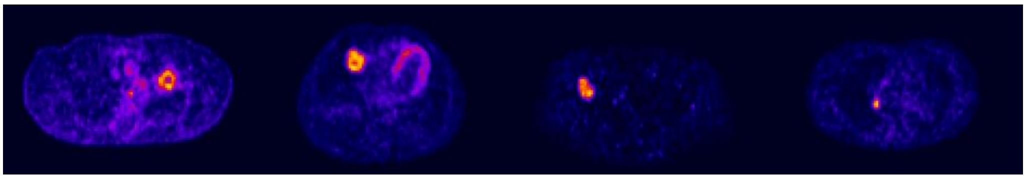


Fig. 3.14 Example of real lesions with shape similar to the phantom inserts: the two left images are ^{18}F and the two right images are ^{68}Ga PET.

3.3.2 Comparison of phantom RFs with real patient RFs

The comparison of RF values extracted by real patient and phantom inserts highlighted important differences when the radionuclide concentration ratio 4:1 was used and lower differences for concentration ratio 16:1. In particular, p-values of Mann-Whitney tests for concentration ratio 4:1 where below 0.05 for about 50% of conventional RFs (table 3.3), whereas for textural RFs the percentage with significative differences was 56% for ^{18}F and 78% for ^{68}Ga (table 3.4). Increasing the concentration ratio between inserts and background to 16:1, more similar to real patients, the percentage of RFs with significative differences reduces to 12-22% for different kind of RFs and radionuclides. Cells highlighted in gray correspond to p values less than 0.05.

	18F		68Ga	
Feature	4:1	16:1	4:1	16:1
CONVENTIONAL_SUVbwmin	0.00	0.00	0.71	0.57
CONVENTIONAL_SUVbwmean	0.00	0.31	0.02	0.07
CONVENTIONAL_SUVbwstd	0.00	0.31	0.02	0.14
CONVENTIONAL_SUVbwmax	0.00	0.94	0.02	0.07
CONVENTIONAL_SUVbwQ1	0.00	0.01	0.02	0.07
CONVENTIONAL_SUVbwQ2	0.00	0.06	0.02	0.07
CONVENTIONAL_SUVbwQ3	0.00	0.59	0.02	0.14
CONVENTIONAL_SUVbwSkewness	0.06	0.06	0.02	0.07
CONVENTIONAL_SUVbwKurtosis	0.48	0.59	0.10	0.57
CONVENTIONAL_SUVbwExcessKurtosis	0.48	0.59	0.10	0.57
CONVENTIONAL_TLG(mL)	0.39	0.94	0.10	0.39
SHAPE_Volume(mL)	0.48	0.09	0.55	0.06
SHAPE_Volume(vx)	0.48	0.39	0.55	0.79
SHAPE_Sphericity[onlyFor3DROI])	0.03	0.00	0.10	0.07
SHAPE_Surface(mm2)[onlyFor3DROI]	0.48	0.82	0.55	0.79
SHAPE_Compacity[onlyFor3DROI]	0.24	0.31	0.55	0.79

Table 3.3 – p-values of Mann-Whitney test of differences between phantom inserts and real patient lesion conventional RFs, for different radionuclides and concentrations. Cells highlighted in gray show p values less than 0.05.

	18F		68Ga	
Feature	4:1	16:1	4:1	16:1
GLCM_Homogeneity[=InverseDifference]	0.00	0.82	0.02	0.06
GLCM_Energy[=AngularSecondMoment]	0.00	0.48	0.02	0.39
GLCM_Contrast[=Variance]	0.00	0.59	0.02	0.07
GLCM_Correlation	0.13	0.06	0.55	0.57
GLCM_Entropy_log10	0.00	0.39	0.02	0.39
GLCM_Entropy_log2[=JointEntropy]	0.00	0.39	0.02	0.39
GLCM_Dissimilarity	0.00	0.59	0.02	0.07
GLRLM_SRE	0.10	0.33	0.04	0.06
GLRLM_LRE	0.10	0.13	0.04	0.06
GLRLM_LGRE	0.10	0.03	0.04	0.11
GLRLM_HGRE	0.10	0.79	0.04	0.23
GLRLM_SRLGE	0.10	0.03	0.04	0.11
GLRLM_SRHGE	0.10	0.79	0.04	0.23
GLRLM_LRLGE	0.10	0.01	0.04	0.11
GLRLM_LRHGE	0.10	0.66	0.04	0.23
GLRLM_GLNU	0.10	0.43	0.07	0.23
GLRLM_RLNU	0.19	0.93	0.79	0.63
GLRLM_RP	0.00	0.39	0.02	0.04
NGLDM_Coarseness	0.82	0.31	0.38	0.79
NGLDM_Contrast	0.00	0.70	0.02	0.07
NGLDM_Busyness	0.00	0.18	0.02	0.04
GLZLM_SZE	0.00	0.24	0.02	0.04
GLZLM_LZE	0.00	0.39	0.02	0.04
GLZLM_LGZE	0.00	0.06	0.02	0.25
GLZLM_HGZE	0.00	0.94	0.02	0.07
GLZLM_SZLGE	0.09	0.39	0.38	0.79
GLZLM_SZHGE	0.00	0.82	0.02	0.07
GLZLM_LZLGE	0.00	0.03	0.02	0.04
GLZLM_LZHGE	0.00	0.31	0.10	0.39
GLZLM_GLNU	0.24	0.82	0.26	1.00
GLZLM_ZLNU	0.00	0.82	0.02	0.25
GLZLM_ZP	0.00	0.39	0.02	0.04

Table 3.4: p-values of Mann-Whitney test of differences between phantom inserts and real patient lesion textural RFs, for different radionuclides and concentrations. Cells highlighted in gray show p values less than 0.05.

Figure 3.15 shows the boxplot of conventional SUV_{max} and lesion volume distributions, highlighting the similarities in the SUV range for phantom concentration ratio of 16:1 and in the volume range for all the phantom acquisitions.

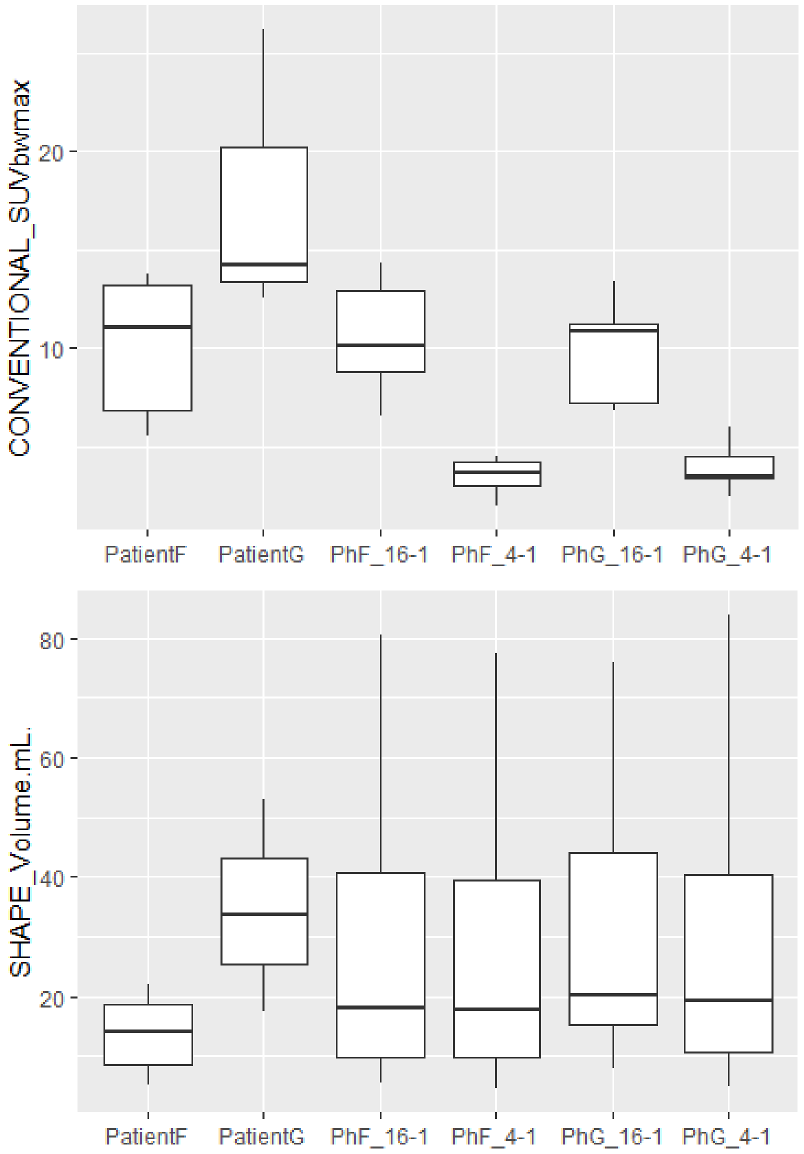


Figure 3.15 – boxplot of conventional SUV max and lesion volume distributions, for 18F patients (PatientF), 68Ga patients (PatientG), 18F phantom with 4:1 (PhF_4-1) and 16:1 (PhF_16-1) concentration ratio and 68Ga phantom with the same concentration ratios (PhG_4-1 and PhG_16-1).

There are more differences between patients and phantoms for 68Ga features, and this could be a consequence of the greatest differences between the sample patients' maximum SUV and phantom maximum SUV.

In figure 3.16 two examples of feature distributions are shown, without significant differences between patients and phantom distributions for concentration ratio 16:1.

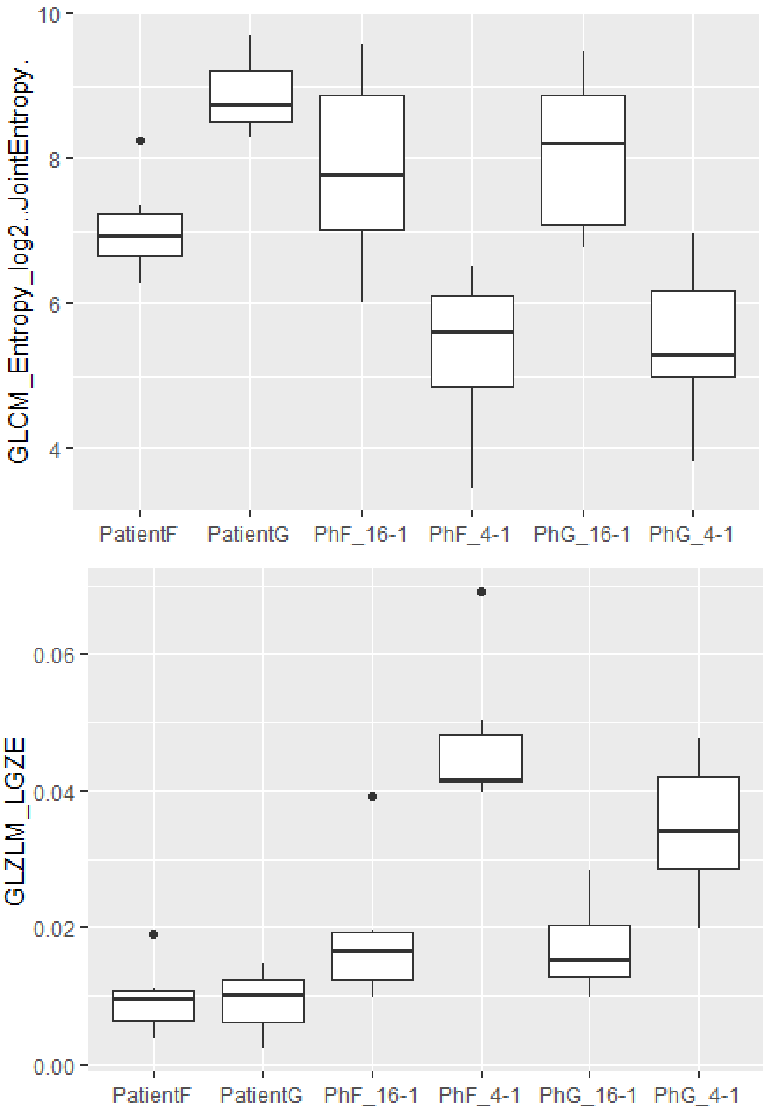
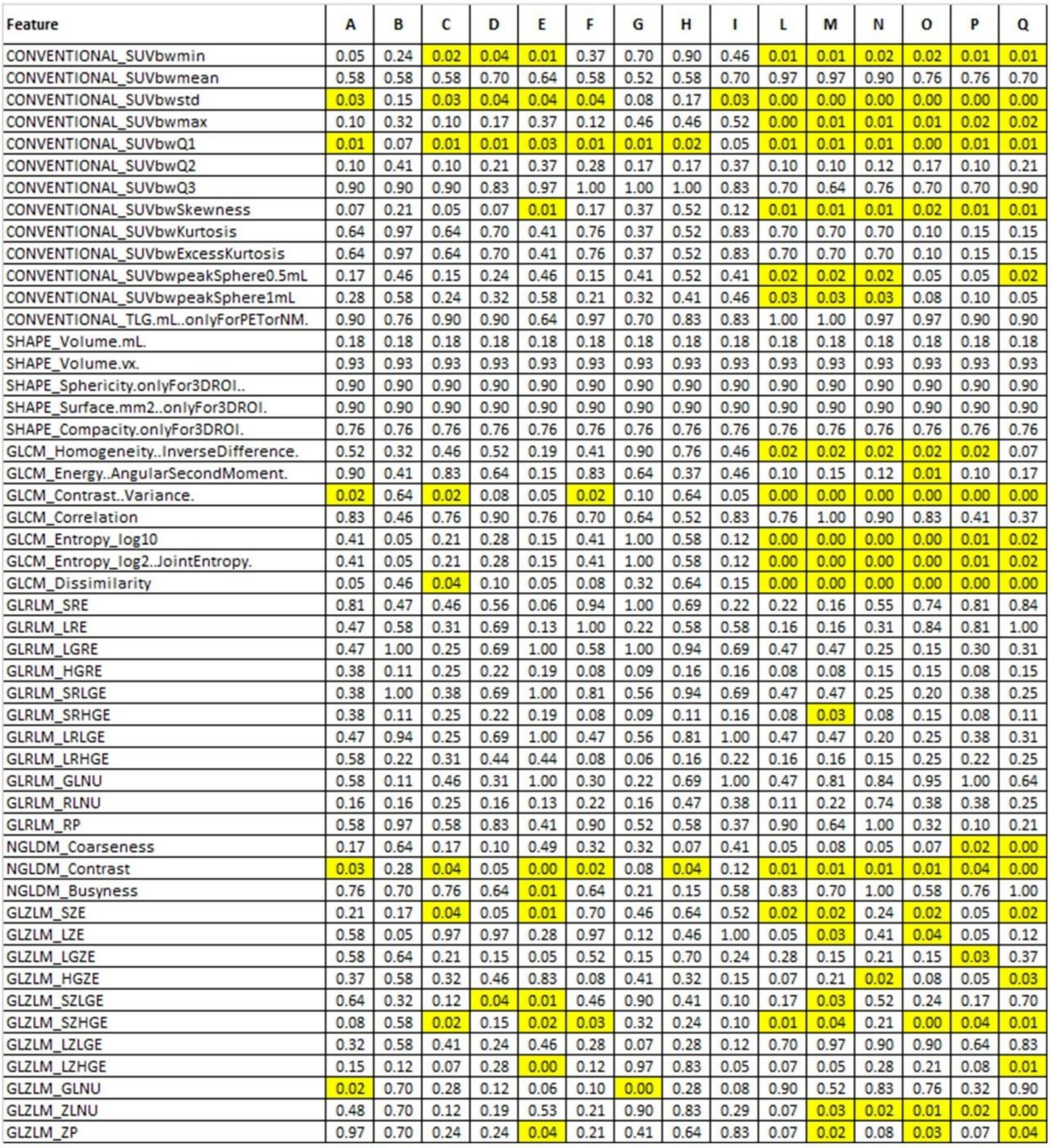


Figure 3.16: boxplot of two textural RFs (GLCM_entropy_log2 and GLZLM_LGZE) distributions, for 18F patients (PatientF), 68Ga patients (PatientG), 18F phantom with 4:1 (PhF_4-1) and 16:1 (PhF_16-1) concentration ratio and 68Ga phantom with the same concentration ratios (PhG_4-1 and PhG_16-1).

3.3.2 Comparison of RFs calculated with ^{68}Ga and ^{18}F filled inserts

To analyze the differences between the RF values calculated for the Ga- and F-filled phantom inserts, Mann Whitney tests were performed for pairs of samples filled with the two radioisotopes, separately for each of the acquisition modes. The corresponding p-value for different RF and for each acquisition mode are shown in figure 3.16 (p values below 0.05 are highlighted in yellow).



Feature	A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q
CONVENTIONAL_SUVbwmin	0.05	0.24	0.02	0.04	0.01	0.37	0.70	0.90	0.46	0.01	0.01	0.02	0.02	0.01	0.01
CONVENTIONAL_SUVbwmean	0.58	0.58	0.58	0.70	0.64	0.58	0.52	0.58	0.70	0.97	0.97	0.90	0.76	0.76	0.70
CONVENTIONAL_SUVbwstd	0.03	0.15	0.03	0.04	0.04	0.04	0.08	0.17	0.03	0.00	0.00	0.00	0.00	0.00	0.00
CONVENTIONAL_SUVbwmax	0.10	0.32	0.10	0.17	0.37	0.12	0.46	0.46	0.52	0.00	0.01	0.01	0.01	0.02	0.02
CONVENTIONAL_SUVbwQ1	0.01	0.07	0.01	0.01	0.03	0.01	0.01	0.02	0.05	0.01	0.01	0.01	0.00	0.01	0.01
CONVENTIONAL_SUVbwQ2	0.10	0.41	0.10	0.21	0.37	0.28	0.17	0.17	0.37	0.10	0.10	0.12	0.17	0.10	0.21
CONVENTIONAL_SUVbwQ3	0.90	0.90	0.90	0.83	0.97	1.00	1.00	1.00	0.83	0.70	0.64	0.76	0.70	0.70	0.90
CONVENTIONAL_SUVbwSkewness	0.07	0.21	0.05	0.07	0.01	0.17	0.37	0.52	0.12	0.01	0.01	0.01	0.02	0.01	0.01
CONVENTIONAL_SUVbwKurtosis	0.64	0.97	0.64	0.70	0.41	0.76	0.37	0.52	0.83	0.70	0.70	0.70	0.10	0.15	0.15
CONVENTIONAL_SUVbwExcessKurtosis	0.64	0.97	0.64	0.70	0.41	0.76	0.37	0.52	0.83	0.70	0.70	0.70	0.10	0.15	0.15
CONVENTIONAL_SUVbwpeakSphere0.5mL	0.17	0.46	0.15	0.24	0.46	0.15	0.41	0.52	0.41	0.02	0.02	0.02	0.05	0.05	0.02
CONVENTIONAL_SUVbwpeakSphere1mL	0.28	0.58	0.24	0.32	0.58	0.21	0.32	0.41	0.46	0.03	0.03	0.03	0.08	0.10	0.05
CONVENTIONAL_TLG.mL..onlyForPETorNM.	0.90	0.76	0.90	0.90	0.64	0.97	0.70	0.83	0.83	1.00	1.00	0.97	0.97	0.90	0.90
SHAPE_Volume.mL.	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
SHAPE_Volume.vx.	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
SHAPE_Sphericity.onlyFor3DROI..	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
SHAPE_Surface.mm2..onlyFor3DROI.	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
SHAPE_Compacity.onlyFor3DROI.	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
GLCM_Homogeneity..InverseDifference.	0.52	0.32	0.46	0.52	0.19	0.41	0.90	0.76	0.46	0.02	0.02	0.02	0.02	0.02	0.07
GLCM_Energy..AngularSecondMoment.	0.90	0.41	0.83	0.64	0.15	0.83	0.64	0.37	0.46	0.10	0.15	0.12	0.01	0.10	0.17
GLCM_Contrast..Variance.	0.02	0.64	0.02	0.08	0.05	0.02	0.10	0.64	0.05	0.00	0.00	0.00	0.00	0.00	0.00
GLCM_Correlation	0.83	0.46	0.76	0.90	0.76	0.70	0.64	0.52	0.83	0.76	1.00	0.90	0.83	0.41	0.37
GLCM_Entropy_log10	0.41	0.05	0.21	0.28	0.15	0.41	1.00	0.58	0.12	0.00	0.00	0.00	0.00	0.01	0.02
GLCM_Entropy_log2..JointEntropy.	0.41	0.05	0.21	0.28	0.15	0.41	1.00	0.58	0.12	0.00	0.00	0.00	0.00	0.01	0.02
GLCM_Dissimilarity	0.05	0.46	0.04	0.10	0.05	0.08	0.32	0.64	0.15	0.00	0.00	0.00	0.00	0.00	0.00
GLRLM_SRE	0.81	0.47	0.46	0.56	0.06	0.94	1.00	0.69	0.22	0.22	0.16	0.55	0.74	0.81	0.84
GLRLM_LRE	0.47	0.58	0.31	0.69	0.13	1.00	0.22	0.58	0.58	0.16	0.16	0.31	0.84	0.81	1.00
GLRLM_LGRE	0.47	1.00	0.25	0.69	1.00	0.58	1.00	0.94	0.69	0.47	0.47	0.25	0.15	0.30	0.31
GLRLM_HGRE	0.38	0.11	0.25	0.22	0.19	0.08	0.09	0.16	0.16	0.08	0.08	0.15	0.15	0.08	0.15
GLRLM_SRLGE	0.38	1.00	0.38	0.69	1.00	0.81	0.56	0.94	0.69	0.47	0.47	0.25	0.20	0.38	0.25
GLRLM_SRHGE	0.38	0.11	0.25	0.22	0.19	0.08	0.09	0.11	0.16	0.08	0.03	0.08	0.15	0.08	0.11
GLRLM_LRLGE	0.47	0.94	0.25	0.69	1.00	0.47	0.56	0.81	1.00	0.47	0.47	0.20	0.25	0.38	0.31
GLRLM_LRHGE	0.58	0.22	0.31	0.44	0.44	0.08	0.06	0.16	0.22	0.16	0.16	0.15	0.25	0.22	0.25
GLRLM_GLNU	0.58	0.11	0.46	0.31	1.00	0.30	0.22	0.69	1.00	0.47	0.81	0.84	0.95	1.00	0.64
GLRLM_RLNU	0.16	0.16	0.25	0.16	0.13	0.22	0.16	0.47	0.38	0.11	0.22	0.74	0.38	0.38	0.25
GLRLM_RP	0.58	0.97	0.58	0.83	0.41	0.90	0.52	0.58	0.37	0.90	0.64	1.00	0.32	0.10	0.21
NGLDM_Coarseness	0.17	0.64	0.17	0.10	0.49	0.32	0.32	0.07	0.41	0.05	0.08	0.05	0.07	0.02	0.00
NGLDM_Contrast	0.03	0.28	0.04	0.05	0.00	0.02	0.08	0.04	0.12	0.01	0.01	0.01	0.01	0.04	0.00
NGLDM_Busyness	0.76	0.70	0.76	0.64	0.01	0.64	0.21	0.15	0.58	0.83	0.70	1.00	0.58	0.76	1.00
GLZLM_SZE	0.21	0.17	0.04	0.05	0.01	0.70	0.46	0.64	0.52	0.02	0.02	0.24	0.02	0.05	0.02
GLZLM_LZE	0.58	0.05	0.97	0.97	0.28	0.97	0.12	0.46	1.00	0.05	0.03	0.41	0.04	0.05	0.12
GLZLM_LGZE	0.58	0.64	0.21	0.15	0.05	0.52	0.15	0.70	0.24	0.28	0.15	0.21	0.15	0.03	0.37
GLZLM_HGZE	0.37	0.58	0.32	0.46	0.83	0.08	0.41	0.32	0.15	0.07	0.21	0.02	0.08	0.05	0.03
GLZLM_SZLGE	0.64	0.32	0.12	0.04	0.01	0.46	0.90	0.41	0.10	0.17	0.03	0.52	0.24	0.17	0.70
GLZLM_SZHGE	0.08	0.58	0.02	0.15	0.02	0.03	0.32	0.24	0.10	0.01	0.04	0.21	0.00	0.04	0.01
GLZLM_LZLGE	0.32	0.58	0.41	0.24	0.46	0.28	0.07	0.28	0.12	0.70	0.97	0.90	0.90	0.64	0.83
GLZLM_LZHGE	0.15	0.12	0.07	0.28	0.00	0.12	0.97	0.83	0.05	0.07	0.05	0.28	0.21	0.08	0.01
GLZLM_GLNU	0.02	0.70	0.28	0.12	0.06	0.10	0.00	0.28	0.08	0.90	0.52	0.83	0.76	0.32	0.90
GLZLM_ZLNU	0.48	0.70	0.12	0.19	0.53	0.21	0.90	0.83	0.29	0.07	0.03	0.02	0.01	0.02	0.00
GLZLM_ZP	0.97	0.70	0.24	0.24	0.04	0.21	0.41	0.64	0.83	0.07	0.02	0.08	0.03	0.07	0.04

Table 3.5 – p-values of the Mann Whitney test of differences between 68Ga and 18F phantom filled inserts, for each combination of reconstruction parameters (p values below 0.05 are highlighted in yellow).

The number of RFs showing significant differences varies greatly, depending on the reconstruction mode used, ranging from 0 to 20. The mean, median (SuvbwQ2) and third quartile values of SUV have p-values greater than 0.05 for all the conditions, while the SUV max, standard deviation of SUV and first quartile in some combinations of parameters have a p-value below 0.05. For the same conditions the number of textural RFs with appreciable differences is also greater. In general, the number of features with significant differences between the two radioisotopes is greater in combinations with a high number of iterations or with a high number of subsets, while it is very low for the standard acquisition conditions of patients and phantoms for quality controls (mode A and G). The only condition for which significant differences are not appreciated is mode B which includes the correction for the point spread function.

3.3.3 Impact of different reconstruction parameters on RFs

For different acquisition times, the majority of RFs showed COV lower than 20%, as shown in figure 3.17.

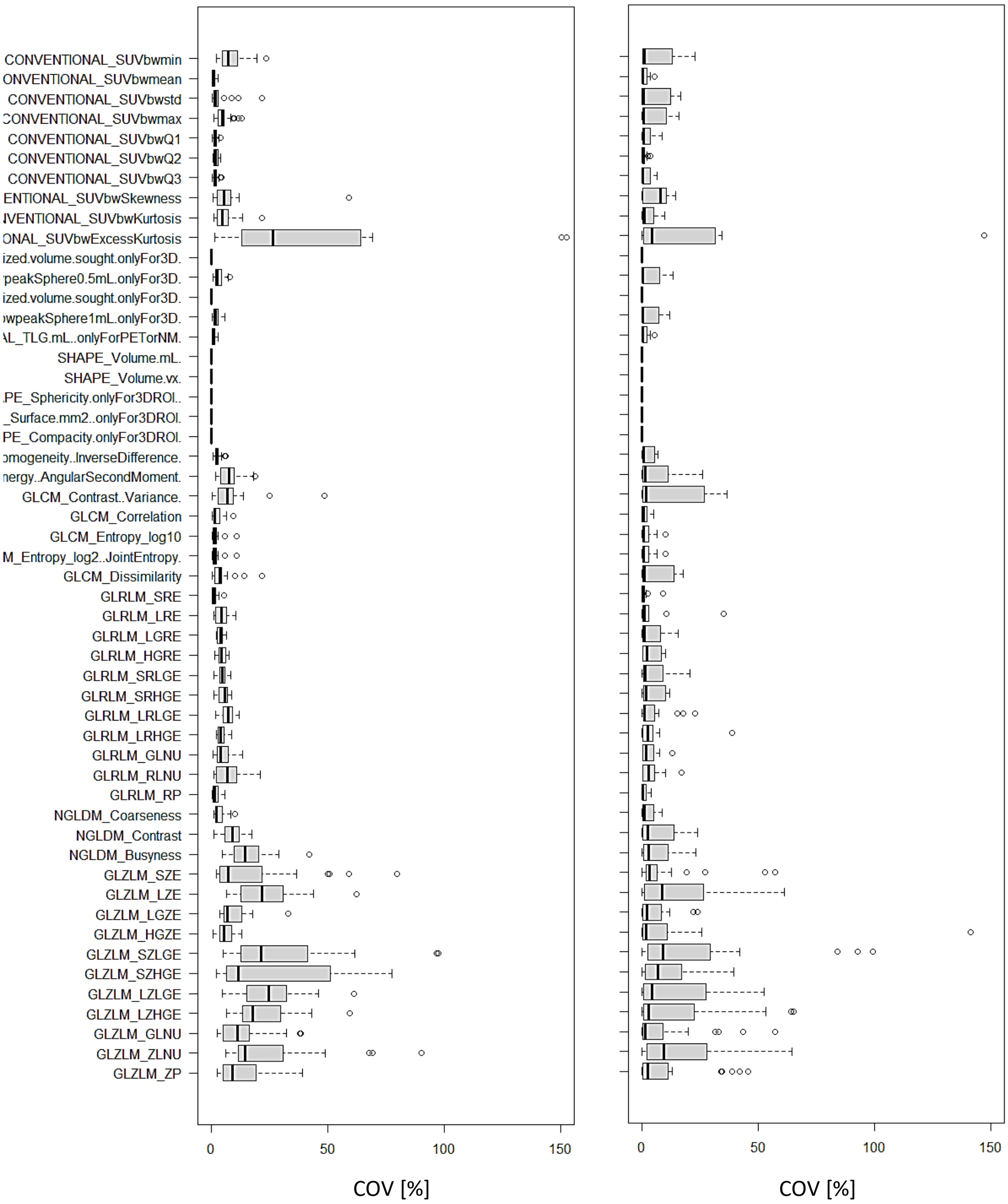


Figure 3.17 – COV distributions for the phantom inserts for different acquisition times (left) and choice of point spread function correction (right).

The only conventional feature with higher COV values was the excess kurtosis, that is the kurtosis value minus 3 and, as a consequence of having mean values close to zero, the COV assumes high values. Higher COV values were also observed for GLZLM features, in general below 50%. A similar behavior of COV values was assessed for the PSF parameter, with a general light increase for all the textural RFs including GLCM and GLRLM. These results highlight that features are in general more sensitive to spatial resolution than to noise, as the impact of different acquisition times is a light increase in noise, whereas the PSF correction has consequences on spatial resolution.

Plot of the feature values towards acquisition time highlight similar trends for ^{18}F and ^{68}Ga radionuclides. An example is shown in figure 3.18 for GLCM_correlation, but similar trends were observed for most of the features, with higher variations for the GLZLM RFs family. As shown, these features are almost constant with a very small increase from 1 minute of acquisition to 5 minutes, with a low impact of the related image noise differences.

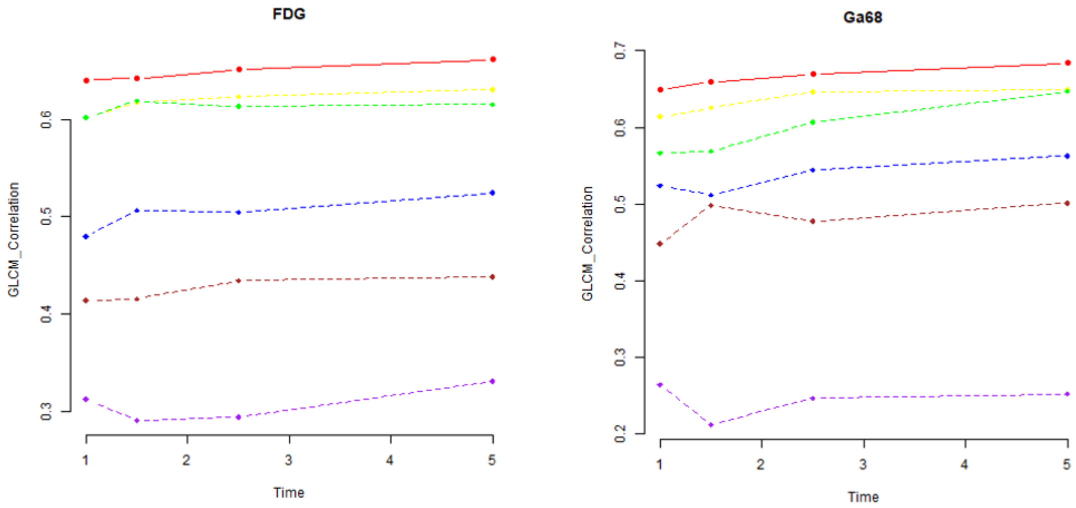


Figure 3.18 – GLCM correlation values as a function of the acquisition time (left: acquisition with FDG of the left, with Ga68 on the right. The colours correspond to different lesion volumes (lower values correspond to lower volumes).

The number of subsets had a major impact on RFs, in particular for the extreme case of 1 subset (condition E). Figure 3.19 show two examples: the conventional SUV max is affected by the choice of the number of subsets and the value increases with a maximum for about 30 subsets, whereas for other RFs a decreasing trend was observed.

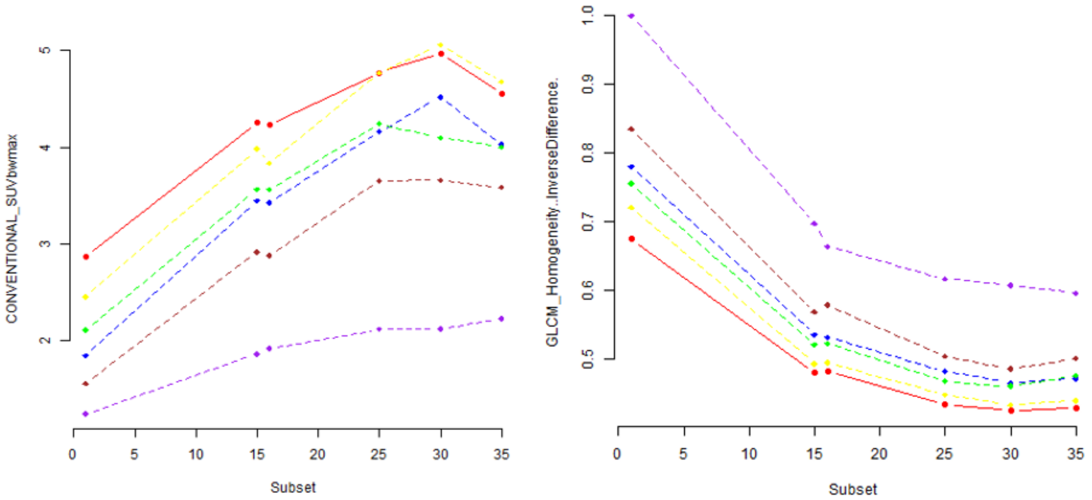


Figure 3.19 – Conventional SUV max and GLCM_homogeneity values towards the number of subset considered for FDG PET images reconstruction, for different radionuclides and lesion volume (lower values correspond to lower volumes).

In figure 3.20 the COV for all the subset conditions and for all except the E parameters combination is presented. For subset greater than 15 most of COV values are below 20%, whereas including the E case several RFs have a COV greater than 50% and some greater than 150%.

The ICC analysis highlighted in general values greater than 0.9 for most of the RFs and reconstruction parameters (Table 3.6). In general, it is possible to say that RFs extracted from the phantom test are in general “robust” with respect to the choice of PET reconstruction parameters. Low ICC values were observed in particular for GLZLM RFs and for the different number of subset and iteration, coherently with the COV distributions.

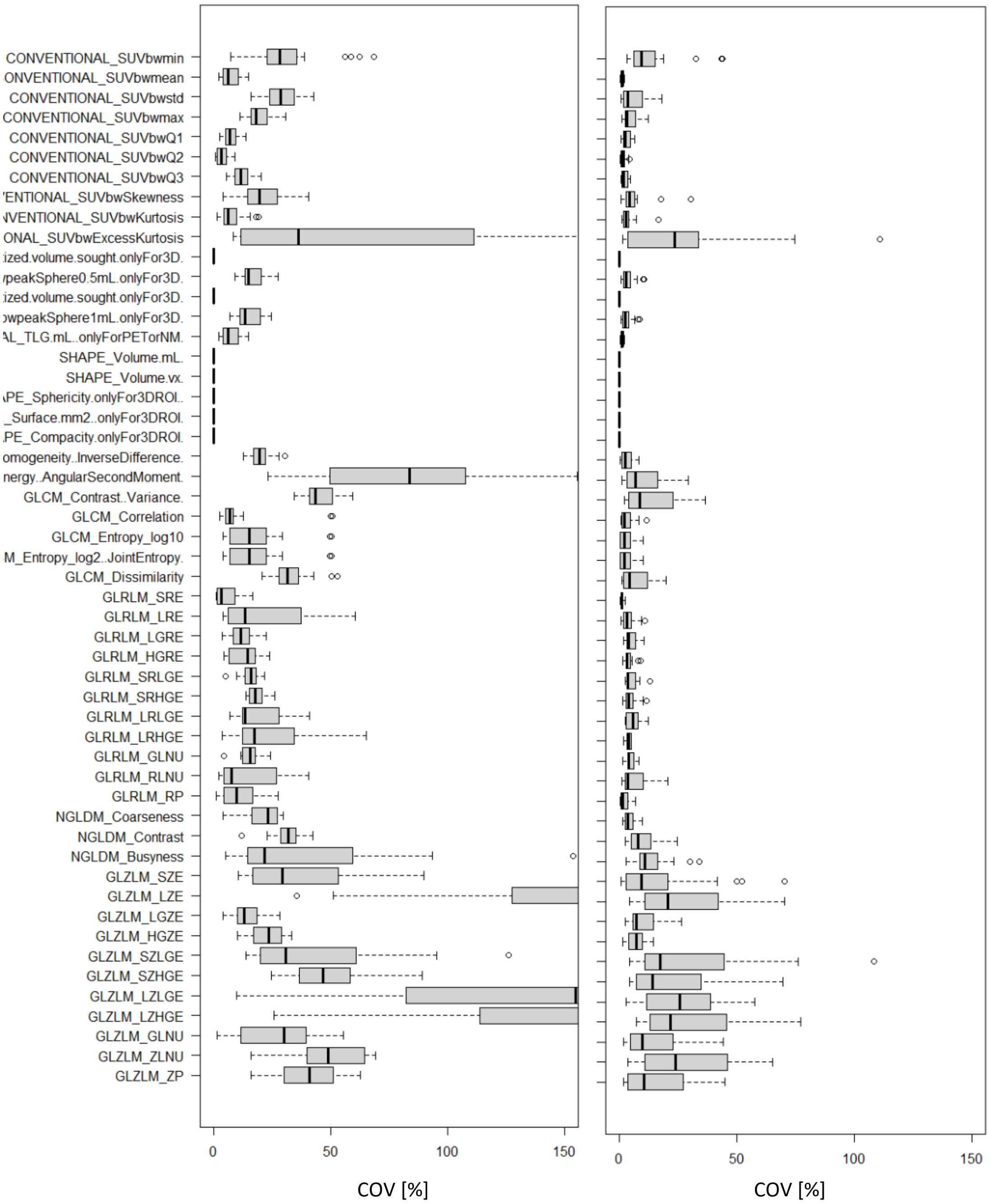


Figure 3.20 – COV distributions for the phantom inserts for different number of subsets: the right graph was obtained excluding the case of 1 subset (E parameters combination of table 3.2).

RF	Acq time	PSF	Subset	Subs -E	Iteration
CONVENTIONAL_SUVbwmin	0.87	0.91	0.39	0.83	0.72
CONVENTIONAL_SUVbwmean	1.00	1.00	0.97	1.00	1.00
CONVENTIONAL_SUVbwstd	1.00	0.98	0.88	1.00	1.00
CONVENTIONAL_SUVbwmax	0.99	0.98	0.89	0.99	0.99
CONVENTIONAL_SUVbwQ1	1.00	0.98	0.94	0.99	0.98
CONVENTIONAL_SUVbwQ2	1.00	1.00	0.99	1.00	1.00
CONVENTIONAL_SUVbwQ3	1.00	1.00	0.94	1.00	1.00
CONVENTIONAL_SUVbwSkewness	0.91	0.96	0.71	0.97	0.97
CONVENTIONAL_SUVbwKurtosis	0.91	0.97	0.82	0.96	0.97
CONVENTIONAL_SUVbwExcessKurtosis	0.91	0.97	0.82	0.96	0.97
CONVENTIONAL_SUVbwpeakSphere0.5mL	1.00	0.99	0.92	1.00	1.00
CONVENTIONAL_SUVbwpeakSphere1mL	1.00	0.99	0.94	1.00	1.00
CONVENTIONAL_TLG.mL..onlyForPETorNM.	1.00	1.00	0.99	1.00	1.00
SHAPE_Volume.mL.	1.00	1.00	1.00	1.00	1.00
SHAPE_Volume.vx.	1.00	1.00	1.00	1.00	1.00
SHAPE_Sphericity.onlyFor3DROI..	1.00	1.00	1.00	1.00	1.00
SHAPE_Surface.mm2..onlyFor3DROI.	1.00	1.00	1.00	1.00	1.00
SHAPE_Compacity.onlyFor3DROI.	1.00	1.00	1.00	1.00	1.00
GLCM_Homogeneity..InverseDifference.	0.99	0.99	0.71	0.98	0.95
GLCM_Energy..AngularSecondMoment.	0.96	0.95	0.18	0.92	0.72
GLCM_Contrast..Variance.	1.00	0.94	0.79	1.00	0.99
GLCM_Correlation	0.99	1.00	0.86	0.98	0.97
GLCM_Entropy_log10	0.99	0.99	0.79	0.99	0.97
GLCM_Entropy_log2..JointEntropy.	0.99	0.99	0.79	0.99	0.97
GLCM_Dissimilarity	1.00	0.97	0.82	0.99	0.99
GLRLM_SRE	0.93	0.93	0.60	0.96	0.90
GLRLM_LRE	0.93	0.77	0.35	0.94	0.85
GLRLM_LGRE	0.99	0.97	0.93	0.98	0.96
GLRLM_HGRE	1.00	0.99	0.95	1.00	1.00
GLRLM_SRLGE	0.98	0.94	0.84	0.97	0.94
GLRLM_SRHGE	1.00	0.99	0.94	1.00	1.00
GLRLM_LRLGE	0.96	0.94	0.76	0.98	0.97
GLRLM_LRHGE	1.00	1.00	0.95	1.00	1.00
GLRLM_GLNU	0.99	0.99	0.95	0.99	1.00
GLRLM_RLNU	0.99	1.00	0.97	0.99	0.99
GLRLM_RP	0.96	0.98	0.51	0.94	0.88
NGLDM_Coarseness	0.99	1.00	0.89	0.99	0.99
NGLDM_Contrast	0.99	0.98	0.77	0.99	0.97
NGLDM_Busyness	0.93	0.98	0.19	0.95	0.91
GLZLM_SZE	0.88	0.96	0.66	0.88	0.90
GLZLM_LZE	0.86	0.89	0.07	0.80	0.56
GLZLM_LGZE	0.94	0.97	0.90	0.93	0.91
GLZLM_HGZE	1.00	0.00	0.00	0.00	0.00
GLZLM_SZLGE	0.33	0.12	0.35	0.53	0.26
GLZLM_SZHGE	0.99	0.97	0.86	0.99	0.99
GLZLM_LZLGE	0.86	0.87	0.11	0.80	0.62
GLZLM_LZHGE	0.89	0.90	0.01	0.00	0.00
GLZLM_GLNU	0.99	0.96	0.00	0.00	0.11
GLZLM_ZLNU	0.99	0.99	0.00	0.00	0.20
GLZLM_ZP	0.99	0.99	0.81	0.98	0.98

Table 3.6– ICC values for RFs towards different combination of reconstruction parameters (dark green ICC>0.9, light green 0.8<ICC<0.9, orange 0.7<ICC<0.8, yellow 0.5<ICC<0.7, pink ICC<0.5).

Relative trends of RFs as a function of the number of iterations used for the PET reconstruction showed in most cases a large variation when this number varies between

3 and 5 and low changes for a number of iterations between 5 and 10 iterations, regardless of the volume of the spheres (Figure 3.21). COV values are not shown, but they are similar to the ones obtained for different subset without the "E" parameter combination (Figure 3.20 right).

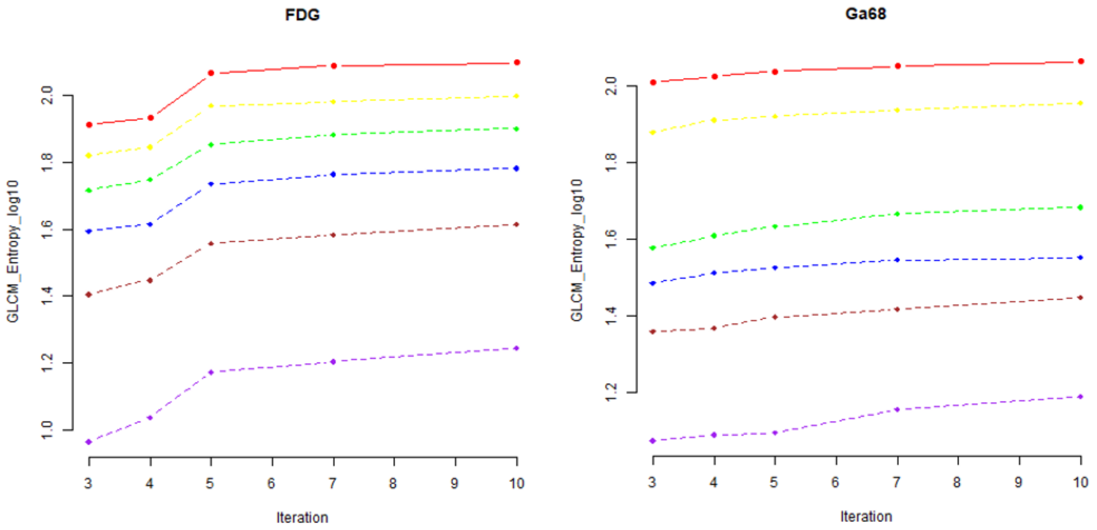


Figure 3.21 – GLCM_Entropy trend towards different number of iterations used for PET reconstruction, for ^{18}F and ^{68}Ga.

3.3.4 Correlation between RFs and lesion volumes

Table 3.7 shows the Pearson correlation coefficients of single features with respect to the insert volumes, including the acquisitions with the same radionuclide and two different concentrations or with the same concentrations and the two radionuclides. When the same concentration is considered, a large number of RFs (more than 60% excluding shape features) correlates with volume. On the other hand, when two different concentrations were considered, only two textural RFs (GLRLM_RLNU and GLZLM_GLNU) are correlated with lesion volumes for both ^{18}F and ^{68}G.

3.4 Discussion

In this study, a three-dimensional phantom was implemented to evaluate the reproducibility of radiomic features of neuroendocrine tumors with different PET image reconstruction modalities and different radionuclides. Although the use of phantoms made with 3D printers is also growing in the field of nuclear medicine, no previous study has focused on this type of tumor and on the comparison between two radioisotopes with different concentrations.

Feature	F_G_4	F_G _16	F_4_16	G_4_16
CONVENTIONAL_SUVbwmin	-0.39	-0.19	-0.45	-0.27
CONVENTIONAL_SUVbwmean	0.92	0.95	0.55	0.53
CONVENTIONAL_SUVbwstd	0.93	0.89	0.46	0.45
CONVENTIONAL_SUVbwmax	0.89	0.85	0.39	0.42
CONVENTIONAL_SUVbwQ1	0.49	0.92	0.61	0.59
CONVENTIONAL_SUVbwQ2	0.87	0.95	0.61	0.55
CONVENTIONAL_SUVbwQ3	0.96	0.96	0.58	0.54
CONVENTIONAL_SUVbwSkewness	-0.44	-0.87	-0.87	-0.44
CONVENTIONAL_SUVbwKurtosis	-0.84	-0.77	-0.82	-0.79
CONVENTIONAL_SUVbwExcessKurtosis	-0.84	-0.77	-0.82	-0.79
CONVENTIONAL_SUVbwpeakSphere0.5mL.	0.91	0.91	0.46	0.47
CONVENTIONAL_SUVbwpeakSphere1mL	0.92	0.91	0.49	0.50
CONVENTIONAL_TLG.mL..onlyForPETorNM.	0.99	0.99	0.84	0.86
SHAPE_Volume.mL.	1.00	1.00	1.00	1.00
SHAPE_Volume.vx.	1.00	1.00	1.00	1.00
SHAPE_Sphericity.onlyFor3DROI..	-0.79	-0.90	-0.74	-0.89
SHAPE_Surface.mm2..onlyFor3DROI.	0.99	1.00	0.99	1.00
SHAPE_Compacity.onlyFor3DROI.	0.97	0.97	0.97	0.97
GLCM_Homogeneity..InverseDifference.	-0.83	-0.69	-0.29	-0.35
GLCM_Energy..AngularSecondMoment.	-0.67	-0.73	-0.43	-0.53
GLCM_Contrast..Variance.	0.92	0.79	0.36	0.30
GLCM_Correlation	0.73	0.81	0.82	0.70
GLCM_Entropy_log10	0.84	0.91	0.58	0.57
GLCM_Entropy_log2..JointEntropy.	0.84	0.91	0.58	0.57
GLCM_Dissimilarity	0.88	0.76	0.30	0.30
GLRLM_SRE	0.55	0.86	0.29	0.48
GLRLM_LRE	-0.30	-0.81	-0.25	-0.27
GLRLM_LGRE	-0.65	-0.86	-0.64	-0.73
GLRLM_HGRE	0.93	0.94	0.63	0.57
GLRLM_SRLGE	-0.64	-0.85	-0.67	-0.74
GLRLM_SRHGE	0.93	0.94	0.62	0.57
GLRLM_LRLGE	-0.51	-0.89	-0.46	-0.57
GLRLM_LRHGE	0.92	0.95	0.67	0.61
GLRLM_GLNU	0.96	0.99	0.69	0.71
GLRLM_RLNU	0.98	0.99	0.95	0.90
GLRLM_RP	0.85	0.51	0.24	0.32
NGLDM_Coarseness	-0.77	-0.86	-0.70	-0.76
NGLDM_Contrast	0.93	0.81	0.33	0.30
NGLDM_Busyness	0.92	0.08	0.48	0.48
GLZLM_SZE	0.81	0.57	0.35	0.34
GLZLM_LZE	0.45	-0.31	0.26	0.07
GLZLM_LGZE	-0.62	-0.85	-0.37	-0.49
GLZLM_HGZE	0.83	0.92	0.50	0.48
GLZLM_SZLGE	0.17	-0.67	-0.02	-0.19
GLZLM_SZHGE	0.78	0.90	0.49	0.43
GLZLM_LZLGE	0.30	-0.45	0.19	-0.02
GLZLM_LZHGE	0.79	0.99	0.54	0.34
GLZLM_GLNU	0.97	0.99	0.90	0.85
GLZLM_ZLNU	0.98	0.99	0.65	0.57
GLZLM_ZP	0.61	0.55	0.16	0.12

Table 3.7 – Pearson correlation coefficients between RFs and lesion volumes. The table columns refer to the following data included in the correlation analysis: F_G_4 ^{18}F and ^{68}Ga with concentration ratio 4:1, F_G_16 ^{18}F and ^{68}Ga with concentration ratio 16:1, F_4_16 ^{18}F with concentration ratio 4:1 and 16:1 and G_4_16 ^{68}G with concentration ratio 4:1 and 16:1. Values above 0.8 or below -0.8 are highlighted in grey.

A similar approach to the one followed in this study was adopted by [Pfaheler 2020] to simulate non-small cell lung cancer, with the fabrication of three different inserts of similar size: one simple, one with two compartments with different concentrations and one with an internal necrotic region. The possible effect of the lesions size was not examined (as instead it was decided to do in this study), while the acquisitions were carried out 4 times on the same scanners and once on other 5 pet scanners. The results are reported in terms of percentage of features with ICC values higher than 0.9 or in other lower ranges, for images obtained from the same acquisition on the same scanner with different noise patterns (reliability), repeated acquisitions on the same scanner (repeatability) and acquisitions on different scanners. All analysis were replicated with fixed bin number (FBN) and fixed bin width (FBW) discretization. For reliability, the percentage of features with ICC above 0.9 is generally higher than 90%, for repeatability it falls in many cases below 40% and for reproducibility it is generally below 20%. In the multicentric evaluation, the results improve if the acquisition criteria defined by the European Nuclear Medicine Research LTD (EARL) [Leijenaar 2015] are followed. The ICC values were lower with the FBN discretization than with the FBW. In our study percentages of features with very high ICC were observed for the different reconstructions used on the same scanner, in analogy to what reported by Pfaheler et al in the reliability section with reconstructions with different noise patterns. At the moment we do not have repeatability and reproducibility data that will be analyzed in the next phases of the study.

The two different radioisotopes used to fill the phantom inserts in this study are both also used in PET examinations of neuroendocrine tumor patients. ^{68}Ga and ^{18}F have a different positron range (PR), that is the distance traveled by the positrons from the point of emission to the annihilation point and depends on the energy of the emitted positron and the electron density of the surrounding medium. In the case of 18F, the positron emission energies are relatively low (maximum positron energy: E_{max} = 0.63 MeV, mean positron energy: E_{mean} = 0.25 MeV), and the mean PR (r_{mean}) in water is only 0.6 mm. This does not induce considerable differences between the measured and true tracer distribution [Alessio 2008], given the spatial resolution of state-of-the-art PET systems ranges from 2 to 4 mm. ^{68}Ga has higher positron emission energies and more complex decay schemes (Emax = 1.9 MeV, r_{mean} = 2.9 mm) and this can lead to deterioration of the perceived spatial resolution resulting in image blurring and loss of image contrast. The consequences of the different PR on image quality and an approach for possible correction were recently analyzed in a study [Kertesz 2022], using the classical NEMA phantom. The greatest differences were observed for spheres smaller than 20 mm and for high values of the number of iterations. In our study the lesions have a minimum dimension greater than 20 mm, although with a necrotic nucleus, but also in our analysis greater differences were highlighted between the features of inserts filled with ^{68}Ga and ^{18}F for the combinations of reconstruction parameters with high number of iterations and subsets, which emphasize the spatial resolution of the image at the expense of a greater presence of noise. Therefore, in the case of multicentric PET radiomics studies with the use of three-dimensional phantoms, it is conceivable that the results obtained with ^{18}F can also be extended to ^{68}Ga in the case of low or intermediate interaction number values, while it is certainly advisable to use a specific radionuclide for reconstructions with greater spatial resolution.

The use of the same shape of lesion with different dimensions allowed to analyze the correlation of the different features with the volume. When similar concentration ratio was used to fill the inserts, in most cases high values of the Pearson coefficients were found. When RFs calculated with different concentration ratios were considered, only two textural RFs show high correlation towards volume. These results can be compared with the analysis of correlation performed in the previous chapter for the two patients with different PRRT outcome (Figures 2.13 and 2.14). In particular it is possible to observe that the two RFs without correlation with volume in all the conditions of figure 3.23 (GLZLM_SZLGE and GLZLM_ZP) were also not correlated with volume in the PRRT patient analysis, with significant differences between responding and not responding lesions.

This study has several limitations, which will be addressed in the continuation of the work. In particular, due to various problems of a practical nature and of access to the PET equipment with high workloads of ordinary examinations, at the time of writing this thesis it was not possible to carry out repeated acquisitions in the same conditions of the phantoms to evaluate the repeatability of the values of obtainable RFs. These evaluations are planned for the coming months and will provide important information on the possibility of using the phantom in multi-center studies. Another limitation is linked to the use of only one shape of insert: in this work we have focused on the differences between fluorine and gallium and on the impact of different lesion sizes, with results of interest in this area. In the future it will be possible to evaluate other forms of NET lesions in particular by creating them starting from patients with different clinical variables of interest.

In conclusions, these implemented and user-friendly 3D-printed inserts for NEMA-phantom allow to evaluate the impact of different acquisition and reconstruction parameters on RFs. This approach appears promising for the standardization of PET protocols and the harmonization of data in multicenter radiomic studies.

4. Lessons learned and future perspectives

When undertaking a study in the field of radiomics, various aspects must be taken into consideration, in particular at a general level the dataset available or to be collected, the software used, the level of experience of the various professional figures involved in the study (both medical and techniques), the presence of evidence already published in the field and the questions to which the study undertaken intends to provide answers.

What is a relevant question in a radiomics study? A relevant question for physicians is usually a question of interest for patient management or for better understanding a disease. For physicists (or computer scientists or engineers) the questions afforded are more often related to aspects that impact on the use of radiomics in general or in specific applications, in particular about methodological aspects. The relevance of the question considered in the radiomics study should always be verified by comparison with the literature but also with discussions with physicians and other scientists. When reviewing the literature about the subject, many useful information is often provided in the supplementary data of the articles, because radiomics studies implies commonly a lot of data results and only a part of them are published in the main part of the paper. Particular attention should be paid to the homogeneity of the data (recruitment, scanner, acquisition protocol, reconstruction protocol) and if the author assesses the validity of their conclusion on an independent dataset.

In general, it is important to verify that the question of the study cannot be answered with already known other simple quantities extracted from the images, such as SUV in nuclear medicine PET and lesion volumes. As an example, in [Welch 2019] a radiomic model was fitted and externally validated using features extracted from previously reported lung and head and neck (H&N) cancer datasets using gross-tumour-volume contours, as well as from images with randomly permuted voxel index values (images without meaningful texture). To determine if the model added benefit, the prognostic accuracy of tumour volume alone was also calculated as a baseline. It was determined that the radiomics signature was a surrogate for tumour volume and that intensity and texture values were not pertinent for prognostication. Many similar examples of misinterpretation have been published.

It is also important to publish "negative results", in the sense of studies where the radiomics analysis did not provide additional information to commonly used quantities. As an example, in another study performed during this doctoral work [Thuiller 2019], we investigated the diagnostic value of both conventional parameters and RFs to distinguish lung NETs, including typical (TC) and atypical carcinoid (AC), from lung neuroendocrine carcinomas (Lu-NECs), including large-cell neuroendocrine carcinomas (LCNECs) and small-cell neuroendocrine carcinomas (SCLCs). We found that RFs do not provide additional information allowing us to discriminate Lu-NECs and Lu-NETs. Only HISTO_Entropy_log10 was selected by a LASSO regression, but it was highly correlated to conventional PET parameters. In this example these results may be explained by the characteristics of the considered tumor sample. First, even if several NETs might present high FDG-avidity, while several NECs hold low FDG-avidity, the broad difference in SUV values in NECs compared with NETs could lead to a broader difference in RF values between the two groups [Orlhac 2015]. Moreover, the volume of tumors in the considered cohort was low (median value of MTV = 7.52 mL) and several studies reported that a radiomic approach does not provide additional information when the lesion metabolic volume is lower than 10 mL [Hatt 2015]. In any case, the publication of these results is useful to the scientific community to select fields where more investigations about radiomics are promising from other where there are not expectable added values.

In the field of radiomics, differently from most of other science subjects, reproducing results obtained from other research groups are at least as important as producing new results. At present, most of the radiomics studies are not confirmed by other teams, in particular for predictive models, and as a consequence there is not translation in clinical practice. This is probably the greatest roadblock for advancing in the field.

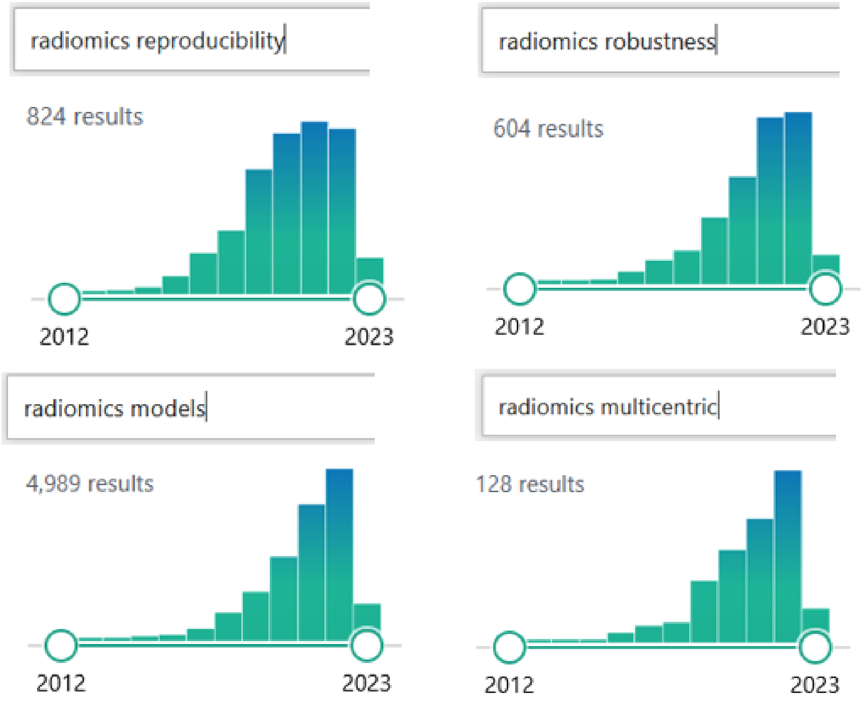
Several studies investigated the potential robustness of RFs, intended as maintenance of information provided by the RF in case of evaluation in different moments (repeatability) and conditions (reproducibility). In particular [Zwanenburg 2019] presents a meta-analysis with evaluation of the impact of single factors affecting reproducibility of image biomarkers:

- patient factors (disease type and site, injected radionuclide, injected tracer activity, tumour motion, breathing patterns);
- image acquisition factors (tracer uptake time, scan duration, 4D breathing frames, static 3D vs gated 4D, scanner differences, test-retest repeatability);
- image reconstruction (reconstruction method, number of iterations, number of subsets, gaussian filter width, partial volume corrections, voxel dimension difference, voxel harmonization);
- segmentation (delineation variability, manual or automatic or semi-automatic method)
- image processing (SUV normalization, image interpolation method, voxel harmonization, factors affecting reproducibility of image biomarkers, discretization method, discretization levels, added noise sensitivity)
- feature computation (texture matrix aggregation and distances, software used)

Several of these factors were investigated in this doctoral thesis and useful information about the consequences on RFs were collected, for both in patient and in phantom studies.

We haven’t the possibility of investigating scanner differences in PET images, although the first part of the study with patient data was performed on an old PET/CT and the second part with phantom on a new digital PET/CT installed in march 2021. Another part of doctoral work not reported in this thesis was focused on radiomics analysis on MRI images of breast triple negative cancers. In that case we have different MRI sequences acquired on two different scanners and a sample of about 40 patients with different response to neo adjuvant chemotherapy. We found that some features were affected by the scanner and other features showed significative differences between responding and not responding patients independently of the scanner used.

Although most of the robustness studies show results in terms of ICC and COV of RFs values for different factors analyzed on patients' samples or phantoms data, as also done in this thesis, the future investigations should focus on samples with patients subdivided according to a clinical variable of interest, with the definition of a model and the relative impact of procedural factors on the model. In other words we have to move from a selection of robust RFs to the identification of robust models. This require the collection of more numerous patient samples, multicentric studies with high quality data and this is still a great challenge at present, and probably we have to wait some years before to see relevant progresses. Most of the published studies are still monocentric or with few centers and number of cases rarely above one hundred. In any case, with some simple queries on Pubmed it is possible to see that the trends of research are actually going in this direction.



radiomics reproducibility

824 results

radiomics robustness

604 results

radiomics models

4,989 results

radiomics multicentric

128 results

Figure 4.1 – Number of Pubmed researches as a function of the publication year extracted with different keywords connected to radiomics: reproducibility, robustness, models and multicentric (searches of February 26th, 2023).

Literature research with keyword “reproducibility” combined with radiomics shows a total of 824 studies with a plateau in the period 2020-2022. Also the research with “robustness” combined with radiomics shows an increase of studies up to 2021 with a similar number in 2022. On the contrary, the keyword “models” with radiomics shows still a progressive increase in the last three years and probably the number of 2023 studies will be greater than 2022. Also, the number of multicentric studies is in progressive increase, although the total number is still very poor and represents less than 2% of the total number of studies.

A systematic review of radiomics in PET nuclear medicine imaging was recently published in two parts [Morland 2022a and b]. Seven studies were found to be conducted on neuroendocrine tumors [Bevilacqua 2021, Mapelli 2020, Thuillier 2019, Werner 2019, Atkinson 2021, Werner 2017, Weber 2020], half of which on pancreatic neuroendocrine tumors. The radiotracers used were 68Ga-DOTA-peptides (6/7) and 18F-FDG (1/7). Four studies aimed at predicting prognosis and four were conducted for diagnostic purposes, particularly for Ki67 prediction.

Study	Aim	Patient sample size	Relevant features
[Bevilacqua 2021]	Predicting grade 1 (G1) and 2 (G2) primary pancreatic neuroendocrine tumour	51	NH_{GLCM}, E_{GLCM}
[Mapelli 2020]	Predicting tumour aggressiveness and outcome in patients candidate to surgery for pancreatic neuroendocrine neoplasms	61	Intensity variability, size zone variability, entropy, homogeneity
[Thuillier 2019]	Differentiate among different histological subtypes of lung neuroendocrine neoplasms	44	No added value of RFs towards conventional quantities
[Werner 2019]	Predicts Overall Survival in Pancreatic Neuroendocrine Tumor Patients Undergoing Peptide Receptor Radionuclide Therapy	31	GLCM entropy
[Atkinson 2021]	Prognostic potential of tumor heterogeneity and tracer avidity in NET	44	First order kurtosis, skewness, and entropy
[Werner 2017]	Prognostic value of tumor heterogeneity as assessed by somatostatin receptor	141	Entropy, Correlation, Short Zone Emphasis and Homogeneity
[Weber 2020]	Prediction of proliferative activity of NETs, potentially allowing non-invasive tumor grading.	100	Entropy and homogeneity

Table 4.1 – Aims and relevant features for recent studies about radiomics in NET patients.

It is not easy to compare the relevant features with significative differences towards the clinical variable of interest for these studies, as there are in general few details about the method used to calculate the features, the software and the compliance with IBSI standard. In general, “entropy” is always present, in some studies it is specified that it is a GLCM entropy and in others no. In the future, it should be always stated the software used for RF calculation and the IBSI compliance, in order to facilitate the comparison of results of different studies. Moreover, the identification of relevant RFs calculated in a standard manner, will open the road to implement robustness studies focused on the relevant biomarkers and on their values for different patient populations. This should be a shift in perspective from the last few years of RFs computing binges towards a more selective approach to single significant RFs, with thorough investigation of their behavior.

Phantoms studies, in my opinion, will still have a role in the next years of radiomics analysis. In this work we made several attempts to obtain our first inserts suitable for in phantom investigation of RFs reproducibility and at the end we implemented a phantom with RFs results comparable with real patients and with reproducibility results coherent with findings of other studies. It is desirable also in this field to move towards a standardization of phantom implementations for radiomics studies, such as a common 3D printing approach and a shared library of 3D printing models, to facilitate the comparison of the results among centers.

Conclusions

At the end of this journey, I can confirm with greater awareness that the workflow of a radiomics study includes various tasks which commonly belong to the knowledge and skills of the specialist in medical physics: from image quality assurance to image processing, from segmentation techniques to statistical processing of results, from phantom implementation to experimental management.

It was possible to obtain the results presented here thanks to an interdisciplinary work involving in particular nuclear medicine physicians, engineers and physicists. The various analysis of real NET patient data and related publications have been published on relevant journals, and several subsequent studies have cited the results. The specific information in terms of the semi-automatic approach to segmentation without threshold and the use of fixed discretization will also be useful to set up the next studies correctly from the beginning and also to define standardized approaches.

The phantom study highlighted in particular the possibility of making inserts for phantoms that are already commercially available with low costs and good representation of the radiomic data of real tumor lesions. The first results are encouraging and have allowed us to formulate interesting conclusions on the relative effects of the size of the lesions, the type of radioisotope used and the effect of various image reconstruction parameters. The natural evolution of the next phases of the study will include a verification of the repeatability of the method, its use on different scanners and the creation of inserts with different shapes and internal heterogeneities.

References

[Alessio 2008] Alessio, A., and MacDonald, L. (2008). Spatially variant positron range modeling derived from CT for PET image reconstruction. IEEE Nucl. Sci. Symp. Conf. Rec. 2008, 3637–3640. doi: 10.1109/NSSMIC.2008.4774106

[Altazi 2017] Altazi BA, Zhang GG, Fernandez DC, Montejo ME, Hunt D, Werner J, et al. Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. J Appl Clin Med Phys. 2017;18:32–48.

[Atkinson 2021] Atkinson, C.; Ganeshan, B.; Endozo, R.; Wan, S.; Aldridge, M.D.; Groves, A.M.; Bomanji, J.B.; Gaze, M.N. Radiomics-Based Texture Analysis of (68)Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With (177)Lu-DOTATATE. Front. Oncol. 2021, 11, 686235.

[Bailly 2016] Bailly C, Bodet-Milin C, Couespel S, Necib H, Kraeber-Bodéré F, Ansquer C, et al. Revisiting the robustness of PET-based textural features in the context of multi-centric trials. PLoS One. 2016;11:154–91.

[Bashir 2017] Bashir, Azad G, Siddique MM, Dhillon S, Patel N, Bassett P, et al. The effects of segmentation algorithms on the measurement of 18F-FDG PET texture parameters in non-small cell lung cancer. EJNMMI Res. 2017;7.

[Berthon 2015] Berthon B, Marshall C, Holmes R, Spezi E. A novel phantom technique for evaluating the performance of PET auto-segmentation methods in delineating heterogeneous and irregular lesions. EJNMMI Phys. 2015;2:13.

[Bevilacqua 2021] Bevilacqua, A.; Calabrò, D.; Malavasi, S.; Ricci, C.; Casadei, R.; Campana, D.; Baiocco, S.; Fanti, S.; Ambrosini, V. A [68Ga]Ga-DOTANOC PET/CT Radiomic Model for Non-Invasive Prediction of Tumour Grade in Pancreatic Neuroendocrine Tumours. Diagnostics 2021, 11, 870.

[Biehl 2006] Biehl KJ, Kong FM, Dehdashti F, Jin JY, Mutic S, El Naqa I, et al. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? Society of Nuclear Medicine. 2006;47:1808–12.

[Binderup 2010] Binderup T, Knigge U, Loft A, Federspiel B, Kjaer A. 18F-fluorodeoxyglucose positron emission tomography predicts survival of patients with neuroendocrine tumors. Clin Cancer Res. Clin Cancer Res. 2010;16:978–85.

[Bozkurt 2017] Bozkurt MF, Virgolini I, Balogova S, Beheshti M, Rubello D, DecristoforoC, et al. Guideline for PET/CT imaging of neuroendocrine neoplasms with 68Ga-DOTA-conjugated somatostatin receptor targeting peptides and 18F–DOPA. Eur J Nucl Med Mol Imaging. (2017) 44:1588–601. doi: 10.1007/s00259-017-3728-y

[Carideo 2019] Carideo L, Prosperi D, Panzuto F, Magi L, Pratesi MS, Rinzivillo M, et al. Role of Combined [68Ga]Ga-DOTA-SST analogues and [18F]FDG PET/CT in the management of GEP-NENs: a systematic review. J Clin Med. 2019 [cited 2019 Dec 11];8:1032. Available from: https://www.mdpi.com/2077-0383/8/7/1032.

[Carles 2017] Carles M, Torres-Espallardo I, Alberich-Bayarri A, et al. Evaluation of PET texture features with heterogeneous phantoms: complementarity and effect of motion and segmentation method. Phys Med Biol. 2017;62:652–668.

[Cerviño 2017] Cerviño L, Soultan D, Cornell M, et al. A novel 3D-printed phantom insert for 4D PET/CT imaging and simultaneous integrated boost radiotherapy. Med Phys. 2017;44:5467–5474.

[Chan 2001[Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Process. 2001;10:266–77.

[Chan 2017] DLH, Pavlakis N, Schembri GP, Bernard EJ, Hsiao E, Hayes A, et al. Dual somatostatin receptor/FDG PET/CT imaging in metastatic neuroendocrine tumours: Proposal for a novel grading scheme with prognostic significance. Theranostics. 2017;7:1149–58.

[Forgacs 2016] Forgacs A, Pall Jonsson H, Dahlbom M, et al. A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images. PLoS ONE. 2016;11:e0164113.

[Forgacs 2019] Forgacs A, Kallos-Balogh P, Nagy F, et al. Activity painting: PET images of freely defined activity distributions applying a novel phantom technique. PLoS ONE. 2019;14:e0207658.

[Foster 2014] Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. Comput. Biol. Med; 2014. p. 76–96.

[Galavis 2010] Galavis PE, 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 Taylor & Francis. 2010;49:1012–6.

[Gerlinger 2019] Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366:883–92 [cited 2019 Dec 11] Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1113205.

[Giovannini 2018] Giovannini E, Giovacchini G, Borsò E, Lazzeri P, Riondato M, Leoncini R, et al. [68 Ga]-Dota Peptide PET/CT in Neuroendocrine Tumors: Main Clinical Applications. Curr Radiopharm. 2018;12:11–22. [cited 2019 Dec 11] Available from: http://www.eurekaselect.com/168288/article

[Granzier 2020] R. Granzier, N. Verbakel, A. Ibrahim, J. Van Timmeren, T. Van Nijnatten, R. Leijenaar, M. Lobbes, M. Smidt, and H. Woodruff, “Mri-based radiomics in breast cancer: Feature robustness with respect to inter-observer segmentation variability,” Scientific reports, vol. 10, no. 1, pp. 1–11, 2020.

[Grillo 2016] Grillo F, Albertelli M, Brisigotti MP, Borra T, Boschetti M, Fiocca R, et al. Grade increases in gastroenteropancreatic neuroendocrine tumor metastases compared to the primary tumor. Neuroendocrinol. 2016; 103:452–9.

[Halt 2009] Halt M, Le Rest CC, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET. IEEE Trans Med Imaging. 2009;28:881–93.

[Hamberg 1994] Hamberg LM, Hunter GJ, Alpert NM, Choi NC, Babich JW, Fischman AJ, The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification? J Nucl Med 1994; 35:1308-1312

[Hatt 2013 Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging. 2013;40:1662–71.

[Hatt 2015] Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, 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 Off Publ Soc Nucl Med. 2015;56:38–44.

[Hatt 2017] Hatt M, Lee JA, Schmidtlein CR, Naqa IE, Caldwell C, Bernardi ED, et al. Classification and evaluation strategies of autosegmentation approaches for PET: Report of AAPM task group No. 211. Med Phys. 2017;44:e1–42.

[Hatt 2018] Hatt M, Laurent B, Ouahabi A, Fayad H, Tan S, Li L, et al. The first MICCAI challenge on PET tumor segmentation. Med Image Anal. 2018;44:177–95.

[Hatt 2023] M. Hatt, A. K. Krizsan, A. Rahmim, T. J. Bradshaw, P. F. Costa, A. Forgacs, R. Seifert, A. Zwanenburg, I. El Naqa, P. E. Kinahan, F. Tixier, A. K. Jha & D. Visvikis Joint ANM/SNMMI guideline on radiomics in nuclear medicine European Journal of Nuclear Medicine and Molecular Imaging volume 50, pages 352–375 (2023)

[IBSI 2019] Alex Zwanenburg, Stefan Leger, Martin Vallières, Steffen Löck, Image biomarker standardisation initiative, https://arxiv.org/abs/1612.07003v11

[Kadrmas 2009] Kadrmas DJ, Casey ME, Black NF, Hamill JJ, Panin VY, Conti M. Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes. IEEE Trans Med Imaging. 2009;28:523–534.

[Kertesz 2022] Hunor Kertész, Thomas Beyer, Vladimir Panin, Walter Jentzen, Jacobo Cal-Gonzalez, Alexander Berger, Laszlo Papp, Peter L Kench, Deepak Bharkhada, Jorge Cabello, Maurizio Conti, Ivo Rausch Implementation of a Spatially-Variant and Tissue-Dependent Positron Range Correction for PET/CT Imaging Front Physiol. 2022 Mar 8;13:818463. doi: 10.3389/fphys.2022.818463

[Koo 2016] Terry K. Koo, PhDa,⁎ and Mae Y. Li, BPSb A Guideline of Selecting and Reporting intraclass Correlation Coefficients for Reliability Research J Chiropr Med. 2016 Jun; 15(2): 155–163.

[Lambin 2017] Philippe Lambin , Ralph T H Leijenaar , Timo M Deist , Jurgen Peerlings , Evelyn E C de Jong , Janita van Timmeren , Sebastian Sanduleanu , Ruben T H M Larue , Aniek J G Even , Arthur Jochems , Yvonka van Wijk , Henry Woodruff , Johan van Soest , Tim Lustberg , Erik Roelofs , Wouter van Elmpt , Andre Dekker , Felix M Mottaghy , Joachim E Wildberger , Sean Walsh Radiomics: the bridge between medical imaging and personalized

Review Nat Rev Clin Oncol . 2017 Dec;14(12):749-762. doi: 10.1038/nrclinonc.2017.141. Epub 2017 Oct 4.

[Leijenaar 2015] Leijenaar RTH, Nalbantov G, Carvalho S, Van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep. 2015;5:1–10.

[Liberini 2021a] Liberini, V., De Santi, B., Rampado, O., Gallio, E., Dionisi, B., Ceci, F., Polverari, G., Thuillier, P., Molinari, F., Deandreis, D. Impact of segmentation and discretization on radiomic features in 68Ga-DOTA-TOC PET/CT images of neuroendocrine tumor (2021) EJNMMI Physics, 8 (1), art. no. 21

[Liberini 2021b] Liberini, V., Rampado, O., Gallio, E., De Santi, B., Ceci, F., Dionisi, B., Thuillier, P., Ciuffreda, L., Piovesan, A., Fioroni, F., Versari, A., Molinari, F., Deandreis, D. 68Ga-DOTATOC PET/CT-Based Radiomic Analysis and PRRT Outcome: A Preliminary Evaluation Based on an Exploratory Radiomic Analysis on Two Patients (2021) Frontiers in Medicine, 7, art. no. 601853

[Lu 2016] Lu L, Lv W, Jiang J, Ma J, Feng Q, Rahmim A, et al. Robustness of radiomic features in [11C]choline and [18F]FDG PET/CT imaging of nasopharyngeal carcinoma: impact of segmentation and discretization. Mol Imaging Biol. 2016; 18:935–45.

[Ly 2018] Lv W, Yuan Q, Wang Q, Ma J, Jiang J, Yang W, et al. Robustness versus disease differentiation when varying parameter settings in radiomics features: application to nasopharyngeal PET/CT. Eur Radiol. 2018;28:3245–54.

[Mapelli 2020] Mapelli, P.; Partelli, S.; Salgarello, M.; Doraku, J.; Pasetto, S.; Rancoita, P.M.V.; Muffatti, F.; Bettinardi, V.; Presotto, L.; Andreasi, V.; et al. Dual Tracer 68Ga-DOTATOC and 18F-FDG PET/Computed Tomography Radiomics in Pancreatic Neuroendocrine Neoplasms: An Endearing Tool for Preoperative Risk Assessment. Nucl. Med. Commun. 2020, 41, 896–905.

[Messerli 2019] Messerli M, Kotasidis F, Burger IA, Ferraro DA, Muehlematter UJ, Weyermann C, et al. Impact of different image reconstructions on PET quantification in non-small cell lung cancer: a comparison of adenocarcinoma and squamous cell carcinoma. Br J Radiol. 2019;92.

[Morland 2022a] David Morland, Elizabeth Katherine Anna Triumbari, Luca Boldrini, Roberto Gatta, Daniele Pizzuto, Salvatore Annunziata Radiomics in Oncological PET Imaging: A Systematic Review-Part 1, Supradiaphragmatic Cancers. Diagnostics (Basel). 2022 May 27;12(6):1329. doi: 10.3390/diagnostics12061329

[Morland 2022b] David Morland, Elizabeth Katherine Anna Triumbari, Luca Boldrini, Roberto Gatta, Daniele Pizzuto, Salvatore Annunziata Radiomics in Oncological PET Imaging: A Systematic Review-Part 2, Infradiaphragmatic Cancers, Blood Malignancies, Melanoma and Musculoskeletal Cancers Review Diagnostics (Basel). 2022 May 27;12(6):1330.

[Nioche 2018] Nioche C, Orlhac F, Boughdad S, Reuze S, Goya-Outi J, Robert C, et al. Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res. 2018;78:4786–9.

[Oberg 2016] Oberg K, Krenning E, Sundin A, Bodei L, Kidd M, Tesselaar M, et al. A Delphic consensus assessment: imaging and biomarkers in gastroenteropancreatic neuroendocrine tumor disease management. Endocr Connect. (2016) 5:174–87. doi: 10.1530/EC-16-0043

[Oberg 2016] Oberg K, Krenning E, Sundin A, Bodei L, Kidd M, Tesselaar M, et al. A delphic consensus assessment: imaging and biomarkers in gastroenteropancreatic neuroendocrine tumor disease management. Endocr Connect. 2016;5:174–87. [cited 2020 Feb 7] http://www.ncbi.nlm.nih.gov/pubmed/27582247.

[Orlhac 2015] 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. Public Library of Science; 2015 [cited 2020 Feb 18];10:e0145063. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26669541.

[Orlhac 2016] Orlhac F, Thézé B, Soussan M, Boisgard R, Buvat I. Multiscale texture analysis: from 18F-FDG PET images to histologic images. J Nucl Med. 2016;57:1823–8. [cited 2019 Dec 11] Available from: http://www.ncbi.nlm.nih.gov/pubmed/27261515.

[Oronsky 2017] Oronsky B, Ma PC, Morgensztern D, Carter CA. Nothing But NET: a review of neuroendocrine tumors and carcinomas. Neoplasia (United States). 2017; 991–1002. [cited 2019 Dec 11]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29091800.

[Ortuno 2006] J Ortuno, P. Gutierrez, J. Rubio, G. Kontaxakis, A. Santos, 3D-OSEM iterative image reconstruction for high-resolution PET using precalculated system matrix Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment Volume 569, Issue 2, 20 December 2006, Pages 440-444

[Papp 2019] Papp L, Rausch I, Grahovac M, Hacker M, Beyer T. Optimized feature extraction for radiomics analysis of 18F-FDG PET imaging. J Nucl Med Off Publ Soc Nucl Med. 2019;60:864–72.

[Parmar 2015] Parmar C, Leijenaar RTH, Grossmann P, Velazquez ER, Bussink J, Rietveld D, et al. Radiomic feature clusters and prognostic signatures specific for lung and head &neck cancer. Sci Rep. 2015;5.

[Pfaehler 2020] Pfaehler E, van Sluis J, Merema BBJ, van Ooijen P, Berendsen RCM, van Velden FHP, et al. Experimental multicenter and multivendor evaluation of the performance of PET radiomic features using 3-dimensionally printed phantom inserts. J Nucl Med. NLM (Medline); 2020; 61:469–76.

[Presotto 2018] Presotto L, Bettinardi V, De Bernardi E, et al. PET textural features stability and pattern discrimination power for radiomics analysis: an "ad-hoc" phantoms study. Phys Med. 2018;50:66–74.

[Strosberg 2017] Strosberg J, El-Haddad G, Wolin E, Hendifar A, Yao J, Chasen B, et al. Phase 3 trial of 177lu-dotatate for midgut neuroendocrine tumors. N Engl J Med. (2017) 376:125–35. doi: 10.1056/NEJMoa1607427

[Suter 2020] Y. Suter, U. Knecht, M. Alão, W. Valenzuela, E. Hewer, P. Schucht, R. Wiest, and M. Reyes, “Radiomics for glioblastoma survival analysis in pre-operative mri: exploring feature robustness, class boundaries, and machine learning techniques,” Cancer Imaging, vol. 20, no. 1, pp. 1–13, 2020.

[Tankyevych 2021] Tankyevych O, Tixier F, Antonorsi N, Filali Razzouki A, Mondon R, Pinto-Leite T, et al. Can alternative PET reconstruction schemes improve the prognostic value of radiomic features in non-small cell lung cancer? Methods. 2021;188:73–83.

[Thie 2004] Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. J Nucl Med. 2004;45(9):1431–1434.

[Thompson 2002] Thompson CJ. Instrumentation. In: Burk ED, ed. Principles and practice of PET, Philadelphia, Pa:Lippincott, 2002; 48-63

[Thuillier 2019] Thuillier, P., Liberini, V., Rampado, O., Gallio, E., De Santi, B., Ceci, F., Metovic, J., Papotti, M., Volante, M., Molinari, F., Deandreis, D. Diagnostic value of conventional pet parameters and radiomic features extracted from 18f-fdg-pet/ct for histologic subtype classification and characterization of lung neuroendocrine neoplasms (2021) Biomedicines, 9 (3), art. no. 281, pp. 1-14.

[Traverso 2018] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys. 2018;102:1143–58.

[Valladares 2020] Alejandra Valladares, Thomas Beyer, Ivo Rausch Physical imaging phantoms for simulation of tumor heterogeneity in PET, CT, and MRI: An overview of existing designs Med Phys. 2020 Apr;47(4):2023-2037.

[Warfield 2004] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004;23:903–21.

[Warfield 2004] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004;23:903–21.

[Weber 2019] Weber M, Kessler L, Schaarschmidt B, Fendler WP, Lahner H, Antoch G, et al. Textural analysis of hybrid DOTATOC-PET/MRI and its association with histological grading in patients with liver metastases from neuroendocrine tumors. Nucl Med Commun. 2020;41:363–9 [cited 2020 Oct 6] Available from: https://pubmed.ncbi.nlm.nih.gov/31977752/.

[Weber 2020] Weber, M.; Kessler, L.; Schaarschmidt, B.; Fendler, W.P.; Lahner, H.; Antoch, G.; Umutlu, L.; Herrmann, K.; Rischpler, C. Textural Analysis of Hybrid DOTATOC-PET/MRI and Its Association with Histological Grading in Patients with Liver Metastases from Neuroendocrine Tumors. Nucl. Med. Commun. 2020, 41, 363–369.

[Werner 2017] Werner, R.A.; Lapa, C.; Ilhan, H.; Higuchi, T.; Buck, A.K.; Lehner, S.; Bartenstein, P.; Bengel, F.; Schatka, I.; Muegge, D.O.; et al. Survival Prediction in Patients Undergoing Radionuclide Therapy Based on Intratumoral Somatostatin-Receptor Heterogeneity. Oncotarget 2017, 8, 7039–7049.

[Werner 2019] Werner RA, Ilhan H, Lehner S, Papp L, Zsótér N, Schatka I, et al. Pre-therapy Somatostatin receptor-based heterogeneity predicts overall survival in pancreatic neuroendocrine tumor patients undergoing peptide receptor radionuclide therapy. Mol Imaging Biol. 2019;21:582–90.

[Werner 2019] Werner, R.A.; Ilhan, H.; Lehner, S.; Papp, L.; Zsótér, N.; Schatka, I.; Muegge, D.O.; Javadi, M.S.; Higuchi, T.; Buck, A.K.; et al. Pre-Therapy Somatostatin Receptor-Based Heterogeneity Predicts Overall Survival in Pancreatic Neuroendocrine Tumor Patients Undergoing Peptide Receptor Radionuclide Therapy. Mol. Imaging Biol. 2019, 21, 582–590.

[Yan 2015] Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med Off Publ Soc Nucl Med. 2015;56:1667–73.

[Zwanenburg 2019] Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging. 2019;46:2638–55. [cited 2019 Dec 11] Available from: http://www.ncbi.nlm.nih.gov/pubmed/31240330.

Publications during the doctoral period

Journal papers

1. O. Rampado, A. Depaoli , F. Marchisio , M. Gatti, D. Racine, V. Ruggeri, I. Ruggirello, F. Darvizeh , P. Fonio, R. Ropolo "Effects of different levels of CT iterative reconstruction on low-contrast detectability and radiation dose in patients of different sizes: an anthropomorphic phantom study" Radiol Med. 2020 Jun 3. doi: 10.1007/s11547-020-01228-5.

2. G. Polverari , F.Ceci , V. Bertaglia, M. Reale, O. Rampado, E. Gallio, R. Passera, V. Liberini, P. Scapoli, V. Arena, M. Racca, A. Veltri, S. Novello, D. Deandreis "18 F-FDG Pet Parameters and Radiomics Features Analysis in Advanced Nsclc Treated with Immunotherapy as Predictors of Therapy Response and Survival". Cancers (Basel). 2020 May 5;12(5):1163. doi: 10.3390/cancers12051163.

3. Liberini, V., De Santi, B., Rampado, O., Gallio, E., Dionisi, B., Ceci, F., Polverari, G., Thuillier, P., Molinari, F., Deandreis, D. Impact of segmentation and discretization on radiomic features in 68Ga-DOTA-TOC PET/CT images of neuroendocrine tumor (2021) EJNMMI Physics, 8 (1), art. no. 21, .

4. Thuillier, P., Liberini, V., Rampado, O., Gallio, E., De Santi, B., Ceci, F., Metovic, J., Papotti, M., Volante, M., Molinari, F., Deandreis, D. Diagnostic value of conventional pet parameters and radiomic features extracted from 18f-fdg-pet/ct for histologic subtype classification and characterization of lung neuroendocrine neoplasms (2021) Biomedicines, 9 (3), art. no. 281, pp. 1-14.

5. Zanca, F., Hernandez-Giron, I., Avanzo, M., Guidi, G., Crijns, W., Diaz, O., Kagadis, G.C., Rampado, O., Lønne, P.I., Ken, S., Colgan, N., Zaidi, H., Zakaria, G.A., Kortesniemi, M. Expanding the medical physicist curricular and professional programme to include Artificial Intelligence (2021) Physica Medica, 83, pp. 174-183.

6. Homayounieh, F., Holmberg, O., Al Umairi, R., Aly, S., Basevičius, A., Costa, P.R., Darweesh, A., Gershan, V., Ilves, P., Kostova-Lefterova, D., Renha, S.K., Mohseni, I., Rampado, O., Rotaru, N., Shirazu, I., Sinitsyn, V., Turk, T., van Ngoc Ty, C., Kalra, M.K., Vassileva, J. Variations in CT utilization, protocols, and radiation doses in COVID-19 pneumonia: Results from 28 countries in the IAEA study (2021) Radiology, 298 (3), pp. E141-E151.

7. Avanzo, M., Porzio, M., Lorenzon, L., Milan, L., Sghedoni, R., Russo, G., Massafra, R., Fanizzi, A., Barucci, A., Ardu, V., Branchini, M., Giannelli, M., Gallio, E., Cilla, S., Tangaro, S., Lombardi, A., Pirrone, G., De Martin, E., Giuliano, A., Belmonte, G., Russo, S., Rampado, O., Mettivier, G. Artificial intelligence applications in medical imaging: A review of the medical physics research in Italy (2021) Physica Medica, 83, pp. 221-241.

8. Liberini, V., Rampado, O., Gallio, E., De Santi, B., Ceci, F., Dionisi, B., Thuillier, P., Ciuffreda, L., Piovesan, A., Fioroni, F., Versari, A., Molinari, F., Deandreis, D. 68Ga-DOTATOC PET/CT-Based Radiomic Analysis and PRRT Outcome: A Preliminary Evaluation Based on an Exploratory Radiomic Analysis on Two Patients (2021) Frontiers in Medicine, 7, art. no. 601853.

9. Philippe THUILLIER, Virginia LIBERINI, Serena GRIMALDI, Osvaldo RAMPADO, Elena GALLIO, Bruno DE SANTI, Emanuela ARVAT, Alessandro PIOVESAN, Roberto FILIPPI, Ronan ABGRAL, Filippo MOLINARI and Desiree DEANDREIS "Prognostic value of whole-body PET volumetric parameters extracted from 68Ga-DOTATOC-PET/CT in well-differentiated

neuroendocrine tumors" Journal of Nuclear Medicine November 2021, jnumed.121.262652; DOI: https://doi.org/10.2967/jnumed.121.262652

10. Davide Tore, Osvaldo Rampado, Carla Guarnaccia, Roberto Mina, Maria Oronzio, Ambra Santonocito, Alessandro Serafini, Giulio Antonino Strazzarino, Laura Gianusso, Sara Bringhen, Paolo Fonio, Alessandro Depaoli "Ultra-Low-Dose Whole-Body Computed Tomography Protocol Optimization for Patients With Plasma Cell Disorders: Diagnostic Accuracy and Effective Dose Analysis From a Reference Center" Front Oncol. 2021 Nov 17;11:769295. doi: 10.3389/fonc.2021.769295. eCollection 2021.

11. Francesco Ria, Loredana D'Ercole, Daniela Origgi, Nicoletta Paruccini, Luisa Pierotti, Osvaldo Rampado, Veronica Rossetti, Sabina Strocchi , Alberto Torresin, Association of Medical Physics Task Group "Statement of the Italian Association of Medical Physics (AIFM) task group on radiation dose monitoring systems" Insights Imaging. 2022 Feb 5;13(1):23. doi: 10.1186/s13244-022-01155-1.

Abstract

1. Liberini V, De Santi B, Gallio E, Rampado O, Dionisi B, Ceci F, Zotta M, Giunta F, Finessi M, Bellò M, Molinari F, Deandreis D "Robustness of Radiomic Features in 68Ga-DOTATOC PET/CT: A Monocentric Experience for Neuroendocrine Tumors", accepted as poster at the 17th Annual ENETS Conference 2020

2. Liberini V, Rampado O, Gallio E, De Santi B, Ceci F, Pilati E, Finessi M, Grimaldi S, Ponzetti A, Molinari F, Ciuffreda L, Deandreis D "Peptide Receptor Radionuclide Therapy (PRRT) Outcome Prediction in Neuroendocrine Tumors (NET): Could Radiomic Features Analysis of 68Ga-DOTATOC PET/CT Have a Role?" accepted as poster at the 17th Annual ENETS Conference 2020

3. Philippe Thuillier, Virginia Liberini, Osvaldo Rampado, Elena Gallio, Bruno De Santi, Francesco Ceci, Annarita Ianniello, Cinzia Baiocco, Andrea Bianchi, Filippo Molinari, Désirée Deandreis "Dual 18F-FDG and 68Ga-DOTATOC PET/CT radiomic analysis in the evaluation of primary pulmonary neuroendocrine tumor", accepted as poster at the 2020 Neuro Endocrine Tumour Research Foundation Symposium

4. Philippe Thuillier, Virginia Liberini, Serena Grimaldi, Osvaldo Rampado, Elena Gallio, Bruno De Santi, Francesco Ceci, Emanuela Arvat, Alessandro Piovesan, Roberto Filippi, Filippo Molinari, Désirée Deandreis "Prognostic value of whole-body PET volumetric parameters extracted from 68Ga-DOTATOC-PET/CT in well-differentiated neuroendocrine tumors", accepted as poster at the 18th Annual ENETS Conference 2021

5. Santonocito, Galioto, Greco, Pelle, Durando, Bartoli, Regini, Rampado, Costa, Fonio "Breast MRI radiomic features for predicting response to neoadjuvant chemotherapy in triple negative breast cancers: a preliminary experience" presented as poster ad the European Society of Breast Imaging , Eusobi 2021

6. O. Rampado, V. Liberini, E. Gallio, F. Ceci, B. Dionisi, G. Polverari, B. De Santi, F. Molinari, D. De Andreis, R. Ropolo "Impact of intensity discretization on radiomics analiysis in 68Ga-DOTATOC PET/CT for Neuroendocrine Tumor" presented as poster at the European Congress of Medical Physics 2021

7. O. Rampado, L. Gianusso, S. Costantino, V. Rossetti, R. Ropolo, "Radiation Risk for Multiple CT Examinations in a Large Multi-Specialist Hospital: a potential role of total DLP from body series" presented as oral presentation at the European Congress of Medical Physics 2021

8. M. Oronzio, O. Rampado, L. Gianusso, R. Ropolo “Correlation of peak skin dose with dose indicators and geometric factors in vascular interventional radiology: a general linear model approach.” presented as oral presentation at the European Congress of Medical Physics 2021

9. O. Rampado, V. Liberini, S. Deagostini, E. Gallio, L. Lamberti, L. Li Donni, P. Thuillier, D. Deandreis “A 3D printed phantom study of the impact of reconstruction parameters and used radioisotopes on PET radiomics features”, presentato come poster al Congresso Europeo di Fisica Medica ECMP 2022, Dublino Agosto 2022

10. L. Li Donni, O. Rampado “Physical characterization of a Fast kVp switching Dual-Energy CT: proposal of a quality control protocol.”, presentato come poster al Congresso Europeo di Fisica Medica ECMP 2022, Dublino Agosto 2022

Appendix 1

In this appendix Lifex features and IBSI features are compared.

GLCM (Grey Level Co-occurrence Matrix)

LifeX	IBSI
GLCM_Homogeneity	Inverse difference
GLCM_Energy	Angular second moment
GLCM_Contrast	Contrast
GLCM_Correlation	Correlation
GLCM_Entropy_log10	/
GLCM_Entropy_log2	Joint entropy
GLCM_Dissimilarity	Dissimilarity

In the IBSI document, GLCM features are 24.

NGLDM (Neighborhood Grey-Level Different Matrix)

In IBSI: NGTDM (*Neighborhood Grey Tone Difference Matrix)*

LifeX	IBSI
NGLDM_Coarseness	Coarseness
NGLDM_Contrast	Contrast
NGLDM_Busyness	Busyness

In the IBSI documento, NGTDM features are 5.

GLRLM (Grey-Level Run Length Matrix)

LifeX	IBSI
GLRLM_SRE	Short runs emphasis
GLRLM_LRE	Long runs emphasis
GLRLM_LGRE	Low grey level run emphasis
GLRLM_HGRE	High grey level run emphasis
GLRLM_SRLGE	Short run low grey level emphasis
GLRLM_SRHGE	Short run high grey level emphasis
GLRLM_LRLGE	Long run low grey level emphasis
GLRLM_LRHGE	Long run high grey level emphasis
GLRLM_GLNUr	Grey level non-uniformity
GLRLM_RLNU	Run length non-uniformity
GLRLM_RP	Run percentage

In the IBSI document, GLRLM features are 16.

GLZLM (Grey-Level Zone Length Matrix)

In IBSI: GLSZM (*Grey Level Size Zone Matrix)*

LifeX	IBSI
GLZLM_SZE	Small zone emphasis
GLZLM_LZE	Large zone emphasis
GLZLM_LGZE	Low grey level zone emphasis
GLZLM_HGZE	High grey level zone emphasis
GLZLM_SZLGE	Small zone low grey level emphasis
GLZLM_SZHGE	Small zone high grey level emphasis
GLZLM_LZLGE	Large zone low grey level emphasis
GLZLM_LZHGE	Large zone high grey level emphasis
GLZLM_GLNUz	Grey level non-uniformity
GLZLM_ZLNU	Zone size non-uniformity
GLZLM_ZP	Zone percentage

In IBSI, GLSZM *features* are 16.

Shape

LifeX	IBSI
SHAPE_Sphericity	Sphericity
SHAPE_Compacity	/ *
SHAPE_Volume	Volume

In IBSI document two measures of compacity are provided, both similar to the one defined in IBSI.

LifeX	IBSI	Equivalenza
SHAPE_Compacity: $SHAPE\_Compacity = \frac{A^{3/2}}{V}$	Compactness 1: $F_{morph.comp.1} = \frac{V}{\pi^{1/2} A^{3/2}}$ Compactness 2: $F_{morph.comp.2} = 36\pi \frac{V^2}{A^3}$	Compactness 1: $F_{morph.comp.1} = \frac{1}{\pi^{1/2} \cdot SHAPE\_Compacity}$ Compactness 2: $F_{morph.comp.2} = \frac{(36\pi)^{\frac{1}{2}}}{SHAPE\_Compacity}$

Compactness 1 e Compactness 2 correspond to the reciprocal of SHAPE_Compacity with a constant multiplicative factor

Histogram features

In IBSI: Intensity histogram features

LifeX	IBSI
HISTO_Skewness	Intensity Skewness
HISTO_Kurtosis	(Excess) intensity kurtosis *
HISTO_Entropy_log10	/
HISTO_Entropy_log2	Discretised intensity entropy
HISTO_Energy	Discretised intensity uniformity

*(Excess) intensity kurtosis = HISTO_Kurtosis – 3 in order to obtain a value of 0 for gaussian distributions.

Conventional indices

LifeX	IBSI
CONV_min	Minimum intensity
CONV_mean	Mean intensity
CONV_max	Maximum intensity
CONV_peak	Local intensity peak
CONV_TLG	Integrated intensity
CONV_RIM_xxx	/

Categories present in IBSI and absent in Lifex:

- Grey Level Distance Zone Matrix (GLDZM)
- Neighbouring Grey Level Dependence Matrix (NGLDM). To be noticed that in LifeX the acronym NGLDM is used to indicate Neighborhood Grey-Level Different Matrix which represent the NGTDM im IBSI (Neighborhood Grey Tone Difference Matrix).