ORIGINAL ARTICLE

The Impact of Preprocessing on the PET-CT Radiomics Features in Non-Small Cell Lung Cancer

Seyyed Ali Hosseini [1,2], Isaac Shiri [3], Ghasem Hajianfar [4], Pardis Ghafarian [5,6*] ⓘ , Mehrdad Bakhshayesh Karam [5], Mohammad Reza Ay [1,2]

[1] Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

[2] Research Center for Science and Technology in Medicine, Tehran University of Medical Sciences, Tehran, Iran

[3] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva 4, Switzerland

[4] Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran

[5] Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[6] PET/CT and Cyclotron Center, Masih Daneshvari Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

*Corresponding Author: Pardis Ghafarian   Received: 06 May 2021 / Accepted: 06 July 2021
Email: pardis.ghafarian@sbmu.ac.ir

Abstract

Purpose: This study aimed to investigate the impact of image preprocessing steps, including Gray Level Discretization (GLD) and different Interpolation Algorithms (IA) on ^{18}F-Fluorodeoxyglucose (^{18}F-FDG) radiomics features in Non-Small Cell Lung Cancer (NSCLC).

Materials and Methods: One hundred and seventy-two radiomics features from the first-, second-, and higher-order statistic features were calculated from a set of Positron Emission Tomography/Computed Tomography (PET/CT) images of 20 non-small cell lung cancer delineated tumors with volumes ranging from 10 to 418 cm3 regarding five intensity discretization schemes with the number of gray levels of 16, 32, 64, 128, and 256, and four Interpolation algorithms, including nearest neighbor, tricubic convolution and tricubic spline interpolation, and trilinear were used. Segmentation was based on 3D region growing-based. The Intraclass Correlation Coefficient (ICC), Overall Concordance Correlation Coefficient (OCCC), and Coefficient Of Variations (COV) were calculated to demonstrate the features' variability and select robust features. ICC and OCCC < 0.5 presented weak reliability, ICC and OCCC between 0.5 and 0.75 illustrated appropriate reliability, values within 0.75 and 0.9 showed satisfying reliability, and values higher than 0.90 indicate exceptional reliability. Besides, features with less than 10% COV have been selected as robust features.

Results: All morphology family (except four features), statistic, and Intensity volume histogram families were not affected by GLD and IA. And the rest of them, 10 and 61 features showed COV ≤ 5% against GLD and IA, respectively. Ten and 80 features showed excellent reliability (ICC values greater than 0.90) against GLD and IA. Eight and 60 features showed OCCC≥0.90 against GLD and IA, respectively. Based on our results Inverse difference normalized and Inverse difference moment normalized from Grey Level Co-occurrence Matrix (GLCM) were the most robust features against GLD and Skewness from intensity histogram family and Inverse difference normalized and Inverse difference moment normalized from GLCM were the most robust features against IA.

Conclusion: Preprocessing can substantially impact the 18F-FDG PET image radiomic features in NSCLC. The impact of gray level discretization on radiomics features is significant and more than Interpolation algorithms.

Keywords: Non-Small Cell Lung Cancer; Gray Level Discretization; Interpolation Algorithms; Radiomics Features; Positron Emission Tomography/Computed Tomography.

1. Introduction

Lung cancer is one of the main reasons for cancer-related deaths globally. In 2018, more than 3.6 million patients with lung cancer and 2.1 million lung cancer-related deaths were recorded in the world [1]. With 70% of lung cancer diagnoses following the opening of signs from limited or metastatic disorder, the five-year survival rate of lung cancer following investigation is observed in just 17% of patients [2, 3]. When the cancer is diagnosed, the survival rate is higher than 50% [4]. Sorrowfully, just 15% of lung cancers are investigated at the first stages, and a reliable and affordable experiment method is still a significant need.

Recently, numerous researchers have investigated the potential of radiomics features to predict patient consequences non-invasively [5-9]. Radiomics aims [5] to enhance the predictive and diagnostic value of medical images by converting images to data [10]. These features intend to quantify tumor characteristics such as intensity, heterogeneity, and shapes associated with clinical results and promote personalized cancer therapy [11]. Radiomics feature extraction framework can be divided into different steps, including data acquisition, image preprocessing, segmentation, radiomics feature extraction, and model advancement.

For radiomics feature extraction, texture investigation is a broadly proven strategy that has revealed diagnostic potential in both Positron Emission Tomography (PET) and Computed Tomography (CT) images to characterize tumor heterogeneity [5]. Notwithstanding promoting research, there are novel trials to succeed in a particular step of the radiomics framework before proceeding from thought to clinical application [12]. A fundamental hurdle is to guarantee medical image features with predictive and/or prognostic value are robust to required image processing steps along with the radiomics principle. Like any biomarkers, the repeatability and reproducibility of radiomics features can be affected by different parameters. For instance, image-acquisition methods, test-retest repeatability, reconstruction algorithm, and multi-center reproducibility all take part in questioning the repeatability and reproducibility radiomics features. Improved repeatability and reproducibility of radiomics features, with respect to different parameters such as pre- and post-processing including segmentation, data acquisition, gray level discretization, and Interpolation algorithms is beneficial. Phantom studies and test-retest research can assess feature reproducibility and repeatability, also can be used to feature selection utilizing the Concordance Correlation Coefficient (CCC) or Intraclass Correlation Coefficient (ICC) threshold values [13]. The Gray Level Discretization (GLD) and Interpolation Algorithms (IA) methods have shown a direct effect on the reproducibility of texture features. The GLD and IA methods have shown a direct effect on the reproducibility of texture features. The effect of preprocessing on the PET texture features is considerable especially GLD and IA, and it is suggested in many studies, a precise examination of the impact of these parameters is necessary before any PET radiomics features clinical application. For example, Shafiq-ul-Hassan et al. [12] Studied the impact of grey level discretization on PET radiomics features in lung cancer phantom and concluded that radiomics researchers should estimate possible imaging biomarkers' dependence on grey level variations. Larue et al. [13] examine the impact of image preprocessing, including grey-level discretization, on radiomic feature values and their stability and selected stable radiomic features with CCC>0.85. Their results showed that image preprocessing especially grey-level discretization, has a large impact on PET radiomics features. Altazi et al. [14] examined the impact of tumor segmentation, reconstruction, and gray level discretization on [18F]-Fluorodeoxyglucose ([18F]-FDG) radiomics features and showed 81.3% of radiomics features scored Dice coefficient > 0.75. Shiri et al. [15] studied the impact of reconstruction on ^{18}F-FDG radiomics features and showed the 45% of features have Coefficient Of Variations (COV) ≤ 0.05. This study aimed to examine the influence of a wide range of gray level variability from 16, 32, 64, 128, and 256 and three different Interpolation algorithms on PET radiomics features in Non-Small Cell Lung Cancer (NSCLC) and selecting robust features against GLD and IA in lung cancer.

2. Materials and Methods

2.1. Patient Images

In this study, the images acquired with Discovery 690 PET/CT scanner (GE Healthcare), It has 24 detector rings cover 15.7 cm axially and 70 cm Transaxial Field Of View (FOV). This scanner is also equipped with a 64-slice CT scan system, which has 58,368 solid-state detectors. In total, 20 patients' images with NSCLC were examined (11 men, and 9 women, mean age: 45 ± 15y). All patients had biopsy-proven NSCLC, Adenocarcinoma (AC), and went

through PET/CT as part of the standard diagnostic staging path. All Patients were injected with 300.0 ± 56.8MBq of ^{18}F-FDG and scanned 50-60 minutes following the injection. The PET acquisition for 20 patients was conducted using the Ordered Subset Expectation Maximization (OSEM) algorithm with the modelling of the Point Spread Function (PSF) and 18 subsets and three iterations. To cover the whole body area (from skull to mid-thigh), seven to eight-bed positions were acquired. The post-reconstruction 6.4mm full width at half maximum (FWHM) filter was applied to the images. Also, low dose CT data were used for Attenuation correction.

2.2. Image Segmentation

All segmentations were applied using the Matlab2019b software. 3D region growing-based segmentation was used for lesions Volume Of Interest (VOI) segmentation. The algorithm is a statistical region growing that needs one or more seeds as input. The vicinity (standard deviation and the mean of the intensity) is measured by the statistical pattern on the seed points regions. The method is iterated on similarly to standard data clustering algorithms that the reproducibility of segmentation with it is approved in a recent study [16] (Figure 1).

2.3. Post-Processing Methods

2.3.1. Gray Level Discretization

All of the image post-processing was applied to the images using the SERA package [18] in Matlab. Two ways to discretization are frequently utilized. The first method includes the discretization to a Fixed Bin Number (FBN) and the other one involves a fixed width of bins. Nevertheless; both approaches have unique features that may advance them properly suited for special goals [17]. In this study, to investigate the impact of Gestational Diabetes Mellitus (GDM) on ^{18}F-FDG radiomics features in NSCLC, all images were processed with different range of gray levels of 16, 32, 64, 128, and 256, fixed bin

numbers with linear Interpolation algorithm. Fixed bin number methods are defined as follows:

$$X_{d,k} = \begin{cases} \left[N_g \frac{X_{gl,k} - X_{gl,min}}{X_{gl,max} - X_{gl,min}} \right] + 1, & X_{gl,k} < X_{gl,max} \\ N_g & , X_{gl,k} = X_{gl,max} \end{cases} \quad (1)$$

At Equation 1 (X_{gl} is the intensity and N_g is the number of bins.)

The FBN method includes a normalizing impact that may be advantageous if intensity factors are unpredictable and where diversity is deemed essential.

Besides, since amounts of numerous radiomics features depend on the number of grey levels located inside the segmented region, the application of an FBN discretization provides for a primary association of feature values over various investigated Region Of Interest (ROIs).

2.3.2. Interpolation Algorithms

Various algorithms are frequently applied for Interpolation. In this study, to examine the impact of the IA on radiomics features, four Interpolation algorithms, including nearest neighbor, tricubic convolution and tricubic spline interpolation, and trilinear were applied to the images with fixed bin size 64 gray level before tumor delineation. Interpolation algorithms interpret the intensities of the images of the primary image grid through an interpolation one. Their center spatially represents the voxels of these grids. Numerous algorithms are frequently utilized for interpolation, including nearest neighbor, tricubic convolution and tricubic spline interpolation, and trilinear. To describe shortly, nearest-neighbor interpolation indicates the most nearly voxel's intensity in the primary grid toward all voxel in the interpolation grid. Trilinear-interpolation utilizes the intensities of the eight most nearly voxels in the primary grid to measure a different interpolated intensity utilizing linear interpolation. Tricubic-spline interpolation and tricubic-convolution induce a broader neighborhood to estimate a continuous, smooth third-order polynomial at interpolation grid in the voxel centers. The distinction between tricubic-spline and tricubic-convolution interpolation prevails in the
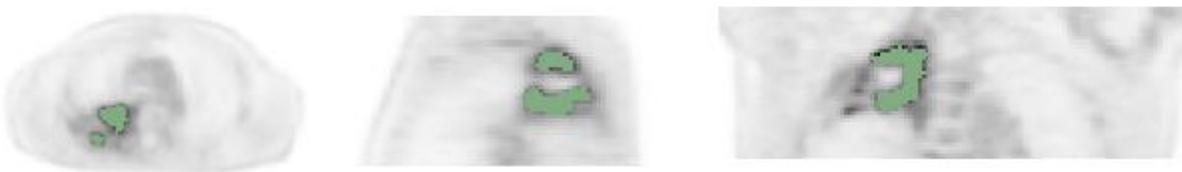

Figure 1. 3D-region growing-based segmentation in Matlab

implementation. Whereas tricubic-convolution approaches the answer utilizing a convolution filter, tricubic spline interpolation estimates the continuous and smooth third-order polynomial at each voxel center [31].

2.4. Feature Extraction

Radiomics features were extracted using SERA code[18]. Overall, One hundred and seventy two radiomics features, including Morphology (n=29), Intensity histogram (n=23), Intensity-based statistics (n=18), Intensity-volume histogram (n=7), Grey level run length matrix (GLRLM) (n=16), Grey level co-occurrence matrix (GLCM) (n=25), Grey level distance zone matrix (GLDZM) (n=16), Grey level size zone matrix (GLSZM) (n=16), Neighboring grey level dependence matrix (NGLDM) (n=17), and Neighborhood grey tone difference matrix (NGTDM) (n=5) were extracted from each lesions. Radiomics features and their family are listed in the Table 1.

Table1. One hundred and seventy-two Radiomics features extracted from patient images

Family	Image Biomarker	Family	Image Biomarker
Morphology	Volume (mesh-based) Volume (counting) Surface area Surface to volume ratio Compactness 1 Compactness 2 Spherical disproportion Sphericity Asphericity Centre of mass shift Maximum 3D diameter Major axis length Minor axis length Least axis length Elongation Flatness Volume density (AABB) Area density (AABB) Volume density (OMBB) Area density (OMBB) Volume density (AEE) Area density (AEE) Volume density (MVEE) Area density (MVEE) Volume density (convex hull) Area density (convex hull) Integrated intensity Moran's I index Geary's C measure	**Co-occurrence matrix (3D, merged)**	Joint maximum Joint average Joint variance Joint entropy Difference average Difference variance Difference entropy Sum average Sum variance Sum entropy Angular second moment Contrast Dissimilarity Inverse difference Inverse difference normalized Inverse difference moment Inverse difference moment normalized cm_inv_diff_mom_norm_3D_comb Inverse variance Correlation Autocorrelation Cluster tendency Cluster shade Cluster prominence Information correlation 1 Information correlation 2
Statistics	Mean Variance Skewness (Excess) kurtosis Median Minimum 10th percentile 90th percentile Maximum Interquartile range Range Mean absolute deviation Robust mean absolute deviation Median absolute deviation Coefficient of variation Quartile coefficient of dispersion Energy Root mean square	**Run length matrix (3D, merged)**	Short runs emphasis Long runs emphasis Low grey level run emphasis High grey level run emphasis Short run low grey level emphasis Short run high grey level emphasis Long run low grey level emphasis Long run high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalized Run length non-uniformity Run length non-uniformity normalized Run percentage Grey level variance Run length variance Run entropy

Family	Image Biomarker	Family	Image Biomarker
Intensity Histogram	Mean Variance Skewness Kurtosis Median Minimum 10th percentile 90th percentile Maximum Mode Interquartile range Range Mean absolute deviation Robust mean absolute deviation Median absolute deviation Coefficient of variation Quartile coefficient of dispersion Entropy Uniformity Maximum histogram gradient Maximum gradient grey level Minimum histogram gradient Minimum gradient grey level	**Size Zone Matrix (3D)**	Small zone emphasis Large zone emphasis Low grey level emphasis High grey level emphasis Small zone low grey level emphasis Small zone high grey level emphasis Large zone low grey level emphasis Large zone high grey level emphasis Grey level non-uniformity Grey level non uniformity normalized Zone size non-uniformity Zone size non-uniformity normalized Zone percentage Grey level variance Zone size variance Zone size entropy
Intensity Volume Histogram	Volume fraction at 10% intensity Volume fraction at 90% intensity Intensity at 10% volume Intensity at 90% volume Volume fraction difference between 10% and 90% intensity Intensity difference between 10% and 90% volume Area under the IVH curve	**Distance Zone Matrix (3D)**	Small distance emphasis Large distance emphasis Low grey level emphasis High grey level emphasis Small distance low grey level emphasis Small distance high grey level emphasis Large distance low grey level emphasis Large distance high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalized Zone distance non-uniformity Zone distance non-uniformity normalized Zone percentage Grey level variance Zone distance variance Zone distance entropy

Family	Image biomarker
Neighborhood grey tone difference matrix (3D)	Coarseness Contrast Busyness Complexity Strength Low dependence emphasis High dependence emphasis Low grey level count emphasis High grey level count emphasis Low dependence low grey level emphasis Low dependence high grey level emphasis High dependence low grey level emphasis High dependence high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalized Dependence count non-uniformity Dependence count non-uniformity normalized Dependence count percentage Grey level variance Dependence count variance Dependence count entropy Dependence count energy

2.5. Statistical Analysis

To demonstrate the impact of GLD and IA on radiomics features, different statistical analyses were performed. ICC, Overall Concordance Correlation Coefficient (OCCC), and Coefficient Of Variations (COV) were calculated for each feature [14,15]. Based on Koo and Li's study [34], two-way random effects with an absolute agreement and multiple raters were used for ICC calculation. ICC and OCCC values < 0.5 demonstrate weak reliability, ICC and OCCC between 0.5 and 0.75 show reasonable reliability, values within 0.75 and 0.9 indicate satisfying reliability, and values higher than 0.90 indicate exceptional reliability [19]. Furthermore, radiomics features with less than 10% COV over different GLD or Segmentation Methods (SM) have been selected as robust features [15]. For ICC [20-22] and OCCC [23, 24] calculation we used 'irr' (0.84.1) and 'epiR' (2.0.19) libraries in R software, respectively.

3. Results

3.1 Impact of Gray Level Discretization

In this study, we had 20 PET NSCLC images with AC. All images were processed and five different gray levels, including 16, 32, 64, 128, and 256 were applied to the images. The ICC, OCCC, and COV values for all shape features were higher than 90% because of the objectivity that the same segmentations in test images were applied onto retest images. Therefore, all morphology, Statistics, and Intensity volume histogram features were eliminated from further analysis. Ten features showed less than 10% COV, including Inverse difference normalized, Inverse

difference moment normalized, Information correlation 2, Correlation of GLCM family, Skewness, Coefficient of variation, and Quartile coefficient of dispersion of Intensity histogram family and Dependence count entropy of NGLDM family. Four features showed $10 \leq COV < 20$, and 103 features (87% of 118 features) showed more than 20% COV. ICC results showed ten features were exceptional reliability (ICC≥ 0.9) such as Inverse difference normalized, Inverse difference moment normalized, Information correlation 2, Correlation of GLCM family, Skewness, Kurtosis, Coefficient of variation, Quartile coefficient of dispersion of Intensity histogram family, Run length non-uniformity of GLRLM family and Coarseness of NGDTM. Eleven features showed $0.9 \leq ICC \geq 0.75$, and 9 features were reasonably reliable. Eighty-four radiomics features (71% of 118) showed less than 0.5 ICC over GLD. OCCC statistical analysis showed eight features have values more than 0.9, such as Inverse difference normalized, Inverse difference moment normalized, Correlation of GLCM, Skewness, Kurtosis, Coefficient of variation, and Quartile coefficient of dispersion of Intensity histogram family, and Coarseness of NGDTM. One and three features showed satisfying reliability and reasonable reliability, respectively.

Figure 2 depicts the percentage COV of each radiomics feature family over different gray levels and Interpolation algorithms. In Figure 3 we represent the ICC and OCCC values concentration over different gray level variability and Figure 4 illustrates bar plots depicting the percentage of four ICC and OCCC categories for different image preprocessing methods over all radiomic features, our results indicate the impact of gray level variability on features is significant, and about 73% and 89% of ICC
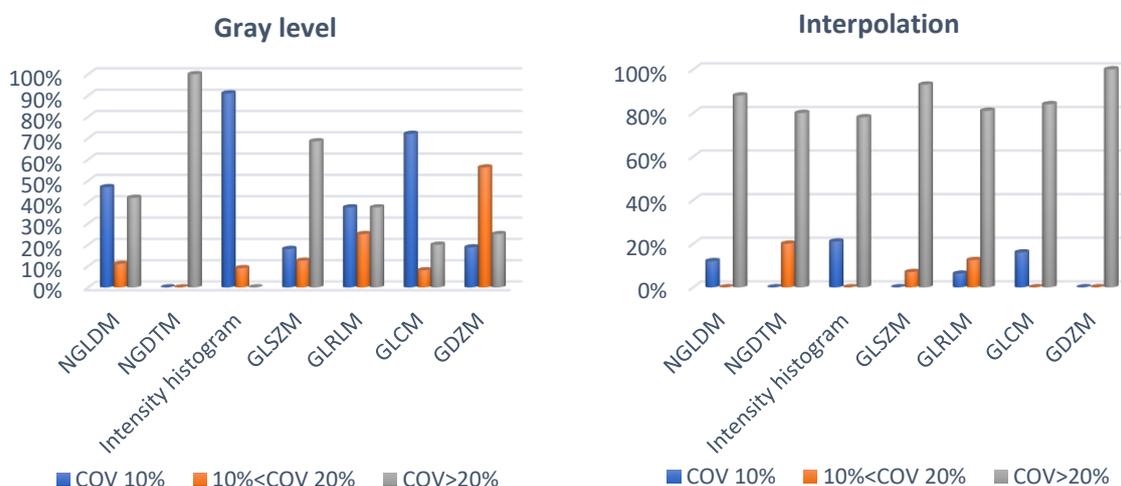

Figure 2. Percentage COV of radiomics features' family (the left figure is for the GLD and the right one is for IA)

and OCCC values are less than 0.5 which shows weak reliability.

Percentage COV has been categorized into three groups: COV lesser than 10 percent are highly robust, and COV between 10 to 20 percent are robust, and COV of more than 20 percent are not robust. Interpolation algorithms between zeros to one. It can be observed that ICC and OCCC values concentration of gray level variability are about 0.25 or less. In Figure 5 we illustrate the probability density (PD) of ICC and OCCC distribution for various radiomic features over different gray levels variability and Interpolation algorithms which is used to present a quantitative analytical description of OCCC and ICC. In PDD, peak value and shape can be utilized

to analyze the ICC results. Precisely, in the current study, we utilize this structure to evaluate how radiomic features are affected over various image post-processing methods. This Figure demonstrates the concentration probability of ICC and OCCC values of gray level impact on radiomics features are lesser than 0.5, and specially OCCC values concentration probability are lower than 0.25. In Figure 6, we represent the variability of various radiomics features with the percentage of COV over different gray levels and Interpolation algorithms. Forty-four features (37% of 118) showed more than 100% COV over different gray level variability.
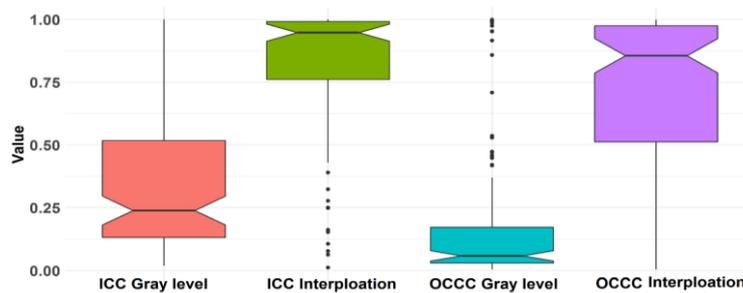
Figure 3. ICC and OCCC values concentration within 0 to 1. (0= not reproducible and 1= highly reproducible)
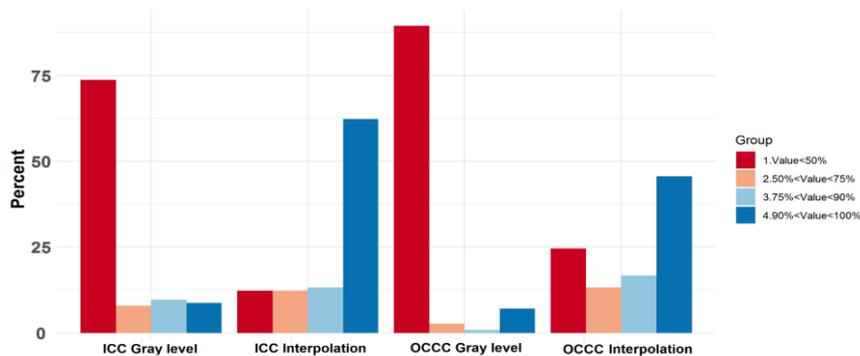
Figure 4. ICC and OCCC bar plots for Image preprocessing including gray level discretization and Interpolation
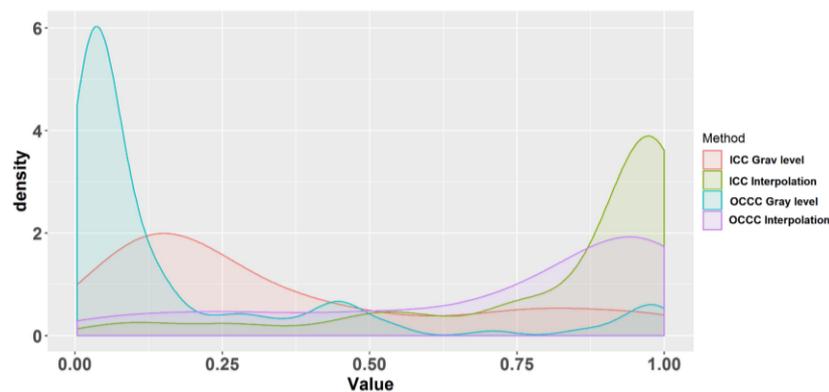

Figure 5. A Probability Density Distribution (PDD) plot compares ICC and OCCC values of various radiomics features over different gray levels and Interpolation algorithms using peak values and shape of each plot. X-axis: ICC and OCCC value; y-axis: density value
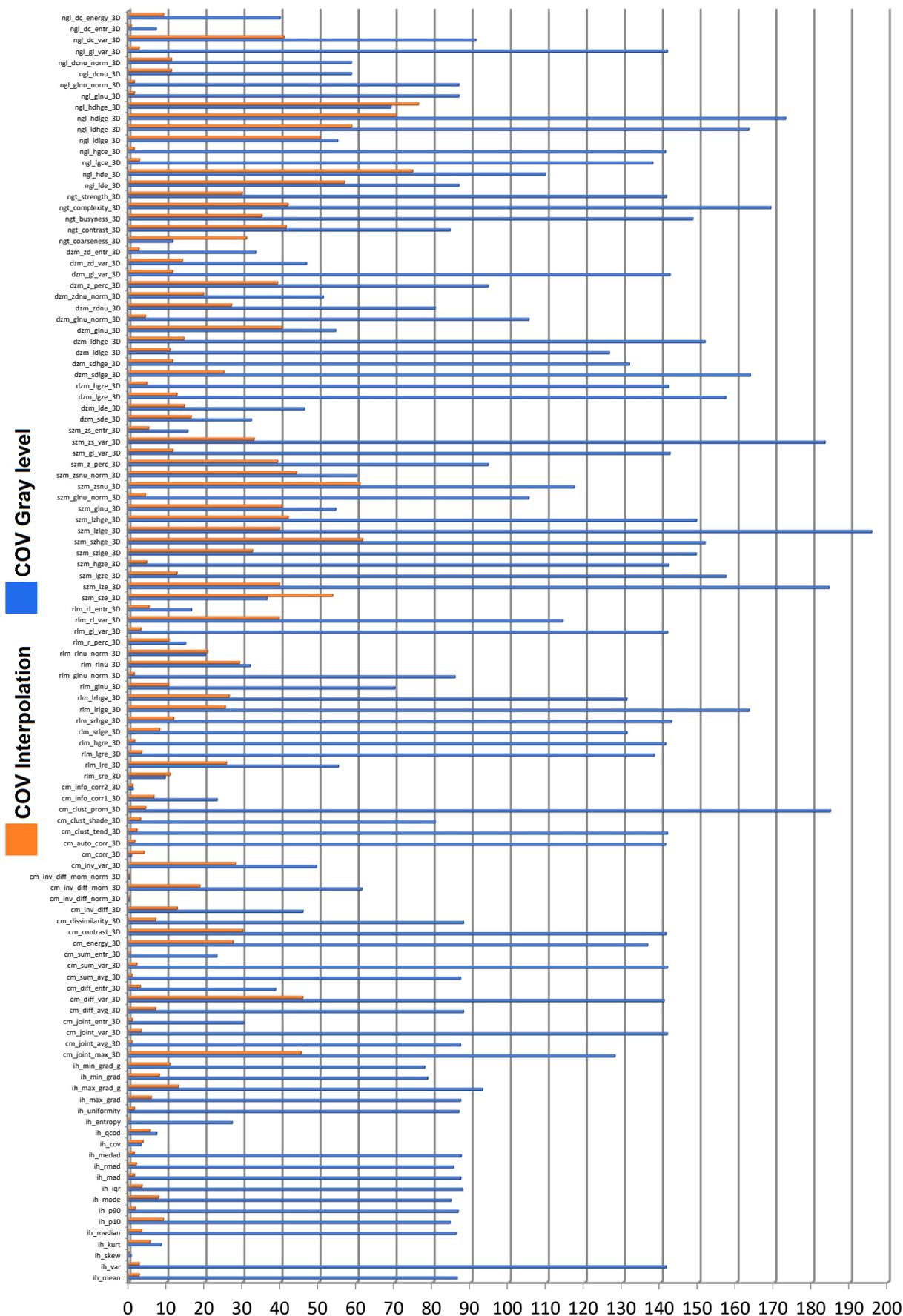
Figure 6. Percentage COV of features over different gray levels and Interpolation algorithms

Figure 7 indicates the ICC and OCCC values (categorized 1 to 4: 1 = low and 4 = highly robust) of various radiomic features extracted from discretization different gray level variability Interpolation algorithms of images.
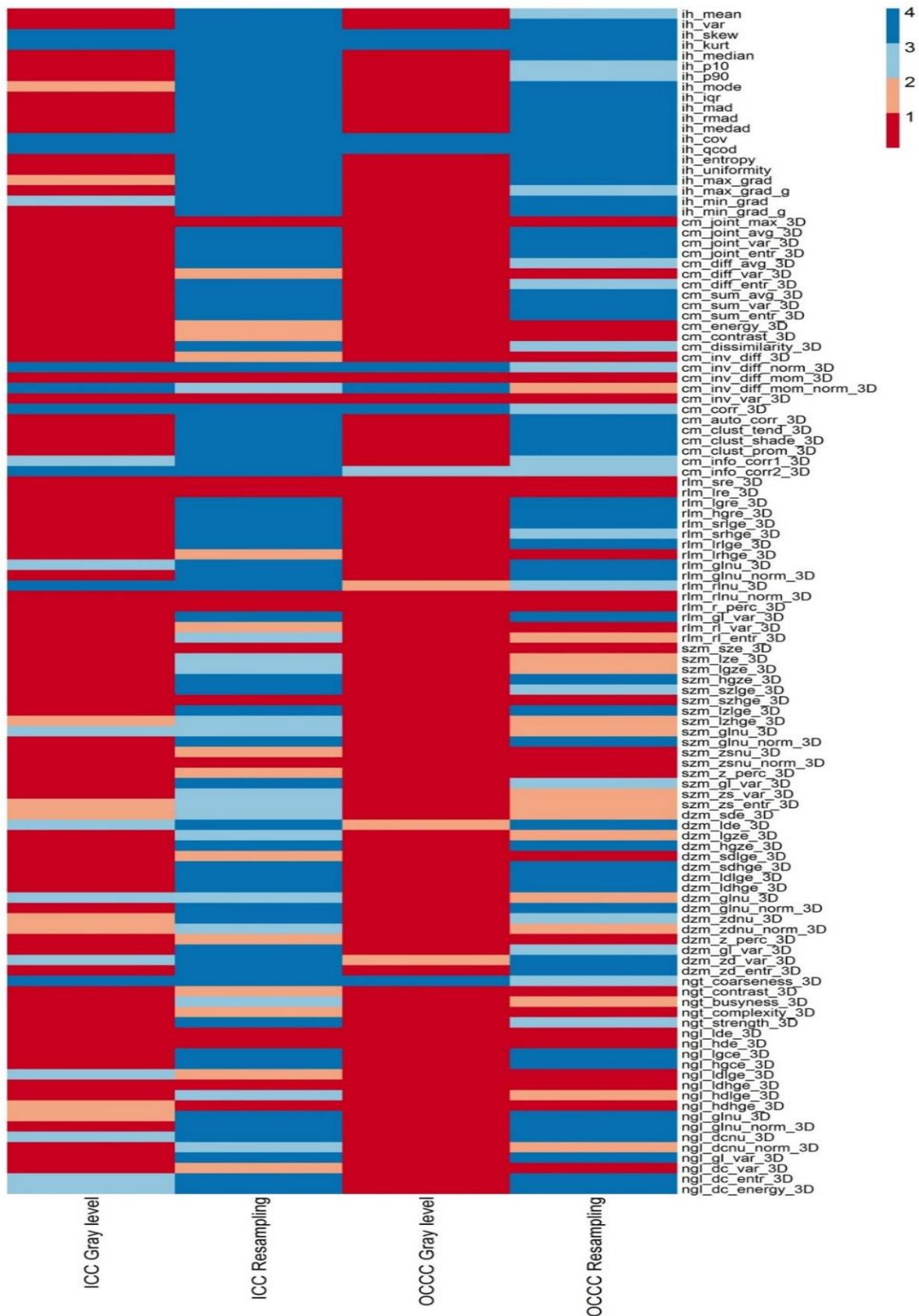

Figure 7. ICC and OCCC heat map of the percentage values of radiomics features categorized in 4 groups over different gray level and Interpolation algorithms. (Categorized 1 to 4: 1 = low and 4= highly robust)

4. Discussion

The consent of radiomics, considering other -omics, outlines robust labels for personalized medicine administrations. One of its possible utilization might be in tracking and predicting clinical outcomes for numerous treatment planning. Oikonomou *et al.* [25] recognized a significant relationship between ^{18}F–FDG PET radiomics features and lung cancer staging.

Although the application of radiomic features as a quantitative marker for diagnosis and prognosis, staging or predicting response to therapy is a growing utilization of FDG PET, examining the robustness, reliability, and reproducibility of such image biomarker within physical or biological factors have determined to be a measure of vast significance. Besides, numerous medical image factors cause distinctive hurdles while extracting and quantifying the tumor's FDG uptake data. The evaluation of reproducibility and repeatability for medical image radiomics features has earned growing attention [26]. Gathering proof recommends the value of taking such investigations into account. Researches have indicated that repeatable radiomic features must be applied for predictive treatment modeling [27]. This study aimed to evaluate PET radiomic features' repeatability in NSCLC patients against different gray levels and Interpolation algorithms. In terms of radiomics features extraction, several tools have been extended for feature extraction [28-30]. Our study was led using the SERA package compatible with the IBSI [31]. IBSI is an independent global collaboration working approaching the regularity of image biomarkers.

In a recent study, Larue *et al.* [33] showed that nearly all of the radiomics features vary in value while changing bin width for gray-level discretization, and in some features, a considerable or a minimal bin width ended in distinctive values in scanning. They showed that feature values fluctuate over various bin widths, but it could not prove that the selection of bin width dramatically influences the stability of radiomic features. The impact of GLD on the predictive ability of radiomic features has not been examined yet.

Our analysis showed that preprocessing, mainly gray level variability, can substantially impact the radiomics features. As we can see in Figures 3, 4, and 5 the concentration of ICC and OCCC values are located in less than 0.5, demonstrating the massive impact of gray level

variability on features. Our results showed ten features showing COV less than 10%, including four GLCM family features, three features of the Intensity histogram family, and four features showed $10 \leq COV \geq 20$. None of the NGTDM and GLSZM families illustrated reproducibility; therefore, all of them were admitted sensitive to gray-level discretization. As a comparison, Altazi *et al.* [14] showed 18% percent of the GLCM family is highly reproducible, and the same showed none of the NGTDM and GLSZM families are robust against GLD.

Shafiq-ul-Hassan *et al.* [32] confirmed that resampling decreased the variability of features from COV>70% to COV<30%. Consequently, we suggest regularly apply resampling before any radiomic study. Our data were resampled to an isotropic voxel size of $2 \times 2 \times 2mm^3$ which was suggested by IBSI using a nearest-neighbor, linear, and cubic interpolation, while in the Shafiq-ul-Hassan *et al.* [32] study, the voxel size of resampling was $1 \times 1 \times 2mm^3$ applying linear interpolation. In another study, Larue *et al.* [33] confirmed that linear interpolation followed by the smallest value of the features range about half of the features and cubic-interpolation for 30% of the features, while nearest-neighbor interpolation showed the most considerable extent, 61% of all. Hence, linear and/or cubic interpolations are preferred over nearest-neighbor interpolation for the $1 \times 1 \times 3mm^3$ voxels resampling.

In our study, the majority of the local texture features extracted from the images showed COV>20% against parameters. These texture features are categorized into separate families. Radiomics features that focus on low-intensity areas and small homogenous inside the tumor mass indicated great sensitivity to gray-level variation. As we observed from our result, forty-four features (25% of all) showed more than 100% COV over different gray level variability. These results can be observed in other studies [14]. The GLCM family showed more robust features against gray level variability with four robust features among texture features. Other texture features are highly influenced by gray-level variation. Furthermore, altogether, about 7% of texture features showed COV<10%. Similarly, 5 of 23 intensity volume histogram features showed COV<10% against GLD. This result was supposed due to; initially, they have great fluctuations because of their absence of measuring meaningful data of uptake heterogeneity inside the segmented lesions. Last but not least, it is due to the sensitive process applied to extract those features.

We calculated shape-based radiomics features to demonstrate the morphological features explaining the distribution of voxel-intensity of segmented lesions externally regarding spatial relations among neighboring voxel; therefore, all of the shape-based radiomics features illustrated insensitivity against GLD, our result are fully match with Altazi *et al.* [14].

On the contrary, Shafiq *et al.* [32] did not report the same results toward PET radiomics features. However, GLCM features indicated more reproducibility following the correction for gray-level and volume dependence.

Our results illustrated that the impact of different Interpolation algorithms on radiomics features is lesser than GLD. As we can see in Figures 3, 4, and 5, the concentrations of ICC and OCCC Interpolation algorithms s' values are located at more than 0.75. Also, our COV results demonstrated that 34% of radiomics features are robust against Interpolation algorithms. This percentage can be increased if we add the morphology, Statistics, and Intensity volume histogram family features to the number of robust features. The primary limitation of this study was the size of the data set. Future research should utilize more extensive data sets to increase the repeatability and reproducibility of radiomics features. The limitation of the current study is the limited number of patient data. However the outcomes and results of current research should be verified using more substantial and multicenter dataset.

5. Conclusion

This study investigated the reproducibility of numerous radiomic features extracted from 18F-FDG PET images of non-small cell lung cancer, adenocarcinoma against various parameters: various gray-level including 16, 32, 64, 128, and 256; besides, different Interpolation algorithms such as Linear, Cubic, and Nearest. Based on our results, most of the radiomic features in this study were extremely affected by GLD. The impact of IA on PET radiomics feature is much lesser than GLD but still is considerable. Hence, we recommend that careful examination of radiomic features' reproducibility is required before employing them in any clinical treatments.

References

1- C. Fitzmaurice *et al.*, "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study," *JAMA oncology*, vol. 3, no. 4, pp. 524-548, 2017.

2- R. Siegel, E. Ward, O. Brawley, and A. Jemal, "Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths." *CA: a cancer journal for clinicians*, vol. 61, no. 4, pp. 212-236, 2011.

3- I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. Tavares, "Automatic 3D pulmonary nodule detection in CT images: a survey." *Computer methods and programs in biomedicine*, vol. 124, pp. 91-107, 2016.

4- J. F. Palma, P. Das, and O. Liesenfeld, "Lung cancer screening: utility of molecular applications in conjunction with low-dose computed tomography guidelines." *Expert review of molecular diagnostics*, vol. 16, no. 4, pp. 435-447, 2016.

5- H. J. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, no. 1, pp. 1-9, 2014.

6- B. Ganeshan, E. Panayiotou, K. Burnand, S. Dizdarevic, and K. Miles, "Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival," *European radiology*, vol. 22, no. 4, pp. 796-802, 2012.

7- Y. Balagurunathan *et al.*, "Reproducibility and prognosis of quantitative features extracted from CT images." *Translational oncology*, vol. 7, no. 1, pp. 72-87, 2014.

8- G. J. Weiss *et al.*, "Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic." *PloS one*, vol. 9, no. 7, p. e100244, 2014.

9- O. Gevaert *et al.*, "Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results." *Radiology*, vol. 264, no. 2, pp. 387-396, 2012.

10- R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data." *Radiology*, vol. 278, no. 2, pp. 563-577, 2016.

11- P. Lambin *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," *Nature reviews Clinical oncology*, vol. 14, no. 12, pp. 749-762, 2017.

12- S. S. Yip and H. J. Aerts, "Applications and limitations of radiomics," *Physics in Medicine & Biology*, vol. 61, no. 13, p. R150, 2016.

13- J. E. Park, S. Y. Park, H. J. Kim, and H. S. Kim, "Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives," *Korean journal of radiology*, vol. 20, no. 7, p. 1124, 2019.

14- M. Shafiq-ul-Hassan *et al.*, "Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels," *Medical physics*, vol. 44, no. 3, pp. 1050-1062, 2017.

15- R. T. Larue *et al.*, "Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study," *Acta oncologica*, vol. 56, no. 11, pp. 1544-1553, 2017.

16- B. A. Altazi *et al.*, "Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms," *Journal of applied clinical medical physics*, vol. 18, no. 6, pp. 32-48, 2017.

17- I. Shiri, A. Rahmim, P. Ghaffarian, P. Geramifar, H. Abdollahi, and A. Bitarafan-Rajabi, "The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies," *European radiology*, vol. 27, no. 11, pp. 4498-4509, 2017.

18- A. Fedorov *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323-1341, 2012.

19- S. Ashrafinia, "Quantitative nuclear medicine imaging using advanced image reconstruction and radiomics," *Johns Hopkins University*, 2019.

20- A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative-feature definitions," *arXiv preprint arXiv*:1612.07003, 2016.

21- A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative." *arXiv preprint arXiv*: 161207003," 2016.

22- T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol. 15, no. 2, pp. 155-163, 2016.

23- K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological methods*, vol. 1, no. 1, p. 30, 1996.

24- J. J. Bartko, "The intraclass correlation coefficient as a measure of reliability," *Psychological reports*, vol. 19, no. 1, pp. 3-11, 1966.

25- P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychological bulletin*, vol. 86, no. 2, p. 420, 1979.

26- I. Lawrence and K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, pp. 255-268, 1989.

27- H. X. Barnhart, M. Haber, and J. Song, "Overall concordance correlation coefficient for evaluating agreement among multiple observers," *Biometrics*, vol. 58, no. 4, pp. 1020-1027, 2002.

28- A. Oikonomou *et al.*, "Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy," *Scientific reports*, vol. 8, no. 1, pp. 1-11, 2018.

29- A. Traverso, L. Wee, A. Dekker, and R. Gillies, "Repeatability and reproducibility of radiomic features: a systematic review," *International Journal of Radiation Oncology* Biology* Physics*, vol. 102, no. 4, pp. 1143-1158, 2018.

30- S. Gourtsoyianni *et al.*, "Primary rectal cancer: repeatability of global and local-regional MR imaging texture features," *Radiology*, vol. 284, no. 2, pp. 552-561, 2017.

31- C. Davatzikos *et al.*, "Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome," *Journal of medical imaging*, vol. 5, no. 1, p. 011018, 2018.

32- J. J. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104-e107, 2017.

33- C. Nioche *et al.*, "LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity," *Cancer research*, vol. 78, no. 16, pp. 4786-4789, 2018.

34- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*; 15(2): pp. 155–63, 2016.