Japanese Journal of Radiology
https://doi.org/10.1007/s11604-021-01100-0

TECHNICAL NOTE

Comparison of three freeware software packages for ^{18}F-FDG PET texture feature calculation

Michele Larobina[1] · **Rosario Megna**[1] · **Raffaele Solla**[1]

Received: 2 October 2020 / Accepted: 1 February 2021

Abstract

Purpose To compare texture feature estimates obtained from ^{18}F-FDG-PET images using three different software packages.
Methods PET images from 15 patients with head and neck cancer were processed with three different freeware software: CGITA, LIFEx, and Metavol. For each lesion, 38 texture features were extracted from each software package. To evaluate the statistical agreement among the features across packages a non-parametric Kruskal–Wallis test was used. Differences in the features between each couple of software were assessed using a subsequent Dunn test. Correlation between texture features was evaluated via the Spearman coefficient.
Results Twenty-three of 38 features showed a significant agreement across the three software ($P < 0.05$). The agreement was better between LIFEx vs. Metavol (36 of 38) and worse between CGITA and Metavol (24 of 38), and CGITA vs. LIFEx (23 of 38). All features resulted correlated ($\rho >= 0.70$, $P < 0.001$) in comparing LIFEx vs. Metavol. Seven of 38 features were found not in agreement and slightly or not correlated ($\rho < 0.70$, $P < 0.001$) in comparing CGITA vs. LIFEx, and CGITA vs. Metavol.
Conclusion Some texture discrepancies across software packages exist. Our findings reinforce the need to continue the standardization process, and to succeed in building a reference dataset to be used for comparisons.

Keywords Positron emission tomography · Oncology · Radiomics · Texture

Introduction

Radiomics refers to the set of image analysis techniques aimed at extracting quantitative information intending to find useful indications for the diagnosis and prognosis of tumor diseases. A relevant part of radiomics is represented by texture analysis. Texture refers to the calculation of features aimed to identify spatial patterns of variation and or repetition of grey tones on the entire image or within a region of interest [1, 2]. The texture is expected to give a measure of the lesion heterogeneity [3], which is likely to be related to the aggressiveness of the disease [4]. However, to date, the results appear to be confounding. It is not clear which indices carry useful information, which of them are stable, and which is the added value compared to conventional indices [5–9].

Texture analysis workflow starts with the definition of a region of interest (ROI) that delineate the tumor volume. Then, features extraction is performed by image-processing software. Next, a statistical analysis is performed, combining extracted features with clinical data for the selection of clinical-informative features and the formulation of a radiomics prediction model. Multiple factors can influence texture extraction, including image acquisition and reconstruction settings, tumor delineation, image pre-processing, and the software used for texture calculation [10]. The adoption of feature extraction software is a critical aspect, as, for the same image, different software packages can return different values for the same texture index. This is because there is a certain variability in the features' nomenclature, in the formulas underlying the names, and in their software implementation. This makes the results obtained by different research groups not comparable and affects the validity and reproducibility of the clinical indication of the features. The

✉ Rosario Megna
rosario.megna@cnr.it

Michele Larobina
michele.larobina@cnr.it

Raffaele Solla
raffaele.solla@cnr.it

[1] Institute of Biostructures and Bioimaging, National Research Council (CNR), Via Tommaso de Amicis, 95, 80145 Napoli, Italy

Published online: 17 February 2021

issue is known, and the image biomarker standardization initiative (IBSI) is an international collaboration which works towards standardizing the extraction of image biomarkers from acquired imaging for the purpose of high-throughput quantitative image analysis [11–13]. IBSI provides guidelines and reference values with which developers can verify their software and declare them IBSI-compliant.

Several research groups have been working on the texture extraction from Positron Emission Tomography (PET) images, and some of the developed software have been made available to the scientific community as freeware or open-source [14–19]. Despite the availability of these software tools, a limited number of studies on the comparison between software dedicated to radiomics have been reported in the literature. Only some of these works refer to PET imaging modality, being mostly addressed to Computed Tomography (CT) [20–22].

This article aims to test three freeware software for texture, to understand and measure the level of agreement when they are used with the same PET image dataset as input.

Materials and methods

Software packages

Three software were considered for comparison: CGITA version 1.4 (Chang-Gung Memorial Hospital, Taiwan), LIFEx version 6.3 (Inserm, CEA, CNRS, Université Paris Sud, France), and Metavol/Ptexture version 20181009/20180909 (Hokkaido University, Japan).

CGITA (https://sites.google.com/site/deanfanglab/software) is a freeware and open-source software developed in Matlab for quantifying tumor heterogeneity with molecular images [14].

LIFEx (http://www.lifex.org) is a freeware software developed in Java for radiomic feature calculation in multimodality imaging for characterization of tumour heterogeneity [15].

Metavol (https://www.metavol.org/) is a freeware software for metabolic tumor volume measurement in FDG PET/CT and Ptexture (https://github.com/metavol/ptexture) is an open-source add-on software for Metavol developed in Python to compute texture features from a segmented lesion volume [16]. In the following, we will use the term Metavol to indicate the two packages—Metavol and Ptexture—together.

The three considered software are packages developed for the processing of PET images, although CGITA and LIFEx, widely used in literature, can be used with other imaging modalities. LIFEx is the only software declaring compliance with the IBSI standard. The main characteristics of the three software packages are summarized in Table 1.

Patient data

We retrospectively analyzed imaging data of 15 patients (9 men and 6 women) with locally advanced head and neck cancer. The mean patient age was 60 ± 10 years. All patients underwent a whole-body ^{18}F-FDG PET/CT before the start of radiotherapy. Most cases (93%) were squamous cell carcinomas. The primary tumor site was larynx in 5 cases, oropharynx in 5 cases, oral cavity in 4 cases, and rhinopharynx in 1 case.

All procedures performed in this study involving human participants were in accordance with international ethical standards detailed in the 1964 Declaration of Helsinki and its later amendments, and according to the Italian Personal Data Protection Code for scientific research.

Image dataset

PET/CT studies were performed on a Discovery LS scanner (GE Healthcare) 45 min after the intravenous administration of 5 MBq/Kg of ^{18}F-FDG. Images were reconstructed using an iterative OSEM algorithm with CT-based attenuation correction. Each slice consists of 128×128 pixels with a pixel size of 3.90625×3.90625 mm^2 and a slice thickness of 4.25 mm. The authors used PET images in the form of anonymized DICOM files. For the segmentation and subsequent feature extraction, only patients' primary tumors were considered. Patient's lesions were segmented using the 40% of the SUV_{max} [23]. The median SUV_{max} was 16.9 (range 9.4—30.6). The median tumor volume was 10.8 mL (range 3.8–40.7 mL).

Digital phantom

To assess the origin of observed differences across packages, we developed a simple digital phantom. The phantom consists of *224* voxels with SUV decreasing from 5.0 (SUV_{max}) to 1.5 (SUV_{min}). The phantom is depicted in Fig. 1. It is composed of 8 slices symmetrically disposed respect to the *z*-axis; voxel dimensions are the same as the patient's PET studies. The *224* voxels are arranged as follows: 12 voxels have $SUV_{max} = 5.0$, 24 voxels have $SUV = 4.5$, 52 voxels have $SUV = 3.5$, and 136 voxels have $SUV = 1.5$. The outside of the phantom is filled with voxels having $SUV = 0$. Setting the threshold at 40% of SUV_{max}, a volume of *88* voxels is expected to be segmented.

Feature extraction

First, for each of the three software, patient's lesions were segmented using the 40% of the SUVmax. Second, for each

Table 1 Summary of the main characteristics of the three software considered in the study

Software package	Software license	ROI/VOI definition and feature calculation settings	N. features calculated	Operating system supported	Note
CGITA	Open-source	3D Region Growing with absolute SUV threshold, or Fuzzy C-mean Imports VOI in PMOD.voi or in DICOM-RT format Fixed bin number discretization: absolute (0-max) or relative (min–max in the VOI) No fixed bin width; no spatial resampling	72	Windows, Linux, and MacOS, with Matlab licence A compiled version for Windows is available	Source code almost not commented Lesion Volume only expressed in mL Last update 2014
LIFEx	Freeware	2D/3D ROI/VOI manually, semi-automated, or by SUV threshold (absolute, %SUV_{max}, Nestle) Imports RTSTRUCT object in DICOM or NIFTI-1 format Fixed bin number discretization: absolute (0-max) or relative (min–max or mean ± 3sd in the ROI/VOI) Fixed bin width discretization Spatial resampling	44	Windows, Linux, MacOS	Possibility to save the matrices used for texture calculation Lesion Volume expressed in mL and in number of voxels Last update 2020
METAVOL	Freeware	2D/3D ROI/VOI semi-automated, or by SUV threshold (relative or fixed) Fixed bin number discretization: absolute (0-max) or relative (min–max in the ROI/VOI) No fixed bin width; No spatial resampling	42	Windows Ptexture add-on is in Python language and runs in every platform	Lesion Volume only expressed in number of voxels Documentation not available Last update 2018

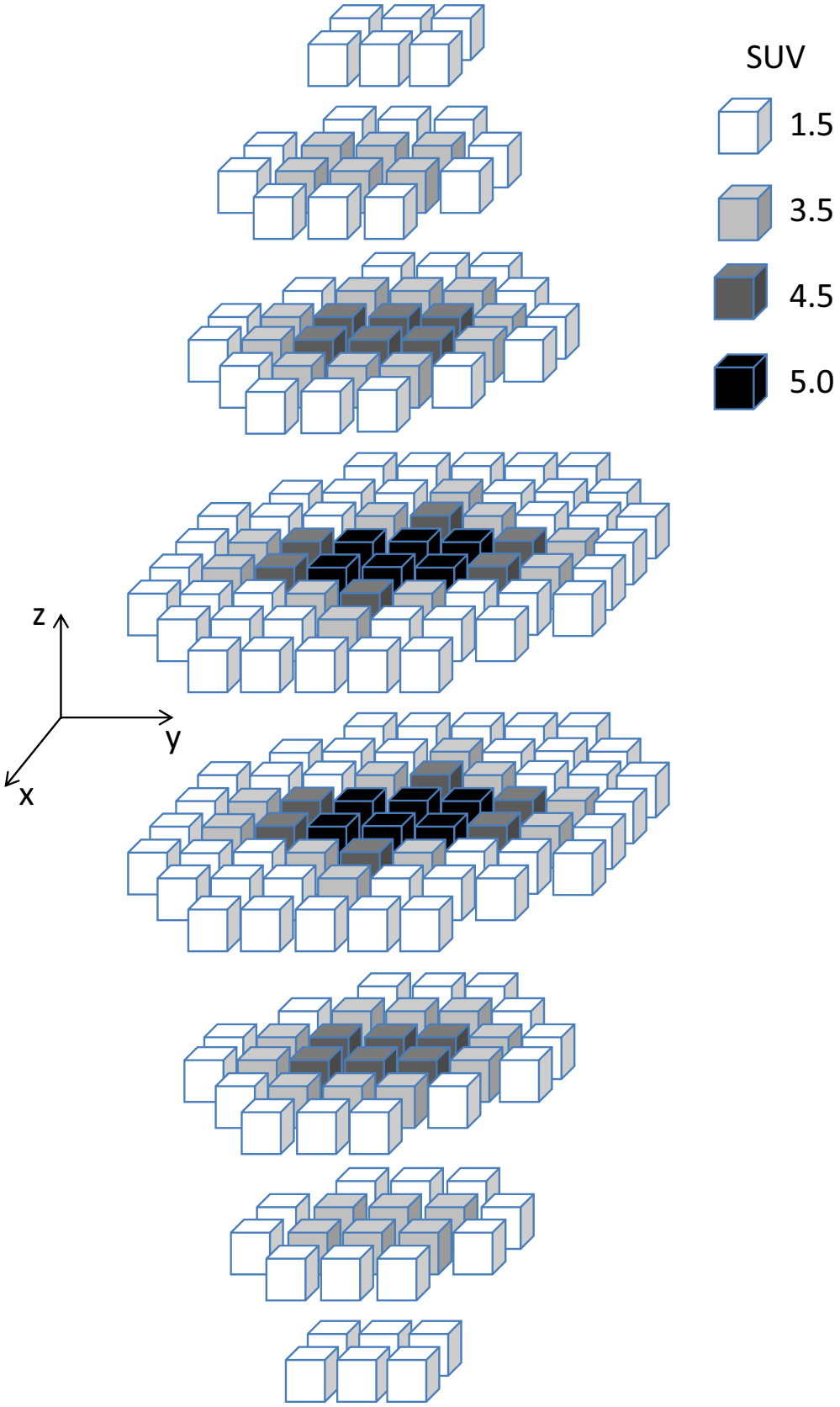
lesion, a total of 38 texture features representing the common group of features among all packages, were extracted. The features list comprise:

- four conventional features: Metabolic Tumor Volume, SUV_{max}, SUV_{mean}, and SUV_{std};
- three histogram-based (histo) features: $Skewness^{histo}$, $Kurtosis^{histo}$, $Entropy^{histo}$;
- six grey level co-occurrence matrix (cm) features: $Homogeneity^{cm}$, $Energy^{cm}$, $Correlation^{cm}$, $Contrast^{cm}$, $Entropy^{cm}$, $Dissimilarity^{cm}$;
- eleven grey level run-length matrix (rlm) features: SRE^{rlm}, LRE^{rlm}, $LGRE^{rlm}$, $HGRE^{rlm}$, $SRLGE^{rlm}$, $SRHGE^{rlm}$, $LRLGE^{rlm}$, $LRHGE^{rlm}$, $GLNU^{rlm}$, $RLNU^{rlm}$, RP^{rlm};
- eleven grey level size-zone matrix (szm) features: SZE^{szm}, LZE^{szm}, $LGZE^{szm}$, $HGZE^{szm}$, $SZLGE^{szm}$, $SZHGE^{szm}$, $LZLGE^{szm}$, $LZHGE^{szm}$, $GLNU^{szm}$, $SZNU^{szm}$, ZP^{szm};
- three neighbourhood grey level difference matrix (ndm) features: $coarseness^{ndm}$, $Contrast^{ndm}$, $Busyness^{ndm}$.

Features were collected using the same setting for all the three software: fixed bin number discretization with 32 grey levels between the minimum and maximum in the lesion volume (relative intensity rescaling); no spatial resampling. Second- and higher-order feature calculations (cm, rlm, szm, ndm) were performed in 3D with an average over 13 directions and a distance set to one voxel. The three groups of texture features originated by the three packages were analyzed and compared.

Statistical analysis

A graphical comparison of range and median values of the 38 features, calculated by the three packages, were reported as boxplots. To evaluate the agreement among features across packages a non-parametric Kruskal–Wallis test was


SUV
1.5
3.5
4.5
5.0
z
y
x

◂**Fig. 1** Sketch of the simple phantom used in the study. The phantom consists of 224 voxels with SUV values from 1.5 (SUV_{min}) to 5.0 (SUV_{max}), placed in 8 slices. In the two central slices, there is a core of 6×2 voxels with SUV = SUV_{max}. This core is surrounded from a first shell of 6×4 voxels with SUV = 4.5. A second shell of 10×4 + 6×2 voxels with SUV = 3.5 surrounds this first shell. Lastly, the third shell of 136 voxels with SUV = 1.5 surrounds the second shell

used. Differences in features between each couple of software packages were assessed using a subsequent Dunn test. To verify whether the characteristics that were not in agreement according to Dunn test were at least correlated, we calculated the Spearman coefficient (ρ), defining correlated those features with $\rho \geq 0.70$ in absolute value. The correlation among features was also evaluated intra-software, as a tool for dimensionality reduction. The Bonferroni correction for multiple comparisons and a significance level of 0.05 were used for all the tests. We computed the Principal Component Analysis (PCA) and plotted the first principal component against the second principal component to have a bi-dimensional representation of the data preserving most of the sample variance. To visualize the weight of the features relative to the two main components we plotted the PCA squared cosines for each of the software packages. Statistical analysis was performed with R software, version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Image dataset

Overall, 23 features of 38 showed a good agreement across the three software. Among these, we found the conventional indices Metabolic Tumor Volume, SUV_{max}, SUV_{mean}, and SUV_{std}. Feature's boxplots are reported in Fig. 2. The features resulted significantly different among the three software via non-parametric Kruskal–Wallis test are listed in Table 2. Agreement percentages between packages, as assessed by post hoc Dunn test, were: 95% (36 of 38 features) for LIFEx vs. Metavol, 63% (24 of 38 features) for CGITA vs. Metavol, and 60% (23 of 38 features) for CGITA vs. LIFEx. Correlation analysis highlighted that all features resulted correlated ($\rho \geq 0.70$, $P < 0.001$) in comparing LIFEx vs. Metavol, while 7 of 38 features (18%) were found not in agreement and not correlated ($\rho < 0.70$, $P < 0.001$) in comparing CGITA vs. LIFEx, and CGITA vs. Metavol. Spearman's correlation coefficients between the features of each couple of software packages are reported in Table 3. The PCA bi-dimensional plot with the position of each subject is reported in Fig. 3. The first two principal components explain more than 58% of the sample variance. Plots of squared cosine computed for each of the three software packages are reported in Fig. 4. The features correlation matrix for each of the three software obtained using the hierarchical clustering method are shown in Fig. 5. PCA plots and correlation matrices confirm the substantial agreement between LIFEx and Metavol features, and the discordance of CGITA features value respect to the other two packages.

Digital phantom

Using the 40% SUV_{max} threshold, LIFEx and Metavol provided a volume of 88 voxels, while CGITA underestimated the volume segmenting 84 voxels. This last package excludes from the segmentation the four voxels in the central slices with SUV = 3.5 attached to the segmented volume only for one face.

Discussion

Texture is a promising image-processing technique in providing biomarkers to support clinical decision making in cancer, but its added value has not yet been clearly demonstrated. An interference factor in the comprehension of the results is represented by the variability in the texture features extracted by different software. On comparing, software users may encounter some difficulties, as: (a) different software can adopt different features names; (b) features with the same name do not always have the same calculation formula; (c) it is not always possible to access the source code to verify the implemented formulas, as for freeware the code used for feature calculation is not available; and (d) software documentation often is incomplete. In this regard, the IBSI collaboration has been working to standardize the nomenclature and the mathematical formulas of a large number of radiomics features [11–13].

Our study was an analysis of the variability in the values of the features extracted from PET images by three radiomics software. Features were selected and matched based on the name assigned to them by each software package. The analysis was performed on 38 features representing the common group of features across the three software. In some cases, the name of the features did not match precisely across packages, but there was clear evidence of their correspondence. In particular, the Energy feature derived by the Grey Level Co-occurrence matrix was called Second Angular Moment in the CGITA software; the eleven features derived by the Grey Level Run-Length matrix were called Voxel-alignment matrix derived in the CGITA software as reported into the publication that describes the package [14]. An overview of the differences in the feature's value calculated by the three software is provided by features boxplot. Kruskal–Wallis and Dunn tests highlighted a significant difference between 15 of 38 features of CGITA respect to LIFEx and 14 of 38

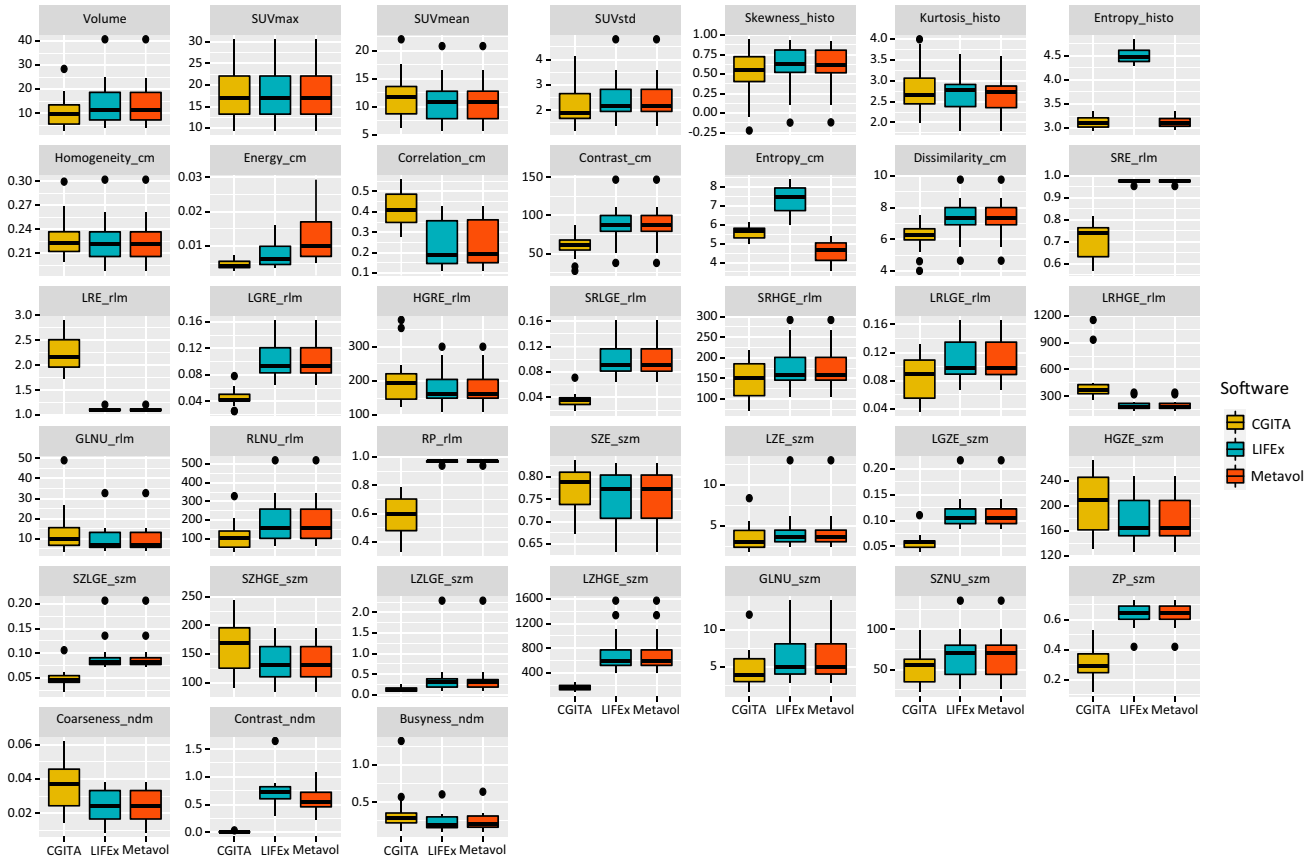


Fig. 2 Boxplot of the 38 common features obtained with the three software packages

Table 2 *P* values of the features statistically significant at Kruskal–Wallis test, and post hoc Dunn test

Features	Kruskal–Wallis	Dunn		
		CGITA–LIFEx	CGITA–Metavol	LIFEx–Metavol
$Entropy^{histo}$	$1.6*10^{-5}$	$9.0*10^{-6}$	ns	$7.3*10^{-6}$
$Energy^{cm}$	$6.3*10^{-4}$	$3.0*10^{-2}$	$8.4*10^{-6}$	ns
$Correlation^{cm}$	$2.5*10^{-3}$	$2.7*10^{-4}$	$7.3*10^{-4}$	ns
$Entropy^{cm}$	$2.0*10^{-7}$	$5.3*10^{-3}$	$7.0*10^{-3}$	$2.0*10^{-9}$
$LGZE^{szm}$	$3.4*10^{-5}$	$1.6*10^{-5}$	$1.3*10^{-5}$	ns
$SZLGE^{szm}$	$3.4*10^{-5}$	$2.1*10^{-5}$	$1.1*10^{-5}$	ns
$LZHGE^{szm}$	$1.6*10^{-5}$	$9.0*10^{-6}$	$7.3*10^{-6}$	ns
ZP^{szm}	$1.6*10^{-5}$	$8.6*10^{-6}$	$7.5*10^{-6}$	ns
SRE^{rlm}	$1.6*10^{-5}$	$7.6*10^{-6}$	$8.7*10^{-5}$	ns
LRE^{rlm}	$1.6*10^{-5}$	$8.1*10^{-6}$	$8.1*10^{-6}$	ns
$LGRE^{rlm}$	$2.7*10^{-5}$	$9.6*10^{-6}$	$1.5*10^{-5}$	ns
$SRLGE^{rlm}$	$1.6*10^{-5}$	$7.3*10^{-6}$	$9.0*10^{-6}$	ns
$LRHGE^{rlm}$	$4.4*10^{-5}$	$1.8*10^{-5}$	$1.8*10^{-5}$	ns
RP^{rlm}	$1.6*10^{-5}$	$7.0*10^{-6}$	$9.3*10^{-6}$	ns
$Contrast^{ndm}$	$9.4*10^{-6}$	$5.6*10^{-7}$	$9.1*10^{-5}$	ns

features of CGITA vs. Metavol (Table 2). The same statistical test showed substantial overlap between 36 of 38 features calculated by LIFEx and Metavol. The only two features significantly different were: $Entropy^{histo}$ and $Entropy^{cm}$. We verified that the two $Entropy^{histo}$ values were correlated and their ratio was constant and equal to $\log_2(e)$. Therefore, the

Table 3 Coefficients of correlation between the features of each couple of software packages

Features	Spearman correlation coefficient (ρ)		
	CGITA–LIFEx	CGITA–Metavol	LIFEx–Metavol
$Entropy^{histo}$	0.90	0.90	1.00
$Energy^{cm}$	0.75	0.71	0.99
$Correlation^{cm}$	0.94	0.95	1.00
$Entropy^{cm}$	0.74	0.71	1.00
$LGZE^{szm}$	0.63	0.63	1.00
$SZLGE^{szm}$	0.74	0.73	1.00
$LZHGE^{szm}$	0.03	0.03	1.00
ZP^{szm}	0.67	0.67	1.00
SRE^{rlm}	0.83	0.83	1.00
LRE^{rlm}	0.83	0.83	1.00
$LGRE^{rlm}$	0.49	0.50	1.00
$SRLGE^{rlm}$	0.55	0.57	0.99
$SRHGE^{rlm}$	0.91	0.91	1.00
RP^{rlm}	−0.15	−0.15	1.00

difference was only due to the change in the base of the logarithm ($\log_2$ for LIFEx and $\log_e$ for Metavol). In the case of $Entropy^{cm}$, the two feature values resulted correlated, but their ratio was not constant, so the change in the logarithm base alone cannot explain the observed difference. It is clear that for this feature there is a difference in the formula implemented by the two software. No significant difference between conventional indices, Metabolic Tumor Volume, SUV_{max}, SUV_{mean}, and SUV_{std} was found across the three software. Regarding the metabolic tumor volume, we found a systematic underestimation in the CGITA segmentation of the fifteen patient's lesion in comparison to the other two software: median of 10.77 mL for CGITA vs. 11.54 mL for both LIFEx and Metavol. Therefore, the median underestimation of CGITA compared to other software was about 7%.

To check the 40% SUV_{max} segmentation output, we developed a simple phantom that confirmed us that CGITA returns a volume underestimation probably because of different region growing rules. Due to the different VOI formats required by the three software packages, it was not possible to use the same lesion VOI for comparative analysis.

Consequently, we decided to vary the threshold for CGITA by plus or minus 5% of the SUV_{max} to assess how small changes in the VOI affect the feature calculations. With the threshold fixed to 35% of the SUVmax, $Contrast^{cm}$ and $Dissimilarity^{cm}$ changed significantly with respect to the 40% SUVmax comparisons. This variation was observed for both CGITA vs. LIFEx and CGITA vs. Metavol comparisons. With the threshold fixed to 45% of the SUV_{max}, only $LZLGE^{szm}$ changed significantly with respect to the 40% SUVmax comparisons. Also in this case, the variation concerned both CGITA vs. LIFEx and CGITA vs. Metavol comparisons. None of the significant features to the Dunn test reported in Table 2 changed significantly varying the segmentation threshold of CGITA. Therefore, an increase of the ROI volume segmented by CGITA as to balance its volume underestimation did not explain the difference in the textural feature calculations observed between the software.

Correlation analysis was performed to establish if the significant features to the Dunn test (i.e., that generate different distributions by two software) were correlated or not. This analysis adds further information on if and how the texture features extracted by different software are related. Features that are not in agreement but are correlated may describe the same image characteristics. The correlation analysis highlighted that among the 15 significant features to the Dunn test in comparing CGITA vs. LIFEx, eight were correlated and seven uncorrelated. Also in comparing CGITA vs. Metavol, we observed the same correlation results. Lastly, $Entropy^{histo}$ and $Entropy^{cm}$, that were significant to the Dunn test in comparing LIFEx vs. Metavol, resulted correlated.

In an attempt to efficiently visualize the features space a PCA analysis was performed. Plotting the first principal component against the second principal component we found that the 95% confidence region related to LIFEx and Metavol are largely superimposed (Fig. 3). At the same time, the overlap with CGITA is marginal. The PCA squared cosine plots calculated for each software package also indicate an overlap between the projections of the LIFEx and Metavol features on the two principal components and a substantial difference in CGITA compared to them. These plots could also be used to identify clusters of features from which to select a reduced number of representative features [24]. Similarly, the feature reduction can be made by analyzing the intra-software correlation matrix with hierarchical clustering order, shown in Fig. 5. To obtain a dimensionality reduction for a software package, users can choose one or a few features from each cluster. Note that the feature clusters obtained from CGITA are dissimilar with respect to those produced from LIFEx and Metavol, confirming what was previously found.

Despite the limitations of a small number of PET studies and an analysis confined to the variability of the extracted radiomic features, we can understand how different software can lead to a different selection of informative features. The effect of these variations on the clinical applications has only recently been investigated [25–27]. Our work highlights the importance of comparing software and, at the same time, reinforces the importance of continuing the standardization process to obtain reproducible and comparable analyses.

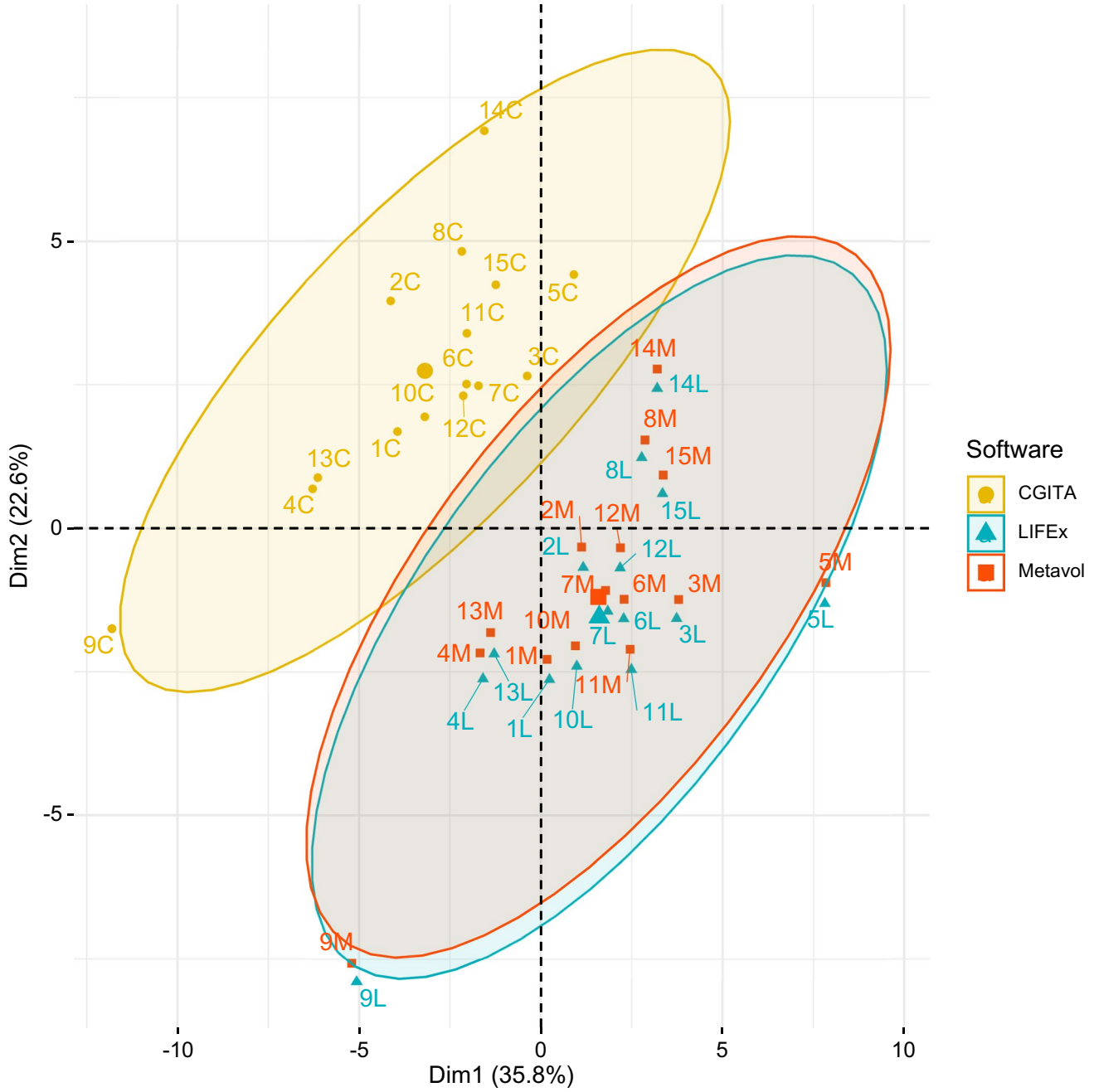


Fig. 3 PCA biplot of the 95% confidence regions (ellipses) obtained with LIFEx, Metavol, and CGITA. Patients are marked with a label indicating the subject (prefix number) and the software package (suffix: L for LIFEx, M for Metavol, and C for CGITA)

Conclusion

Three freeware texture analysis software were compared using as input ^{18}F-FDG PET images of fifteen patients with head and neck cancer. The analysis of the 38 features extracted revealed a significant discrepancy of CGITA compared to Metavol and LIFEx, while this last two software resulted substantially interchangeable. Our findings reinforce the need to continue the standardization process, and to succeed in building a phantom to be used as reference data for comparisons.

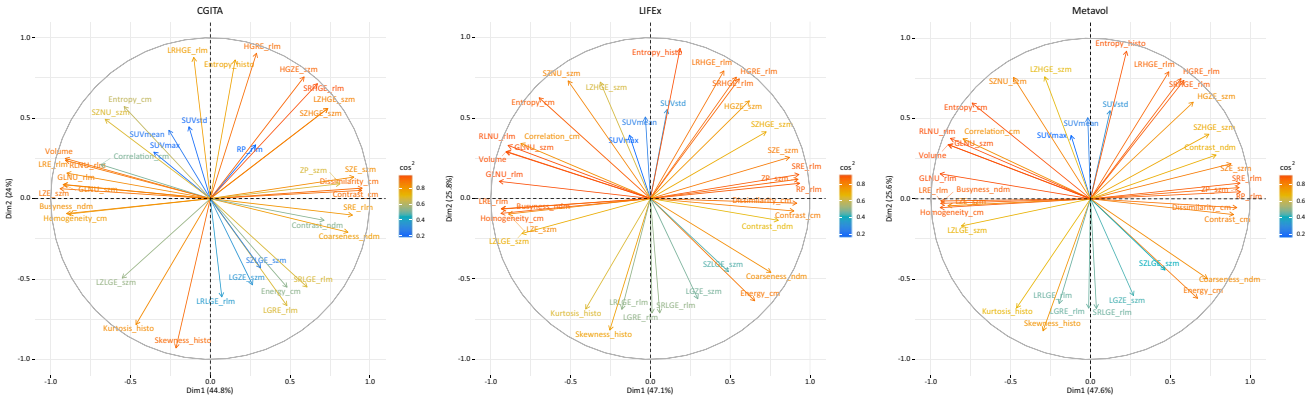


Fig. 4 PCA $\cos^2$ plots calculated for the three software packages. Color palette from blue to red indicates a greater contribution of the features to the principal components

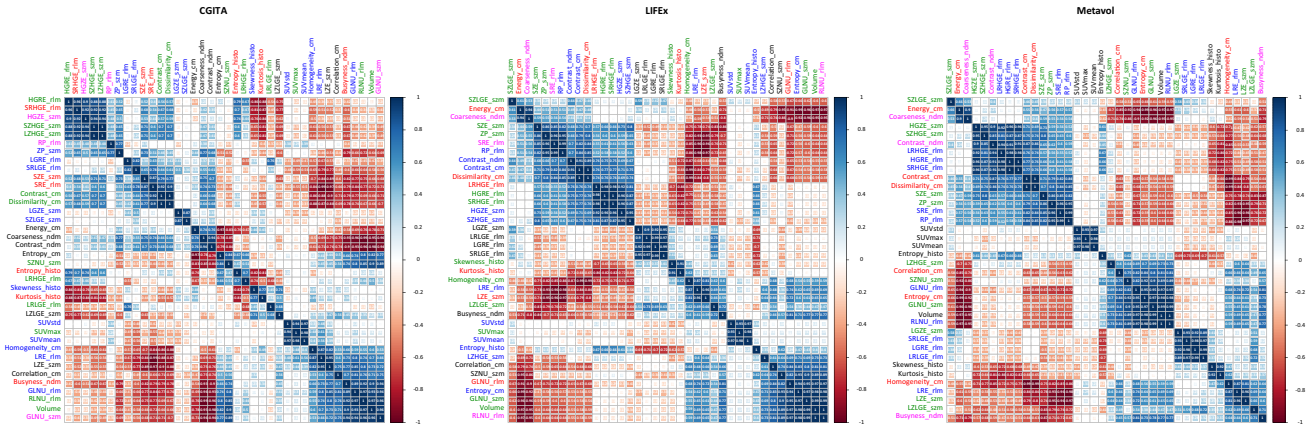


Fig. 5 Correlation matrices for each of the three software obtained using the hierarchical clustering method

Author contributions ML and RM conceptualized the paper; ML and RM evaluated and reported the imaging findings; RM carried out the statistical analysis; RS provided medical advice; ML, RM, and RS drafted the manuscript; all the authors revised the paper and approved its final version.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards All the procedures performed in this study involving human participants were in accordance with international ethical standards detailed in the 1964 Declaration of Helsinki and its later amendments, and according to the Italian Personal Data Protection Code for scientific research.

Informed consent Nothing to declare.

References

1. Gonzalez RC, Woods RE. Digital image processing. Truskey: Third Edition - Pearson Prentice Hall; 2008. p. 827–56.
2. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;3(5):4006.
3. Heppner GH. Tumor heterogeneity. Cancer Res. 1984;44(6):2259–65.
4. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81–94.
5. Kitajima K, Suenaga Y, Sugimura K. Present and future role of FDG-PET/CT imaging in the management of head and neck carcinoma. Jpn J Radiol. 2015;33(12):776–89.
6. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in pet: where do we stand. J Nucl Med. 2015;56(11):1642–4.
7. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21(2):249–57.
8. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE. Quantitative radiomics: impact of stochastic effects on textural

feature analysis implies the need for standards. J Med Imaging (Bellingham). 2015;2(4):041002.
9. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future. Eur J Nucl Med Mol Imaging. 2017;44(1):151–65.
10. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. Insights Imaging. 2020;11(1):91.
11. Image biomarker standardization initiative reference manual. https://arxiv.org/pdf/1612.07003.pdf. Accessed 2 Oct 2020.
12. 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(13):2638–55.
13. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295(2):328–38.
14. Fang YH, Lin CY, Shih MJ, Wang HM, Ho TY, Liao CT, et al. Development and evaluation of an open-source software package "CGITA" for quantifying tumor heterogeneity with molecular images. Biomed Res Int. 2014;2014:248505.
15. Nioche C, Orlhac F, Boughdad S, Reuzé 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(16):4786–9.
16. Hirata K, Kobayashi K, Wong KP, Manabe O, Surmak A, Tamaki N, et al. A semi-automated technique determining the liver standardized uptake value reference for tumor delineation in FDG PET-CT. PLoS ONE. 2014;9(8):e105682.
17. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77(21):e104–7.
18. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. Med Phys. 2015;42(3):1341–53.
19. Apte AP, Iyer A, Crispin-Ortuzar M, Pandya R, van Dijk LV, Spezi E, et al. Technical note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research. Med Phys. 2015. https://doi.org/10.1002/mp.13046.
20. Foy JJ, Robinson KR, Li H, Giger ML, Al-Hallaq H, Armato SG. Variation in algorithm implementation across radiomics software. J Med Imaging (Bellingham). 2018;5(4):044505.
21. Liang ZG, Tan HQ, Zhang F, Rui Tan LK, Lin L, Lenkowicz J, et al. Comparison of radiomics tools for image analyses and clinical prediction in nasopharyngeal carcinoma. Br J Radiol. 2019;92(1102):20190271.
22. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, Riesterer O, Pruschy M, Studer G, et al. Post-radiochemotherapy PET radiomics in head and neck cancer—The influence of radiomics implementation on the reproducibility of local control tumor models. Radiother Oncol. 2017;125(3):385–91.
23. Im HJ, Bradshaw T, Solaiyappan M, Cho SY. Current methods to define metabolic tumor volume in positron emission tomography: which one is better? Nucl Med Mol Imaging. 2018;52(1):5–15.
24. Song F, Guo Z, Mei D. Feature selection using principal component analysis. International conference on system science, engineering design and manufacturing informatization, Yichang, 2010. p. 27–30.
25. Lu L, Sun SH, Yang H, Guo P, Schwartz LH, et al. Radiomics prediction of EGFR status in lung cancer-our experience in using multiple feature extractors and the cancer imaging archive data. Tomography. 2020;6(2):223–30.
26. Fornacon-Wood I, Mistry H, Ackermann CJ, Blackhall F, McPartlin A, Faivre-Finn C, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. Eur Radiol. 2020;30(11):6241–50.
27. Foy JJ, Armato SG, Al-Hallaq H. Effects of variability in radiomics software packages on classifying patients with radiation pneumonitis. J Med Imaging. 2020;7(1):014504.